FEJÉR MONOTONICITY IN CONVEX OPTIMIZATION

Let S be a nonempty closed and convex set in a real Hilbert space $\mathcal{H}$ with norm $\|\cdot\|$. A sequence $(x_n)_{n\geq 0}$ of points in $\mathcal{H}$ is said to be *Fejér monotone* with respect to S (or simply *S-Fejérian*) if

$$(\forall \overline{x} \in S)(\forall n \in \mathbf{N}) \ \|x_{n+1} - \overline{x}\| \leq \|x_n - \overline{x}\| \,. \tag{1}$$

In words, each point in the sequence is not further from any point in S than its predecessor. Given $x_0 \in \mathcal{H}$, a typical example of S-Fejérian sequence is that generated by the algorithm

$$(\forall n \in \mathbf{N}) \ x_{n+1} = Tx_n,$$

where $T\colon \mathcal{H} \to \mathcal{H}$ is a nonexpansive operator, i.e.,

$$(\forall (x,y) \in \mathcal{H}^2) \ \|Tx - Ty\| \leq \|x - y\| \,, \tag{2}$$

with nonempty fixed point set S. Under suitable assumptions, the sequence of successive approximations $(x_n)_{n\geq 0}$ converges to a point in S [20].

In convex optimization, one frequently encounters algorithms whose orbits $(x_n)_{n\geq 0}$ are Fejér monotone with respect to the solution set. In order to simplify and standardize the convergence proofs of this broad class of algorithms, it is important to investigate the notion of Fejér monotonicity and to bring out some general convergence principles. These are precisely the objectives of the present article.

Notation and assumptions. Throughout, the sequence $(x_n)_{n\geq 0}$ is Fejér monotone with respect to a nonempty closed and convex set S in a real Hilbert space $\mathcal{H}$ with scalar product $\langle \cdot \mid \cdot \rangle$, norm $\|\cdot\|$, and distance d. For every $n \in \mathbf{N}$, p_n denotes be the projection of x_n onto S, i.e., the unique point $p_n \in S$ such that $\|x_n - p_n\| = d(x_n, S)$. Recall that p_n is characterized by the variational inequality

$$(\forall \overline{x} \in S) \ \langle \overline{x} - p_n \mid x_n - p_n \rangle \leq 0. \tag{3}$$

The expressions $x_n \rightharpoonup x$ and $x_n \to x$ denote respectively the weak and strong convergence of $(x_n)_{n\geq 0}$ to x. $\mathfrak{W}$ and $\mathfrak{S}$ denotes respectively the sets of weak and strong cluster points of $(x_n)_{n\geq 0}$. Finally, Id denotes the identity operator on $\mathcal{H}$.

Fejér monotone $\to$ Fejér monotone sequence
Fejérian$\to$ Fejérian$\to$ Fejérian

Basic Convergence Properties. By way of preamble, some immediate consequences of (1) are stated below.

PROPOSITION 1 The following assertions hold.

i) $(x_n)_{n\geq 0}$ is bounded.
ii) $(\forall \overline{x} \in S)$ $(\|x_n - \overline{x}\|)_{n\geq 0}$ converges.
iii) $(d(x_n, S))_{n\geq 0}$ is nonincreasing.
iv) $(\forall \overline{x} \in S)$ $x_n \to \overline{x} \Leftrightarrow \underline{\lim} \|x_n - \overline{x}\| = 0 \Leftrightarrow S \cap \mathfrak{S} \neq \emptyset$.

□

Weak convergence. In general, Fejér monotone sequences do not converge, even weakly (consider for instance the $\{0\}$-Fejérian sequence $((-1)^n x_0)_{n\geq 0}$ with $x_0 \neq 0$). By virtue of Proposition 1i), $\mathfrak{W} \neq \emptyset$ and a necessary condition for $(x_n)_{n\geq 0}$ to converge weakly to a point in S is $\mathfrak{W} \subset S$. A remarkable consequence of Fejér monotonicity is that this condition is also sufficient. To see this, take y_1 and y_2 in $\mathfrak{W}$, say $x_{k_n} \rightharpoonup y_1$ and $x_{l_n} \rightharpoonup y_2$, and $\overline{x} \in S$. By Proposition 1ii),

$$\lim \|x_{k_n} - \overline{x}\|^2 = \lim \|x_{l_n} - \overline{x}\|^2 \,.$$

Therefore, by expanding,

$$\lim \|x_{k_n}\|^2 - \lim \|x_{l_n}\|^2 = 2\langle \overline{x} \mid y_1 - y_2 \rangle.$$

It follows that

$$S \subset \{x \in \mathcal{H} : \langle x \mid y_1 - y_2 \rangle = \alpha\} \,, \tag{4}$$

where $\alpha = (\lim \|x_{k_n}\|^2 - \lim \|x_{l_n}\|^2)/2$. Thus, $(y_1, y_2) \in S^2 \Rightarrow \alpha = \langle y_1 \mid y_1 - y_2 \rangle = \langle y_2 \mid y_1 - y_2 \rangle \Rightarrow y_1 = y_2$. Consequently, the bounded sequence $(x_n)_{n\geq 0}$ cannot have more than one weak cluster point in S. This fundamental property will be recorded as:

PROPOSITION 2 $(x_n)_{n\geq 0}$ converges weakly to a point in S if and only if $\mathfrak{W} \subset S$. □

Two additional properties are worth mentioning in connection with weak convergence.

- Let $\overline{\operatorname{aff}}\, S$ be the closed affine hull of S. If $y_1 \neq y_2$, then (4) asserts that S is contained in a closed affine hyperlane. If $\overline{\operatorname{aff}}\, S = \mathcal{H}$, $\mathfrak{W}$ reduces to a singleton and $(x_n)_{n\geq 0}$ therefore converges weakly.

