International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-97-051 / August

Constraint Aggregation in Infinite-Dimensional Spaces and Applications

Arkadii V. Kryazhimskii (kryazhim@iiasa.ac.at)
Andrzej Ruszczyński (rusz@iiasa.ac.at)

Approved by
Gordon J. MacDonald (macdon@iiasa.ac.at)
Director, *IIASA*

Contents

1 Introduction **1**

2 Problem setting. Preliminaries **2**

3 Constraint aggregation **4**

4 Optimal control of linear systems **7**

5 Convex processes **14**

6 Games **19**

7 Stochastic programming **22**

8 Conclusions **25**

Abstract

An aggregation technique for constraints with values in Hilbert spaces is suggested. The technique allows to replace the original optimization problem by a sequence of subproblems having scalar or finite-dimensional constraints. Applications to optimal control, games and stochastic programming are discussed in detail.

Key Words: Constrained optimization in vector spaces, aggregation, optimal control, games, stochastic programming.

Constraint Aggregation in Infinite-Dimensional Spaces and Applications

Arkadii V. Kryazhimskii * (kryazhim@iiasa.ac.at)
Andrzej Ruszczyński (rusz@iiasa.ac.at)

1 Introduction

Optimization problems in vector spaces constitute convenient models of various applied problems in control theory, game theory, stochastic optimization and many other areas. There are, however, well-known difficulties associated with constraints having values in infinite-dimensional spaces: constraint qualification conditions are often in conflict with the desire to have an easy representation of Lagrange multipliers. This occurs, for example, in the optimal control theory [4, 7] and in stochastic programming [12, 18, 19, 6]. Numerical solution of such problems (or their finite-dimensional approximations), usually poses a great challenge, and much efforts are devoted to the ways of dealing with infinite-dimensional constraints.

Our objective is to show a new possibility to drastically reduce the complexity of such optimization problems: aggregation of constraints to one or finitely many scalar equations or inequalities. Together with an abstract theory we shall present some applications of the new approach to control, games and stochastic programming. All of them, with a single exception, are concerned with convex optimization problems in Hilbert spaces; in the exceptional case (a game problem in section 6) we operate in the space dual to the space of continuous functions. In order to cover both situations with minimum generality (generality and complexity are not our aims in this paper), we describe the method in the context of convex optimization in a space that is dual to a separable Banach space.

The idea of aggregation was inspired by the *extremal shift control principle* used in differential games [9]. In the context of finite-dimensional optimization it has been developed in [10, 5]. In [11] a relevant approximate solution method for linear optimal control was analyzed.

We give the problem setting in the next section, in which we also decribe some useful schemes of establishing weak* convergence of approximate solutions to the solution set. Section 3 is devoted to the description of the abstract constraint aggregation method, and to the convergence analysis. The next two sections, 4 and 5, apply the method to optimal control problems. In section 6 we consider an application to game theory, and section 7 discusses an application to multistage stochastic programming.

Throughout, we write $\|\cdot\|$ for the norm in a Banach space Ξ and $\langle\cdot,\cdot\rangle$ for the duality between Ξ and Ξ^* ($\langle g,x\rangle$ is the value of the functional $x \in \Xi^*$ at the element $g \in \Xi$).

*This author has been partially supported by the RFFI Grant 97-01-01060.

In particular, if Ξ is a Hilbert space, $\langle\cdot,\cdot\rangle$ stands for the scalar product in Ξ. The same notation will be used for the norm and scalar product in a finite-dimensional Euclidean space.

2 Problem setting. Preliminaries

Let Ξ be a separable Banach space, and let $\mathcal{X} = \Xi^*$ be its dual. Let X be a closed bounded set in the space $\mathcal{X}$ and let f be a convex weakly* lower semicontinuous functional on $\mathcal{X}$. We consider the minimization problem

$$\min\ f(x), \tag{2.1}$$

$$Ax = b, \tag{2.2}$$

$$h(s,x) \le 0 \quad \text{for } \mu\text{-a.a.} \quad s \in S \tag{2.3}$$

$$x \in X. \tag{2.4}$$

Here A is a linear operator from $\mathcal{X}$ to some Hilbert space $\mathcal{H}$, and b is a fixed element in $\mathcal{H}$. We assume that $x \mapsto \|Ax - b\|^2$ is weakly* lower semicontinuous. The functionals $h(s,\cdot)$, $s \in S$, are assumed to be convex and weakly* lower semicontinuous on X, and the parameter set S is a measurable space with a σ-additive measure μ. Furthermore, we assume that for each $x \in X$ the functions $s \mapsto h(s,x)$ are μ-measurable, and there exists a μ-measurable function $\overline{h} : S \to \mathbb{R}$ such that $|h(s,x)| \le \overline{h}(s)$ for all $x \in X$, $s \in S$, and

$$\int_S (\overline{h}(s))^2\, \mu(ds) < \infty. \tag{2.5}$$

Finally, we assume that the feasible set of the problem (2.1)–(2.4) is nonempty.

Since the unit ball in $\mathcal{X}$ is weakly* compact [8, Thm. 6, p. 179], and the set X is weakly* closed, the latter set is weakly* compact. Moreover, due to the weak* lower semicontinuity of $x \mapsto \|Ax - b\|^2$ and $h(s,\cdot)$ ($s \in S$), the feasible set of (2.1)–(2.4) is weakly* compact. By assumption, the objective function f is weakly* lower semicontinuous. Therefore the problem (2.1)–(2.4) has a nonempty solution set X^*. We denote by f^* the minimum value: $f^* = f(x)$, $x \in X^*$.

Remark 2.1. If $\mathcal{X}$ is a Hilbert space, we as usual identify Ξ with $\mathcal{X} = \Xi^*$, and the weak* topology in $\mathcal{X}$ with the weak topology in $\mathcal{X}$. In this case, the assumptions imposed above are satisfied if the set X is convex, closed and bounded, the linear operator A is bounded, and the functionals f and $h(s,\cdot)$ ($s \in S$) are convex and bounded on a neighborhood of X (in the strong topology).

For simplicity we start our analysis from two cases in which the problem (2.1)–(2.4) has only equality or inequality constraints, respectively. These are

$$\min f(x),\ \text{s.t. } Ax = b,\ x \in X, \tag{2.6}$$

and

$$\min\ f(x),\ \text{s.t. } h(s,x) \le 0\ \mu\text{-a.a. } s \in S,\ x \in X. \tag{2.7}$$

Let us first recall some standard definitions. A sequence $\{x^k\} \subset \mathcal{X}$ is called *strongly convergent* to a set $Y \subset \mathcal{X}$ if

$$\lim_{k\to\infty} \inf_{y \in Y} \|x^k - y\| = 0,$$

and is said to *weakly* converge* to Y if for every finite set of elements $g_1, \ldots, g_m \in \Xi$ one has

$$\inf_{y\in Y} \max_{1\le j\le m} |\langle g_j, x^k - y\rangle| \to 0 \quad \text{as} \quad k \to \infty.$$

The convergence analysis of the methods discussed in the paper will be based on the following simple observations.

Lemma 2.1. *If a sequence* $\{x^k\} \subset X$ *satisfies the conditions*

$$\limsup_{k\to\infty} f(x^k) \le f^*, \tag{2.8}$$

$$\lim_{k\to\infty} \|Ax^k - b\| = 0, \tag{2.9}$$

then $\{x^k\}$ *weakly* converges to the solution set* X^* *of the problem (2.6).*

Proof. We use standard arguments. Suppose the contrary. Then there is a subsequence $\{x^{k_i}\}$ such that for certain $g_1, \ldots, g_m \in \Xi$,

$$\inf_{x\in X^*} \max_{1\le j\le m} \mid \langle g_j, x^{k_i} - x\rangle \mid > \varepsilon > 0. \tag{2.10}$$

Since X is weakly* compact in $\mathcal{X}$, with no loss of generality we may assume that $\{x^{k_i}\}$ weakly* converges to some $x' \in X$. Hence (2.8) implies $f(x') \le f^*$. The weak* lower semicontinuity of $x \mapsto \|Ax - b\|^2$ and (2.9) yield $Ax' = b$. Consequently, $x' \in X^*$. Setting $x = x'$ in (2.10), we arrive at a contradiction. □

For the problem (2.7) the convergence scheme differs only in the details. For each $x \in X$ and $s \in S$ we define

$$h^+(s, x) = \max(0, h(s, x)).$$

Lemma 2.2. *If a sequence* $\{x^k\} \subset X$ *satisfies condition (2.8) and*

$$\lim_{k\to\infty} \int_S (h^+(s, x^k))^2 \, \mu(ds) = 0, \tag{2.11}$$

then $\{x^k\}$ *weakly* converges to the solution set* X^* *of the problem (2.6).*

Proof. The proof is almost identical to the proof of Lemma 2.1. Only at the end (instead of $Ax' = b$) we need to show that

$$h(s, x') \le 0 \quad \text{for } \mu\text{-almost all } s \in S.$$

To this end we notice that the functionals $x \to (h^+(s, x))^2$ are weakly* lower semicontinuous. This combined with (2.11) and (2.5) implies that

$$\int_S (h^+(s, x'))^2 \, \mu(ds) = 0,$$

which completes the proof. □

Combining the proofs of Lemmas 2.1 and 2.2 one easily obtains a formulation for the problem (2.1)–(2.4) with both equality and inequality constraints.

Lemma 2.3. *Let a sequence* $\{x^k\}$ *from* X *satisfy (2.8), (2.9), and (2.11). Then* $\{x^k\}$ *weakly* converges to the solution set* X^* *of the problem (2.1)–(2.4).*

We will separate the case where functional f is *strongly convex*, that is, there exists $\kappa > 0$ such that for all $x_1, x_2 \in \mathcal{X}$ and $\beta \in [0, 1]$ one has

$$f(\beta x_1 + (1-\beta)x_2) \le \beta f(x_1) + (1-\beta)f(x_2) - \beta(1-\beta)\kappa\|x_1 - x_2\|^2. \tag{2.12}$$

Lemma 2.4. *Let* f *be strongly convex and a sequence* $\{x^k\}$ *in* $\mathcal{X}$ *satisfy (2.8), (2.9), and (2.11). Then the solution set* X^* *contains only one element, and* $\{x^k\}$ *strongly converges to* X^**.*

Proof. The strong convexity of f obviously implies that X^* is a singleton: $X^* = \{x^*\}$. Let us show that $\{x^k\}$ strongly converges to x^*. Suppose the contrary, i.e., that there is a subsequence $\{x^{k_i}\}$ such that

$$\|x^{k_i} - x^*\| > \varepsilon > 0. \tag{2.13}$$

With no loss of generality, we may assume that $\{x^{k_i}\}$ weakly* converges in $\mathcal{X}$ to some $x' \in X$. By (2.8), (2.9), and (2.11), x' is feasible in the problem (2.1)–(2.4) and $f(x' = f^*$. Therefore, $x' = x^*$. The next argument is close to [22]. Obviously, the sequence $\{(x^{k_i} + x^*)/2\}$ weakly* converges to x^*. Due to the weak* lower semicontinuity of f,

$$f^* = f(x^*) \le \liminf_{i\to\infty} f\left(\frac{x^{k_i} + x^*}{2}\right).$$

On the other hand, by the convexity of f

$$\limsup_{i\to\infty} f\left(\frac{x^{k_i} + x^*}{2}\right) \le \limsup_{i\to\infty}\left(\frac{1}{2}f(x^{k_i}) + \frac{1}{2}f(x^*)\right) = f^*.$$

