ISSN 0965-5425, Computational Mathematics and Mathematical Physics, 2011, Vol. 51, No. 5, pp. 741–750. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © E.A. Nurminski, 2011, published in Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 2011, Vol. 51, No. 5, pp. 791–801.

Fejer Algorithms with an Adaptive Step

E. A. Nurminski

Institute for Automation and Control Processes, Far East Branch, Russian Academy of Sciences, ul. Radio 5, Vladivostok, 690041 Russia

e-mail: nurmi@dvo.ru

Received June 15, 2010; in final form, September 14, 2010

Abstract—For Fejer processes with attractants, a general adaptive scheme for step multiplier control is proposed and the convergence of this class of algorithms to stationary points is proved. Numerical results demonstrating that the convergence rate is generally linear are presented.

DOI: 10.1134/S0965542511050137

Keywords: Fejer processes, convex optimization, variational inequalities, decomposition, stepsize control.

INTRODUCTION

In [1, 2] Fejer processes with diminishing disturbances were used as models of decomposition algorithms for solving optimization and other problems. This approach provides ample opportunities for designing algorithms for distributed and parallel computations and makes it possible to overcome many difficulties associated with taking into account complicated constraints. However, disturbances with controlled step multipliers satisfying classical conditions of smallness and the divergent series condition were considered in [1, 2]. In practice, these conditions are known to lead to rather slow convergence, so the acceleration of such methods is of great theoretical and practical importance. In this context, an adaptive algorithm for step multiplier control in Fejer processes was proposed in [3] and its convergence in the simplest case was proved. Numerical experiments demonstrated considerable convergence acceleration. In this paper, the convergence of such algorithms is proved in the general case of strongly Fejer processes with attractant disturbances, which can be generated by the gradients of objective functions in optimization problems or by the operators of variational inequalities.

1. BASIC CONCEPTS AND NOTATION

Let E denote a finite-dimensional Euclidean space of vectors with the inner product xy and the corresponding norm $\|x\| = \sqrt{xx}$. The dimension of this space is denoted by $\mathrm{dim}(E)$.

The sum $A + B$, where $A, B \subset E$, is understood as $A + B = \{a + b : a \in A, b \in B\}$. For notational simplicity, a singleton $\{a\}$ is designated as a, if this does not lead to confusion.

The unit ball $\{x : \|x\| \leq 1\}$ is denoted by B, and the neighborhood of zero is understood as an arbitrary open subset of E containing the origin. Of course, the open balls $B_\varepsilon = \{x : \|x\| < \varepsilon\}$, $\varepsilon > 0$ can be used as such neighborhoods. The norm of a set A is defined as $\|A\| = \sup_{a \in A}\|a\|$.

As is customary, the convex hull of a finite family of vectors $\{a^i : i \in \mathcal{N}\} \subset E$, where $\mathcal{N} = \{1, 2, \ldots, N\}$, is defined and denoted as

$$\mathrm{co}\{a^i : i \in \mathcal{N}\} = \left\{a = \sum_{i \in \mathcal{N}} \lambda_i a^i : \sum_{i \in \mathcal{N}} \lambda_i = 1, \lambda_i \geq 0, i \in \mathcal{N}\right\} = \{a = A\lambda : \lambda \in \Delta\},$$

where $\Delta = \left\{\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N) : \sum_{i=1}^N \lambda_i = 1, \lambda_i \geq 0, i \in \mathcal{N}\right\}$ is a standard simplex and A is a $\mathrm{dim}(E) \times N$ matrix whose columns are the vectors a^i, $i \in \mathcal{N}$.

The set of all subsets of E is denoted by 2^E. A point-to-set mapping $G : E \to 2^E$ is said to be upper semicontinuous at a point x if, for any neighborhood U of zero, there exists a neighborhood V of zero such that $G(y) \subset G(x) + U$ for $y \in x + V$. A point-to-set mapping G is said to be upper semicontinuous if it is upper semicontinuous at any point $x \in E$.

Given a convex set X, its normal cone $N_X^+(x)$ at a point x is defined as $N_X^+(x) = \{p : p(y - x) \geq 0,\ y \in X\}$. Note that $N_X^+(x) = \{0\}$ for $x \in \mathrm{int}(X)$.

For brevity, the support function of a set A is denoted by $(A)_p$. Obviously, the operation $(\cdot)_p$ is positive linear: $(\lambda A)_p = \lambda(A)_p$ for $\lambda \geq 0$ and $(A + B)_p = (A)_p + (B)_p$.