- Suppose that $x_n \rightharpoonup \overline{x} \in S$ and let $x \in \mathcal{H}$. Then the identities

$$(\forall n \in \mathbf{N}) \quad \|x_n - x\|^2 = \|x_n - \overline{x}\|^2 + 2\langle x_n - \overline{x} \mid \overline{x} - x\rangle + \|\overline{x} - x\|^2$$

together with Proposition 1ii) imply that $(\|x_n - x\|)_{n\geq 0}$ converges.

Strong convergence. As evidenced by the classical counterexample of [13], $x_n \rightharpoonup \overline{x} \in S \not\Rightarrow x_n \to \overline{x} \in S$. Accordingly, strong convergence conditions for Fejér monotone sequences must be identified.

First, consider the projected sequence $(p_n)_{n\geq 0}$. It follows from (1) and (3) that for every $(m,n) \in \mathbf{N}^2$

$$\begin{aligned}\|p_n - p_{n+m}\|^2 &= \|p_n - x_{n+m}\|^2 \\ &\quad + 2\langle\, p_n - x_{n+m} \mid x_{n+m} - p_{n+m}\rangle \\ &\quad + \|x_{n+m} - p_{n+m}\|^2 \\ &\leq d(x_n,S)^2 - d(x_{n+m},S)^2 \\ &\quad + 2\langle\, p_n - p_{n+m} \mid x_{n+m} - p_{n+m}\rangle \\ &\leq d(x_n,S)^2 - d(x_{n+m},S)^2.\end{aligned}$$

Consequently, since $(d(x_n,S))_{n\geq 0}$ converges by Proposition 1iii), $(p_n)_{n\geq 0}$ is a Cauchy sequence. This establishes:

Proposition 3 $(p_n)_{n\geq 0}$ converges strongly. □

This result, which is of interest in its own right, also leads to a simple criterion for the strong convergence of $(x_n)_{n\geq 0}$ to a point in S. Indeed, suppose that $\underline{\lim}\, d(x_n,S) = 0$. Then, thanks to Proposition 1iii), $d(x_n,S) \to 0$, i.e., $x_n - p_n \to 0$. On the other hand, by Proposition 3, $p_n \to \overline{x}$ with $\overline{x} \in S$ since S is closed. One thus obtains:

Proposition 4 $(x_n)_{n\geq 0}$ converges strongly to a point in S if and only if $\underline{\lim}\, d(x_n,S) = 0$. □

Going back to (4), assume now that $(y_1,y_2) \in \mathfrak{S}^2$. Then $\alpha = (\|y_1\|^2 - \|y_2\|^2)/2$ and (4) therefore becomes

$$\begin{aligned}S &\subset \left\{x \in \mathcal{H} : \left\langle x - \frac{y_1+y_2}{2} \;\middle|\; y_1 - y_2\right\rangle = 0\right\} \\ &= \{x \in \mathcal{H} : \|x - y_1\| = \|x - y_2\|\}\,.\end{aligned} \tag{5}$$

In words, if $(x_n)_{n\geq 0}$ possesses two distinct strong cluster points y_1 and y_2, S is contained in the closed affine hyperplane whose elements are equidistant from y_1 and y_2. If $\overline{\operatorname{aff}}\, S = \mathcal{H}$, it results from (5) that $(x_n)_{n\geq 0}$ possesses at most one strong cluster point. This happens in particular when the interior of S is nonempty (Slater condition). In this case, however, a sharper result holds, namely $(x_n)_{n\geq 0}$ converges strongly [22].

Linear convergence. Proposition 1iii) asserts that $(d(x_n,S))_{n\geq 0}$ is nonincreasing. Assume now that it decreases at a linear rate, say

$$(\exists\, \kappa \in]0,1[)(\forall n \in \mathbf{N})\; d(x_{n+1},S) \leq \kappa d(x_n,S). \tag{6}$$

Then, in view of Proposition 4, $x_n \to \overline{x} \in S$. On the other hand, for every $(m,n) \in \mathbf{N}^2$, (1) yields

$$\begin{aligned}\|x_n - x_{n+m}\| &\leq \|x_n - p_n\| + \|x_{n+m} - p_n\| \\ &\leq 2d(x_n,S).\end{aligned}$$

Thus $\|x_n - \overline{x}\| \leq 2d(x_n,S)$ and one reaches the following conclusion.

Proposition 5 Suppose that (6) holds. Then $(x_n)_{n\geq 0}$ converges linearly to a point $\overline{x} \in S$: $(\forall n \in \mathbf{N})\ \|x_n - \overline{x}\| \leq 2\kappa^n d(x_0,S)$. □

Geometric Construction. In order to make the above theoretical convergence results more readily applicable in concrete problems, it will henceforth be assumed that $(x_n)_{n\geq 0}$ has been generated by the following algorithm.

Algorithm 1: General Fejérian scheme
0. Take $x_0 \in \mathcal{H}$ and set $n = 0$.
1. Generate a closed affine half-space H_n such that $S \subset H_n$.
2. Compute the projection $P_n x_n$ of x_n onto H_n and take $\lambda_n \in [0,2]$.
3. Set $x_{n+1} = x_n + \lambda_n\,(P_n x_n - x_n)$.
4. Set $n = n+1$ and go to step 1.

The relaxation parameter λ_n determines the position of the update x_{n+1} on the closed segment between the current iterate x_n and its reflection $r_n = 2P_n x_n - x_n$ with respect to H_n (see Fig. 1). In some problems, it is possible to significantly accelerate the progression of the iterates towards a solution by proper choice of the relaxation sequence $(\lambda_n)_{n\geq 0}$ [5].

