Submitted to OMS

Fejer processes with diminishing disturbances and decomposition of constrained nondifferentiable optimization problems *

Nurminski E.A. †

Abstract

Iterative processes based on Fejer mappings with diminishing problem-specific shifts in the arguments are considered. Such structure allows fine-tuning of Fejer processes by directing them toward selected subsets of attracting sets. Use of various Fejer operators provides ample opportunities for decomposition and parallel computations. Subgradient projection algorithms with sequential and simultaneous projections on segmented constraints are considered as an example. To speed up convergence of this type of algorithms the novel step-size rule is proposed.

Keywords: Fejer processes, convex optimization, decomposition, stepsize control, subgradient method

Introduction

Fejer processes and their generalizations are well known as an algorithmic model for many convex optimization techniques [1]. They are also widely used for solving convex feasibility problem (CFP) of finding an element of convex set represented by a system of convex inequalities [2]. These two areas of application to a certain extent are considered independent of each other — advances in solving CFP via Fejer-type methods involving decomposition and parallelization so far are hardly applicable to optimization problems. and vice-versa: optimization techniques based on Fejer processes as such are of a little interest to CFP because of resources limitations. Typically CFP is viewed as a large-scale problem with traditional applications in medical imaging, computer tomography, pattern recognition and like [4]. In such cases it is difficult to apply even simplest optimization techniques with anything short of triviality for taking constraints into account.

The aim of this paper is to show that these two directions however can nevertheless be combined under fairly permissible conditions. It means that for instance for solving convex optimization problems one can use Fejer processes of a general nature to take care of feasibility on one hand, and use an optimum-directed displacements of almost any kind in the arguments of Fejer operators to attain optimality on the other. The same idea can be be applied and for another problems, solution of variational inequalities par example. In this aspect the article continues the the investigation started in [8].

To begin with we develop convergence theory of such combined processes under classical "divergence sum" conditions, however the paper is concluded with envelope step-size control (ESC) schema which shows some promises to speed up convergence.

*Supported by RFBR grant 09-01-90413-Ukr_f_a

†IACP, Vladivostok

1 Notations and preliminaries

Let E denotes a finite dimensional euclidian space with inner product xy and norm $\|x\| = \sqrt{xx}$. The unit ball $\{x : \|x\| \leq 1\}$ will be denoted as B. Convex hull of a family of vectors $a^i, i = 1, 2, \ldots, N$ will be denoted as

$$\mathrm{co}\,\{a^i, i = 1, 2, \ldots, N\} = \{a = \sum_{i=1}^{N} \lambda_i a^i, \sum_{i=1}^{N} \lambda_i = 1, \lambda_i \geq 0, i = 1, 2, \ldots, N\} = \{a = A\lambda, \lambda \in \Delta\},$$

where $\Delta = \{\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N) : \sum_{i=1}^{N} \lambda_i = 1, \lambda_i \geq 0, i = 1, 2, \ldots, N\}$ — standard simplex, X — the matrix of column vectors $x^i, i = 1, 2, \ldots, N$.

For notational convenience we denote as $\{a_k\}_q^p$ a segment of the sequence $\{a_k\}$ such that $k = p, p+1, \ldots, q$ and to simplify notations even more let $\{a_k\}_1^\infty = \{a_k\}$.

In what follows we will make a heavy use of certain convergence conditions which proved to be rather convenient for the analysis of iterative algorithms [6] especially in the field of nondifferentiable optimization. From the point of view of these conditions an algorithm is a rule for constructing an infinite sequence of points $\{x^k\}$ which should converge to some target set $X_\star$. This target set is preselected prior to running algorithm and in practice may be a solution set of a given optimization or feasibility problem, fixed points of a given operator, set of points satisfying necessary optimality conditions and like.

Convergence for a certain subsequence of $\{x^k\}$ is granted if following conditions are fulfilled:

A1 Sequence $\{x^k\}$ is bounded.

A2 There is a continuous function $W(x)$ such that if there exists a limit point $x' \notin X_\star$ then there is another limit point x'' such that $W(x'') < W(x'')$.

It is easy to show that under these conditions there exist a limit point $x^\star \in X_\star$. Indeed, denote a set of limit points of sequence $\{x^k\}$ as $\bar{X}$. It is closed bounded set and because of continuity of W the set $\bar{W} = \{W(x), x \in \bar{X}\}$ is closed and bounded as well. Let $w_\star = \min w : w \in \bar{W}$ and $\{x^{k_t}\}$ be the corresponding subsequence such that $\lim_{t\to\infty} W(x^{k_t}) = w_\star$. Without loss of generality one can assume that there is a limit $\lim_{t\to\infty} x^{k_t} = x^\star$. Obviously $x^\star \in X$ otherwise according to condition **A2** there is another limit point $\bar{x}^\star$ with $W(\bar{x}^\star) < W(x^\star) = w_\star$ which contradicts the definition of $w_\star$.

Conditions **A1-A2** are insufficient however to prove that all limit points of $\{x^k\}$ belong to $X_\star$. For the latter it is necessary to request more stringent monotonicity of the sequence $\{W(x^k)\}$ and ask for specific features of the the $W_\star = \{W(x), x \in X_\star\}$. The resulting conditions may look as follows:

B1 Sequence $\{x^k\}$ is bounded.

