Averaged Optimization and Finite-Time Thermodynamics

Anatoly Tsirlin $^{1,*}$ and Ivan Sukin $^{2,*}$

Ailamazyan Program Systems Institute of Russian Academy of Sciences, Petra Pervogo st., 4a, Veskovo, Yaroslavl oblast 152021, Russia

* Correspondence: tsirlin@sarc.botik.ru (A.T.); ivsukin@gmail.com (I.S.); Tel.: +7-915-973-44-42 (I.S.)

Received: 1 July 2020; Accepted: 18 August 2020; Published: 20 August 2020

Abstract: The paper considers typical extremum problems that contain mean values of control variables or some functions of these variables. Relationships between such problems and cyclic modes of dynamical systems are explained and optimality conditions for these modes are found. The paper shows how these problems are linked to the field of finite-time thermodynamics.

Keywords: averaged; optimization; thermodynamics; heat transfer; cyclic mode

1. Introduction

Averaging plays a very important role in optimization problems applied to engineering. Here are a few examples:

1. Assume that there is a bound on the source flow $g_s$ and this constraint lowers the possible optimal value of the production. If we introduce some buffer (container) in such a way that the source flow is its feed, we can raise the possible value of the feed flow to the actual process without violating this constraint. Actual values of the feed flow will oscillate between values that are greater than and less than $g_s$. Only the mean value of the feed flow will be bounded in this case. Using this approach we can replace the strict constraint on $g_s$ by the averaged one. If we use such a buffer to store the product flow, we can maximize not this flow itself, but its mean value (Figure 1).

![Figure 1. Flowsheet of a simple process with averaging of both source and product flows.](image-url)

Let us assume that the relationship between production rate $g$ and consumption $q$ has the form presented at Figure 2.
With the help of buffers this relationship could be improved on interval from 0 to \( q_1 \). The process must operate with consumption \( q_1 \) during some fraction of time and with zero consumption during the remaining time. Relationship between average production and average consumption is represented by dashed slope line at Figure 2. This is the way pumps of water towers operate.

2. Controls often can have only discrete values. For example, the light switch can be either on or off. None of these discrete values satisfy the constraints of the original problem. If there are devices that smooth out any oscillations of control variables, the optimal mode can correspond to the switching strategy that maintains given average values of flows. This kind of switching is the basis of electronic light dampers.

3. In a heat engine the working fluid periodically makes contacts with the hot and cold sources, and the properties of these contacts must be chosen such that the average properties of the working fluid satisfy the constraints of the optimum cycle problem.

Averaged problems arise in finite-time thermodynamics for two main reasons:

1. Many processes are periodic and their constraints must be satisfied on average per cycle.
2. Interactions of thermodynamic systems are characterized by values of extensive variables \( X \) (volume, amount of substance, internal energy, entropy), and flows of mass and energy emerging in these interactions depend on intensive variables \( y \) (temperature, pressure, molar fraction). The rate of change of extensive variables depend on a flow, and of course on \( y \). This means that the governing equations for thermodynamic interactions have the form:

\[
\frac{dX}{dt} = F(y)
\]  

(1)

The right hand side of (1) does not contain \( X \) and this means that the increase in extensive variables during some given amount of time depends only on the mean value of \( F \). It does not depend on the order in which intensive variables have different values, if the mean value of \( F \) remains constant. Equations such as (1) are called Lyapunov-type. They allow us to formulate the problem of optimal control for thermodynamic systems in averaged form.

Below we will consider some of these problems. The last section contains applications of methods developed in the paper to finite-time thermodynamics.

We consider dynamical systems characterized by a finite number of variables. By a steady-state mode of a system we mean a mode such that, for every variable \( y_\nu(t) \) characterizing the system, there exists a period \( T_\nu \) such that the average value of \( y_\nu(t) \) over this period is constant in time. Formally, this can be written as [1]

\[
\frac{1}{T_\nu} \int_{t-T_\nu}^{t} y_\nu(\tau)d\tau = \overline{y}_\nu
\]  

(2)

Clearly, static modes, under which \( y_\nu(t) \) are constant for all \( \nu \), satisfy this definition.
Another, more general, subclass of steady-state modes is formed by modes for which there exists a period $T$ that is a multiple of all periods $T_\nu$. Such modes are called cyclic ones.

There are also steady modes for which there does not exist a common period $T$ for all variables $y_\nu(t)$. This corresponds to the case where the ratio of at least two periods $T_\nu$ and $T_\mu$ is irrational. Such modes are called quasi-cyclic steady-state modes.

If the system is affected by external factors represented by stationary random processes and the mean values of the variables characterizing the system tend to some limits as the period $T$ of averaging increases, then the steady mode is said to be stochastic.

A switch to a non-static steady-state mode may be caused either by the absence of a static mode admitted by the operating conditions of the system or by the fact that the efficiency of the system in a static mode is lower than in other modes.

Consider some examples.

1. The human organism in a steady mode in the absence of external perturbations is characterized by a constant temperature, a constant composition of arterial blood, etc. However, some factors such as the blood pressure and the lung volume periodically change. This is related to the “structure” of the respiration and circulation organs.

2. A system consisting of a pump connected with a tank (e.g., a water tower) and consumers operates so that, even if the liquid consumption $G$ is constant, the pump is sometimes completely switched off (and the liquid does not flow into the tank) and is sometimes switched on and operates with delivery higher than $G$, with the average delivery being $G$. If the dependence of the pump delivery $g$ on the power expenditure $S$ is described by a strictly convex function, then the average pump delivery is higher than that in the static mode at the same average power expenditure.

In the rest of the paper, we mainly consider cyclic steady-state modes, among which two limit classes are distinguished. The first class includes modes in which each of the periods $T_\nu$ significantly exceeds the time of relaxation processes in the system. Moreover, each of the static steady states is assumed to be stable. In this case, we can neglect the dynamics of the system and assume that under variation of the mode variables, the state variables change in accordance with the static characteristics. Such modes are said to be quasi-static.

The second class is formed by sliding steady-state modes, in which all or some of the control variables vary with frequency so high that, due to the inertia of the object, the state variables remain virtually constant, and their values depend only on the averaged influence of the control variables.

Although static modes are a special case of cyclic modes, below by cyclic modes we mean modes under which at least one variable of the process changes periodically in time. A cyclic mode is said to be efficient if the passage to this mode improves the efficiency of the process in comparison with the static mode.

3. Cyclic modes are typical of systems with no admissible static modes. Often a system has no static modes if the set $V$ of admissible values of variables is non-convex; e.g., this set may include only discrete values. This is so, for example, in a heat engine in which a working fluid contacts a heat source whose temperature can take only two values, $T_+$ (a hot source) and $T_-$ (a cold source), and the average power over a cycle is required to be maximal under certain constraints.

Cyclic processes may be organized not only in time; variables may also depend on a spatial coordinate. In this case, the parameters of the system are constant in each section of the apparatus and vary periodically from section to section.

When passing from a static mode to a cyclic mode, one needs to replace the objective function by its mean value over the cycle and to replace all or some constraints imposed at each moment of time by averaged constraints. Thus, this passage involves an operation of averaging. Before passing to a cyclic mode, one must answer the following questions:
1. Does there exist a cyclic mode satisfying the constraints of the problem?
2. Is the transition from the optimal static mode to the cyclic mode efficient?
3. What is the gain in the optimality criterion from this passage?
4. What are the optimal forms of variation of the control and state variables, optimality conditions, computational algorithms?

It is desirable to answer questions 1—3 without solving problem 4, which is rather difficult in most cases.

Usually the problem of choosing an optimal static mode of an apparatus reduces to the problem of finding the extremal value of the objective function under certain equality and inequality constraints on the variables, i.e., to a non-linear programming problem. The transition to cyclic modes extends the set of possible solutions and, depending on a particular setting, leads either to an averaged non-linear programming problem or to a variational control problem. If the problem has an optimal static mode, then we refer to this problem as the initial problem.

In the rest of the paper we consider various methods for constructing problems with larger sets of admissible solutions as compared with the initial problem; we show that there are relationships between such extended problems, which allows one to estimate solutions and values of some of them by solving others.

2. Averaged Optimization Problems and Their Optimality Conditions

In this section, we consider various methods for introducing averaging into a non-linear programming (NLP) problem and obtain optimality conditions for averaged problems. To obtain these conditions, we use a trick based on reducing any averaged problem to a canonical form and deriving necessary optimality conditions for a particular problem from those for a general problem.

