International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Constraint Aggregation Principle in Convex Optimization

Ermoliev, Y.M., Kryazhimskiy, A.V. and Ruszczynski, A.

IIASA Working Paper

WP-95-015

February 1995

Working Paper

Constraint Aggregation Principle in Convex Optimization

Yuri M. Ermoliev
Arkadii Kryazhimskii
Andrzej Ruszczyński

WP-95-015
February 1995

IIASA
International Institute for Applied Systems Analysis ◻ A-2361 Laxenburg ◻ Austria
Telephone: +43 2236 807 ◻ Fax: +43 2236 71313 ◻ E-Mail: info@iiasa.ac.at

Constraint Aggregation Principle in Convex Optimization

Yuri M. Ermoliev
Arkadii Kryazhimskii
Andrzej Ruszczyński

WP-95-015
February 1995

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Abstract

A general constraint aggregation technique is proposed for convex optimization problems. At each iteration a set of convex inequalities and linear equations is replaced by a single inequality formed as a linear combination of the original constraints. After solving the simplified subproblem, new aggregation coefficients are calculated and the iteration continues. This general aggregation principle is incorporated into a number of specific algorithms. Convergence of the new methods is proved and speed of convergence analyzed. It is shown that in case of linear programming, the method with aggregation has a polynomial complexity. Finally, application to decomposable problems is discussed.

Key words: Nonsmooth optimization, constraint aggregation, subgradient methods, polynomial algorithms, decomposition.

Constraint Aggregation Principle in Convex Optimization

Yuri M. Ermoliev
Arkadii Kryazhimskii
Andrzej Ruszczyński

1 Introduction

In this article we consider the problem of minimizing a non-differentiable function subject to a large number of linear and nonlinear constraints. The study of such problems is motivated by new challenges arising in the analysis of large scale systems with the lack of smooth behavioral properties. Our objective is to develop a general constraint aggregation principle which can be used in a variety of methods for solving convex optimization problems of the form

$$\min f(x) \tag{1.1}$$

$$h_j(x) \leq 0, \quad j = 1, \ldots, m, \tag{1.2}$$

$$x \in X. \tag{1.3}$$

We assume throughout this paper that the functions $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $h_j : \mathbb{R}^n \mapsto \mathbb{R}$, $j = 1, \ldots, m$, are convex and $X \subset \mathbb{R}^n$ is convex and compact. We also assume that the feasible set defined by (1.2)-(1.3) is non-empty which guarantees that the problem has an optimal solution.

We are especially interested in the case when the dimensions m and n are very large, so that the direct application of such procedures as *bundle methods* (see, e.g., [4, 3]) may become difficult. An alternative approach is provided by non-monotone *subgradient methods* dating back to [1, 9, 13], which found numerous applications and extensions to various classes of problems, such as discrete optimization, design of robust decomposition procedures, general problems of stochastic optimization, min-max and more general game-theoretical problems (see [2, 7, 8, 14]).

However, treating general convex constraints within subgradient methods may also be difficult. This requires either projection (impractical for large m and n) or additional penalty/multiplier iterations, which for general f and h_j have (so far) only theoretical importance in this framework.

We are going to follow a different path. We shall assume that the structure of X is simple and the main difficulty comes from the large number of constraints (1.2). To overcome it, we shall replace the original problem by a sequence of problems in which the complicating constraints (1.2) are represented by one *surrogate inequality*

$$\sum_{j=1}^{m} s_j^k h_j(x) \leq 0, \tag{1.4}$$

where $s^k \geq 0$ are iteratively modified *aggregation coefficients.* In this way a substantial simplification of (1.1)-(1.3) is achieved, because (1.4) inherits linearity or differentiability properties of (1.2). We shall show how to update the aggregation coefficients and how to use the solution of the simplified problems to arrive at the solution of (1.1)-(1.3). It is worth noting here that our approach is fundamentally different from the aggregation provided by Lagrange multipliers and works also for problems for which duality does not hold. It is also different from the mostly heuristic aggregation methods discussed in [11].

In section 2 we shall consider the simplified version of the problem having only linear constraints and we shall develop a continuous time feedback rule using constraint aggregation. In section 3 a reformulation of this procedure is presented, which allows for various extensions and generalizations. We shall prove a simple lemma which is an analog of the Lyapunov function approach in the study of convergence of non-monotone dynamic (iterative) processes. Using this result, the convergence of the aggregation procedure in the simplified case is proved, under the assumption that the problem of minimizing f subject to the surrogate constraint can be solved easily. In section 4 this assumption is relaxed: it is shown that it is sufficient to make a step in the direction of the subgradient and project the result on the non-stationary aggregate constraint. Section 5 generalizes the aggregation principle to convex non-linear inequalities. Two aggregation schemes are analyzed: direct aggregation (1.4) and aggregation combined with linearization, which replaces (1.2) with one linear inequality. In section 6, we further follow this idea and develop a class of linearization methods with aggregation. Section 7 is devoted to the analysis of the speed of convergence. We derive estimates of the rate of convergence and we prove that our method has a polynomial complexity for linear programming problems. Finally, in section 8 we interpret the method from the dual viewpoint and we sketch out a possible application to decomposition techniques.

2 The continuous time procedure

Let us consider the simplified version of the problem:

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

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

$$x \in X. \tag{2.3}$$

Following [6], we consider the dynamic system

$$\dot{x}(t) = u(t), \quad x(0) = 0 \tag{2.4}$$

operating on the time interval $\mathbb{R}_+ = [0, \infty)$. Let us assume that

$$u(t) \in X \tag{2.5}$$

for all $t \geq 0$. The control function $u(\cdot)$ will be constructed in such a way that for the corresponding trajectory $x(\cdot)$ of the system (2.4) the ratio $x(t)/t$ will approach the optimal solution set X^* of (2.1)-(2.3) as $t \to \infty$.

In general, for a feedback control

$$u(t) = U(t, x(t))$$

the closed system (1.2) may have no trajectories understood as solutions of the above differential equation (if, for example, U is discontinuous). We shall, therefore, redefine the notion of trajectory using the concept of *δ-trajectories* coming from the theory of differential games [5].