B2 For $\{x^{k_t}\} \to x'$ when $t \to \infty$ with $x' \notin X_\star$ there exists $\epsilon > 0$ such that for any t

$$m_t = \inf\{m : \|x^{k_t} - x^m\| > \epsilon\} < \infty. \tag{1}$$

B3 There is a continuous function $W(x)$ such that

$$\limsup_{t\to\infty} W(x^{m_t}) < \lim_{t\to\infty} W(x^{k_t}) = W(x') \tag{2}$$

for any sequences $\{x^{k_t}\}, \{x^{m_t}\}$ satisfying **B2**.

B4 The set $W_\star = \{W(x^\star), x^\star \in X_\star\}$ is such that $\mathbb{R} \setminus W_\star$ is everywhere dense.

B5 If$\{x^{k_t}\} \to x^\star \in X_\star$ then $\|x^{k_t+1} - x^{k_t}\| \to 0$ when $t \to \infty$.

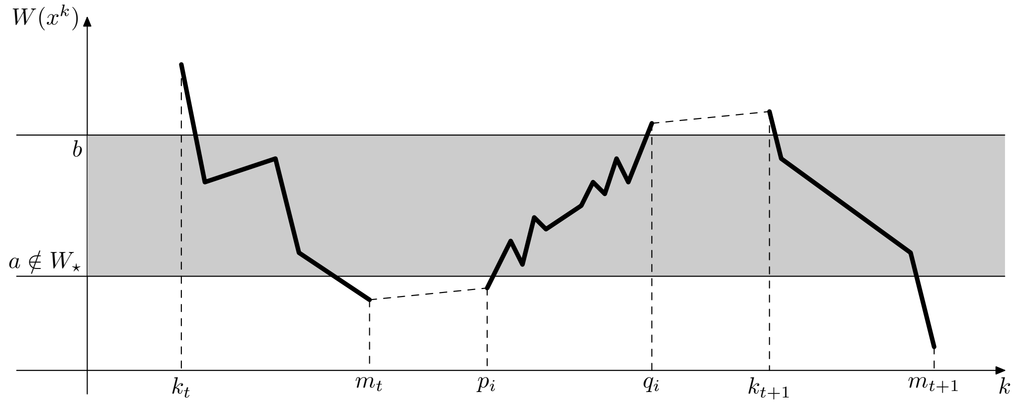

Figure 1: Subsequences involved: $x^{k_t} \to x' \notin X_\star$, $\|x^{m_t} - x^{m_t}\| > \epsilon$, $\limsup_{t\to\infty} W(x^{m_t}) < \lim_{t\to\infty} W(x^{k_t})$, $x^{p_k} \to \bar{x}' \notin X_\star$, $\|x^{p_k} - x^{q_k}\| > \epsilon$, $\limsup_{k\to\infty} W(x^{q_k}) \geq \lim_{k\to\infty} W(x^{p_k}) = W(\bar{x}')$.

Conditions **B2-B3** can be considered as a constructive way to fulfill **A2** so it follows from above that $\{x^k\}$ has at least one limit point in $X_\star$. To prove that there are no limit points out of $X_\star$ it is easy to come to contradiction.

Indeed, if there is a limit point $x' \notin X_\star$ then according to **B2-B3** there is a second limit point x'' such that $\|x'' - x'\| \geq \epsilon$ and $W(x'') < W(x')$. According to **B4** it is possible to select subinterval $[a, b] \subset (W(x''), W(x'))$ such that $a \notin W_\star$. Select from $\{x^k\}$ subsequences $\{x^{p_t}\}, \{x^{q_t}\}$ such that $p_t < q_t$, $W(x^{p_t}) \leq a$, $W(x^s) \geq a$ for $p_k < s \leq q_t$ and $W(x^{q_t}) \geq (a+b)/2$. Fig. 1 may clarify the role of the relevant subsequences.

Without loss of generality it can be assumed that $x^{p_t} \to \bar{x}$. It is clear that $\bar{x} \notin X_\star$ otherwise from $W(x^{p_t}) \leq a < W(x^{p_t+1})$ and $\|x^{p_t+1} - x^{p_t}\| \to 0$ it follows that $W(x^{p_t}) \to a = W(\bar{x})$ and $\bar{x}$ can not belong to $X_\star$ due to the special choice of a.

Define as in (1)

$$r_t = \inf\{r : \|x^{p_t} - x^r\| > \epsilon\} < \infty.$$

where ϵ is sufficiently small that $|W(\bar{x}) - W(x)\| \leq (b-a)/4$ for all $\|\bar{x} - x\| < 4\epsilon$. For t large enough $\|x^{p_t} - \bar{x}\| \leq \epsilon$ and

$$\|x^s - \bar{x}\| = \|x^{p_t} - \bar{x} - x^{p_t} + x^s\| \leq \|x^{p_t} - \bar{x}\| + \|x^{p_t} - x^s\| \leq 2\epsilon$$

for all $p_t < s < r_t$ and consequently $|W(x^s) - W(\bar{x})| \leq (b-a)/4$. Hence $W(x^s) \leq W(\bar{x}) + (b-a)/4 < a + (b-a)/2 = (a+b)/2 \leq W(x^{q_t})$ for s such that $p_t \leq s < r_t$ and therefore $r_t \leq q_t$. By construction $W(x^s) \geq a$ for $p_t < s \leq q_t$ hence $W(x^{r_t}) \geq a$ and

$$\limsup_{t\to\infty} W(x^{r_t}) \geq a \geq W(\bar{x})$$

which contradicts **B3.** This contradiction proves the assertion that all limit points of the sequence $\{x^k\}$ belong to $X_\star$. ■

The development of conditions **A1–A2** and **B1–B5** were initiated by the pioneering work of W.Zangwill [13] but was especially tailored to deal with non-monotone algorithms of nondifferentialble optimization. They and their variants were successfully applied to prove convergence of many deterministic and even stochastic algorithms. where the local nature of these conditions happened to be very useful. The ideas of this theory can also be traced down to Lyapunov conditions for continuous-time dynamical systems with (2) be the analog of negative sign of full derivative of Lyapunov function along the trajectory of system, described by ordinary differential equations. For this reason we will sometimes call $W(\cdot)$ a Lyapunov function of the process $\{x^k\}$.