2.1. Averaging of Functions Included in the Formulation of an Optimization Problem

Consider an initial NLP problem \([2]\) in the form

\[
f_0(x) \to \max / f_i(x) = 0, \quad i = 1, \ldots, m, \quad x \in V_x. \tag{3}
\]

On the set \(V_x\), we define a probability measure \(p(x)\) such that

\[
\int_{V_x} p(x)dx = 1, \quad p(x) \geq 0. \tag{4}
\]

The average value of the function \(f(x)\) on the interval \([0, \tau]\) can be calculated as follows:

\[
\bar{f}(x) = \frac{1}{\tau} \int_0^\tau f(x(t))dt = \int_{V_x} f(x)p(x)dx. \tag{5}
\]

Let us assume that \(x\) varies with time or one solves the problem \((3)\) and maximizes the mean value of \(f_0\), but not the value of this function itself. If functions \(f_i\) vanish on average, then we will arrive at a problem of the form

\[
f_0(x) \to \max / f_i(x) = 0, \quad i = 1, \ldots, m. \tag{6}
\]

A sought solution of problem \((6)\) is a measure \(p(x)\) on \(V_x\) rather than a vector \(x\). The variable \(x\) is called a randomized one, and \(p(x)\) is called a generalized solution. Following A.D. Ioffe and V.M. Tikhomirov \([3]\), we call the value of the objective functional at the optimal solution as the value of a problem.

2.2. Convex Hulls—Carathéodory’s Theorem

The notion of *convexity* is very important for optimization problems.
1. The convex hull of a set $V$ is the minimum convex set $CoV$ such that $V \subset CoV$.

2. The set of points lying on or below the graph of a function is called its hypograph. The convex hull $Co f$ of a function $f$ is the upper boundary of the convex hull of its hypograph.

3. Alternatively, the convex hull of a function $f$ is the minimum convex function defined on the convex hull of the domain of $f$. For every $\hat{x}$ from the domain of $f$ the following holds: $Co f(\hat{x}) \geq f(\hat{x})$.

Carathéodory’s theorem is the most important theorem of convex analysis and geometry. It states that coordinates of every point of the convex hull of the set $V \subset \mathbb{R}^n$ could be calculated as the weighted arithmetic mean of some points of $V$ and the maximum necessary number of these points is no more than $n + 1$. The beautiful exposition of this theorem is given in [4].

2.3. Optimal Distribution in An Averaged NLP Problem

Let us take some $x_0 \in V_x$. If $p(x) = \delta(x - x_0)$, then problem (6) coincides with the initial problem. If the set of admissible solutions of a problem includes the set of admissible solution of the initial NLP problem and the optimality criteria in both problems coincide on the set of admissible solutions of the NLP problem, then the former problem is called an extension of the NLP problem.

First, consider the special form of problem (6) with $f_i(x) = x_i$:

$$f_0(x) \rightarrow \max_{x_i = 0}, \quad i = 1, n.$$  \hspace{1cm} (7)

The value of the problem (7) is equal to the ordinate of the convex hull of the function $f_0(x)$ on the set $V_x$ at the point $x = 0$. According to Carathéodory’s theorem, constructing any ordinate of the convex hull of a function of $n$ variables requires averaging at most $n + 1$ ordinates of the function $f_0(x)$; therefore, we can rewrite problem (7) in the form

$$\sum_{v=0}^{n} \gamma_v f_0(x^v) \rightarrow \max \left/ \sum_{v=0}^{n} \gamma_v x^v_j = 0, \quad i = 1, n, \quad \gamma_v \geq 0, \quad \sum_{v=0}^{n} \gamma_v = 1. \right.$$ \hspace{1cm} (8)

Let us return to problem (6) and try to reduce it to simple calculation. We need to calculate the ordinate of a convex hull of the given function. Please note that problem (6) can be solved in two stages. At the first stage, we find the maximum of the function $f_0(x)$ subject to the constraint $f(x) = C$, where $C$ takes all values for which the level surface $f(x) = C$ intersects $V_x$. The problem

$$f_0(x) \rightarrow \max_{x_i = 0} f_i(x) = C, \quad i = 1, m, \quad x \in V_x$$ \hspace{1cm} (9)

is a non-linear programming problem. Solving (9), we obtain a set of conditionally optimal solutions $x^*(C)$ and the corresponding values of the reachability function $f_0^*(C) = f_0(x^*(C))$ of the non-linear programming problem.

The following assertion holds: The optimal distribution $p^*(x)$ in problem (6) is concentrated at the points $x^*(C)$. In other words, one needs to average only over conditionally optimal values of $f_0$.

2.4. Necessary Conditions of Optimality—Kuhn-Tucker Theorem

The Kuhn-Tucker theorem generalizes Lagrange multipliers method to problems with inequality constraints:

$$f_0(x) \rightarrow \max_x f_i(x) = 0, \quad \varphi_j(x) \geq 0, \quad i = 1, k, \quad j = k + 1, m,$$ \hspace{1cm} (10)

where all functions are smooth.
The theorem states that there is nonzero vector of Lagrange multipliers with components \( \lambda_i, \mu_j \leq 0 \) such that Lagrange function

\[
L = \lambda_0 f_0(x) + \sum_i \lambda_i f_i(x) + \sum_j \mu_j \varphi_j(x) = R(\lambda, x) + \sum_j \mu_j \varphi_j(x)
\]

is stationary on the optimal solution of the problem (10). The multiplier \( \lambda_0 \) could equal to zero or one. In the former case the solution is called degenerate.

It follows from this theorem that when \( \varphi_j(x) = x_j \) we have inequality \( \frac{\partial R}{\partial x_j} \leq 0 \) for the optimal solution. More detailed explanation could be found in [5].

2.5. Reduction to an Ordinary NLP Problem

The above considerations allow us to formulate the second stage in solving problem (6). This stage is the maximization of the average value of the function \( f_0^*(C) \) with the constraint that the vector \( C \) has zero mean, i.e.,

\[
\overline{f_0}(C) \rightarrow \max \quad \overline{C}_i = 0, \quad i = 1, m, \quad C_i \in V_C.
\]

This problem is similar to the problem (7). Its value, and hence the value of problem (6), is equal to the ordinate of the convex hull of the reachability function \( f_0^*(C) \) at \( C = 0 \):

\[
\sup_{x \in D} \overline{f_0}(x) = \sup \overline{f_0}(C) \quad \overline{C} = 0, \quad C \in V_C.
\]

Since the vector \( C \) is \( m \)-dimensional, the number of base points \( C^v \) in problem (9) is at most \( m + 1 \). Thus, the distribution \( p(C) \) in problem (9) can be sought in the form

\[
p(C) = \sum_{v=0}^{m} \gamma_v \delta(C - C^v).
\]

Since each of the base values \( C^v \) corresponds to a conditionally optimal solution \( x^*(C^v) \), the optimal distribution \( p(x) \) is also concentrated at no more than \( m + 1 \) points:

\[
p(x) = \sum_{v=0}^{m} \gamma_v \delta(x - x^v).
\]

Substituting the distribution (15) into the expressions for \( \overline{f_0}(x) \) and \( \overline{f_i}(x) \), we reduce problem (6) to the form

\[
I = \sum_{v=0}^{m} \gamma_v f_0(x^v) \rightarrow \max \quad \sum_{v=0}^{m} \gamma_v f_i(x^v) = 0, \\
i = 1, m, \quad x^v \in V_x, \quad \gamma_v \geq 0, \quad \sum_{v=0}^{m} \gamma_v = 1.
\]

Thus, we have reduced the problem to an ordinary NLP problem whose variables are the base values \( x^v \) of the vector \( x \) and the weight factors \( \gamma_v \).

2.6. Relationship between Averaged NLP Problem and the Lagrangian Function of the NLP Problem without Averaging

The Lagrangian function

\[
\mathcal{R} = \sum_{v=0}^{m} \gamma_v f_0(x^v) + \sum_{i=1}^{m} \lambda_i \sum_{v=0}^{m} \gamma_v f_i(x^v) + \Lambda \left( 1 - \sum_{v=0}^{m} \gamma_v \right) = \\
= \sum_{v=0}^{m} \gamma_v \left[ f_0(x^v) + \sum_{i=1}^{m} \lambda_i f_i(x^v) - \Lambda \right] + \Lambda
\]

(17)
of problem (16) is related to the Lagrangian function

\[ R = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \]  

(18)
of the initial NLP problem by

\[ \overline{R} = \sum_{\nu=0}^{m} \gamma_\nu (R(x^\nu, \lambda) - \Lambda) + \Lambda. \]  

(19)

Since \( \sum_\nu \gamma_\nu = 1 \), the Lagrangian function of the averaged problem equals the average value of the Lagrangian function of the initial problem over all base values \( x^\nu \). Some of the weight factors \( \gamma_\nu \) may vanish; then the number of base points is less than \( m + 1 \).