Hence

$$\lim_{i\to\infty} f\left(\frac{x^{k_i} + x^*}{2}\right) = f^*.$$

By (2.12) with $\beta = 1/2$

$$\frac{\kappa}{4}\|x^{k_i} - x^*\|^2 \le \frac{1}{2}f(x^{k_i}) + \frac{1}{2}f(x^*) - f\left(\frac{x^{k_i} + x^*}{2}\right)$$

and (2.8) yields $\lim_{i\to\infty}\|x^{k_i} - x^*\|^2 = 0$. A contradiction with (2.13) completes the proof. □

3 Constraint aggregation

Here we extend the finite-dimensional constraint aggregation method of [5] to the problem (2.1)–(2.4). We start from the problem (2.6) with the equality constraint. Let us fix $x^0 \in X$ such that $f(x^0) \le f^*$ (in particular, x^0 may be a minimizer of f over X), and define a sequence $\{x^k\}$ of approximate solutions to (2.1)–(2.4) by

$$x^{k+1} = x^k + \tau_k(u^k - x^k), \tag{3.1}$$

where u^k is a solution of:

$$\min\ f(u), \tag{3.2}$$

$$\langle Ax^k - b, Au - b\rangle \le 0, \tag{3.3}$$

$$u \in X, \tag{3.4}$$

and

$$\tau_k = \arg \min_{0\le\tau\le1} \left\|(1-\tau)Ax^k + \tau Au^k - b\right\|^2. \tag{3.5}$$

Theorem 3.1. *The sequence $\{x^k\}$ generated by (3.1)–(3.5) weakly* converges to the solution set X^* of the problem (2.6).*

Proof. It is sufficient to verify the properties (2.8), (2.9). One verifies them using the same arguments as in the proof of Theorem 3.3 in [5]. The first one is obvious, because the subproblem (3.2)–(3.4) is a relaxation of (2.6), so $f(u^k) \le f^*$. By the convexity of f we have $f(x^{k+1}) \le (1-\tau_k)f(x^k) + \tau_k f(u^k) \le f^*$. To prove the second one we use (3.3) to obtain the key inequality:

$$\begin{aligned}
&\|(1-\tau)Ax^k + \tau Au^k - b\|^2 \\
&\quad = (1-\tau)^2\|Ax^k - b\|^2 + 2\tau(1-\tau)\langle Ax^k - b, Au^k - b\rangle + \tau^2\|Au^k - b\|^2 \\
&\quad \le (1-2\tau)\|Ax^k - b\|^2 + 2K_A\tau^2,
\end{aligned} \tag{3.6}$$

where K_A is an upper bound on $\|Ax - b\|^2$ in X. Therefore

$$\begin{aligned}
\|Ax^{k+1} - b\|^2 &\le \min_{0\le\tau\le1}\left((1-2\tau)\|Ax^k - b\|^2 + 2K_A\tau^2\right) \\
&\le \left(1 - \frac{\|Ax^k - b\|^2}{2K_A}\right)\|Ax^k - b\|^2.
\end{aligned}$$

This proves (2.9). □

For the problem (2.7) with inequality constraints we modify the algorithm by replacing (3.2)–(3.4) with

$$\min\ f(u), \tag{3.7}$$

$$\int_S h^+(s, x^k)h(s, u)\,\mu(ds) \le 0, \tag{3.8}$$

$$u \in X, \tag{3.9}$$

and by using the corresponding stepsize rule

$$\tau_k = \arg \min_{0\le\tau\le1} \int_S \left(h^+(s, (1-\tau)x^k + \tau u^k)\right)^2 \mu(ds). \tag{3.10}$$

Theorem 3.2. *The sequence (x^k) generated by (3.1), (3.7)–(3.9) and (3.10) weakly converges to the solution set X^* of the problem (2.7).*

Proof. By Lemma 2.2 it is sufficient to verify properties (2.8), (2.11). We follow the proof of Theorem 5.1 in [5].

For each $s \in S$, by the convexity of $h(s, \cdot)$,

$$\begin{aligned}
h(s, (1-\tau)x^k + \tau u^k) &\le (1-\tau)h(s, x^k) + \tau h(s, u^k) \\
&\le (1-\tau)h^+(s, x^k) + \tau h(s, u^k).
\end{aligned}$$

In the above inequality, for the parameter values s for which the left hand side is positive, the right hand side has a larger absolute value. Therefore

$$\begin{aligned}
&\int_S \left(h^+(s, (1-\tau)x^k + \tau u^k)\right)^2 \mu(ds) \\
&\quad \le \int_S \left((1-\tau)h^+(s, x^k) + \tau h(s, u^k)\right)^2 \mu(ds) \\
&\quad \le (1-\tau)^2 \int_S \left(h^+(s, x^k)\right)^2 \mu(ds) + \tau^2 \int_S \left(h(s, u^k)\right)^2 \mu(ds),
\end{aligned}$$

where in the last inequality we used (3.8). The rest is similar to the proof of Theorem 3.1. □

Finally, for the problem (2.1)–(2.4) with both equality and inequality constraints, the algorithm combines (3.2)–(3.4) and (3.7)–(3.9). The subproblem takes on the form

$$\min\ f(u), \tag{3.11}$$
$$\langle Ax^k - b, Au - b\rangle \le 0, \tag{3.12}$$
$$\int_S h^+(s, x^k) h(s, u)\, \mu(ds) \le 0, \tag{3.13}$$
$$u \in X, \tag{3.14}$$

and the stepsize rule is modified accordingly:

$$\tau_k = \arg\min_{0\le\tau\le 1} \left[\gamma \left\|(1-\tau)Ax^k + \tau Au^k - b\right\|^2 + \int_S \left(h^+(s, (1-\tau)x^k + \tau u^k)\right)^2 \mu(ds)\right], \tag{3.15}$$

where the scaling factor $\gamma > 0$.

Theorem 3.3. *The sequence $\{x^k\}$ generated by (3.1), (3.11)–(3.14) and (3.15) weakly* converges to the solution set X^* of the problem (2.1)–(2.4).*

The proof is identical.

Let us point out a modification of the constraint aggregation method, in which the inequality (3.13) is replaced by a more general constraint. Namely, consider the auxiliary problem

$$\min\ f(u), \tag{3.16}$$
$$\langle Ax^k - b, Au - b\rangle \le 0, \tag{3.17}$$
$$g^k(u) \le 0, \tag{3.18}$$
$$x \in X. \tag{3.19}$$

Here $g^k(\cdot)$ is a scalar function on X such that for every $u \in X$

$$\int_S h^+(s, x^k) h(s, u) \mu(ds) \le g^k(u), \tag{3.20}$$

and for every u feasible in the original problem (2.1)–(2.4) inequality (3.18) is satisfied.

Theorem 3.4. *Let the sequence $\{x^k\}$ be generated by (3.16)–(3.19) and (3.15). Then $\{x^k\}$ weakly* converges to the solution set X^* of the problem (2.1)–(2.4).*

The proof is identical.

Remark 3.1. It also clear from the analysis that the aggregate inequality constraints (3.12) can be replaced by equations. It is also possible to aggregate constraints in subgroups, analogously to the two groups (3.12) and (3.14) in the last case. The stepsize rule has to be then modified in a corresponding way, similarly to (3.15). The detailed description of these technical details and the convergence proof will be omitted, because they are obvious at this stage.

Remark 3.2. By virtue of Lemma 2.4 in all the above results the convergence of $\{x^k\}$ to X^* is strong if the functional f is strongly convex.

4 Optimal control of linear systems

In this section we employ the constraint aggregation technique to solve a problem of optimal control for a linear system, under mixed constraints on state and control variables. The problem formulation is as follows:

$$\min\ f(y,v), \tag{4.1}$$
$$\dot{y}(t) = Cy(t) + Dv(t) \quad \text{for a.a.} \quad t \in [0,1], \tag{4.2}$$
$$y(0) = y_0, \tag{4.3}$$
$$(y(t), v(t)) \in M \quad \text{for a.a.} \quad t \in [0,1]. \tag{4.4}$$

Here

$$f(y,v) = \int_0^1 \varphi(t, y(t), v(t))dt, \tag{4.5}$$

C and D are $n \times n$ and $n \times m$ matrices, respectively, $y_0 \in \mathbb{R}^n$, and M is a convex compactum in $\mathbb{R}^n \times \mathbb{R}^n$. The constraint (4.4) (which does not carry information on the classes of functions y and v) is formally understood as

$$(y,v) \in \Omega$$

with

$$\Omega = \{(y,v) \in \mathcal{A}([0,1],\mathbb{R}^n) \times \mathcal{L}^\infty([0,1],\mathbb{R}^m) : (y(t),v(t)) \in M \text{ for a.a } t \in [0,1]\}, \tag{4.6}$$

where $\mathcal{A}([0,1],\mathbb{R}^n)$ is the set of all absolutely continuous functions from $[0,1]$ to $\mathbb{R}^n$. The function φ in (4.5) is continuous, and for each $t \in [0,1]$ the function $(y,v) \mapsto \varphi(t,y,v)$ is convex. We assume the admissible set of problem (4.1)–(4.4) to be nonempty.

Remark 4.1. The solution method described below works in a more general setting. The system equation (4.2) may be nonstationary and nonhomogeneous, that is, C and D may be bounded measurable matrix functions of time t, and the right hand side of (4.2) may contain an additional measurable time-dependent term. Also the set M in (4.6) may be a bounded measurable map with compact and convex values. Finally, the function φ in (4.5) may be only bounded and measurable in t. Wishing to avoid unnecessary technical details, we concentrate on the simplest case (4.1)–(4.4).

We shall reduce the optimal control problem (4.1)–(4.4) to a convex optimization problem (2.6) with linear constraints. Define the spaces: $\mathcal{X} = \mathcal{L}^2([0,1],\mathbb{R}^n) \times \mathcal{L}^2([0,1],\mathbb{R}^m)$, and $\mathcal{H} = \mathcal{L}^2([0,1],\mathbb{R}^n)$. Thus, we treat arguments (y,v) of problem (4.1)–(4.4) as elements of the Hilbert space $\mathcal{X}$. Let us rewrite the equation (4.2) in an integral form:

$$y(\theta) = y_0 + \int_0^\theta (Cy(t) + Dv(t))\, dt, \quad \theta \in [0,1].$$

In a more general notation it reads

$$A(y,v) = b.$$

Here A is a linear bounded operator from $\mathcal{X}$ to $\mathcal{H}$ given by

$$A(y,v)(\theta) = y(\theta) - \int_0^\theta (Cy(t) + Dv(t))\, dt \quad \theta \in [0,1], \tag{4.7}$$

and

$$b(\theta) = y_0, \quad \theta \in [0,1]. \tag{4.8}$$

The set (4.6) is naturally transformed into

$$X = \{(y,v) \in \mathcal{X} : (y(t), v(t)) \in M \text{ for a.a } t \in [0,1]\}. \tag{4.9}$$

Obviously, X is closed, convex and bounded in $\mathcal{X}$. Thus the optimization problem (2.6) where the operator A, element b and set X are given by (4.7), (4.8) and (4.9) satisfies all the assumptions of section 3, and can be solved with the constraint aggregation method described in Theorem 3.1.

The next reduction theorem is evident.

Theorem 4.1. *Let the operator $A : \mathcal{X} \mapsto \mathcal{H}$, the element $b \in \mathcal{H}$ and the set $X \subset \mathcal{X}$ be defined by (4.7), (4.8) and (4.9). Then the optimal control problem (4.1)–(4.4) is equivalent to the convex optimization problem (2.6) in the following sense:*
(i) *each solution $x^* = (y^*, v^*)$ of the problem (4.1)–(4.4) solves the problem (2.6),*
(ii) *if $x^* = (y^*, v^*)$ solves the problem (2.6), then there exists a solution $x^{**} = (y^{**}, v^{**})$ of the problem (4.1)–(4.4) such that $v^{**} = v^*$ and $y^{**}(t) = y^*(t)$ for a.a. $t \in [0,1]$.*

In the remaining part of this section the convex optimization problem (2.6) is understood as described in Theorem 4.1. Its solution set will be denoted by X^*; for the solution set of the optimal control problem (4.1)–(4.4) we will use the notation Ω^*.