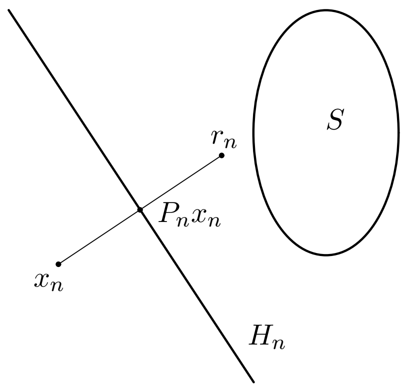

Fig. 1: A Fejérian iteration.

Hereafter, two properties of the relaxation sequence will be considered, namely

$$\sum_{n\geq 0} \lambda_n(2-\lambda_n) = +\infty \tag{7}$$

and

$$(\lambda_n)_{n\geq 0} \text{ lies in } [\varepsilon, 2-\varepsilon], \text{ where } \varepsilon \in]0,1[. \tag{8}$$

Now fix $\overline{x} \in S$. Then, for every $n \in \mathbf{N}$,

$$\begin{aligned} \|x_{n+1} - \overline{x}\|^2 &= \|x_n - \overline{x}\|^2 + \lambda_n^2 \|P_n x_n - x_n\|^2 \\ &\quad +2\lambda_n \langle x_n - \overline{x} \mid P_n x_n - x_n \rangle \\ &\leq \|x_n - \overline{x}\|^2 \\ &\quad -\lambda_n(2-\lambda_n) d(x_n, H_n)^2. \end{aligned} \tag{9}$$

Consequently, $(x_n)_{n\geq 0}$ is S-Fejérian and

$$\sum_{n\geq 0} \lambda_n(2-\lambda_n) d(x_n, H_n)^2 < +\infty. \tag{10}$$

Furthermore, if $(\lambda_n)_{n\geq 0}$ lies in $[0, 2-\varepsilon]$ for some $\varepsilon \in]0,1[$, then the series $\sum_{n\geq 0} \|x_{n+1} - x_n\|^2$ and $\sum_{n\geq 0} \langle \overline{x} - x_n \mid x_{n+1} - x_n \rangle$ converge [6], [15].

In view of (10), the next two convergence results are immediate consequences of Propositions 2 and 4, respectively.

Proposition 6 $(x_n)_{n\geq 0}$ converges weakly to a point in S if one of the conditions below is fulfilled.

i) (10) $\Rightarrow \mathfrak{W} \subset S$.
ii) (7) is in force and $\underline{\lim}\, d(x_n, H_n) = 0 \Rightarrow \mathfrak{W} \subset S$.
iii) (8) is in force and $\sum_{n\geq 0} d(x_n, H_n)^2 < +\infty \Rightarrow \mathfrak{W} \subset S$.

firmly nonexpansive → *firmly nonexpansive operator*
fixed point → *fixed point problem*

□

Proposition 7 $(x_n)_{n\geq 0}$ converges strongly to a point in S if one of the conditions below is fulfilled.

i) (10) $\Rightarrow \underline{\lim}\, d(x_n, S) = 0$.
ii) (7) is in force and $\underline{\lim}\, d(x_n, H_n) = 0 \Rightarrow \underline{\lim}\, d(x_n, S) = 0$.
iii) (8) is in force and $\sum_{n\geq 0} d(x_n, H_n)^2 < +\infty \Rightarrow \underline{\lim}\, d(x_n, S) = 0$.

□

To investigate linear convergence, assume that

$$(\exists\, \eta \in]0,1[)(\forall n \in \mathbf{N})\; d(x_n, H_n) \geq \eta d(x_n, S) \tag{11}$$

and that (8) holds. Then $\overline{x} = p_n$ in (9) supplies

$$\begin{aligned} d(x_{n+1}, S)^2 &\leq \|x_{n+1} - p_n\|^2 \\ &\leq d(x_n, S)^2 - \varepsilon^2 d(x_n, H_n)^2 \\ &\leq (1 - \varepsilon^2\eta^2) d(x_n, S)^2. \end{aligned}$$

Whence, Proposition 5 yields:

Proposition 8 Suppose that (8) and (11) hold. Then $(x_n)_{n\geq 0}$ converges linearly to a point $\overline{x} \in S$: $(\forall n \in \mathbf{N})\; \|x_n - \overline{x}\| \leq 2\kappa^n d(x_0, S)$ with $\kappa = (1 - \varepsilon^2\eta^2)^{1/2}$. □

Applications. Several convex optimization methods are now presented. They are shown to be Fejér monotone and their convergence is established on the basis of the general results stated above. For brevity, only weak convergence is considered; however, strong and linear convergence results can be derived in a like manner under suitable assumptions. In each problem, the solution set S is assumed to be nonempty.

Fixed Points of Nonlinear Operators. For every $n \in \mathbf{N}$, let $T_n \colon \mathcal{H} \to \mathcal{H}$ be a *firmly nonexpansive* operator, i.e.,

$$(\forall (x,y) \in \mathcal{H}^2)\; \langle T_n x - T_n y \mid x - y \rangle \geq \|T_n x - T_n y\|^2, \tag{12}$$

and let $\operatorname{Fix} T_n = \{x \in \mathcal{H} : T_n x = x\}$ be its fixed point set. The problem under consideration is to find a common *fixed point* of the family $(T_n)_{n\geq 0}$, i.e.,

$$\begin{cases} \text{Find} & \overline{x} \in \mathcal{H} \\ \text{s. t.} & (\forall n \in \mathbf{N})\; T_n \overline{x} = \overline{x}. \end{cases} \tag{13}$$

Let $S = \bigcap_{n\geq 0} \operatorname{Fix} T_n$ and

$$H_n = \{x \in \mathcal{H} :\ \langle x - T_n x_n \mid x_n - T_n x_n \rangle \leq 0\}.$$

It then follows from (12) that $S \subset \operatorname{Fix} T_n \subset H_n$. Thus, Algorithm 1 takes the following form.