2. PRELIMINARY RESULTS

The main object of study in this paper is Fejer mappings.

Definition 1. A mapping $F : E \to E$ is called Fejer with respect to a given nonempty closed set $V \subset E$ if

$$\|F(x) - v\| \leq \|x - v\| \tag{1}$$

for any $v \in V$.

Obviously, $F(v) = v$ for all $v \in V$. What is meant by the set V is usually clear from the definition of F, and the dependence of F on V is usually omitted.

Applying Fejer mappings, we can define the Fejer processes

$$x^{k+1} = F(x^k), \quad k = 0, 1, \ldots, \tag{2}$$

which are frequently used as models of iterative algorithms for solving various problems (see, e.g., [4]). Here, V is the desired solution set of the problem in question and the main point in the substantiation of the applicability of (2) is to prove that this process converges to V.

As a rule, a stronger property than that of F being a Fejer mapping is needed to prove the convergence of process (2). Accordingly, following [1], we use the definition below.

Definition 2. A Fejer mapping F is called *locally strong* if, for any $\bar{x} \notin V$, there exists a neighborhood U of zero and a number $\alpha \in [0, 1)$ such that $\|F(x) - v\| \leq \alpha\|x - v\|$ for all $v \in V$ and $x \in \bar{x} + U$.

As before, the dependence of F on V is omitted in obvious cases.

2.1. Convergence of Fejer Processes with Disturbances

For locally strong Fejer mappings, we can show that process (2) is resilient with respect to diminishing additive disturbances (see [1]); i.e., the process

$$x^{k+1} = F(x^k + z^k), \quad k = 0, 1, \ldots, \tag{3}$$

where F is a locally strong Fejer mapping and z^k is an arbitrary diminishing disturbance ($z^k \to 0$), converges to V. Convergence to V is understood in the sense that each limit point of the sequence $\{x^k\}$ belongs to V. Results of this type can be used as a complement to well-known algorithms (see the survey in [5]), for example, in the solution of convex feasibility problems.

Moreover, stronger convergence of process (3) to V can be achieved if the disturbances z^k have a special form. To describe such special disturbances, we introduce the following concept.

Definition 3. A point-to-set mapping $G : E \to 2^E$ is called a local attractant of $Z \subset V$ if $g(z - x) \geq 0$ for all $x \in V\setminus Z$, $g \in G(x)$, and $z \in Z$.

Generally speaking, to guarantee stronger convergence results for process (3), this definition has to be strengthened.

Definition 4. A local attractant G is called a *strong local attractant* if, for every $x' \in V\setminus Z$, there exists a neighborhood U of zero and a number $\delta > 0$ such that

$$g(z - x) \geq \delta > 0$$

for all $z \in Z$, $x \in x' + U$, and $g \in G(x)$.

Without explicit mention, we assume in what follows that G is locally uniformly bounded, i.e., for every x, there exists a neighborhood U of zero and a constant $C < \infty$ such that $\|G(y)\| \leq C$ for $y \in x + U$. In this case, for strong local attractants, the convergence result for process (3) can be strengthened.

Theorem 1 (see [1]). *Let F be a locally strong Fejer mapping with respect to the set V, $G : E \to 2^E$ be a strong local attractant of a set $Z \subset V$; and the sequence $\{x^s\}$ defined by the recurrence relation*

$$x^{s+1} = F(x^s + \lambda_s g^s), \quad g^s \in G(x^s), \tag{4}$$

where $\lambda_k \to +0$ and $\sum \lambda_k = \infty$, be bounded. Then any limit point of this sequence belongs to Z.

A version of this theorem that is more useful for the development of decomposition algorithms provides the opportunity to use a collection of Fejer operators $\mathcal{F} = \{\phi_i, i \in \mathcal{M} = \{1, 2, \ldots, \mathcal{M}\}\}$, each being locally strong with respect to its own set V_i $(i \in \mathcal{M})$ such that their intersection is the feasible set V.

Theorem 2 (see [1]). *Let $D(\cdot)$ be a strong local attractant of $Z \subset V$ and $\mathcal{F} = \{\phi_i, i \in \mathcal{M}\}$ be a finite family of continuous Fejer operators that are locally strong with respect to the corresponding $V_i, i \in \mathcal{M}$. For any $x \notin V = \cap_{i \in \mathcal{M}} V_i$, suppose that there is $k(x) \in \{\mathcal{M}\}$ such that $\phi_{k(x)}$ is locally strong at x. If the sequence $\{x^s\}$ generated according to the rule*

$$\begin{gathered} x^{s+1} = F_s(x^s + \lambda_s d^s), \quad d^s \in D(x^s), \quad F_s = \phi_{k_s}, \\ \text{where} \quad k_s \in \mathcal{M} \quad \text{is such that} \quad x^s + \lambda_s d^s \notin V_{k_s}, \end{gathered} \tag{5}$$

is bounded, then it converges to Z as $\lambda_k \to +0$ and $\sum \lambda_k = \infty$.