For any $\rho > 0$ the ρ-trajectory $x_\rho(\cdot)$ under the feedback rule U is defined as follows:

$$x_\rho(t) = x_\rho(t_k) + U(t_k, x_\rho(t_k))(t - t_k), \tag{2.6}$$

where $x_\rho(0) = 0$, $t_k = k\rho$, $t \in [t_k, t_{k+1}]$, $k = 0, 1, 2, \ldots$. A function $x(\cdot) : \mathbb{R}_+ \mapsto \mathbb{R}^n$ is called a trajectory under U if for every bounded subinterval of $\mathbb{R}_+$, $x(\cdot)$ is the uniform limit of a sequence of $x_\rho(\cdot)$ with $\rho \to 0$.

Let us note that (2.6) yields

$$\frac{x_\rho(t)}{t} = \frac{1}{t} \int_0^t \dot{x}_\rho(\tau) d\tau.$$

Hence, by (2.5) and the convexity and closedness of X,

$$\frac{x_\rho(t)}{t} \in X.$$

A feedback ensuring that the above ratio approaches X^* is defined as follows:

$$U(t, x) \in \arg\min \{f(u) : \ u \in X, \ \langle Ax - tb, Au - b \rangle \leq 0\}. \tag{2.7}$$

Note that $U(t, x) \neq \emptyset$, since the feasible set (2.2)-(2.3) is non-empty. Let the constant K_A be such that $|Ax - b|^2 \leq K_A$ for all $x \in X$.

Theorem 2.1. *For any $x_\rho(t)$ under the feedback (2.7) the following inequalities hold for all $t \geq 0$:*

$$|A(x_\rho(t)/t) - b|^2 \leq K_A^2 \rho / t,$$
$$f(x_\rho(t)/t) \leq f^*,$$

where f^* denotes the optimal value of (2.1)-(2.3).

Proof. Since $\langle Ax_\rho(t_k) - t_k b, Au(t_k) - b \rangle \leq 0$, then for $t \in [t_k, t_{k+1}]$

$$\begin{aligned} &|Ax_\rho(t) - tb|^2 \\ &= |Ax_\rho(t_k) - t_k b|^2 + 2\langle Ax_\rho(t_k) - t_k b, Au(t_k) - b \rangle (t - t_k) + |Au(t_k) - b|^2 (t - t_k)^2 \\ &\leq |Ax_\rho(t_k) - t_k b|^2 + K_A (t - t_k)^2. \end{aligned} \tag{2.8}$$

The first assertion follows then by induction.

Next, for $t \in [t_k, t_{k+1}]$,

$$\int_0^t f(\dot{x}_\rho(\tau))d\tau = \int_0^{t_k} f(\dot{x}_\rho(\tau))d\tau + f^*(t - t_k),$$

because $f(u(t_k)) \leq f^*$. By induction,

$$\int_0^t f(\dot{x}_\rho(\tau))d\tau \leq tf^*.$$

Dividing by t and using the convexity of f we get

$$f\left(\frac{x_\rho(t)}{t}\right) \leq \frac{1}{t}\int_0^t f(\dot{x}_\rho(\tau))d\tau \leq f^*,$$

which proves the second assertion. □

3 The basic iterative procedure

Let us take the following denotations in the continuous time procedure: $x^k = x_\rho(t_k)/t_k$, $t_k = k\rho$, $\tau_k = 1/(k+1)$, $k = 0, 1, 2, \ldots$ (for $k = 0$ we set $x^0 = 0$). Then (2.6)-(2.7) can be rewritten as

$$x^{k+1} = x^k + \tau_k(u^k - x^k), \quad k = 0, 1, 2, \ldots, \tag{3.1}$$

where u^k is the solution of the subproblem

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

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

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

This reformulation is the starting point for our developments. Procedure (3.1)-(3.4) can be viewed as an iterative constraint aggregation method. The initial equality constraints (2.2) are replaced by a sequence of non-stationary scalar inequalities (3.3). Obviously, (3.3) is a relaxation of (2.2), so u^k exists and $f(u^k) \leq f^*$. It is rather clear that the sequence $\{x^k\}$ can converge to the optimal solution set X^* only if $|Ax^k - b|^2 \to 0$ as $k \to \infty$. This can be analyzed by the following key inequality:

$$\begin{aligned} |Ax^{k+1} - b|^2 &= |(1-\tau_k)(Ax^k - b) + \tau_k(Au^k - b)|^2 \\ &= (1-\tau_k)^2|Ax^k - b|^2 + 2\tau_k(1-\tau_k)\langle Ax^k - b, Au^k - b\rangle + \tau_k^2|Au^k - b|^2 \\ &\leq (1 - 2\tau_k)|Ax^k - b|^2 + 2K_A\tau_k^2, \end{aligned} \tag{3.5}$$

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

The convergence of various iterative algorithms analyzed in this paper can be proved by using the following simple lemma (similar results along these lines can be found in [1, 8]).

Lemma 3.1. *Let the sequences* $\{\beta_k\}$, $\{\tau_k\}$, $\{\delta_k\}$ *and* $\{\gamma_k\}$ *satisfy the inequality*

$$0 \le \beta_{k+1} \le \beta_k - \tau_k \delta_k + \gamma_k. \tag{3.6}$$

If

(i) $\liminf \delta_k \ge 0$;

(ii) *for every subsequence* $\{k_i\} \subset \mathbb{N}$ *one has* $[\liminf \beta_{k_i} > 0] \Rightarrow [\liminf \delta_{k_i} > 0]$;

(iii) $\tau_k \ge 0$, $\lim \tau_k = 0$, $\sum_{k=0}^{\infty} \tau_k = \infty$;

(iv) $\lim \gamma_k / \tau_k = 0$,

then $\lim_{k\to\infty} \beta_k = 0$.

Proof. Suppose that $\liminf \delta_k = \delta > 0$. Then (3.6) for large k yields $\beta_{k+1} \le \beta_k - \tau_k\delta/2 + \gamma_k \le \beta_k - \tau_k\delta/4$. This contradicts (iii). Therefore there is a subsequence $\{k_i\}$ such that $\beta_{k_i} \to 0$. Suppose that there is another subsequence $\{s_j\}$ such that $\beta_{s_j} \ge \beta > 0$ for $j = 0,1,2,\ldots$. With no loss of generality we may assume that $k_1 < s_1 < k_2 < s_2 \ldots$. By (i), (iii) and (iv), for all sufficiently large j there must exist indices $r_j \in [k_j, s_j]$ such that $\beta_{r_j} > \beta/2$ and $\beta_{r_{j+1}} > \beta_{r_j}$. But then, by (ii), $\liminf \delta_{r_j} = \delta > 0$ and we obtain a contradiction with (3.6) for large j. □

We are now ready to prove the convergence of our method.

