A two-level decomposition–aggregation approach for large-scale optimal control problems

I.S. Litvinchev*

Computing Center, Russian Academy of Sciences, Vavilov 40, Moscow 117967, Russia

Received 19 January 1993

Abstract

A new approach is proposed for the solution of large-scale constrained optimal control problems. The method is based on aggregation–disaggregation ideas and is within the class of feasible decomposition methods. It is shown that the problem of updating the disaggregation in each iteration decomposes into independent subproblems of lower dimension. If the original problem has block- or block-separable structure, then the subproblems are formulated in accordance with the blocks.

Keywords: Large-scale constrained optimal control problems; Decomposition; Aggregation; First-order methods

1. Introduction

Mathematical models of many physical and engineering systems are frequently of high dimension or possess interacting dynamic phenomena. In recent years there has been an increasing amount of research in the development of efficient techniques for solving large-scale optimal control problems [12, 5, 9, 2, 14].

Concerning the decomposition–coordination approaches [12, 2], the main idea is to decompose the original problem into a set of separate subproblems by defining a set of coordinating variables. Then a two- or multi-level structure is used for the solution. At the lower level, the set of subproblems have to be solved independently while at the higher level(s) the coordinating variables have to be updated until we reach the final convergence of the problem. The decomposition–coordination methods are usually applied to the special-structured block-separable problems, where the objective and the constraints are additively separable with respect to the block variables.

* e-mail: litv@sms.ccas.msk.su.
The main idea of aggregation [5] is to substitute the original large problem for a smaller one. In iterative aggregation algorithms [14] an iterative adjustment of the aggregated problem is used in order to obtain an optimal or suboptimal solution of the original problem.

In this paper an iterative decomposition–aggregation approach is proposed for the constrained optimal control problems. The method is intended for problems where the main difficulty is the large number of controls and hard control constraints. A typical example of such a system is a dynamical resource allocation problem, where we have relatively simple individual subsystems gathered by control (resource) constraints.

The underlying idea of the approach is as follows. The original large problem is transformed to a smaller (aggregated) one by the respective linear transformation of the variables, called aggregation. In our case the aggregation of the controls is considered. The number of aggregated variables is less than the number of original variables. Some of the coefficients of this linear aggregating transformation are treated as parameters. The aim is to choose the parameters such that the disaggregated solution of the aggregated problem will be optimal for the original one. This approach results in a two-level solution scheme: the first level solves the aggregated problem for the fixed parameters, while the second level adjusts the parameters. It is shown that if we choose the block matrix of the linear aggregation, then the problem of the adjustment of the parameters in each iteration decomposes into independent subproblems of small dimension. Generally, this decomposition depends on the structure of the aggregation, but not on the structure of the original problem. This gives the way to construct the decomposition for nonseparable problems. The aggregated problem plays here the role of the master problem. If the original problem has some special structure, then it is natural to associate the structure of the aggregation rule with the structure of the original problem. It is shown that if the original problem has block-, block-separable or block structure with the coupling variables, then the subproblems to adjust the aggregation parameters are formulated in accordance with the blocks. If the bounds of the interconnections in the original problem are known, this extra information also can be used in the subproblems. In each iteration we have the disaggregated solution feasible to the original problem. Moreover, the upper and the lower bounds of the optimal objective are calculated and the gap between these bounds tends to zero.

For the block-separable original problem, the special aggregation rule is considered, such that the aggregated problem itself decomposes into independent block subproblems. This leads to the highly decentralized method, because only independent subproblems are solved in each iteration.

This paper is considered to be a generalization of [7, 8] and it is divided as follows. In Section 2 the original problem is formulated and the aggregation–disaggregation rules are introduced. In Section 3 the optimality criteria for the disaggregated solution are formulated and the first algorithm is considered. In Section 4 the first algorithm is modified in order to utilize some special structures of the original problem. The examples are discussed in Section 5.

2. Problem formulation

Let $\mathbb{R}^n$ be the n-dimensional Euclidean space with the inner product denoted by $(x, y)$, and let $L^2_m[0, T]$ be the Hilbert space of m-dimensional vector-functions square integrable on $[0, T]$.
with the inner product denoted by \( \langle x, y \rangle \). In this paper we consider the following optimal control problem: Find \( u(t) \in L^2_n[0, T] \) such that

\[
J(u(\cdot)) = Q(x(T)) + \int_0^T f(x, u, t) \, dt \to \max
\]

subject to

\[
\begin{align*}
\dot{x}(t) &= A(t)x(t) + b(u, t), \quad x(0) = x_0, \\
v(x(T)) &\leq 0, \\
p(x, u, t) &\leq 0, \quad u(t) \geq 0,
\end{align*}
\]

where \( x \in \mathbb{R}^N, u \in \mathbb{R}^n; b(\cdot, \cdot), v(\cdot) \) and \( p(\cdot, \cdot, \cdot) \) are \( N-, V- \) and \( R- \) dimensional vector-functions, respectively; \( A(t) \) is a square \( N \times N \) real matrix.

Throughout this paper we assume that the original problem (2.1)-(2.4) can be converted to a concave programming problem in the Hilbert space, such that the saddle point and the Kuhn–Tucker theorems hold. For example, one may use Ter-Krikorov [13] conditions, which guarantee this property:

(a) \(- p(x, u, t), - b(u, t) \) and \( f(x, u, t) \) are continuously differentiable and concave with respect to \( x, u \) and monotonously increasing with respect to \( x \);

(b) \(- v(\cdot) \) and \( Q(\cdot) \) are continuously differentiable, concave and monotonously increasing with respect to \( \gamma \);

(c) the Slater conditions for (2.3), (2.4) hold;

(d) \( x(t) \geq \varepsilon > 0, p(0, 0, t) \leq 0 \).

Note that under these conditions the maximum principle for the original problem holds [13].

**Remark 2.1.** In the literature one can find the other conditions to guarantee the optimality in the Kuhn–Tucker form [1, 10]. In this paper we do not try to state the respective conditions in the most general or the simplest form. To formulate the results given below, it is essential only that one can write the optimality conditions in the Kuhn–Tucker form. Note that if the original problem is not mixed constrained, but has the separate control and (or) state constraints, then the respective conditions are simplified [4].

A vector \( u(t) \) is broken up into \( I \) (\( 1 \leq I \leq n \)) disjoint subvectors \( u_i(t), i = 1, 2, \ldots, I \), with \( J_i \) components in the \( i \)th subvector:

\[
u(t) = (u_1(t) \ldots u_i(t) \ldots u_J(t)), \quad u_i(t) = (u_1^i(t) \ldots u_j^i(t) \ldots u_{J_i}^i(t)),
\]

\[
\sum_i J_i = n.
\]

We introduce a vector of the aggregated controls \( U(t) = (U^1(t) \ldots U^I(t) \ldots U^I(t)) \) and a linear disaggregation \( u = \alpha U \) or in detailed form

\[
u_j^i(t) = \alpha_j^i(t) U^i(t), \quad i = 1, 2, \ldots, I, \quad j = 1, 2, \ldots, J_i.
\]

Here the disaggregation parameters \( \alpha_j^i(t) \) satisfy the following condition:

\[
\alpha \in A = \{ \alpha_j^i(t): \sum_j \alpha_j^i(t) = 1, \quad \alpha_j^i(t) \geq 0, \quad i = 1, 2, \ldots, I, \quad j = 1, 2, \ldots, J_i, \quad t \in [0, T] \}.
\]

Unless otherwise specified, the symbol \( \sum \) denotes the summation from \( j = 1 \) to \( j = J_i \).
Note that if \( z \in A \), then it follows from (2.5), (2.6) that \( U_i(t) = \sum u'_j(t), \, i = 1, 2, \ldots, I \). Moreover, if \( z \in A \) and \( U \geq 0 \), then the disaggregated control \( u = zU \geq 0 \).

Fixing \( z \in A \) and substituting (2.5) into (2.1)-(2.4), we get the aggregated problem

\[
\bar{J}(z(\cdot), U(\cdot)) = Q(x(T)) + \int_0^T f(x, zU, t) \, dt \to \max,
\]

\[
\dot{x}(t) = A(t)x(t) + b(zU, t), \quad x(0) = x_0, \quad (2.7)
\]

\[
v(x(T)) \leq 0,
\]

\[
p(x, zU, t) \leq 0, \quad U(t) \geq 0.
\]

The aggregated problem (2.7) has only \( I \) macrocontrols \( U_i(t), \, i = 1, 2, \ldots, I \), instead of \( n = \sum_i J_i \) controls \( u'_j(t) \) in the original problem.

Denote by \( P \) the set of all admissible controls of the original problem, \( P(z) \) the set of all optimal controls of the aggregated problem (2.7) and let \( \theta(z) \) be the extremal-value functional of (2.7) depending on the choice of \( z \in A \).

**Assumption 2.2.** There exists \( \varepsilon > 0 \) such that for all \( z \in A \) the aggregated problem has an optimal control \( \hat{U}(t) \in P(z) \) such that \( \hat{U}_i(t) \geq \varepsilon, \, i = 1, 2, \ldots, I, \, t \in [0, T] \).