This theorem validates sequential gradient projection methods of the form

$$x^{k+1} = F_k(x^k - \lambda_k g^k), \quad k = 0, 1, \ldots, \tag{6}$$

as applied to the optimization problems

$$\min_{x \in X} f(x) \quad \text{with} \quad X = \bigcap_{i=1}^{\mathcal{M}} V_i,$$

where the Fejer operators F_k are the projections onto the respective sets V_i, with respect to which the point $x^k - \lambda_k g^k$ is infeasible. In many cases, the sets V_i can be chosen so that the projection onto them is fairly easy to implement.

Algorithms (6) with $g^k = 0$ are well known as applied to convex feasibility problems, i.e., to finding some $x \in X$ (see, e.g., [5, 9]), but their application to optimization problems has not been substantiated in this formulation thus far. A analogue of algorithm (6) can be seen in constraint linearization algorithms, where the feasible set X defined by the inequality $X = \{x : h(x) \le 0\}$ is represented as the intersection of a generally continual set of half-spaces $H_y = \{x : h(y) + g(x - y) \le 0,\ g \in \partial h(y),\ y \in E\}$ (concerning the solution of variational inequalities, see, for example, [8]). Dealing with such an infinite system of linear constraints, later, we have to choose a finite or countable set of half-spaces, followed by their aggregation and other operations, which apparently have a negative effect on the convergence of the algorithm. Possibly, this is why these algorithms are not widely used, although a detailed comparison with algorithms (6) is worth to be made. In a limited form, the idea of separately processing the constraints was also used in the split feasibility problem (see [10]). The computational efficiency of algorithm (6) is reduced by rather abstract conditions imposed on the step multipliers λ_k; these conditions are not associated with a particular problem, and they were also used in [8].

The main goal of this paper is to improve the adaptive scheme proposed in [3] for the control of λ_k in algorithms (6). Although this scheme was found to have a satisfactory (linear) convergence rate in preliminary numerical experiments, it was rigorously substantiated only for gradient-type methods in problems without constraints.

2.2. Necessary Stationarity Conditions

To describe sets of limit points of converging Fejer processes with attractants, we use conditions similar to necessary optimality ones written in the form of variational inequalities. Indeed, since Fejer operators have much in common with projective operators, they are especially suitable for solving projective equations of the form

$$x = \Pi_X(x - \lambda g), \quad g \in G(x), \tag{7}$$

where the projective operator $\Pi_X(\cdot)$ is traditionally defined as

$$\|x - \Pi_X(x)\| = \min_{z \in X} \|x - z\|, \quad \Pi_X(x) \in X. \tag{8}$$

Constrained optimization problems with a feasible set X and the variational inequalities

$$g(x^* - x) \geq 0, \quad g \in G(x^*) \quad \text{for all} \quad x \in X, \tag{9}$$

which comprise a large fraction of problems in computational mathematics, can be reduced to Eq. (7).

To relate projective equations (7) to the theory of Fejer operators with attractants under development, we have to compare their stationarity conditions and the determining relations of corresponding problems in computational mathematics. The necessary optimality conditions are well studied (see, e.g., [6, p. 142]) for convex constrained optimization problems of the form

$$\min_{x \in X} f(x), \tag{10}$$

where X is a convex closed set and f is a convex closed eigenfunction that is finite on some open set $X' \supset X$. In terms of normal cones $N_X^+(x)$ of X, we have the following result.

Theorem 3. *Let $N_X^+(x)$ be the normal cone of a convex closed set X, $\partial f(x)$ at the point x, and let $\partial f(x)$ be the subdifferential of a function f at the point x. Then the point x^* is a solution of problem* (10) *if and only if* $0 \in \partial f(x^*) + N_X^+(x^*)$.

In what follows, we need the continuity property of normal cones and subdifferentials. Again, using the traditional formalism of convex analysis, it can be shown that the set-valued mapping $\partial f(x) + N^+(x)$ is upper semicontinuous.

To analyze convergence with nontrivial operators F, we need general convergence conditions, which are briefly described below.

