ISSN 1064-5624, Doklady Mathematics, 2008, Vol. 78, No. 2, pp. 755–758. © Pleiades Publishing, Ltd., 2008.
Original Russian Text © E.A. Nurminski, 2008, published in Doklady Akademii Nauk, 2008, Vol. 422, No. 5, pp. 601–604.

MATHEMATICS

Fejer Processes with Diminishing Disturbances

E. A. Nurminski

Presented by Academician S.K. Korovin November 12, 2007

Received June 3, 2007

DOI: 10.1134/S106456240805030X

Fejer processes are widely used to solve systems of convex inequalities. The present-day theory of Fejer processes and applications can be found in [1]. In the general form, the Fejer process is an iterative scheme of the form $x^{s+1} = F(x^s)$, $s = 0, 1, 2, \ldots$, where $F(x)$ is a Fejer operator. A distinctive feature of Fejer operators and processes based on this scheme is that they are attracted and eventually converge to certain sets. For this reason, they are used as theoretical models of computational algorithms. These processes are especially useful when applied to decomposition and parallel computations. For this reason, they are widely applied to large-scale problems in computer tomography, radiation therapy scheduling, pattern recognition, image processing, and other areas associated with processing large amounts of data. In this paper, we analyze the behavior of Fejer processes with a diminishing disturbance generated by a small shift in the argument of the Fejer operator. It is shown that, if $F(x)$ is a locally strongly Fejer operator, then a diminishing disturbance does not prevent convergence to an attracting set. At the same time, such a disturbance can be used to furnish the process with additional properties that ensure convergence to certain subsets of the attracting set. In particular, based on this scheme, a new decomposition principle for optimization problems can be suggested that does not require that the constraints possess a specific structure.

CONVERGENCE OF FEJER SEQUENCES WITH DISTURBANCES

The consideration below is associated primarily with the finite-dimensional Euclidean space E equipped with the inner product xy and the norm $\|x\| = \sqrt{xx}$. Situations where a vector x is multiplied by a scalar factor α are usually clear from the context. A standard N-dimensional simplex is denoted by $\Delta_N = \left\{ w_i \geq 0,\ i = 1, 2, \ldots, N;\ \sum_{i=1}^{N} w_i = 1 \right\}$. The neighborhood of zero is an arbitrary open set containing $0 \in E$.

In the standard manner, the Fejer operator is defined with respect to a given set V as follows.

Definition 1. An operator F: $E \to E$ is called Fejer (with respect to a given set V) if $F(v) = v$ for $v \in V$ and

$$\|F(x) - v\| \leq \|x - v\| \quad (1)$$

for all $v \in V$.

The set V is usually clear from the context and is hereafter assumed to be closed and bounded. In addition to the definition, we also assume that $F(x)$ is continuous on an open extension of V.

Given a Fejer operator F and an initial point x^0, we can construct an iterative Fejer process $x^{s+1} = F(x^s)$ $(s = 0, 1, \ldots)$ that models a computational algorithm for determining a point or points of V (the feasibility problem). Property (1), more exactly its various stronger versions, guarantees that the elements of $\{x^s\}$ converge to V in a certain sense. More than 100 publications on this subject were overviewed in [2] (and that list is far from being complete). As a rule, for a Fejer process to converge sufficiently strongly, it must have stronger properties than those in Definition 1, such as quasi-Fejerness, quasi-compressibility, etc. Bearing in mind the subsequent applications, we propose the following property of F, which is also stronger than (1).

Definition 2. A Fejer operator F is called locally strongly Fejer at a point $\bar{x} \notin V$ if there exists a neighborhood U of zero such that

$$\|F(x) - v\| \leq \alpha \|x - v\| \quad (2)$$

Institute for Automation and Control Processes, Far East Division, Russian Academy of Sciences, ul. Radio 5, Vladivostok, 690041 Russia
e-mail: nurmi@dvo.ru

for all $v \in V$, $x \in \bar{x} + U$, and some $\alpha < 1$.

An operator $F(x)$ is called a locally strongly Fejer operator if it is locally strongly Fejer at any point $x \notin V$. Of course, the attraction coefficient α depends on the point chosen.

Theorem 1. *Let V be a closed and bounded set; F be a locally strongly Fejer operator; the sequence $\{x^s\}$ generated by the recurrence relations*

$$x^{s+1} = F(x^s + z^s), \quad s = 0, 1, 2, \dots \tag{3}$$

be bounded; and $z^s \to 0$ as $s \to \infty$.

Then, for arbitrary x^0, all the limit points of $\{x^s\}$ belong to V.

This result can be extended to nonstationary Fejer sequences of the form

$$x^{s+1} = F_s(x^s + z^s), \quad s = 0, 1, 2, \dots, \tag{4}$$

where F_s is chosen from a finite family of Fejer operators, which allows us to use the apparatus of Fejer processes in decomposition algorithms.