Algorithm 2: Common fixed point	
0.	Take $x_0 \in \mathcal{H}$ and set $n = 0$.
1.	Take $\lambda_n \in [0, 2]$.
2.	Set $x_{n+1} = x_n + \lambda_n (T_n x_n - x_n)$.
3.	Set $n = n + 1$ and go to step 1.

Noting that $d(x_n, H_n) = \|(\mathrm{Id} - T_n)x_n\|$, several convergence results can be derived by direct application of Propositions 6-8. In particular, in the case of a single nonexpansive operator T (see (2)), the algorithm below is pertinent.

Algorithm 3: Fixed point	
0.	Take $x_0 \in \mathcal{H}$ and set $n = 0$.
1.	Take $\lambda_n \in [0, 1]$.
2.	Set $x_{n+1} = x_n + \lambda_n (T x_n - x_n)$.
3.	Set $n = n + 1$ and go to step 1.

Proposition 9 If $\sum_{n\geq 0} \lambda_n(1 - \lambda_n) = +\infty$, any sequence generated by Algorithm 3 converges weakly to a fixed point of T. $\square$

Indeed, the assignments $T_n \leftarrow (\mathrm{Id} + T)/2$ and $\lambda_n \leftarrow 2\lambda_n$ in Algorithm 2 yield Algorithm 3 as T_n is firmly nonexpansive [3], [5] and $\operatorname{Fix} T_n = \operatorname{Fix} T$. Next, observe that $(d(x_n, H_n))_{n\geq 0} = (\|(\mathrm{Id} - T)x_n\|/2)_{n\geq 0}$ is nonincreasing by (2). Hence, $\underline{\lim}\, d(x_n, H_n) = 0 \Rightarrow (\mathrm{Id} - T)x_n \to 0$ and it results from the demiclosedness of $\mathrm{Id} - T$ [20] that $x_{k_n} \rightharpoonup x \Rightarrow (\mathrm{Id} - T)x = 0$. Thus, Proposition 9 follows from Proposition 6ii).

Zeros of Monotone Maps. In connection with set-valued maps $A, B\colon \mathcal{H} \rightrightarrows \mathcal{H}$ a few definitions and facts need to be recalled [2], [27]. First, A is characterized by its graph $\mathrm{gr}A = \{(x, u) \in \mathcal{H}^2 :\ u \in Ax\}$. The inverse A^{-1} of A has graph $\{(u, x) \in \mathcal{H}^2 :\ (x, u) \in \mathrm{gr}A\}$ and the linear combination $A + \gamma B$ ($\gamma \in \mathbf{R}$) has graph

$$\{(x, u + \gamma v) :\ (x, u) \in \mathrm{gr}A, (x, v) \in \mathrm{gr}B\}.$$

maximal monotone→ *maximal monotone map*
proximal point algorithm

A is monotone if

$$(\forall (x, u) \in \mathrm{gr}A)(\forall (y, v) \in \mathrm{gr}A) \quad \langle x - y \mid u - v \rangle \geq 0.$$

If A is monotone and if there exists no monotone map $B \neq A$ such that $\mathrm{gr}A \subset \mathrm{gr}B$ then A is *maximal monotone*. In this case

- $\mathrm{gr}A$ is weakly-strongly closed: for every sequence $((y_n, v_n))_{n\geq 0}$ in $\mathcal{H}^2$

$$\begin{cases} ((y_n, v_n))_{n\geq 0} \text{ is in } \mathrm{gr}A \\ y_n \rightharpoonup y \\ v_n \to v \end{cases} \Rightarrow\ (y, v) \in \mathrm{gr}A. \tag{14}$$

- For every $\gamma \in]0, +\infty[$, the resolvent of A, $J_\gamma^A = (\mathrm{Id} + \gamma A)^{-1}$, is a single-valued firmly nonexpansive operator defined on $\mathcal{H}$ [17], [23].

Of broad interest is the problem of finding a zero of a maximal monotone map $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ [23], i.e.,

$$\begin{cases} \text{Find} & \overline{x} \in \mathcal{H} \\ \text{s. t.} & 0 \in A\overline{x}. \end{cases} \tag{15}$$

For every $\gamma \in]0, +\infty[$, the solution set $S = A^{-1}0$ can be written as $S = \{x \in \mathcal{H} :\ x \in x + \gamma Ax\} = \operatorname{Fix} J_\gamma^A$. Thus, given $(\gamma_n)_{n\geq 0}$ in $]0, +\infty[$, the equilibrium problem (15) can be cast in the form of the common fixed point problem (13) with $(T_n)_{n\geq 0} = (J_{\gamma_n}^A)_{n\geq 0}$. Algorithm 2 is then known as the (relaxed) *proximal point algorithm* [17], [23].

Algorithm 4: Proximal point	
0.	Take $x_0 \in \mathcal{H}$ and set $n = 0$.
1.	Take $\gamma_n \in]0, +\infty[$ and $\lambda_n \in [0, 2]$.
2.	Set $x_{n+1} = x_n + \lambda_n \big(J_{\gamma_n}^A x_n - x_n\big)$.
3.	Set $n = n + 1$ and go to step 1.