2.3. Convergence Conditions

In a sense, the used general convergence conditions for iterative processes (see [7]) are a discrete version of the conditions for Lyapunov asymptotic stability. Specifically, for the sequence $\{x^s\}$ generated by an iterative process to have a limit point in a given set X_*, it is sufficient that the following conditions hold.

Condition A1. The sequence $\{x^s\}$ is bounded.

Condition A2. For each subsequence $\{x^{n_k}\}$ converging to $x' \notin X_*$, there exists $\epsilon > 0$ such that, for every n_k, there is an index $m_k > n_k$ satisfying $\|x^{n_k} - x^s\| \leq \epsilon$ for $n_k \leq s < m_k$ and $\|x^{n_k} - x^{m_k}\| > \epsilon$.

Condition A3. There exists a continuous function $W(x)$ such that, for any subsequence $\{x^{n_k}\}$ converging to $x' \notin X_*$, which exists by Condition **A2** for the corresponding subsequence $\{x^{m_k}\}$, there is a subsequence $\{p_k\}$ with $n_k < p_k \leq m_k$ such that

$$\limsup_{k \to \infty} W(x^{p_k}) < \liminf_{k \to \infty} W(x^{m_k}).$$

For each limit point of $\{x^s\}$ to belong to X_*, it is sufficient that the following two additional conditions hold.

Condition A4. If $x^{n_k} \to x^* \in X_*$, then $\|x^{n_k+1} - x^{n_k}\| \to 0$.

Condition A5. The set $W(X_*) = \{W(x^*) : x^* \in X_*\}$ is such that $R \backslash W(X_*)$ is everywhere dense.

Meaningfully, Condition **A2** prevents the iterative process $\{x^s\}$ from being stuck at points that are not in X_*. Condition **A3** implicitly forbids limit cycles that do not pass through points of X_* and is an analogue of the condition that the total derivative of the Lyapunov function along the trajectory of the dynamic process is negative (see the stability theory of systems described by ordinary differential equations). Condition **A4** prevents the points of $\{x^s\}$ from jumping away from the limit set X_*. Condition **A5**, in conjunction with **A1**–**A4**, guarantees the convergence of $\{W(x^s)\}$. Together with **A3**, this prevents $\{x^s\}$ from having limit points that are not in X_*.

3. METHOD FOR ADAPTIVE STEPWISE STEPSIZE CONTROL

The idea behind adaptive stepsize control in Fejer algorithms with attractants was proposed in [3] and can be described as follows. Consider the iterative process

$$x^{k+1} = F(x^k + \lambda_k g^k), \quad g^k \in G(x^k), \tag{11}$$

where $G(x)$ is a point-to-set attractant satisfying the conditions to be specified later. We analyze the convergence of process (11) to the set X_* defined by the stationarity conditions

$$X_* = \{x^* : 0 \in G(x^*) + N_X^+(x^*)\}, \tag{12}$$

where X is the set of stationary points of the Fejer mapping F.

Process (11) can be represented as

$$x^{k+1} = x^k + \lambda_k d^k, \quad d^k = [(F(x^k + \lambda_k g^k) - x^k]/\lambda_k, \quad g^k \in G(x^k), \tag{13}$$

where d^k can be assumed to belong to some mappings D_k. Equivalently, $d^k = (x^{k+1} - x^k)/\lambda_k$.

For notational simplicity, we set $D(p,q) = \mathrm{co}\{d^p, d^{p+1}, \ldots, d^q\}$..

Given a sequence $\theta_m \to +0$, $m = 0, 1, \ldots$, the corresponding sequences of indices $\{k_m\}$ and numbers $\{\lambda_k\}$ are defined as follows:

Set $m = 0$ and $k_0 = 0$ and choose an initial step $\lambda_0 > 0$. Let $q \in (0,1)$.

For given m and k_m, define k_{m+1} as an index satisfying the conditions

$$0 \notin D(k_m, k) + \theta_m B, \quad k_m \le k < k_{m+1}, \quad 0 \in D(k_m, k_{m+1}) + \theta_m B, \tag{14}$$

with $\lambda_k = \lambda_{k_m}$ for $k_m \le k < k_{m+1}$. Set

$$\lambda_{k_{m+1}} = q\lambda_{k_m}. \tag{15}$$

Increment $m : m \to m+1$ and continue the iteration of algorithm (11).

Meaningfully, condition (14) is used to determine the moment when some segment of $\{x^k\}$ begins to satisfy a substitute for stationarity condition (14). As λ_k decreases to zero, which is guaranteed by the factor $q < 1$, this segment of the sequence shrinks to an arbitrarily small neighborhood where stationarity condition (12) is satisfied.