Among related work the "gradient related" algorithms of Bertsecas [11] also bear some relation to **B3-B4** with $W(x) = f(x)$, but as the name suggests it requires differentiability of objective function.

2 General convergence theory

In this section two sets of results are presented — first for Fejer processes with arbitrary but fading out disturbances in their arguments of Fejer operators, second — for directed Fejer processes with diminishing disturbances taken to be attractors of some other sets. Those attractors act basically within stationary set of primary Fejer operator and finally direct the iterative process toward some specific target subsets. We begin with a few definitions related to Fejer processes. They are called by different names in the literature, so there might be a collision of definitions: quasi-nonexpansive [3], weakly Fejer [1], attractive [2], etc. However for the sake of simplicity we just call them Fejer processes with some adjectives. The above mentioned processes have the following simple recursive structure

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

where $F(\cdot)$ is a Fejer operator.

Definition 1 *Operator $F : E \to E$ is called Fejer if for any $v \in V$*

$$\|F(x) - v\| \leq \|x - v\|. \tag{4}$$

It immediately follows from the definition that any $v \in V$ is a fixed point of F. Set V is embedded into the definition of F and it is usually clear from the context what set is considered.

To ensure convergence the stronger attractivity is typically necessary and a convenient definition for the purposes of this work is as follows:

Definition 2 *F is called locally strong Fejer if for any $\bar{x} \notin X$ there exists an neighborhood U and $\alpha < 1$ such that $\|F(x) - v\| \leq \alpha\|x - v\|$ for any $v \in V$ and $x \in \bar{x} + U$.*

2.1 Arbitrary disturbances

For the purposes of further applications we consider a slight modification of (3):

$$x^{k+1} = F(x^k + z^k), \tag{5}$$

where z^k — an arbitrary (for a moment) diminishing ($z^k \to 0$) disturbance. It can be shown that the presence of this kind of disturbance does not prevent convergence of (5).

Theorem 1 *If F is a locally strong Fejer operator, $z^k \to 0$ and $\{x^k\}$ is bounded then any limit point of $\{x^k\}$ belongs to V.*

Proof. The proof consists in checking conditions **B1-B5** with the target set $X_\star = V$ and Lyapunov function $W(x) = \min_{z \in V} \|x - z\|$ — distance to V. Conditions **B4-B5** are therefore satisfied in an obvious way, and **B1** is fulfilled by the theorem assumptions. To check **B2** assume that there is a subsequence $x^{k_t} \to x' \notin V$ when $t \to \infty$. Assume that contrary to **B2** for all $\epsilon > 0$ there is t large enough that $\|x^{k_t} - x^k\| \leq \epsilon$ for all $k > k_t$. It implies that $\|x^{k_t} - x'\| \leq \epsilon$ so

$$\|x^{k_t} - x^k + x^k - x'\| \geq \|x^k - x'\| - \|x^{k_t} - x^k\| \geq \|x^k - x'\| - \epsilon,$$

therefore $\|x^k - x'\| \leq 2\epsilon$ for all $k \geq t_t$. For t large enough and $\|x^k + z^k - x'\| \leq 4\epsilon$ holds as well. As $\epsilon > 0$ can be considered as arbitrary small it ensures that $x^k + z^k \notin V + 4\epsilon B$ as well. Then

$$W(x^{k+1}) = \inf_{v \in V} \|x^{k+1} - v\| \leq \|x^{k+1} - v\| = \|F(x^k + z^k) - v\|^2$$
$$\alpha^2 \|x^k + z^k - v\|^2 \leq$$

by taking inf of the right-hand side of the last inequality obtain

$$W(x^{k+1}) \leq W(x^k) - \gamma^2 \tag{6}$$

for all $k \geq k_t$ and hence

$$W(x^N) \leq W(x^{k_t}) - \gamma^2(N - k_t) \to -\infty \tag{7}$$

when $N \to \infty$, which is an obvious contradiction and thus proves **B2**.

Condition **B3** can be proved by using the same estimate (6) but for $k_t \leq k < m_t$, where

$$m_t = \min\{m : \|x^{k_t} - m^m\| > \epsilon, m > k_t\} < \infty$$

according to already proved **B2**. Notice that for $k_t \leq k < m_t$ the distance $\|x^{k_t} - x^k\| \leq \epsilon$ and hence (7) holds and for $N = m_t$. Then

$$W(x^{m_t}) \leq W(x^{k_t}) - \gamma^2(m_t - k_t) \leq W(x^{k_t}) - \gamma^2.$$

By computing corresponding limits obtain

$$\limsup_{t\to\infty} W(x^{m_t}) < \lim_{t\to\infty} W(x^{k_t}) = W(x')$$

which proves **B3** and hence proves the theorem. ■

This result can be generalized for the case of a family of Fejer operators, that is for the sequence

$$x^{s+1} = \tilde{F}_s(x^s + z^s), \quad s = 0, 1, \ldots, \tag{8}$$

where $\tilde{F}_s$ is selected from some finite family of Fejer operators.