Obviously, if \( \hat{U} \) is an optimal control of (2.7) for some \( z \in A \), then the disaggregated control \( \hat{u} = z\hat{U} \) is a feasible control of the original problem. In the following we shall construct an iterative adjustment process for the aggregated problem (2.7) in order to find \( z \in A \) such that if \( \hat{U} \in P(z) \), then the disaggregated control \( \hat{u} = z\hat{U} \) is an optimal control of the original problem. Note that if \( \hat{u} \) is an optimal control of (2.1)-(2.4) and \( \sum \hat{u}_j(t) > 0, \, i = 1, 2, \ldots, I, \, t \in [0, T] \), such an \( \hat{z} \) always exists. For example, we may put \( \hat{u}_j(t) = \hat{u}_j(t) [\sum \hat{u}_j(t)]^{-1}, \, i = 1, 2, \ldots, I, \, j = 1, 2, \ldots, J_i \).

For our aggregated problem (2.7) we consider the dual problem

\[
Q(x(T)) + \int_0^T [f(x, zU, t) + b(zU, t) + A(t)x(t) - \dot{x}(t), z(t)] \, dt - (v(x(T)), \mu) \to \min,
\]

\[
\dot{\lambda}(t) = A^T(t)\lambda(t) - \frac{\partial p(x, zU, t)}{\partial x} \eta(t) + \frac{\partial f(x, zU, t)}{\partial x},
\]

\[
\lambda_i(t) = \left( \lambda(t), \frac{\partial b(zU, t)}{\partial U^i} \right) - \left( \eta(t), \frac{\partial p(x, zU, t)}{\partial U^i} \right) + \frac{\partial f(x, zU, t)}{\partial U^i} = 0, \quad i = 1, 2, \ldots, I, \quad (2.8)
\]

\[
\eta(t) \geq 0, \quad \mu \geq 0, \quad U(t) \geq 0, \quad t \in [0, T].
\]

Here \( \lambda(t) \) are the costate variables, \( \eta(t) \) and \( \mu \) are the Lagrange multipliers for the mixed and terminal constraints in (2.7), respectively. We have \( \lambda_i(t) = 0 \) in (2.8) since \( \hat{U}(t) > 0 \) from Assumption 2.2. Denote by \( D(z) \) the set of all optimal solutions of the dual problem (2.8).
3. The optimality criteria and the algorithm

The following theorem establishes the optimality criterion for the disaggregated control in terms of the aggregated problem.

**Theorem 3.1.** Let $\mathbf{z} \in A$, $\mathbf{U} \in P(\mathbf{z})$ and $\mathbf{\hat{z}}$ be the respective vector of the state variables in (2.7). The disaggregated control $\mathbf{\hat{u}} = \mathbf{\hat{z}} \mathbf{U}$ is an optimal control of the original problem (2.1)–(2.4) if and only if there exists a triplet $\mathbf{\hat{z}}(t), \mathbf{\hat{U}}(t), \mathbf{\hat{u}}$ such that the element $[\mathbf{\hat{z}}, \mathbf{\hat{U}}, \mathbf{\hat{z}}, \mathbf{\hat{U}}, \mathbf{\hat{u}}] \in D(\mathbf{\hat{z}})$ and

$$\mathbf{A}_{ij}(t) = \left[ \begin{array}{c} \mathbf{z}(t), \frac{\partial b(u, t)}{\partial u_i} \\ \frac{\partial p(x, u, t)}{\partial u_j} \end{array} \right] + \frac{\partial f(x, u, t)}{\partial u_j} \right|_{\mathbf{\hat{z}}, \mathbf{\hat{u}}} \leq 0,$$

$$i = 1, 2, \ldots, I, \quad j = 1, 2, \ldots, J_i, \quad t \in [0, T].$$

**Proof.** Let (3.1) be satisfied. By straightforward calculation we obtain $\mathbf{A}_{ij}(t) = \mathbf{A}_{ij}(t)|_{\mathbf{\hat{z}}, \mathbf{\hat{U}}, \mathbf{\hat{u}}, \mathbf{\hat{z}}, \mathbf{\hat{U}}, \mathbf{\hat{u}}}$. By the Kuhn-Tucker complementary slackness conditions for the aggregated problem we have

$$0 = \mathbf{U}(t) \mathbf{A}(t) = \mathbf{U}(t) \sum \mathbf{z}_i(t) \mathbf{A}_{ij}(t) = \sum \mathbf{u}_i(t) \mathbf{A}_{ij}(t),$$

$$0 = (p(\mathbf{z}, \mathbf{u}(t), \mathbf{z}), \mathbf{\hat{U}}(t), \mathbf{\hat{u}}(t)).$$

Moreover, since $\mathbf{u}_i(t) \geq 0$ and $\mathbf{A}_{ij}(t) \leq 0$, it follows from $\sum \mathbf{u}_i(t) \mathbf{A}_{ij}(t) = 0$ that $\mathbf{u}_i(t) \mathbf{A}_{ij}(t) = 0$, $i = 1, 2, \ldots, I, j = 1, 2, \ldots, J_i, t \in [0, T]$. Note that if we consider the dual problem to (2.1)–(2.4), then (3.1) is the restriction of this dual problem, calculated for $\mathbf{\hat{z}}, \mathbf{\hat{U}}, \mathbf{\hat{u}}$.

Thus, we have that a pair $\mathbf{\hat{z}}, \mathbf{\hat{u}}$ is a feasible solution of the original problem, the element $[\mathbf{\hat{z}}, \mathbf{\hat{U}}, \mathbf{\hat{z}}, \mathbf{\hat{U}}, \mathbf{\hat{u}}]$ is a feasible solution of its dual and the complementarity conditions are satisfied. Then it follows by the Kuhn-Tucker theorem [13] that $\mathbf{\hat{u}}$ is an optimal control of the original problem.

The other part of the theorem is obvious, because if $\mathbf{\hat{u}}$ is an optimal control of the original problem with $\sum \mathbf{u}_i(t) > 0$ for all $i$, then $\mathbf{U}(t) = \sum \mathbf{u}_i(t), i = 1, 2, \ldots, I$, is an optimal control of the aggregated problem with $\mathbf{z}_i(t) = \mathbf{u}_i(t)/\mathbf{U}(t)$. Moreover, (3.1) is satisfied since it is a part of the restrictions of the dual to the original problem.

In order to find $\mathbf{\hat{z}} \in A$ such that the conditions of Theorem 3.1 are satisfied, we shall propose an iterative procedure. In each iteration of this procedure the aggregated problem is solved and then the disaggregation parameters are updated. First of all we investigate some extremal properties of $\mathbf{\hat{z}}$, defined in Theorem 3.1.

Introduce the auxiliary problem

$$\max \{ \theta(x) | x \in A \}.$$  \hspace{1cm} (3.2)

Obviously, if $\mathbf{\hat{z}}$ is the extremal solution of (3.2) and $\mathbf{\hat{U}} \in P(\mathbf{\hat{z}})$, then the disaggregated control $\mathbf{\hat{u}} = \mathbf{\hat{z}} \mathbf{\hat{U}}$ is an optimal control of the original problem. The auxiliary problem (3.2) has a very simple feasible set, but $\theta(x)$ is not a concave functional in general, so it is rather difficult to find the global maximum in (3.2). We shall show later that under some differentiability assumptions, the necessary conditions of optimality are also sufficient for the auxiliary problem.
Assumption 3.2. (a) \( P(\alpha) \) and \( D(\alpha) \) are uniformly bounded for \( \alpha \in A \).
(b) \( f, Q, b, v \) and \( p \) have the first-order partial derivatives with respect to \( x \) and \( u \), satisfying the Lipschitz condition for every \( t \in [0, T] \). Moreover, these derivatives are uniformly bounded for \( t \in [0, T] \) if \( x \) and \( u \) belong to any bounded subset of \( \mathbb{R}^N \times \mathbb{R}^n \).

It is not hard to see that under this assumption the marginal value theorem [6] holds for the family of the aggregated problems, depending on the choice of \( \alpha \in A \). For simple consideration we restrict ourselves in this paper only to the case of a unique optimal solution of the aggregated problem and its dual for all \( \alpha \in A \).

If \( P(\alpha) \) and \( D(\alpha) \) are singletons for all \( \alpha \in A \), then \( \theta(\alpha) \) is a Fréchet-differentiable functional, such that

\[
\nabla_2 \theta(\tilde{\alpha}) = \nabla_2 \mathcal{L}(\tilde{x}, \tilde{U}, \tilde{\lambda}, \tilde{\eta}, \tilde{\mu}),
\]

where \( \mathcal{L} \) is the standard Lagrange functional associated with the aggregated problem. In our case \( \mathcal{L} \) coincides with the objective functional of the dual aggregated problem (2.8). By straightforward calculation we get

\[
\nabla \theta(\tilde{\alpha}) = \begin{cases} \tilde{\zeta}_{ij}^t(t), & i = 1, 2, \ldots, I, j = 1, 2, \ldots, J_i \\ \tilde{\zeta}_{ij}^t(t) = \tilde{U}^i(t)\tilde{A}_{ij}^i(t), \end{cases}
\]

where \( \tilde{A}_{ij}^i(t) \) has been defined in (3.1).