Theorem 3.2. *Consider the algorithm (3.1)-(3.4) and assume that* $\tau_k \in [0,1]$, $\lim_{k\to\infty} \tau_k = 0$, $\sum_{k=0}^{\infty} \tau_k = \infty$ *and* $x^0 \in X$. *Then every accumulation point of the sequence* $\{x^k\}$ *is a solution of (2.1)-(2.3).*

Proof. By (3.5) and Lemma 3.1, taking $\beta_k = |Ax^k - b|^2$, $\delta_k = 2\beta_k$ and $\gamma_k = 2K_A\tau_k^2$, we obtain $|Ax^k - b| \to 0$ as $k \to \infty$. Since X is compact, we must have

$$\liminf f(x^k) \ge f^*.$$

Next, by the convexity of f,

$$f(x^{k+1}) \le (1-\tau_k) f(x^k) + \tau_k f(u^k) \le (1-\tau_k) f(x^k) + \tau_k f^*.$$

Thus,

$$\max(f(x^{k+1}) - f^*, 0) \le (1-\tau_k)\max(f(x^k) - f^*, 0). \tag{3.7}$$

Using Lemma 3.1 again, with $\beta_k = \delta_k = \max(0, f(x^k) - f^*)$, we obtain $f(x^k) \to f^*$. Our assertion follows then from the compactness of X. □

Clearly, in view of the compactness of X we also have $\mathrm{dist}\big(x^k, X^*\big) \to 0$, where X^* is the set of optimal solutions of (2.1)-(2.3). This general remark applies to many other results in this paper.

The stepsizes in the basic procedure can also be generated in a more systematic way:

$$\tau_k = \arg\min_{0\le\tau\le1} |(1-\tau)(Ax^k - b) + \tau(Au^k - b)|^2. \tag{3.8}$$

Theorem 3.3. *Consider the algorithm (3.1)-(3.4) and assume that $f(x^0) \leq f^*$ and the stepsizes τ_k are defined by (3.8). Then every accumulation point of the sequence $\{x^k\}$ is a solution of (2.1)-(2.3).*

Proof. From the key inequality (3.5) we obtain

$$|Ax^{k+1} - b|^2 \leq \min_{0\leq\tau\leq 1} \left((1-2\tau)|Ax^k - b|^2 + 2K_A\tau^2\right)$$
$$\leq \left(1 - \frac{|Ax^k - b|^2}{2K_A}\right)|Ax^k - b|^2.$$

Therefore $|Ax^k - b|^2 \to 0$. Since $f(x^k) \leq f^*$ for all k, the required result follows. □

4 The subgradient projection method

The basic procedure requires the solution of the auxiliary subproblems (3.2)-(3.4), which may still prove difficult. We shall show that it is sufficient to use individual subgradients of f within the constraint aggregation scheme to achieve convergence.

Consider the same procedure as before:

$$x^{k+1} = (1-\tau_k)x^k + \tau_k u^k, \tag{4.1}$$

but with u^k generated in a simpler fashion:

$$u^k = \Pi_k(x^k - \lambda_k g^k). \tag{4.2}$$

Here $g^k \in \partial f(x^k)$ and $\Pi_k(\cdot)$ denotes the orthogonal projection onto the set

$$X_k = \{x \in X : \ \langle Ax^k - b, Ax - b\rangle \leq 0\}.$$

Theorem 4.1. *Assume $\tau_k \in [0,1]$, $\lambda_k \geq 0$, $\tau_k \to 0$, $\lambda_k \to 0$, $\sum_{k=0}^{\infty}\tau_k = \infty$, $\sum_{k=0}^{\infty}\tau_k\lambda_k = \infty$. Then every accumulation point of the sequence generated by the method (4.1)-(4.2) is a solution of (2.1)-(2.3).*

Proof. For every optimal solution x^* we have

$$|x^{k+1} - x^*|^2 \leq \left((1-\tau_k)|x^k - x^*| + \tau_k|u^k - x^*|\right)^2$$
$$= (1-\tau_k)^2|x^k - x^*|^2 + 2\tau_k(1-\tau_k)|x^k - x^*||u^k - x^*| + \tau_k^2|u^k - x^*|^2. \tag{4.3}$$

Since $x^* \in X_k$, from the non-expansiveness of $\Pi_k(\cdot)$ one obtains

$$|u^k - x^*|^2 = \left|\Pi_k(x^k - \lambda_k g^k) - \Pi_k(x^*)\right|^2 \leq |x^k - x^*|^2 - 2\lambda_k\langle g^k, x^k - x^*\rangle + C\lambda_k^2,$$

where C is an upper bound on $|g^k|^2$. Thus

$$|u^k - x^*|^2 \leq |x^k - x^*|^2 - 2\lambda_k\delta_k + C\lambda_k^2 \tag{4.4}$$

with $\delta_k = f(x^k) - f^*$. Taking the square root of both sides and using the concavity of $\sqrt{\cdot}$ we get

$$|u^k - x^*| \le |x^k - x^*| \left(1 - \frac{2\lambda_k\delta_k - C\lambda_k^2}{|x^k - x^*|^2}\right)^{\frac{1}{2}} \le |x^k - x^*| - \frac{2\lambda_k\delta_k - C\lambda_k^2}{2|x^k - x^*|}.$$

Multiplication by $|x^k - x^*|$ yields

$$|u^k - x^*||x^k - x^*| \le |x^k - x^*|^2 - \lambda_k\delta_k + C\lambda_k^2/2. \tag{4.5}$$

Combining (4.3), (4.4) and (4.5), after elementary simplifications, for every $x^* \in X^*$ one has

$$|x^{k+1} - x^*|^2 \le |x^k - x^*|^2 - 2\tau_k\lambda_k\delta_k + C\tau_k\lambda_k^2.$$