To solve the problem (2.6) we employ the constraint aggregation method (3.1)–(3.5) (see Theorem 3.1).

We shall specify u^k in (3.2)–(3.4) in terms of a certain optimal control problem. We start with a specification of the term $\langle Ax^k - b, Au - b\rangle$ in the aggregate constraint (3.3). Let $x^k = (y^k, v^k)$, and for $\theta \in [0,1]$

$$r^k(\theta) = A(y^k, v^k)(\theta) - b(\theta) = y^k(\theta) - y_0 - \int_0^\theta (Cy^k(t) + Dv^k(t))\,dt. \tag{4.10}$$

Referring to (4.7), for $u = (z,w) \in \mathcal{X}$ we have

$$\begin{aligned} \langle A(y^k, v^k) - b, A(z,w)\rangle &= \int_0^1 \left\langle r^k(\theta), z(\theta) - \int_0^\theta (Cz(t) + Dw(t))\,dt \right\rangle d\theta \\ &= \int_0^1 \langle r^k(\theta), z(\theta)\rangle\, d\theta - \int_0^1 \int_t^1 \langle r^k(\theta), Cz(t) + Dw(t)\rangle\, d\theta\, dt \\ &= \int_0^1 \left(\langle r^k(t), z(t)\rangle - \left\langle \int_t^1 r^k(\theta)\,d\theta, Cz(t) + Dw(t) \right\rangle \right) dt. \end{aligned}$$

Introducing

$$q_C^k(t) = r^k(t) - C^T \int_t^1 r^k(\theta)\,d\theta, \tag{4.11}$$

$$q_D^k(t) = -D^T \int_t^1 r^k(\theta)\,d\theta, \tag{4.12}$$

where T denotes the transposition, we obtain

$$\langle A(y^k, v^k) - b, A(z,w)\rangle = \int_0^1 \left(\langle q_C^k(t), z(t)\rangle + \langle q_D^k(t), w(t)\rangle \right) dt. \tag{4.13}$$

Let us define for $\theta \in [0,1]$

$$\eta(\theta) = \int_0^\theta \Big(\langle q_C^k(t), z(t)\rangle + \langle q_D^k(t), w(t)\rangle \Big)\, dt,$$

and let

$$\eta_1^k = \langle A(y^k, v^k) - b, b\rangle = \Big\langle \int_0^1 r^k(\theta)\, d\theta, y_0 \Big\rangle. \tag{4.14}$$

We arrive at the following specification of (3.2), (3.4): $u^k = (z^k, w^k)$ is an optimal control in the problem

$$\min f(z,w) \tag{4.15}$$

$$\dot\eta(t) = \Big(q_C^k(t)\Big)^T z(t) + \Big(q_D^k(t)\Big)^T w(t) \quad \text{a.a.} \quad t \in [0,1], \tag{4.16}$$

$$\eta(0) = 0, \quad \eta(1) = \eta_1^k, \tag{4.17}$$

$$(z,w) \in \mathcal{X}, \quad (z(t), w(t)) \in M \quad \text{a.a.} \quad t \in [0,1]. \tag{4.18}$$

To sum up, we describe the constraint aggregation algorithm as follows.

Step 0. Fix $(y^0, v^0) \in X$ such that $f(y^0, v^0) \le f^*$ (in particular, (y^0, v^0) may be a minimizer of f over X).

Step k.

(i) Given the kth approximate solution $(y^k, v^k) \in X$, build functions r^k (4.10), q_C^k (4.11), q_D^k (4.12) and calculate the bound η_1 (4.14).

(ii) Find measurable functions (z^k, w^k) which constitute the optimal controls in the problem (4.15)–(4.18).

(iii) Calculate the function $\rho^k \in \mathcal{L}([0,1], \mathbb{R}^n)$ given by

$$\rho^k(\theta) = A(z^k, w^k)(\theta) - b(\theta) = z^k(\theta) - y_0 - \int_0^\theta (Cz^k(t) + Dw^k(t))\, dt, \tag{4.19}$$

and the stepsize

$$\tau_k = \arg\min_{0 \le \tau \le 1} \Big\| (1-\tau) r^k + \tau \rho^k \Big\|^2.$$

Note that $r^k \perp \rho^k$ by (4.13) and (4.17).

(iv) Form the $(k+1)$th approximate solution

$$(y^{k+1}, v^{k+1}) = (y^k, v^k) + \tau_k((z^k, w^k) - (y^k, v^k)).$$

Let us call the above algorithm *the constraint aggregation algorithm for problem (4.1)–(4.4)* .

The following convergence theorem is a direct consequence of Theorem 3.1.

Theorem 4.2. *The sequence* $\{(y^k, v^k)\}$ *generated by the constraint aggregation algorithm for problem (4.1)–(4.4)*

(i) *weakly converges in* $\mathcal{X} = \mathcal{L}^2([0,1], \mathbb{R}^n) \times \mathcal{L}^2([0,1], \mathbb{R}^m)$ *to the solution set* Ω^* *of this problem; and*

(ii) *strongly converges in* $\mathcal{X}$ *to* Ω^* *if the functional* f *is strongly convex.*

Remark 4.2. Obviously f is strongly convex if $(y, v) \mapsto \varphi(t, y, s)$ is *uniformly strongly convex*, i.e. there exists $\kappa > 0$ such that for all $t \in [0, 1]$, $(y_1, v_1), (y_2, v_2) \in \mathbb{R}^n \times \mathbb{R}^m$, and $\beta \in [0, 1]$, the following inequality holds

$$\begin{aligned}\varphi(t, \beta y_1 + (1-\beta)y_2, \beta v_1 + (1-\beta)v_2) \quad \leq \quad & \beta\varphi(t, y_1, v_1) + (1-\beta))\varphi(t, y_2, v_2) - \\ & \beta(1-\beta)\kappa\|(y_1 - y_2, v_1 - v_2)\|^2.\end{aligned}$$

A typical example is $\varphi(t, y, v) = \varphi_0(t, y, v) + d_1\|y\|^2 + d_2\|v\|^2$ where $d_1, d_2 > 0$, φ_0 is continuous, and $\varphi_0(t, \cdot, \cdot)$ is convex for each $t \in [0, 1]$.

Proof of Theorem 4.2. We have already shown that the sequence $\{x^k\} = \{(y^k, v^k)\}$ is constructed via a specification of the constraint aggregation method (3.1), (3.2)–(3.4) of section 3. Therefore, by Theorem 3.1, this sequence weakly converges to the solution set X^* of the problem (2.6), and, by Remark 3.2, strongly converges to X^*, if the functional f is strongly convex. Since the problem (2.6) is equivalent to (4.1)–(4.4), as stated in Theorem 4.1, the proof is complete. □

Let us consider stronger types of convergence of sequences of form $\{\bar{y}^k, v^k\}$, which require a uniform convergence of $\{\bar{y}^k\}$. Let us introduce space $\mathcal{Y} = \mathcal{C}([0, 1], \mathbb{R}^n) \times \mathcal{L}^2([0, 1], \mathbb{R}^m)$. We shall say that a sequence $\{\bar{y}^k, v^k\}$ *$\mathcal{Y}$-strongly converges* to Ω^* if the sequence of $\mathcal{Y}$-distances from $(\bar{y}^k, v^k)$ to Ω^* converges to zero, and *$\mathcal{Y}$-weakly converges* to Ω^* if for every finite collection $g_1, ..., g_m$ from $\mathcal{L}^2([0, 1], \mathbb{R}^m)$ one has

$$\inf_{(y,v)\in\Omega^*} \left\{ \|y^k - y\|_c + \max_{1\leq j\leq m} |\langle g_j, v^k - v\rangle| \right\} \to 0,$$

where $\|\cdot\|_c$ stands for the norm in $\mathcal{C}([0, 1], \mathbb{R}^n)$. These types of convergence hold for the functions $\bar{y}^k$ defined as the solutions to the Cauchy problems

$$\dot{y}(t) = Cy(t) + Dv^k(t) \quad \text{a.a.} \quad t \in [0, 1], \quad y(0) = y_0. \tag{4.20}$$

Adding the computation of $\bar{y}^k$ to Step k of the above algorithm we define *the modified constraint aggregation algorithm* for the problem (4.1)–(4.4).

Remark 4.3. Note that for the sequence $\{(y^k, \bar{y}^k, v^k)\}$ formed by this algorithm for the problem (4.1)–(4.4), the sequence $\{(\bar{y}^k, v^k)\}$ is a sequence of *control processes* [14], i.e., for each k the function $\bar{y}^k$ is the trajectory of the system (4.2) corresponding to the control v^k.

To formulate the result, we need the next definition. A functional f will be called *uniformly strongly convex in control* if there exists $\kappa > 0$ such that for all $y \in \mathcal{C}([0, 1], \mathbb{R}^n)$, $v_1, v_2 \in \mathcal{L}^2([0, 1], \mathbb{R}^m)$ and $\beta \in [0, 1]$, the following inequality holds

$$f(y, \beta v_1 + (1-\beta)v_2) \leq \beta f(y, v_1) + (1-\beta)f(y, v_2) - \beta(1-\beta)\kappa\|v_1 - v_2\|_2^2$$

where $\|\cdot\|_2$ is the norm in $\mathcal{L}^2([0, 1], \mathbb{R}^m)$.

Remark 4.4. Obviously f is uniformly strongly convex in control if $v \mapsto \varphi(t, y, v)$ is *uniformly strongly convex*, i.e. there exists a $\kappa > 0$ such that for all $t \in [0, 1]$, $y \in \mathbb{R}^n$, $v_1, v_2 \in \mathbb{R}^m$ and $\beta \in [0, 1]$, it holds that

$$\varphi(t, y, \beta v_1 + (1-\beta)v_2) \leq \beta\varphi(t, y, v_1) + (1-\beta)\varphi(t, y, v_2) - \beta(1-\beta)\kappa\|v_1 - v_2\|^2.$$

A typical example is $\varphi(t, y, v) = \varphi_0(t, y, v) + d_0\|v\|^2$ where $d_0 > 0$, φ_0 is continuous, and $\varphi_0(t, \cdot, \cdot)$ is convex for each $t \in [0, 1]$.