The following theorem shows that, although \( \theta(\alpha) \) is not a concave functional in general, the first-order necessary conditions of optimality are also sufficient for (3.2).

**Theorem 3.3.** Let \( \tilde{\alpha} \in A \) be a stationary point of (3.2), i.e., \( \langle \nabla_2 \theta(\tilde{\alpha}), \alpha - \tilde{\alpha} \rangle \leq 0 \) for all \( \alpha \in A \). This stationarity condition can be reformulated in the following way:

\[
S(\tilde{\alpha}) = \max_{\alpha} \left\{ \sum_1^I \sum_1^{J_i} \int_0^T \tilde{U}^i(t)\tilde{A}_{ij}^i(t)\tilde{a}_{ij}^i(t)dt \mid \alpha \in A \right\} = 0.
\]

Let Assumptions 2.2 and 3.2 hold. Then the disaggregated control \( \tilde{u} = \tilde{\alpha}\tilde{U} \) is an optimal control of the original problem.

**Proof.** By Theorem 3.1 we have that \( \tilde{u} \) is an optimal control of the original problem iff (3.1) holds. Suppose that \( S(\tilde{\alpha}) = 0 \), but (3.1) is not satisfied. Then there exists a pair \((i_0, j_0)\) and a subinterval \([\tau_1, \tau_2] \subseteq [0, T]\) such that \( \tilde{A}_{ij}^i(t) > 0 \) for a.a. \( t \in [\tau_1, \tau_2] \). From the dual aggregated problem we have \( \tilde{A}^i(t) = \sum\tilde{A}_{ij}^i(t) = 0, \) \( i = 1, 2, \ldots, I. \) Since \( \sum\tilde{A}_{ij}^i(t) = 1, \) it follows that \( \max_j \tilde{A}_{ij}^i(t) > 0 \) for every \( i = 1, 2, \ldots, I \) and \( t \in [0, T] \). By our assumption \( \max_j \tilde{A}_{ij}^i(t) > 0 \) for a.a. \( t \in [\tau_1, \tau_2] \).

Let \( \tilde{A}_{ij}^i(t) = \max_j \tilde{A}_{ij}^i(t). \) Define \( \tilde{a}_{ij}^i(t) \) such that

\[
\tilde{a}_{ij}^i(t) = \begin{cases} 1 & \text{if } j = j_i(t), \\ 0 & \text{if } j \neq j_i(t), \end{cases}
\]

Obviously, \( \tilde{\alpha} \in A \) and from the definition of \( S(\tilde{\alpha}) \) we have

\[
S(\tilde{\alpha}) = \sum_1^I \left\{ \int_0^T \tilde{U}^i(t)\tilde{A}_{ij}^i(t)\tilde{a}_{ij}^i(t)dt \right\} = \sum_1^I \left\{ \int_0^T \tilde{U}^i(t)\tilde{A}_{ij}^i(t, j_i(t))dt \right\} = I_1 + I_2 + I_3,
\]
where
\[ I_1 = \sum_1^T \int_0^{T_1} \tilde{U}'(t) \tilde{A}'_{(t)}(t) \, dt, \quad I_2 = \sum_1^T \int_0^{T_2} \tilde{U}'(t) \tilde{A}'_{(t)}(t) \, dt, \quad I_3 = \sum_1^T \tilde{U}'(t) \tilde{A}'_{(t)}(t) \, dt. \]

Since \( \tilde{U}'(t) > 0 \), by the definition of \( j_i(t) \) we have \( I_1 \geq 0, I_3 \geq 0 \). Moreover, by our assumption \( \tilde{A}'_{(t)} > 0 \) for a.a. \( t \in [\tau_1, \tau_2] \) and hence \( I_2 > 0 \). Therefore we have \( S(\hat{x}) > 0 \) which contradicts (3.4).

Thus, if \( S(\hat{x}) = 0 \), then it follows that (3.1) is satisfied and hence by Theorem 3.1 that \( \hat{u}(t) \) is an optimal control of the original problem. \[ \square \]

It follows from Theorem 3.3 that if you find a stationary point of (3.2), solve the aggregated problem and find the disaggregated control, then you get an optimal control of the original problem.

In order to find a stationary point of (3.2), one may use the first-order methods, such as feasible directions, projected gradient, conditional gradient methods and so on [15]. The main subproblem arising in these methods is to find the direction of the ascent. In order to find this direction, one has to maximize an additively separable function (quadratic in projected gradient or linear in the other methods) on \( A \).

In our case this problem decomposes into \( I \) independent subproblems due to the special structure of the set \( A: A = A_1 \times A_2 \times \cdots \times A_I \), where
\[ A_i = \{ x_i(t) | \sum x_i(t) = 1, x_i(t) \geq 0, j = 1, 2, \ldots, J_i, t \in [0, T] \}. \]

Describe, for example, one iteration of the conditional gradient method. Let \( \hat{x} \in A \) be given from the previous iteration.

(S1) Solve the aggregated problem (2.7) and its dual for \( \alpha = \hat{x} \) and find their optimal solutions \( \hat{x}, \hat{U} \) and \( \hat{x}, \hat{U}, \hat{A} \). Let \( \hat{u} = \hat{x}\hat{U} \).

(S2) Compute \( V_\alpha \theta(\hat{x}) \) in (3.3). Solve \( I \) independent subproblems
\[ \max_{z_i} S_i(z_i) = \max_{z_i} \left\{ \int_0^T \sum \tilde{U}'(t) \tilde{A}'_{(t)}(t) \, dt \mid z_i \in A_i \right\}. \]

Let \( \tilde{x}(t) \) be an optimal solution of these subproblems. If \( \tilde{x} = \hat{x} \) or \( S_i(\tilde{x}_i) = 0, i = 1, 2, \ldots, I \), then stop. Otherwise go to (S3) with \( \tilde{s} = \tilde{x} - \hat{x} \).

(S3) Find \( \tilde{\rho} \): \( \theta(\tilde{x} + \tilde{\rho}s) = \max \{ \theta(\tilde{x} + \rho\tilde{s}) | 0 \leq \rho \leq 1 \} \). Set \( \alpha = \tilde{x} + \tilde{\rho}s \) and go to (S1).

Of course, one may use in (S3) another appropriate rule to choose the step-length \( \tilde{\rho} \).

Note that since there are no differential equations in the definition of \( A \), we get in (S2) the independent subproblems for each \( t \in [0, T] \), too.

The original problem has \( n = \sum_i J_i \) controls. In the proposed decomposition-aggregation method we deal with the problems of lower dimension: the aggregated problem in (S1) has \( I \) controls, each of \( I \) independent subproblems in (S2) has \( J_i, i = 1, 2, \ldots, I \), unknown functions. In order to calculate \( \theta(\tilde{x} + \rho \tilde{s}) \) for some \( 0 \leq \rho \leq 1 \) in (S3) one has to solve the aggregated problem for \( \alpha = \tilde{x} + \rho \tilde{s} \).

The convergence of the first-order methods to a stationary point was established, for example, in [15], where the gradient of the objective functional was assumed to be Lipschitz continuous. It holds for \( V_\alpha \theta(\alpha) \) in (3.3) if Assumptions 2.2 and 3.2 are satisfied and \( P(\alpha), D(\alpha) \) are singletons.
In the proposed technique we did not use any structural properties of the original problem, such as separability, block-diagonal structure of the constraints and so on. The decomposition properties arise due to the special structure of the disaggregation, given by (2.5), (2.6).

However, in many practical problems we have complex systems which possess some special structures. For example, in the resource allocation problems we often deal with the number of individual dynamical subsystems, dividing a common resource. In this case the main difficulty is due to the large dimension of the control vector and hard control constraints. In the following section we shall show how to use some structural properties of the constraints and the bounds of interconnection in the decomposition-aggregation scheme.

4. Decomposition-aggregation technique for particular classes of the original problem

First of all we modify the problem of finding the direction of the ascent in (S2) in order to use some structural properties of the original problem.

Denote by $\Omega$ the subset of $L^2_a[0, T]$ such that the following assumption is satisfied:

**Assumption 4.1.** (a) $\Omega$ is a closed bounded set;
(b) $\Omega \supseteq P$, where $P$ is the set of all feasible controls of the original problem (2.1)–(2.4).

Consider the problem

$$
\delta(\hat{x}) = \max_u \left\{ \int_0^T \sum_i \sum_j u_i^j(t) \hat{A}_i^j(t) dt \mid u(t) \in \Omega, t \in [0, T] \right\},
$$

where $\hat{x} \in A$ and $\hat{A}_i^j(t)$ has been defined in (3.1).

We assume that there exists an optimal solution $\hat{u}(t)$ of (4.1) such that $\hat{U}_i^j(t) = \sum \hat{A}_i^j(t) > 0$, $i = 1, 2, \ldots, I, t \in [0, T]$. Define

$$
\hat{a}_i^j(t) = \hat{A}_i^j(t)/\hat{U}_i^j(t), \quad \hat{s}_i^j(t) = \hat{U}_i^j(t)[\hat{A}_i^j(t) - \hat{x}_i^j(t)]/\hat{U}_i^j(t),
$$

$i = 1, 2, \ldots, I, \quad j = 1, 2, \ldots, J_i$.

