QUASI-FEJÉRIAN ANALYSIS OF SOME OPTIMIZATION ALGORITHMS

Patrick L. Combettes[a*]

[a]Laboratoire d'Analyse Numérique, Université Pierre et Marie Curie – Paris 6, 4 Place Jussieu, 75005 Paris, France
and
City College and Graduate Center, City University of New York, New York, NY 10031, USA

A quasi-Fejér sequence is a sequence which satisfies the standard Fejér monotonicity property to within an additional error term. This notion is studied in detail in a Hilbert space setting and shown to provide a powerful framework to analyze the convergence of a wide range of optimization algorithms in a systematic fashion. A number of convergence theorems covering and extending existing results are thus established. Special emphasis is placed on the design and the analysis of parallel algorithms.

1. INTRODUCTION

The convergence analyses of convex optimization algorithms often follow standard patterns. This observation suggests the existence of broad structures within which these algorithms could be recast and then studied in a simplified and unified manner.

One such structure relies on the concept of Fejér monotonicity: a sequence $(x_n)_{n\geq 0}$ in a Hilbert space $\mathcal{H}$ is said to be a *Fejér (monotone) sequence* relative to a *target set* $S \subset \mathcal{H}$ if

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

In convex optimization, this basic property has proven to be an efficient tool to analyze various optimization algorithms in a unified framework, e.g., [8], [9], [10], [13], [20], [22], [29], [30], [31], [45], [54], [63], [64], [69]; see also [24] for additional references and an historical perspective. In this context, the target set S represents the set of solutions to the problem under consideration and (1) states that each iterate generated by the underlying solution algorithm cannot be further from any solution point than its predecessor.

In order to derive unifying convergence principles for a broader class of optimization algorithms, the notion of Fejér monotonicity can be extended in various directions. In this paper, the focus will be placed on three variants of (1).

Definition 1.1 Relative to a nonempty *target set* $S \subset \mathcal{H}$, a sequence $(x_n)_{n\geq 0}$ in $\mathcal{H}$ is

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

- *Quasi-Fejér of Type I* if

$$(\exists\,(\varepsilon_n)_{n\geq 0}\in \ell_+\cap\ell^1)(\forall x\in S)(\forall n\in\mathbb{N})\ \ \|x_{n+1}-x\|\leq\|x_n-x\|+\varepsilon_n. \tag{2}$$

- *Quasi-Fejér of Type II* if

$$(\exists\,(\varepsilon_n)_{n\geq 0}\in \ell_+\cap\ell^1)(\forall x\in S)(\forall n\in\mathbb{N})\ \ \|x_{n+1}-x\|^2\leq\|x_n-x\|^2+\varepsilon_n. \tag{3}$$

- *Quasi-Fejér of Type III* if

$$(\forall x\in S)(\exists\,(\varepsilon_n)_{n\geq 0}\in \ell_+\cap\ell^1)(\forall n\in\mathbb{N})\ \ \|x_{n+1}-x\|^2\leq\|x_n-x\|^2+\varepsilon_n. \tag{4}$$

The concept of quasi-Fejér sequences goes back to [34], where it was introduced through a property akin to (4) for sequences of $\mathbb{R}^N$-valued random vectors (see also [33] for more recent developments in that direction). Instances of quasi-Fejér sequences of the above three types also appear explicitly in [2], [47], and [49].

The goal of this paper is to study the properties of the above types of quasi-Fejér sequences and to exploit them to derive convergence results for numerous optimization algorithms. Known results will thus be recast in a common framework and new extensions will be obtained in a straightforward fashion. In Section 2, it is shown that most common types of nonlinear operators arising in convex optimization belong to a so-called $\mathfrak{T}$ class whose properties are investigated. The asymptotic properties of quasi-Fejér sequences are discussed in Section 3. In particular, necessary and sufficient conditions for weak and strong convergence are established and convergence estimates are derived. In Section 4, a generic quasi-Fejér algorithm is constructed by iterating $\mathfrak{T}$-class operator with errors and introducing relaxation parameters. Convergence results for this algorithm are obtained through the analysis developed in Section 3 and applied to specific optimization problems in Section 5. In Section 6, a general inexact, parallel, block-iterative algorithm for countable convex feasibility problems is derived from the generic algorithm constructed in Section 4 and analyzed. All the algorithms discussed up to this point are essentially perturbed Fejér monotone algorithms. Section 7 concerns the projected subgradient method and constitutes a different field of applications of the analysis of Section 3.

Notation. Throughout $\mathcal{H}$ is a real Hilbert space with scalar product $\langle\cdot\mid\cdot\rangle$, norm $\|\cdot\|$, and distance d. Given $x\in\mathcal{H}$ and $\rho\in\,]0,+\infty[$, $B(x,\rho)$ is the closed ball of center x and radius ρ. The expressions $x_n\rightharpoonup x$ and $x_n\to x$ denote respectively the weak and strong convergence to x of a sequence $(x_n)_{n\geq 0}$ in $\mathcal{H}$, $\mathfrak{W}(x_n)_{n\geq 0}$ its set of weak cluster points, and $\mathfrak{S}(x_n)_{n\geq 0}$ its set of strong cluster points. $\overline{\operatorname{aff}}\,S$, $\overline{\operatorname{conv}}\,S$, and $\operatorname{conv}S$ are respectively the closed affine hull, the closed convex hull, and the convex hull of a set S. d_S is the distance function to the set S and, if S is closed and convex, P_S is the projector onto S. The sets $\operatorname{dom}A=\{x\in\mathcal{H}\mid Ax\neq\varnothing\}$, $\operatorname{ran}A=\{u\in\mathcal{H}\mid(\exists\,x\in\mathcal{H})\ u\in Ax\}$, and $\operatorname{gr}A=\{(x,u)\in\mathcal{H}^2\mid u\in Ax\}$ are respectively the domain, the range, and the graph of a set-valued operator $A\colon\mathcal{H}\to 2^{\mathcal{H}}$; the inverse A^{-1} of A is the set-valued operator with graph $\{(u,x)\in\mathcal{H}^2\mid u\in Ax\}$. The *subdifferential* of a function $f\colon\mathcal{H}\to\mathbb{R}$ is the set-valued operator

$$\partial f\colon\mathcal{H}\to 2^{\mathcal{H}}\colon x\mapsto\big\{u\in\mathcal{H}\mid(\forall y\in\mathcal{H})\ \ \langle y-x\mid u\rangle+f(x)\leq f(y)\big\} \tag{5}$$

and the elements of $\partial f(x)$ are the *subgradients* of f at x. The *lower level set* of f at height $\eta \in \mathbb{R}$ is $\operatorname{lev}_{\leq \eta} f = \{x \in \mathcal{H} \mid f(x) \leq \eta\}$. $\operatorname{Fix} T = \{x \in \mathcal{H} \mid Tx = x\}$ denotes the set of fixed points of an operator $T \colon \mathcal{H} \to \mathcal{H}$. Finally, ℓ_+ denotes the set of all sequences in $[0, +\infty[$ and ℓ^1 [resp. ℓ^2] the space of all absolutely [resp. square] summable sequences in $\mathbb{R}$.

2. NONLINEAR OPERATORS

Convex optimization algorithms involve a variety of (not necessarily linear) operators. In this respect, recall that an operator $T \colon \mathcal{H} \to \mathcal{H}$ with $\operatorname{dom} T = \mathcal{H}$ is *firmly nonexpansive* if

$$(\forall (x, y) \in \mathcal{H}^2) \;\; \|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(T - \operatorname{Id})x - (T - \operatorname{Id})y\|^2; \tag{6}$$

nonexpansive if

$$(\forall (x, y) \in \mathcal{H}^2) \;\; \|Tx - Ty\| \leq \|x - y\|; \tag{7}$$

and *quasi-nonexpansive* if

$$(\forall (x, y) \in \mathcal{H} \times \operatorname{Fix} T) \;\; \|Tx - y\| \leq \|x - y\|. \tag{8}$$

Clearly, (6) $\Rightarrow$ (7) $\Rightarrow$ (8). Now let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator. Then the *resolvent* of index $\gamma \in]0, +\infty[$ of A is (the single-valued operator) $(\operatorname{Id} + \gamma A)^{-1}$. Next, let $f \colon \mathcal{H} \to \mathbb{R}$ be a continuous convex function such that $\operatorname{lev}_{\leq 0} f \neq \varnothing$ and let g be a selection of ∂f. Then the operator

$$G_f^g : x \mapsto \begin{cases} x - \dfrac{f(x)}{\|g(x)\|^2} g(x), & \text{if } \; f(x) > 0 \\ x & \text{if } \; f(x) \leq 0 \end{cases} \tag{9}$$

is a *subgradient projector* onto $\operatorname{lev}_{\leq 0} f$.

The above operators are closely related to the so-called class $\mathfrak{T}$ of [9]. Given $(x, y) \in \mathcal{H}^2$, we shall use the notation

$$H(x, y) = \big\{u \in \mathcal{H} \mid \langle u - y \mid x - y\rangle \leq 0\big\}. \tag{10}$$

Definition 2.1 [9, Def. 2.2] $\mathfrak{T} = \big\{T \colon \mathcal{H} \to \mathcal{H} \,|\, \operatorname{dom} T = \mathcal{H}, (\forall x \in \mathcal{H}) \operatorname{Fix} T \subset H(x, Tx)\big\}$.

Proposition 2.2 [9, Prop. 2.3] *Consider the following statements:*

(i) T *is the projector onto a nonempty closed and convex subset of* $\mathcal{H}$.

(ii) T *is the resolvent of a maximal monotone operator* $A \colon \mathcal{H} \to 2^{\mathcal{H}}$.

(iii) $\operatorname{dom} T = \mathcal{H}$ *and* T *is firmly nonexpansive.*

(iv) *T is a subgradient projector onto* $\operatorname{lev}_{\leq 0} f$, *where* $f\colon \mathcal{H} \to \mathbb{R}$ *is a continuous convex function such that* $\operatorname{lev}_{\leq 0} f \neq \varnothing$.

(v) $\operatorname{dom} T = \mathcal{H}$ *and* $2T - \operatorname{Id}$ *is quasi-nonexpansive.*

(vi) $T \in \mathfrak{T}$.

Then:

$$\begin{array}{ccccc} \text{(i)} & \Rightarrow & \text{(ii)} & \Leftrightarrow & \text{(iii)} \\ \Downarrow & & & & \Downarrow \\ \text{(iv)} & \Rightarrow & \text{(v)} & \Leftrightarrow & \text{(vi)}. \end{array}$$

Some properties of $\mathfrak{T}$-class operators are described below.

Proposition 2.3 *Every T in $\mathfrak{T}$ satisfies the following properties.*

(i) $(\forall (x,y) \in \mathcal{H} \times \operatorname{Fix} T)\ \ \|Tx - x\|^2 \leq \langle y - x \mid Tx - x\rangle$.

(ii) *Put* $T' = \operatorname{Id} + \lambda(T - \operatorname{Id})$, *where* $\lambda \in [0,2]$. *Then* $(\forall (x,y) \in \mathcal{H} \times \operatorname{Fix} T)\ \ \|T'x - y\|^2 \leq \|x - y\|^2 - \lambda(2-\lambda)\|Tx - x\|^2$.

(iii) $(\forall x \in \mathcal{H})\ \ \|Tx - x\| \leq d_{\operatorname{Fix} T}(x)$.

(iv) $\operatorname{Fix} T = \bigcap_{x \in \mathcal{H}} H(x, Tx)$.

(v) $\operatorname{Fix} T$ *is closed and convex.*

(vi) $(\forall \lambda \in [0,1])\ \ \operatorname{Id} + \lambda(T - \operatorname{Id}) \in \mathfrak{T}$.

Proof. From Definition 2.1, we get

$$(\forall (x,y) \in \mathcal{H} \times \operatorname{Fix} T)\ \ \langle y - Tx \mid x - Tx\rangle \leq 0. \tag{11}$$

and (i) ensues. (ii): Fix $(x,y) \in \mathcal{H} \times \operatorname{Fix} T$. It follows from (i) that

$$\begin{aligned} \|T'x - y\|^2 &= \|x - y\|^2 - 2\lambda\langle y - x \mid Tx - x\rangle + \lambda^2\|Tx - x\|^2 \\ &\leq \|x - y\|^2 - \lambda(2-\lambda)\|Tx - x\|^2. \end{aligned} \tag{12}$$

(iii): Fix $x \in \mathcal{H}$. Then (ii) with $\lambda = 1$ implies

$$(\forall y \in \operatorname{Fix} T)\ \ \|Tx - x\| \leq \|y - x\|. \tag{13}$$

Now take the infimum over all $y \in \operatorname{Fix} T$ (with the usual convention $\inf \varnothing = +\infty$). (iv)-(vi): See [9, Prop. 2.6]. □

The next proposition provides a generalization of the operator averaging process described in item (vi) that will be an essential component in the design of block-iterative parallel algorithms in Section 6.

Proposition 2.4 *Let I be a countable index set, $(T_i)_{i\in I}$ a family of operators in $\mathfrak{T}$, and $(\omega_i)_{i\in I}$ strictly positive real numbers such that $\sum_{i\in I}\omega_i = 1$. Suppose that $\bigcap_{i\in I}\operatorname{Fix} T_i \neq \varnothing$ and let*

$$T\colon \mathcal{H}\to\mathcal{H}\colon x\mapsto x+\lambda(x)L\big(x,(T_i)_{i\in I},(\omega_i)_{i\in I}\big)\left(\sum_{i\in I}\omega_i T_i x - x\right), \tag{14}$$

where, for every $x\in\mathcal{H}$, $\lambda(x)\in\left]0,1\right]$ and

$$L\big(x,(T_i)_{i\in I},(\omega_i)_{i\in I}\big) = \begin{cases} 1 & \text{if } x\in\bigcap_{i\in I}\operatorname{Fix} T_i\\ \dfrac{\sum_{i\in I}\omega_i\|T_i x - x\|^2}{\|\sum_{i\in I}\omega_i T_i x - x\|^2} & \text{otherwise.}\end{cases} \tag{15}$$

Then $\operatorname{Fix} T = \bigcap_{i\in I}\operatorname{Fix} T_i$ and $T\in\mathfrak{T}$.

Proof. Fix $(x,y)\in\mathcal{H}\times\bigcap_{i\in I}\operatorname{Fix} T_i$. We first observe that the series $\sum_{i\in I}\omega_i(T_i x - x)$ converges absolutely since, by (13), $\sum_{i\in I}\omega_i\|T_i x - x\|\leq\|y-x\|$. Moreover, by Proposition 2.3(i),

$$(\forall i\in I)\ \ \|T_i x - x\|^2\leq\langle y-x \mid T_i x - x\rangle. \tag{16}$$

Hence, since the function $\|\cdot\|^2$ is convex and continuous, we arrive at the chain of inequalities

$$\begin{aligned}\left\|\sum_{i\in I}\omega_i T_i x - x\right\|^2 &\leq \sum_{i\in I}\omega_i\|T_i x - x\|^2 \leq \left\langle y-x \,\middle|\, \sum_{i\in I}\omega_i T_i x - x\right\rangle\\ &\leq \|y-x\|\cdot\left\|\sum_{i\in I}\omega_i T_i x - x\right\|,\end{aligned} \tag{17}$$

from which we deduce that

$$\begin{aligned} x\in\operatorname{Fix}\sum_{i\in I}\omega_i T_i &\Leftrightarrow \left\|\sum_{i\in I}\omega_i T_i x - x\right\| = 0 \Leftrightarrow \sum_{i\in I}\omega_i\|T_i x - x\|^2 = 0\\ &\Leftrightarrow x\in\bigcap_{i\in I}\operatorname{Fix} T_i.\end{aligned} \tag{18}$$

Hence, $L(x)$ in (15) is a well-defined number in $\left[1,+\infty\right[$ and

$$\operatorname{Fix} T = \operatorname{Fix}\sum_{i\in I}\omega_i T_i = \bigcap_{i\in I}\operatorname{Fix} T_i. \tag{19}$$

Since $\operatorname{dom} T = \mathcal{H}$, it remains to show that $\operatorname{Fix} T\subset H(x,Tx)$ to establish that $T\in\mathfrak{T}$. To this end, we derive from (16) that

$$\begin{aligned}
\langle y - x \mid Tx - x\rangle &= \lambda(x)L(x)\sum_{i\in I}\omega_i\langle y - x \mid T_i x - x\rangle \\
&\geq \lambda(x)L(x)\sum_{i\in I}\omega_i\|T_i x - x\|^2 \\
&= \frac{1}{\lambda(x)}\left(\lambda(x)L(x)\left\|\sum_{i\in I}\omega_i T_i x - x\right\|\right)^2 \\
&= \|Tx - x\|^2/\lambda(x) \\
&\geq \|Tx - x\|^2. \qquad (20)
\end{aligned}$$

It follows that $\langle y - Tx \mid x - Tx\rangle \leq 0$ and, in turn, that $y \in H(x, Tx)$. Since by virtue of (19) y is an arbitrary fixed point of T, we conclude that $\operatorname{Fix} T \subset H(x, Tx)$. □

Remark 2.5 Taking $\lambda(x) = 1/L(x)$ in (14) yields $\sum_{i\in I}\omega_i T_i \in \mathfrak{T}$.