Let us find conditions that must hold for those \( x^\nu \) that have non-zero weights in (19). For this purpose, we apply the Kuhn–Tucker theorem and write the optimality conditions for problem (16) with respect to the variables \( \gamma_\nu \):

\[ \frac{\partial R}{\partial \gamma_\nu} \delta \gamma_\nu \leq 0. \]  

(20)

Since \( \gamma_\nu \) are bounded only from below (\( \gamma_\nu \geq 0 \)), it follows that \( \delta \gamma_\nu \geq 0 \); therefore,

\[ \frac{\partial R}{\partial \gamma_\nu} = R(x^\nu) - \Lambda \leq 0, \]  

(21)

or \( R(x^\nu) \geq \Lambda \). If \( \gamma_\nu^* > 0 \), then \( \delta \gamma_\nu \) may be of any sign, and so inequality (21) transforms into the equality

\[ R(x^\nu) = \Lambda. \]  

(22)

Thus, for all \( x^\nu \) involved in the averaged problem with non-zero weights, the Lagrangian function \( R \) of the initial non-linear programming problem attains an absolute maximum. Of course, this maximum is the same for all \( x^\nu \).

The requirements that the function \( R \) must take the same value at all points \( x^\nu \) and that this value must be maximum give equations for the variables to be found. Thus, applying Kuhn–Tucker theorem the problem (6), we obtain the vector of Lagrange multipliers \( \lambda \) for which the function \( R \) attains an absolute maximum with respect to the variables \( x_\nu \in V_x \) and \( \gamma_\nu \in V_\gamma \) at an element of the set \( D \) of admissible solutions to problem (6), and these multipliers \( \lambda \) satisfy the condition

\[ R(\lambda^*, \gamma_\nu^*, x^\nu^*) = \inf_{\lambda \in V_\lambda} \sup_{\gamma_\nu, x^\nu} R(\lambda, \gamma_\nu, x^\nu) = \inf_{\lambda \in V_\lambda} \sup_{x \in V_x} R(\lambda, x). \]  

(23)

Thus, when the attainability function \( f_0^*(C) \) coincides with its convex hull at \( C = 0 \), the transition to the averaged problem is not efficient (the values of the NLP problem and problem (6) coincide). By virtue of (23), we can look for the value of the averaged problem in the form \( \inf_{\lambda \in V_\lambda} \sup_{x \in V_x} R(\lambda, x) \). If the extended problem is inefficient, then we say that it is equivalent to the initial problem.

In the general case, the dimension of the vector of unknown variables and the computational complexity of problem (6) are much greater than those for the NLP problem. However, in many cases, we are interested not in the solution but in the value of the averaged problem, which shows the gain obtained by transition to the averaged setting. Some methods for estimating the value of problem (6) from above and below were proposed in [6].

2.7. Other Forms of Averaged Extensions of the NLP Problem

Problem (6) is not the only possible extension of the NLP problem by averaging. The optimality criteria, relations, and constraints in real-life problems often include the mean values of variables \( x \) rather than the variables themselves. For example, the performance of a distillation column
is characterized by the mean not current composition of output flows, because these flows are accumulated in some containers or apparatuses at the exit of the column (or attached to the column). Below, we describe several possible modifications of the averaged extension [7].

1. **Problem of maximizing a function of the mean value of the argument.** When \( D \) is the set of admissible solutions of the initial NLP problem, i.e., \( D \) is defined by the condition \( f(x) = 0 \), and \( \bar{x} \) is the mean value of the vector \( x \) on the set \( D \), we have:

\[
f_0(\bar{x}) \to \sup / p(x) = 0 \quad \forall x \notin D.
\]

Since the set of values \( \bar{x} \) satisfying this condition is the convex hull of \( D \), problem (24) is equivalent to the NLP problem on the convex hull of \( D \):

\[
f_0(x) \to \sup / x \in \text{Co } D.
\]

2. **Problem of maximizing the mean value of a function under constraints imposed on the mean value of the argument:**

\[
\overline{f_0(x)} \to \sup / f_i(\bar{x}) = 0, \quad i = \overline{1,m}
\]

or, in more detail,

\[
\int_{V_x} f_0(x) p(x) dx \to \sup / f_i \left( \int_{V_x} x \cdot p(x) dx \right) = 0, \quad i = \overline{1,m}.
\]

3. **Problem of maximizing a function of the mean value of \( x \) under averaged constraints:**

\[
f_0(\bar{x}) \to \sup / \overline{f_i(x)} = 0, \quad i = \overline{1,m}.
\]

Each of the above problems is an extension of the non-linear programming problem, and the solutions of these problems are distributions \( p(x) \).

**Averaged problems with two types of variables.** An NLP problem can be extended only with respect to some components of the solution rather than with respect to the whole solution. In practice, this situation occurs when the problem is solved repeatedly and some components (we denote them by \( x \)) can vary from one solution to another, while the remaining components must be chosen only once and then fixed. We denote the latter group of variables by \( y \). For example, \( x \) may be the operating conditions of the process (such as flow, pressure, temperature, etc.) and \( y \) may be the design parameters of an apparatus.

If we denote

\[
\overline{f(y, x)}^x = \int_{C_x} f(y, x) p(x) dx,
\]

a problem in which averaging is performed over only part of variables has the form

\[
f_0(y, x)^x \to \sup / \overline{f_i(y, x)}^x = 0, \quad i = \overline{1,m}.
\]

One need to find the vector \( y \) and distribution \( p(x) \) in (30).

For each fixed \( y \), this problem coincides with the usual setting of problem (6). If we separate the randomized variables \( x \in E \) and the deterministic variables \( y \in E^3 \) in the Lagrangian function \( R \) of the initial NLP problem, then we can write optimality conditions with respect to \( x \) by analogy with problem (6) in the form (see (23))

\[
R(\lambda, \gamma_y, y, x^{\nu}) = \sup_{x \in V_x} R(\lambda, y, x), \quad v = \overline{0,m}.
\]
In this case, if we denote the admissible set of (30) as $D_x(y)$, for each $y \in V_y$, there exist $\lambda(y)$ such that

$$\inf_{\lambda} \sup_{x \in V_x} R(\lambda, y, x) = \sup_{x \in D_x(y)} f_0(y, x).$$

(32)

The Lagrangian function attains an absolute maximum at the base values of $x$.

At the same time, for a fixed function $p(x)$, problem (30) becomes a usual non-linear programming problem with respect to the variables $y$. The Kuhn–Tucker conditions hold for this problem, which include in this case the complementary slackness conditions as well as the requirement that the function $R(\lambda, \gamma_v, y, x^v)$ be stationary with respect to $y$, which in turn, leads to the equations

$$\frac{\partial}{\partial y_j} \left[ \sum_{\nu=0}^{m} \gamma_{\nu} R(\lambda, y, x^\nu) \right] = 0, \quad j = 1, s.$$  

(33)

where $R$ is the Lagrangian function for the NP problem.

Averaged problems with two types of variables are in a sense close to optimal control problems, and optimality conditions for such problems are close to the Pontryagin maximum principle.

2.8. The Algorithm for Obtaining Optimality Conditions in Averaged Problems

By an averaged problem of static optimization we mean any NLP problem in which either functions or variables are averaged with respect to all or part of the variables.

As shown above, the settings of averaged problems are very diverse. The reason for this is that a problem may contain both the mean values of functions and functions of the mean values of variables. Moreover, averaging may involve only part of the variables. Under these conditions, it is inexpedient to study each possible setting of an averaged problem. It is significantly more convenient to obtain optimality conditions for some canonical form of an averaged problem and apply them to each particular problem after having reduced the latter to this canonical form [8].

Before obtaining optimality conditions, we must answer the following two questions:

1. Is the optimal distribution, which is one of the components of the solution of an averaged problem, always concentrated at finitely many base points?
2. If the answer to the previous question is “yes,” then what is the limit number of these points?