The following theorem establishes some properties of the direction $\hat{s}(t) = \{\hat{s}_i^j(t)\}$.

**Theorem 4.2.** Let Assumptions 2.2, 3.2 and 4.1 hold. Then:

(a) If $\delta(\hat{x}) > 0$, then there exists $\rho > 0$ such that $\theta(\hat{x}) < \theta(\hat{x} + \rho \hat{s})$, i.e., $\hat{s}$ is the direction of the ascent for $\theta(\hat{x})$ in $x = \hat{x}$;

(b) If $\delta(\hat{x}) = 0$, then $\hat{u} = \hat{x} \hat{U}$ is an optimal control of the original problem (2.1)–(2.4);

(c) For all $\hat{x} \in A$ there exists $M > 0$, such that $S(\hat{x}) \geq M \delta(\hat{x})$, where $S(\hat{x})$ has been defined in (3.4).

**Proof.** From the complementary slackness conditions for the aggregated problem we have

$$
0 = \hat{U}_i^j(t) \sum \hat{a}_i^j(t) \hat{A}_i^j(t) = \sum \hat{A}_i^j(t) \hat{A}_i^j(t), \quad i = 1, 2, \ldots, I.
$$

Since $\hat{u}(t) \in P$, then according to Assumption 4.1, $\hat{u}(t) \in \Omega$ and hence $\delta(\hat{x}) \geq 0$. 

Let us denote $\theta(\rho) = \theta(\dot{x} + \rho \dot{s})$. From Assumptions 2.2 and 3.2 and since $P(\dot{x})$ and $D(\dot{x})$ are singletons, we have that the marginal value theorem holds and therefore

$$
(\theta(0))' = \frac{\partial \mathcal{L}(\rho)}{\partial \rho} \bigg|_{\rho=0},
$$

where $\mathcal{L}(\rho)$ is the standard Lagrange functional of the aggregated problem (2.7) for $x = \dot{x} + \rho \dot{s}$. By straightforward calculations we obtain

$$
(\theta(0))' = \int_0^T \sum_i \sum_j \frac{\hat{U}_i(t)}{\hat{U}_i(t)} \left[ \dot{\mathcal{L}}_i^j(t) - \mathcal{L}_i^j(t) \right] \dot{\mathcal{L}}_i^j(t) \dot{\mathcal{L}}_i^j(t) dt
$$

$$
= \int_0^T \sum_i \sum_j \ddot{u}_i^j(t) \ddot{\mathcal{L}}_i^j(t) dt = \delta(\dot{x}).
$$

Recalling the definition of $(\theta(0))'$ we get the first part of the theorem.

To prove the second part of the theorem, assume that $\delta(\dot{x}) = 0$, but the disaggregated control $\dot{u}$ is not optimal for the original problem. Then by Theorem 3.1 there exist a pair $(i_0, j_0)$ and a subinterval $[\tau_1, \tau_2] \subseteq [0, T]$, such that $\mathcal{L}_i^{j_0}(t) > 0$ for a.a. $t \in [\tau_1, \tau_2]$. Construct $\dot{x} \in A$ as in the proof of Theorem 3.3. From the definition of $\dot{x}$ we have $\sum \ddot{x}_i(t) \ddot{\mathcal{L}}_i^j(t) > 0$ for a.a. $t \in [\tau_1, \tau_2]$ and $\sum \ddot{x}_i(t) \ddot{\mathcal{L}}_i^j(t) > 0$ for $i \neq i_0$, $t \in [0, T]$). Under Assumption 2.2 there exists $\tilde{U} > 0$ such that $\dot{u} = \dot{x} \tilde{U} \in P$ and therefore from Assumption 4.1 we have $\dot{u} \in \Omega$. Since $\tilde{U} > 0$,

$$
0 < \int_0^T \sum_i \ddot{u}_i(t) \ddot{\mathcal{L}}_i^j(t) dt = \int_0^T \sum_i \sum_j \ddot{u}_i^j(t) \ddot{\mathcal{L}}_i^j(t) dt \leq \delta(\dot{x}),
$$

and this contradicts $\delta(\dot{x}) = 0$. Hence, if $\delta(\dot{x}) = 0$, then (3.1) is satisfied and by Theorem 3.1 $\dot{u}(t)$ is an optimal control of the original problem.

To prove the last part of Theorem 4.2, we write the following sequence of inequalities:

$$
S(\dot{x}) = \max_{\dot{z} \in A} \int_0^T \sum_j \ddot{z}_j(t) \ddot{\mathcal{L}}_j^i(t) dt = \sum_{i \in A_i} \max_{\dot{z} \in A_i} \int_0^T \ddot{z}_j(t) \ddot{\mathcal{L}}_j^i(t) dt
$$

$$
= \sum_{i \in A_i} \int_0^T \frac{\hat{U}_i(t)}{\hat{U}_i(t)} \hat{U}_i(t) \sum_j \ddot{z}_j(t) \ddot{\mathcal{L}}_j^i(t) dt
$$

$$
= \sum_{i \in A_i} \int_0^T \frac{\hat{U}_i(t)}{\hat{U}_i(t)} \max_{\dot{z} \in A_i} \hat{U}_i(t) \ddot{z}_j(t) \ddot{\mathcal{L}}_j^i(t) dt
$$

$$
\geq \left( \min_{i \in A_i} \frac{\hat{U}_i(t)}{\hat{U}_i(t)} \right) \int_0^T \sum_{i \in A_i} \max_{\dot{z} \in A_i} \sum_j \ddot{z}_j(t) \ddot{\mathcal{L}}_j^i(t) dt
$$

$$
\geq \frac{\epsilon}{D} \int_0^T \sum_i \ddot{u}_i(t) \ddot{\mathcal{L}}_i^j(t) dt = \frac{\epsilon}{D} \delta(\dot{x}),
$$

where $\hat{U}_i(t) \leq D$ for $i = 1, 2, \ldots, I$, $t \in [0, T]$, and such $D > 0$ exists since Assumption 4.1 is satisfied. We can place the operator $\max_{\dot{z} \in A_i}$ into the integrand since the condition $\dot{z} \in A_i$ is formulated independently for each $t \in [0, T]$. The first inequality in this sequence holds since $\max \{ \sum_j \ddot{z}_j(t) \ddot{\mathcal{L}}_j^i(t) | \dot{z} \in A_i \} \geq 0$ for each $i = 1, 2, \ldots, I$, $t \in [0, T]$. The last inequality holds since $\dot{x} \in A$. \( \square \)
It follows from Theorem 4.2 that if the problem (4.1) is solved, then one can either state the optimality of the disaggregated solution or construct the direction of the ascent $\delta$.

Consider the following iterative process: 

$$a_{k+1} = a_k + \rho_k \delta_k,$$

where $k$ is the number of iterations, $a_k \in A$, $\delta_k$ has been defined in Theorem 4.2 and $\rho_k$ is such that $\theta(a_k) < \theta(a_{k+1})$. It was established in [11] that if the conditions (a)-(c) of Theorem 4.2 are satisfied, then the iterative process, described above, converges to a stationary point of (3.2). Thus, in the algorithm described in the second section, we may use the modified step (S2') instead of (S2):

(S2') Compute $V\theta(\hat{x})$ in (3.3). Solve (4.1) and find its optimal solution $\hat{u}$. Let $\hat{U}^i(t) = \sum_{j} \hat{u}_{ij}^i(t) > 0$, $i = 1, 2, ..., I$, $t \in [0, T]$. Define $\hat{s}_{ij}^i(t) = \hat{U}(t)[\hat{u}_{ij}^i(t) - \hat{u}_{ij}^i(t)]/\hat{U}^i(t)$, $\hat{s}_{ij}^i(t) = \hat{u}_{ij}^i(t)/\hat{U}^i(t)$, $i = 1, 2, ..., I$, $j = 1, 2, ..., J_i$. If $\delta(\hat{x}) = 0$, then stop, otherwise go to (S3) with $s = \delta$.

Note that in each iteration we have the disaggregated control $\hat{u}_k$, feasible for the original problem, and thus our algorithm belongs to the class of so-called "feasible methods" [12]. Moreover, if the optimality criterion is not satisfied, then $J(\hat{u}_k) = \theta(\hat{x}_k) < \theta(\hat{x}_{k+1}) = J(\hat{u}_{k+1})$, i.e., the objective functional of the original problem increases monotonously.

Below we shall consider some particular classes of the original problem where we can construct $\Omega$ such that (4.1) decomposes into independent subproblems. In this context, note that if $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_L$, then due to the additive-separability of the goal functional of (4.1), this problem decomposes into $L$ low-order subproblems. Since the main property of $\Omega$ in Assumption 4.1 is $\Omega \subseteq P$, we may use in the definition of $\Omega$ some constraints of the original problem.