In [3] the stepsize control scheme proposed was substantiated in the simple case where F_k are identity operators and, in fact, $d^k = g^k$ and $D_k(x) = G(x)$, although numerical experiments demonstrated that this scheme also applies to general scheme (11). Below, the stepsize control procedure (14) is substantiated in the general case.

For the proof, we need several auxiliary results.

Lemma 1. *If the sequence* $\{x^k\}$ *generated by* (11*) with stepsize control* (14) *is bounded, then all its limit points belong to* X.

Proof. It is easy to see that, if $\{x^k\}$ is bounded, then the index sequence of $\{k_m\}$ defined by (14) is infinite. Assuming the opposite, for some $\bar{m}$, we would have $0 \notin D(k_{\bar{m}}, k) + \theta_{\bar{m}} B$ for all $k > k_{\bar{m}}$. Therefore, since $D(k_{\bar{m}}, k)$ is monotone with respect to k by the inclusion, we have

$$\bar{0} \notin \mathrm{ri}\left[\lim_{k\to\infty} D(k_{\bar{m}}, k)\right] + \theta_{\bar{m}} B = \mathrm{ri}(\bar{D}) + \theta_{\bar{m}} B,$$

where the limit of sets is understood as $\bar{D} = \bigcap_{k > k_{\bar{m}}} D(k_{\bar{m}}, k)$. Accordingly, there exists p with $\|p\| = 1$ such that

$$(\bar{D} + \theta_{\bar{m}} B)_p \le 0,$$

or

$$(\bar{D})_p \le -\theta_{\bar{m}} < 0.$$

Moreover, $\lambda_k = \lambda_{\bar{m}} > 0$ for all $k > k_{\bar{m}}$ and, accordingly,

$$-\left\|x^k - x^{k_{\bar{m}}}\right\| \leq (x^k - x^{k_{\bar{m}}})p = \sum_{s=k_{\bar{m}}}^{k-1}(x^{s+1} - x^s)p$$

$$= \sum_{s=k_{\bar{m}}}^{k-1}\frac{d^s}{\lambda_s}p \leq \sum_{s=k_{\bar{m}}}^{k-1}\frac{1}{\lambda_s}(\bar{D})_p \leq -\frac{\theta_{\bar{m}}}{\lambda_{\bar{m}}}(k-1-k_{\bar{m}}) \to -\infty$$

as $k \to \infty$. Therefore, $\left\|x^k\right\| \to \infty$ as $k \to \infty$, which contradicts the boundedness of $\{x^k\}$.

In turn, the unboundedness of $\{k_m\}$ implies that $\lambda_k \to 0$ as $k \to \infty$. Consequently, by Theorem 3,

$$\lim_{s\to\infty} x^{k_s} = x' \in V$$

for any converging subsequence $\{x^{k_s}\}$.

Now let $\lim_{k\to\infty} x^{n_k} = x'$. Then, by Lemma 1, $x' \in X_*$. The behavior of $\{x^k\}$ in the neighborhood of x' can be determined more accurately by applying the following lemma.

Lemma 2. *If* $x^{n_k} \to x'$ *and* $0 \notin G(x') + N_X^+(x')$, *then there exists* $\varepsilon > 0$ *such that* $\left\|x^{n_k} - x^{m_k}\right\| > \varepsilon$ *for some* $m_k > n_k$ *for all* k.

Proof. Assume the opposite; that is, there exists a subsequence $\{x^{n_k}\} \to x' \notin X_\star$ such that, for any $\varepsilon > 0$, there is k satisfying $\left\|x^s - x^{n_k}\right\| \leq \varepsilon$ for all $s > n_k$. Generally speaking, this means that the entire sequence $\{x^s\}$ converges to x'.

Since the set $G(x') + N_X^+(x')$ is closed and bounded away from 0, there exist p and δ such that $p \neq 0, \|p\| = 1$, and $py \geq \delta$ for all $y \in G(x') + N_X^+(x') + U$, where U is a neighborhood of zero. Then, for $s > n_k$ and sufficiently large k with $\bar{x}^s = x^s - \lambda_s g^s$, we have

$$\frac{1}{\lambda_s}(\bar{x}^s - x^s) = g^s \in G(x') + U,$$

since G is upper semicontinuous, and it holds that $\bar{x}^s - F_s(\bar{x}^s) \in N_X^+(x') + U$ since F is a Fejer mapping.