Proposition 10 Suppose that

$$\begin{cases} (\gamma_n)_{n\geq 0} \text{ is in } [\varepsilon, +\infty[\\ (\lambda_n)_{n\geq 0} \text{ is in } [\varepsilon, 2 - \varepsilon] \end{cases} \text{where } \varepsilon \in]0, 1[. \tag{16}$$

Then any sequence generated by Algorithm 4 converges weakly to a zero of A. $\square$

This result is a consequence of Proposition 6iii). Indeed, for every $n \in \mathbf{N}$, define $y_n = x_n +$

$(x_{n+1}-x_n)/\lambda_n$, $v_n=(x_n-x_{n+1})/(\gamma_n\lambda_n)$ and note that $v_n\in Ay_n$. Now suppose $d(x_n,H_n)\to 0$. Then, thanks to (16), $x_{n+1}-x_n\to 0$ and, in turn, $v_n\to 0$ and $y_n-x_n\to 0$. Hence, $x_{k_n}\rightharpoonup x \Rightarrow y_{k_n}\rightharpoonup x\Rightarrow 0\in Ax$ by (14).

Weak convergence can also be achieved under variants of (16), e.g., $\sum_{n\geq 0}\gamma_n^2=+\infty$ and $(\forall n\in\mathbf{N})\ \lambda_n=1$ [2]. Such results can be deduced from Proposition 6 as well.

Zeros of the Sum of Two Monotone Maps. Take two maximal monotone maps $A,B\colon\mathcal{H}\rightrightarrows\mathcal{H}$. An extension of (15) that captures a wide body of optimization and applied mathematics problems is [27]

$$\begin{cases}\text{Find} & \overline{x}\in\mathcal{H}\\ \text{s. t.} & 0\in A\overline{x}+B\overline{x}.\end{cases}\tag{17}$$

In instances when $A+B$ is maximal monotone, one can approach this problem via Algorithm 4. Naturally, for this approach to be numerically viable, the resolvents of $A+B$ should be computable relatively easily. A more widely applicable alternative is to devise an *operator splitting algorithm*, in which the operators A and B are employed in separate steps [16]. Two Fejérian splitting algorithms are described below.

First, suppose that B is (single-valued and) *co-coercive* in the sense that $B^{-1}-\alpha\mathrm{Id}$ is monotone for some $\alpha\in\,]0,+\infty[$, i.e.,

$$(\forall(x,y)\in\mathcal{H}^2)\ \langle Bx-By\mid x-y\rangle\geq \alpha\,\|Bx-By\|^2.\tag{18}$$

Given $\gamma\in\,]0,2\alpha]$, it follows from (18) that $\mathrm{Id}-\gamma B$ is nonexpansive. Moreover, the solution set $S=(A+B)^{-1}0$ can be written as $S=\{x\in\mathcal{H}\colon x-\gamma Bx\in x+\gamma Ax\}=\operatorname{Fix}T$ where $T=J_\gamma^A\circ(\mathrm{Id}-\gamma B)$ is nonexpansive as the composition of two nonexpansive operators. Algorithm 3 can then be implemented by alternating a forward step involving B with a backward (proximal) step involving A.

Algorithm 5: Forward-backward method
0. Take $\gamma\in\,]0,2\alpha]$, $x_0\in\mathcal{H}$, and set $n=0$.
1. Set $x_{n+\frac12}=x_n-\gamma Bx_n$ and take $\lambda_n\in[0,1]$.
2. Set $x_{n+1}=x_n+\lambda_n\left(J_\gamma^A x_{n+\frac12}-x_n\right)$.
3. Set $n=n+1$ and go to step 1.

As a corollary of Proposition 9 we obtain:

PROPOSITION 11 If $\sum_{n\geq 0}\lambda_n(1-\lambda_n)=+\infty$, any sequence generated by Algorithm 5 converges weakly to a zero of $A+B$. □

The second algorithm is centered around the operator $T=J_\gamma^A\circ(2J_\gamma^B-\mathrm{Id})+\mathrm{Id}-J_\gamma^B$, where $\gamma\in\,]0,+\infty[$. This operator possesses two nice properties: it is firmly nonexpansive and $y\in\operatorname{Fix}T\Leftrightarrow J_\gamma^B y\in(A+B)^{-1}0$ [16]. Whence, by putting $T_n\leftarrow T$ in Algorithm 2, one obtains the *Douglas-Rachford method* [8], [16].

Algorithm 6: Douglas-Rachford method
0. Take $\gamma\in\,]0,+\infty[$, $x_0\in\mathcal{H}$, and set $n=0$.
1. Set $x_{n+\frac12}=J_\gamma^B x_n$ and take $\lambda_n\in[0,2]$.
2. Set $x_{n+1}=x_n+\lambda_n\left(J_\gamma^A\left(2x_{n+\frac12}-x_n\right)-x_{n+\frac12}\right)$.
3. Set $n=n+1$ and go to step 1.