Then also (after minimizing both sides with respect to x^*)

$$\text{dist}(x^{k+1}, X^*)^2 \le \text{dist}(x^k, X^*)^2 - 2\tau_k\lambda_k\delta_k + C\tau_k\lambda_k^2. \tag{4.6}$$

Similarly to Theorem 3.2, $|Ax^k - b| \to 0$. Therefore the sequence $\{\delta_k\}$ satisfies conditions (i) and (ii) of Lemma 3.1 with $\beta_k = \text{dist}(x^k, X^*)^2$. By virtue of Lemma 3.1 (with $\tau_k\lambda_k$ playing the role of τ_k), $\text{dist}(x^k, X^*) \to 0$ as $k \to \infty$. □

5 Convex inequalities

To show how the aggregation principle extends to nonlinear constraints, let us consider the problem

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

$$h_j(x) \le 0, \quad j = 1, \ldots, m, \tag{5.2}$$

$$x \in X. \tag{5.3}$$

We assume that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is convex, $X \subset \mathbb{R}^n$ is compact and convex, and the functions $h_j : \mathbb{R}^n \mapsto \mathbb{R}$, $j = 1, \ldots, m$, are convex. Denote $h_j^+(x) = \max(0, h_j(x))$, $h(x) = (h_1(x), \ldots, h_m(x))$ and $h^+(x) = (h_1^+(x), \ldots, h_m^+(x))$.

The basic iterative procedure (3.1)-(3.4) can be adapted to our problem as follows:

$$x^{k+1} = x^k + \tau_k(u^k - x^k), \quad k = 0, 1, 2, \ldots, \tag{5.4}$$

where u^k is the solution of the subproblem

$$\min f(u) \tag{5.5}$$

$$\langle h^+(x^k), h(u) \rangle \le 0, \tag{5.6}$$

$$u \in X. \tag{5.7}$$

Note that if linear constraints $Ax = b$ are written as inequalities $Ax - b \le 0$ and $-Ax + b \le 0$, then the surrogate inequalities (5.6) and (3.3) are identical.

Theorem 5.1. *Consider the algorithm (5.4)-(5.7) and assume that* $\tau_k \in [0,1]$, $\lim_{k\to\infty} \tau_k = 0$, $\sum_{k=0}^{\infty} \tau_k = \infty$ *and* $x^0 \in X$. *Then every accumulation point of the sequence* $\{x^k\}$ *is a solution of (5.1)-(5.3).*

Proof. Let us derive a counterpart of the key inequality (3.5). By the convexity of $h(\cdot)$,

$$h((1-\tau_k)x^k + \tau_k u^k) \le (1-\tau_k)h(x^k) + \tau_k h(u^k) \le (1-\tau_k)h^+(x^k) + \tau_k h(u^k).$$

In the above vector inequality, for the indices j for which the left hand side is positive, the right hand side has a larger absolute value. Therefore

$$\begin{aligned}|h^+((1-\tau_k)x^k + \tau_k u^k)|^2 &\le |(1-\tau_k)h^+(x^k) + \tau_k h(u^k)|^2 \\ &\le (1-\tau_k)^2|h^+(x^k)|^2 + \tau_k^2|h(u^k)|^2,\end{aligned}$$

where in the last inequality we used (5.6). Consequently,

$$|h^+(x^{k+1})|^2 \le (1-2\tau_k)|h^+(x^k)|^2 + 2K_h\tau_k^2, \tag{5.8}$$

with K_h being an upper bound on $|h(x)|^2$ in X. By Lemma 3.1, $\lim_{k\to\infty} h^+(x^k) = 0$. The remaining part of the proof is identical with the proof of Theorem 3.2. □

Again, similar to theorem 3.3 the stepsizes in the basic procedure (5.4) can be generated in a more systematic way:

$$\tau_k = \arg\min_{0\le\tau\le1} |h^+((1-\tau)x^k + \tau u^k)|^2. \tag{5.9}$$

Theorem 5.2. *Consider the algorithm (5.4)-(5.7) and assume that* $f(x^0) \le f^*$ *and the stepsizes* τ_k *are defined by (5.9). Then every accumulation point of the sequence* $\{x^k\}$ *is a solution of (5.1)-(5.3).*

Proof. From the inequality (5.8) we obtain

$$|h^+(x^{k+1})|^2 \le \min_{0\le\tau\le1}\left((1-2\tau)|h^+(x^k)|^2 + 2K_h\tau^2\right) \le \left(1 - \frac{|h^+(x^k)|^2}{2K_h}\right)|h^+(x^k)|^2. \tag{5.10}$$

Therefore $|h^+(x^k)|^2 \to 0$. Since $f(x^k) \le f^*$ for all k, the required result follows. □

For continuously differentiable constraint functions $h(\cdot)$ the subproblems can be further simplified by replacing (5.6) with the inequality

$$\langle h^+(x^k), h(x^k) + \nabla h(x^k)(u - x^k)\rangle \le 0, \tag{5.11}$$

where $\nabla h(\cdot)$ denotes the Jacobian of $h(\cdot)$. Let us note that by the convexity of $h(\cdot)$, (5.11) is still a relaxation of (5.2).

Theorem 5.3. *Consider the algorithm (5.4), (5.5), (5.11), (5.7) and assume that* $\tau_k \in [0,1]$, $\lim_{k\to\infty} \tau_k = 0$, $\sum_{k=0}^{\infty} \tau_k = \infty$ *and* $x^0 \in X$. *Then every accumulation point of the sequence* $\{x^k\}$ *is a solution of (5.1)-(5.3).*

Proof. We shall again derive an inequality similar to (5.8). By the uniform continuity of $\nabla h(\cdot)$ in X,

$$h(x^k + \tau_k(u^k - x^k)) \le h(x^k) + \tau_k \nabla h(x^k)(u^k - x^k) + o(\tau_k, u^k, x^k)$$
$$\le (1-\tau_k)h^+(x^k) + \tau_k(h(x^k) + \nabla h(x^k)(u^k - x^k)) + o(\tau_k, u^k, x^k),$$

where $o(\tau, u, x)/\tau \to 0$, when $\tau \to 0$, uniformly over $u, x \in X$. Proceeding exactly as in the proof of Theorem 5.1 and using (5.11), we obtain

$$|h^+(x^{k+1})|^2 \le (1 - 2\tau_k)|h^+(x^k)|^2 + o_1(\tau_k), \tag{5.12}$$

with $o_1(\tau)/\tau \to 0$, when $\tau \to 0$. Applying Lemma 3.1, we get $\lim_{k\to\infty} h^+(x^k) = 0$. The remaining part of the proof is analogous to the proof of Theorem 3.2. □