Theorem 4.3. *Let the sequence* $\{(y^k, \bar{y}^k, v^k)\}$ *be formed by the modified constraint aggregation algorithm for the problem (4.1)–(4.4) Then the corresponding sequence of control processes* $\{(\bar{y}^k, v^k)\}$
(i) $\mathcal{Y}$*-weakly converges to the solution set* Ω^* *of this problem; and*
(ii) $\mathcal{Y}$*-strongly converges to* Ω^* *if the functional* f *is uniformly strongly convex in control.*

Proof. Let us prove (i) by contradiction. Suppose that $\{(\bar{y}^k, v^k)\}$ is not $\mathcal{Y}$-weakly convergent to Ω^*. Then there is a subsequence $\{(y^{k_i}, \bar{y}^{k_i}, v^{k_i})\}$ such that for certain $g_1, \dots, g_m \in \mathcal{L}^2([0,1], \mathbb{R}^m)$,

$$\inf_{(y,v)\in\Omega^*} \left\{ \|\bar{y}^{k_i} - y\|_c + \max_{1\le j\le m} |\, \langle g_j, v^{k_i} - v\rangle \,| \right\} > \varepsilon > 0. \tag{4.21}$$

Since the set X given by (4.9) (closed and bounded in $\mathcal{X}$) is weakly compact in $\mathcal{X}$ and the sequence $\{\bar{y}^k\}$ is contained in a compact set in $\mathcal{C}([0,1], \mathbb{R}^m)$ (of equicontinuous functions), with no loss of generality we may assume that

$$y^{k_i} \to y_* \quad \text{weakly in} \quad \mathcal{L}^2([0,1], \mathbb{R}^n), \tag{4.22}$$

$$v^{k_i} \to v_* \quad \text{weakly in} \quad \mathcal{L}^2([0,1], \mathbb{R}^m), \tag{4.23}$$

$$\bar{y}^{k_i} \to \bar{y}_* \quad \text{in} \quad \mathcal{C}([0,1], \mathbb{R}^n). \tag{4.24}$$

Relations (4.23), (4.24) and assumption (4.21) imply that

$$(\bar{y}_*, v_*) \notin \Omega^*. \tag{4.25}$$

By Theorem 4.2, $\{y^k, v^k\}$ converges to X^* weakly in $\mathcal{X}$, which together with (4.22), (4.23) yields

$$(y_*, v_*) \in X^*. \tag{4.26}$$

This implies in particular that the equality constraint $A(y_*, v_*) = b$ in problem (2.6) is satisfied. Referring to the definitions (4.7), (4.8) of the operator A and the element b, we specify:

$$y_*(\theta) - \int_0^\theta (Cy_*(t) + Dv_*(t))\, dt = y_0 \quad \text{a.a} \quad \theta \in [0,1].$$

Hence

$$y_*(\theta) = y_{**}(\theta) \quad \text{a.a} \quad \theta \in [0,1], \tag{4.27}$$

where y_{**} is the solution to the Cauchy problem

$$\dot{y}(t) = Cy(t) + Dv_*(t) \quad \text{a.a.} \quad t \in [0,1], \quad y(0) = y_0.$$

On the other hand, the weak convergence (4.23) and the fact that $\bar{y}^k$ are solutions to the Cauchy problems (4.20) imply that $\bar{y}^{k_i} \to y_{**}$ in $\mathcal{C}([0,1], \mathbb{R}^n)$. Consequently, $\bar{y}_* = y_{**}$ and, by (4.27), $\bar{y}_*(\theta) = y_*(\theta)$ for a.a. $\theta \in [0,1]$. These facts together with (4.26) imply by Theorem 4.1 that $(\bar{y}_*, v_*) \in \Omega^*$, which contradicts (4.25).

Let us now prove (ii). Suppose this is not true, i.e., there is a subsequence $\{\bar{y}^{k_i}, v^{k_i}\}$ such that

$$\inf_{(y,v)\in\Omega^*} \left\{ \|\bar{y}^{k_i} - y\|_c + \|v^{k_i} - v\| \right\} > \varepsilon > 0. \tag{4.28}$$

As above, with no loss of generality we assume that the relations (4.22), (4.23), (4.24) take place. Thus $(\bar{y}_*, v_*) \in \Omega^*$. To complete the proof by contradiction with (4.28) it is sufficient to show that

$$v^{k_i} \to v_* \quad \text{strongly in} \quad \mathcal{L}^2([0,1], \mathbb{R}^m). \tag{4.29}$$

Relations (4.24), (4.23), the weak lower semicontinuity of $f : \mathcal{X} \mapsto \mathbb{R}$ and the fact that $(\bar{y}_*, v_*) \in \Omega^*$ yield that

$$f(\bar{y}^{k_i}, v^{k_i}) \to f(\bar{y}_*, v_*) = f^*.$$

The continuity of φ and the convergence (4.24) imply

$$| f(\bar{y}^{k_i}, v^{k_i}) - f(\bar{y}_*, v^{k_i}) | \to 0$$

Hence

$$f(\bar{y}_*, v^{k_i}) \to f(\bar{y}_*, v_*).$$

The last relation, by virtue of the strong convexity of the functional $v \mapsto f(\bar{y}_*, v)$, guarantees the convergence (4.29), which contradicts (4.28). □

Let us end this section with some observations concerning the auxiliary problem (4.15)–(4.18) that needs to be solved at each iteration of the method. Both direction vectors, z^k and w^k, which correspond to the state trajectory and the control function in the original problem, are control functions in the auxiliary problem. Therefore the constraints (4.18) in the auxiliary problem apply only to the controls. Moreover, the state equation (4.16) is one-dimensional, and particularly simple, because the 'aggregate state' does not appear at the right hand side. We shall show how these features can be exploited in solving (4.15)–(4.18).

Using the standard terminology of control theory (see [14]), we call pairs $(y, z) \in \Omega$ satisfying (4.2) and (4.3) *control processes.* Notation intM will stand for the interior of M.

Lemma 4.1. *Assume that there exists a control process $(\widetilde{y}, \widetilde{v})$ such that $(\widetilde{y}(t), \widetilde{v}(t)) \in \text{int}M$ for all $t \in [0, 1]$. Furthermore, let the function $(y, v) \mapsto \varphi(t, y, v)$ be strictly convex for every $t \in [0, 1]$. Then*

(i) *For every $t \in [0, 1]$ and every $\lambda \in \mathbb{R}$, there is a unique minimizer $(z^k(\lambda, t), w^k(\lambda, t))$ in the problem*

$$\min_{(z,w)\in M} \left[\varphi(t, z, w) + \lambda\Big((q_C^k(t))^T z + (q_D^k(t))^T w \Big) \right], \tag{4.30}$$

and the map $(\lambda, t) \mapsto (z^k(\lambda, t), w^k(\lambda, t))$ is continuous;

(ii) *There exists λ^k such that*

$$\int_0^1 \left(\left(q_C^k(t) \right)^T z^k(\lambda^k, t) + \left(q_D^k(t) \right)^T w^k(\lambda^k, t) \right) dt = \eta_1^k; \tag{4.31}$$

(iii) *$u^k = (z^k(\lambda^k, \cdot), w^k(\lambda^k, \cdot))$ solves the problem (5.19)–(5.24).*

Proof. Assertion (i) follows from the strict convexity of $\varphi(t, \cdot, \cdot)$. To prove (ii) suppose that (4.31) is not satisfied for $\lambda^k = 0$, for example,

$$\int_0^1 \left(\left(q_C^k(t) \right)^T z^k(0, t) + \left(q_D^k(t) \right)^T w^k(0, t) \right) dt > \eta_1^k \tag{4.32}$$

(the opposite inequality is treated similarly). For the control process $(\widetilde{y}, \widetilde{v})$ we have $\langle A(y^k, v^k) - b, A(\widetilde{y}, \widetilde{v}) \rangle = \eta_1^k$, or, equivalently (see (4.13)),

$$\int_0^1 \left(\left(q_C^k(t) \right)^T \widetilde{y}(t) + \left(q_D^k(t) \right)^T \widetilde{v}(t) \right) dt = \eta_1^k. \tag{4.33}$$

Define $z^k_\varepsilon(t) = \widetilde{y}(t) - \varepsilon(z^k(0,t) - \widetilde{y}(t))$ and, similarly, $w^k_\varepsilon(t)) = \widetilde{v}(t) - \varepsilon(w^k(0,t) - \widetilde{v}(t))$, where $\varepsilon > 0$. Since $(\widetilde{y}(t), \widetilde{v}(t)) \in intM$ for all $t \in [0,1]$, we have $(z^k_\varepsilon(t), w^k_\varepsilon(t)) \in M$ for all $t \in [0,1]$, if $\varepsilon > 0$ is sufficiently small. In view of (4.32) and (4.33),

$$\int_0^1 \left(\left(q^k_C(t)\right)^T z^k_\varepsilon(t) + \left(q^k_D(t)\right)^T w^k_\varepsilon(t) \right) dt < \eta^k_1.$$

Therefore, the minimum value in the problem

$$\min_{(z,w)\in\Omega} \int_0^1 \left(\left(q^k_C(t)\right)^T z(t) + \left(q^k_D(t)\right)^T w(t) \right) dt,$$

is smaller than $\eta^k_1 - \delta$ with some positive δ. Hence, the minimum value in the perturbed problem

$$\min_{(z,w)\in\Omega} \left[\alpha f(z,w) + \int_0^1 \left(\left(q^k_C(t)\right)^T z(t) + \left(q^k_D(t)\right)^T w(t) \right) dt \right],$$

is smaller than $\eta^k_1 - \delta/2$, provided $\alpha > 0$ is sufficiently small. In the latter problem, $(z^k(1/\alpha,\cdot), w^k(1/\alpha,\cdot))$ is obviously a minimizer. Setting α so small that $\alpha\|f(z,w)\| < \delta/2$ for all $(z,w) \in \Omega$, we get

$$\int_0^1 \left(\left(q^k_C(t)\right)^T z^k(1/\alpha,t) + \left(q^k_D(t)\right)^T w^k(1/\alpha,t) \right) dt < \eta^k_1. \tag{4.34}$$

Since the map $(\lambda,t) \mapsto (z^k(\lambda,t), w^k(\lambda,t))$ is continuous, (4.34) and (4.32) yield (4.31) for some $\lambda^k \in (0, 1/\alpha)$.

The equality (4.31) shows that $(z^k(\lambda^k,\cdot), w^k(\lambda^k,\cdot))$ is feasible in the problem (5.19)–(5.24). For every (z,w) feasible in this problem we have

$$\begin{aligned} & f(z^k(\lambda^k,\cdot), w^k(\lambda^k,\cdot)) + \lambda^k \eta^k_1 \\ = \;& f(z^k(\lambda^k,\cdot), w^k(\lambda^k,\cdot)) + \lambda^k \int_0^1 \left(\left(q^k_C(t)\right)^T z^k(\lambda^k,t) + \left(q^k_D(t)\right)^T w^k(\lambda^k,t) \right) dt \\ \le \;& f(z,w) + \lambda^k \int_0^1 \left(\left(q^k_C(t)\right)^T z(t) + \left(q^k_D(t)\right)^T w(t) \right) dt \;=\; f(z,w) + \lambda^k \eta^k_1, \end{aligned}$$

hence, $f(z^k(\lambda^k,\cdot), w^k(\lambda^k,\cdot)) \le f(z,w)$. This proves that $(z^k(\lambda^k,\cdot), w^k(\lambda^k,\cdot))$ solves the problem (5.19)–(5.24). □

Remark 4.5. Lemma 4.1 shows that if the solutions $(z^k(\lambda,t), w^k(\lambda,t))$ of the finite-dimensional problems (4.30) are given explicitly, then the major operation in the constraint aggregation algorithm (Step k, (ii)) is reduced to an one-dimensional algebraic equation (4.31). For example, for a linear-quadratic integrand

$$\varphi(t,y,v) = c^T y + \alpha\|y\|^2 + d^T v + \beta\|v\|^2$$

($\alpha > 0$, $\beta > 0$) and with box constraints

$$M = \{(y,v) \in R^n \times R^m : y^-_i \le y_i \le y^+_i \ (i = 1,\ldots,n),\ v^-_j \le v_j \le v^+_j \ (j = 1,\ldots,m)\},$$

we have

$$z^k(\lambda,t) = \Big[-(c + \lambda q^k_C(t))/(2\alpha) \Big]_{y^-}^{y^+}$$

and

$$w^k(\lambda,t) = \Big[-(d + \lambda q^k_D(t))/(2\beta) \Big]_{v^-}^{v^+},$$

where $[\,\cdot\,]^u_l$ denotes the orthogonal projection on the box $[l,u]$. These formulae can be substituted into the algebraic equation (4.31).