Theorem 2. *Let $\mathcal{F} = \{P_1, P_2, \dots, P_m\}$ be a family of operators P_i such that, for any $x \notin V$, there exists P_i that is locally strongly Fejer at x, $z^s \to 0$ as $s \to \infty$, and $F_s = P_{i_s}$ (where P_{i_s} is a locally strongly Fejer operator at x^s).*

Then, if the sequence $\{x^s\}$ defined by (4) *is bounded, all its limit points belong to V.*

Theorems 1 and 2 show that, under rather mild conditions, the diminishing disturbance does not prevent convergence to some set of fixed points, which is the main result in the theory of Fejer processes. Below, we present results concerning how small disturbances z^s can be used to make processes (3) and (4) converge to certain subsets of V.

We introduce the concept of a localized attractant as a vector field that is directed inside V toward a subset of V.

Definition 3. A point-to-set mapping Φ: $V \to 2^E$ is called a localized attractant of $Z \subset V$ at a point x if $x \in V \backslash Z$ implies $g(z - x) \geq 0$ for all $g \in \Phi(x)$ and $z \in Z$.

In fact, a somewhat stronger property is necessary to substantiate the desired convergence.

Definition 4. An attractant Φ is called a strong localized attractant at a point x' if $x' \in V \backslash Z$ implies that there exists a neighborhood U of zero such that

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

for all $z \in Z$, $x \in x' + U$, and $g \in \Phi(x)$ and for some $\delta > 0$.

Given a fixed Z, we say that Φ is a strong localized attractant if the above property holds at each point of $V \backslash Z$. Obviously, the constant δ depends on the point chosen.

Theorem 3. *Let F be a locally strongly Fejer operator; Φ be a bounded and upper semicontinuous localized strong attractant of $Z \subset V$; and the sequence $\{x^s\}$ be generated by*

$$x^{s+1} = F(x^s + \lambda_s \Phi(x^s)), \tag{5}$$

where x_0 is an arbitrary initial state, $\lambda_s \to +0$, and $\sum \lambda_s = \infty$.

Then, if $\{x^s\}$ is bounded, all its limit points belong to Z.

As in the case of Theorem 1, this result can be extended to nonstationary Fejer operators.

Theorem 4. *Let F_s be a locally strongly Fejer operator at x^s chosen from a finite family $\mathcal{F} = \{P_1, P_2, \dots, P_m\}$ of continuous operators P_i such that $P_i(v) = v$ ($i = 1, 2, \dots, m$) for all $v \in V$ and, for any $x \notin V$, there exists P_i that is locally strongly Fejer at x; $z^s \to 0$ as $s \to \infty$; Φ be a bounded upper semicontinuous strong localized attractant $Z \subset V$; and the sequence $\{x^s\}$ be generated by*

$$x^{s+1} = F_s(x^s + \lambda_s \Phi(x^s)), \tag{6}$$

where x^0 is an arbitrary initial state, $\lambda_s \to +0$, and $\sum \lambda_s = \infty$.

Then, if $\{x^s\}$ is bounded, all its limit points belong to Z.

Theorems 1–4 are nontrivial and do not follow from the existing theory of Fejer processes, since, generally speaking, none of the processes given by (3)–(6) is Fejer. The proofs of the theorems are based on general convergence conditions for iterative processes [3].

Theorems 1–4 involve the rather restrictive and difficult-to-check (at first glance) global condition that the sequence $\{x^s\}$ is bounded. However, the algorithmic schemes (in Theorems 1–4) are easy to modify with the help of a retract R: $E \to \tilde{V}$ that returns the process $\{x^s\}$ to a bounded set $\tilde{V}$ such that $\tilde{V} \subset V + U$. For example,

$$x^{s+1} = \tilde{F}(x^s + z^s) = \begin{cases} F(x^s + z^s), & x^s \in V + U \\ R(x^s) = y^s \in \tilde{V} & \text{otherwise}, \end{cases} \tag{7}$$

where U is a bounded neighborhood of zero and y^s is chosen arbitrarily in $\tilde{V}$.

GRADIENT PROJECTION METHOD WITH DECOMPOSITION OF THE CONSTRAINT SYSTEM

In practice, the above results can be applied as follows. On the one hand, the wide range of Fejer operators can be used to solve feasibility problems, which guarantee that the limit points of the constructed sequence $\{x^s\}$ belong to the feasible set V even in the presence of a disturbance z^s, $s = 0, 1, 2, \ldots$. On the other hand, attractants can be used to improve a feasible point so as to make it closer to the set Z of distinguished points in V. Examples are points that solve an optimization or other problem on V. In this case, many methods are available for determining, say, relaxation directions in which the current feasible point approaches the solution set. The essence of Theorems 3 and 4 is that, under their assumptions, schemes for deriving feasible points and algorithms for achieving the distinguished subset can be fairly easily combined.