Now, let us modify the objective functional of the subproblem (4.1) in order to handle the block-separable formulation. For this, we shall use the inequality

$$\varphi(x, u, t) + \left(\nabla_x \varphi(x, u, t), x - \tilde{x}\right) + \left(\nabla_u \varphi(x, u, t), u - \tilde{u}\right) - \varphi(x, u, t) \geq 0$$

which holds for a concave differentiable scalar function $\varphi(x, u, t)$. Denote by $\nabla_x p(x, u, t)$, $\nabla_u p(x, u, t)$ and $\nabla_u b(u, t)$ the matrices of the respective partial derivatives. Then we can derive the following sequence of inequalities:

$$\int_0^T (u(t), \dot{\hat{x}}(t)) \, dt \geq \int_0^T \left[ f(x, u, t) - f(\hat{x}, \hat{u}, t) - \left(\nabla_x f(\hat{x}, \hat{u}, t), x - \hat{x}\right) + \left(\nabla_u f(\hat{x}, \hat{u}, t), \hat{u}\right) \right] \, dt$$

$$+ \int_0^T \left( \dot{\hat{x}}(t), b(u, t) - b(\hat{u}, t) + \nabla_u b(\hat{u}, t)\hat{u}(t) \right) \, dt$$

$$- \int_0^T \left( \hat{u}(t), p(x, u, t) - p(\hat{x}, \hat{u}, t) - \nabla_x p(\hat{x}, \hat{u}, t)(x - \hat{x}) + \nabla_u p(\hat{x}, \hat{u}, t)\hat{u}(t) \right) \, dt$$

$$= \left\{ \int_0^T (\nabla_u f(\hat{x}, \hat{u}, t) + \hat{\varphi}(t), \nabla_u b(\hat{u}, t) - \hat{\varphi}(t) \nabla_u p(\hat{x}, \hat{u}, t) \hat{u}(t)) \, dt \right\}$$

$$+ \left\{ \int_0^T (\hat{\varphi}(t), p(\hat{x}, \hat{u}, t)) \, dt \right\}$$

$$- \int_0^T \left[ (\nabla_x f(\hat{x}, \hat{u}, t), (x - \hat{x})) - \hat{\varphi}(t) \nabla_x p(\hat{x}, \hat{u}, t)(x - \hat{x}) - (\hat{\varphi}(t), b(u, t)) \right.$$

$$- b(\hat{u}(t)) \right] \, dt + \int_0^T \left[ f(x, u, t) - f(\hat{x}, \hat{u}, t) - (\hat{\varphi}(t), p(x, u, t)) \right] \, dt.$$
From the complementarity conditions for the aggregated problem and its dual we have that the terms in the figured braces are equal to zero.

By the differential equation for the costate variable $\lambda(t)$ in (2.8) we have the equality

$$-(\frac{d\tilde{x}(t)}{dt}, x - \hat{x}) = (A^T(t)\lambda(t), x - \hat{x}) - \tilde{h}(t) \mathbf{V}_x p(\hat{x}, \hat{u}, t)(x - \hat{x}) + (\mathbf{V}_x f(\hat{x}, \hat{u}, t), x - \hat{x}).$$

Let now a pair $x(t), u(t)$ be such that $\hat{x}(t) = A(t)x(t) + b(u, t), x(0) = x_0$. Then, integrating by parts the above equality, we get

$$\int_0^T \left[ (\mathbf{V}_x f(\hat{x}, \hat{u}, t), x - \hat{x}) - \tilde{h}(t) \mathbf{V}_x p(\hat{x}, \hat{u}, t)(x - \hat{x}) - (\lambda(t), b(u, t) - b(\hat{u}, t)) \right] dt = - (\hat{\lambda}(T), x(T) - \hat{x}(T)).$$

Substituting this expression in (4.2) and using the equality for $\hat{\lambda}(T)$ from the dual aggregated problem (2.8) we have

$$\int_0^T (u(t), \hat{A}(t)) dt \geq \int_0^T \left[ f(x, u, t) - f(\hat{x}, \hat{u}, t) - (\tilde{h}(t), p(x, u, t)) \right] dt$$

$$+ (\mathbf{V}_x Q(\hat{x}(T))) - \tilde{h} \mathbf{V}_x v(\hat{x}(T)), x(T) - \hat{x}(T)) \geq Q(x(T)) - Q(\hat{x}(T)) - (\tilde{\mu}, v(x(T)) - v(\hat{x}(T)))$$

$$+ \int_0^T \left[ f(x, u, t) - f(\hat{x}, \hat{u}, t) - (\tilde{h}(t), p(x, u, t)) \right] dt$$

$$= Q(x(T)) - (\tilde{\mu}, v(x(T))) + \int_0^T \left[ f(x, u, t) - (\tilde{h}(t), p(x, u, t)) \right] dt$$

$$- [Q(\hat{x}(T)) + \int_0^T f(\hat{x}, \hat{u}, t) dt]$$

(4.3)

for any $x(t), u(t)$ such that $\hat{x}(t) = A(t)x(t) + b(u, t), x(0) = x_0$. Note that the last term in brackets in (4.3) equals $\theta(\hat{x})$.

Let now the vectors $v(\cdot)$ and $p(\cdot)$ be broken up into disjoint subvectors: $v(\cdot) = \{v_1(\cdot), v_2(\cdot)\}$ and $p(\cdot) = \{p_1(\cdot), p_2(\cdot)\}$. Let $\Omega$ be the set of controls such that the following conditions are satisfied:

$$\hat{x}(t) = A(t)x(t) + b(u, t), \quad x(0) = x_0,$$

$$p_2(x, u, t) \leq 0, \quad v_2(x(T)) \leq 0, \quad u(t) \geq 0.$$

(4.4)

Obviously, $\Omega \supseteq P$, where $P$ is the set of feasible controls of the original problem (2.1)-(2.4).

Then, since $\tilde{\mu} \geq 0$, $\tilde{h}(t) \geq 0$, we have by (4.3), (4.4) that

$$\int_0^T (u(t), \hat{A}(t)) dt \geq Q(x(T)) - (\tilde{\mu}, v_1(x(T)))$$

$$+ \int_0^T \left[ f(x, u, t) - (\tilde{h}_1(t), p_1(x, u, t)) \right] dt - \theta(\hat{x}),$$

(4.5)

where $\tilde{\mu}_1$ and $\tilde{h}_2(t)$ are respective parts of the Lagrange multipliers.
Recalling the definition of $\delta(\dot{x})$ in (4.1), we have for $\Omega$ in (4.4):

$$
\delta(\dot{x}) \geq \max \left\{ Q(x(T)) - (\mu_1, v_1(x(T))) + \int_0^T \left[ f(x, u, t) - (\eta_1(t), p_1(x, u, t)) \right] dt \right\}
$$

$$
= \max \left\{ \dot{x}(t) = A(t)x(t) + b(u, t), \quad x(0) = x_0, \right. \\
\left. p_2(x, u, t) \leq 0, \quad v_2(x(T)) \leq 0, \quad u(t) \geq 0 \right\} - \theta(\dot{x}) = \pi(\dot{x}).
$$

(4.6)

The following theorem shows that in order to construct the direction of the ascent in the step (S2') of the modified algorithm, we can use the extremal problem in the right-hand side of (4.6).

**Theorem 4.3.** Let $\hat{x}(t), \hat{u}(t)$ be the optimal solution of the extremal problem in (4.6), such that $\hat{U}^i(t) = \sum \hat{u}_i^i(t) > 0, \quad i = 1, 2, \ldots, I, t \in [0, T]$. Construct $\hat{s}(t)$ as in Theorem 4.2. If $\pi(\dot{x}) = 0$, then $\hat{u} = \hat{x} \hat{U}$ is an optimal control of the original problem. If $\pi(\dot{x}) > 0$, then $\hat{s}(t)$ is the direction of the ascent for $\theta(x)$ in $x = \dot{x}$.

**Proof.** Denote by $\mathcal{L}(\hat{\mu}_1, \hat{\eta}_1(\cdot), u(\cdot))$ the objective functional of the extremal problem in (4.6). Then by the saddle point theorem we get

$$
\max_{u \in \Omega} \mathcal{L}(\hat{\mu}_1, \hat{\eta}_1(\cdot), u(\cdot)) \geq \max_{u \in \Omega} \min_{\mu_1 \geq 0, \eta_1(\cdot) \geq 0} \mathcal{L}(\mu_1, \eta_1(\cdot), u(\cdot)) = J(u^*) \geq J(\hat{u}) = \theta(\dot{x}),
$$

(4.7)

where $u^*$ is an optimal control of (2.1)–(2.4). Thus we have $\pi(\dot{x}) \geq J(u^*) - J(\hat{u}) \geq 0$. Hence, if $\pi(\dot{x}) = 0$, then $\hat{u}$ is an optimal control of (2.1)–(2.4).