The necessary optimality conditions given below yield an affirmative answer to the first question and allow one to determine the limit number of base points.

Let $y$ denote the vector of deterministic variables, and let $x$ be the vector of randomized variables. For the former, we must find an optimal value, and for the latter, an optimal measure. The canonical form of the averaged problem is

$$F_0(f(x, y), y, x) \to \max$$

(34)

under the constraints

$$F_v(f(x, y), \varphi(x, y), x) = 0, \quad v = 1, r,$$

$$F_v(f(x, y), \varphi(x, y), x) \geq 0, \quad v = r + 1, m.$$  

(35)

Here the bar over the symbol of a function denotes averaging over the set $V_x$ of randomized variables $x$, which is assumed to be compact.

Suppose that the vector $x$ has dimension $k$ and the vector function $f$ has dimension $n$. The function $F$ is assumed to be continuously differentiable with respect to all its variables, and $f$ and $\varphi$ are continuous in $x$ and continuously differentiable in $y$.

In [8], one of the authors (A.T.) proved that the optimal measure $p^*(x)$ on the set of randomized variables is concentrated at no more than $L + 1$ base points, where $L = n + k$. Thus,

$$p^*(x) = \sum_{l=0}^{L} \gamma_l \delta(x - x^l), \quad \gamma_l \geq 0, \quad \sum_{l=0}^{L} \gamma_l = 1.$$  

(36)
Therefore, for the optimal solution, we have

\[ f^*(x, y) = \sum_{l=0}^{L} \gamma_l f(x^l, y), \quad x = \sum_{l=0}^{L} \gamma_l x^l, \]  

(37)

and constraints (35) take the form

\[ F_v(\bar{f}, \varphi(x^l, y), \bar{x}) = 0, \quad v = 1, r, \]
\[ F_v(\bar{f}, \varphi(x^l, y), \bar{x}) \geq 0, \quad v = r + 1, m. \]  

(38)

for all values of \( x^l \).

These expressions turn problem (34), (35) into an ordinary NLP problem with respect to \( \gamma_l, y \) and \( x^l \). The Kuhn–Tucker conditions reduce to the following: the Lagrangian function

\[ R = F_0(\bar{f}, y, \bar{x}) + \sum_{v=1}^{m} \lambda_v F_v(\bar{f}, \varphi(x^l, y), x) \]  

(39)

of this problem is stationary with respect to \( x^l \) and \( y \) and is unimprovable with respect to \( \gamma_l \) (we assume the solution is non-degenerate, so \( \lambda_0 = 1 \)). To write down the optimality conditions, we introduce the notation

\[ a_j = \frac{\partial R}{\partial f_j}, \quad \beta_i = \frac{\partial R}{\partial x_i}, \quad r_{\mu l} = \frac{\partial R}{\partial \varphi_{\mu l}}(x^l, y). \]  

(40)

Using this notation, we can write the condition that \( R \) is unimprovable with respect to \( \gamma_l \) as follows: the expression

\[ C(x) = \sum_j a_j f_j(x, y^*) + \sum_i \beta_i x_i \]  

(41)

attains its maximum with respect to \( x \in V_x \) at the points \( x^l \), so that

\[ x^{l*} = \arg \max_x C(x), \quad l = 1, L; \]  

(42)

the condition that \( R \) is stationary with respect to \( y \) has the form

\[ \nabla_y \left[ \sum_j a_j f_j(x, y) + F_0(\bar{f}, x, y) + \sum_{\mu l} r_{\mu l} \varphi_{\mu l}(x^l, y) \right] = 0. \]  

(43)

The maximality of \( C(x) \), together with equations (42), constraints (35), and the complementary slackness conditions

\[ \sum_{v=r+1}^{m} \lambda_v F_v(\bar{f}^*, \varphi^*, x^*) = 0, \quad \lambda_v \geq 0, \quad v = r + 1, m \]  

(44)

allows one to find a solution \( \gamma^*_l, y^*, x^l \).

When formulating a specific averaged problem, one

1. writes the conditions of the problem in the canonical form (34), (35);
2. separates the randomized and deterministic variables;
3. calculates the total number \( L \) of averagings, which is equal to the sum of the dimensions of the vector of randomized variables and of the vector of functions to be averaged;
4. constructs the functions \( R \) and \( C \) and substitutes them into expressions (42)–(44).

For example, in problem (26), we have

\[ F_0 = \bar{f}_0(x), \quad F_v = f_v(\bar{x}), \quad v = 1, m. \]  

(45)
The number $L$ equals $k$, and
\[ R = \lambda_0 f_0(x) + \sum_{v=1}^{m} \lambda_v f_v(x). \] (46)

In (42), we have $a_0 = \lambda_0 = 1$, $a_v = 0$ for $v > 0$ and
\[ \beta_i = \sum_{v=1}^{m} \lambda_v \left( \frac{\partial f(x)}{\partial x_i} \right)_x, \quad i = 1, k. \] (47)

At the base points $x_i$, the number of which does not exceed $k + 1$, the expression
\[ C(x) = f_0(x) + \sum_{i=1}^{k} x_i \sum_{v=1}^{m} \lambda_v \left( \frac{\partial f(x)}{\partial x_i} \right)_x \] (48)
attains its maximum, and conditions (35) hold, which have the form
\[ f_v \left( \sum_{i=0}^{k} \gamma_i x^i \right) = 0, \quad v = 1, m. \] (49)

3. Non-Stationary Problems of Averaged Optimization

Consider an extremal problem of the form
\[ \mathcal{J}_0 = \frac{1}{\tau} \int_{0}^{\tau} f_0(J(t), u(t)) dt \to \max_u \] (50)
subject to the constraints
\[ \mathcal{J}_v = \frac{1}{\tau} \int_{0}^{\tau} f_v(J(t), u(t)) dt = 0, \quad v = 1, m, \] (51)
where the functions $f_v : \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \to \mathbb{R}$, $v = 0, \ldots, m$, are continuous in $J$ and $u$, $u \in V_u \subset \mathbb{R}^{k_1}$ is a measurable function, the set $V_u$ is compact, and $J(t) \in V_J \subset \mathbb{R}^{k_2}$ is a given measurable function of time. With $J(t)$ we can associate a probability measure (distribution) $p(J)$. If $J(t)$ takes a value $J^k$ on a part of the interval $(0, \tau)$ of relative length $\alpha_k$, then $p(J)$ contains a term of the form $\alpha_k \delta(J - J^k)$. The length of the interval $(0, \tau)$ may tend to infinity, and $J(t)$ may be a stationary random process with distribution $p(J)$.

The distribution $p(J)$ can be written in the form
\[ p(J) = \mathcal{P}(J) + \sum_k \alpha_k \delta(J - J^k). \] (52)

For problem (50), (51), let $\alpha \tau$ be the length of the part of $(0, \tau)$ on which $J(t)$ takes one of the constant values $J^k$; we have $\alpha = \sum_k \alpha_k$. We refer to $\alpha \tau$ as the total constancy interval of $J(t)$. The remaining part $(1 - \alpha)\tau$ is called the interval of variation of the parameter $J$.

**Theorem 1.** Let $u^*(t)$ be an optimal solution; then there exists a non-zero vector $\lambda = \{\lambda_0, \ldots, \lambda_m\}$ with $\lambda_0 \in \{0, 1\}$ such that

- on the interval of variation of the parameter $J(t)$
  \[ u^*(J, \lambda) = \arg\max_{u \in V_u} \sum_{v=0}^{m} \lambda_v f_v(J, u); \] (53)
on the total constancy interval of \( J(t) \), the optimal solution switches between at most \( n + 1 \) base values \( u^j \), and each of these values satisfies the condition

\[
u^j = \arg \max_{u \in V} \sum_k a_k \lambda_v f_v(J^k, u), \quad j = 0, n; \quad (54)
\]

- the portions \( \gamma_j \) of the constancy interval \( \alpha \tau \) on which \( u^*(t) \) takes the respective values \( u^j \) satisfy the conditions

\[
\int_{\gamma_j} \overline{p}(J) f_v(J, u^*(J)) \, dJ + \sum_{j=0}^n \gamma_j \sum_k a_k f_v(J^k, u^j) = 0, \quad v = \overline{1, n},
\]

\[
\sum_{j=0}^n \gamma_j = 1, \quad \gamma_j \geq 0; \quad (55)
\]

- the vector of multipliers \( \lambda_v, v = \overline{1, n} \), is determined by the conditions

\[
\lambda^* = \arg \min_{\lambda} \left[ \int_{\gamma_j} \overline{p}(J) \sum_{v=0}^n \lambda_v f_v(J, u^*(J, \lambda)) \, dJ + \sum_{j=0}^n \gamma_j \sum_{v=0}^n \lambda_v \sum_k a_k f_v(J^k, u^j(\lambda)) \right]. \quad (56)
\]

Thus, on the constancy intervals, the optimal solution of a problem with non-stationary parameters coincides with the solution of an averaged mathematical programming problem, and on the interval of variation of the parameter, it varies as the solution of a problem with integral constraints. This theorem was proved in [9].