Therefore,

$$p(x^{s+1} - x^s) = \lambda_s p\left(\frac{F(\bar{x}^s) - \bar{x}^s}{\lambda_s} + \frac{\bar{x}^s - x^s}{\lambda_s}\right) = \lambda_s p w^s,$$

where $w^s \in G(x') + N_X^+(x') + U$. Consequently,

$$p(x^{s+1} - x^s) \geq \lambda_s\delta > 0, \quad s \geq n_k.$$

Without loss of generality, we can assume that $\lambda_s \geq q\lambda_{n_k}$ for $s > n_k$, since, in this case, the step can be divided at most once. Summing up the last inequalities over $s > n_k$ from n_k to $m-1$ yields

$$p(p^m - x^{n_k}) \geq q\lambda_{n_k}(m - n_k) \to \infty$$

as $m \to \infty$, which contradicts the boundedness of $\{x^s, s \geq n_k\}$ and, hence, proves the lemma.

Using Lemmas 1 and 2, we can prove the final assertion concerning the convergence of algorithms (6).

Theorem 4. *Let the sequence* $\{x^k\}$ *be generated by* (11), *where* F *is a locally strong Lipschitz continuous Fejer operator with respect to X, G is a bounded upper semicontinuous attractant of the set* $X_* = \{x^* : G(x^*) + N_X^+(x^*)\}$, *and the sequence of step multipliers is determined by* (14) *and* (15). *Then all its limit points belong to* X_*.

Proof. Lemmas 1 and 2 guarantee that the limit points of $\{x^k\}$ are feasible and Condition **A2** is satisfied at each of them. Since **A4** holds by construction, it remains to check the fulfillment of Conditions **A3** and **A5**.

Define $W(x) = \inf_{x^*\in X_*}\|x - x^*\|^2$. Then Condition **A5** holds automatically. Furthermore,

$$\begin{aligned}\|x^{s+1} - x^*\|^2 &= \|F(x^s - \lambda_s g^s) - x^*\|^2 \le \alpha\|x^s - \lambda_s g^s - x^*\|^2\\ &\le \alpha(\|x^s - x^*\|^2 - 2q\lambda_{n_k} g^s(x^s - x^*) + \lambda_s^2\|g^s\|^2)\\ &\le \alpha(\|x^s - x^*\|^2 - 2q\lambda_{n_k}\delta + q^2\lambda_{n_k}^2 C) \le \alpha\|x^s - x^*\|^2 - \gamma\lambda_{n_k}\end{aligned}$$

with a constant $\gamma > 0$. Calculating the infimum, we obtain

$$W(x^{s+1}) \le \alpha W(x^s) - \gamma\lambda_{n_k} \le W(x^s) - \gamma\lambda_{n_k}.$$

Adding these inequalities in s from n_k to $m_k - 1$ gives

$$W(x^{m_k}) \le W(x^{n_k}) - \gamma\lambda_{n_k}(m_k - n_k - 1). \tag{16}$$

On the other hand,

$$\varepsilon < \|x^{m_k} - x^{n_k}\| \le \sum_{s=n_k}^{m_k-1}\|x^{s+1} - x^s\| \le Lq\lambda_{n_k}(m_k - n_k) \le \kappa\lambda_{n_k}(m_k - n_k),$$

whence $\lambda_{n_k}(m_k - n_k) > \varepsilon / \kappa$.. Substituting this estimate into (16), we obtain

$$W(x^{m_k}) \le W(x^{n_k}) - \varepsilon/\kappa + \gamma\lambda_{n_k} \le W(x^{n_k}) - \varepsilon/2\kappa$$

for sufficiently large k. Passage to the limit yields

$$\limsup_{k\to\infty} W(x^{m_k}) \le \lim_{k\to\infty} W(x^{n_k}) - \varepsilon/2\kappa < W(x'),$$

which proves that Condition **A3** holds. The proof of the theorem is completed.

4. NUMERICAL EXPERIMENTS

The goal of the numerical experiments conducted with the algorithms involving adaptive stepwise step-size control was primarily to analyze the qualitative character of convergence for problems of various classes. In this context, we considered constrained and unconstrained convex optimization problems of different structure and a variational inequality that is not reduced to an optimization problem. As a result, the general idea of stepsize control was tested using problems traditionally considered in different areas of computational mathematics and the first conclusions were made about the practical applicability of this approach.