Theorem 2 *Let $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ is a family of operators F_i such, that for any $x \notin V$ there exists F_i, locally strong Fejer in x. If $z^s \to 0$ when $s \to \infty$, and $\tilde{F}_s = F_{i_s}$, where i_s is such that F_{i_s} — locally strong Fejer in x^s. Then if the sequence $\{x^s\}$ is bounded, all its limit points belong to V.*

Proof. The proof of this theorem is analogous to that of the theorem 1, as we have for $\tilde{F}_s$ and any $x^s \in x' + U, x' \notin V$ the uniform estimate

$$\|\tilde{F}_s(x^s) - v\| = \|F_{i_s}(x^s) - v\| \leq \max_{i \in I_s} \alpha_i \|x^s - v\| = \alpha\|x^s - v\|,$$

with $\alpha < 1$, where I_s is a set of locally strong Fejer at the point x^s operators F_i with corresponding constants $\alpha_i < 1, i \in I_s$. ■

2.2 Directed Fejer processes

By the special choice of disturbances z^s the sequence $\{x^s\}$ can be directed toward selected subset of V in both cases (5) and (8). We denote such subset as $Z \subset V$. For that we define the notion of restricted attractant as set-valued mapping of a space of variables into itself with a special property that lousily speaking it is directed toward some $Z \subset V$ at $x \in V$. More precisely,

Definition 3 *Set-valued map $\Phi : V \to E$ is called a locally restricted attractant of $Z \subset V$ if $g(z-x) \geq 0$ for all $x \in V \setminus Z, g \in \Phi(x) \quad z \in Z$.*

In fact as for Fejer operators we need a stronger definition:

Definition 4 *A restricted attractant Φ is called strong restricted attractant of Z if for each $x' \in V \setminus Z$ there exists a neighborhood of zero U such that,*

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

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

Of course in the definition above the neighborhood U and the constant δ depend on x', what we did not write down explicitely to simplify notation. The following theorem holds.

Theorem 3 *Let F is locally strong Fejer, Φ — locally strong restricted attractant $Z \subset V$, upper-semicontiniuos on some open superset $\tilde{V} \supset V$ of V, and a sequence $\{x^s\}$ is generated by*

$$x^{s+1} = F(x^s + \lambda_s \Phi(x^s)), \tag{9}$$

with arbitrary x^0 and $\lambda_s \to +0, \sum \lambda_s = \infty$. Then if $\{x^s\}$ bounded, any limit point of $\{x^s\}$ belongs to Z.

Proof.. To proof the theorem the same **B1**–**B5** conditions can be applied with $W(x) = \inf_{v \in Z} \|x - v\|$. As **B1** is included into the theorem assumption, and **B4**–**B5** are trivially fulfilled we need to check **B2**–**B3** only. We start with **B2**: let $\{x^{n_k}\}$ is a certain subsequence of $\{x^s\}$, converging to $x' \notin Z$. As shown above independently of attracting properties of $D(\cdot)$ the point x' at least belongs to V. Then there exists $\epsilon, \delta > 0$ such that $d(z - x) \geq \delta > 0$ for all $z \in Z, x \in x' + 4\epsilon B$ and $d \in D(x)$.

By upper-semicontinuity of $D(\cdot)$ sets $D(x)$ are uniformly bounded for $x \in x' + 4\epsilon B$ that is $\|d\| \leq C < \infty$ for all $d \in D(x), x \in x' + 4\epsilon B$ therefore as $\lambda_s \to 0$ then $\lambda_s d^s \to 0$ as well. As for sufficiently large k $x^{n_k} \in x' + \epsilon B$ and if for all $s > n_k$ points $\{x^s\}_{n_k}^S \subset x^{n_k} + \epsilon B$, then $\{x^s\}_{n_k}^S \subset x' + 2\epsilon B$ and $\{x^s + \lambda_s d_s\}_{n_k}^S \subset x' + 4\epsilon B$

Hence for $s \geq n_k$

$$\|x^{s+1} - z\|^2 = \|F(x^s + \lambda_s d^s) - z\|^2 \leq \|x^s + \lambda_s d^s - z\|^2 = \|x^s - z\|^2 + \lambda_s^2 \|d^s\|^2 - 2\lambda_s d^s (z - x^s) \leq$$
$$\|x^s - z\|^2 + \lambda_s^2 C^2 - 2\lambda_s \delta = \|x^s - z\|^2 - \lambda_s(2\delta - \lambda_s C^2) \leq \|x^s - z\|^2 - \delta\lambda_s$$

for sufficiently large n_k. It follows from here that $\|x^{s+1} - z\| \leq \|x^s - z\| - \delta\lambda_s/2$ and by computing infinums obtain

$$W(x^{s+1}) \leq W(x^s) - \delta\lambda_s/2. \tag{10}$$

Summing (10) leads to contradiction

$$0 \geq W(x^S) \leq W(x^{n_k}) - \delta/2 \sum_{s=n_k}^{S} \lambda_s \to -\infty$$

and hence proves **B2.** Now it is easy to demonstrate that the condition **B3** holds as well. If $m_k > n_k$ is defined as in **B3** as

$$\|x^{n_k} - x^{m_k}\| > \epsilon, \|x^{n_k} - x^s\| < \epsilon$$

for $n_k \leq s < m_k$ then $x^{m_k - 1} \in x' + 2\epsilon B$ and therefore the estimate (10) holds for $s = m_k - 1$. Hence

$$W(x^{m_k}) \leq W(x^{n_k}) \leq \delta/2 \sum_{s=n_k}^{m_k - 1} \lambda_s \tag{11}$$