**Example 1.** Consider the problem of maximizing the average power \( \overline{p} \) of a heat engine in which the working fluid contacts a source of variable temperature \( T_0(t) \). This problem has the form

\[
\overline{p} = \frac{1}{\tau} \int_0^\tau q(T_0(t), T(t)) \, dt \to \max
\]

subject to the constraint

\[
\overline{\sigma} = \frac{1}{\tau} \int_0^\tau \frac{q(T_0(t), T(t))}{T(t)} \, dt = 0. \quad (58)
\]

Here \( T(t) \) is the temperature of the working substance, \( q \) is the heat flux from the source to the working fluid, and \( \overline{\sigma} \) is the mean rate of variation of the entropy of the working substance. A substantiation of the setting (56), (57) can be found in [10–12]. The optimality conditions (53) imply the following relation for the interval of variation of \( T_0(t) \):

\[
\frac{1}{T^2} \frac{\partial q(T_0, T)}{\partial T} = \frac{1}{T} = \text{const} \quad (59)
\]

In particular, for the Newtonian law \( q(T_0, T) = \beta(T_0 - T) \) of heating, (59) implies

\[
T^*(T_0) = m \sqrt{T_0}, \quad (60)
\]

where \( m \) is the constant equal to the mean value of the square root of the source temperature.

For example, suppose that \( T_0(t) \) has a uniform distribution (for a regular function \( T_0(t) \), this means that the source temperature depends linearly on time) and \( T_{02} \) and \( T_{01} \) are the maximal and minimal source temperatures, respectively. Then

\[
T^*(T_0) = \frac{2(T_{02}^{3/2} - T_{01}^{3/2})}{3(T_{02} - T_{01})} \sqrt{T_0}. \quad (61)
\]
The maximum power is given by

$$p_{\text{max}} = \beta \left[ \frac{T_{02} + T_{01}}{2} - \frac{4}{9} \left( \frac{T_{02}^{3/2} - T_{01}^{3/2}}{T_{02} - T_{01}} \right) \right].$$  \hspace{1cm} (62)

Thus, a heat engine with one source may have non-zero power if the variance of the source temperature is positive.

For some laws $q(T_0, T)$, the optimal temperature $T^*(t)$ may switch between two base values on intervals of constancy of the parameter $T_0$.

4. Estimation of the Performance of Cyclic Modes

Suppose that the dynamics of a system is characterized by the differential equations

$$x_\nu = f_\nu(x, u, a), \quad \nu = 0, \ldots, m, \quad (63)$$

whose right-hand sides do not explicitly depend on $t$. Here, as in the preceding sections, $x$ denotes the state variables, $u$ are the control ones, and $a$ denotes parameters to be optimized. As a rule, boundary conditions are not fixed for equations (63), but the state variables are required to vary periodically:

$$x_\nu(\tau) = x_\nu(0) \Rightarrow \int_0^\tau f_\nu(x, u, a) dt = 0, \quad \nu = 0, \ldots, m. \quad (64)$$

The performance averaged over the cycle plays the role of the optimality criterion for such a cyclic process and can be written in the form

$$I = \frac{1}{\tau} \int_0^\tau f_0(x, u, a) dt \rightarrow \text{max}. \quad (65)$$

The duration $\tau$ of each cycle is one of the components of the vector $a$; in the general case, it is not fixed. The parameters and controls are subject to constraints $a \in V_a$ and $u \in V_u$, in addition to the integral constraints (64), which follow from the periodicity of the process, the problem usually contains integral constraints determined by given mean rates of consumption of some resources (resource constraints):

$$J_j = \int_0^\tau \phi_j(x, u, a) dt = 0, \quad j = 1, \ldots, r. \quad (66)$$

It is assumed that each of the functions determining the problem is continuous in all its variables and is continuously differentiable with respect to $x$ and $a$.

**Optimality conditions.** Optimality conditions for problem (63)–(66) can be obtained by using the maximum principle [6]. Namely, if an optimal solution $x^*, a^*, u^*$ exists and is non-degenerate, then there exist a non-zero vector $\lambda$ and a differentiable vector function $\psi(t)$ such that the function

$$R = \frac{1}{\tau} f_0 + \sum_\nu \left[ \psi_\nu x_\nu + (\psi_\nu + \lambda_\nu) f_\nu \right] + \sum_j \lambda_j \phi_j, \quad \hspace{1cm} (67)$$

is stationary with respect to $x$ and attains a maximum with respect to $u$, and the integral $S$ of this function is locally unimprovable with respect to $a$. Thus,

$$\frac{\partial R}{\partial x_i} = 0 \quad \Rightarrow \quad \psi_i = -\frac{\partial}{\partial x_i} \left\{ \frac{1}{\tau} f_0 + \sum_\nu (\psi_\nu + \lambda_\nu) f_\nu + \sum_j \lambda_j \phi_j \right\}. \quad (68)$$
Since the values \( x_v(\tau) \) and \( x_v(0) \) are not fixed, it follows that \( \psi_v(\tau) \) and \( \psi_v(0) \) vanish. Introducing the notation \( \tilde{\psi}_v = \psi_v + \lambda_v \) and taking into account the equality \( \dot{\tilde{\psi}}_v = \dot{\psi}_v \), we can rewrite condition (68) in the form

\[
\dot{\tilde{\psi}}_v = -\frac{\partial}{\partial x_i} \left\{ \frac{1}{\tau} f_0 + \sum_v \psi_v f_v + \sum_j \lambda_j \phi_j \right\} = -\frac{\partial}{\partial x_i} H. \tag{69}
\]

For these equations, since \( \psi(0) \) and \( \psi(\tau) \) vanish, the costate variables satisfy the periodicity conditions

\[
\tilde{\psi}_v(0) = \tilde{\psi}_v(\tau) \Rightarrow \int_0^\tau \frac{\partial H}{\partial x_v} dt = 0, \quad v = 1, 2, \ldots. \tag{70}
\]

The conditions of maximality of \( R \) with respect to \( u \) have the form

\[
u^*(t) = \arg \max_{u \in V_u} \left\{ f_0(\tau) + \sum_v \tilde{\psi}_v f_v + \sum_j \lambda_j \phi_j \right\}. \tag{71}\]

Finally, the optimality conditions with respect to each component \( a_k \) of the vector \( a \), including the duration \( \tau \) of the cycle, yield the inequalities

\[
\frac{\partial S}{\partial a_k} \delta a_k \leq 0, \quad k = 1, 2, \ldots \tag{72}
\]

Here \( \delta a \) is the cone of variations of the vector \( a \) that are admissible with respect to the inclusion \( a \in V_a \).

Please note that the phase trajectory corresponding to an optimal cyclic process has no self-intersections [13].

5. Estimation of the Efficiency of Transition to a Cyclic Process

5.1. Conditions of Equivalence and Efficiency of a Cyclic Extension

The optimal cyclic mode problem (63)–(66) (we refer to it as Problem C) is an extension of a non-linear programming problem. Indeed, imposing the additional constraints \( x = \text{const} \) and \( u = \text{const} \) on the solution of this problem, we obtain the following optimal static mode problem (Problem S):

\[
I_S = f_0(x, u, a) \rightarrow \max \quad \begin{cases} 
 f_v(x, u, a) = 0, \quad \phi_j(x, u, a) = 0 \\
 u \in V_u, \quad a \in V_a, \quad v = 1, 2, \ldots, \quad j = 1, 2
\end{cases} \tag{73}
\]

Since the set of admissible solutions of problem (63)–(66) is larger than that of Problem S, it follows that

\[
I^*_S \leq I^*_C. \tag{74}
\]

where \( I^*_C \) denotes the value of the optimal cyclic mode problem.