As in Algorithm 5, B is activated at step 1 and A at step 2. Convergence is established as in Proposition 9:

PROPOSITION 12 If $\sum_{n\geq 0}\lambda_n(2-\lambda_n)=+\infty$, any sequence generated by Algorithm 6 converges weakly and the image of the weak limit under J_γ^B is a zero of $A+B$. □

Variational Inequalities. Let $B\colon\mathcal{H}\to\mathcal{H}$ be a single-valued maximal monotone operator, let $\varphi\colon\mathcal{H}\to\,]-\infty,+\infty]$ be a proper, lower-semicontinuous, convex function, and let $\partial\varphi\colon\mathcal{H}\rightrightarrows\mathcal{H}$ be its subdifferential, i.e.,

$$\partial\varphi(x)=\bigcap_{y\in\mathcal{H}}\{u\in\mathcal{H}\colon\langle y-x\mid u\rangle+\varphi(x)\leq\varphi(y)\}.$$

operator splitting algorithm
co-coercive → co-coercive operator
Douglas-Rachford method
variational inequality

Then $\partial\varphi$ is maximal monotone [2] and, upon taking $A = \partial\varphi$ in (17), one arrives at the *variational inequality* problem

$$\begin{cases} \text{Find} & \overline{x} \in \mathcal{H} \\ \text{s. t.} & (\forall x \in \mathcal{H}) \\ & \langle \overline{x} - x \mid B\,\overline{x} \rangle + \varphi(\overline{x}) \le \varphi(x). \end{cases} \tag{19}$$

In this context, the resolvent J_γ^A reduces to Moreau's prox mapping [18]

$$\operatorname{prox}_\gamma^\varphi \colon x \mapsto \arg\min_{y \in \mathcal{H}} \varphi(y) + \frac{1}{2\gamma} \|y - x\|^2 .$$

As a special instance of (17), the variational inequality problem (19) can be solved via the forward-backward method (Algorithm 5) and Proposition 11 then yields:

PROPOSITION 13 Suppose that (18) is in force. Take $\gamma \in]0, 2\alpha]$, $x_0 \in \mathcal{H}$, and let

$$(\forall n \in \mathbf{N}) \;\; x_{n+1} = x_n + \lambda_n \left(\operatorname{prox}_\gamma^\varphi (x_n - \gamma B x_n) - x_n \right), \tag{20}$$

where $(\lambda_n)_{n\ge 0}$ is in $[0,1]$ and $\sum_{n\ge 0} \lambda_n (1 - \lambda_n) = +\infty$. Then $(x_n)_{n\ge 0}$ converges weakly to a solution of (19). □

A noteworthy situation is when $\varphi = \iota_Q$, where ι_Q is the indicator function of a nonempty closed convex set Q, i.e.,

$$\iota_Q \colon x \mapsto \begin{cases} 0 & \text{if } x \in Q \\ +\infty & \text{if } x \notin Q. \end{cases} \tag{21}$$

It follows that $\partial\iota_Q = N_Q$, where N_Q is the normal cone to Q, i.e.,

$$N_Q\, x = \bigcap_{y \in Q} \{ u \in \mathcal{H} : \langle y - x \mid u \rangle \le 0 \},$$

if $x \in Q$, and $N_Q\, x = \emptyset$ otherwise. In addition, (19) reads

$$\begin{cases} \text{Find} & \overline{x} \in Q \\ \text{s. t.} & (\forall x \in Q) \;\; \langle \overline{x} - x \mid B\,\overline{x} \rangle \le 0, \end{cases} \tag{22}$$

and $\operatorname{prox}_\gamma^{\iota_Q} = P_Q$ is the projector onto Q.

Differentiable Optimization. A standard convex programming problem is to minimize a proper, lower-semicontinuous, convex function $f \colon \mathcal{H} \to]-\infty, +\infty]$ over a nonempty closed convex set $Q \subset \mathcal{H}$, i.e.,

$$\text{Find } \overline{x} = \arg\min_{x \in Q} f(x). \tag{23}$$

In view of (21), (23) is equivalent to finding a global minimizer of $\iota_Q + f$, i.e., by Fermat's rule, to finding a zero of $\partial(\iota_Q + f)$. If 0 lies in the interior of $Q - \{x \in \mathcal{H} : f(x) < +\infty\}$, then $\partial(\iota_Q + f) = \partial\iota_Q + \partial f$ [2] and (23) is therefore of the form (17) with $A = N_Q$ and $B = \partial f$. This occurs in particular when f is finite and continuous at a point in Q.

Now suppose that f is differentiable. Then $\partial f = \{\nabla f\}$ is single-valued and (23) can further be reduced to (22) with $B = \nabla f$. The forward-backward scheme (20) then becomes the *projected gradient algorithm*

$$(\forall n \in \mathbf{N}) \;\; x_{n+1} = x_n + \lambda_n \left(P_Q (x_n - \gamma \nabla f(x_n)) - x_n \right).$$

Proposition 13 provides conditions for weak convergence to a minimizer of f over Q.

Convex Feasibility Problems. Given a family $(S_i)_{i\in I}$ of intersecting nonempty closed and convex subsets of $\mathcal{H}$, the *convex feasibility problem* reads [3], [5], [6], [15]

$$\text{Find } \overline{x} \in S = \bigcap_{i \in I} S_i. \tag{24}$$

At iteration n, select a nonempty finite index set $I_n \subset I$ and, for every $i \in I_n$, let $p_{i,n}$ be an approximate projection of x_n onto S_i, i.e., the projection of x_n onto a closed affine half-space $H_{i,n}$ containing S_i. Then

$$H_{i,n} = \{ x \in \mathcal{H} : \langle x - p_{i,n} \mid x_n - p_{i,n} \rangle \le 0 \}.$$

Let

$$H_n = \left\{ x \in \mathcal{H} : \sum_{i \in I_n} w_{i,n} \langle x - p_{i,n} \mid x_n - p_{i,n} \rangle \le 0 \right\}$$

where the weights $(w_{i,n})_{i\in I_n}$ are in $]0,1]$ and satisfy $\sum_{i\in I_n} w_{i,n} = 1$. Then $S \subset \bigcap_{i\in I_n} S_i \subset \bigcap_{i\in I_n} H_{i,n} \subset H_n$ and $P_n x_n = x_n + L_n (x_{n+\frac{1}{2}} - x_n)$,

projected gradient algorithm
convex feasibility problem

where $x_{n+\frac{1}{2}} = \sum_{i\in I_n} w_{i,n} p_{i,n}$ and

$$L_n = \begin{cases} \dfrac{\sum_{i\in I_n} w_{i,n} \|p_{i,n} - x_n\|^2}{\left\|x_{n+\frac{1}{2}} - x_n\right\|^2} & \text{if } x_{n+\frac{1}{2}} \neq x_n \\ 1 & \text{else.} \end{cases} \tag{25}$$

Algorithm 1 then turns into Algorithm 7.