To prove the second part of the theorem, define $\theta(\rho) = \theta[\dot{x} + \rho \hat{s}]$. As in Theorem 4.2, we can derive

$$
(\theta(0))' = \delta(\dot{x})
$$

Then by (4.6) we have

$$
(\theta(0))' = \delta(\dot{x}) \geq \mathcal{L}(\hat{\mu}_1, \hat{\eta}_1(\cdot), \hat{u}(\cdot)) - \theta(\dot{x}) = \pi(\dot{x}).
$$

Hence, if $\pi(\dot{x}) > 0$, then $(\theta(0))' > 0$ and $\hat{s}$ is the direction of the ascent.

Bearing in mind the resource allocation problems, consider the original dynamical system, composed of $L$ individual subsystems, gathered by the control (resource) constraints, such that the problem (2.1)–(2.4) can be written in the form

$$
Q(x(T)) + \int_0^T f(x, u, t) dt \rightarrow \max, \\
\dot{x}_i(t) = A_i(t)x_i + b_i(u_i, t), \quad x_i(0) = x_{i0},
$$

$$
p_i(x_i, u_i, t) \leq 0, \quad v_i(x_i(T)) \leq 0, \quad u_i(t) \geq 0, \quad i = 1, 2, \ldots, L,
$$

$$
q(x, u, t) \leq 0, \quad d(x(T)) \leq 0, \quad x = (x_1, \ldots, x_L), \quad u = (u_1, \ldots, u_L).
$$

(4.8)
It is assumed that all the previous assumptions are fulfilled for the problem (4.8).

Let $\Omega$ be the set of controls, such that the following conditions are satisfied:
\[
\begin{align*}
\dot{x}_i(t) &= A_i(t)x_i + b_i(u_i, t), \quad x_i(0) = x_{i0}, \\
p_i(x_i, u_i, t) &\leq 0, \quad v_i(x_i(T)) \leq 0, \quad u_i(t) \geq 0, \\
l = 1, 2, \ldots, L.
\end{align*}
\]

(4.9)

Obviously, $\Omega \supseteq P$ and the problem (4.1) decomposes into $L$ independent subproblems of the form
\[
\begin{align*}
\int_0^T (u_i(t), \hat{A}_i(t)) \, dt \rightarrow \max, \\
\dot{x}_i(t) &= A_i(t)x_i + b_i(u_i, t), \quad x_i(0) = x_{i0}, \\
p_i(x_i, u_i, t) &\leq 0, \quad v_i(x_i(T)) \leq 0, \quad u_i(t) \geq 0,
\end{align*}
\]

(4.10)

where $\hat{A}_i(t)$ is the part of $\hat{A}(t) = \{\hat{A}_i(t), i = 1, 2, \ldots, L, j = 1, 2, \ldots, J_i\}$ corresponding to $u_i(t)$ and $\hat{A}_i^j(t)$ is defined similarly to (3.1).

Suppose that in the problem (4.8) the objective functional and the coupling constraints have the following additively separable form:
\[
\begin{align*}
Q(x(T)) &= \sum_i Q_i(x_i(T)), \\
f(x, u, t) &= \sum_i f_i(x_i, u_i, t), \\
q(x, u, t) &= \sum_i q_i(x_i, u_i, t), \\
d(x(T)) &= \sum_i d_i(x_i(T)).
\end{align*}
\]

Construct $\Omega$ as in (4.9). Note that (4.9) is a particular case of (4.4) and thus, due to the block-separability of the involved functions, the problem in the right-hand side of (4.6) decomposes into $L$ independent subproblems of the form
\[
\begin{align*}
Q_i(x_i(T)) - (\hat{\mu}, d_i(x_i(T))) + \int_0^T \left[ f_i(x_i, u_i, t) - (\hat{\eta}(t), q_i(x_i, u_i, t)) \right] \, dt \rightarrow \max, \\
\dot{x}_i &= A_i(t)x_i + b_i(u_i, t), \quad x_i(0) = x_{i0}, \\
p_i(x_i, u_i, t) &\leq 0, \quad v_i(x_i(T)) \leq 0, \quad u_i(t) \geq 0.
\end{align*}
\]

(4.11)

Then, based on Theorem 4.3 one either constructs the direction of the ascent or states the optimality of the disaggregated control.

Note that the form of the subproblems (4.10),(4.11) depends only on the structural properties of the original problem. The structure of the disaggregation (2.5), (2.6) is essential in the step (S2') only when one constructs the direction $s$ for the known optimal solutions of the subproblems. Recall that in the algorithm described in the second section, we obtain subproblems due to the special structure of the disaggregation.

From (4.7) it follows that
\[
\pi(\hat{s}) + \theta(\hat{s}) \geq J(u^*) \geq \theta(\hat{s})
\]
and thus in each iteration of the proposed algorithm we have the upper and the lower bounds of the optimal value of the objective functional of the original problem. Moreover, the lower bound increases monotonously from iteration to iteration and coincides with the upper bound for the optimal solution.

Based on Theorem 4.3 and combining the inequalities (4.2), (4.3), we can construct the different subproblems for the partial separable cases. Let, for example, \( Q(x(T)) \) and \( v(x(T)) \) be nonseparable, while \( f(x, u, t) \) and \( q(x, u, t) \) have the standard block-separable form. Construct \( \Omega \) as in (21). Then, using the first inequality in (4.3), we get

\[
\delta(\dot{x}) \geq \max \left\{ \left( \begin{array}{c} \left( \dot{x}, (\dot{x} - \hat{\eta} - q(x, u, t)) \right) \\
\left( \dot{x} - A_{i}(t)x_{i}(t) + b_{i}(u_{i}, t), x(0) = x_{i0}, \right) \\
p_{i}(x_{i}, u_{i}, t) \leq 0, v_{i}(x_{i}(T)) \leq 0, u_{i}(t) \geq 0, \\
l = 1, 2, \ldots, L \end{array} \right) \right\} \\
- (V_{x}Q(\dot{x}(T)) - \mu V_{x}d(\dot{x}(T)), \dot{x}(T)) + \int_{0}^{T} [f(x, u, t) - (\hat{\eta}(t), q(x, u, t))] dt = \pi_{Qd}(\dot{x}). \tag{4.12} \]

Due to the block-separability of \( f(x, u, t) \) and \( q(x, u, t) \), the problem in the right-hand side of (4.12) decomposes into \( L \) independent subproblems of the form

\[
(V_{x}Q(\dot{x}(T)) - \mu V_{x}d(\dot{x}(T)), x_{i}(T)) + \int_{0}^{T} [f_{i}(x_{i}, u_{i}, t) - (\hat{\eta}(t), q_{i}(x_{i}, u_{i}, t))] dt \to \max, \\
\dot{x}_{i} = A_{i}(t)x_{i} + b_{i}(u_{i}, t), x_{i}(0) = x_{i0}, \tag{4.13} \\
p_{i}(x_{i}, u_{i}, t) \leq 0, v_{i}(x_{i}(T)) \leq 0, u_{i}(t) \geq 0.
\]

It is not hard to verify that if \( \pi_{Qd}(\dot{x}) = 0 \), then \( \hat{u} \) is an optimal control of (4.8). If \( \pi_{Qd}(\dot{x}) > 0 \), then as in Theorem 4.3, one can construct the direction of the ascent. Moreover, by (4.2), (4.3) we have \( \pi(\dot{x}) \leq \pi_{Qd}(\dot{x}) \) and thus we get the upper and lower bounds

\[
\pi_{Qd}(\dot{x}) + \theta(\dot{x}) \geq J(u^{*}) \geq \theta(\dot{x})
\]
as before.

Now consider the case when the mixed constraints in (4.8) have the form

\[
p_{i}(x_{i}, u_{i}, t) \leq 0, v_{i}(x_{i}(T)) + c_{i}(x(T)) \leq 0, \\
q(x, u, t) \leq 0, d(x(T)) \leq 0, \quad l = 1, 2, \ldots, L.
\]
Assume that for all \( x, u \geq 0 \) we have the inequalities for the interconnection terms:

\[
    r_l(x, u, t) \geq S_l(t), \quad c_l(x(T)) \geq \beta_l,
\]

\( l = 1, 2, \ldots, L, \quad t \in [0, T] \),

where \( S_l(t) \) and \( \beta_l \) are given.

Let \( \Omega \) be the set of controls, such that the following conditions are satisfied:

\[
    \dot{x}_i = A_i(t)x_i + b_i(u_i, t), \quad x_i(0) = x_{i0},
\]

\[
    p_l(x_i, u_i, t) \leq -S_l(t), \quad v_i(x_i(T)) \leq -\beta_l, \quad u_i(t) \geq 0,
\]

\( l = 1, 2, \ldots, L \).

Obviously, \( \Omega \supseteq P \). Then the problem (4.1) decomposes into independent subproblems of the form

\[
    \int_0^T (u_i(t), \dot{\Delta}_i(t)) dt \rightarrow \max,
\]

\[
    \dot{x}_i = A_i(t)x_i + b_i(u_i, t), \quad x_i(0) = x_{i0},
\]

\[
    p_l(x_i, u_i, t) \leq -S_l(t), \quad v_i(x_i(T)) \leq -\beta_l, \quad u_i(t) \geq 0.
\]

(4.14)

Of course, one may use the other bounds of the interconnections to construct the proper set \( \Omega \). For example, it is possible to use the inequality \( r_l(x, u, t) \geq S_l(x_i, u_i, t) \). Moreover, using the inequalities similar to (4.2), (4.3), one can modify the objective functional of (4.14) in order to use the separability properties of the involved functions, as before.