One of the problems in designing cyclic processes consists of distinguishing a class of problems for which inequality (74) turns into an equality, i.e., the cyclic extension is equivalent to the static problem. An important role in solving this problem is played by the Lagrangian function of Problem S,

\[
R_S = f_0(x, u, a) + \sum_v \lambda_v f_v(x, u, a) + \sum_j \xi_j \phi_j(x, u, a) \tag{75}
\]

To determine whether a cyclic process is equivalent to a static one or efficient without solving problem (63)–(66), we form averaged problems, which are in turn extensions for Problem S or C or for both. Comparing the values of these problems with \( I^*_C \), we find conditions for the equivalence of a cyclic extension.
1. An upper bound for $I_\mathrm{C}^*$ and sufficient conditions for the equivalence of a cyclic extension. Let us enlarge the set of admissible solutions of Problem C by removing the differential equations (63). We obtain Problem $\bar{S}$, which we call an estimating problem:

$$I_\bar{S} = f_0(x, u, a)^{x, u} \rightarrow \max \left/ \frac{f_v(x, u, a)}{v = \overline{V}, \ u \in V_u, \ a \in V_a, \ j = \overline{J}, r. \right.$$  \hspace{1cm} (76)

Clearly,

$$I_\bar{S} \geq I_{\mathrm{C}}^*,$$  \hspace{1cm} (77)

and Problem $\bar{S}$ is an averaged extension of Problem S with the variables $x$ and $u$ and the parameters $a$. The roles of the variables $x$ and $u$ in the conditions of Problem $\bar{S}$ are similar, and we unite these variables and denote them by $y = (x, u)$. In shorthand notation, this problem has the form

$$I_\bar{S} = f_0(y, a)^{y} \rightarrow \max \left/ f_v(y, a)^{y} = 0, \ \overline{\varphi}_j(y, a)^{y} = 0, \ v = \overline{V}, \ a \in V_a, \ j = \overline{J}, r. \right.$$  \hspace{1cm} (78)

The value of problem (78) as an extension of the optimal static mode problem can be expressed in terms of the function $R_\bar{S}$ as

$$I_\bar{S} = \inf_{\lambda} \sup_{y} R_\bar{S}(y, a^*, \lambda, \xi).$$  \hspace{1cm} (79)

For determining the vector of parameters, we have the condition

$$\left[ \frac{\partial}{\partial a} R_\bar{S}(y, a, \lambda, \xi)^y \right]_{a = a^*} = 0.$$  \hspace{1cm} (80)

If $a^*$ lies inside $V_u$, then condition (80) reduces to the condition of stationarity of $R_\bar{S}$ with respect to $a$.

If the value $I_\bar{S}^*$ given by (79) equals $I_{\mathrm{C}}^*$ (i.e., Problem $\bar{S}$ has a unique base solution), then inequalities (74) and (77) imply $I_{\mathrm{C}}^* = I_{\bar{S}}^*$; i.e., the static mode cannot be improved by passing to a cyclic mode. If $I_\bar{S}^* > I_{\mathrm{C}}^*$, then the difference $\Delta_\bar{S}$ between these values gives an upper bound for the possible gain from the passage to a cyclic mode.

2. A lower bound for $I_\mathrm{C}^*$. Quasi-static and sliding modes. Consider the case when $x(t)$ and $u(t)$ vary so that the time derivatives of $x(t)$ can be neglected. Then the relations between $x$ and $u$ are given, as in the static case, by $f(x(t), u(t), a) = 0$ for all $t$. The corresponding modes are said to be quasi-static. The problem of an optimal choice of $x(t)$ and $u(t)$ under the quasi-static conditions (Problem QS) has the form

$$I_{\mathrm{QS}} = \frac{1}{T} \int_0^T f_0(x, u, a)dt \rightarrow \max / f(x, u, a) = 0, \ \int_0^T \varphi(x, u, a)dt = 0, \ u \in V_u, \ a \in V_a.$$  \hspace{1cm} (81)

or, in shorthand notation,

$$I_{\mathrm{QS}} = f_0(y, a)^y \rightarrow \max \varphi(y, a)^y = 0, \ y \in V_y, \ a \in V_a.$$  \hspace{1cm} (81)

Here $y = (x, u)$, and the set $V_y$ is determined by the conditions $u \in V_u, a \in V_a,$ and $f(x, u, a) = 0$. Since any solution of Problem QS is admissible for Problem C, it follows that

$$I_{\mathrm{QS}} \leq I_{\mathrm{C}}^*.$$  \hspace{1cm} (82)
At the same time, the value $I^*_{QS}$ of Problem QS, being the value of an averaged problem, is given by the expression

$$I^*_{QS} = \inf_{\xi} \left\{ \sup_y \left[ f_0(y, a^*) + \xi \varphi(y, a^*) \right] \right\} \frac{f(y, a^*)}{f(y, a^*)} = 0, \; u \in V_u \right\}. \quad (83)$$

Here $a^*$ is the optimal value of $a$ subject to the constraint

$$\frac{\partial}{\partial a} \left[ f_0(y, a) + (\xi \varphi(y, a)) \right] + \sum_{i=0}^{r} \lambda_i f\left( y_i, a \right) \right] \delta a \leq 0. \quad (84)$$

in which $\delta a$ is the set of variations allowed by the inclusion $a \in V_a$.

We choose the Lagrange multipliers $\lambda_i$ in (84) so that $f_0(y_i, a) = 0$ for any base value $y_i$ of the vector $y$. The number of base values of $y$ is determined by the dimension $r$ of the vector function $\varphi$; thus, the problem takes the form

$$\tilde{f}_0 = \sum_{i=0}^{r} \gamma_if_0(y_i, a), \; \varphi = \sum_{i=0}^{r} \gamma_i \varphi(y_i, a), \; \sum_{i=0}^{r} \gamma_i = 1, \; \gamma_i \geq 0. \quad (85)$$

Consider the case when the control vector in the steady state of the system changes with a frequency so high that the state vector $x$ remains virtually constant. Such a mode is called a sliding steady mode. The optimization problem for such a mode is formulated as

$$I_{SL} = \tilde{f}_0(b, u)u \to \sup / f(b, u)u = 0, \; u \in V_u, \; \varphi(b, u)u = 0, \; b \in V_b. \quad (86)$$

This problem is known as Problem SL. In (86), $b$ denotes the vector formed by $x$ and $a$. This mode is the limit case of the cyclic mode, so we have

$$I^*_{SL} \leq I^*_C.$$

Problem (86) is an averaged extension of Problem S with two types of variables; its value is given by

$$I^*_{SL} = \min_{\lambda, \xi} \max_u R(u, b^*; \lambda, \xi) = \min_{\lambda, \xi} \max_u \left\{ f_0(u, b^*) + \lambda f(u, b^*) + \xi \varphi(u, b^*) \right\}, \quad (87)$$

where $b^*$ satisfies the condition

$$\frac{\partial}{\partial b} R(u, b^*; \lambda, \xi)u \delta b \leq 0. \quad (88)$$

The number of base values of the vector function $u$ in Problem SL is at most $m + r + 1$.

A necessary condition for the efficiency of the transition to a cyclic mode can be stated in terms of $I^*_{QS}$ and $I^*_{SL}$. Consider the quantity

$$I_K = \max \left[ I^*_{QS}, I^*_{SL} \right].$$

If $I_K$ is greater than $I^*_S$, then the passage to a cyclic mode is efficient, and the difference

$$\Delta_K = I_K - I^*_S$$

provides a lower bound for the efficiency.
5.2. The Frequency Criterion for the Efficiency of the Passage to a Cyclic Mode

Suppose that an optimal static mode \( x^0, u^0 \) in Problem S is known. As above, it is required to determine whether the cyclic extension of Problem S is efficient. In [14], a frequency criterion for the efficiency of a cyclic mode was proposed. This criterion is based on the analysis of the increment in the optimality criterion \( I \) as compared to its maximal static value \( I^0 \) for small harmonic oscillations of the control about \( u^0 \).

Let \( \lambda^0 \) and \( \mu^0 \) be the values of Lagrange multipliers \( \lambda \) and \( \mu \) corresponding to the optimal static mode in the Lagrangian function

\[
R = f_0(x, u) + \sum_i \lambda_i f_i(x, u) + \sum_j \mu_j \varphi_j(x, u)
\]

for Problem S.