4.1. Convex Optimization Problems

The first test problem was the minimization of a linear objective function on the set given by the set of convex quadratic cylindrical constraints

$$\min cx$$
$$\sum_{i=1,\ i\ne k}^{n} x_i^2 \le 1, \quad k = 1, 2, \ldots, n. \tag{17}$$

The problem was solved for the low dimension $n = 4$, but, even in this case, the determination of an extreme point on the boundary is a rather complicated problem. Figure 1 shows the convergence of the gradient method as applied to this problem with sequential projection onto the system of constraints.

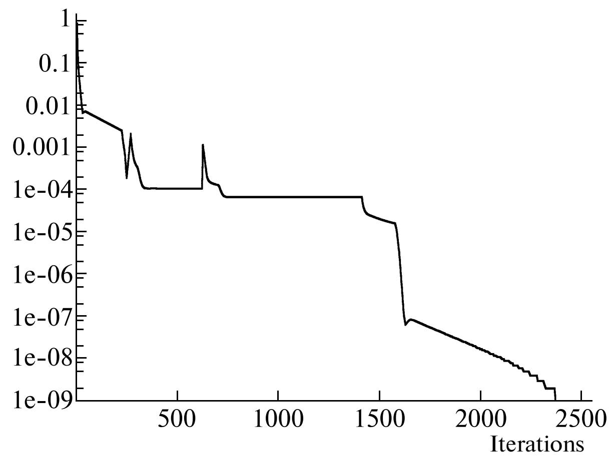

Fig. 1.

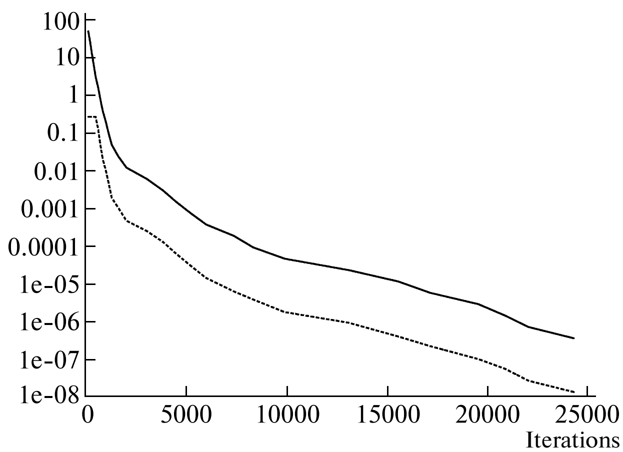

Fig. 2.

In this case, a gradient step was taken from the current iteration point $\bar{x}^{k+1} = x^k - \lambda_k c$ and the result was then projected onto the $j_k$kth constraint defined by the relation $j_k = 1 + k \bmod n$:

$$\|x^{k+1} - \bar{x}^{k+1}\| = \min\|x - \bar{x}^{k+1}\|, \quad x \in C(j_k), \quad C(j_k) = \left\{x : \sum_{i=1, i \neq j_k}^{n} x_i^2 \leq 1\right\}.$$

This strategy in convex feasibility problems is known as the round-robin policy. Figure 1 shows only the iterations corresponding to feasible points, which makes it possible to estimate the type of algorithm convergence with respect to the search for an optimal value of the objective function. It can be seen that, despite the considerable lacunas due to infeasible points (horizontal lines in the plot), the algorithm generally exhibits a linear convergence rate.

Numerical experiments were also performed with the TR48 problem, whose dual is known as a rather complicated problem for subgradient algorithms. The goal is to minimize the piecewise linear function

$$-f(x) = \sum_{i=1}^{48} s_i x_i + \sum_{j=1}^{48} d_j \min_{i=1,\ldots,48} (a_{ij} - x_i), \tag{18}$$

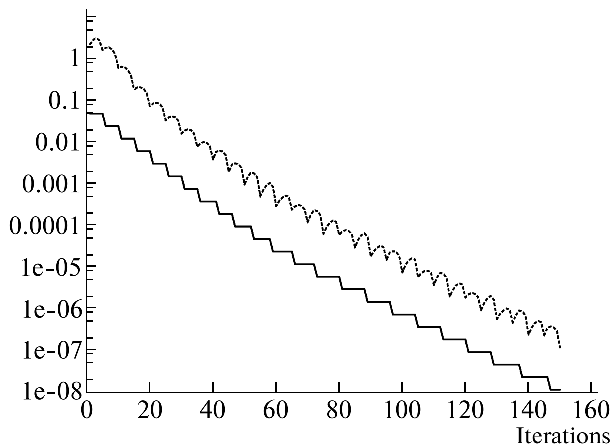

Fig. 3.