A couple of additional definitions will be required. An operator $T\colon \mathcal{H} \to \mathcal{H}$ is said to be *demiclosed* at $y \in \mathcal{H}$ if for every $x \in \mathcal{H}$ and every sequence $(x_n)_{n\geq 0}$ in $\mathcal{H}$ such that $x_n \rightharpoonup x$ and $Tx_n \to y$, we have $Tx = y$ [13]; *demicompact* [resp. demicompact at $y \in \mathcal{H}$] if, for every bounded sequence $(x_n)_{n\geq 0}$ such that $(Tx_n - x_n)_{n\geq 0}$ converges strongly [resp. converges strongly to y], we have $\mathfrak{S}(x_n)_{n\geq 0} \neq \varnothing$ [57].

Remark 2.6 Take an operator $T\colon \mathcal{H} \to \mathcal{H}$. Then:

- If T is nonexpansive, then $T - \mathrm{Id}$ is demiclosed on $\mathcal{H}$ [13].
- One will easily check that T is demicompact if its range is boundedly compact (its intersection with any closed ball is compact), e.g., T is the projector onto a boundedly compact convex set. Other examples will be found in [57].

3. PROPERTIES OF QUASI-FEJÉR SEQUENCES

As we shall find in this section, most of the asymptotic properties of Fejér sequences remain valid for quasi-Fejér sequences.

3.1. Basic properties

First, we need

Lemma 3.1 *Let* $\chi \in\,]0,1]$, $(\alpha_n)_{n\geq 0} \in \ell_+$, $(\beta_n)_{n\geq 0} \in \ell_+$, *and* $(\varepsilon_n)_{n\geq 0} \in \ell_+ \cap \ell^1$ *be such that*

$$(\forall n \in \mathbb{N})\quad \alpha_{n+1} \leq \chi\alpha_n - \beta_n + \varepsilon_n. \qquad (21)$$

Then

(i) $(\alpha_n)_{n\geq 0}$ *is bounded.*

(ii) $(\alpha_n)_{n\geq 0}$ *converges.*

(iii) $(\beta_n)_{n\geq 0} \in \ell^1$.

(iv) *If* $\chi \neq 1$, $(\alpha_n)_{n\geq 0} \in \ell^1$.

Proof. Put $\varepsilon = \sum_{n\geq 0} \varepsilon_n$ and $\alpha = \underline{\lim}\, \alpha_n$. (i): We derive from (21) that

$$(\forall n \in \mathbb{N})\ \ 0 \leq \alpha_{n+1} \leq \chi^{n+1}\alpha_0 + \gamma_n, \ \ \text{where} \ \ \gamma_n = \sum_{k=0}^{n} \chi^{n-k}\varepsilon_k. \tag{22}$$

Hence, $(\alpha_n)_{n\geq 0}$ lies in $[0, \alpha_0 + \varepsilon]$. (ii): It follows from the previous inclusion that $\alpha \in [0, \alpha_0 + \varepsilon]$. Now extract from $(\alpha_n)_{n\geq 0}$ a subsequence $(\alpha_{k_n})_{n\geq 0}$ such that $\alpha = \lim \alpha_{k_n}$ and fix $\delta \in \,]0, +\infty[$. Then we can find $n_0 \in \mathbb{N}$ such that $\alpha_{k_{n_0}} - \alpha \leq \delta/2$ and $\sum_{m\geq k_{n_0}} \varepsilon_m \leq \delta/2$. However, by (21),

$$(\forall n \geq k_{n_0})\ \ 0 \leq \alpha_n \leq \alpha_{k_{n_0}} + \sum_{m\geq k_{n_0}} \varepsilon_m \leq \delta/2 + \alpha + \delta/2 = \alpha + \delta. \tag{23}$$

Hence $\overline{\lim}\, \alpha_n \leq \underline{\lim}\, \alpha_n + \delta$ and, since δ can be arbitrarily small in $]0, +\infty[$, the whole sequence $(\alpha_n)_{n\geq 0}$ converges to α. (iii): It follows from (21) that, for every $N \in \mathbb{N}$, $\beta_N \leq \alpha_N - \alpha_{N+1} + \varepsilon_N$ and, in turn, $\sum_{n=0}^{N} \beta_n \leq \alpha_0 - \alpha_{N+1} + \sum_{n=0}^{N} \varepsilon_n \leq \alpha_0 + \varepsilon$. Hence $\sum_{n\geq 0} \beta_n \leq \alpha_0 + \varepsilon$. (iv): Suppose $\chi \in \,]0, 1[$. Then the sequence $(\gamma_n)_{n\geq 0}$ of (22) is the convolution of the two ℓ^1-sequences $(\chi^n)_{n\geq 0}$ and $(\varepsilon_n)_{n\geq 0}$. As such, it is therefore in ℓ^1 and the inequalities in (22) force $(\alpha_n)_{n\geq 0}$ in ℓ^1 as well. □

Let us start with some basic relationships between (2), (3), and (4).

Proposition 3.2 *Let $(x_n)_{n\geq 0}$ be a sequence in $\mathcal{H}$ and let S be a nonempty subset of $\mathcal{H}$. Then the three types of quasi-Fejér monotonicity of Definition 1.1 are related as follows:*

(i) *Type I $\Rightarrow$ Type III $\Leftarrow$ Type II.*

(ii) *If S is bounded, Type I $\Rightarrow$ Type II.*

Proof. It is clear that Type II $\Rightarrow$ Type III. Now suppose that $(x_n)_{n\geq 0}$ satisfies (2). Then

$$\begin{aligned}(\forall x \in S)(\forall n \in \mathbb{N})\ \ \|x_{n+1} - x\|^2 &\leq \big(\|x_n - x\| + \varepsilon_n\big)^2 \\ &\leq \|x_n - x\|^2 + 2\varepsilon_n \sup_{l\geq 0} \|x_l - x\| + \varepsilon_n^2.\end{aligned} \tag{24}$$

Hence, since $(\forall x \in S)\ \sup_{l\geq 0} \|x_l - x\| < +\infty$ by Lemma 3.1(i) and $(\varepsilon_n)_{n\geq 0} \in \ell^1 \subset \ell^2$, (4) holds. To show (ii), observe that (24) yields

$$(\forall x \in S)(\forall n \in \mathbb{N})\ \ \|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 + 2\varepsilon_n \sup_{z\in S}\sup_{l\geq 0} \|x_l - z\| + \varepsilon_n^2. \tag{25}$$

Therefore, if S is bounded, $\sup_{z\in S}\sup_{l\geq 0}\|x_l - z\| < +\infty$ and (3) ensues. □

Our next proposition collects some basic properties of quasi-Fejér sequences of Type III.

Proposition 3.3 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type III relative to a nonempty set S in $\mathcal{H}$. Then*

(i) *$(x_n)_{n\geq 0}$ is bounded.*

(ii) *$(x_n)_{n\geq 0}$ is quasi-Fejér of Type III relative to* $\operatorname{conv} S$.

(iii) *For every $x \in \overline{\operatorname{conv}}\, S$, $(\|x_n - x\|)_{n\geq 0}$ converges.*

(iv) *For every $(x, x') \in (\overline{\operatorname{conv}}\, S)^2$, $(\langle x_n \mid x - x'\rangle)_{n\geq 0}$ converges.*

Proof. Suppose that $(x_n)_{n\geq 0}$ satisfies (4). (i) is a direct consequence of Lemma 3.1(i). (ii): Take $x \in \operatorname{conv} S$, say $x = \alpha y_1 + (1-\alpha)y_2$, where $(y_1, y_2) \in S^2$ and $\alpha \in [0, 1]$. Then there exist two sequences $(\varepsilon_{1,n})_{n\geq 0}$ and $(\varepsilon_{2,n})_{n\geq 0}$ in $\ell_+ \cap \ell^1$ such that

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \|x_{n+1} - y_1\|^2 \leq \|x_n - y_1\|^2 + \varepsilon_{1,n} \\ \|x_{n+1} - y_2\|^2 \leq \|x_n - y_2\|^2 + \varepsilon_{2,n}. \end{cases} \tag{26}$$

Now put $(\forall n \in \mathbb{N})\ \ \varepsilon_n = \max\{\varepsilon_{1,n}, \varepsilon_{2,n}\}$. Then $(\varepsilon_n)_{n\geq 0} \in \ell_+ \cap \ell^1$ and

$$\begin{aligned}(\forall n \in \mathbb{N}) \ \ \|x_{n+1} - x\|^2 &= \|\alpha(x_{n+1} - y_1) + (1-\alpha)(x_{n+1} - y_2)\|^2 \\ &= \alpha\|x_{n+1} - y_1\|^2 + (1-\alpha)\|x_{n+1} - y_2\|^2 - \alpha(1-\alpha)\|y_1 - y_2\|^2 \\ &\leq \alpha\|x_n - y_1\|^2 + (1-\alpha)\|x_n - y_2\|^2 - \alpha(1-\alpha)\|y_1 - y_2\|^2 + \varepsilon_n \\ &= \|\alpha(x_n - y_1) + (1-\alpha)(x_n - y_2)\|^2 + \varepsilon_n \\ &= \|x_n - x\|^2 + \varepsilon_n. \end{aligned} \tag{27}$$

(iii): Start with $y \in \operatorname{conv} S$, say $y = \alpha y_1 + (1-\alpha)y_2$, where $(y_1, y_2) \in S^2$ and $\alpha \in [0, 1]$. Then $(\|x_n - y_1\|)_{n\geq 0}$ and $(\|x_n - y_2\|)_{n\geq 0}$ converge by Lemma 3.1(ii) and so does $(\|x_n - y\|)_{n\geq 0}$ since

$$(\forall n \in \mathbb{N}) \ \ \|x_n - y\|^2 = \alpha\|x_n - y_1\|^2 + (1-\alpha)\|x_n - y_2\|^2 - \alpha(1-\alpha)\|y_1 - y_2\|^2. \tag{28}$$

Next, take $x \in \overline{\operatorname{conv}}\, S$, say $y_k \to x$ where $(y_k)_{k\geq 0}$ lies in $\operatorname{conv} S$. It remains to show that $(\|x_n - x\|)_{n\geq 0}$ converges. As just shown, for every $k \in \mathbb{N}$, $(\|x_n - y_k\|)_{n\geq 0}$ converges. Moreover,

$$\begin{aligned}(\forall k \in \mathbb{N}) \quad -\|y_k - x\| &\leq \underline{\lim}\, \|x_n - x\| - \lim{}_n \|x_n - y_k\| \\ &\leq \overline{\lim}\, \|x_n - x\| - \lim{}_n \|x_n - y_k\| \\ &\leq \|y_k - x\|. \end{aligned} \tag{29}$$

Taking the limit as $k \to +\infty$, we conclude that $\lim_n \|x_n - x\| = \lim_k \lim_n \|x_n - y_k\|$. (iv): Fix $(x, x') \in (\overline{\operatorname{conv}}\, S)^2$. Then

$$(\forall n \in \mathbb{N}) \ \ \langle x_n \mid x - x'\rangle = \big(\|x_n - x'\|^2 - \|x_n - x\|^2 - \|x - x'\|^2\big)/2 + \langle x \mid x - x'\rangle. \tag{30}$$

However, as the right-hand side converges by (iii), we obtain the claim. □

Not unexpectedly, sharper statements can be formulated for quasi-Fejér sequences of Types I and II.

Proposition 3.4 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type II relative to a nonempty set S in $\mathcal{H}$. Then $(x_n)_{n\geq 0}$ quasi-Fejér of Type II relative to $\overline{\operatorname{conv}}\, S$.*

Proof. Suppose that $(x_n)_{n\geq 0}$ satisfies (3). By arguing as in the proof of Proposition 3.3(ii), we obtain that $(x_n)_{n\geq 0}$ is quasi-Fejér of Type II relative to $\operatorname{conv} S$ with the same error sequence $(\varepsilon_n)_{n\geq 0}$. Now take $x \in \overline{\operatorname{conv}}\, S$, say $y_k \to x$ where $(y_k)_{n\geq 0}$ lies in $\operatorname{conv} S$. Then, for every $n \in \mathbb{N}$, we obtain

$$(\forall k \in \mathbb{N})\ \ \|x_{n+1} - y_k\|^2 \leq \|x_n - y_k\|^2 + \varepsilon_n \tag{31}$$

and, upon taking the limit as $k \to +\infty$, $\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 + \varepsilon_n$. □

A Fejér monotone sequence $(x_n)_{n\geq 0}$ relative to a nonempty set S may not converge, even weakly: a straightforward example is the sequence $((-1)^n x)_{n\geq 0}$ which is Fejér monotone with respect to $S = \{0\}$ and which does not converge for any $x \notin S$. Nonetheless, if S is closed and convex, the projected sequence $(P_S x_n)_{n\geq 0}$ always converges strongly [8, Thm. 2.16(iv)], [24, Prop. 3] (see also [65, Rem. 1], where this result appears in connection with a fixed point problem). We now show that quasi-Fejér sequence of Types I and II also enjoy this remarkable property.

Proposition 3.5 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type I relative to a nonempty set S in $\mathcal{H}$ with error sequence $(\varepsilon_n)_{n\geq 0}$. Then the following properties hold.*

(i) $(\forall n \in \mathbb{N})\ \ d_S(x_{n+1}) \leq d_S(x_n) + \varepsilon_n$.

(ii) $(d_S(x_n))_{n\geq 0}$ *converges.*

(iii) *If* $(\exists\, \chi \in \left]0,1\right[)(\forall n \in \mathbb{N})\ \ d_S(x_{n+1}) \leq \chi d_S(x_n) + \varepsilon_n$, *then* $(d_S(x_n))_{n\geq 0} \in \ell^1$.

(iv) *If S is closed and convex, then*

  (a) $(x_n)_{n\geq 0}$ *is quasi-Fejér of Type II relative to the set* $\{P_S x_n\}_{n\geq 0}$ *with error sequence* $(\varepsilon'_n)_{n\geq 0}$, *where*

$$(\forall n \in \mathbb{N})\ \ \varepsilon'_n = 2\varepsilon_n \sup_{(l,k)\in\mathbb{N}^2} \|x_l - P_S x_k\| + \varepsilon_n^2. \tag{32}$$

  (b) $(P_S x_n)_{n\geq 0}$ *converges strongly.*

Proof. (i): Take the infimum over $x \in S$ in (2). (i) $\Rightarrow$ (ii): Use Lemma 3.1(ii). (iii): Use Lemma 3.1(iv). (iv): (a): Since $(x_n)_{n\geq 0}$ is bounded by Proposition 3.3(i) and P_S is (firmly) nonexpansive by Proposition 2.2, $\{P_S x_n\}_{n\geq 0}$ is bounded. The claim therefore follows from Proposition 3.2(ii) and (25). (b): By Proposition 2.2, $P_S \in \mathfrak{T}$. Therefore Proposition 2.3(ii) with $\lambda = 1$ yields

$$(\forall (m,n) \in \mathbb{N}^2) \;\; \|P_S x_{n+m} - P_S x_n\|^2 \leq \|x_{n+m} - P_S x_n\|^2 - d_S(x_{n+m})^2. \tag{33}$$

On the other hand, we derive from (a) that

$$(\forall (m,n) \in \mathbb{N}^2) \;\; \|x_{n+m} - P_S x_n\|^2 \leq \|x_n - P_S x_n\|^2 + \sum_{k=n}^{n+m-1} \varepsilon'_k. \tag{34}$$

Upon combining (33) and (34), we obtain

$$(\forall (m,n) \in \mathbb{N}^2) \;\; \|P_S x_{n+m} - P_S x_n\|^2 \leq d_S(x_n)^2 - d_S(x_{n+m})^2 + \sum_{k \geq n} \varepsilon'_k. \tag{35}$$

However, since $(\varepsilon'_n)_{n\geq 0} \in \ell_+ \cap \ell^1$, $\lim \sum_{k\geq n} \varepsilon'_k = 0$. It therefore follows from (ii) that $\lim_{m,n} \|P_S x_{n+m} - P_S x_n\| = 0$, i.e, $(P_S x_n)_{n\geq 0}$ is a Cauchy sequence. ☐

Proposition 3.6 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type II relative to a nonempty set S in $\mathcal{H}$ with error sequence $(\varepsilon_n)_{n\geq 0}$. Then the following properties hold.*

(i) $(\forall n \in \mathbb{N}) \;\; d_S(x_{n+1})^2 \leq d_S(x_n)^2 + \varepsilon_n$.

(ii) $(d_S(x_n))_{n\geq 0}$ *converges.*

(iii) *If* $(\exists\, \chi \in \,]0,1[)(\forall n \in \mathbb{N}) \;\; d_S(x_{n+1})^2 \leq \chi d_S(x_n)^2 + \varepsilon_n$, *then* $(d_S(x_n))_{n\geq 0} \in \ell^2$.