where $\sum_{s=n_k}^{m_k-1} \lambda_s$ has strictly positive estimate from below:

$$\epsilon < \|x^{m_k} - x^{n_k}\| \leq \sum_{s=n_k}^{m_k-1} \lambda_s \|d^s\| \leq C \sum_{s=n_k}^{m_k-1} \lambda_s$$

and consequently $\sum_{s=n_k}^{m_k-1} \lambda_s > \epsilon/C$. After substitution of this estimate in (11) and taking corresponding limits obtain

$$\limsup_{k\to\infty} W(x^{m_k}) \leq \lim_{k\to\infty} W(x^{n_k}) - \epsilon/C < W(x')$$

which completes the proof. ■

This theorem can also be generalized for the case of a finite family of Fejer operators $\mathcal{F} = \{F_i, i = 1, 2, \ldots, N\}$, each of which is locally strong at some superset X_i of V and intersection of X_i coincides with V.

Theorem 4 *Let $\mathcal{F} = \{F_i, i = 1, 2, \ldots, N\}$ is a finite family of locally strong on correspondent X_i continuous Fejer operators, and for any $x \notin V$ there exists $i(x) \in \{1, 2, \ldots, n\}$ such that $F_{i(x)}$ is locally strong at x. Then the combined process*

$$x^{s+1} = F_{i(x^s)}(x^s + \lambda_s d^s), \; d^k \in D(x^k) \tag{12}$$

if bounded converges to the set Z if $\lambda_s \to +0, \sum \lambda_s = \infty$ and $D(\cdot)$ is some locally strong restricted attractant of $Z \subset V$.

The proof of this theorem is basically the same as of theorem 3 with some minor technical changes.

3 Subgradient projection decomposition algorithms

As an immediate application of the results obtain above we can develop subgradient projection decomposition algorithms for constrained convex nondifferentiable problem

$$\min f(x), x \in C, \tag{13}$$

when convex set C can be factored into the intersection of another convex sets: $C = \cap_{i=1}^m C_i$. It is quite common case in practice and very often C_i are "simple" sets — hemispaces, balls, boxes, hyperplanes etc. For such sets it is easy to develop for C_i corresponding locally strong Fejer operators F_i and it is clear that $\mathcal{F} = \{F_i, i = 1, 2, \ldots, m\}$ will provides us with the right family of operators to apply theorem 4.

Most often Fejer operators are constructed with the help of projection operation $\|\Pi_X(x) - x\| = \min_{v \in X} \|v - x\|$, the popular choice is for instance

$$F(x) = x + \rho(\Pi_X(x) - x), \tag{14}$$

where $\rho \in (0, 2)$ is a relaxation parameter. It is easy to show that in this case F is a locally strong Fejer operator with respect to X and therefore the family of $F_i = x + \rho_i(\Pi_{C_i}(x) - x)$ will be the right collection of operators to apply the theorem 4. The natural choice for locally strong restricted attractant for solution set Z of (13) is the anti-subgradient $D(x) = -\partial f(x)$ — it satisfies all requirements of the theorem 4 for "normal", i.e. finite convex functions f.

Applying the algorithmic schema of the theorem 4 with $F_i = \Pi_i$ obtain for instance subgradient projection decomposition algorithm

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

where Π_s projection on the set C_{i_s} such that $x^s \notin C_{i_s}$ and one can use in this algorithm the relaxed projections (14). As such this algorithm reminds very much the sequential projection algorithm of Motzkin-Schoenberg-Agmon for CFP however the important difference is that here we are solving optimization problem (13). It is worth to mention here the recent work [3] were the similar algorithm was suggested, however under much more restricted conditions as with respect to objective function as to the stepsizes λ_s.

It can also be shown that under not very restrictive assumptions the simultenious projection operator

$$F(x) = \sum_{i=1}^{m} w_i \Pi_i(x)$$

with positive weights $w_i > 0, \sum_{i=1}^m w_i = 1$ is also locally strong Fejer and therefore can be used to develop simultenious projection algorithms with concurrent projections.

4 Envelope step-size control

As numerical experiments and theoretical analysis show step-size rule used in the theorem 3 results in slow convergence and there were many attempts to speed it up. Here we present simple and rather general idea for step-size control in gradient-like methods which, as preliminary tests demonstrate gives much better results. For simplicity we consider iterative processes

$$x^{k+1} = x^k - \lambda_k d^k,\ d^k \in D(x^k), \tag{15}$$

where $D(x)$ is a set-valued mapping whose properties will be specified later. To simplify notations let $D(p,q) = \mathrm{co}\,\{d^p, d^{p+1}, \dots, d^q\}$.

For a given sequence $\theta_m \to +0, m = 0, 1, \dots$ determine corresponding sequences of indeces $\{k_m\}$ and numbers $\{\lambda_k\}$ by the following recursive relationships:

- Set $k_0 = 0$ and pick up initial $\lambda_0 > 0$. Let $q \in (0,1)$.
- For given m and k_m determine k_{m+1} as the index which satisfies conditions

$$0 \notin D(k_m, s) + \theta_m B, k_m \leq k < k_{m+1}, \quad 0 \in D(k_m, k_{m+1}) + \theta_m B \tag{16}$$

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

$$\lambda_{k_{m+1}} = q\lambda_{k_m}. \tag{17}$$

In other words, condition (16) detects the first instance when $\{x^k\}$ seems to be cycling, λ_k is kept constant between k_m and k_{m+1}, and according to (17) at $k = k_{m+1}$ it is multiplied by q. This idea — to keep stepsize constant whilst we seems to be moving in the right direction and decrease it othewise, is by no means new and can be traced back as far as Armijo [14]. Recently as a certain heuristic to improve current approximate solution it was propagated in [15] with successful applications in image processing in tomography.