In a neighborhood of the optimal static mode and the corresponding Lagrange multipliers, we calculate the first and second derivatives of the functions that determine the problem with respect to \( x \) and \( u \) (if \( x \) and \( u \) are vectors, then these derivatives are matrices):

\[
A = \frac{\partial f}{\partial x}, \quad B = \frac{\partial f}{\partial u}, \quad P = \frac{\partial^2 R}{\partial x^2}, \quad Q = \frac{\partial^2 R}{\partial x \partial u}, \quad H = \frac{\partial^2 R}{\partial u^2}, \quad K = \frac{\partial \varphi}{\partial x}, \quad M = \frac{\partial \varphi}{\partial u}.
\]

In a neighborhood of the optimal static mode, the increment of the functional \( I \) under small variations \( \delta x(t) \) and \( \delta u(t) \) is given by

\[
\Delta I = \frac{1}{2T} \int_0^T (\delta x'(t)P \delta x(t) + \delta x'(t)Q \delta u(t) + \delta u'(t)Q^\prime \delta x(t) + \delta u'(t)H \delta u(t)) dt.
\]

The transition to a cyclic mode is efficient if there is a variation \( \delta u \) such that the quantity \( \Delta I \) is positive under the linearized constraints (63), i.e.,

\[
\delta x = \delta A \delta x + \delta B \delta u, \quad \delta x(T) = \delta x(0) = 0.
\]  \hspace{1cm} (89)

To get rid of these constraints, we consider only harmonic variations, i.e., those of the form

\[
\delta u(t) = \sum_{\nu = -\infty}^{\infty} u_{\nu} e^{i \nu \frac{2\pi t}{T}}.
\]

Applying the Fourier transform to the linear differential constraints (89), we obtain

\[
\delta x(i\omega) = \delta u(i\omega) \frac{B}{i\omega E - A} = \delta u(i\omega)W(i\omega).
\]

Here \( E \) is the identity matrix (\( E = 1 \) for scalar \( x \)). It is assumed that the matrix \( A \) has no eigenvalues with zero real part; otherwise, small deviations \( \delta u(t) \) may correspond to large deviations \( \delta x(t) \), and the linearization may be incorrect.

Let us express the quantity \( \Delta I \) by Parseval’s identity in the frequency domain, replacing \( \delta x(i\omega) \) by its expression in terms of \( \delta u \). The increment in the criterion under harmonic oscillations of the control with frequencies that are multiples of \( 2\pi / T \) takes the form

\[
\Delta I = \frac{1}{2} \int_{-\infty}^{\infty} \delta u'(-i\omega) A(\omega) \delta u(i\omega) d\omega.
\]

Here \( A(\omega) \) is defined by the matrices \( P, Q, \) and \( H \) and the relation between \( \delta u \) and \( \delta x \); it is easy to show that

\[
A(\omega) = W'(-i\omega)PW(i\omega) + Q'W(i\omega) + W'(-i\omega)Q + H,
\]
where the prime denotes transposition.

For the scalar problem, we have

$$A(\omega) = P|W(i\omega)|^2 + 2Q \text{Re} W(i\omega) + H$$

If the matrix $A(\omega)$ for some $\omega$ is such that the integrand in the expression for $\Delta I$ is positive for at least one vector $\delta u$, then the static mode can be improved and the passage to a cyclic mode is efficient.

For the scalar problem, we have

$$\Delta I = \frac{1}{2} \int_{-\infty}^{\infty} |\delta u(i\omega)|^2 A(\omega) d\omega,$$

and the static mode improves if $A(\omega)$ is positive for some $\omega$.

5.3. Lyapunov Problems

For an important class of problems, the inequality (77) turns into an equality. In these problems, the functions $f_0, f,$ and $\varphi$ in relations (63)–(66) depend only on $u$ and $a$, so that

$$\dot{x} = f(u, a)$$

(90)

Such equations are called Lyapunov-type equations, and the corresponding problems are known as Lyapunov problems. If we discard equations (63), which have the form (90), in Problem C, thereby transitioning to Problem S, then we can find its solution $u^*(t), a^*$. Substituting this solution into equation (90), we determine an optimal trajectory. Clearly, in this problem, $I^*_S = I^*_C$, $u^*(t)$ takes at most $m + r + 1$ base values, and the function $x^*(t)$ is a polygonal line with at most $m + r$ (internal) vertices.

Problems that include, in addition to Lyapunov-type equations, equations of the form

$$\dot{y}_v = f_v(u, a)F_v(x_v)$$

can also be reduced to Lyapunov problems. Indeed, such equations can be reduced to the form (90) by the change

$$y_v(x_v) = \int \frac{dx_v}{F_v(x_v)}$$

(91)

so that $\dot{y}_v = f_v(u, a)$. The optimal solution $y_v(t)$ is piecewise linear, and $x_v(t)$ can be found from (91) by solving the equation

$$\frac{dx_v}{F_v(x_v)} = y_v(t) dt.$$

6. Average Optimization in Finite-Time Thermodynamics

The field of finite-time thermodynamics is one of the most important examples of application of averaged optimization techniques. The reasons for this are the following:

1. Problems of optimal thermodynamic cycles.

There are a very important kind of thermodynamic systems — intermediary ones. These systems contact different subsystems (reservoirs) alternately while producing power and thus lowering the irreversibility arising from a continuous contact of the above-mentioned subsystems. The main example here is the heat engine, where the working fluid contacts two sources of different temperature.

One of the most essential problems in finite-time thermodynamics is the problem of maximum average power of heat engines, when the average rate of the heat flow from the hot source is given.
Similar problems arise also in absorption-desorption systems, where the working fluid contacts with the multi-component mixture and picks one component out from one source, releasing it at another one.

In reverse cycles, the working fluid obtains the energy from the exterior system. Upon contact with the source that loses energy or matter in the regular cycle, the working fluid enriches it with the corresponding resource.

In all of these problems, the working fluid restores its state at the beginning of every cycle. One needs to average all of the variables determining the process.

2. Relations between intensive and extensive variables are Lyapunov-type equations. Thermodynamic variables are divided into two classes: intensive (temperature, pressure, chemical potential, . . . ) and extensive (volume, internal energy, entropy, amount of substance, . . . ) ones. Flow rates of transport processes between subsystems depend only on intensive variables. This value determines in turn the rate of change of extensive variables. This means that equations determining the change of state of the thermodynamic system have the form [10–12]:

\[
\frac{dZ_j}{dt} = F_j(u_i, u_j).
\]  

(92)

Here \(i\) and \(j\) are indices of the contacting subsystems, \(u\) is the vector of intensive variables, \(Z\) is the vector of extensive variables. Equations of this type are called Lyapunov-type equations earlier in this paper. The right hand side of these equations does not depend on \(Z\), and the increase of \(Z\) is determined by the average value of the function \(F\). As we have shown above, one can obtain the limiting capabilities of systems characterized by Lyapunov-type equations using techniques of the averaged optimization.

7. Example: Averaged Optimization of a Heat Engine

7.1. Maximum Average Power Output

We will assume that there is a heat engine with constant temperature of sources \(T_+\) and \(T_-\) [15]. If we denote the temperature of the source contacting with the working fluid at the moment as \(T_n\) and the temperature of the working fluid itself as \(T\), we will obtain that the average output per cycle is

\[
\bar{p} = q(T_n, T).
\]  

(93)

Now we can formulate the averaged optimization problem, given that the average entropy generation within the working fluid per cycle is zero:

\[
\frac{q(T_n, T)}{T} \rightarrow \max_T \quad \frac{q(T_n, T)}{T} = 0, \quad T_n = (T_+; T_-), \quad T > 0.
\]  

(94)

This is the problem in the form (6). Using the algorithm described earlier (see (42)–(44)) we find the number of base points is two. This means that the Lagrange function

\[
R = q(T_n, T) + \lambda \frac{q(T_n, T)}{T} = q(T_n, T) \left(1 + \frac{\lambda}{T}\right)
\]  

(95)

has two maxima, so

\[
T_1 = \arg \max_T q(T_+, T) \left(1 + \frac{1}{T}\right),
\]

\[ T_2 = \arg \max_T q(T_-, T) \left(1 + \frac{1}{T}\right). \]  

(96)
Both maxima are global and therefore they must be equal \[16\]. It means that the Lagrange multiplier is the solution of
\[
q(T_+, T_1) \left( 1 + \frac{\lambda}{T_1} \right) = q(T_+, T_2) \left( 1 + \frac{\lambda}{T_2} \right).
\]
(97)

When the heat transfer law is linear
\[
q(T_+, T) = \alpha_+ (T_+ - T), \quad q(T_-, T) = \alpha_- (T_- - T),
\]
(98)
solution of equations (94)–(96) with notation \(\alpha = \frac{\alpha_+ \alpha_-}{(\sqrt{\alpha_+} + \sqrt{\alpha_-})^2}\) leads to
\[
p_{\text{max}} = \alpha \left( \sqrt{T_+} - \sqrt{T_-} \right)^2.
\]
(99)

The relationship between entropy generation and heat flows is shown at Figure 3. It is clear from this picture that the point of maximum power output lies on the convex hull of original output curves, so it is attainable only when averaged control is used.