In this context, as an application, we consider the convex programming problem

$$f_\star = \min_{x \in V} f(x) = f(x^\star), \quad x^\star \in Z \subset V, \tag{8}$$

where f is a convex finite objective function and V is a convex feasible set. A general solution method for problem (8) is to combine gradient steps and the projection onto V:

$$x^{s+1} = \Pi(x^s - \lambda_s g^s), \quad g^s \in \partial f(x^s), \tag{9}$$

where Π is the projection onto V, but the last operation is laborious for a general set and is rarely used, except for very simple V. However, V is nearly always the intersection of a family of convex subsets V_i, $i = 1, 2, \ldots, N$: $V = \bigcap_{i=1}^{N} V_i$. At least for feasibility problems, a range of Fejer algorithms are available that use only individual projections onto the elements V_i in the representation $V = \bigcap_{i=1}^{N} V_i$. In contrast to V, the elements V_i can be relatively simple sets, for example, half-spaces, linear manifolds, rays, spheres, etc., so that the projection onto them is easy to implement computationally.

To apply Theorem 2 to projectors, it suffices to show that they are locally strong Fejer.

Theorem 5. *Let V be a closed bounded set representable as the intersection of a finite or infinite family of convex subsets*: $V = \bigcap_{\tau \in T} V_\tau$. *Denote by $\Pi_\tau(x)$ the projection of a point x onto V_τ.*

Then, if $x \notin V_{\tau'}$ for some $\tau' \in T$, the operator $F = \Pi_{\tau'}$ is locally strongly Fejer at x.

Theorem 2 implies that, for a finite set $T = \{1, 2, \ldots, N\}$, the operator F_s constructed by choosing at the point x^s the operator $F_s = \Pi_{i_s}$ with $x^s \notin V_{i_s}$ guarantees that the simple iteration $x^{s+1} = F_s(x^s + z^s)$ as applied to the feasibility problem converges under fairly weak conditions on the disturbance z^s. Note that the method for choosing V_{i_s} is of no importance. Therefore, in terms of convergence theory, nearly all row-action methods [4], such as cyclic projection, farthest set projection, intermittent methods, maximum residual, etc., are covered by Theorem 2.

However, Theorem 4 provides an additional advantage due to the attractant $\Phi(x^s) = -\partial f(x_s)$, i.e., a subdifferential mapping of the objective function in problem (8). Since $0 \le f(x) - f_\star \le g(x - z)$ for $g \in \partial f(x)$, $x \in V \cap Z$, the mapping $\Phi(\cdot)$ is an upper semicontinuous bounded localized strong attractant for Z.

Letting $z^s = \lambda_s g^s$, where $g^s \in \Phi(x^s) = -\partial f(x^s)$, and applying Theorem 4 yields the convergence of various alternating gradient projections onto the decomposition elements of V:

$$x^{s+1} = \Pi_s(x^s - \lambda_s g^s), \quad s = 0, 1, 2, \ldots, \tag{10}$$

where Π_s is the projector onto a set V_{i_s} such that $x^s \notin V_{i_s}$, $\lambda_s \to +0$, and $\sum_{s=0}^{\infty} \lambda_s = \infty$. Based on this approach, the gradient projection method (9) can be decomposed according to (10) so that the hard-to-implement projection onto V is replaced by the projection onto V_i.

Using Theorem 5, we can show that the operator

$$F(x) = \sum_{i=1}^{N} w_i \Pi_i(x), \tag{11}$$

where Π_i is the projection onto V_i and $w = (w_1, w_2, w_N) \in \Delta_N$, is a locally strong Fejer operator.

Theorem 6. *Let the operator F given by* (11) *be such that* $\sum_{i:\, x \notin V_i} w_i \ge \gamma > 0$. *Then $F(x)$ is a locally strongly Fejer operator at the point x.*

Due to this result, we can substantiate and use parallel version of the gradient projection decomposition method with a Fejer operator of the form (11)

$$x^{s+1} = \sum_{i=1}^{N} w_i^s \Pi_i(x^s - \lambda_s g^s),$$

where the projections can be performed in parallel. The conditions imposed on the weights w_i^s in Theorem 6 are satisfied, for example, when all of them are uniformly bounded away from zero: $w_i^s \geq \epsilon > 0$. For step multipliers λ_s, it sufficient that $\lambda_s \to +0$, $\sum_{s=0}^{\infty} \lambda_s = \infty$, which are traditional conditions for nonsmooth gradient schemes. However, numerical experience suggests that these conditions lead to rather slow convergence and such methods need to be further improved.

ACKNOWLEDGMENTS

This work was supported by the Far East Division of the Russian Academy of Sciences, project no. 06-III-A-01-459.

REFERENCES

1. V. V. Vasin and I. I. Eremin, *Operators and Fejer Iterative Processes: Theory and Applications* (Ural. Otd. Ross. Akad. Nauk, Yekaterinburg, 2005) [in Russian].
2. H. H. Bauschke and J. M. Borwein, SIAM Rev. **38**, 367–426 (1996).
3. E. A. Nurminskii, *Numerical Methods in Convex Optimization* (Nauka, Moscow, 1991) [in Russian].
4. Y. Censor, SIAM Rev. **23**, 444–466 (1988).