4.1 Convergence

The following theorem establishes convergence of the process (15) with stepsize rules (16) - (17).

Theorem 5 *Let $D(x)$ is convex-valued, locally bounded upper-semicontiniuos set-valued locally strong attractant of $X_\star$. Then if the sequence $\{x^k\}$ is bounded then all its limit points belong to $X_\star$.*

Proof. It is easy to show that as in this case there is no Fejer operator, or it can be assumed to be an identity, $X_\star = \{x^\star : 0 \in D(x^\star)\}$. Notice that $d^k = 0$ for some k and further on represents a trivial case, so assume that $d^k \neq 0$ for all k. Next we establish that the sequence $\{k_m\}$, defined by (16) is infinite, or, in other words, λ_k is decreased in accordance with (17) infinitely many times.

Indeed if λ_k is decreased only finite number of times, then there is M such that for all $k > k_M$

$$0 \notin D(k_M + 1, k) + 2\delta_M B = D_k + 2\delta_M B$$

for some $\delta_M > 0$. By monotonicy D_k with respect to inclusion there is a Kuratowski limit $\lim_{k\to\infty} D_k = \tilde{D}$ with $0 \notin \bar{G} + \delta_M B$, where $\bar{D}$ — closure of $\tilde{D}$. Sets $\tilde{D}$ and $\bar{D}$ are of course convex, and therefore it exists $v \in \bar{D}$ such that

$$v\bar{d} \geq \|v\|^2$$

for all $\bar{d} \in \bar{D} + \delta_M B$. By representing $\bar{d}$ as $d + \delta_M z, d \in \bar{D}, z \in B$, obtain

$$vd \geq \|v\|^2 - \delta_M vz,$$

for all $z \in U$. After taking supremum of the right-hand side with respect to $z \in B$ obtain

$$vd \geq \|v\|^2 + \delta_M \|v\| > \delta_M \|v\| > 0$$

or

$$\bar{v}d > \delta_M, \tag{18}$$

where $\bar{v} = v/\|v\|$. As $\bar{D} \supset D_k$ for all $k \geq k_M$ the inequality (18) holds for all $d^k, k > k_M$ and hence

$$\|x^{k_M} - x^K\| \geq (x^{k_M} - x^K)\bar{v} = \sum_{k=k_M}^{K-1} \lambda_M d^k \bar{v} \geq \lambda_M \gamma (K - 1 - k_M)\delta_M \to \infty$$

when $K \to \infty$, which contradicts the boundness of $\{x^k\}$. and this contradiction proves that $\lambda_k \to 0$ and also $\|x^{k+1} - x^k\| \to 0$ and hence **B4** is fulfilled.

In what follows we show that **B2** is fulfilled as well.

Assume that $\{x^{n_k}\}$ — a certain subsequence, which converges to $x' \notin V$. Then $0 \notin D(x')$ and by upper-semi-continuity of $D(\cdot)$ there exists $\epsilon, \delta > 0$ such that

$$0 \notin \text{co}\,\{D(x), \|x' - x\| \leq 4\epsilon\} + \delta B. \tag{19}$$

Consider n_k large enough that $\|x^{n_m} - x'\| \leq \epsilon$ for $m \geq k$. Without loss of generality we can assume that the corresponding $2\theta_m < \delta$. Then if $\{x^l\}$ remains in the 4ϵ-neighborhood of x' for $l > n_k$ then not more then one change in value of λ_k can occur. In other words among indeces l such that $l \geq n_k$ and $\|x^l - x'\| \leq \epsilon$ there is not more then one $l \in \{k_m\}$. In fact, if there were 2 such indeces k'_m and $k_{m'+1}$ it would contradict (19):

$$0 \in D(k_{m'}, k_{m'+1}) \subset \text{co}\,\{D(x), \|x' - x\| \leq 4\epsilon\} + \delta B.$$

Therefore the assumption that $\{x^l, l \geq n_k\} \in \{x : \|x - x'\| \leq 4\epsilon\}$ contradicts the infiniteness of $\{k_m\}$ and hence for all k there exist $m_k \geq n_k$ such that

$$\|x^{m_k} - x'\| > \epsilon, \quad \|x^l - x'\| \leq \epsilon \text{ for } n_k \leq l < m_k.$$

Finally we show that **B3** is fulfilled as well. Assume as above then the sequence $\{x^k\}$ has a subsequence $\{x^{n_k}\} \to x' \notin V$. Then, according to **B2** for any $\epsilon > 0$ small enough there exists $\{x^{m_k}\}$ such that,

$$\|x^{n_k} - x^s\| \leq \epsilon, n_k < s \leq m_k,$$

Estimate $W(x^{m_k}) = \min_{x^\star \in X_\star} \|x^{m_k} - x^\star\|$:

$$\begin{gathered} W(x^{m_k}) \leq \|x^{m_k} - x^\star\|^2 = \|x^{m_k} - x^{n_k} + x^{n_k} - x^\star\|^2 = \\ \|x^{n_k} - x^\star\|^2 + 2(x^{n_k} - x^\star)(x^{m_k} - x^{n_k}) + \|x^{m_k} - x^{n_k}\|^2 \leq \\ \|x^{n_k} - x^\star\|^2 + 2(x^{n_k} - x^\star)(x^{m_k} - x^{n_k}) + \epsilon^2 \leq \end{gathered}$$