5 Convex processes

In this section, we solve a more complicated problem of optimal control employing the constraint aggregation method specialized for both equality and inequality constraints. Instead of a linear control system (4.1)–(4.4), we treat a more general *convex process* described by a differential inclusion with a convex graph (see [2]). The problem formulation is as follows:

$$\min\ f(y) \tag{5.1}$$

$$\dot y(t) \in F(y(t)) \quad \text{a.a.} \quad t \in [0,1], \tag{5.2}$$

$$y(0) = y_0, \tag{5.3}$$

$$y \in \Omega. \tag{5.4}$$

Here F is a (set-valued) map from $\mathbb{R}^n$ into the set of all nonempty convex compacta in $\mathbb{R}^n$, $y_0 \in \mathbb{R}^n$,

$$\Omega = \{x \in \mathcal{A}([0,1], \mathbb{R}^n) : (x(t), \dot x(t)) \in M \text{ for a.a } t \in [0,1]\}, \tag{5.5}$$

and

$$f(x) = \int_0^1 \varphi(t, x(t), \dot x(t))\, dt. \tag{5.6}$$

In (5.5) M is a convex compactum in $\mathbb{R}^n \times \mathbb{R}^n$. The function φ in (5.6) is continuous, and for each $t \in [0,1]$ the function $(y, v) \mapsto \varphi(t, y, v)$ is convex.

Let us return to the set-valued map F. Let $M_0 = \{y \in \mathbb{R}^n : (y, v) \in M\}$. We assume that the set-valued map $y \mapsto F(y)$ has a convex graph on M_0, i.e. for every $y_1, y_2 \in M_0$, every $v_1 \in F(y_1)$, $v_2 \in F(y_2)$ and every $\beta \in [0,1]$, one has $\beta v_1 + (1-\beta) v_2 \in F(\beta x_1 + (1-\beta) x_2)$.

Remark 5.1. Let F be determined through the velocity sets of the linear control system (4.2), i.e. $F(x) = \{Cx + Dv : v \in V\}$ where V is a convex compactum in $\mathbb{R}^m$. Then F has a convex graph on M_0.

Remark 5.2. The fact that F has a convex graph on M_0 implies that F is continuous on M_0.

We assume that the feasible set of problem (5.1)–(5.4) is nonempty.

Remark 5.3. The solution method described below works in a more general situation where F and M are nonstationary, measurably dependent on time. We however focus on the simplest case.

We first reduce problem (5.1)–(5.4) to a convex optimization problem of the form (2.1)–(2.4) with equality and inequality constraints. Let us rewrite (5.2) in an integral form:

$$y(\theta) = y_0 + \int_0^\theta v(t)\, dt, \quad \theta \in [0,1], \tag{5.7}$$

where

$$v(t) \in F(y(t)) \quad \text{for a.a. } t \in [0,1].$$

The latter is equivalent to the system of inequalities

$$\langle s, v(t)\rangle \le \rho(s, F(y(t))) \quad \text{for } s \in \Sigma^{n-1}, \quad \text{and a.a. } t \in [0,1]. \tag{5.8}$$

Here Σ^{n-1} is the unit sphere in $\mathbb{R}^n$, $\Sigma^{n-1} = \{s \in \mathbb{R}^n : \|s\| = 1\}$, and $\rho(\cdot, B)$ is the support function of the set B defined on Σ^{n-1} by

$$\rho(s, B) = \sup\{\langle s, v\rangle : v \in B\}$$

(see [16]). For reasons that will become clear later, it is convenient to consider on Σ^{n-1} a measure ν proportional to the Lebesgue measure and such that $\nu(\Sigma^{n-1}) = 1$.

Remark 5.4. The fact that F has a convex graph easily implies that for every $s \in \Sigma^{n-1}$ the function $y \mapsto \rho(s, F(y))$ is concave on M_0.

Remark 5.5. It follows from the continuity of F on M_0 (Remark 5.2) that the function $t \mapsto \rho(s, F(y(t)))$ is continuous for every $y \in \mathcal{C}([0,1], \mathbb{R}^n)$ taking values in M_0.

Since $s \mapsto \rho(s, F(y))$ is continuous, (5.8) is equivalent to

$$(s, v(t)) \leq \rho(s, F(y(t))) \quad \text{for a.a. } (s,t) \in \Sigma^{n-1} \times [0,1], \tag{5.9}$$

where a.a. is understood with respect to the product of the measure ν on Σ^{n-1} and the Lebesgue measure on $[0,1]$. Let us define $\mathcal{X} = \mathcal{L}^2([0,1], \mathbb{R}^n) \times \mathcal{L}^2([0,1], \mathbb{R}^n)$, and $\mathcal{H} = \mathcal{L}^2([0,1], \mathbb{R}^n)$. Treating the functions y and v as elements of $\mathcal{X}$, we may rewrite (5.7), (5.9) as

$$A(y, v) = b,$$

$$h(s, t, y, v) \leq 0 \quad \text{for a.a. } (s,t) \in S = \Sigma^{n-1} \times [0,1].$$

Here A is a linear bounded operator from $\mathcal{X}$ to $\mathcal{H}$ given by

$$A(y, v)(\theta) = y(\theta) - \int_0^\theta v(t)\, dt, \quad \theta \in [0,1], \tag{5.10}$$

$$b(\theta) = y_0, \quad \theta \in [0,1], \tag{5.11}$$

and

$$h(s, t, y, v) = \langle s, v(t)\rangle - \rho(s, F(y(t))). \tag{5.12}$$

The set (5.5) is naturally transformed into

$$X = \{(y, v) \in \mathcal{X} : (y(t), v(t)) \in M \text{ for a.a } t \in [0,1]\}. \tag{5.13}$$

Obviously, X is closed, convex, and bounded in $\mathcal{X}$. Let us show that the functions $(y, v) \mapsto h(s, t, y, v)$, $(s,t) \in \Sigma^{n-1} \times [0,1]$, which are defined by (5.12), satisfy the conditions of section 3 with $S = \Sigma^{n-1} \times [0,1]$ and with the measure μ on S being the product of the measure ν on Σ^{n-1} and the Lebesgue measure on $[0,1]$. By Remark 5.4 these functions are convex on X. By the continuity of the map F on M_0 (Remark 5.2) this map is bounded on M_0. Hence the family $\rho(\cdot, F(y))$, $y \in M_0$, is equicontinuous. This and the continuity of $\rho(s, F(\cdot))$ following from Remark 5.2 imply that the function $(s, y) \mapsto \rho(s, F(y))$ is continuous on $\Sigma^{n-1} \times M_0$. Therefore, for each $(y, v) \in X$ the function $(s,t) \mapsto \rho(s, F(y(t)))$ is bounded and measurable. We conclude that for each $(y, v) \in X$ the function $(s, t) \mapsto h(s, t, y, v)$ together with $(s, t) \mapsto h^+(s, t, y, v) = \max\{h(s, t, y, v), 0\}$, belong to the space $\mathcal{L}^2_S = \mathcal{L}^2(S, \mu, \mathbb{R})$, and $\overline{h}(s, t) = \sup\{| h(s, t, y, s) | : (y, s) \in X\}$ lies in $\mathcal{L}^2_S$, too.

Thus, the optimization problem (2.1)–(2.4) where the operator A, element b, functions h and set X are given by (5.10), (5.11), (5.12), (5.13) satisfies all the assumptions of section 3, and can be solved with the constraint aggregation method described in Theorem 3.3.

Theorem 5.1. *Let the operator $A : \mathcal{X} \mapsto \mathcal{H}$, element $b \in \mathcal{H}$, function h and set $X \in \mathcal{X}$ be defined by (5.10), (5.11), (5.12) and (5.13). The optimal control problem (5.1)–(5.4) is equivalent to the convex optimization problem (2.1)–(2.4) in the following sense:*
(i) *each solution $x^* = (y^*, v^*)$ of problem (5.1)–(5.4) solves problem (2.1)–(2.4),*
(ii) *if $x^* = (y^*, v^*)$ solves problem (2.1)–(2.4), then there exists a solution $x^{**} = (y^{**}, v^{**})$ of problem (5.1)–(5.4) such that $v^{**} = v^*$ and $y^{**}(t) = y^*(t)$ for a.a. $t \in [0, 1]$.*

The theorem is obvious.

In the rest of this section the convex optimization problem (2.1)–(2.4) is understood as described in Theorem 5.1. Its solution set will be denoted by X^*; for the solution set of the initial problem (5.1)–(5.4) we shall use the notation Ω^*. For solving problem (2.1)–(2.4) we employ the constraint aggregation method (3.1), (3.11)–(3.14) and (3.15) of section 2 (see Theorem 3.3). We shall first specify u^k in (3.11)–(3.14). Let $x^k = (y^k, v^k)$. The term $(Ax^k - b, Au - b)$ in (3.12), where $u = (z, w) \in \mathcal{X}$, has the form

$$\langle A(y^k, v^k) - b, A(z, w) \rangle = \int_0^1 \Big(\langle r^k(t), z(t) \rangle + \langle q^k(t), w(t) \rangle \Big) \, dt, \tag{5.14}$$

where

$$r^k(\theta) = A(y^k, v^k)(\theta) = y^k(\theta) - \int_0^\theta v^k(t) \, dt, \tag{5.15}$$

$$q^k(t) = - \int_t^1 r^k(\theta) \, d\theta. \tag{5.16}$$

We obtain this arguing like in the previous section (with $C = 0$ and D equal to the identity matrix).

The integral in the constraint (3.13) has the form

$$\begin{aligned} &\int_{\Sigma^{n-1} \times [0,1]} h^+(s, t, y^k, v^k) h(s, t, z, w) \, \nu(ds) \, dt \\ &= \int_0^1 \int_{\Sigma^{n-1}} h^k(s, t) \Big(\langle s, w(t) \rangle - \rho(s, F(z(t))) \Big) \, \nu(ds) \, dt \end{aligned} \tag{5.17}$$

where

$$h^k(s, t) = h^+(s, t, y^k, v^k) = \max\{0, \langle s, v^k(t) \rangle - \rho(s, F(y^k(t)))\}. \tag{5.18}$$

We arrive at the following specification of (3.11), (3.14): $u^k = (z^k, w^k)$ is an optimal control in the problem

$$\min \int_0^1 \varphi(t, z(t), w(t)) \, dt \tag{5.19}$$

$$\dot{\eta}(t) = \Big(r^k(t) \Big)^T z(t) + \Big(q^k(t) \Big)^T w(t) \quad \text{a.a.} \quad t \in [0, 1], \tag{5.20}$$

$$\eta(0) = 0, \quad \eta(1) = \eta_1^k = \Big\langle \int_0^1 r^k(\theta) \, d\theta, y_0 \Big\rangle, \tag{5.21}$$

$$\dot{\xi}(t) = - \int_{\Sigma^{n-1}} h^k(s, t) \rho(s, F(z(t)) - w(t)) \, \nu(ds) \quad \text{a.a.} \quad t \in [0, 1], \tag{5.22}$$

$$\xi(0) = 0, \quad \xi(1) \leq 0, \tag{5.23}$$

$$(z(t), w(t)) \in M \quad \text{a.a.} \quad t \in [0, 1]. \tag{5.24}$$

The constraint aggregation algorithm (3.1), (3.11)–(3.14). (3.15) can be summarized as follows.

Step 0. Fix $(y^0, v^0) \in X$ such that $\int_0^1 \varphi(t, y(t), v(t))\, dt \leq f^*$ (in particular, (y^0, v^0) may be the minimizer of the objective functional in X).

Step k.

(i) Given the kth approximate solution $(y^k, v^k) \in \mathcal{X}$, build functions r^k (5.15), q^k (5.16), h^k (5.18).

(ii) Find measurable functions (z^k, w^k) which constitute the optimal controls of the problem (5.19), (5.24).