![Figure 3. Relationship between entropy generation and heat flows and its convex hull. Here \(q^h\) and \(q^c\) are heat exchange rates upon contact with the hot and cold reservoirs, respectively, and \(\sigma^h, \sigma^c\) are the corresponding entropy generation rates. The optimal solution is attained when \(\sigma^c = \sigma_1, \sigma^h = \sigma_2\), \(q^1 = q^c(\sigma_1), q^2 = q^h(\sigma_2)\) and \(p_{\text{max}} = q^1 + q^2\).](image)

7.2. Maximum Efficiency

When the power output \(p_0\) is given, the problem of maximum efficiency is equivalent to the problem of minimum entropy generation within the system. Using again that the average entropy generation within the working fluid is zero for a cyclic process, we obtain the problem:

\[
- \sigma = \left( \frac{q(T_n, T)}{T_n} \right) \to \max, \quad \frac{q(T_n, T)}{T_n} = p_0, \quad \frac{q(T_n, T)}{\sqrt{T_n}} = 0, \quad T_n = (T_+; T_-), \quad T > 0.
\]
(100)

One may notice that this problem allows three base points in general, because there are three averaging operations in (100). This is the case when the entropy generation as a function of \(q(T_+, T)\) is not convex. We will not consider this case here, because this function is convex for most of heat transfer laws.
Another possibility corresponds to two base points. In this case, we have the following equations for $T_1$ and $T_2$:

\[
T_1 = \arg \max_T \left[ q(T_+, T) \left( \frac{1}{T_+} + \lambda + \frac{\mu}{T} \right) - \lambda p_0 \right],
\]

\[
T_2 = \arg \max_T \left[ q(T_-, T) \left( \frac{1}{T_-} + \lambda + \frac{\mu}{T} \right) - \lambda p_0 \right].
\] (101)

These maxima must be equal, which leads to:

\[
q(T_+, T) \left( \frac{1}{T_+} + \lambda + \frac{\mu}{T} \right) = q(T_-, T) \left( \frac{1}{T_-} + \lambda + \frac{\mu}{T} \right).
\] (102)

The averaged constraints must also be satisfied:

\[
\gamma q(T_+, T_1) + (1 - \gamma) q(T_-, T_2) = p_0, \\
\gamma \frac{q(T_+, T_1)}{T_1} + (1 - \gamma) \frac{q(T_-, T_2)}{T_2} = 0.
\] (103)

Equations (101)–(103) allow one to find values of $T_1$, $T_2$, $\lambda$, $\mu$ and $\gamma$.

For the linear heat transfer law we have the following value of maximum efficiency:

\[
\eta_{\text{max}}(p) = \frac{1}{2} \left( \frac{p}{\alpha T_+} + \eta_c \right) \pm \sqrt{\left( \frac{1}{4} \left( \frac{p}{\alpha T_+} + \eta_c \right) \right)^2 - \frac{p}{\alpha T_+} - \frac{p}{\alpha T_+}}.
\] (104)

When $p \to 0$, the value of (104) approaches $\eta_c$ (Carnot efficiency) and when $p = p_{\text{max}}$ (99), we have

\[
\eta_{\text{max}}(p_{\text{max}}) = 1 - \sqrt{\frac{T_-}{T_+}} = 1 - \sqrt{1 - \eta_c},
\] (105)

which is the well-known result of Novikov [17], Chambadal [18], Curzon and Ahlborn [19]. Important results for other types of heat transfer laws and different processes are presented in [20–22].

8. Results

We obtained the general necessary conditions of optimality for averaged optimization problems. These conditions can be written down using the algorithmic procedure given in the paper, which allows one to use them for problems of any structure. We showed how these techniques can be applied to the problems of finite-time thermodynamics leading to new results in the field.

Author Contributions: Conceptualization, A.T. and I.S.; methodology, A.T.; validation, I.S.; writing–original draft preparation, A.T.; writing–review and editing, I.S.; supervision, A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tsirlin, A.M. Problems and methods of averaged optimization. Proc. Steklov Inst. Math. 2008, 261, 270–286. [CrossRef]
2. Kaplinski, A.M.; Propoi, A.I. Stochastic Approach to Nonlinear Programming Problems. Automat. Remote Control 1970, 31, 448–459.
3. Ioffe, A.; Tikhomirov, V. Theory of Extremal Problems; Elsevier North-Holland: New York, NY, USA, 1979; p. 459.
4. Rockafellar, R.T. Convex Analysis; Princeton University Press: Princeton, NJ, USA, 1996; p. 472.
5. Boltyanski, V.G; Martini, H.; Soltan, V. Geometric Methods and Optimization Problems; Springer: New York, NY, USA, 1999; p. 432.
6. Tsirlin, A. Averaged Optimization Methods and Their Applications (Metody usrednennoj optimizatsii i ikh prilozhenija); Fizmatlit: Moscow, Russia, 1997. (In Russian)
7. Tsirlin A. Optimal Cycles and Cyclic Modes (Optimal’nye tsikly i tsiklicheskie rezhimy); Energoatomizdat: Moscow, Russia, 1985. (In Russian)
8. Tsirlin, A.M. Conditions for Optimality of Solutions to Average Problems in Mathematical Programming. Sov. Phys. Dokl. 1992, 37, 117–119.
9. Tsirlin, A.M. The Optimal Conditions for Averaged Problems with Time-Dependent Parameters. Dokl. Math. 2000, 62, 297–299.
10. Rozonoer, L.I.; Tsirlin, A.M. Optimal Control of Thermodynamic Processes. I. Automat. Remote Control 1983, 44, 55–62.
11. Rozonoer, L.I.; Tsirlin, A.M. Optimal Control of Thermodynamic Processes. II. Automat. Remote Control 1983, 44, 209–220.
12. Rozonoer, L.I.; Tsirlin, A.M. Optimal Control of Thermodynamic Processes. III. Automat. Remote Control 1983, 44, 314–326.
13. Zevin, A.A. Optimal Control of Periodic Processes. Automat. Remote Control 1980, 41, 304–308.
14. Guardabassi, G.; Locatelli, A.; Rinaldi, S. Periodic Optimization of Continuous Systems. In Proceedings of the International Conference on Cybernetics and Society, Washington, DC, USA, 9–12 October 1972; pp. 261–263.
15. Tsirlin, A. Minimum Dissipation Processes in Irreversible Thermodynamics (Protsessy minimalnoj dissipatsii v neobratimoj termodinamike); Lan: Saint-Petersburg, Russia, 2020; p. 400. (In Russian)
16. Tsirlin, A. Optimization Methods in Irreversible Thermodynamics and Microeconomics (Metody optimizatsii v neobratimoj termodinamike i mikroekonomike); Fizmatlit: Moscow, Russia, 2003. (In Russian)
17. Novikov, I.I. The efficiency of atomic power stations (a review). J. Nucl. Energy 1958, 7, 125–128. [CrossRef]
18. Chambadal, P. Atomic Power Stations (Les centrales nucleaires); Colin: Paris, France, 1957; p. 188. (In French)
19. Curzon, F.L.; Ahlborn, B. Efficiency of a Carnot engine at maximum power output. Am. J. Phys. 1975, 43, 22–24. [CrossRef]
20. Berry, R.; Kazakov, V.; Sieniutycz, S.; Szwast, Z.; Tsirlin, A. Thermodynamic Optimization of Finite-Time Processes; Wiley: Chichester, UK, 1999.
21. Boehme, B.; Sofieva, Y.N.; Tsirlin, A.M. On the characteristic of steady state for some types of dynamic plants. Automat. Remote Control 1979, 40, 5–11.
22. Kuznetsov, A.G.; Rudenko, A.V.; Tsirlin, A.M. Optimal control in thermodynamic systems with sources of finite capacity. Automat. Remote Control 1985, 46, 20–32.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).