for any $x^\star \in X_\star$. Taking into account that

$$x^{m_k} - x^{n_k} = \sum_{s=n_k}^{m_k-1} (x^{s+1} - x^s) = \sum_{s=n_k}^{m_k-1} \lambda_s d^s,$$

obtain

$$\begin{gathered} W(x^{m_k} \leq \|x^{n_k} - x^\star\|^2 + 2\sum_{s=n_k}^{m_k-1} (x^{n_k} - x^\star)\lambda_s d^s + \epsilon^2 = \\ \|x^{n_k} - x^\star\|^2 + 2\sum_{s=n_k}^{m_k-1} (x^{n_k} - x^s + x^s - x^\star)\lambda_s d^s + \epsilon^2 \leq \\ \|x^{n_k} - x^\star\|^2 + 2\sum_{s=n_k}^{m_k-1} (x^s - x^\star)\lambda_s d^s + 2\sum_{s=n_k}^{m_k-1} \lambda_s \|x^{n_k} - x^s\| \|d^s\| + \epsilon^2 \leq \\ \|x^{n_k} - x^\star\|^2 + 2\sum_{s=n_k}^{m_k-1} (x^s - x^\star)\lambda_s d^s + 2\epsilon D \sum_{s=n_k}^{m_k-1} \lambda_s + \epsilon^2. \end{gathered}$$

As $D(\cdot)$ is an attractant $(x^s - x^\star)\lambda_s d^s \leq -\gamma$ for some $\gamma > 0$ and hence

$$W(x^{m_k} \leq \|x^{n_k} - x^\star\|^2 - 2\gamma \sum_{s=n_k}^{m_k-1} \lambda_s + 2\epsilon D \sum_{s=n_k}^{m_k-1} \lambda_s + \epsilon^2 \leq$$

Assuming $\epsilon < \gamma/(2D)$ the last inequality can be strengthen to

$$W(x^{m_k} \leq \|x^{n_k} - x^\star\|^2 - \gamma \sum_{s=n_k}^{m_k-1} \lambda_s + \epsilon^2. \tag{20}$$

The sum $\sum_{s=n_k}^{m_k-1} \lambda_s$ can be estimated from below:

$$\epsilon < \|x^{m_k} - x^{n_k}\| \leq \sum_{s=n_k}^{m_k-1} \lambda_s d^s\| \leq D \sum_{s=n_k}^{m_k-1} \lambda_s$$

and after substitution that in (20) obtain

$$W(x^{m_k} \leq \|x^{n_k} - x^\star\|^2 - \gamma\epsilon/D + \epsilon^2 \leq \|x^{n_k} - x^\star\|^2 - \gamma\epsilon/(2D)$$

for arbitrary $x^\star \in V$. Computing infinum of the right-hand side with respect to $x^\star \in X_\star$ yields:

$$W(x^{m_k}) \leq W(x^{n_k}) - \gamma\epsilon/(2D).$$

Passing to the limit when $k \to \infty$ produces

$$\limsup_{k\to\infty} W(x^{m_k}) \leq \lim_{k\to\infty} W(x^{n_k}) - \gamma\epsilon/(2D) = W(x') - \gamma\epsilon/(2D) < W(x'),$$

which proves **B3** and hence the whole theorem. ■

A few words about computational issues related to the practical use of ESC. To apply this step-size rule we need to perform repetitive checks of inclusion

$$0 \in D(k_m, k) + \delta_m B,\ k = k_m + 1, k_m + 2, \ldots. \tag{21}$$

A natural way to check (21) is to solve least norm problem for the set $D(k_m, k)$ and compare its result with δ_m. This is nontrivial problem, however the algorithm [7] demonstrated in our tests quite adequate performance. Much can also be gained from the incremental growth of the set $D(k_m, k)$, when least norm solution on the previous iteration can be used as a good starting point for the next which can be easily incorporated in the algorithm [7].

4.2 Illustrative example

It is interesting to consider even tiny illustrative example of use of envelope step-size control (ESC). Let us apply subgradient method for minimization of piece-wise linear function

$$f(x) = \max\{y_i, i = 1, 2, 3\}, (y_1, y_2, y_3) = y = Ax$$

of 2-dimensional vector $x = (x_1, x_2)$ defined by the matrix

$$A = \begin{vmatrix} 1 & 0 \\ -0.5 & -0.4 \\ -3 & 0.2 \end{vmatrix}.$$

The origin $x^\star = (0, 0)$ with $f(x^\star) = 0$ is a trivial solution of this problem. The algorithm

$$x^{k+1} = x^k - \lambda_k g^k, g^k \in \partial f(x^k), k = 0, 1, \ldots$$

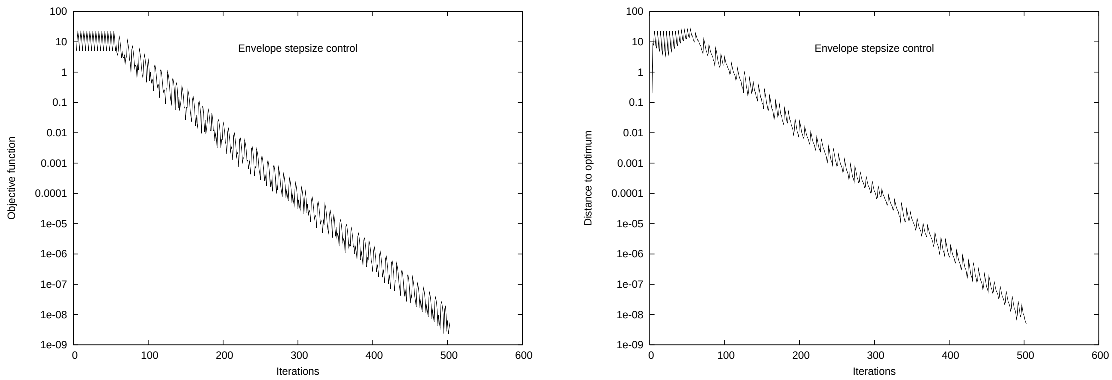

Figure 2: Subgradient algorithm with ESC. Step-size multiplier 0.5. Objective function (left) and distance to the optimum (right).