Similar results can be obtained for the subgradient projection method

$$x^{k+1} = (1-\tau_k)x^k + \tau_k u^k, \tag{5.13}$$

where

$$u^k = \Pi_k(x^k - \lambda_k g^k), \tag{5.14}$$

$g^k \in \partial f(x^k)$ and $\Pi_k(\cdot)$ denotes the orthogonal projection onto the set

$$X_k = \{x \in X : \ \langle h^+(x^k), h(x)\rangle \le 0\} \tag{5.15}$$

(if the aggregation scheme (5.6) is used), or onto the set

$$X_k = \{x \in X : \ \langle h^+(x^k), h(x^k) + \nabla h(x^k)(x - x^k)\rangle \le 0\} \tag{5.16}$$

(if linearization of smooth constraints is employed). In the identical way as Theorems 5.1 and 5.3 we obtain the following results.

Theorem 5.4. *Assume* $\tau_k \in [0,1]$, $\lambda_k \ge 0$, $\tau_k \to 0$, $\lambda_k \to 0$, $\sum_{k=0}^{\infty} \tau_k = \infty$, $\sum_{k=0}^{\infty} \tau_k\lambda_k = \infty$. *Then every accumulation point of the sequence generated by the method (5.13)-(5.15) is a solution of (5.1)-(5.3). If additionally, the constraint functions* h_j, $j = 1, \dots, m$ *are continuously differentiable, every accumulation point of the sequence generated by the method (5.13), (5.14), (5.16) is a solution of (5.1)-(5.3).*

Let us suppose now that the optimal value f^* in (5.1)-(5.3) is known. Then we can reformulate the problem as a system of inequalities:

$$f(x) - f^* \le 0 \tag{5.17}$$

$$h_j(x) \le 0, \ j = 1, \dots, m, \tag{5.18}$$

$$x \in X. \tag{5.19}$$

We can, formally, think of it as a problem of minimizing a function identically equal to 0 subject to (5.17)-(5.19). Our aggregation procedure (5.4)-(5.7) replaces our problem by a sequence of simple systems having the form:

$$\max(0, f(x^k) - f^*)f(u) + \sum_{j\in J_k} h_j(x^k)h_j(u) \le 0,$$

$$u \in X,$$

where $J_k = \{j : \ h_j(x^k) > 0\}$. Linearization of smooth constraints is also possible here.

6 The linearization procedure

Let us consider the problem (5.1)-(5.3) again, but with the additional assumption that the functions f and h_j, $j = 1, \dots, m$, are convex and continuously differentiable. The set X is assumed to be convex and compact. We shall extend the idea of linearization used in (5.11) to the objective function, constructing the following procedure:

$$x^{k+1} = x^k + \tau_k(u^k - x^k), \quad k = 0, 1, 2, \dots, \tag{6.1}$$

where u^k is the solution of the subproblem

$$\min \langle \nabla f(x^k), u - x^k \rangle \tag{6.2}$$

$$\langle h^+(x^k), h(x^k) + \nabla h(x^k)(u - x^k) \rangle \leq 0, \tag{6.3}$$

$$u \in X. \tag{6.4}$$

The above algorithm is a modification of the classical Frank-Wolfe method of nonlinear programming, but with constraint linearization, aggregation and with non-monotone stepsize selection.

Theorem 6.1. *Assume that $\tau_k \in [0, 1]$, $\tau_k \to 0$, $\sum_{k=0}^{\infty} \tau_k = \infty$. Then every accumulation point of the sequence generated by the method (6.1)-(6.4) is a solution of (5.1)-(5.3).*

Proof. Proceeding exactly as in the proof of Theorem 5.3, we obtain $h^+(x^k) \to 0$, as $k \to \infty$. Next, by the differentiability of f,

$$f(x^{k+1}) \leq f(x^k) + \tau_k \langle \nabla f(x^k), u^k - x^k \rangle + o(\tau_k, x^k, u^k), \tag{6.5}$$

where $o(\tau, x, u)/\tau \to 0$ when $\tau \to 0$, uniformly for $x, u \in X$. Since every solution x^* is feasible for (6.2)-(6.4),

$$\langle \nabla f(x^k), u^k - x^k \rangle \leq \langle \nabla f(x^k), x^* - x^k \rangle \leq f^* - f(x^k).$$

Combining the last two inequalities we get

$$f(x^{k+1}) - f^* \leq (1 - \tau_k)(f(x^k) - f^*) + o(\tau_k, x^k, u^k).$$

By Lemma 3.1, $\max(0, f(x^k) - f^*) \to 0$, as $k \to \infty$. The proof is complete. □

Let us note that if X has a simple structure, e.g.

$$X = \left\{ x \in \mathbb{R}^n : \ x_j^{\min} \leq x_j \leq x_j^{\max}, \ j = 1, \dots, n \right\},$$

then (6.2)-(6.4) can be solved analytically.

Finally, it is worth mentioning that in a similar fashion one can derive a family of algorithms, using different direction-finding subproblems, such as trust-region methods, methods with regularization and averaging (see, e.g., [12]), etc.

7 Rate of convergence and complexity

Let us now analyze the speed of convergence of the methods with constraint aggregation in the case of harmonic stepsizes:

$$\tau_k = \frac{\alpha}{k+1}, \quad k = 0, 1, 2, \ldots, \tag{7.1}$$

where $0 < \alpha \leq 1$. We start from the following technical lemma.

Lemma 7.1. *Let the sequence $\{\tau_k\}$ be defined by (7.1).*

(i) If the sequence $\{\beta_k\}$ satisfies for $k = 0, 1, 2, \ldots$ the inequality

$$0 \leq \beta_{k+1} \leq (1 - 2\tau_k)\beta_k + C\tau_k^2 \tag{7.2}$$

and $\beta_0 \leq C/2$, then for all $k \geq 0$ and all $p \in [2\alpha(1-\alpha), 2\alpha)$ such that $p \leq 1$ one has

$$\beta_k \leq \frac{M(p)}{(k+1)^p}, \tag{7.3}$$

where $M(p) = C\alpha^2/(2\alpha - p)$.