(iii) Calculate the stepsize τ_k by minimizing with respect to $\tau \in [0, 1]$ the expression

$$\int_0^1 \int_{\Sigma^{n-1}} \Big(\max\{0, \langle s, (1-\tau)v^k(t) + \tau w^k(t)\rangle - \rho(s, F((1-\tau)y^k(t) + \tau z^k(t)))\Big)^2 \nu(ds)\, dt.$$

(iv) Form the $(k + 1)$st approximate solution

$$(y^{k+1}, v^{k+1}) = (y^k, v^k) + \tau_k((z^k, w^k) - (y^k, v^k)).$$

Let us call the above algorithm *the constraint aggregation algorithm for problem (5.1)–(5.4)* .

Theorem 5.2. *The sequence* $\{(y^k, v^k)\}$ *formed by the constraint aggregation algorithm for problem (5.1)–(5.4)*

(i) *weakly converges in* $\mathcal{X} = \mathcal{L}^2([0, 1], \mathbb{R}^n) \times \mathcal{L}^2([0, 1], \mathbb{R}^m)$ *to the solution set* Ω^* *of this problem; and*

(ii) *strongly converges in* $\mathcal{X}$ *to* Ω^* *if the functional* f *is strongly convex.*

The proof of Theorem 5.2 is identical to that of Theorem 4.2. Here, instead of Theorem 3.1, we refer to Theorem 3.3 and use the reduction Theorem 5.1.

There are intriguing connections of the method just described with Steiner selections of convex maps. We shall use them to develop a modified constraint aggregation method for convex processes.

Recall that the *Steiner point* of a convex compact set $B \subset \mathbb{R}^n$ is defined as

$$\mathrm{St}(B) = n \int_{\Sigma^{n-1}} s\rho(s, B)\, \nu(ds),$$

where ν is a uniform measure on Σ^{n-1} satisfying $\nu(\Sigma^{n-1}) = 1$, and $\rho(\cdot, B)$ is the support function of B (see [21] and [20, §3.4]).

Let us define the mappings:

$$F^k(t) = F(y^k(t)), \quad F_+^k(t) = conv\left\{F^k(t) \cup \{v^k(t)\}\right\}.$$

It is evident that

$$h^k(s, t) = \max\{0, \langle s, v^k(t)\rangle - \rho(s, F(y^k(t)))\} = \rho(s, F_+^k(t)) - \rho(s, F^k(t)).$$

Therefore the first component of the inner integral at the right side of (5.17) can be transformed as follows:

$$\begin{aligned} &\int_{\Sigma^{n-1}} h^k(s, t)\langle s, w(t)\rangle\, \nu(ds) \\ &= \left\langle w(t), \int_{\Sigma^{n-1}} s\Big(\rho(s, F_+^k(t)) - \rho(s, F^k(t))\Big)\, \nu(ds)\right\rangle \\ &= \frac{1}{n}\Big\langle w(t), \mathrm{St}(F_+^k(t)) - \mathrm{St}(F^k(t))\Big\rangle. \end{aligned} \qquad (5.25)$$

Consider the second component of the inner integral in the right hand side of (5.17). We have

$$
\begin{aligned}
&\int_{\Sigma^{n-1}} h^k(s,t)\rho(s,F(z(t))\,\nu(ds) \\
&\quad= \int_{\Sigma^{n-1}} \Big(\rho(s,F_+^k(t)) - \rho(s,F^k(t))\Big)\rho(s,F(z(t))\,\nu(ds) \\
&\quad= \int_{\Sigma^{n-1}} \max_{p\in F(z(t))} \Big\langle \Big(\rho(s,F_+^k(t)) - \rho(s,F^k(t))\Big)s, p\Big\rangle\,\nu(ds) \\
&\quad\geq \max_{p\in F(z(t))} \int_{\Sigma^{n-1}} \Big\langle \Big(\rho(s,F_+^k(t)) - \rho(s,F^k(t))\Big)s, p\Big\rangle\,\nu(ds) \\
&\quad= \max_{p\in F(z(t))} \frac{1}{n}\langle \mathrm{St}(F_+^k(t)) - \mathrm{St}(F^k(t)), p\rangle \\
&\quad= \frac{1}{n}\rho(\mathrm{St}(F_+^k(t)) - \mathrm{St}(F^k(t)), F(z(t))).
\end{aligned}
$$

In view of (5.25), the expression (5.17) can be estimated as follows:

$$
\begin{aligned}
&\int_{\Sigma^{n-1}\times[0,1]} h^+(s,t,y^k,v^k)h(s,t,z,w)\,\nu(ds)\,dt \\
&\leq \frac{1}{n}\int_0^1 \langle w(t), \mathrm{St}(F_+^k(t)) - \mathrm{St}(F^k(t))\rangle\,dt - \frac{1}{n}\int_0^1 \rho\left(\mathrm{St}(F_+^k(t)) - \mathrm{St}(F^k(t)), F(z(t))\right)\,dt \\
&= -\frac{1}{n}\int_0^1 \rho\left(\mathrm{St}(F_+^k(t)) - \mathrm{St}(F^k(t)), F(z(t)) - w(t)\right)\,dt \;=\; g^k(z,w).
\end{aligned}
$$

We see that the estimate (3.20) holds. Furtheremore, for every $u = (z,w)$ feasible in the problem (5.1)–(5.4) we have $w(t) \in F(z(t))$ for a.a. $t \in [0,1]$. Consequently, $\rho(s, F(z(t)) - w(t)) \geq 0$ for all $s \in \Sigma^{n-1}$ and a.a. $t \in [0,1]$, and we get $g^k(z,w) \leq 0$ (see (3.18)). With a view to Theorems 5.1 and 3.4, we arrive at the following modification of the constraint aggregation algorithm (5.19)–(5.24): $u^k = (z^k, w^k)$ is an optimal control in the problem

$$
\begin{aligned}
&\min \int_0^1 \varphi(t, z(t), w(t))\,dt, \\
&\dot\eta(t) = \left(r^k(t)\right)^T z(t) + \left(q^k(t)\right)^T w(t) \quad \text{a.a.} \quad t\in[0,1], \\
&\eta(0) = 0, \quad \eta(1) = \eta_1^k = \Big\langle \int_0^1 r^k(\theta)\,d\theta, y_0\Big\rangle, \\
&\dot\xi(t) = -\frac{1}{n}\rho\Big(\mathrm{St}(F_+^k(t)) - \mathrm{St}(F^k(t)), F(z(t)) - w(t)\Big), \\
&\xi(0) = 0, \quad \xi(1) \leq 0, \\
&(z(t), w(t)) \in M \quad \text{a.a.} \quad t\in[0,1].
\end{aligned}
$$

Let us call the above algorithm the *Steiner constraint aggregation algorithm* for problem (5.19)–(5.24).

The reduction Theorem 5.1 and Theorem 3.4 yield the following result.

Theorem 5.3. *The sequence $\{(y^k, v^k)\}$ generated by the Steiner constraint aggregation algorithm for problem (5.1)–(5.4)*
(i) *weakly converges in $\mathcal{X}$ to the solution set Ω^* of this problem; and*
(ii) *strongly converges to Ω^* if f is strongly convex.*

Consider stronger types of convergence of sequences $\{(\bar{y}^k, v^k)\}$, which require uniform convergence of $\{\bar{y}^k\}$. Like in the previous section, we introduce the space $\mathcal{Y} = \mathcal{C}([0,1], \mathbb{R}^n) \times \mathcal{L}^2([0,1], \mathbb{R}^n)$ and define *$\mathcal{Y}$-strong* and *$\mathcal{Y}$-weak convergence* of a sequence $\{(\bar{y}^k, v^k)\}$ to the solution set Ω^*. These types of convergence hold for the functions $\bar{y}^k$ given by

$$\bar{y}^k(\theta) = y_0 + \int_0^1 v^k(t)\, dt, \quad \theta \in [0,1]. \tag{5.26}$$

Adding the operation (5.26) to Step k of the constraint aggregation algorithm we define *the modified constraint aggregation algorithm for problem (5.1)–(5.4).*

Remark 5.6. Note that for the sequence $\{(y^k, \bar{y}^k, v^k)\}$ formed by this algorithm, the components $\bar{y}^k$ are trajectories of the system (5.2).

The functional f given by (5.6) will be called *uniformly strongly convex with respect to state velocity* if there exists $\kappa > 0$ such that for all $y, v_1, v_2 \in \mathcal{L}^2([0,1], \mathbb{R}^n)$, and $\beta \in [0,1]$, the following inequality holds

$$F(y, \beta v_1 + (1-\beta) v_2) \le \beta F(y, v_1) + (1-\beta) F(y, v_2) - \beta(1-\beta)\kappa \|v_1 - v_2\|_2^2$$

where

$$F(y, v) = \int_0^1 \varphi(t, y(t), v(t))\, dt,$$

and $\|\cdot\|_2$ is the norm in $\mathcal{L}^2([0,1], \mathbb{R}^m)$.

Remark 5.7. Obviously f is uniformly strongly convex with respect to state velocity if $v \mapsto \varphi(t, y, v)$ is *uniformly strongly convex* (see Remark 4.4).

Theorem 5.4. *Let the sequence $\{(y^k, \bar{y}^k, v^k)\}$ be formed by the modified constraint aggregation algorithm for the problem (5.1)–(5.4) Then the sequence $\{(\bar{y}^k, v^k)\}$*
(i) *$\mathcal{Y}$-weakly converges to the solution set Ω^* of this problem; and*
(ii) *$\mathcal{Y}$-strongly converges to Ω^* if the functional f is uniformly strongly convex with respect to state velocity.*

The proof is similar to that of Theorem 4.3. An identical modification and result are possible also for the Steiner constraint aggregation method.

6 Games

Let us now consider a zero-sum game defined as follows. There are compact sets $S \subset \mathbb{R}^n$ and $T \subset \mathbb{R}^m$, which we shall call *sets of pure strategies* of players A and B, respectively. The *payoff function* $r : S \times T \to \mathbb{R}$ assigns to each pair of pure strategies (s,t) the amount that A has to pay B. We assume that r is continuous.

A *mixed strategy* of A is a probability measure μ on S; similarly, a mixed strategy of B is a probability measure ν on T (see [1, §3.1.6]). The set of all probability measures on S, or T, will be denoted $\Pi(S)$ and $\Pi(T)$, respectively. For each pair of mixed strategies $(\mu, \nu) \in \Pi(S) \times \Pi(T)$ the expected payoff is given by

$$R(\mu, \nu) = \int_{S \times T} r(s,t)\, \mu(ds)\, \nu(dt). \tag{6.1}$$

It is well-defined, because r is continuous.

The problem is to find a *saddle point* of the game, that is, a pair of measures $(\mu^*, \nu^*) \in \Pi(S) \times \Pi(T)$ such that

$$R(\mu^*, \nu) \le R(\mu^*, \nu^*) \le R(\mu, \nu^*) \quad \text{for all} \quad (\mu, \nu) \in \Pi(S) \times \Pi(T). \tag{6.2}$$

Such a saddle point exists [1, §7.7.2].

A *symmetric* game [13] is a game in which $m = n$, $S = T$, and the function r is skew-symmetric in the following sense:

$$r(s,t) = -r(t,s), \quad \text{for all} \quad (s,t) \in S \times S.$$

In a symmetric game, for each $\mu \in \Pi(S)$ we have $R(\mu, \mu) = 0$. In particular, this must hold for μ^* and for ν^*. Setting $\mu = \nu^*$ and $\nu = \mu^*$ in (6.2) we see that $R(\mu^*, \nu^*) = 0$ and that it is sufficient to look for symmetric saddle points of form (μ^*, μ^*). This simplifies (6.2) to the system of inequalities

$$R(\mu^*, \nu) \le 0, \quad \text{for all} \quad \nu \in \Pi(S). \tag{6.3}$$

Remark 6.1. Every zero-sum game with a payoff function $r : S \times T \to \mathbb{R}$ can be reduced to a symmetric game on $(S \times T) \times (S \times T)$, by defining a skew-symmetric payoff function

$$w((s_1, t_1), (s_2, t_2)) = r(s_1, t_2) - r(s_2, t_1).$$

It is clear that every symmetric saddle point of this game corresponds to a saddle point in (6.2), and vice versa. Therefore we shall concentrate on symmetric games, to simplify the notation.