where the 48×48 matrix A and the 48-dimensional vectors s and d are defined as in [11]. A Fortran and MATLAB-compatible code for this problem, together with the necessary data, can be found at the site [12] of the Joint Russian–Ukrainian Project of the Russian Foundation for Basic Research (project no. 09-01-90413-Ukr_f_a) and the State Fund for Fundamental Research (project no. F28.1.005) in the section “Tests”⟶ “Test functions”⟶ “medium.” Figure 2 shows the convergence to zero of the relative error in the determined optimal value of the objective function (solid curve) and the decrease in the step multiplier (broken curve) for the solution of the TR48 problem. It is well seen that these characteristics decrease linearly, though not very fast. This suggests that the number of iterations between the step divisions is uniformly bounded, which is apparently characteristic of convex polyhedral functions as functions with a sharp minimum.

4.2. Variational Inequalities

As an example, consider the variational inequality $G(x)(y - x) \geq 0$, $x \in X$,, where

$$G(x) = (1 + \|x\|)(x - x^*), \quad X = \{x \leq 0\}, \tag{19}$$

with the point $x^* = (1, 2, \dots, n)$.. The solution of this inequality is the point $(0, \dots, 0)$. Figure 3 demonstrates the convergence of the iterative sequence generated by algorithm (5) involving stepwise stepsize control with the nonpositivity constraints taken into account sequentially. As before, the j_kth constraint $x_{j_k} \leq 0$, where $j_k = 1 + k \bmod n$, was checked at the kth iteration step. If it was violated, the trivial projection onto the half-space $H(j_k) = \{x : x_{j_k} \leq 0\}$ was performed. Figure 3 shows the Euclidean distance to the solution (solid curve) and a decreasing stepsize (broken curve). As in the above experiments with convex optimization problems, it is well seen that the algorithm converges linearly.

CONCLUSIONS

Both the theoretical substantiation of the adaptive stepsize control method in Fejer processes with attractants and the numerical results suggest that this approach is widely applicable to problems of various structures. Of special practical interest is that the experimentally observed convergence rate is linear, which considerably accelerates Fejer processes as compared with the classical conditions that the step multipliers are small and their sum diverge. However, this finding still requires rigorous theoretical substantiation.

REFERENCES

1. E. A. Nurminskii, “The Use of Additional Diminishing Disturbances in Fejer Models of Iterative Algorithms,” Zh. Vychisl. Mat. Mat. Fiz. **48**, 2121–2128 (2008) [Comput. Math. Math. Phys. **48**, 2154–2161 (2008)].

2. E. A. Nurminskii, "Fejer Processes with Diminishing Disturbances," Dokl. Akad. Nauk **422**, 601–604 (2008) [Dokl. Math. **78**, 755–758 (2008)].
3. E. A. Nurminski, "Envelope Stepsize Control for Iterative Algorithms Based on Fejer Processes with Attractants," Optmizat. Methods Software **25** (1), 97–108 (2010).
4. V. V. Vasin and I. I. Eremin, *Operators and Fejer-Type Iteration Processes: Theory and Applications* (Ural. Otd. Ross. Akad. Nauk, Yekaterinburg, 2005) [in Russian].
5. H. H. Bauschke and J. M. Borwein, "On Projection Algorithms for Solving Convex Feasibility Problems," SIAM Rev. **38**, 367–426 (1996).
6. B. N. Pshenichnyi, *Convex Analysis and Optimization Problems* (Nauka, Moscow, 1980) [in Russian].
7. E. A. Nurminskii, *Numerical Methods of Convex Optimization* (Nauka, Moscow, 1991) [in Russian].
8. M. Fukushima, "A Relaxed Projection Method for Variational Inequalities," Math. Program. **35**, 58–70 (1986).
9. C. L. Byrne, "A Unified Treatment of Some Iterative Algorithms in Signal Processing and Image Reconstruction," Inverse Problems **20**, 103–120 (2004).
10. Q. Yang, "The Relaxed CQ Algorithm Solving the Split Feasibility Problem," Inverse Problems **20**, 1261–1266 (2004).
11. *Nonsmooth Optimization: Proceedings of IASA Workshop, March 28–April 8, 1977* (Int. Inst. Appl. Systems Analysis, IIASA Proc. Ser.: Vol. 3), Ed. by C. Lemarechal and R. Mifflin (Pergamon, Oxford, 1978).
12. http://elis.dvo.ru (Joint Russian–Ukrainian Project "Subgradient Methods of Accelerated Convergence in Convex Programming Problems" of the Russian Foundation for Basic Research, Russia, and the State Fund for Fundamental Research, Ukraine).