(iv) *If S is closed and convex, then $(P_S x_n)_{n\geq 0}$ converges strongly.*

Proof. Analogous to that of Proposition 3.5, except that $(\varepsilon'_n)_{n\geq 0} = (\varepsilon_n)_{n\geq 0}$ in (iv). ☐

3.2. Weak convergence

The following proposition records some elementary weak topology properties of quasi-Fejér sequences.

Proposition 3.7 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type III relative to a nonempty set S in $\mathcal{H}$. Then*

(i) $\mathfrak{W}(x_n)_{n\geq 0} \neq \varnothing$.

(ii) $\big(\forall (x,x') \in (\mathfrak{W}(x_n)_{n\geq 0})^2\big)(\exists\, \alpha \in \mathbb{R}) \;\; S \subset \big\{y \in \mathcal{H} \mid \langle y \mid x - x'\rangle = \alpha\big\}$.

(iii) *If $\overline{\operatorname{aff}}\, S = \mathcal{H}$ (for instance $\operatorname{int} S \neq \varnothing$), then $(x_n)_{n\geq 0}$ converges weakly.*

(iv) *If $x_n \rightharpoonup x \in \overline{\operatorname{conv}}\, S$, then $(\|x_n - y\|)_{n\geq 0}$ converges for every $y \in \mathcal{H}$.*

Proof. (i) follows from Proposition 3.3(i). (ii): Take two points x and x' in $\mathfrak{W}(x_n)_{n\geq 0}$, say $x_{k_n} \rightharpoonup x$ and $x_{l_n} \rightharpoonup x'$, and $y \in S$. Since

$$(\forall n \in \mathbb{N}) \;\; \|x_n - y\|^2 - \|y\|^2 = \|x_n\|^2 - 2\langle y \mid x_n\rangle, \tag{36}$$

it follows from Proposition 3.3(iii) that $\beta = \lim \|x_n - y\|^2 - \|y\|^2$ is well defined. Therefore

$$\beta = \lim \|x_{k_n}\|^2 - 2\langle y \mid x\rangle = \lim \|x_{l_n}\|^2 - 2\langle y \mid x'\rangle \tag{37}$$

and we obtain the desired inclusion with $\alpha = (\lim \|x_{k_n}\|^2 - \lim \|x_{l_n}\|^2)/2$. (iii): In view of (ii), if $\overline{\operatorname{aff}}\, S = \mathcal{H}$ then

$$\big(\forall (x,x') \in (\mathfrak{W}(x_n)_{n\geq 0})^2\big)(\exists\, \alpha \in \mathbb{R})(\forall y \in \mathcal{H}) \;\; \langle y \mid x - x'\rangle = \alpha. \tag{38}$$

Consequently $\mathfrak{W}(x_n)_{n\geq 0}$ reduces to a singleton. Since $(x_n)_{n\geq 0}$ lies in a weakly compact set by virtue of Proposition 3.3(i), it therefore converges weakly. (iv): Take $y \in \mathcal{H}$. Then the identities

$$(\forall n \in \mathbb{N}) \;\; \|x_n - y\|^2 = \|x_n - x\|^2 + 2\langle x_n - x \mid x - y\rangle + \|x - y\|^2 \tag{39}$$

together with Proposition 3.3(iii) imply that $(\|x_n - y\|)_{n\geq 0}$ converges. □

The following fundamental result has been known for Fejér monotone sequences for some time [13, Lem. 6]. In the present context, it appears in [2, Prop. 1.3].

Theorem 3.8 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type III relative to a nonempty set S in $\mathcal{H}$. Then $(x_n)_{n\geq 0}$ converges weakly to a point in S if and only if $\mathfrak{W}(x_n)_{n\geq 0} \subset S$.*

Proof. Necessity is straightforward. To show sufficiency, suppose $\mathfrak{W}(x_n)_{n\geq 0} \subset S$ and take x and x' in $\mathfrak{W}(x_n)_{n\geq 0}$. Since $(x, x') \in S^2$, Proposition 3.7(ii) asserts that $\langle x \mid x - x'\rangle = \langle x' \mid x - x'\rangle$ (this identity could also be derived from Proposition 3.3(iv)), whence $x = x'$. In view of Proposition 3.3(i), the proof is complete. □

3.3. Strong convergence

There are known instances of Fejér monotone sequences which converge weakly but not strongly to a point in the target set [7], [9], [38], [42]. A simple example is the following: any orthonormal sequence $(x_n)_{n\geq 0}$ in $\mathcal{H}$ is Fejér monotone relative to $\{0\}$ and, by Bessel's inequality, satisfies $x_n \rightharpoonup 0$; however, $1 \equiv \|x_n\| \not\to 0$. The strong convergence properties of quasi-Fejér sequences must therefore be investigated in their own rights. We begin this investigation with some facts regarding the strong cluster points of quasi-Fejér sequences of Type III. The first two of these facts were essentially known to Ermol'ev [32].

Proposition 3.9 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type III relative to a nonempty set S in $\mathcal{H}$. Then*

(i) $\big(\forall (x,x') \in (\mathfrak{S}(x_n)_{n\geq 0})^2\big) \;\; S \subset \big\{y \in \mathcal{H} \mid \langle y - (x+x')/2 \mid x - x'\rangle = 0\big\}$.

(ii) *If* $\overline{\operatorname{aff}} S = \mathcal{H}$ *(for instance* $\operatorname{int} S \neq \varnothing$*), then* $\mathfrak{S}(x_n)_{n\geq 0}$ *contains at most one point.*

(iii) $(x_n)_{n\geq 0}$ *converges strongly if there exist* $x \in S$, $(\varepsilon_n)_{n\geq 0} \in \ell_+ \cap \ell^1$, *and* $\rho \in \,]0, +\infty[$ *such that*

$$(\forall n \in \mathbb{N}) \;\; \|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - \rho\|x_{n+1} - x_n\| + \varepsilon_n. \tag{40}$$

Proof. (i): Take x and x' in $\mathfrak{S}(x_n)_{n\geq 0}$, say $x_{k_n} \to x$ and $x_{l_n} \to x'$, and $y \in S$. Then $\lim \|x_{k_n} - y\| = \|x - y\|$ and $\lim \|x_{l_n} - y\| = \|x' - y\|$. Hence, by Proposition 3.3(iii), $\|x - y\| = \|x' - y\|$ or, equivalently, $\langle y - (x + x')/2 \mid x - x'\rangle = 0$. Since $\mathfrak{S}(x_n)_{n\geq 0} \subset \mathfrak{W}(x_n)_{n\geq 0}$, this identity could also be obtained through Proposition 3.7(ii) where $\alpha = (\|x\|^2 - \|x'\|^2)/2$. (ii) follows from (i) or, alternatively, from Proposition 3.7(iii). (iii): By virtue of Lemma 3.1(iii), $(\|x_{n+1} - x_n\|)_{n\geq 0} \in \ell^1$ and $(x_n)_{n\geq 0}$ is therefore a Cauchy sequence. □

We now extend to quasi-Fejér sequences of Types I and II a strong convergence property that was first identified in the case of Fejér sequences in [64] (see also [8, Thm. 2.16(iii)] and the special cases appearing in [53] and [55, Sec. 6]).

Proposition 3.10 *Let* $(x_n)_{n\geq 0}$ *be a quasi-Fejér sequence of Type I or II relative to a set* S *in* $\mathcal{H}$ *such that* $\operatorname{int} S \neq \varnothing$. *Then* $(x_n)_{n\geq 0}$ *converges strongly.*

Proof. Take $x \in S$ and $\rho \in \,]0, +\infty[$ such that $B(x, \rho) \subset S$. Proposition 3.2(ii) asserts that $(x_n)_{n\geq 0}$ is quasi-Fejér of Type II relative to the bounded set $B(x, \rho)$. Hence,

$$(\exists\, (\varepsilon_n)_{n\geq 0} \in \ell_+ \cap \ell^1)(\forall z \in B(x, \rho))(\forall n \in \mathbb{N}) \;\; \|x_{n+1} - z\|^2 \leq \|x_n - z\|^2 + \varepsilon_n. \tag{41}$$

Now define a sequence $(z_n)_{n\geq 0}$ in $B(x, \rho)$ by

$$(\forall n \in \mathbb{N}) \;\; z_n = \begin{cases} x & \text{if } \; x_{n+1} = x_n \\ x - \rho \dfrac{x_{n+1} - x_n}{\|x_{n+1} - x_n\|} & \text{otherwise.} \end{cases} \tag{42}$$

Then (41) yields $(\forall n \in \mathbb{N}) \;\; \|x_{n+1} - z_n\|^2 \leq \|x_n - z_n\|^2 + \varepsilon_n$ and, after expanding, we obtain

$$(\forall n \in \mathbb{N}) \;\; \|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - 2\rho\|x_{n+1} - x_n\| + \varepsilon_n. \tag{43}$$

The strong convergence of $(x_n)_{n\geq 0}$ then follows from Proposition 3.9(iii). □

For quasi-Fejér sequences, a number of properties are equivalent to strong convergence to a point in the target set. Such equivalences were already implicitly established in [41] for Fejér monotone sequences relative to closed convex sets (see also [8] and [24]).

Theorem 3.11 *Let* $(x_n)_{n\geq 0}$ *be a quasi-Fejér sequence of Type III relative to a nonempty set* S *in* $\mathcal{H}$. *Then the following statements are equivalent:*

(i) $(x_n)_{n\geq 0}$ *converges strongly to a point in* S.

(ii) $\mathfrak{W}(x_n)_{n\geq 0} \subset S$ *and* $\mathfrak{S}(x_n)_{n\geq 0} \neq \varnothing$.

(iii) $\mathfrak{S}(x_n)_{n\geq 0} \cap S \neq \varnothing$.

If S *is closed and* $(x_n)_{n\geq 0}$ *is quasi-Fejér of Type I or II relative to* S*, each of the above statements is equivalent to*

(iv) $\underline{\lim}\, d_S(x_n) = 0$.

Proof. (i) ⇒ (ii): Clearly, $x_n \to x \in S \Rightarrow \mathfrak{W}(x_n)_{n\geq 0} = \mathfrak{S}(x_n)_{n\geq 0} = \{x\} \subset S$. (ii) ⇒ (iii): Indeed, $\mathfrak{S}(x_n)_{n\geq 0} \subset \mathfrak{W}(x_n)_{n\geq 0}$. (iii) ⇒ (i): Fix $x \in \mathfrak{S}(x_n)_{n\geq 0} \cap S$. Then $x \in \mathfrak{S}(x_n)_{n\geq 0}$ $\Rightarrow \underline{\lim}\, \|x_n - x\| = 0$. On the other hand, $x \in S$, and it follows from Proposition 3.3(iii) that $(\|x_n - x\|)_{n\geq 0}$ converges. Thus, $x_n \to x$.
Now assume that S is closed. (iv) ⇒ (i): If $(x_n)_{n\geq 0}$ is quasi-Fejér of Type I with error sequence $(\varepsilon_n)_{n\geq 0}$ then

$$(\forall x \in S)(\forall (m,n) \in \mathbb{N}^2)\ \ \|x_n - x_{n+m}\| \ \leq\ \|x_n - x\| + \|x_{n+m} - x\| \ \leq\ 2\|x_n - x\| + \sum_{k=n}^{n+m-1} \varepsilon_k \tag{44}$$

and therefore

$$(\forall (m,n) \in \mathbb{N}^2)\ \ \|x_n - x_{n+m}\| \ \leq\ 2d_S(x_n) + \sum_{k\geq n} \varepsilon_k. \tag{45}$$

Likewise, if $(x_n)_{n\geq 0}$ is quasi-Fejér of Type II with error sequence $(\varepsilon_n)_{n\geq 0}$ then

$$(\forall x \in S)(\forall (m,n) \in \mathbb{N}^2)\ \ \|x_n - x_{n+m}\|^2 \ \leq\ 2\big(\|x_n - x\|^2 + \|x_{n+m} - x\|^2\big) \ \leq\ 4\|x_n - x\|^2 + 2\sum_{k=n}^{n+m-1} \varepsilon_k \tag{46}$$

and therefore

$$(\forall (m,n) \in \mathbb{N}^2)\ \ \|x_n - x_{n+m}\|^2 \leq 4d_S(x_n)^2 + 2\sum_{k\geq n} \varepsilon_k. \tag{47}$$

Now suppose $\underline{\lim}\, d_S(x_n) = 0$. Then Propositions 3.5(ii) and 3.6(ii) yield $\lim d_S(x_n) = 0$ in both cases. On the other hand, by summability, $\lim \sum_{k\geq n} \varepsilon_k = 0$ and we derive from (45) and (47) that $(x_n)_{n\geq 0}$ is a Cauchy sequence in both cases. It therefore converges strongly to some point $x \in \mathcal{H}$. By continuity of d_S, we deduce that $d_S(x) = 0$, i.e., $x \in \overline{S} = S$. (i) ⇒ (iv): Indeed, $(\forall x \in S)(\forall n \in \mathbb{N})\ \ d_S(x_n) \leq \|x_n - x\|$. □

Remark 3.12 With the additional assumption that S is convex, the implication (iv) ⇒ (i) can be established more directly. Indeed, Propositions 3.5(ii) and 3.6(ii) yield $x_n - P_S x_n \to 0$ while Propositions 3.5(iv)(b) and 3.6(iv) guarantee the existence of a point $x \in S$ such that $P_S x_n \to x$. Altogether, $x_n \to x$.

3.4. Convergence estimates

In order to compare algorithms or devise stopping criteria for them, it is convenient to have estimates of their speed of convergence. For quasi-Fejér sequences of Type I or II it is possible to derive such estimates.

Theorem 3.13 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type I [resp. Type II] relative to a nonempty set S in $\mathcal{H}$ with error sequence $(\varepsilon_n)_{n\geq 0}$. Then*

(i) *If $(x_n)_{n\geq 0}$ converges strongly to a point $x \in S$ then*

$$\begin{aligned}&(\forall n \in \mathbb{N})\ \ \|x_n - x\| \leq 2d_S(x_n) + \sum_{k\geq n}\varepsilon_k\\&\big[\text{resp. } (\forall n \in \mathbb{N})\ \ \|x_n - x\|^2 \leq 4d_S(x_n)^2 + 2\sum_{k\geq n}\varepsilon_k\big].\end{aligned} \tag{48}$$

(ii) *If S is closed and*

$$\begin{aligned}&(\exists\, \chi \in \,]0,1[)(\forall n \in \mathbb{N})\ \ d_S(x_{n+1}) \leq \chi d_S(x_n) + \varepsilon_n,\\&\big[\text{resp. } (\exists\, \chi \in \,]0,1[)(\forall n \in \mathbb{N})\ \ d_S(x_{n+1})^2 \leq \chi d_S(x_n)^2 + \varepsilon_n\big],\end{aligned} \tag{49}$$

then $(x_n)_{n\geq 0}$ converges strongly to a point $x \in S$ and

$$\begin{aligned}&(\forall n \in \mathbb{N} \smallsetminus \{0\})\ \ \|x_n - x\| \leq 2\chi^n d_S(x_0) + 2\sum_{k=0}^{n-1}\chi^{n-k-1}\varepsilon_k + \sum_{k\geq n}\varepsilon_k\\&\big[\text{resp. } (\forall n \in \mathbb{N} \smallsetminus \{0\})\ \ \|x_n - x\|^2 \leq 4\chi^n d_S(x_0)^2 + 4\sum_{k=0}^{n-1}\chi^{n-k-1}\varepsilon_k + 2\sum_{k\geq n}\varepsilon_k\big].\end{aligned} \tag{50}$$

Proof. (i): Take the limit as $m \to +\infty$ in (45) [resp. in (47)]. (ii): It follows from (49) and Proposition 3.5(iii) [resp. Proposition 3.6(iii)] that $\lim d_S(x_n) = 0$. Therefore, by Theorem 3.11, there exists a point $x \in S$ such that $x_n \to x$. For every $n \in \mathbb{N}$, we obtain from (22) the estimate $d_S(x_{n+1}) \leq \chi^{n+1}d_S(x_0) + \sum_{k=0}^{n}\chi^{n-k}\varepsilon_k$ [resp. $d_S(x_{n+1})^2 \leq \chi^{n+1}d_S(x_0)^2 + \sum_{k=0}^{n}\chi^{n-k}\varepsilon_k$] which, together with (i), provides (50). □

For Type I sequences, item (i) appears in [47, Thm. 5.3]. Item (ii) owes its relevance to the fact that the right-hand side of (50) converges to zero as $n \to +\infty$ since $(\varepsilon_n)_{n\geq 0} \in \ell_+ \cap \ell^1$ and $\chi \in \,]0,1[$ (as seen in the proof of Lemma 3.1(iv), its two first terms belong to ℓ^1). Sharper estimates require additional assumptions on $(\varepsilon_n)_{n\geq 0}$.