was applied with the starting point $x^0 = (5, 7)$ and rather large initial step-size $\lambda = \|x^0\| = 8.6$. The results of application of ESC are shown on the Fig. 2 where the linear rate of convergence can be seen "in general". At the same time it is seen that the algorithm is not monotone neither in terms of objective function nor in terms of traditional indicator of convex optimization — euclidian distance to the optimum. It makes clear that the proof of convergence for such algorithm can be obtained only by non-traditional arguments like conditions **B1-B5** and also the very notion of the rate of convergence should be modified to cover such cases.

Despite extremal simplicity of this test it is interesting to compare results, obtain by ESC with theoretical recommendations [9] for another stepsize rule which guarantees linear rate of convergence. In this work subgradient method with normalized subgradient

$$x^{k+1} = x^k - h_k g^k / \|g^k\|, k = 0, 1, \dots, \quad g^k \in \partial f(x^k),$$

is considered and the usage of $h_{k+1} = h_k \sin(\phi), h_0 \geq \|x^0 - x^\star\| \cos(\phi)$ is proposed. The angle ϕ is defined from the condition that for $g \in \partial f(x), x \neq x^\star$ the following inequality holds

$$g(x - x^\star) \geq \cos(\phi) \|g\| \|x - x^\star\| \tag{22}$$

with $\phi \in [\pi/4, \pi/2)$, which determines the stepsize multiplicator $q_f = \sin(\phi)$. For given function the maximal angle ϕ is determined by subdifferential of this function on the line $y_1 = y_3$ and is equal to the convex combination of the first and the third rows of A. By direct computation for (22), obtain $\cos(\phi) = 0.016609$, what gives $q_f = 0.99986$. After the same 500 iterations the initial stepsize will be decreased to only $0.93335 h_0$ which in fact means no practical convergence occurs.

Conclusions

In this paper we developed a theoretical framework for combining Fejer processes and other problem-specific attracting mappings to create new algorithms with new opportunities for decomposition and parallel computations. It is shown that this can be done with scaling down the impact of the later with diminishing scale factors. To make resulting processes still converge to solution sets the classical diverging series condition is sufficient. As an example result a subgradient projection algorithm can be decomposed and parallelized with respect to groups of constrains. To speed up convergence the envelope step-size control schema is suggested for which an illustrative example exhibits a linear-like convergence.

References

[1] Vasin V.V., Eremin I.I. Operators and Fejer Iterative Processes: Theory and Applications (Ural. Otd. Ross. Acad. Nauk, Yekaterinburg, 2005) [in Russian].

[2] Bauschke H.H., Borwein J.M. On Projection Algorithms for Solving Convex Feasibility Problems, SIAM Revs. 1996. V. 38(3), P. 367-426.

[3] Hirstoaga S.A., Iterative selection methods for common fixed point problems, J. Math. Anal. Appl., 2006. V. 324, P. 1020-1035.

[4] Y. Censor, M. Jiang and A.K. Louis (Editors), Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT), Edizioni della Normale, Pisa, Italy, 2008.

[5] Nurminski E.A. Fejer Processes with Diminishing Disturbances, Doklady Mathematics, 2008, Vol. 78, No. 2., pp. 1-4.

[6] Nurminski E.A. Numerical methods of convex optimization Moscow: Nauka.- 1991.-168 P. [in Russian]

[7] Nurminski E.A Convergence of the Suitable Affine Subspace Method for Finding the Least Distance to a Simplex, Computational Mathematics and Mathematical Physics, Vol. 45 No. 11, 2005, pp. 1915-1922.

[8] Nurminski E.A. The use of additional small disturbances in Fejer models of iterative algorithms // Zhurn. Vychisl. Mathem. Matem. Physiki, 2008, v. 48, issue 12, pp. 2121-2128 [In Russian].

[9] Shor N.Z., Gamburd P.R. On convergence of generalized gradient descent method, Kibernetika, 1971, 6, 82-84.

[10] Polyak B.T. Minimization of un-smooth functionals, USSR Computational Mathematics and Mathematical Physics, 1969, v.9, pp. 14-29.

[11] Nedi'c A., Bertsekas D.P. Incremental subgradient methods for non-differentiable optimization, SIAM Journal on Optimization, 2001, v. 12, pp.109-138.

[12] Mijangos E. Approximate subgradient methods for nonlinearly constrained network flow problems, Journal of Optimization Theory and Applications, 2006, 128(1), 166-190.

[13] Zangwill W.I. Convergence conditions for nonlinear programming algorithms. Management Science, 1969, 16(1), 1-13.

[14] Armijo L. Minimization of functions having continious partial derivatives // Pacific J. Math., 1966, v. 16, pp. 1-3.

[15] Butnariu D., Davidi R., Herman G.T., Kazantsev I.G. Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems, IEEE J. Sel. Top. Sign. Process. 2007, 1, 540-546.