Let us reduce a symmetric game to a problem with inequality constraints of form (2.7). Define $\mathcal{X} = \mathbf{r}ca(S)$—the linear space of all regular countably additive functions on the σ-algebra $\mathcal{B}$ of all Borel sets in S. We treat $\mathcal{X}$ as the space dual to the space Ξ of continuous functions $g : S \to \mathbb{R}$ with the norm $\|g\| = \max_{s \in S} |g(s)|$.

We define $f \equiv 0$ and $X = \Pi(S)$; the latter set is convex and weakly* compact in $\mathcal{X}$ (Prohorov's theorem, see [3]).

Finally, we define the functional $h : S \times \mathcal{X} \to \mathbb{R}$ by

$$h(t, \mu) = \int_S r(s,t)\, \mu(ds). \tag{6.4}$$

It is linear with respect to μ and bounded, because r is continuous. Let us observe that the inequalities (6.3) can be equivalently expressed as

$$h(t, \mu^*) \le 0 \quad \text{for all} \quad t \in S. \tag{6.5}$$

Due to the continuity of $h(\cdot, \mu)$, the quantifier 'for all' can be replaced by 'for almost all with respect to the Lebesgue measure on S'. Moreover, the linear functionals $h(t, \cdot)$, $t \in S$, given by (6.4), are weakly* continuous.

With these definitions all assumptions of Section 3 are satisfied, and the problem of finding a saddle point of a symmetric game becomes an instance of the general formulation (2.7). Let us specify for it the method with constraint aggregation (3.1), (3.7)–(3.9) and (3.10). Denoting by μ^k the current iterate (which corresponds to x^k in section 3), we obtain from (6.4) the following specification of the left hand side of the aggregate constraint (3.8):

$$\int_S h^+(t, \mu^k) h(t, u)\, dt = \int_{S \times S} r(s,t) h^+(t, \mu^k)\, u(ds)\, dt.$$

We arrive at the following subproblem: u^k is a solution of the system

$$\int_{S\times S} r(s,t)h^+(t,\mu^k)\,u(ds)\,dt \le 0, \tag{6.6}$$

$$u \in \Pi(S). \tag{6.7}$$

The next iterate μ^{k+1} is defined by

$$\mu^{k+1} = (1-\tau_k)\mu^k + \tau_k u^k, \tag{6.8}$$

where

$$\tau_k = \arg\min_{\tau\in[0,1]} \int_S \Big(\max\Big(0, (1-\tau)\int_S r(s,t)\mu^k(ds) + \tau\int_S r(s,t)u^k(ds)\Big)\Big)^2\,dt. \tag{6.9}$$

The next theorem is a simple consequence of Theorem 3.2.

Theorem 6.1. *Let the game have a skew-symmetric and continuous payoff function $r : S\times S \to \mathbb{R}$, where S is a compact set in $\mathbb{R}^n$. Then for every $\mu^0 \in \Pi(S)$ the sequence of probability measures $\{\mu^k\}$ defined by (6.6)–(6.9) weakly converges to the set of probability measures μ^* such that (μ^*,μ^*) is a saddle point of the game.*

Let us analyse in more detail the subproblem (6.6)–(6.7). Observe that due to the absence of any objective function ($f\equiv 0$), there is much freedom in specifying u^k. We shall show that some solutions to (6.6)–(6.7) can be defined in a closed form.

One-step lookahead correction

Let us define $\alpha_k = \int_S h^+(t,\mu^k)\,dt$. If μ^k is not a solution of the game, i.e., if it does not satisfy (6.5), we have $\alpha_k > 0$. Thus we can define

$$u^k(ds) = \alpha_k^{-1}h^+(s,\mu^k)ds, \tag{6.10}$$

that is, u^k is a probability measure with the density $h^+(\cdot,\mu^k)/\alpha_k$.

Since r is skew-symmetric, in (6.6) we have

$$\begin{aligned}\int_{S\times S} r(s,t)h^+(t,\mu^k)\,u^k(ds)\,dt &= \alpha_k^{-1}\int_{S\times S} r(s,t)h^+(t,\mu^k)h^+(s,\mu^k)\,ds\,dt\\ &= \alpha_k^{-1}\int_{s\ge t}\Big((r(s,t)+r(t,s)\Big)h^+(t,\mu^k)h^+(s,\mu^k)\,ds\,dt = 0.\end{aligned}$$

Consequently, the measure (6.10) is a solution of (6.6)–(6.7). We shall call it a *one-step lookahead correction*, because u^k is a good B's response to the strategy μ^k employed by A:

$$\begin{aligned}R(\mu^k,u^k) &= \int_{S\times S} r(s,t)\,\mu^k(ds)\,u^k(dt) = \int_S h(t,\mu^k)\,u^k(dt)\\ &= \alpha_k^{-1}\int_S\Big(h^+(t,\mu^k)\Big)^2\,dt > 0.\end{aligned}$$

The step (6.8) can be thus interpreted as combining μ^k with a strategy u^k that beats μ^k. It bears some similarity with the *method of fictious play* of [15], but our correction is a mixed strategy rather than pure, and we employ the minimizing stepsize (6.9) instead of $\tau_k = 1/(k+1)$.

More generally, we may interpret the aggregate constraint (6.6) as the requirement that u^k be a sufficient response to the one-step lookahead strategy having the density $\alpha_k^{-1} h^+(\cdot, \mu^k)$. Indeed, subject to a normalizing constant, inequality (6.6) requires that the payoff of the game be non-positive. This allows us to consider other candidates for u^k.

Two-step lookahead correction

Let us define for $s \in S$

$$\lambda^k(s) = \int_S r(s,t) h^+(t, \mu^k)\, dt, \quad \lambda_-^k(s) = \max(0, -\lambda^k(s)).$$

If μ^k is not a solution, then, similarly to the one-step case, $\alpha_k = \int_S h^+(t, \mu^k)\, dt > 0$.

Suppose that $\lambda^k(s) \geq 0$ for all $s \in S$. Then for all $u \in \Pi(S)$ we have

$$0 \leq \int_S \lambda^k(s) u(ds) = \int_{S \times S} r(s,t) h^+(t, \mu^k)\, dt\, u(ds).$$

This means that the strategy $\nu(dt) = \alpha_k^{-1} h^+(t, \mu^k) dt$ provides a nonegative payoff for B, irrespective of A's strategy, that is, ν is a solution of the game.

If this is not the case, we must have $\gamma_k = \int_S \lambda_-^k(s)\, ds > 0$. Consequently, we may define the probability measure

$$u^k(ds) = \gamma_k^{-1} \lambda_-^k(s) ds \tag{6.11}$$

to guarantee that (6.6) is satisfied as a strict inequality

$$\int_{S \times S} r(s,t) h^+(t, \mu^k)\, u^k(ds)\, dt = \int_S \lambda^k(s)\, u^k(ds) = -\gamma_k^{-1} \int_S \left(\lambda_-^k(s)\right)^2 ds < 0.$$

Let us observe that the measure (6.11) is a good response of Player A to the strategy (6.10) applied by B. Therefore we shall call it a *two-step lookahead correction.*

7 Stochastic programming

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{\mathcal{F}_t\}$, $t = 1, \ldots, T$ be a filtration in $\mathcal{F}$. A sequence of random variables

$$x = (x_1, \ldots, x_T),$$

where $x_t : \Omega \to \mathbb{R}^{n_t}$, $n_t > 0$, will be called a *policy.* We consider only policies that have a bounded variance and we define the decision space

$$\mathcal{X} = \mathcal{L}^2(\Omega, \mathbb{R}^{n_1}) \times \cdots \times \mathcal{L}^2(\Omega, \mathbb{R}^{n_T}). \tag{7.1}$$

It is a Hilbert space with the scalar product

$$\langle x, y \rangle_{\mathcal{X}} = \sum_{t=1}^{T} \mathbb{E} \langle x_t, y_t \rangle.$$

A policy is *nonanticipative*, if each x_t is $\mathcal{F}_t$-measurable, $t = 1, \ldots, T$ (x is adapted to $\{\mathcal{F}_t\}$). It is evident that the set of all nonanticipative policies is a closed linear subspace of $\mathcal{X}$; it will be denoted $\mathcal{M}$.

Let G be a map from Ω to the set of all subsets of $\mathbb{R}^n$ (identified with $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_T}$), where $n = \sum_{t=1}^{T} n_t$. We assume that G is measurable, and that with probability one its value $G(\omega)$ is a nonempty convex compactum in $\mathbb{R}^n$. A policy x is called *feasible*, if

$$x(\omega) \in G(\omega) \quad \text{with probability 1.} \tag{7.2}$$

The set of all feasible policies will be denoted X. It is convex, closed and bounded in $\mathcal{X}$.

Finally, we have a functional $f : \mathcal{X} \to \mathbb{R}$, defined as

$$f(x) = \mathbb{E}\,\varphi(x(\omega), \omega), \tag{7.3}$$

where $\varphi : \mathbb{R}^n \times \Omega \to \mathbb{R}$. We assume that $\varphi(\cdot, \omega)$ is convex for $\mathbb{P}$-almost all $\omega \in \Omega$, and $\varphi(z, \cdot)$ is measurable for all $z \in \mathbb{R}^n$. Furthermore, we assume that there exists $\varepsilon > 0$ such that for all x such that $\inf_{y \in X} \|x - y\|_{\mathcal{X}} < \varepsilon$ one has

$$|\varphi(x(\omega), \omega)| \le \overline{\varphi}(\omega), \quad \text{with probability 1,}$$

and $\mathbb{E}\overline{\varphi} < \infty$. Then the functional (7.3) is convex and bounded in the ε-neighborhood of X. We see that all assumptions of Section 3 are satisfied.

The *multistage stochastic programming problem* is to minimize (7.3) in the set of policies that are nonanticipative and feasible. Its optimal value, as before, will be denoted by f^*.