Corollary 3.14 *Let $(x_n)_{n\geq 0}$ be a quasi-Fejér sequence of Type I relative to a nonempty closed and convex set S in $\mathcal{H}$ with error sequence $(\varepsilon_n)_{n\geq 0}$, and let $(\varepsilon'_n)_{n\geq 0}$ be as in (32). Then*

(i) *If $(x_n)_{n\geq 0}$ converges strongly to a point $x \in S$ then*

$$(\forall n \in \mathbb{N})\ \ \|x_n - x\|^2 \leq 4d_S(x_n)^2 + 2\sum_{k\geq n}\varepsilon'_k. \tag{51}$$

(ii) *If* $(\exists\, \chi \in \left]0,1\right[)(\forall n \in \mathbb{N})\ d_S(x_{n+1})^2 \leq \chi d_S(x_n)^2 + \varepsilon'_n$, *then* $(x_n)_{n\geq 0}$ *converges strongly to a point* $x \in S$ *and*

$$(\forall n \in \mathbb{N} \smallsetminus \{0\})\ \ \|x_n - x\|^2 \leq 4\chi^n d_S(x_0)^2 + 4\sum_{k=0}^{n-1} \chi^{n-k-1}\varepsilon'_k + 2\sum_{k\geq n} \varepsilon'_k. \tag{52}$$

Proof. The claim follows from Proposition 3.5(iv)(a) and Theorem 3.13 since $(\forall n \in \mathbb{N})$ $d_S(x_n) = d_{\{P_S x_k\}_{k\geq 0}}(x_n)$. ☐

In the case of Fejér monotone sequences, Corollary 3.14 captures well-known results that originate in [41] (see also [8] and [24]). Thus, (i) furnishes the estimate $(\forall n \in \mathbb{N})$ $\|x_n - x\| \leq 2d_S(x_n)$ while (ii) states that if

$$(\exists\, \chi \in \left]0,1\right[)(\forall n \in \mathbb{N})\ \ d_S(x_{n+1}) \leq \chi d_S(x_n), \tag{53}$$

then $(x_n)_{n\geq 0}$ converges linearly to a point in S: $(\forall n \in \mathbb{N})\ \ \|x_n - x\| \leq 2\chi^n d_S(x_0)$.

4. ANALYSIS OF AN INEXACT $\mathfrak{T}$-CLASS ALGORITHM

Let $S \subset \mathcal{H}$ be the set of solutions to a given problem and let T_n be an operator in $\mathfrak{T}$ such that $\operatorname{Fix} T_n \supset S$. Then, for every point x_n in $\mathcal{H}$ and every relaxation parameter $\lambda_n \in [0,2]$, Proposition 2.3(ii) guarantees that $x_n + \lambda_n(T_n x_n - x_n)$ is not further from any solution point than x_n is. This remark suggests that a point in S can be constructed via the iterative scheme $x_{n+1} = x_n + \lambda_n(T_n x_n - x_n)$. Since in some problems one may not want – or be able – to evaluate $T_n x_n$ exactly, a more realistic algorithmic model is obtained by replacing $T_n x_n$ by $T_n x_n + e_n$, where e_n accounts for some numerical error.

Algorithm 4.1 At iteration $n \in \mathbb{N}$, suppose that $x_n \in \mathcal{H}$ is given. Then select $T_n \in \mathfrak{T}$, $\lambda_n \in [0,2]$, and set $x_{n+1} = x_n + \lambda_n(T_n x_n + e_n - x_n)$, where $e_n \in \mathcal{H}$.

The convergence analysis of Algorithm 4.1 will be greatly simplified by the following result, which states that its orbits are quasi-Fejér relative to the set of common fixed points of the operators $(T_n)_{n\geq 0}$.

Proposition 4.2 *Suppose that* $F = \bigcap_{n\geq 0} \operatorname{Fix} T_n \neq \varnothing$ *and let* $(x_n)_{n\geq 0}$ *be an arbitrary orbit of Algorithm 4.1 such that* $(\lambda_n \|e_n\|)_{n\geq 0} \in \ell^1$. *Then*

(i) $(x_n)_{n\geq 0}$ *is quasi-Fejér of Type I relative to* F *with error sequence* $(\lambda_n \|e_n\|)_{n\geq 0}$.

(ii) $\big(\lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2\big)_{n\geq 0} \in \ell^1$.

(iii) *If* $\overline{\lim}\, \lambda_n < 2$, *then* $(\|x_{n+1} - x_n\|)_{n\geq 0} \in \ell^2$.

Proof. Fix $x \in F$ and put, for every $n \in \mathbb{N}$, $z_n = x_n + \lambda_n(T_n x_n - x_n)$. (i): For every $n \in \mathbb{N}$, $x \in \operatorname{Fix} T_n$ and, since $T_n \in \mathfrak{T}$, Proposition 2.3(ii) yields

$$\|z_n - x\|^2 \leq \|x_n - x\|^2 - \lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2. \tag{54}$$

Whence,

$$(\forall n \in \mathbb{N})\ \ \|z_n - x\| \leq \|x_n - x\| \tag{55}$$

and therefore

$$(\forall n \in \mathbb{N})\ \ \|x_{n+1} - x\| \leq \|z_n - x\| + \lambda_n\|e_n\| \leq \|x_n - x\| + \lambda_n\|e_n\|, \tag{56}$$

which shows that $(x_n)_{n\geq 0}$ satisfies (2). (ii): Set

$$(\forall n \in \mathbb{N})\ \ \varepsilon'_n(x) = 2\lambda_n\|e_n\| \sup_{l\geq 0} \|x_l - x\| + \lambda_n^2\|e_n\|^2. \tag{57}$$

Then it follows from the assumption $(\lambda_n\|e_n\|)_{n\geq 0} \in \ell^1$ and from Proposition 3.3(i) that $(\varepsilon'_n(x))_{n\geq 0} \in \ell^1$. Using (56), (54), and (55), we obtain

$$\begin{aligned}(\forall n \in \mathbb{N})\ \ \|x_{n+1} - x\|^2 &\leq \big(\|z_n - x\| + \lambda_n\|e_n\|\big)^2 \\ &\leq \|x_n - x\|^2 - \lambda_n(2 - \lambda_n)\|T_n x_n - x_n\|^2 + \varepsilon'_n(x)\end{aligned} \tag{58}$$

and Lemma 3.1(iii) allows us to conclude $(\lambda_n(2 - \lambda_n)\|T_n x_n - x_n\|^2)_{n\geq 0} \in \ell^1$. (iii): By assumption, there exist $\delta \in \left]0,1\right[$ and $N \in \mathbb{N}$ such that $(\lambda_n)_{n\geq N}$ lies in $[0, 2-\delta]$. Hence, for every $n \geq N$, $\lambda_n \leq (2-\delta)(2-\lambda_n)/\delta$ and, in turn,

$$\begin{aligned}\|x_{n+1} - x_n\|^2 &\leq \big(\lambda_n\|T_n x_n - x_n\| + \lambda_n\|e_n\|\big)^2 \\ &\leq 2\lambda_n^2\|T_n x_n - x_n\|^2 + 2\lambda_n^2\|e_n\|^2 \\ &\leq \frac{2(2-\delta)}{\delta}\lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2 + 2\lambda_n^2\|e_n\|^2.\end{aligned} \tag{59}$$

In view of (ii) and the fact that $(\lambda_n\|e_n\|)_{n\geq 0} \in \ell^2$, the proof is complete. □

We are now ready to prove

Theorem 4.3 *Suppose that* $\varnothing \neq S \subset \bigcap_{n\geq 0} \operatorname{Fix} T_n$ *and let* $(x_n)_{n\geq 0}$ *be an arbitrary orbit of Algorithm 4.1. Then* $(x_n)_{n\geq 0}$ *converges weakly to a point* x *in* S *if*

(i) $(\lambda_n\|e_n\|)_{n\geq 0} \in \ell^1$ *and* $\mathfrak{W}(x_n)_{n\geq 0} \subset S$.

The convergence is strong if any of the following assumptions is added:

(ii) S *is closed and* $\underline{\lim}\, d_S(x_n) = 0$.

(iii) $\operatorname{int} S \neq \varnothing$.

(iv) *There exist a strictly increasing sequence* $(k_n)_{n\geq 0}$ *in* $\mathbb{N}$ *and an operator* $T\colon \mathcal{H} \to \mathcal{H}$ *which is demicompact at* 0 *such that* $(\forall n \in \mathbb{N})$ $T_{k_n} = T$ *and* $\sum_{n\geq 0} \lambda_{k_n}(2 - \lambda_{k_n}) = +\infty$.

(v) *S is closed and convex, $(\lambda_n)_{n\geq 0}$ lies in $[\delta, 2-\delta]$, where $\delta \in \,]0,1[$, and*

$$(\exists\, \chi \in \,]0,1])(\forall n \in \mathbb{N}) \;\; \|T_n x_n - x_n\| \geq \chi d_S(x_n). \tag{60}$$

In this case, for every integer $n \geq 1$, we have

$$\|x_n - x\|^2 \leq 4(1-\delta^2\chi^2)^n d_S(x_0)^2 + 4\sum_{k=0}^{n-1}(1-\delta^2\chi^2)^{n-k-1}\varepsilon'_k + 2\sum_{k\geq n}\varepsilon'_k, \tag{61}$$

where $\varepsilon'_k = 2\lambda_k\|e_k\| \sup_{(l,m)\in\mathbb{N}^2}\|x_l - P_S x_m\| + \lambda_k^2\|e_k\|^2$.

Proof. First, we recall from Propositions 4.2(i) and 3.2(i) that $(x_n)_{n\geq 0}$ is quasi-Fejér of Types I and III relative to S. Hence, (i) is a direct consequence of Theorem 3.8. We now turn to strong convergence. (ii) follows from Theorem 3.11. (iii) is supplied by Proposition 3.10. (iv): Proposition 4.2(ii) yields

$$\sum_{n\geq 0}\lambda_{k_n}(2-\lambda_{k_n})\|Tx_{k_n} - x_{k_n}\|^2 < +\infty. \tag{62}$$

Since $\sum_{n\geq 0}\lambda_{k_n}(2-\lambda_{k_n}) = +\infty$, it therefore follows that $\underline{\lim}\,\|Tx_{k_n} - x_{k_n}\| = 0$. Hence, the demicompactness of T at 0 gives $\mathfrak{S}(x_n)_{n\geq 0} \neq \varnothing$, and the conclusion follows from Theorem 3.11. (v): The assumptions imply

$$(\forall n \in \mathbb{N}) \;\; \lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2 \geq \delta^2\chi^2 d_S(x_n)^2. \tag{63}$$

Strong convergence therefore follows from Proposition 4.2(ii) and (ii). On the other hand, (58) yields

$$\begin{aligned}(\forall (k,n) \in \mathbb{N}^2) \;\; \|x_{n+1} - P_S x_k\|^2 \;\; &\leq \;\; \|x_n - P_S x_k\|^2 \\ &\qquad - \lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2 + \varepsilon'_n\end{aligned} \tag{64}$$

where, just as in Proposition 3.5(iv)(a), $(\varepsilon'_n)_{n\geq 0} \in \ell^1$. Hence, we derive from (63) and (64) that

$$\begin{aligned}(\forall n \in \mathbb{N}) \;\; d_S(x_{n+1})^2 \;\; &\leq \;\; \|x_{n+1} - P_S x_n\|^2 \\ &\leq \;\; \|x_n - P_S x_n\|^2 - \lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2 + \varepsilon'_n \\ &\leq \;\; (1-\delta^2\chi^2)d_S(x_n)^2 + \varepsilon'_n.\end{aligned} \tag{65}$$

Corollary 3.14(ii) can now be invoked to complete the proof. □

Remark 4.4 The condition $(\lambda_n\|e_n\|)_{n\geq 0} \in \ell^1$ cannot be removed in Theorem 4.3. Indeed take $(\forall n \in \mathbb{N})$ $T_n = \mathrm{Id}$, $\lambda_n = 1$, and $e_n = x_0/(n+1)$ in Algorithm 4.1. Then $\|x_n\| \to +\infty$.

Remark 4.5 Suppose that $e_n \equiv 0$. Then Algorithm 4.1 describes all Fejér monotone methods [9, Prop. 2.7]. As seen in Theorem 4.3, stringent conditions must be imposed to achieve strong convergence to a point in the target set with such methods. In [9], a slight modification of Algorithm 4.1 was proposed that renders them strongly convergent without requiring any additional restrictions on the constituents of the problem.

5. APPLICATIONS TO PERTURBED OPTIMIZATION ALGORITHMS

Let S be a nonempty set in $\mathcal{H}$ representing the set of solutions to an optimization problem. One of the assumptions in Theorem 4.3 is that $S \subset \bigcap_{n\geq 0} \operatorname{Fix} T_n$. This assumption is certainly satisfied if

$$(\forall n \in \mathbb{N}) \;\; \operatorname{Fix} T_n = S. \tag{66}$$

In this section, we shall consider applications of Algorithm 4.1 in which (66) is fulfilled. In view of Proposition 2.3(v), S is therefore closed and convex.

5.1. Krasnosel'skiĭ-Mann iterates

Let T be an operator in $\mathfrak{T}$ with at least one fixed point. To find such a point, we shall use a perturbed version of the Krasnosel'skiĭ-Mann iterates.

Algorithm 5.1 At iteration $n \in \mathbb{N}$, suppose that $x_n \in \mathcal{H}$ is given. Then select $\lambda_n \in [0, 2]$, and set $x_{n+1} = x_n + \lambda_n(Tx_n + e_n - x_n)$, where $e_n \in \mathcal{H}$.

Theorem 5.2 *Let $(x_n)_{n\geq 0}$ be an arbitrary orbit of Algorithm 5.1. Then $(x_n)_{n\geq 0}$ converges weakly to a fixed point x of T if*

(i) $(\|e_n\|)_{n\geq 0} \in \ell^1$, *$T - \operatorname{Id}$ is demiclosed at 0, and $(\lambda_n)_{n\geq 0}$ lies in $[\delta, 2-\delta]$ for some $\delta \in \,]0, 1[$.*

The convergence is strong if any of the following assumptions is added:

(ii) $\underline{\lim}\, d_{\operatorname{Fix} T}(x_n) = 0$.

(iii) $\operatorname{int} \operatorname{Fix} T \neq \varnothing$.

(iv) *T is demicompact at 0.*

(v) $(\exists\, \chi \in \,]0, 1])(\forall n \in \mathbb{N}) \;\; \|Tx_n - x_n\| \geq \chi d_{\operatorname{Fix} T}(x_n)$. *In this case, the convergence estimate (61) holds with $S = \operatorname{Fix} T$.*

Proof. It is clear that Algorithm 5.1 is a special case of Algorithm 4.1 and that Theorem 5.2 will follow from Theorem 4.3 if "$T - \operatorname{Id}$ is demiclosed at 0 and $(\lambda_n)_{n\geq 0}$ lies in $[\delta, 2-\delta]$" $\Rightarrow$ $\mathfrak{W}(x_n)_{n\geq 0} \subset \operatorname{Fix} T$. To show this, take $x \in \mathfrak{W}(x_n)_{n\geq 0}$, say $x_{k_n} \rightharpoonup x$. Since $(\forall n \in \mathbb{N})\; \delta \leq \lambda_{k_n} \leq 2-\delta \Rightarrow \lambda_{k_n}(2-\lambda_{k_n}) \geq \delta^2$, it follows from Proposition 4.2(ii) that $(T - \operatorname{Id})x_{k_n} \to 0$. The demiclosedness of $T - \operatorname{Id}$ at 0 therefore implies $x \in \operatorname{Fix} T$. ☐

Remark 5.3 For T firmly nonexpansive and $\lambda_n \equiv 1$, Theorem 5.2(i) is stated in [49, Rem. 5.6.4].

5.2. Successive approximations with a nonexpansive operator

Let $R\colon \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator with $\operatorname{dom} R = \mathcal{H}$ and $\operatorname{Fix} R \neq \varnothing$. Then a fixed point of R can be obtained via Theorem 5.2 with $T = (\mathrm{Id} + R)/2$, which is firmly nonexpansive [39, Thm. 12.1] (hence in $\mathfrak{T}$ by Proposition 2.2 and, furthermore, nonexpansive so that $T - \mathrm{Id}$ is demiclosed on $\mathcal{H}$ by Remark 2.6) with $\operatorname{Fix} R = \operatorname{Fix} T$. This substitution amounts to halving the relaxations in Algorithm 5.1 and leads to

Algorithm 5.4 At iteration $n \in \mathbb{N}$, suppose that $x_n \in \mathcal{H}$ is given. Then select $\lambda_n \in [0,1]$, and set $x_{n+1} = x_n + \lambda_n(Rx_n + e_n - x_n)$, where $e_n \in \mathcal{H}$.

A direct application of Theorem 5.2 would require the relaxation parameters to be bounded away from 0 and 1. We show below that the nonexpansivity of R allows for a somewhat finer relaxation strategy.