(ii) If the sequence $\{\delta_k\}$ satisfies for $k = 0, 1, 2, \ldots$ the inequality

$$0 \leq \delta_{k+1} \leq (1 - \tau_k)\delta_k, \tag{7.4}$$

then for all $k \geq 0$

$$\delta_k \leq \frac{\delta_0}{(k+1)^\alpha}. \tag{7.5}$$

Proof. For $k = 0$ (7.3) is true, because $p \geq 2\alpha(1-\alpha)$ implies $\beta_0 \leq C/2 \leq M(p)$. Assuming that (7.3) holds for k, we shall prove it for $k+1$. From (7.2) we get

$$\begin{aligned} \beta_{k+1} &\leq \left(1 - \frac{2\alpha}{k+1}\right)\frac{M(p)}{(k+1)^p} + \frac{C\alpha^2}{(k+1)^2} \\ &= \left(1 - \frac{p}{k+1}\right)\frac{M(p)}{(k+1)^p} + \frac{C\alpha^2}{(k+1)^2} - \frac{(2\alpha - p)M(p)}{(k+1)^{p+1}} \\ &\leq \left(1 - \frac{p}{k+1}\right)\frac{M(p)}{(k+1)^p}, \end{aligned} \tag{7.6}$$

where in the last inequality we used the definition of $M(p)$. Let us now note that the concavity of the logarithm implies

$$\ln\left(1 - \frac{p}{k+1}\right) + p\ln\left(1 + \frac{1}{k+1}\right) \leq 0,$$

so

$$1 - \frac{p}{k+1} \leq \left(\frac{k+1}{k+2}\right)^p. \tag{7.7}$$

Using this inequality in (7.6) we conclude that

$$\beta_{k+1} \le \frac{M(p)}{(k+2)^p},$$

which completes the induction step for (7.3).

To prove (7.5) (which is trivial for $k = 0$), we use (7.4) and (7.7) with $p = \alpha$ to get

$$\delta_{k+1} \le \left(1 - \frac{\alpha}{k+1}\right) \frac{\delta_0}{(k+1)^\alpha} \le \frac{\delta_0}{(k+2)^\alpha}.$$

By induction (7.5) holds for all $k \ge 0$. □

We can now derive rate of convergence estimates for our basic procedure.

Theorem 7.2. *Consider the problem (5.1)-(5.3) and the method (5.4)-(5.5) with the stepsizes (7.1). Then for all $k \ge 0$ and all $p \in [2\alpha(1-\alpha), 2\alpha)$ such that $p \le 1$ the sequence $\{x^k\}$ satisfies the inequalities*

$$|h^+(x^k)|^2 \le \frac{M_h(p)}{(k+1)^p}, \tag{7.8}$$

and

$$f(x^k) - f^* \le \frac{M_f}{(k+1)^\alpha}, \tag{7.9}$$

where $M_h(p) = 2K_h\alpha^2/(2\alpha - p)$, $K_h = \max_{x \in X} |h(x)|^2$, and $M_f = \max(f(x^0) - f^*, 0)$.

Proof. The result can be obtained from inequalities (5.8) and (3.7) by applying Lemma 7.1 with $\beta_k = |h^+(x^k)|^2$ and $\delta_k = \max(0, f(x^k) - f^*)$. □

It is worth noting that the rate of convergence estimates (7.8) and (7.9) are invariant with respect to the scaling of the variables and of the constraint and objective functions.

A similar result can be obtained for the method with constraint linearization.

Theorem 7.3. *Consider problem (5.1)-(5.3) and the method (5.4), (5.5), (5.7), (5.11) with the stepsizes (7.1). Assume additionally that the Jacobian $\nabla h(\cdot)$ is Lipschitz continuous on X. Then the estimates (7.8) and (7.9) hold with $M_h(p) = (K_h + LD^2)\alpha^2/(2\alpha - p)$, where $K_h = \max_{x \in X} |h(x)|^2$, L is the Lipschitz constant of $\nabla h(\cdot)$ and D is the diameter of X.*

Proof. It is sufficient to observe that in (5.12) one has $o_1(\tau_k) = (K_h + LD^2)\tau_k^2$. Application of Lemma 7.1 with $\beta_k = |h^+(x^k)|^2$ yields the required result. □

It is also clear that the estimate (7.8) holds for the constraint violation in the subgradient projection method and in the linearization method. In the latter case, however, we can again estimate the speed of convergence of objective values.

Theorem 7.4. *Consider problem (5.1)-(5.3) and the method (6.1)-(6.4) with the stepsizes (7.1). Assume additionally that the Jacobian $\nabla h(\cdot)$ and the gradient $\nabla f(\cdot)$ are Lipschitz continuous on X. Then the estimates (7.8) hold with $M_h(p) = (K_h + L_h D^2)\alpha^2/(2\alpha - p)$, where $K_h = \max_{x\in X} |h(x)|^2$, L_h is the Lipschitz constant of $\nabla h(\cdot)$ and D is the diameter of X. Moreover, we have for every $p \in (0, \alpha)$ the estimate*

$$f(x^k) - f^* \le \frac{M_f(p)}{(k+1)^p}, \tag{7.10}$$

where $M_f(p) = \max(f(x^0) - f^*, L_f D^2/(\alpha - p))$ and L_f is the Lipschitz constant of $\nabla f(\cdot)$.

Proof. The estimate of the constraint violation can be proved as in Theorem 7.3. To prove (7.10) we observe that in the inequality (6.5) one has

$$o(\tau, u, x) \le L_f D^2 \tau^2.$$

Application of Lemma 7.1 to this inequality (with 2α replaced by α) yields the required result. □

Equally strong results can be obtained for the method with stepsize selection via directional minimization (5.9).