Following [17, 24] let us equivalently represent the nonanticipativity condition $x \in \mathcal{M}$ in a form of a linear constraint. Let $\mathbb{E}_t$ denote the operation of conditional expectation with respect to the σ-subfield $\mathcal{F}_t$, $t = 1, \ldots, T$. Directly from the definition of the conditional expectation, the nonanticipativity condition is equivalent to the equations $x_t = \mathbb{E}_t\, x_t$ a.s., for $t = 1, \ldots, T$. This allows to formulate the multistage stochastic programming problem as a special case of (2.6):

$$\min \mathbb{E}\,\varphi(x(\omega), \omega), \tag{7.4}$$
$$x_t - \mathbb{E}_t x_t = 0, \quad t = 1, \ldots, T, \tag{7.5}$$
$$x(\omega) \in G(\omega), \quad \text{a.s.} \tag{7.6}$$

Let us construct the aggregate constraint of form (3.3). Consider the linear operator $A : \mathcal{X} \to \mathcal{X}$ given by

$$Ax = (A_1 x_1, \ldots, A_T x_T),$$

where the linear operators $A_t : \mathcal{L}^2(\Omega, \mathbb{R}^{n_t}) \to \mathcal{L}^2(\Omega, \mathbb{R}^{n_t})$ are defined as follows:

$$A_t x_t = x_t - \mathbb{E}_t x_t, \quad t = 1, \ldots, T.$$

Since $\langle Ax, x - Ax \rangle_{\mathcal{X}} = 0$, it is clear that $I - A$ is the orthogonal projection on $\mathcal{M}$, and A—on the orthogonal complement $\mathcal{M}^{\perp}$. Therefore

$$\langle Ax^k, Au \rangle_{\mathcal{X}} = \langle Ax^k, u \rangle_{\mathcal{X}} = \sum_{t=1}^{T} \mathbb{E}\,\langle A_t x_t^k, u_t \rangle.$$

Let us define $r_t^k = A_t x_t^k = x_t^k - \mathbb{E}_t x_t^k$. The subproblem (3.2)–(3.4) takes on the form:

$$\min \mathbb{E}\,\varphi(u(\omega), \omega), \tag{7.7}$$
$$\mathbb{E} \sum_{t=1}^{T} \langle r_t^k, u_t \rangle = 0, \tag{7.8}$$
$$u(\omega) \in G(\omega), \quad \text{a.s.} \tag{7.9}$$

The next solution is defined by (3.1) and (3.5). Setting $\rho^k = Au^k$, that is,

$$\rho_t^k = u_t^k - \mathbb{E}_t u_t^k, \quad t = 1, \ldots, T,$$

we see from (7.8) that $\rho^k \perp r^k$. Therefore the expression that is minimized in (3.5) can be transformed into

$$\|(1-\tau)Ax^k + \tau Au^k\|_{\mathcal{X}}^2 = \|(1-\tau)r^k + \tau\rho^k\|_{\mathcal{X}}^2 = (1-\tau)^2\|r^k\|_{\mathcal{X}}^2 + \tau^2\|\rho^k\|_{\mathcal{X}}^2,$$

which yields the minimizing stepsize

$$\begin{aligned} \tau_k &= \|r^k\|_{\mathcal{X}}^2 \Big(\|r^k\|_{\mathcal{X}}^2 + \|\rho^k\|_{\mathcal{X}}^2\Big)^{-1} \\ &= \Big(\sum_{t=1}^T \mathbb{E}\,\|r_t^k(\omega)\|^2\Big)\Big(\sum_{t=1}^T \mathbb{E}\,\|r_t^k(\omega)\|^2 + \sum_{t=1}^T \mathbb{E}\,\|\rho_t^k(\omega)\|^2\Big)^{-1}. \end{aligned} \tag{7.10}$$

The next theorem is a direct consequence of Theorem 3.1.

Theorem 7.1. *Let $x^0 \in X$, $f(x^0) \le f^*$, and let the sequence $\{x^k\}$ be defined by (3.1), where u^k is a solution (7.7)–(7.9) and τ_k is given by (7.10). Then $\{x^k\}$ is weakly convergent to the set of solutions of the problem (7.4)–(7.6).*

The main advantage of the constraint aggregation is the possibility to reduce a large (potentially infinite) number of constraints (7.5) to a scalar constraint (7.8), which may prove advantageous computationally. But there is also a theoretical advantage: the possibility to substantially weaken constraint qualification conditions.

Let us at first mention that the classical constraint qualification for problems of form (7.4)–(7.6) requires that the set $\{Ax : x \in X\}$ contains a neighborhood of zero in $\mathcal{X}$ (see, e.g., [7, Thm. 5, §1.1]). This is impossible if X is defined by the scenario constraints (7.2) with $G(\omega) \ne \mathbb{R}^n$ and $\mathcal{X}$ is given by (7.1). Using the spaces $\mathcal{L}^\infty(\Omega, \mathbb{R}^{n_t})$ is a remedy, but it leads to difficulties resulting from the structure of the dual space $(\mathcal{L}^\infty)^*$ (see [18, 19]).

Aggregation allows to avoid all these complications, and much weaker conditions are sufficient for the application of Kuhn-Tucker conditions to the auxiliary problems (7.7)–(7.9).

Lemma 7.1. *Assume that there exisits a nonanticipative policy $\widetilde{x}$ such that $\widetilde{x}(\omega) \in \operatorname{int} G(\omega)$ with probability 1. Then for each $k \ge 0$ there exists $\lambda^k \in \mathbb{R}$, such that*

(i) $u^k(\omega) \in \operatorname{Arg}\min_{v \in G(\omega)} \Big[\varphi(v,\omega) + \lambda^k\langle r^k(\omega), v\rangle\Big]$ *a.s.*;

(ii) λ^k *maximizes in* $\mathbb{R}$ *the dual functional*

$$g(\lambda) = \mathbb{E} \min_{v \in G(\omega)} \Big[\varphi(v,\omega) + \lambda\langle r^k(\omega), v\rangle\Big].$$

Proof. We starrt with the observation that there exists a strictly positive random variable $\varepsilon : \Omega \to \mathbb{R}_+$ such that $\widetilde{x}(\omega) + \varepsilon(\omega)B_n \subset G(\omega)$ with probability one, where B_n is the unit ball in $\mathbb{R}^n$. Moreover, since $\widetilde{x} \in \mathcal{M}$ and $r^k \in \mathcal{M}^\perp$, we have

$$\mathbb{E}\,\langle r^k(\omega), \widetilde{x}(\omega)\rangle = 0.$$

Since $y(\omega) = \widetilde{x}(\omega) \pm \varepsilon(\omega) r^k(\omega)/\|r^k(\omega)\| \in G(\omega)$ a.s., and

$$\langle r^k, y\rangle_{\mathcal{X}} = \mathbb{E}\,\langle r^k(\omega), y(\omega)\rangle = \pm\mathbb{E}\Big\{\varepsilon(\omega)\|r^k(\omega)\|\Big\},$$

the origin is in the interior of the (convex) set $\{\langle r^k, u\rangle_{\mathcal{X}} : u(\omega) \in G(\omega),\ \text{a.s.}\}$, that is, the constraint qualification for (7.7)–(7.9) holds. Thus, the necessary and sufficient conditions in the Kuhn-Tucker form (see, e.g., [7, Thm. 5, §1.1]) are satisfied for this problem: there exists λ^k such that

$$u^k \in \operatorname{Arg}\min_{u\in X} \Big[f(u) + \lambda^k \langle r^k, u\rangle_{\mathcal{X}} \Big],$$

and λ^k maximizes on $\mathbb{R}$ the dual functional

$$g(\lambda) = \min_{u\in X} \Big[f(u) + \lambda \langle r^k, u\rangle_{\mathcal{X}} \Big].$$

All minimizers u in the last problem are measurable selections of the multifunction $\omega \mapsto \operatorname{Arg}\min_{v\in G(\omega)} \Big[\varphi(v,\omega) + \lambda\langle r^k(\omega), v\rangle \Big]$ (see, e.g., [7, §8.3, Prop. 2]). □

Remark 7.1. It is a matter of obvious modifications to develop an approach that employs aggregation of nonanticipativity constraints in groups, for example corresponding to time stages $t = 1, \dots, T$. This would mean aggregating the constraints (7.5) to T equations

$$\mathbb{E}\,\langle r_t^k, u_t\rangle = 0, \quad t = 1, \dots, T.$$

The analysis is similar.

8 Conclusions

The idea of constraint aggregation allows to replace a convex optimization problem with infinitely-dimensional constraints by a sequence of simpler problems having finite-dimensional constraints. Solutions to these problems are used to construct a sequence which is weakly* convergent to a solution of the original problem. The idea is fairly general and applies to many classes of problems. We have developed some new and simple approaches to such remote application areas like optimal control, games and stochastic programming.

In the application areas of interest, weak* convergence turns out to be either natural (when we speak of convergence of probability measures in the section on games) or equivalent to weak convergence (in the sections on optimal control and stochastic programming). For control problems strong convergence of trajectories follows, and in all cases strong convexity guarantees strong convergece to the solution set.

One of the features of the aggregation is the possibility to enjoy some advantages of the duality theory in convex programming—in particular decomposition—although the existence of Lagrange multipliers in the original problem is not guaranteed. In some cases the auxiliary problems can be solved in a closed form, which substantially simplifies the entire algorithm.

References

[1] J.-P. Aubin (1979), *Mathematical Methods of Game and Economic Theory*, North-Holland, Amsterdam.

[2] J.-P. Aubin, H. Frankowska, and C.Olech (1986), *Controllability of convex processes*, SIAM J. Control Optimiz., **24**, No 6, 192–211.

[3] P. Billingsley (1968), *Convergence of Probability Measures*, John Wiley & Sons, New York.

[4] A.Ya. Dubovitskii and A.A. Milljutin (1965), *Necessary conditions for a weak extremum in optimal control problems with mixed constraints of the inequality type*, U.S.S.R. Comput. Math. and Math. Phys. 5 (1965) 1–80.

[5] Yu. M. Ermoliev, A.V. Kryazhimskii, and A. Ruszczyński (1997), *Constraint aggregation principle in convex optimization*, Mathematical Programming 76, pp. 353–372.

[6] S.D. Flåm (1995), *Corrigendum: Lagrange multipliers in stochastic programming*, SIAM J. Control and Optimization 33, pp. 667–671.

[7] A.D. Ioffe and V.D. Tihomirov (1979), *Theory of Extremal Problems*, North–Holland, Amsterdam.

[8] L.V. Kantorovich and G.P. Akilov (1982), *Functional Analysis*, Pergamon Press, Oxford.

[9] N.N. Krasovskii, and A.I.Subbotin (1988), *Game-Theoretical Control Problems*, Springer, Berlin.

[10] A.V. Kryazhimskii (1994) *Convex optimization via feedbacks*, International Institute for Applied System Analysis, Working Paper WP-94-109.

[11] A.V. Kryazhimskii, and Yu.S.Osipov (1987), *To a regularization of a convex extremal problem with inaccurately given constraints. An application to an optimal control problem with state constraints*, in: Some Methods of Positional and Program Contral, Urals Sci. Center, Sverdlovsk, 34 - 54 (in Russian).

[12] M. Eisner and P. Olsen (1974), *Duality in stochastic programming interpreted as L.P. in L_p space*, SIAM J. Appl. Math. 28, pp. 779–792.

[13] J. von Neumann and O. Morgenstern (1944), *Theory of Games and Economic Behaviour*, Princeton.

[14] L.S. Pontryagin, V.G. Boltianskii, R.V Gamkrelidze and E.F. Mishchenko (1962), *The Mathematical Theory of Optimal Processes*, Interscience, New York.

[15] J. Robinson (1951) *An iterative method for solving a game*, Annals of Mathematics, v. 54, pp. 296–301.

[16] R.T. Rockafellar (1970), *Convex Analysis*, University Press, Princeton.

[17] R.T. Rockafellar and R. J.-B. Wets (1975), *Stochastic convex programming: Kuhn-Tucker conditions*, Journal of Mathematical Economics, 2 , pp. 349–370.

[18] R.T. Rockafellar and R.J.-B. Wets (1977), *Measures as Lagrange multipliers in multistage stochastic programming*, J. Math. Anal. Appl. 60 (1977) 301–313.

[19] R.T. Rockafellar and R.J.-B. Wets (1978), *The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints,* SIAM J. Contr. Optim. 16, pp. 16–36.

[20] R. SCHNEIDER (1996), *Convex Bodies: The Brunn–Minkowski Theory*, Cambridge University Press.

[21] G.C. SHEPHARD (1968), *The uniqueness theorem for the Steiner point of a convex region*, J. London Math. So. 43, pp. 439–444.

[22] F.P.VASILIEV (1981), *Solution Methods to Extremal Problems*, Nauka, Moscow (in Russian).

[23] J. WARGA (1975), *Optimal Control of Differential and Functional Equations*, Academic Press, New York.

[24] R. J.-B. WETS (1975), *On the relation between stochastic and deterministic optimization*, in: A. Bensoussan and J.L. Lions (Eds.), *Control Theory, Numerical Methods and Computer Systems Modelling*, Springer Verlag Lecture Notes in Economics and Mathematical Systems, vol. 107, Berlin. pp. 350-361