Theorem 5.5 *Let $(x_n)_{n\geq 0}$ be an arbitrary orbit of Algorithm 5.4. Then $(x_n)_{n\geq 0}$ converges weakly to a fixed point of R if*

(i) $(\lambda_n\|e_n\|)_{n\geq 0} \in \ell^1$ *and* $\big(\lambda_n(1-\lambda_n)\big)_{n\geq 0} \notin \ell^1$.

The convergence is strong if any of the following assumptions is added:

(ii) $\underline{\lim}\, d_{\operatorname{Fix} R}(x_n) = 0$.

(iii) $\operatorname{int} \operatorname{Fix} R \neq \varnothing$.

(iv) *R is demicompact at* 0.

Proof. $R - \mathrm{Id}$ is demiclosed on $\mathcal{H}$ by Remark 2.6 and an inspection of the proof of Theorem 5.2 shows that it is sufficient to demonstrate that

$$Rx_n - x_n \to 0. \tag{67}$$

By Proposition 4.2(ii) $\sum_{n\geq 0}\lambda_n(1-\lambda_n)\|Rx_n - x_n\|^2 < +\infty$. Hence, $\sum_{n\geq 0}\lambda_n(1-\lambda_n) = +\infty$ $\Rightarrow \underline{\lim}\,\|Rx_n - x_n\| = 0$. However,

$$(\forall n \in \mathbb{N}) \quad Rx_{n+1} - x_{n+1} = Rx_{n+1} - Rx_n + (1-\lambda_n)(Rx_n - x_n) - \lambda_n e_n \tag{68}$$

and therefore

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \|Rx_{n+1} - x_{n+1}\| &\leq \|Rx_{n+1} - Rx_n\| + (1-\lambda_n)\|Rx_n - x_n\| + \lambda_n\|e_n\| \\ &\leq \|x_{n+1} - x_n\| + (1-\lambda_n)\|Rx_n - x_n\| + \lambda_n\|e_n\| \\ &\leq \|Rx_n - x_n\| + 2\lambda_n\|e_n\|. \end{aligned} \tag{69}$$

Consequently, as $\sum_{n\geq 0}\lambda_n\|e_n\| < +\infty$, Lemma 3.1(ii) secures the convergence of the sequence $(\|Rx_n - x_n\|)_{n\geq 0}$ and (67) is established. □

Remark 5.6 When $e_n \equiv 0$, Theorem 5.5 is related to several known results:

- The weak convergence statement corresponds to the Hilbert space version of [40, Cor. 3] (see also [65]).
- For constant relaxation parameters, strong convergence under condition (ii) covers the Hilbert space version of [58, Cor. 2.2] and strong convergence under condition (iv) corresponds to the Hilbert space version of [58, Cor. 2.1].
- Strong convergence under condition (iii) was obtained in [53] and [64] by replacing "$\big(\lambda_n(1-\lambda_n)\big)_{n\geq 0}\notin \ell^1$" by "$\lambda_n\equiv 1$" in (i).

Remark 5.7 Theorem 5.5 authorizes nonsummable error sequences. For instance, for n large, suppose that $\|e_n\|\leq (1+\sqrt{1-1/n})/n^{\kappa}$, where $\kappa\in\,]0,1]$, and that the relaxation rule is $\lambda_n=(1-\sqrt{1-1/n})/2$. Then $\sum_{n\geq 0}\|e_n\|$ may diverge but $\sum_{n\geq 0}\lambda_n\|e_n\|<+\infty$ and $\sum_{n\geq 0}\lambda_n(1-\lambda_n)=+\infty$.

5.3. Gradient method

In the error-free case ($e_n\equiv 0$), it was shown in [24] that convergence results could be derived from Theorem 5.5 for a number of algorithms, including the Forward-Backward and Douglas-Rachford methods for finding a zero of the sum of two monotone operators, the prox method for solving variational inequalities, and, in particular, the projected gradient method. Theorem 5.5 therefore provides convergence results for perturbed versions of these algorithms. As an illustration, this section is devoted to the case of the perturbed gradient method. A different analysis of the perturbed gradient method can be found in [51].

Consider the unconstrained minimization problem

$$\text{Find}\;\; x\in\mathcal{H}\;\;\text{such that}\;\; f(x)=\bar{f},\;\;\text{where}\;\; \bar{f}=\inf f(\mathcal{H}). \tag{70}$$

The standing assumption is that $f\colon\mathcal{H}\to\mathbb{R}$ is a continuous convex function and that the set S of solutions of (70) is nonempty, as is the case when f is coercive; it is also assumed that f is differentiable and that, for some $\alpha\in\,]0,+\infty[$, $\alpha\nabla f$ is firmly nonexpansive (it follows from [6, Cor. 10] that this is equivalent to saying that ∇f is $(1/\alpha)$-Lipschitz, i.e., that $\alpha\nabla f$ is nonexpansive).

Algorithm 5.8 Fix $\gamma\in\,]0,2\alpha]$ and, at iteration $n\in\mathbb{N}$, suppose that $x_n\in\mathcal{H}$ is given. Then select $\lambda_n\in[0,1]$ and set $x_{n+1}=x_n-\lambda_n\gamma\big(\nabla f(x_n)+e_n\big)$, where $e_n\in\mathcal{H}$.

Theorem 5.9 *Let $(x_n)_{n\geq 0}$ be an arbitrary orbit of Algorithm 5.8. Then $(x_n)_{n\geq 0}$ converges weakly to point in S if $(\lambda_n\|e_n\|)_{n\geq 0}\in\ell^1$ and $\big(\lambda_n(1-\lambda_n)\big)_{n\geq 0}\notin\ell^1$.*

Proof. Put $R=\operatorname{Id}-\gamma\nabla f$. Then

$$\begin{aligned}(\forall (x,y)\in\mathcal{H}^2)\;\;\|Rx-Ry\|^2 &= \|x-y\|^2-2\gamma\langle x-y\mid \nabla f(x)-\nabla f(y)\rangle\\ &\quad+\gamma^2\|\nabla f(x)-\nabla f(y)\|^2\\ &\leq \|x-y\|^2-\gamma(2\alpha-\gamma)\|\nabla f(x)-\nabla f(y)\|^2.\end{aligned}\tag{71}$$

Hence R is nonexpansive and Algorithm 5.8 is a special case of Algorithm 5.4. As $\operatorname{Fix} R=(\nabla f)^{-1}(\{0\})=S$, the claim follows from Theorem 5.5(i). ☐

Remark 5.10 Strong convergence conditions can be derived from Theorem 5.5(ii)-(iv). Thus, it follows from item (ii) that weak convergence in Theorem 5.9 can be improved to strong convergence if we add the correctness condition [23], [48]:

$$\lim f(x_n) = \bar{f} \quad \Rightarrow \quad \underline{\lim}\, d_S(x_n) = 0. \tag{72}$$

Indeed, by convexity

$$(\forall x \in S)(\forall n \in \mathbb{N}) \;\; 0 \leq f(x_n) - \bar{f} \leq \langle x_n - x \mid \nabla f(x_n)\rangle \leq \sup_{l\geq 0} \|x_l - x\| \cdot \|\nabla f(x_n)\|. \tag{73}$$

Consequently, with the same notation as in the above proof, it follows from (72) that (67) $\Leftrightarrow \nabla f(x_n) \to 0 \Rightarrow f(x_n) \to \bar{f} \Rightarrow \underline{\lim}\, d_S(x_n) = 0$.

5.4. Inconsistent convex feasibility problems

Let $(S_i)_{i\in I}$ be a finite family of nonempty closed and convex sets in $\mathcal{H}$. A standard convex programming problem is to find a point in the intersection of these sets. In instances when the intersection turns out to be empty, an alternative is to look for a point which is closest to all the sets in a least squared distance sense, i.e., to minimize the *proximity function*

$$f = \frac{1}{2}\sum_{i\in I} \omega_i d_{S_i}^2, \quad \text{where} \quad (\forall i \in I)\; \omega_i > 0 \;\; \text{and} \;\; \sum_{i\in I} \omega_i = 1. \tag{74}$$

The resulting problem is a particular case of (70). We shall denote by S the set of minimizers of f over $\mathcal{H}$ and assume that it is nonempty, as is the case when one of the sets in $(S_i)_{i\in I}$ is bounded since f is then coercive. Naturally, if $\bigcap_{i\in I} S_i \neq \varnothing$, then $S = \bigcap_{i\in I} S_i$.

To solve the (possibly inconsistent) convex feasibility problem (70)/(74), we shall use the following parallel projection algorithm.

Algorithm 5.11 At iteration $n \in \mathbb{N}$, suppose that $x_n \in \mathcal{H}$ is given. Then select $\lambda_n \in [0,2]$ and set $x_{n+1} = x_n + \lambda_n\big(\sum_{i\in I}\omega_i(P_{S_i}x_n + e_{i,n}) - x_n\big)$, where $(e_{i,n})_{i\in I}$ lies in $\mathcal{H}$.

Theorem 5.12 *Let $(x_n)_{n\geq 0}$ be an arbitrary orbit of Algorithm 5.11. Then $(x_n)_{n\geq 0}$ converges weakly to point in S if $(\lambda_n\|\sum_{i\in I}\omega_i e_{i,n}\|)_{n\geq 0} \in \ell^1$ and $\big(\lambda_n(2-\lambda_n)\big)_{n\geq 0} \notin \ell^1$.*

Proof. We have $\nabla f = \sum_{i\in I}\omega_i(\mathrm{Id} - P_{S_i})$. Since the operators $(P_{S_i})_{i\in I}$ are firmly nonexpansive by Proposition 2.2, so are the operators $(\mathrm{Id} - P_{S_i})_{i\in I}$ and, in turn, their convex combination ∇f. Hence, Algorithm 5.11 is a special case of Algorithm 5.8 with $\alpha = 1$, $\gamma = 2$, and $(\forall n \in \mathbb{N})\; e_n = \sum_{i\in I}\omega_i e_{i,n}$. The claim therefore follows from Theorem 5.9. ◻

Remark 5.13 Let us make a couple of comments.

- Theorem 5.12 extends [18, Thm. 4], where $e_{i,n} \equiv 0$ and the relaxations parameters are bounded away from 0 and 2 (see also [27] where constant relaxation parameters are assumed).

- Algorithm 5.8 allows for an error in the evaluation of each projection. As noted in Remark 5.7, the average projection error sequence $(\sum_{i\in I}\omega_i e_{i,n})_{n\geq 0}$ does not have to be absolutely summable.

Remark 5.14 Suppose that the problem is consistent, i.e., $\bigcap_{i\in I} S_i \neq \varnothing$.

- If $e_{i,n}\equiv 0$, $\lambda_n\equiv 1$, and $\omega_i = 1/\operatorname{card} I$, Theorem 5.12 was obtained in [4] (see also [66, Cor. 2.6] for a different perspective).
- If I is infinite (and possibly uncountable), a more general operator averaging process for firmly nonexpansive operators with errors is studied in [35] (see also [16] for an error-free version with projectors).
- If the projections can be computed exactly, a more efficient weakly convergent parallel projection algorithm to find a point in $\bigcap_{i\in I} S_i$ is that proposed by Pierra in [59], [60]. It consists in taking T in Algorithm 5.1 as the operator defined in (14) with $(\forall i\in I)\ T_i = P_{S_i}$ and relaxations parameters in $]0,1]$. The large values achieved by the parameters $(L(x_n))_{n\geq 0}$ result in large step sizes that significantly accelerate the algorithm, as evidenced in various numerical experiments (see Remark 6.2 for specific references). This type of extrapolated scheme was first employed in the parallel projection method of Merzlyakov [52] to solve systems of affine inequalities in $\mathbb{R}^N$; the resulting algorithm was shown to be faster than the sequential projection algorithms of [1] and [54]. An alternative interpretation of Pierra's algorithm is the following: it can be obtained by taking T in Algorithm 5.1 as the subgradient projector defined in (9), where f is the proximity function defined in (74). A generalization of Pierra's algorithm will be proposed in Section 6.1.

5.5. Proximal point algorithm

Many optimization problems – in particular (70) – reduce to the problem of finding a zero of a monotone operator $A\colon \mathcal{H}\to 2^{\mathcal{H}}$, i.e., to the problem

$$\text{Find } x\in\mathcal{H} \text{ such that } 0\in Ax. \tag{75}$$

It will be assumed henceforth that $0\in\operatorname{ran} A$ and that A is maximal monotone.

The following algorithm, which goes back to [50], is known as the (relaxed) inexact proximal point algorithm.

Algorithm 5.15 At iteration $n\in\mathbb{N}$, suppose that $x_n\in\mathcal{H}$ is given. Then select $\lambda_n\in[0,2]$, $\gamma_n\in\,]0,+\infty[$, and set $x_{n+1} = x_n + \lambda_n\big((\mathrm{Id}+\gamma_n A)^{-1}x_n + e_n - x_n\big)$, where $e_n\in\mathcal{H}$.

Theorem 5.16 *Let $(x_n)_{n\geq 0}$ be an arbitrary orbit of Algorithm 5.15. Then $(x_n)_{n\geq 0}$ converges weakly to point in $A^{-1}0$ if $(\|e_n\|)_{n\geq 0}\in\ell^1$, $\inf_{n\geq 0}\gamma_n > 0$, and $(\lambda_n)_{n\geq 0}$ lies in $[\delta, 2-\delta]$, for some $\delta\in\,]0,1[$.*

Proof. It follows from Proposition 2.2 that Algorithm 5.15 is a special case of Algorithm 4.1. Moreover, $(\forall n\in\mathbb{N})\ \operatorname{Fix}(\mathrm{Id}+\gamma_n A)^{-1} = A^{-1}0$. Hence, in view of Theorem 4.3(i), we need to show $\mathfrak{W}(x_n)_{n\geq 0}\subset A^{-1}0$. For every $n\in\mathbb{N}$, define $y_n = (\mathrm{Id}\,+$

$\gamma_n A)^{-1}x_n$ and $v_n = (x_n - y_n)/\gamma_n$, and observe that $(y_n, v_n) \in \operatorname{gr} A$. In addition, since $\inf_{n\geq 0} \lambda_n(2-\lambda_n) \geq \delta^2$, it follows from Proposition 4.2(ii) that $x_n - y_n \to 0$. Therefore, since $\inf_{n\geq 0} \gamma_n > 0$, we get $v_n \to 0$. Now take $x \in \mathfrak{W}(x_n)_{n\geq 0}$, say $x_{k_n} \rightharpoonup x$. Then $y_{k_n} \rightharpoonup x$ and $v_{k_n} \to 0$. However, as A is maximal monotone, $\operatorname{gr} A$ is weakly-strongly closed, which forces $0 \in Ax$. □

Remark 5.17 Theorem 5.16 can be found in [28, Thm. 3] and several related results can be found in the literature.

- The unrelaxed version (i.e., $\lambda_n \equiv 1$) is due to Rockafellar [68, Thm. 1]. There, it was also proved that $x_{n+1} - x_n \to 0$. This fact follows immediately from Proposition 4.2(iii).
- Perturbed proximal point algorithms are also investigated in [3], [12], [14], [44], [46], and [55].

Remark 5.18 As shown in [42], an orbit of the proximal point algorithm may converge weakly but not strongly to a solution point. In this regard, two comments should be made.

- Strong convergence conditions can be derived from Theorem 4.3(ii)-(v). Thus, the convergence is strong in Theorem 5.16 in each of the following cases:
  - $\sum_{n\geq 0} \|(\operatorname{Id} + \gamma_n A)^{-1}x_n - x_n\|^2 < +\infty \Rightarrow \underline{\lim}\, d_{A^{-1}0}(x_n) = 0$. This condition follows immediately from item (ii). For accretive operators in nonhilbertian Banach spaces and $\lambda_n \equiv 1$, a similar condition was obtained in [55, Sec. 4].
  - $\operatorname{int} A^{-1}0 \neq \varnothing$. This condition follows immediately from item (iii) and can be found in [55, Sec. 6].
  - $\operatorname{dom} A$ is boundedly compact. This condition follows from item (iv) if $(\gamma_n)_{n\geq 0}$ contains a constant subsequence and, more generally, from the argument given in the proof of Theorem 6.9(iv).

  Additional conditions will be found in [12] and [68].
- A relatively minor modification of the proximal point algorithm makes it strongly convergent without adding any specific condition on A. See [71] and Remark 4.5 for details.