In the block-separable case we may construct the aggregation such that the aggregated problem itself decomposes into independent subproblems. For example, let the original problem have the form

\[
    \sum_l \left[ Q_l(x_l(T)) + \int_0^T f_l(x_l, u_l, t) dt \right] \rightarrow \max,
\]

\[
    \dot{x}_l(t) = A_l(t)x_l + b_l(u_l, t), \quad x_l(0) = x_{l0},
\]

\[
    p_l(x_l, u_l, t) \leq 0, \quad v_l(x_l(T)) \leq 0, \quad u_l(t) \geq 0, \quad l = 1, 2, \ldots, L,
\]

\[
    \sum_l q^{k}_l(x_l, u_l, t) \leq R^k(t), \quad k = 1, 2, \ldots, K,
\]

where \( R^k(t), k = 1, 2, \ldots, K \), are smooth enough one-dimensional functions, such that \( R^k(t) \neq 0 \), \( t \in [0, T] \). Introducing the new controls \( y_l^k(t) \), we may reformulate the last restrictions of the original problem in the form

\[
    q^k_l(x_l, u_l, t) \leq y^k_l(t), \quad \sum_l y^k_l(t) = R^k(t),
\]

\( k = 1, 2, \ldots, K, \quad l = 1, 2, \ldots, L \).
We introduce a vector of the aggregated controls \( Y(t) = (Y^1(t) \ldots Y^K(t)) \) and a linear disaggregation
\[
y^k_l(t) = \alpha^k_l(t) Y^k(t), \quad k = 1, 2, \ldots, K, \quad l = 1, 2, \ldots, L,
\]
\[
\alpha \in A = \{ \alpha^k_l(t) : \sum_l \alpha^k_l(t) = 1, \quad k = 1, 2, \ldots, K, \quad t \in [0, T] \}.
\]
Obviously, if \( \alpha \in A \), then \( Y^k(t) = \sum_l y^k_l(t) \), \( k = 1, 2, \ldots, K \).

Fixing \( \alpha \in A \) and substituting the respective disaggregated controls into the original problem, we get the aggregated problem. The last restrictions of the aggregated problem are
\[
\alpha^k_l(t) \sum_l \alpha^k_l(t) Y^k(t) = R^k(t),
\]
\( l = 1, 2, \ldots, L, \quad k = 1, 2, \ldots, K \).

Let \( \hat{\eta}^k_l(t) \) and \( \hat{\delta}^k_l(t) \) be the respective unique Lagrange multipliers for these restrictions. Denote by \( \mathcal{L} \) the Lagrange functional associated with our aggregated problem. Then, from the necessary condition \( \partial \mathcal{L} / \partial Y^k = 0 \) we have \( \hat{\delta}^k_l(t) Y^k(t) = \sum_l \hat{\delta}^k_l(t) \hat{\eta}^k_l(t) \). Since \( \alpha \in A \), \( \sum_l \hat{\delta}^k_l(t) = 1 \) and thus
\[
\hat{\delta}^k_l(t) = \sum_l \hat{\delta}^k_l(t) \hat{\eta}^k_l(t), \quad k = 1, 2, \ldots, K.
\]

Note that \( \hat{\eta}^k_l(t) \) are the Lagrange multipliers for the binding constraints of the aggregated problem.

Consider now the solution of the aggregated problem. Since \( \hat{\alpha} \in A \), from the restriction \( \alpha^k_l(t) Y^k(t) = R^k(t) \) of the aggregated problem we have \( \hat{Y}^k(t) = R^k(t) \) for all \( \hat{\alpha} \in A \). Substituting this expression into the aggregated problem, we have that due to the block-separable structure, the rest of the aggregated problem decomposes into \( L \) independent subproblems of the form
\[
Q_l(x_l(T)) + \int_0^T f_l(x_l, u_l, t) \, dt \to \max,
\]
\[
x_l(t) = A_l(t) x_l + b_l(u_l, t), \quad x_l(0) = x_{10},
\]
\[
p_l(x_l, u_l, t) \leq 0, \quad v_l(x_l(T)) \leq 0, \quad u_l(t) \geq 0,
\]
\[
q^k_l(x_l, u_l, t) \leq \hat{\delta}^k_l(t) R^k(t), \quad k = 1, 2, \ldots, K.
\]
Calculating the Lagrange multipliers \( \hat{\eta}^k_l(t) \) in these independent subproblems, we can easily find the multipliers \( \hat{\delta}^k_l(t) \) for the binding constraints of the aggregated problem, using the above formula.

In order to update disaggregation parameters \( \alpha \), we can use the subproblems (4.11). In our case (4.11) becomes
\[
Q_l(x_l(T)) + \int_0^T \left[ f_l(x_l, u_l, t) - \sum_k y^k_l(t) \hat{\eta}^k_l(t) \right] \, dt \to \max,
\]
\[
x_l(t) = A_l(t) x_l + b_l(u_l, t), \quad x_l(0) = x_{10},
\]
\[
p_l(x_l, u_l, t) \leq 0, \quad v_l(x_l(T)) \leq 0, \quad u_l(t) \geq 0,
\]
\[
q^k_l(x_l, u_l, t) \leq y^k_l(t), \quad k = 1, 2, \ldots, K.
\]
Thus, in the block-separable case we can construct the aggregation such that only independent subproblems are solved in each iteration of the proposed method.
5. Examples

5.1. Consider at first the following illustrative resource allocation problem:

\[
\sum_{j} \left[ \int_{0}^{T} \gamma_{j}(1 - u_{j}(t))x_{j}(t) \, dt \right] \rightarrow \max,
\]

\[
\dot{x}_{j}(t) = \beta_{j}u_{j}(t), \quad x_{j}(0) = \kappa_{j},
\]

\[
\sum_{j} u_{j}(t) \leq \omega(t), \quad u_{j}(t) \geq 0, \quad j = 1, 2, \ldots, J,
\]

where \( \gamma_{j} > 0, \beta_{j} > 0, \omega(t) > 0, \kappa_{j} < 0 \). A similar problem was investigated in [14].

The Hamiltonian of this problem is

\[
H(t) = \sum_{j} [\lambda_{j}(t)\beta_{j}u_{j}(t) + \gamma_{j}(1 - u_{j}(t))x_{j}(t)],
\]

where \( \lambda_{j}(t) \) is the costate variable, which satisfies

\[
\dot{\lambda}_{j}(t) = - \gamma_{j}(1 - u_{j}(t)), \quad \lambda_{j}(T) = 0.
\]

According to the maximum principle we obtain the following linear programming problem for each \( t \in [0, T] \):

\[
\sum_{j} [\beta_{j}\lambda_{j}(t) - \gamma_{j}x_{j}(t)]u_{j}(t) \rightarrow \max,
\]

\[
\sum_{j} u_{j}(t) \leq \omega(t), \quad u_{j}(t) \geq 0, \quad j = 1, 2, \ldots, J.
\]

Here the possibility of the optimal bang-bang control exists and the switching points depend on the values of \( \beta_{j}\lambda_{j} - \gamma_{j}x_{j} \). In order to find these switching points you have to investigate the two-point boundary value problems, formulated for each \( j = 1, 2, \ldots, J \) from (5.1),(5.2). If \( J \) is sufficiently large, this is not an easy problem.

Consider now the decomposition–aggregation method for (5.1). We introduce one macrocontrol \( U(t) = \sum_{j} u_{j}(t) \) and the set \( A = \{x_{j}(t): \sum x_{j}(t) = 1, x_{j}(t) \geq 0, j = 1, 2, \ldots, J \} \). The aggregated problem is

\[
\sum_{j} \left[ \int_{0}^{T} \gamma_{j}(1 - \hat{x}_{j}(t)U(t))x_{j}(t) \, dt \right] \rightarrow \max,
\]

\[
\dot{\hat{x}}_{j} = \beta_{j}\hat{x}_{j}(t)U(t), \quad \hat{x}_{j}(0) = \kappa_{j},
\]

\[0 \leq U(t) \leq \omega(t).
\]

Let the optimal solution of the aggregated problem be \( \hat{U}(t) = \omega(t) > 0 \). Then from the dual aggregated problem we can derive

\[
\dot{\eta}(t) = \sum_{j} [ - \gamma_{j}\hat{x}_{j}(t)\dot{\hat{x}}_{j}(t) + \lambda_{j}(t)\beta_{j}\hat{x}_{j}(t)],
\]

where \( \hat{x}_{j}(t) \) and \( \dot{\hat{x}}_{j}(t) \) are easily calculated for \( \hat{U}(t) = \omega(t) \).
Let $\Omega = \{u_j(t): 0 \leq u_j(t) \leq \omega(t), j = 1, 2, \ldots, J\}$. Obviously, $\Omega \supseteq P$, where $P$ is the set of all feasible controls of (5.1). Then the $j$th local subproblem can be formulated in accordance with (4.1) in the form

$$\int_0^T \hat{A}_j(t) u_j(t) \, dt \rightarrow \max,$$

$$0 \leq u_j(t) \leq \omega(t),$$

where $\hat{A}_j(t)$ is defined in the sense of (3.1):

$$\hat{A}_j(t) = \hat{\lambda}_j(t) \beta_j - \gamma_j \hat{x}_j(t) - \hat{\eta}(t).$$

Hence, for each $t \in [0, T]$ we have the following subproblems:

$$u_j(t)[\hat{\lambda}_j(t) \beta_j - \gamma_j \hat{x}_j(t) - \hat{\eta}(t)] \rightarrow \max,$$

$$0 \leq u_j(t) \leq \omega(t).$$

These independent subproblems are much easier to solve than the original multidimensional problem (5.1).