Algorithm 7: Convex feasibility
0. Take $x_0 \in \mathcal{H}$ and set $n = 0$.
1. Take a nonempty finite set $I_n \subset I$.
2. Compute approximate projections $(p_{i,n})_{i\in I_n}$ of x_n onto $(S_i)_{i\in I_n}$.
3. Take $(w_{i,n})_{i\in I_n}$ in $]0,1]$ such that $\sum_{i\in I_n} w_{i,n} = 1$.
4. Set $x_{n+\frac{1}{2}} = \sum_{i\in I_n} w_{i,n} p_{i,n}$, L_n as in (25).
5. Take $\lambda_n \in [0, 2L_n]$.
6. Set $x_{n+1} = x_n + \lambda_n \left(x_{n+\frac{1}{2}} - x_n\right)$.
7. Set $n = n + 1$ and go to step 1.

Weak convergence to a point in S follows from Proposition 6 under various assumptions on the control sequence $(I_n)_{n\geq 0}$ and the approximate projections $((p_{i,n})_{i\in I_n})_{n\geq 0}$ [5], [6], [15].

Nondifferentiable Optimization. Suppose that f is subdifferentiable in (23), i.e., $(\forall x \in \mathcal{H})$ $\partial f(x) \neq \emptyset$, and that its minimum value $\overline{f}$ over Q is known. Then (23) can be viewed as a special case of (24) with two sets, namely $S_1 = Q$ and $S_2 = \{x \in \mathcal{H} : \ f(x) \leq \overline{f}\}$. Now take

$$H_{2,n} = \left\{x \in \mathcal{H} : \ \langle x - x_n \mid u_n\rangle \leq \overline{f} - f(x_n)\right\}$$

where $u_n \in \partial f(x_n)$. Then $S_2 \subset H_{2,n}$ and

$$p_{2,n} = \begin{cases} x_n + \dfrac{\overline{f} - f(x_n)}{\|u_n\|^2} u_n & \text{if } x_n \notin S_2 \\ x_n & \text{otherwise} \end{cases}$$

is called a *subgradient projection* of x_n onto S_2 [3], [5]. If Algorithm 7 is implemented by alternating a relaxed subgradient projection onto S_2 with an exact projection onto S_1, i.e.,

$$(\forall n \in \mathbf{N}) \ x_{n+1} = P_Q\left(x_n + \lambda_n\left(p_{2,n} - x_n\right)\right),$$

one obtains the subgradient projection method of [21]. Weak convergence to a solution of (23) under the assumptions that ∂f maps bounded sets into bounded sets, $(\lambda_n)_{n\geq 0}$ is in $[0, 2]$, and (8), follows from Proposition 6iii) [3], [5].

subgradient projection

Inconsistent Convex Feasibility Problems. When $\bigcap_{i\in I} S_i = \emptyset$ and I is finite, (24) can be replaced by the minimization problem

$$\text{Find } \overline{x} = \arg\min_{x\in\mathcal{H}} \frac{1}{2} \sum_{i\in I} w_i d(x, S_i)^2 \tag{26}$$

where $(w_i)_{i\in I}$ is in $]0, 1]$ and $\sum_{i\in I} w_i = 1$. Let $(P_i)_{i\in I}$ be the projectors onto $(S_i)_{i\in I}$, let $T = \sum_{i\in I} w_i P_i$, and let S be the solution set of (26). Then T is firmly nonexpansive and $S = \operatorname{Fix} T$ [5]. By reiterating a previous argument, one obtains:

Proposition 14 Take $x_0 \in \mathcal{H}$, $(\lambda_n)_{n\geq 0}$ in $[0, 2]$ such that $\sum_{n\geq 0} \lambda_n(2 - \lambda_n) = +\infty$, and let

$$(\forall n \in \mathbf{N}) \ x_{n+1} = x_n + \lambda_n \left(\sum_{i\in I} w_i P_i x_n - x_n\right).$$

Then $(x_n)_{n\geq 0}$ converges weakly to a solution of (26). □

Historical Notes and Comments. In 1922, L. Fejér considered the following problem [12]: given a closed subset $S \subset \mathbf{R}^p$ and a point $y \notin S$ can one find a point $x \in \mathbf{R}^p$ such that

$$(\forall \overline{x} \in S) \ \|x - \overline{x}\| < \|y - \overline{x}\|.$$

Inspired by this work, T.S. Motzkin and I.J. Schoenberg adopted in their 1954 paper [19] the term Fejér monotone to describe sequences satisfying (1). In this paper (see also [1]), an algorithm was developed to solve systems of linear inequalities in $\mathbf{R}^p$ by successive projections onto the half-spaces defining the polyhedral solution set S. The concept of Fejér monotonicity was shown to be an adequate tool to study convergence of this algorithm. Basic facts such as (5) and (9) can already be found in [19] and [1], respectively.