6. APPLICATIONS TO BLOCK-ITERATIVE PARALLEL ALGORITHMS

6.1. The algorithm

A common feature of the algorithms described in Section 5 is that $(\forall n \in \mathbb{N})\ \operatorname{Fix} T_n = S$. These algorithms therefore implicitly concern applications in which the target set S is relatively simple. In many applications, however, the target set is not known explicitly but merely described as a countable (finite or countably infinite) intersection of closed convex sets $(S_i)_{i\in I}$ in $\mathcal{H}$. The underlying problem can then be recast in the form of the *countable convex feasibility problem*

$$\text{Find } x \in S = \bigcap_{i\in I} S_i. \tag{76}$$

Here, the tacit assumption is that for every index $i \in I$ it is possible to construct relatively easily at iteration n an operator $T_{i,n} \in \mathfrak{T}$ such that $\operatorname{Fix} T_{i,n} = S_i$. Thus, S is not dealt with directly but only through its supersets $(S_i)_{i\in I}$. In infinite dimensional spaces, a classical method fitting in this framework is Bregman's periodic projection algorithm [11] which solves (76) iteratively in the case $I = \{1,\dots,m\}$ via the sequential algorithm

$$(\forall n \in \mathbb{N}) \;\; x_{n+1} = P_{S_{n \,(\mathrm{mod}\, m)+1}} x_n. \tag{77}$$

As discussed in Remark 5.14, an alternative method to solve this problem is Auslender's parallel projection scheme [4]

$$(\forall n \in \mathbb{N}) \;\; x_{n+1} = \frac{1}{m}\sum_{i=1}^{m} P_{S_i} x_n. \tag{78}$$

Bregman's method utilizes only one set at each iteration while Auslender's utilizes all of them simultaneously and is therefore inherently parallel. In this respect, these two algorithms stand at opposite ends in the more general class of parallel block-iterative algorithms, where at iteration n the update is formed by averaging projections of the current iterate onto a block of sets $(S_i)_{i\in I_n\subset I}$. The practical advantage of such a scheme is to provide a flexible means to match the computational load of each iteration to the distributed computer resources at hand.

The first block parallel projection algorithm in a Hilbert space setting was proposed by Ottavy [56] with further developments in [15] and [22]. Variants and extensions of (77) and (78) involving more general operators such as subgradient projectors, nonexpansive and firmly nonexpansive operators have also been investigated [5], [13], [17], [36], [63], [72] and unified in the form of block-iterative algorithms at various levels of generality in [8], [19], [21], and [43]. For recent extensions of (77) in other directions, see [67] and the references therein.

Building upon the framework developed in [8], a general block-iterative scheme was proposed in [45, Algo. 2.1] to bring together the algorithms mentioned above. An essentially equivalent algorithm was later devised in [23, Algo. 7.1] within a different framework. The following algorithm employs yet another framework, namely the $\mathfrak{T}$ operator class, and, in addition, allows for errors in the computation of each operator.

Algorithm 6.1 Fix $(\delta_1, \delta_2) \in \,]0,1[^2$ and $x_0 \in \mathcal{H}$. At every iteration $n \in \mathbb{N}$,

$$x_{n+1} = x_n + \lambda_n L_n \left(\sum_{i\in I_n} \omega_{i,n}(T_{i,n}x_n + e_{i,n}) - x_n \right) \tag{79}$$

where:

① $\varnothing \neq I_n \subset I$, I_n finite.

② $(\forall i \in I_n)$ $T_{i,n} \in \mathfrak{T}$ and $\operatorname{Fix} T_{i,n} = S_i$.

③ $(\forall i \in I_n)$ $e_{i,n} \in \mathcal{H}$ and $e_{i,n} = 0$ if $x_n \in S_i$.

④ $(\forall i \in I_n)$ $\omega_{i,n} \in [0,1]$, $\sum_{i \in I_n} \omega_{i,n} = 1$, and

$$(\exists\, j \in I_n) \begin{cases} \|T_{j,n}x_n - x_n\| = \max_{i \in I_n} \|T_{i,n}x_n - x_n\| \\ \omega_{j,n} \geq \delta_1. \end{cases}$$

⑤ $\lambda_n \in [\delta_2/L_n, 2-\delta_2]$, where

$$L_n = \begin{cases} \dfrac{\sum_{i \in I_n} \omega_{i,n}\|T_{i,n}x_n - x_n\|^2}{\|\sum_{i \in I_n} \omega_{i,n}T_{i,n}x_n - x_n\|^2} & \text{if } x_n \notin \bigcap_{i \in I_n} S_i \text{ and } \sum_{i \in I_n} \omega_{i,n}\|e_{i,n}\| = 0, \\ 1 & \text{otherwise.} \end{cases}$$

Remark 6.2 The incorporation of errors in the above recursion calls for some comments.

- The vector $e_{i,n}$ stands for the error made in computing $T_{i,n}x_n$. With regard to the convergence analysis, the global error term at iteration n is $\lambda_n \sum_{i \in I_n} \omega_{i,n}e_{i,n}$. Thus, the individual errors $(e_{i,n})_{i \in I_n}$ are naturally averaged and can be further controlled by the relaxation parameter λ_n.

- If $e_{i,n} \equiv 0$, Algorithm 6.1 essentially relapses to [23, Algo. 7.1] and [45, Algo. 2.1]. If we further assume that at every iteration n the index set I_n is a singleton, then it reduces to the exact $\mathfrak{T}$-class sequential method of [9, Algo. 2.8].

- If, for some index $j \in I_n$, it is possible to verify that $\|T_{j,n}x_n - x_n\| \neq \max_{i \in I_n} \|T_{i,n}x_n - x_n\|$, the associated error $e_{j,n}$ can be neutralized by setting $\omega_{j,n} = 0$.

- Suppose that $\sum_{i \in I_n} \omega_{i,n}\|e_{i,n}\| = 0$, meaning that for each selected index i, either $T_{i,n}x_n$ is computed exactly or the associated error $e_{i,n}$ is neutralized (see previous item). Then extrapolated relaxations up to $(2-\delta_2)L_n$ can be used, where L_n can attain very large values. In numerical experiments, this type of extrapolated overrelaxations has been shown to induce very fast convergence [20], [21], [25], [37], [52], [60], [61].

6.2. Convergence

Let us first recall a couple of useful concepts.

Definition 6.3 [19] The control sequence $(I_n)_{n\geq 0}$ in Algorithm 6.1 is *admissible* if there exist strictly positive integers $(M_i)_{i\in I}$ such that

$$(\forall (i,n)\in I\times\mathbb{N})\ \ i\in\bigcup_{k=n}^{n+M_i-1} I_k. \tag{80}$$

Definition 6.4 [8, Def. 3.7] Algorithm 6.1 is *focusing* if for every index $i\in I$ and every generated suborbit $(x_{k_n})_{n\geq 0}$,

$$\begin{cases} i\in\bigcap_{n\geq 0} I_{k_n} \\ x_{k_n}\rightharpoonup x \\ T_{i,k_n}x_{k_n}-x_{k_n}\to 0 \end{cases} \quad\Rightarrow\quad x\in S_i. \tag{81}$$

The notion of a focusing algorithm can be interpreted as an extension of the notion of demiclosedness at 0. Along the same lines, it is convenient to introduce the following extension of the notion of demicompactness at 0.

Definition 6.5 Algorithm 6.1 is *demicompactly regular* if there exists an index $i\in I$ such that, for every generated suborbit $(x_{k_n})_{n\geq 0}$,

$$\begin{cases} i\in\bigcap_{n\geq 0} I_{k_n} \\ \sup_{n\geq 0}\|x_{k_n}\|<+\infty \\ T_{i,k_n}x_{k_n}-x_{k_n}\to 0 \end{cases} \quad\Rightarrow\quad \mathfrak{S}(x_{k_n})_{n\geq 0}\neq\varnothing. \tag{82}$$

Such an index is an *index of demicompact regularity.*

The most relevant convergence properties of Algorithm 6.1 are summarized below. This theorem appears to be the first general result on the convergence of inexact block-iterative methods for convex feasibility problems.

Theorem 6.6 *Suppose that $S\neq\varnothing$ in (76) and let $(x_n)_{n\geq 0}$ be an arbitrary orbit of Algorithm 6.1. Then $(x_n)_{n\geq 0}$ converges weakly to a point in S if*

(i) $\big(\lambda_n\|\sum_{i\in I_n}\omega_{i,n}e_{i,n}\|\big)_{n\geq 0}\in\ell^1$*, Algorithm 6.1 is focusing, and the control sequence $(I_n)_{n\geq 0}$ is admissible.*

The convergence is strong if any of the following assumptions is added:

(ii) *Algorithm 6.1 is demicompactly regular.*

(iii) $\operatorname{int} S\neq\varnothing$.

(iv) *There exists a suborbit $(x_{k_n})_{n\geq 0}$ and a sequence $(\chi_n)_{n\geq 0}\in\ell_+\smallsetminus\ell^2$ such that*

$$(\forall n\in\mathbb{N})\ \ \max_{i\in I_{k_n}}\|T_{i,k_n}x_{k_n}-x_{k_n}\|^2\geq\chi_n d_S(x_{k_n}). \tag{83}$$

Proof. We proceed in several steps. Throughout the proof, y is a fixed point in S and $\beta(y) = \sup_{n\geq 0} \|x_n - y\|$. If $\beta(y) = 0$, all the statements are trivially true; we therefore assume otherwise.

Step 1: Algorithm 6.1 is a special instance of Algorithm 4.1.
Indeed, for every $n \in \mathbb{N}$, we can write (79) as

$$x_{n+1} = x_n + \lambda_n(T_n x_n + e_n - x_n), \tag{84}$$

where

$$e_n = \sum_{i\in I_n} \omega_{i,n} e_{i,n} \tag{85}$$

and

$$T_n x_n = x_n + L_n \left(\sum_{i\in I_n} \omega_{i,n} T_{i,n} x_n - x_n \right). \tag{86}$$

It follows from the definition of L_n in ⑤ that the operator T_n takes one of two forms, namely

$$\begin{cases} T_n \colon x \mapsto \displaystyle\sum_{i\in I_n} \omega_{i,n} T_{i,n} x & \text{if } \displaystyle\sum_{i\in I_n} \omega_{i,n}\|e_{i,n}\| \neq 0 \\ T_n \colon x \mapsto x + L\big(x, (T_{i,n})_{i\in I_n}, (\omega_{i,n})_{i\in I_n}\big) \left(\displaystyle\sum_{i\in I_n} \omega_{i,n} T_{i,n} x - x \right) & \text{otherwise,} \end{cases} \tag{87}$$

where the function L is defined in (15). In view of ②, ④, Proposition 2.4, and Remark 2.5, we conclude that in both cases $T_n \in \mathfrak{T}$.

Step 2: $S \subset \bigcap_{n\geq 0} \operatorname{Fix} T_n$.
It follows from (76), (80), ②, and Proposition 2.4 that

$$S = \bigcap_{i\in I} S_i = \bigcap_{n\geq 0} \bigcap_{i\in I_n} S_i \subset \bigcap_{n\geq 0} \bigcap_{\substack{i\in I_n \\ \omega_{i,n}>0}} \operatorname{Fix} T_{i,n} = \bigcap_{n\geq 0} \operatorname{Fix} T_n. \tag{88}$$

Step 3: $(\|x_{n+1} - x_n\|)_{n\geq 0} \in \ell^2$.
The claim follows from Step 1, (85), and Proposition 4.2(iii) since ⑤ $\Rightarrow \overline{\lim}\, \lambda_n < 2$.

Step 4: $\lim \max_{i\in I_n} \|T_{i,n} x_n - x_n\| = 0$.
To see this, we use successively ⑤, (86), and the inequality $L_n \geq 1$ to derive

$$\begin{aligned} (\forall n \in \mathbb{N}) \;\; \lambda_n(2 - \lambda_n)\|T_n x_n - x_n\|^2 &\geq \frac{\delta_2^2}{L_n} \|T_n x_n - x_n\|^2 \\ &= \delta_2^2 L_n \left\| \sum_{i\in I_n} \omega_{i,n} T_{i,n} x_n - x_n \right\|^2 \\ &\geq \delta_2^2 \left\| \sum_{i\in I_n} \omega_{i,n} T_{i,n} x_n - x_n \right\|^2. \end{aligned} \tag{89}$$

By virtue of Step 1 and Proposition 4.2(i), $(x_n)_{n\geq 0}$ is a quasi-Fejér sequence of Type I relative to S and therefore $\beta(y) < +\infty$. Moreover, ② implies that, for every $n \in \mathbb{N}$, $(T_{i,n})_{i\in I_n}$ lies in $\mathfrak{T}$ and $y \in \bigcap_{i\in I_n} \operatorname{Fix} T_{i,n}$. Hence, we can argue as in (17) to get

$$(\forall n \in \mathbb{N})\quad \left\|\sum_{i\in I_n} \omega_{i,n} T_{i,n} x_n - x_n\right\| \geq \frac{1}{\beta(y)} \sum_{i\in I_n} \omega_{i,n} \|T_{i,n} x_n - x_n\|^2 \geq \frac{\delta_1}{\beta(y)} \max_{i\in I_n} \|T_{i,n} x_n - x_n\|^2, \tag{90}$$

where the second inequality is deduced from ④. Now, since Proposition 4.2(ii) implies that $\lim \lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2 = 0$, it follows from (89) that $\lim \|\sum_{i\in I_n} \omega_{i,n} T_{i,n} x_n - x_n\| = 0$ and then from (90) that $\lim \max_{i\in I_n} \|T_{i,n} x_n - x_n\| = 0$.

Step 5: $\mathfrak{W}(x_n)_{n\geq 0} \subset S$.
Fix $i \in I$ and $x \in \mathfrak{W}(x_n)_{n\geq 0}$, say $x_{k_n} \rightharpoonup x$. Then it is enough to show $x \in S_i$. By (80), there exist an integer $M_i > 0$ and a strictly increasing sequence $(p_n)_{n\geq 0}$ in $\mathbb{N}$ such that

$$(\forall n \in \mathbb{N})\quad k_n \leq p_n \leq k_n + M_i - 1 \quad \text{and} \quad i \in I_{p_n}. \tag{91}$$

Hence, by Cauchy-Schwarz,

$$(\forall n \in \mathbb{N})\quad \|x_{p_n} - x_{k_n}\| \leq \sum_{l=k_n}^{k_n+M_i-2} \|x_{l+1} - x_l\| \leq \sqrt{M_i - 1}\sqrt{\sum_{l\geq k_n} \|x_{l+1} - x_l\|^2} \tag{92}$$

and Step 3 yields $x_{p_n} - x_{k_n} \to 0$. Thus, $x_{p_n} \rightharpoonup x$, while $i \in \bigcap_{n\geq 0} I_{p_n}$ and, by Step 4, $T_{i,p_n} x_{p_n} - x_{p_n} \to 0$. Therefore, (81) forces $x \in S_i$.

Step 6: Weak convergence.
Combine Steps 1, 2, and 5, Theorem 4.3(i), and (85).

Step 7: Strong convergence.
(ii): Let i be an index of demicompact regularity. According to (80), there exists a strictly increasing sequence $(k_n)_{n\geq 0}$ in $\mathbb{N}$ such that $i \in \bigcap_{n\geq 0} I_{k_n}$, where i is an index of demicompact regularity, and Step 4 implies $T_{i,k_n} x_{k_n} - x_{k_n} \to 0$. Since $(x_{k_n})_{n\geq 0}$ is bounded, (82) yields $\mathfrak{S}(x_{k_n})_{n\geq 0} \neq \varnothing$. Therefore, strong convergence results from Step 5 and Theorem 3.11. (iii) follows from Step 1 and Theorem 4.3(iii). (iv) follows from Theorem 4.3(ii). Indeed, using (89), (90), and (83), we get

$$\left(\frac{\beta(y)}{\delta_1\delta_2}\right)^2 \sum_{n\geq 0} \lambda_{k_n}(2-\lambda_{k_n})\|T_{k_n} x_{k_n} - x_{k_n}\|^2 \geq \sum_{n\geq 0} \chi_n^2 d_S(x_{k_n})^2. \tag{93}$$

Hence, since $\sum_{n\geq 0} \lambda_n(2-\lambda_n)\|T_n x_n - x_n\|^2 < +\infty$ by Proposition 4.2(ii) and $(\chi_n)_{n\geq 0} \notin \ell^2$ by assumption, we conclude that $\underline{\lim}\, d_S(x_{k_n}) = 0$. ☐

6.3. Application to a mixed convex feasibility problem

Let $(f_i)_{i\in I^{(1)}}$ be a family of continuous convex functions from $\mathcal{H}$ into $\mathbb{R}$, $(R_i)_{i\in I^{(2)}}$ a family of firmly nonexpansive operators with domain $\mathcal{H}$ and into $\mathcal{H}$, and $(A_i)_{i\in I^{(3)}}$ a family of maximal monotone operators from $\mathcal{H}$ into $2^{\mathcal{H}}$. Here, $I^{(1)}$, $I^{(2)}$, and $I^{(3)}$ are possibly empty, countable index sets.

In an attempt to unify a wide class of problems, we consider the *mixed convex feasibility problem*

$$\text{Find } x \in \mathcal{H} \text{ such that } \begin{cases} (\forall i \in I^{(1)}) & f_i(x) \leq 0 \\ (\forall i \in I^{(2)}) & R_i x = x \\ (\forall i \in I^{(3)}) & 0 \in A_i x, \end{cases} \tag{94}$$

under the standing assumption that it is consistent. Problem (94) can be expressed as

$$\text{Find } x \in S = \bigcap_{i \in I} S_i, \text{ where } I = I^{(1)} \cup I^{(2)} \cup I^{(3)}$$
$$\text{and } (\forall i \in I) \ S_i = \begin{cases} \operatorname{lev}_{\leq 0} f_i & \text{if } i \in I^{(1)} \\ \operatorname{Fix} R_i & \text{if } i \in I^{(2)} \\ A_i^{-1} 0 & \text{if } i \in I^{(3)}. \end{cases} \tag{95}$$

In this format, it is readily seen to be a special case of problem (76) (the closedness and convexity of S_i in each case follows from well known facts).