Consider for example the special case of the problem (5.1) with $\omega(t) = \omega = \text{constant}$. Let the initial values of $\hat{x}_j$, $j = 1, 2, \ldots, J$, be constants. If $|\kappa_j|$, $j = 1, 2, \ldots, J$, are large enough, then the optimal control of the aggregated problem is $\hat{U}(t) = \omega$. We have

$$\hat{x}_j(t) = \beta_j \hat{x}_j \omega t + \kappa_j, \quad \hat{\lambda}_j(t) = \gamma_j(1 - \hat{\alpha}_j \omega)(T - t)$$

and

$$\hat{A}_j(t) = t(\sum \gamma_s \beta_s \hat{a}_s - \gamma_j \beta_j) + [\gamma_j \beta_j(1 - \hat{\alpha}_j \omega) T - \gamma_j \kappa_j]$$

$$- \sum \hat{a}_s[\gamma_s \beta_s(1 - \hat{\alpha}_s \omega) T - \gamma_s \kappa_s].$$

It follows from (5.5) that if $\gamma_j \beta_j = a = \text{constant}$, then $\hat{A}_j(t) = \hat{\lambda}_j = \text{constant}$. Hence, there exists an optimal control without the switching points in the local subproblems. If, moreover, $\gamma_j \kappa_j = b = \text{constant}$, then we have

$$\hat{A}_j = a \omega T(\sum \hat{a}_s^2 - \hat{\lambda}_j).$$

Put $\hat{x}_j = 1/J, j = 1, 2, \ldots, J$. In this case we have $\hat{A}_j = 0, j = 1, 2, \ldots, J$ and it follows from Theorem 3.1 that the disaggregated solution $\hat{u}_j = \omega/J, j = 1, 2, \ldots, J$, is an optimal control of the original problem (5.1). The respective value of the objective function is $(1 - \omega/J)(0.5a \omega T^2 + hJ \omega T)$.

Note that if $\hat{A}_j = 0, j = 1, 2, \ldots, J$, then the local subproblems (5.4) have a nonunique solution. For example, we may choose the bang-bang control

$$\hat{u}_j(t) = \omega, \quad t \in (t_j^0, t_j^* + 1), \quad \hat{u}_j(t) = 0, \quad t \in [0, T] \setminus (t_j^*, t_j^* + 1),$$

$$t_0^* = 0, \quad t_j^* = jT/J, \quad j = 1, 2, \ldots, J.$$

Then in the step (S2') we have

$$\hat{U}(t) = \sum \hat{u}_j(t) = \omega > 0, \quad \hat{s}_j(t) = \hat{U}(t)[\hat{x}_j(t) - \hat{x}_j]/\hat{U}(t) = \hat{x}_j(t) - 1/J,$$

$$\hat{\lambda}_j(t) = \hat{u}_j(t)/\omega.$$
and $x_j(t) = 1/J + \rho(\hat{x}_j(t) - 1/J)$. In the step (S3) we have $\theta(\rho) = \theta[1/J + \rho(\hat{x}_j(t) - 1/J)]$. It is not hard to get the analytical dependence of $\theta(\rho)$ and conclude that $\max \{\theta(\rho) | 0 \leq \rho \leq 1\} = \theta(1)$, such that $x_j(t) = \hat{x}_j(t)$.

For the next step we have $\rho = 0$ and $\hat{x}$ is the optimal solution of the problem (3.2). It follows from Theorem 3.3 that the bang-bang control (5.6) is an optimal control of the original problem (5.1). The respective value of the objective functional is the same as above. Thus, if the original problem has a nonunique optimal solution, then the different stopping rules may give the different optimal solutions.

5.2. The second problem concerns the optimal control of the linear system under the random disturbances [3]. It is assumed that there are a number of interconnected linear systems and we try to complicate the observation of this complex system, creating the noises in the observation channels. One of the possible approaches is to maximize the dispersion, choosing the intensities of the Gaussian white noises. The problem is formulated in the following way (see [3] for details):

$$\sum_{j=1}^{n} \sigma_j x_j(T) + \int_{0}^{T} \sum_{j=1}^{n} [\gamma_j(x_j(t))^2 + \beta_j(u_j(t))^2] \, dt \to \min,$$

$$\dot{x}_j(t) = \left( \sum_{i=1}^{n} a_{ij} u_j(t) \right)^{-1}, \quad x_j(0) = x_{j0}, \quad j = 1, 2, \ldots, n,$$

$$\sum_{j=1}^{n} u_j(t) \leq \omega(t), \quad u_j(t) \geq 0,$$

where the control $u_j(t)$ is proportional to the intensity of the white noise in the respective observation channel. The scalar parameters $\sigma_j, \gamma_j$ and $\beta_j$ are positive, and $\omega(t)$ is a given positive function. The positive coefficients of the interconnections $a_{ij}$ show that the noise created in one subsystem affects the other subsystems.

The one macrocontrol $U(t) = \sum u_j(t)$ is introduced and the set $A$ of the disaggregation parameters is defined by

$$A = \{x_j(t): \sum x_j(t) = 1, x_j(t) \geq 0, j = 1, 2, \ldots, n\}.$$

In the aggregated problem we have the restriction $0 \leq U(t) \leq \omega(t)$. It is not hard to verify that due to the positivity of the coefficients in the original problem, the optimal macrocontrol $\hat{U}(t)$ in the aggregated problem is strictly positive for all $k \in A$. Hence, we can find the unique Lagrange multiplier from the respective equality constraint of the dual aggregated problem. The aggregated problem was solved by the projected gradient method. The respective projection on the set $0 \leq U(t) \leq \omega(t)$ is calculated analytically.

To compute the direction of the descent, we use the subproblem (4.1). The set $\Omega$ in the control space was defined as follows:

$$\Omega = \{u(t): 0 \leq u_j(t) \leq \omega(t), \quad j = 1, 2, \ldots, n\}.$$

Obviously, $\Omega \subseteq P$, where

$$P = \{u(t): \sum_{j} u_j(t) \leq \omega(t), \quad u_j(t) \geq 0, i = 1, 2, \ldots, n\}$$

is the set of feasible controls of the original problem (5.7). For the set $\Omega$ defined in (5.8), the auxiliary problem (4.1) decomposes into $n$ independent subproblems. The $j$th subproblem has the linear
integral objective and one interval constraint $0 \leq u(t) \leq \omega(t)$. This problem is solved analytically. Note the respective independent subproblems we have for each $t \in [0, T]$.

Numerical experiments were done on an IBM PC/XT with an INTEL 8088 processor for $n = 10$, $T = 1$, $\sigma_j = 0$, $\sigma_j = 0.5$, $j = 1, 2, \ldots, 10$, $\omega(t) = \exp(t)$. The integer coefficients $\beta_j, \gamma_j$ were chosen casually from the interval $[1, 9]$. The coefficients of the interconnections $a_{ij}$ were $a_{ij} = 10(0.5)^{|i-j|}$, $i, j = 1, 2, \ldots, 10$. This rule shows that the adjacent subsystems interact stronger than the distant subsystems. The aggregated problem was solved by the projected gradient method after time discretization with 50 nodes per time interval. The stopping rule was $\delta(a') \leq 0.0001$, where $l$ is the number of iterations. The initial values of the disaggregation parameters were chosen constant and equal to each other.

Some optimal disaggregated controls, computed after 10 iterations for $\beta = (4, 4, 6, 3, 2, 4, 8, 2, 6, 5)$ and $\gamma = (7, 7, 8, 6, 3, 1, 3, 2, 7, 1)$, are graphically shown in Fig. 1. The total CPU-time was around 10 min. The difference between $\exp(t)$ and the optimal aggregated control is shown in Fig. 2. The values of $\theta(a')$ were as follows:

$$\theta(a^0) = 7.791, \quad \theta(a^3) = 4.543, \quad \theta(a^6) = 2.378, \quad \theta(a^9) = 1.037.$$ 

The total CPU-time is practically independent of the values of $\beta_j, \gamma_j$ and depends on the initial values of the disaggregation parameters. For example, for the initial values $\alpha_{10} = 1$, $\alpha_j = 0$, $j \neq 10$ only 5 iterations were necessary to get the optimal control with the above $\beta$ and $\gamma$.

![Fig. 1.](image1)

![Fig. 2.](image2)
References