In the 1960s, I.I. Eremin extended the use of Fejér monotonicity to more general convex problems in Hilbert spaces. A summary of his publications covering the period 1961-1967 is given in [9]. By the end of the 1960's, most results on Fejér monotonicity in Hilbert spaces were essentially known and one can find them scattered in the Soviet literature in the context of specific convex programming problems. Thus, (4) appears in [10], Proposition 2 in [4], Propositions 4 and 5 in [14], and Proposition 8 in [14] and [21]. It should be noted that Proposition 2 has been implicitly rediscovered many times and that it seems to originate in [24].

Recently, Fejér monotonicity has been reserved a featured role in several convex optimization papers [3], [6], [15], [25], [26]. It has also proven a valuable tool in more applied disciplines such as biology, economics, and engineering [5], [11]. Some extensions of the notion of Fejér monotonicity are discussed in [7].

Acknowledgement. This work was supported by the National Science Foundation under grant MIP-9705504.

References

[1] Agmon, S.: 'The relaxation method for linear inequalities', *Canad. J. Math.* **6** (1954), 382–392.

[2] Aubin, J.-P., and Cellina, A.: *Differential inclusions*, Springer, 1984.

[3] Bauschke, H.H., and Borwein, J.M.: 'On projection algorithms for solving convex feasibility problems', *SIAM Rev.* **38** (1996), 367–426.

[4] Brègman, L.M.: 'The method of successive projection for finding a common point of convex sets', *Soviet Math. Dokl.* **6** (1965), 688–692.

[5] Combettes, P.L.: 'The convex feasibility problem in image recovery': *Advances in Imaging and Electron Physics*, Vol. 95, Academic, 1996, pp. 155–270.

[6] Combettes, P.L.: 'Hilbertian convex feasibility problem: Convergence of projection methods', *Appl. Math. Optim.* **35** (1997), 311–330.

[7] Combettes, P.L.: 'Quasi-Fejérian analysis of some optimization algorithms', in D. Butnariu, Y. Censor, and S. Reich (eds.): *Inherently Parallel Algorithms for Feasibility and Optimization and Their Applications*, Elsevier, 2001.

[8] Eckstein, J., and Bertsekas, D.P.: 'On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators', *Math. Programming* **55** (1992), 293–318.

[9] Eremin, I.I.: 'Methods of Fejér approximations in convex programming', *Math. Notes* **3** (1968), 139–149.

[10] Eremin, I.I.: 'On the speed of convergence in the method of Fejér approximations', *Math. Notes* **4** (1968), 522–527.

[11] Eremin, I.I., and Mazurov, V.D.: *Nonstationary processes of mathematical programming*, Nauka, 1979.

[12] Fejér, L.: 'Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen', *Math. Ann.* **85** (1922), 41–48.

[13] Genel, A., and Lindenstrauss, J.: 'An example concerning fixed points', *Israel J. Math.* **22** (1975), 81–86.

[14] Gubin, L.G., Polyak, B.T., and Raik, E. V.: 'The method of projections for finding the common point of convex sets', *USSR Comput. Math. Math. Phys.* **7** (1967), 1–24.

[15] Kiwiel, K.C., and Łopuch, B.: 'Surrogate projection methods for finding fixed points of firmly nonexpansive mappings', *SIAM J. Optim.* **7** (1997), 1084–1102.

[16] Lions, P.L., and Mercier, B.: 'Splitting algorithms for the sum of two nonlinear operators', *SIAM J. Numer. Anal.* **16** (1979), 964–979.

[17] Martinet, B.: 'Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox', *C. R. Acad. Sci. Paris Sér. A Math.* **274** (1972), 163–165.

[18] Moreau, J.-J.: 'Fonctions convexes duales et points proximaux dans un espace hilbertien', *C. R. Acad. Sci. Paris Sér. A Math.* **255** (1962), 2897–2899.

[19] Motzkin, T.S., and Schoenberg, I.J.: 'The relaxation method for linear inequalities', *Canad. J. Math.* **6** (1954), 393–404.

[20] Opial, Z.: 'Weak convergence of the sequence of successive approximations for nonexpansive mappings', *Bull. Amer. Math. Soc.* **73** (1967), 591–597.

[21] Polyak, B.T.: 'Minimization of unsmooth functionals', *USSR Comput. Math. Math. Phys.* **9** (1969), 14–29.

[22] Raik, E.: 'A class of iterative methods with Fejér-monotone sequences', *Eesti NSV Tead. Akad. Toimetised Füüs.-Mat.* **18** (1969), 22–26.

[23] Rockafellar, R.T.: 'Monotone operators and the proximal point algorithm', *SIAM J. Control Optim.* **14** (1976), 877–898.

[24] Schaefer, H.: 'Über die Methode sukzessiver Approximationen', *Jahresber. Deutsch. Math.-Verein.* **59** (1957), 131–140.

[25] Schott, D.: 'A general iterative scheme with applications to convex optimization and related fields', *Optimization* **22** (1991), 885–902.

[26] Schott, D.: 'Iterative solution of convex problems by Fejér-monotone methods', *Numer. Funct. Anal. Optim.* **16** (1995), 1323–1357.

[27] Zeidler, E.: *Nonlinear functional analysis and its applications II/B – Nonlinear monotone operators*, Springer, 1990.

Patrick L. Combettes
Laboratoire d'Analyse Numérique
Université Pierre et Marie Curie – Paris 6
75005 Paris, France
and
City College and Graduate School
City University of New York

New York, NY 10031, USA
E-mail address: plc@ann.jussieu.fr

AMS 1991 Subject Classification: 47H05, 49M45, 65J15, 90C25.
Key words and phrases: Fejér monotonicity, convex programming, convergence.