To solve (94), we shall draw the operators $T_{i,n}$ in Algorithm 6.1 from a pool of subgradient projectors, firmly nonexpansive operators, and resolvents. Since such operators conform to ② in Algorithm 6.1 by Proposition 2.2, this choice is legitimate.

Algorithm 6.7 In Algorithm 6.1 set for every $i \in I$ the operators $(T_{i,n})_{n \geq 0}$ as follows.

- If $i \in I^{(1)}$, $(\forall n \in \mathbb{N})$ $T_{i,n} = G_{f_i}^{g_i}$, where g_i is a selection of ∂f_i (see (9)).
- If $i \in I^{(2)}$, $(\forall n \in \mathbb{N})$ $T_{i,n} = R_i$.
- If $i \in I^{(3)}$, $(\forall n \in \mathbb{N})$ $T_{i,n} = (\mathrm{Id} + \gamma_{i,n} A_i)^{-1}$, where $\gamma_{i,n} \in \left]0, +\infty\right[$.

The next assumption will ensure that Algorithm 6.7 is well behaved asymptotically.

Assumption 6.8 The subdifferentials $(\partial f_i)_{i \in I^{(1)}}$ map bounded sets into bounded sets and, for every $i \in I^{(3)}$ and every strictly increasing sequence $(k_n)_{n \geq 0}$ in $\mathbb{N}$ such that $i \in \bigcap_{n \geq 0} I_{k_n}$, $\inf_{n \geq 0} \gamma_{i,k_n} > 0$.

Theorem 6.9 *Suppose that $S \neq \varnothing$ in (95) and let $(x_n)_{n \geq 0}$ be an arbitrary orbit of Algorithm 6.7. Then $(x_n)_{n \geq 0}$ converges weakly to a point in S if*

(i) $\big(\lambda_n \| \sum_{i \in I_n} \omega_{i,n} e_{i,n} \|\big)_{n \geq 0} \in \ell^1$, *Assumption 6.8 is satisfied, and the control sequence $(I_n)_{n \geq 0}$ is admissible.*

The convergence is strong if any of the following assumptions is added:

(ii) *For some $i \in I^{(1)}$ and some $\eta \in \left]0, +\infty\right[$, $\operatorname{lev}_{\leq \eta} f_i$ is boundedly compact.*

(iii) *For some* $i \in I^{(2)}$, R_i *is demicompact at* 0.

(iv) *For some* $i \in I^{(3)}$, $\operatorname{dom} A_i$ *is boundedly compact.*

Proof. Since Algorithm 6.7 is a special case of Algorithm 6.1, we shall derive this theorem from Theorem 6.6. (i): It suffices to show that Assumption 6.8 implies that Algorithm 6.7 is focusing. To this end, fix $i \in I$ and a suborbit $(x_{k_n})_{n\geq 0}$ such that $i \in \bigcap_{n\geq 0} I_{k_n}$, $x_{k_n} \rightharpoonup x$, and $T_{i,k_n}x_{k_n} - x_{k_n} \to 0$. According to (81), we must show $x \in S_i$. Let us consider three cases.

(a) $i \in I^{(1)}$: we must show $f_i(x) \leq 0$. Put $\alpha = \overline{\lim} f(x_{k_n})$. Then, since f is weak lower semicontinuous, $f_i(x) \leq \alpha$ and it is enough to show that $\alpha \leq 0$. Let us extract from $(x_{k_n})_{n\geq 0}$ a subsequence $(x_{l_{k_n}})_{n\geq 0}$ such that $\lim f_i(x_{l_{k_n}}) = \alpha$ and $(\forall n \in \mathbb{N})$ $f_i(x_{l_{k_n}}) > 0$ (if no subsequence exists then clearly $\alpha \leq 0$). Then, by Assumption 6.8, $T_{i,k_n}x_{k_n} - x_{k_n} \to 0 \Rightarrow G_{f_i}^{g_i}x_{l_{k_n}} - x_{l_{k_n}} \to 0 \Rightarrow \lim f_i(x_{l_{k_n}})/\|g_i(x_{l_{k_n}})\| = 0 \Rightarrow \lim f_i(x_{l_{k_n}}) = 0$, where the last implication follows from the boundedness of $(x_{l_{k_n}})_{n\geq 0}$ and Assumption 6.8. Therefore $\alpha \leq 0$.

(b) $i \in I^{(2)}$: we must show $x \in \operatorname{Fix} R_i$. Certainly $T_{i,k_n}x_{k_n} - x_{k_n} \to 0 \Rightarrow (R_i - \operatorname{Id})x_{k_n} \to 0$. Since R_i is firmly nonexpansive, it is nonexpansive. $R_i - \operatorname{Id}$ is therefore demiclosed by Remark 2.6 and the claim ensues.

(c) $i \in I^{(3)}$: we must show $(x, 0) \in \operatorname{gr} A_i$. For every $n \in \mathbb{N}$, define $y_n = (\operatorname{Id} + \gamma_{i,k_n}A_i)^{-1}x_{k_n}$ and $v_n = (x_{k_n} - y_n)/\gamma_{i,k_n}$. Then $\big((y_n, v_n)\big)_{n\geq 0}$ lies in $\operatorname{gr} A_i$ and $T_{i,k_n}x_{k_n} - x_{k_n} \to 0 \Rightarrow y_n - x_{k_n} \to 0 \Rightarrow y_n \rightharpoonup x$. On the other hand, Assumption 6.8 ensures that $y_n - x_{k_n} \to 0 \Rightarrow v_n \to 0$. Since $\operatorname{gr} A_i$ is weakly-strongly closed, we conclude that $(x, 0) \in \operatorname{gr} A_i$.

Let us now show that the three advertised instances of strong convergence yield demicompact regularity and are therefore consequences of Theorem 6.6(ii). Let us fix $i \in I$, a closed ball B, and a suborbit $(x_{k_n})_{n\geq 0}$ such that $i \in \bigcap_{n\geq 0} I_{k_n}$, B contains $(x_{k_n})_{n\geq 0}$, and $T_{i,k_n}x_{k_n} - x_{k_n} \to 0$. We must show $\mathfrak{S}(x_{k_n})_{n\geq 0} \neq \varnothing$. (ii): As shown in (a), $\overline{\lim} f(x_{k_n}) \leq 0$ and therefore the tail of $(x_{k_n})_{n\geq 0}$ lies in the compact set $B \cap \operatorname{lev}_{\leq \eta} f_i$. (iii) is clear. (iv): Define $(y_n)_{n\geq 0}$ as in (c) and recall that $y_n - x_{k_n} \to 0$. Hence $(y_n)_{n\geq 0}$ lies in some closed ball B' and $\mathfrak{S}(x_{k_n})_{n\geq 0} = \mathfrak{S}(y_n)_{n\geq 0}$. Moreover,

$$(\forall n \in \mathbb{N})\ \ y_n \in \operatorname{ran}(\operatorname{Id} + \gamma_{i,k_n}A_i)^{-1} = \operatorname{dom}(\operatorname{Id} + \gamma_{i,k_n}A_i) = \operatorname{dom} A_i. \tag{96}$$

Hence, $(y_n)_{n\geq 0}$ lies in the compact set $B' \cap \operatorname{dom} A_i$ and the desired conclusion ensues. ☐

Remark 6.10 To place the above result in its proper context, a few observations should be made.

- Theorem 6.9 combines and, through the incorporation of errors, generalizes various results on the convergence of block-iterative subgradient projection (for $I^{(2)} = I^{(3)} = \varnothing$) and firmly nonexpansive iteration (for $I^{(1)} = I^{(3)} = \varnothing$) methods [8], [19], [21], [22], [45].

- For $I^{(1)} = I^{(2)} = \varnothing$, the resulting inexact block-iterative proximal point algorithm appears to be new. If, in addition, $I^{(3)}$ is a singleton Theorem 6.9(i) reduces to Theorem 5.16; if we further assume that $\lambda_n \equiv 1$, Theorem 6.9 captures some convergence properties of Rockafellar's inexact proximal point algorithm [68].
- Concerning strong convergence, although we have restricted ourselves to special cases of Theorem 6.6(ii), it is clear that conditions (iii) and (iv) in Theorem 6.6 also apply here. At any rate, these conditions are certainly not exhaustive.
- To recover results on projection algorithms, one can set $(f_i)_{i\in I^{(1)}} = (d_{S_i})_{i\in I^{(1)}}$, $(R_i)_{i\in I^{(2)}} = (P_{S_i})_{i\in I^{(2)}}$, and $(A_i)_{i\in I^{(3)}} = (N_{C_i})_{i\in I^{(3)}}$, where N_{C_i} is the normal cone to S_i.

7. PROJECTED SUBGRADIENT METHOD

The algorithms described in Section 4–6 are quasi-Fejér of Type I. In this section, we shall investigate a class of nonsmooth constrained minimization methods which are quasi-Fejér of Type II. As we shall find, the analysis developed in Section 3 will also be quite useful here to obtain convergence results in a straightforward fashion.

Throughout, $f\colon \mathcal{H} \to \mathbb{R}$ is a continuous convex function, C is a closed convex subset of $\mathcal{H}$, and $\bar{f} = \inf f(C)$. Under consideration is the problem

$$\text{Find } \; x \in C \; \text{ such that } \; f(x) = \bar{f} \tag{97}$$

under the standing assumption that its set S of solutions is nonempty, as is the case when $\operatorname{lev}_{\leq \eta} f \cap C$ is nonempty and bounded for some $\eta \in \mathbb{R}$.

In nonsmooth minimization, the use of projected subgradient methods goes back to the 1960's [70]. Our objective here is to establish convergence results for a class of relaxed, projected approximate subgradient methods in Hilbert spaces. Recall that, given $\epsilon \in [0, +\infty[$, the ϵ-subdifferential of f at $x \in \mathcal{H}$ is obtained by relaxing (5) as follows

$$\partial_\epsilon f(x) = \big\{u \in \mathcal{H} \mid (\forall y \in \mathcal{H}) \;\; \langle y - x \mid u\rangle + f(x) \leq f(y) + \epsilon\big\}. \tag{98}$$

Algorithm 7.1 At iteration $n \in \mathbb{N}$, suppose that $x_n \in \mathcal{H}$ is given. Then select $\alpha_n \in [0, +\infty[$, $\epsilon_n \in [0, +\infty[$, $\lambda_n \in [0, 2]$, $u_n \in \partial_{\epsilon_n} f(x_n)$, and set

$$x_{n+1} = x_n - \alpha_n u_n + \lambda_n\big(P_C(x_n - \alpha_n u_n) - x_n + \alpha_n u_n\big). \tag{99}$$

Let us open the discussion with some basic facts about this algorithm.

Proposition 7.2 *Let $(x_n)_{n\geq 0}$ be an arbitrary orbit of Algorithm 7.1 such that*

$$\Big(\alpha_n^2\|u_n\|^2 + 2\alpha_n\big(\bar{f} - f(x_n)\big)^+ + 2\alpha_n\epsilon_n\Big)_{n\geq 0} \in \ell^1. \tag{100}$$

Then

(i) $(x_n)_{n\geq 0}$ *is quasi-Fejér of Type II relative to* S.

(ii) $\big(\lambda_n(2-\lambda_n)d_C(x_n-\alpha_n u_n)^2\big)_{n\geq 0}\in\ell^1$.

(iii) *If* $(\alpha_n)_{n\geq 0}\notin\ell^1$, $\lim\epsilon_n=0$, *and* $\lim\alpha_n\|u_n\|^2=0$, *then* $\underline{\lim}\, f(x_n)\leq\bar{f}$.

Proof. (i): Fix $x\in S$ and set $(\forall n\in\mathbb{N})\ y_n=x_n-\alpha_n u_n$. Then (98) yields

$$(\forall n\in\mathbb{N})\ \ \|y_n-x\|^2 = \|x_n-x\|^2+\alpha_n^2\|u_n\|^2-2\alpha_n\langle x_n-x\mid u_n\rangle \leq \|x_n-x\|^2+\alpha_n^2\|u_n\|^2+2\alpha_n\big(\bar{f}-f(x_n)+\epsilon_n\big). \tag{101}$$

On the other hand, since $x\in S\subset C=\operatorname{Fix}P_C$ and $P_C\in\mathfrak{T}$ by Proposition 2.2, Proposition 2.3(ii) yields

$$(\forall n\in\mathbb{N})\ \ \|x_{n+1}-x\|^2\leq\|y_n-x\|^2-\lambda_n(2-\lambda_n)d_C(y_n)^2. \tag{102}$$

Altogether

$$(\forall n\in\mathbb{N})\ \ \|x_{n+1}-x\|^2 \leq \|x_n-x\|^2-\lambda_n(2-\lambda_n)d_C(y_n)^2+\alpha_n^2\|u_n\|^2 +2\alpha_n\big(\bar{f}-f(x_n)\big)+2\alpha_n\epsilon_n. \tag{103}$$

In view of (100), $(x_n)_{n\geq 0}$ satisfies (3). (ii) follows from (103), (100), and Lemma 3.1(iii). (iii): We derive from (103) that

$$(\forall n\in\mathbb{N})\ \ \alpha_n\big(2(f(x_n)-\bar{f})-2\epsilon_n-\alpha_n\|u_n\|^2\big)\leq\|x_n-x\|^2-\|x_{n+1}-x\|^2. \tag{104}$$

Hence,

$$\sum_{n\geq 0}\alpha_n\big(2(f(x_n)-\bar{f})-2\epsilon_n-\alpha_n\|u_n\|^2\big)\leq\|x_0-x\|^2 \tag{105}$$

and, since $\sum_{n\geq 0}\alpha_n=+\infty$ by assumption, we get $\underline{\lim}\,2(f(x_n)-\bar{f})-2\epsilon_n-\alpha_n\|u_n\|^2\leq 0$. The remaining assumptions impose $\underline{\lim}\, f(x_n)\leq\bar{f}$. □

The following theorem describes a situation in which weak and strong convergence can be achieved in Algorithm 7.1.

Theorem 7.3 *Let* $(x_n)_{n\geq 0}$ *be an arbitrary orbit of Algorithm 7.1. Then* $(x_n)_{n\geq 0}$ *converges weakly to a point in* S *and* $\lim f(x_n)=\bar{f}$ *if*

(i) $(\alpha_n)_{n\geq 0}\in\ell^\infty\smallsetminus\ell^1$, $(\epsilon_n)_{n\geq 0}\in\ell^1$, *and, furthermore, there exist* $\delta\in\,]0,1[$, $\kappa\in\,]0,+\infty[$, $(\beta_n)_{n\geq 0}\in\ell_+\cap\ell^1$, *and* $(\gamma_n)_{n\geq 0}\in\ell_+\cap\ell^1$ *such that*

$$(\forall n\in\mathbb{N})\ \ \bar{f}-\gamma_{n+1}\leq f(x_{n+1})\leq f(x_n)-\kappa\alpha_n\|u_n\|^2+\beta_n \tag{106}$$

and $(\forall n\in\mathbb{N})\ \delta\leq\lambda_n\leq 2-\delta$.

The convergence is strong if any of the following assumptions is added:

(ii) $\operatorname{int} S\neq\varnothing$.