Theorem 7.5. *Consider the algorithm (5.4)-(5.7) and assume that $f(x^0) \le f^*$ and the stepsizes τ_k are defined by (5.9). Then the following estimate holds*

$$|h^+(x^k)|^2 \le \frac{2K_h}{k+1},\ k = 0, 1, 2, \ldots \tag{7.11}$$

Proof. For $k = 0$ our assertion is true. Assuming that it holds for k we shall prove it for $k+1$. If $|h^+(x^k)|^2 \le 2K_h/(k+2)$, from (5.10) we get $|h^+(x^{k+1})|^2 \le |h^+(x^k)|^2 \le 2K_h/(k+2)$. It remains to consider the case $|h^+(x^k)|^2 \in [2K_h/(k+2), 2K_h/(k+1)]$. But then, by (5.10), one has

$$|h^+(x^{k+1})|^2 \le \left(1 - \frac{1}{k+2}\right)\frac{2K_h}{k+1} = \frac{2K_h}{k+2},$$

as required. □

Let us now pass to the complexity analysis of the basic procedure (3.1)-(3.4) in the case of the linear programming problem

$$\min\langle c, x\rangle \tag{7.12}$$

$$Ax = b, \tag{7.13}$$

$$x^{\min} \le x \le x^{\max}. \tag{7.14}$$

If stepsizes (7.1) are used, we assume for simplicity that $\alpha = 1$, so $\langle c, x^k\rangle \le f^*$ for all $k \ge 1$. Suppose that we need to find an approximate solution such that

$$\|Ax^k - b\|^2 \le \epsilon,$$

where $\epsilon > 0$ is some selected accuracy. Then it follows from (7.8) with $p = 1$ that it is sufficient to carry out

$$k(\epsilon) = \frac{2K_A}{\epsilon} - 1$$

iterations, where K_A is the upper bound on $|Ax - b|^2$ subject to (7.14). By theorem 7.5, the same estimate is true when stepsizes are generated by minimizing the constraint violation, as in (3.8).

Assuming that a is the upper bound on the absolute value of the entries in A and d is the upper bound on $x_j^{\max} - x_j^{\min}$ we get

$$k(\epsilon) \leq Cn^2m/\epsilon,$$

with $C = 2a^2d^2$. Next, each iteration consists in solving the linear program

$$\min\langle c, u\rangle$$

$$\langle A^T s^k, u\rangle \leq \langle s^k, b\rangle$$

$$x^{\min} \leq x \leq x^{\max},$$

where $s^k = Ax^k - b$. Forming the aggregate constraint costs $m(2n+1)$ multiplications. The subproblem requires at most $2n$ multiplications and divisions to solve. Consequently, the effort to find a solution with accuracy ϵ can be bounded by a polynomial function of the problem's dimensions m and n:

$$E(m, n, \epsilon) \leq 2Cn^3m^2/\epsilon.$$

For sparse problems we get a better bound by noting that the aggregate constraint is formed in $2\nu + m$ multiplications, where ν is the number of nonzeros of A.

8 The dual viewpoint

So far we did not assume any constraint qualification conditions; the algorithms discussed in the previous sections converge without that (if a solution exists). Let us now look closer at the properties of the basic procedure (3.1)-(3.4) for the problem (2.4)-(2.6) under the following assumption.

Constraint Qualification Condition. The set X is polyhedral, or at some feasible point $\tilde{x}$ one has $\mathrm{ri}\,(X \setminus \{\tilde{x}\}) \cap \{d : \ Ad = 0\} \neq \emptyset$.

Let us consider the Lagrangian

$$L(x, \pi) = f(x) + \langle \pi, Ax - b\rangle \tag{8.1}$$

and the dual function

$$D(\pi) = \min_{x \in X} L(x, \pi).$$

Under the Constraint Qualification Condition, there exists π^* such that

$$D(\pi^*) = \max_{\pi \in \mathbb{R}^m} D(\pi) = f^*. \tag{8.2}$$

This result can be obtained as follows. At first, we note that at the optimal solution x^* one has $g \in N_C(x^*)$, where $g \in \partial f(x^*)$ and $N_C(x^*)$ is the normal cone to the set defined by (2.5)-(2.6) (see [10]). Then we use convexity and Constraint Qualification Condition to get $N_C(x^*) = N_X(x^*) + \{A^T\pi : \ \pi \in \mathbb{R}^m\}$, so $g + A^T\pi^* \in N_X(x^*)$ for some π^*. The latter is equivalent to x^* being the minimizer of $L(\cdot, \pi^*)$ in X, which in turn yields (8.2).

The Constraint Qualification Condition also guarantees (in the same way) the existence of an optimal Lagrange multiplier μ^k corresponding to the aggregate constraint (3.3) at each iteration of the method. Let us define

$$s^k = Ax^k - b \tag{8.3}$$

and

$$\pi^k = \mu^k s^k. \tag{8.4}$$

It follows that u^k minimizes the Lagrangian for (3.2)-(3.4), which turns out to be the Lagrangian (8.1) at π^k:

$$u^k \in \arg\min_{u \in X} \left[f(u) + \mu^k \langle Ax^k - b, Au - b \rangle \right] = \arg\min_{u \in X} L(u, \pi^k).$$

Therefore, our basic algorithm in the dual space can be interpreted as follows:

$$\pi^k = \arg\max \left\{ D(\pi) : \ \pi = \mu s^k, \ \mu \geq 0 \right\}, \tag{8.5}$$

$$w^k = Au^k - b \in \partial D(\pi^k), \tag{8.6}$$

$$s^{k+1} = (1 - \tau_k)s^k + \tau_k w^k. \tag{8.7}$$

We also have

$$f(u^k) = \min_{u \in X} L(u, \pi^k) = D(\pi^k). \tag{8.8}$$

We can use this result to prove an interesting property of the sequence $\{\pi^k\}$ generated by (8.5)-(8.7).

Theorem 8.1. *Let the Constraint Qualification Condition and the assumptions of Theorem 3.2 be satisfied. Then for the basic procedure (3.1)-(3.4) with* $\tau_k \in [0, 1]$, $\lim_{k\to\infty} \tau_k = 0$ *and* $\sum_{k=0}^{\infty} \tau_k = \infty$ *we have*

$$\limsup_{k\to\infty} D(\pi^k) = f^* = \max_{\pi \in \mathbb{R}^n} D(\pi)$$

and

$$\lim_{k\to\infty} s^k = 0.$$

Proof. By the convexity of f,

$$f(x^{k+1}) \leq f(x^k) - \tau_k(f(x^k) - f(u^k)).$$

Suppose that $\limsup f(u_k) = v < f^*$. By Theorem 3.2, $\lim f(x^k) = f^*$, so for large k the last inequality yields

$$f(x^{k+1}) \leq f(x^k) - \tau_k(f^* - v)/2.$$

Since $\sum \tau_k = \infty$, we obtain $f(x^k) \to -\infty$, a contradiction. Therefore we must have $v = f^*$ and the required result follows from (8.8) and(8.2). □

A stronger result can be obtained for the sequence of averages.