(iii) *C is boundedly compact.*

(iv) *For some $z \in C$,* $\operatorname{lev}_{\leq f(z)} f$ *is boundedly compact.*

Proof. Set $\zeta = \bar{f} - \sup_{n\geq 0} \gamma_n$. Then (106) yields

$$(\forall n \in \mathbb{N})\ \ 0 \leq f(x_{n+1}) - \zeta \leq \big(f(x_n) - \zeta\big) - \kappa\alpha_n\|u_n\|^2 + \beta_n. \tag{107}$$

Hence, it follows from Lemma 3.1(ii) that $(f(x_n))_{n\geq 0}$ converges and from Lemma 3.1(iii) that $(\alpha_n\|u_n\|^2)_{n\geq 0} \in \ell^1$. Hence, since for every $n \in \mathbb{N} \smallsetminus \{0\}$ $\bar{f} - f(x_n) \leq \gamma_n$,

$$\sum_{n\geq 1} \alpha_n\|u_n\|^2 + 2\big(\bar{f} - f(x_n)\big)^+ + 2\epsilon_n \leq \sum_{n\geq 1} \alpha_n\|u_n\|^2 + 2\gamma_n + 2\epsilon_n < +\infty \tag{108}$$

and we deduce from the boundedness of $(\alpha_n)_{n\geq 0}$ that (100) holds. In view of Proposition 7.2(i), Proposition 3.2(i), and Theorem 3.8, to establish weak convergence to a solution, we must show $\mathfrak{W}(x_n)_{n\geq 0} \subset S$. Fix $x \in \mathfrak{W}(x_n)_{n\geq 0}$, say $x_{k_n} \rightharpoonup x$. Note that the assumptions on $(\lambda_n)_{n\geq 0}$ and Proposition 7.2(ii) imply that $\lim d_C(x_n - \alpha_n u_n) = 0$. On the other hand, since $(\alpha_n)_{n\geq 0} \in \ell^\infty$ and $\lim \alpha_n\|u_n\|^2 = 0$, we have $\lim \alpha_n\|u_n\| = 0$ and, therefore, $x_{k_n} - \alpha_{k_n} u_{k_n} \rightharpoonup x$. However, since C is convex, d_C is weak lower semicontinuous and it follows that $d_C(x) \leq \underline{\lim}\, d_C(x_{k_n} - \alpha_{k_n} u_{k_n}) = 0$. As C is closed, we obtain $x \in C$ and, in turn, $\bar{f} \leq f(x)$. The weak lower semicontinuity of f then yields

$$\bar{f} \leq f(x) \leq \underline{\lim}\, f(x_{k_n}) = \lim f(x_n) \leq \bar{f}, \tag{109}$$

where the last inequality is furnished by Proposition 7.2(iii). Consequently $f(x) = \bar{f} = \lim f(x_n)$. Since $x \in C$, we have thus shown $x \in S$, which completes the proof of (i).

Let us now prove the strong convergence statements. By virtue of Theorem 3.11, since $\mathfrak{W}(x_n)_{n\geq 0} \subset S$, it is enough to show $\mathfrak{S}(x_n)_{n\geq 0} \neq \varnothing$. (ii) Apply Proposition 3.10. (iii): As seen above, $P_C x_n - x_n \to 0$. On the other hand, P_C is demicompact (see Remark 2.6). Hence $\mathfrak{S}(x_n)_{n\geq 0} \neq \varnothing$. (iii): Let B be a closed ball containing $(x_n)_{n\geq 0}$. Since $\lim f(x_n) = \bar{f} \leq f(z)$, the tail of $(x_n)_{n\geq 0}$ lies in the compact set $B \cap \operatorname{lev}_{\leq f(z)} f$ and therefore $\mathfrak{S}(x_n)_{n\geq 0} \neq \varnothing$. □

Remark 7.4 A few comments are in order.

- Theorem 7.3 generalizes [26, Prop. 2.2], in which $\dim \mathcal{H} < +\infty$, $C = \mathcal{H}$ (hence $\gamma_n \equiv 0$), and $\beta_n \equiv 0$.
- The second inequality in (106) with $\beta_n \equiv 0$ was interpreted in [26] as an approximate Armijo rule.
- The first inequality in (106) is trivially satisfied with $\gamma_n \equiv 0$ if $(x_n)_{n\geq 0}$ lies in C. This is true in particular when $\lambda_n \equiv 1$.

Remark 7.5 Suppose that $\inf f(C) > \inf f(\mathcal{H})$ and fix $(\gamma_n)_{n\geq 0} \in \ell_+ \cap \ell^2 \smallsetminus \ell^1$. A classical subgradient projection method is described by the recursion [70]

$$x_0 \in C \;\text{ and }\; (\forall n \in \mathbb{N})\;\; x_{n+1} = P_C\left(x_n - \frac{\gamma_n}{\|u_n\|}u_n\right), \;\text{ where }\; u_n \in \partial f(x_n). \tag{110}$$

This algorithm is readily seen to be a particular implementation of Algorithm 7.1 with $\epsilon_n \equiv 0$, $\lambda_n \equiv 1$, and $\alpha_n = \gamma_n/\|u_n\|$. Moreover, $(\alpha_n\|u_n\|)_{n\geq 0} \in \ell^2$ and, since $(f(x_n))_{n\geq 0}$ lies in $\left[\bar{f}, +\infty\right[$, (100) holds. Consequently, Proposition 7.2(i) asserts that any sequence $(x_n)_{n\geq 0}$ conforming to (110) is quasi-Fejér of Type II relative to S. Moreover, if ∂f maps the bounded subsets of C to bounded sets, then $(\alpha_n)_{n\geq 0} \notin \ell^1$ and $\lim \alpha_n\|u_n\|^2 = 0$. As a result, Proposition 7.2(iii) implies that $\underline{\lim}\, f(x_n) = \bar{f}$. Thus, if C is boundedly compact, we can extract a subsequence $(x_{k_n})_{n\geq 0}$ such that $\lim f(x_{k_n}) = \bar{f}$ and which converges strongly to some point $x \in C$. Hence, $x \in \mathfrak{S}(x_n)_{n\geq 0} \cap S$ and it follows from Theorem 3.11 that $x_n \to x$. For $\dim \mathcal{H} < +\infty$ and $C = \mathcal{H}$, this result was established in [26, Prop. 5.1]. One will also find in [2] a convergence analysis of the ϵ-subgradient version of (110).

Remark 7.6 A third method for solving (97) is Polyak's subgradient method [62], which assumes that $\bar{f}$ is known. It proceeds by alternating a relaxed subgradient projection onto $\operatorname{lev}_{\leq \bar{f}} f$ with a projection onto C and can therefore be regarded as a special case of Algorithm 6.1 with two sets.

REFERENCES

1. S. Agmon, The relaxation method for linear inequalities, *Canadian Journal of Mathematics* **6** (1954) 382–392.
2. Ya. I. Alber, A. N. Iusem, and M. V. Solodov, On the projected subgradient method for non-smooth convex optimization in a Hilbert space, *Mathematical Programming* **81** (1998) 23–35.
3. P. Alexandre, V. H. Nguyen, and P. Tossings, The perturbed generalized proximal point algorithm, *RAIRO Modélisation Mathématique et Analyse Numérique* **32** (1998) 223–253.
4. A. Auslender, *Méthodes Numériques pour la Résolution des Problèmes d'Optimisation avec Contraintes* Doctoral thesis (Université de Grenoble, France, 1969).
5. J.-B. Baillon, *Comportement asymptotique des contractions et semi-groupes de contractions* Doctoral thesis (Université Paris 6, France, 1978).
6. J.-B. Baillon and G. Haddad, Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones, *Israel Journal of Mathematics* **26** (1977) 137–150.
7. H. H. Bauschke, Projection algorithms: Results and open problems, *these proceedings*.
8. H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review* **38** (1996) 367–426.
9. H. H. Bauschke and P. L. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Mathematics of Operations Research* to appear.
10. H. H. Bauschke, F. Deutsch, H. Hundal, and S. H. Park, Fejér monotonicity and weak convergence of an accelerated method of projections, in: M. Théra, ed., *Constructive, Experimental, and Nonlinear Analysis* (Canadian Mathematical Society, 2000) 1–6.

11. L. M. Brègman, The method of successive projection for finding a common point of convex sets, *Soviet Mathematics – Doklady* **6** (1965) 688–692.
12. H. Brézis and P. L. Lions, Produits infinis de résolvantes, *Israel Journal of Mathematics* **29** (1978) 329–345.
13. F. E. Browder, Convergence theorems for sequences of nonlinear operators in Banach spaces, *Mathematische Zeitschrift* **100** (1967) 201–225.
14. R. E. Bruck and S. Reich, Nonexpansive projections and resolvents of accretive operators in Banach spaces, *Houston Journal of Mathematics* **3** (1977) 459–470.
15. D. Butnariu and Y. Censor, Strong convergence of almost simultaneous block-iterative projection methods in Hilbert spaces, *Journal of Computational and Applied Mathematics* **53** (1994) 33–42.
16. D. Butnariu and S. D. Flåm, Strong convergence of expected-projection methods in Hilbert spaces, *Numerical Functional Analysis and Optimization* **16** (1995) 601–636.
17. Y. Censor and A. Lent, Cyclic subgradient projections, *Mathematical Programming* **24** (1982) 233–235.
18. P. L. Combettes, Inconsistent signal feasibility problems: Least-squares solutions in a product space, *IEEE Transactions on Signal Processing* **42** (1994) 2955–2966.
19. P. L. Combettes, Construction d'un point fixe commun à une famille de contractions fermes, *Comptes Rendus de l'Académie des Sciences de Paris* **I320** (1995) 1385–1390.
20. P. L. Combettes, The convex feasibility problem in image recovery, in: P. Hawkes, ed., *Advances in Imaging and Electron Physics* **95** (1996) 155–270.
21. P. L. Combettes, Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections, *IEEE Transactions on Image Processing* **6** (1997) 493–506.
22. P. L. Combettes, Hilbertian convex feasibility problem: Convergence of projection methods, *Applied Mathematics and Optimization* **35** (1997) 311–330.
23. P. L. Combettes, Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM Journal on Control and Optimization* **38** (2000) 538–565.
24. P. L. Combettes, Fejér monotonicity in convex optimization, in: C. A. Floudas and P. M. Pardalos, eds., *Encyclopedia of Optimization* (Kluwer, Boston, MA, 2001).
25. P. L. Combettes and H. Puh, Extrapolated projection method for the euclidean convex feasibility problem, research report (City University of New York, 1993).
26. R. Correa and C. Lemaréchal, Convergence of some algorithms for convex minimization, *Mathematical Programming* **62** (1993) 261–275.
27. A. R. De Pierro and A. N. Iusem, A parallel projection method for finding a common point of a family of convex sets, *Pesquisa Operacional* **5** (1985) 1–20.
28. J. Eckstein and D. P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Mathematical Programming* **55** (1992) 293–318.
29. I. I. Eremin, The relaxation method of solving systems of inequalities with convex functions on the left sides, *Soviet Mathematics – Doklady* **6** (1965) 219–222.
30. I. I. Eremin, Methods of Fejér approximations in convex programming, *Mathematical Notes of the Academy of Sciences of the USSR* **3** (1968) 139–149.
31. I. I. Eremin and V. D. Mazurov, *Nonstationary Processes of Mathematical Program-*

ming (Nauka, Moscow, 1979).
32. Yu. M. Ermol'ev, On the method of generalized stochastic gradients and quasi-Fejér sequences, *Cybernetics* **5** (1969) 208–220.
33. Yu. M. Ermol'ev, Stochastic quasigradient methods, in: Yu. Ermol'ev and R. J.-B. Wets, eds., *Numerical Techniques for Stochastic Optimization* 141–185 (Springer-Verlag, New York, 1988).
34. Yu. M. Ermol'ev and A. D. Tuniev, Random Fejér and quasi-Fejér sequences, *Theory of Optimal Solutions – Akademiya Nauk Ukrainskoĭ SSR Kiev* **2** (1968) 76–83; translated in: *American Mathematical Society Selected Translations in Mathematical Statistics and Probability* **13** (1973) 143–148.
35. S. D. Flåm, Successive averages of firmly nonexpansive mappings, *Mathematics of Operations Research* **20** (1995) 497–512.
36. U. García-Palomares, Parallel-projected aggregation methods for solving the convex feasibility problem, *SIAM Journal on Optimization* **3** (1993) 882–900.
37. U. M. García-Palomares and F. J. González-Castaño, Incomplete projection algorithms for solving the convex feasibility problem, *Numerical Algorithms* **18** (1998) 177–193.
38. A. Genel and J. Lindenstrauss, An example concerning fixed points, *Israel Journal of Mathematics* **22** (1975) 81–86.
39. K. Goebel and W. A. Kirk, *Topics in Metric Fixed Point Theory* (Cambridge University Press, Cambridge, 1990).
40. C. W. Groetsch, A note on segmenting Mann iterates, *Journal of Mathematical Analysis and Applications* **40** (1972) 369–372.
41. L. G. Gubin, B. T. Polyak, and E. V. Raik, The method of projections for finding the common point of convex sets, *USSR Computational Mathematics and Mathematical Physics* **7** (1967) 1–24.
42. O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM Journal on Control and Optimization* **29** (1991) 403–419.
43. K. C. Kiwiel, Block-iterative surrogate projection methods for convex feasibility problems, *Linear Algebra and Its Applications* **215** (1995) 225–259.
44. K. C. Kiwiel, A projection-proximal bundle method for convex nondifferentiable minimization, in: M. Théra and R. Tichatschke, eds., *Ill-Posed Variational Problems and Regularization Techniques* Lecture Notes in Economics and Mathematical Systems **477**, 137–150 (Springer-Verlag, New York, 1999).
45. K. C. Kiwiel and B. Łopuch, Surrogate projection methods for finding fixed points of firmly nonexpansive mappings, *SIAM Journal on Optimization* **7** (1997) 1084–1102.
46. B. Lemaire, Coupling optimization methods and variational convergence, in: K. H. Hoffmann, J. B. Hiriart-Urruty, C. Lemaréchal, and J. Zowe, eds., *Trends in Mathematical Optimization* Schriftenreihe zur Numerischen Mathematik **84**, 163–179 (Birkhäuser, Boston, MA, 1988).
47. B. Lemaire, Bounded diagonally stationary sequences in convex optimization, *Journal of Convex Analysis* **1** (1994) 75–86.
48. E. S. Levitin and B. T. Polyak, Constrained minimization methods, *USSR Computational Mathematics and Mathematical Physics* **6** (1966) 1–50.
49. B. Martinet, *Algorithmes pour la Résolution de Problèmes d'Optimisation et de Min-*

imax Doctoral thesis (Université de Grenoble, France, 1972).
50. B. Martinet, Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox, *Comptes Rendus de l'Académie des Sciences de Paris* **A274** (1972) 163–165.
51. B. Martinet, Perturbation des méthodes d'optimisation – Applications, *RAIRO Analyse Numérique* **12** (1978) 153–171.
52. Y. I. Merzlyakov, On a relaxation method of solving systems of linear inequalities, *USSR Computational Mathematics and Mathematical Physics* **2** (1963) 504–510.
53. J. J. Moreau, Un cas de convergence des itérées d'une contraction d'un espace hilbertien, *Comptes Rendus de l'Académie des Sciences de Paris* **A286** (1978) 143–144.
54. T. S. Motzkin and I. J. Schoenberg, The relaxation method for linear inequalities, *Canadian Journal of Mathematics* **6** (1954) 393–404.
55. O. Nevanlinna and S. Reich, Strong convergence of contraction semigroups and of iterative methods for accretive operators in Banach spaces, *Israel Journal of Mathematics* **32** (1979) 44–58.
56. N. Ottavy, Strong convergence of projection-like methods in Hilbert spaces, *Journal of Optimization Theory and Applications* **56** (1988) 433-461.
57. W. V. Petryshyn, Construction of fixed points of demicompact mappings in Hilbert space, *Journal of Mathematical Analysis and Applications* **14** (1966) 276–284.
58. W. V. Petryshyn and T. E. Williamson, Strong and weak convergence of the sequence of successive approximations for quasi-nonexpansive mappings, *Journal of Mathematical Analysis and Applications* **43** (1973) 459–497.
59. G. Pierra, Méthodes de projections parallèles extrapolées relatives à une intersection de convexes, research report (INPG, Grenoble, France, 1975).
60. G. Pierra, Decomposition through formalization in a product space, *Mathematical Programming* **28** (1984) 96–115.
61. G. Pierra and N. Ottavy, New parallel projection methods for linear varieties and applications, research report, presented at the XIIth International Symposium on Mathematical Programming, Tokyo, 1988.
62. B. T. Polyak, Minimization of unsmooth functionals, *USSR Computational Mathematics and Mathematical Physics* **9** (1969) 14–29.
63. E. Raik, Fejér type methods in Hilbert space, *Eesti NSV Teaduste Akadeemia Toimetised – Füüsika-Matemaatika* **16** (1967) 286–293.
64. E. Raik, A class of iterative methods with Fejér-monotone sequences, *Eesti NSV Teaduste Akadeemia Toimetised – Füüsika-Matemaatika* **18** (1969) 22–26.
65. S. Reich, Weak convergence theorems for nonexpansive mappings in Banach spaces, *Journal of Mathematical Analysis and Applications* **67** (1979) 274–276.
66. S. Reich, A limit theorem for projections, *Linear and Multilinear Algebra* **13** (1983) 281–290.
67. S. Reich and A. J. Zaslavski, Convergence of generic infinite products of nonexpansive and uniformly continuous operators, *Nonlinear Analysis* **36** (1999) 1049–1065.
68. R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization* **14** (1976) 877–898.
69. D. Schott, Weak convergence of iterative methods generated by strongly Fejér monotone mappings, *Rostocker Mathematisches Kolloquium* **51** (1997) 83–96.

70. N. Z. Shor, *Minimization Methods for Non-Differentiable Functions* (Springer-Verlag, New York, 1985).
71. M. V. Solodov and B. F. Svaiter, Forcing strong convergence of proximal point iterations in a Hilbert space, *Mathematical Programming* **87** (2000) 189–202.
72. P. Tseng, On the convergence of products of firmly nonexpansive mappings, *SIAM Journal on Optimization* **2** (1992) 425–434.