Theorem 8.2. *Let the Constraint Qualification Condition and the assumptions of Theorem 3.2 be satisfied. Consider the basic procedure (3.1)-(3.4) with* $\tau_k \in [0,1]$, $\lim_{k\to\infty} \tau_k = 0$ *and* $\sum_{k=0}^{\infty} \tau_k = \infty$ *and the sequence* $\{\bar{\pi}^k\}$ *defined as follows:*

$$\bar{\pi}^{k+1} = (1-\tau_k)\bar{\pi}^k + \tau_k \pi^k, \quad k = 0, 1, \ldots,$$

where the sequence $\{\pi^k\}$ *is generated by (8.5)-(8.7). Then*

$$\lim_{k\to\infty} D(\bar{\pi}^k) = f^* = \max_{\pi \in \mathbb{R}^n} D(\pi).$$

Proof. From the concavity of the dual function

$$D(\bar{\pi}^{k+1}) \geq (1-\tau_k)D(\bar{\pi}^k) + \tau_k D(\pi^k).$$

The convexity of f yields

$$f(x^{k+1}) \leq (1-\tau_k)f(x^k) + \tau_k f(u^k).$$

Subtracting the last two inequalities and noting that $D(\pi^k) = f(u^k)$ we obtain

$$f(x^{k+1}) - D(\bar{\pi}^{k+1}) \leq (1-\tau_k)\left(f(x^k) - D(\bar{\pi}^k)\right).$$

By lemma 3.1,

$$\lim_{k\to\infty} \max\left(0, f(x^k) - D(\bar{\pi}^k)\right) = 0.$$

Since $f(x^k) \to f^*$ by theorem 3.2, the required result follows. □

It is obvious that analogous results hold for the case of convex inequalities discussed in section 5, in both versions of aggregation: the nonlinear one (5.6) and the linearized one (5.11). We focused here on linear constraints because of the application to decomposable problems of the form:

$$\min \sum_{i=1}^{I} f_i(x_i) \tag{8.9}$$

$$\sum_{i=1}^{I} A_i x_i = b \tag{8.10}$$

$$x_i \in X_i, \; i = 1, \ldots, I. \tag{8.11}$$

We assume that the functions $f_i : \mathbb{R}^{n_i} \mapsto \mathbb{R}$ are convex, the sets $X_i \subset \mathbb{R}^{n_i}$ are convex and compact, A_i are matrices of dimension $m \times n_i$, $b \in \mathbb{R}^m$, $i = 1, \ldots, I$.

Our approach converts (8.9)-(8.11) into a sequence of decomposable subproblems with only one linking constraint:

$$\min \sum_{i=1}^{I} f_i(u_i) \tag{8.12}$$

$$\sum_{i=1}^{I} \langle s^k, A_i u_i \rangle \leq \langle s^k, b \rangle, \tag{8.13}$$

$$u_i \in X_i, \; i = 1, \ldots, I, \tag{8.14}$$

where $s^k = \sum_{i=1}^{I} A_i x^k - b$. These subproblems can be easily solved by duality and decomposition:

$$\max_{\mu \geq 0} d^k(\mu)$$

with

$$d^k(\mu) = -\mu \langle s^k, b \rangle + \sum_{i=1}^{I} d_i^k(\mu),$$

$$d_i^k(\mu) = \min_{u_i \in X_i} \left[f_i(u_i) + \mu \langle s^k, A_i u_i \rangle \right].$$

In this way we avoid complicated coordination procedures for updating multipliers associated with the linking constraints (8.10).

For non-separable functions f appearing as objectives in place of (8.9), we may use the projection method, whose subproblems consist in minimizing a separable objective $\sum_{i=1}^{I} |u_i - \bar{x}_i^k|^2$, where $\bar{x}^k = x^k - \lambda_k g^k$, $g_k \in \partial f(x^k)$, subject to the constraints (8.13)-(8.14). Again, these subproblems are easily decomposable.

When the objective f is smooth, the subproblems of the linearization method have decomposable objectives of the form $\sum_{i=1}^{I} \langle \nabla_{x_i} f(x^k), u_i - x_i^k \rangle$, and constraints (8.13)-(8.14). In many special cases all these subproblems can be solved analytically.

References

[1] Yu.M. Ermoliev, "Methods for solving nonlinear extremal problems", *Kibernetika* 4(1966) 1-17 (in Russian).

[2] Yu.M. Ermoliev, *Methods of Stochastic Programming*, Nauka, Moscow, 1976 (in Russian).

[3] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.

[4] K.C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Springer-Verlag, Berlin, 1985.

[5] N.N. Krasovskii and A.I. Subbotin, *Game-Theoretical Control Problems*, Springer-Verlag, Berlin, 1988.

[6] A. Kryazhimskii, "Convex optimization via feedbacks", working paper WP-94-109, IIASA, Laxenburg, 1994.

[7] V.S. Mikhalevich, A.M. Gupal and V.I. Norkin, *Methods for Non-Convex Optimization*, Nauka, Moscow, 1987 (in Russian).

[8] E.A. Nurminski, *Numerical Methods for Convex Optimization*, Nauka, Moscow, 1991 (in Russian).

[9] B.T. Polyak, "A general method for solving extremal problems", *Doklady Akad. Nauk SSSR* 174 (1967) 33-36 (in Russian).

[10] R.T. Rockafellar, "Lagrange multipliers and duality", *SIAM Review*, 35(1993) 183-238.

[11] D.F. Rogers, R.D. Plante, R.T. Wong and J.R. Evans, "Aggregation and disaggregation techniques and methodology in optimization", *Operations Research* 39 (1991) 553-582.

[12] A. Ruszczyński, "A linearization method for nonsmooth stochastic programming problems", *Mathematics of Operations Research* 12 (1987) 32-49.

[13] N.Z. Shor, "On the structure of algorithms for numerical solution of optimal planning and design problems", PhD Dissertation, Kiev, 1964 (in Russian).

[14] N.Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, 1985.