Algorithms based on unions of nonexpansive maps

Matthew K. Tam

Abstract In this work, we consider a framework for the analysis of iterative algorithms which can be described in terms of a structured set-valued operator. More precisely, at each point in the ambient space, we assume that the value of the operator can be expressed as a finite union of values of single-valued paracontracting operators. Our main result, which shows that the associated fixed point iteration is locally convergent around strong fixed points, generalises a theorem due to Bauschke and Noll (Arch Math 102(6):589–600, 2014).

Keywords Set-valued map · Fixed point iteration · Paracontracting operator · Sparsity

1 Preliminaries

Our setting throughout is the Euclidean space $X = \mathbb{R}^k$ equipped with norm $\| \cdot \|$. Following Elser et al. [5], we define the class of paracontracting operators as follows.

Definition 1 (Paracontracting operators [5]) A continuous operator, $S : X \to X$, is said to be paracontracting if

$$\|S(x) - y\| < \|x - y\| \quad (\forall x \in X \setminus \text{Fix } S)(\forall y \in \text{Fix } S),$$  \hspace{1cm} (1)
where \( \text{Fix } S = \{ x \in X : S(x) = x \} \) denotes the fixed point set of the (single-valued) operator \( S \).

The class of paracontracting operators includes, in particular, all \textit{averaged nonexpansive} and \textit{firmly nonexpansive} operators; see [2, §4]. In the literature, a possibly discontinuous operator satisfying (1) is called \textit{strictly quasinonexpansive}. Indeed, a paracontracting operator could be equally well described as a continuous, strictly quasinonexpansive operator.

In this work, we consider a structured set-valued operator, \( T : X \rightrightarrows X \), whose values can be expressed as the union of paracontracting operators taken from a finite collection of such operators. To be precise, let \( I \) denote a finite index set. We consider operators \( T : X \rightrightarrows X \) of the form

\[
T(x) := \{ T_i(x) : i \in \phi(x) \} \quad \forall x \in X,
\]

where, for each \( i \in I \), the single-valued operator \( T_i : X \to X \) is paracontracting, and the \textit{active selector}, \( \phi : X \rightrightarrows I \), dictates which operators from the collection \( \{T_i\}_{i \in I} \) are active and satisfies the following two properties.

\begin{enumerate}[	ext{(P1)}]
  \item \( \phi(x) \) is non-empty for every \( x \in X \).
  \item \( \phi \) is \textit{outer semicontinuous} (osc) [1, §3.2], that is, for all \( x \in X \) it holds that
    \[
    \phi(x) \supseteq \limsup_{y \to x} \phi(y) := \{ i \in I : \exists n \to x, \exists i_n \to i \text{ with } i_n \in \phi(x_n) \},
    \]
\end{enumerate}

where the limit superior is taken in the sense of \textit{Painlevé–Kuratowski} [1, §3.1].

Note that, for set-valued maps, there are two notions of fixed points which both coincide for single-valued operators; the \textit{fixed point set} denoted \( \text{Fix } T := \{ x : x \in T(x) \} \), and the \textit{strong fixed point set} denoted \( \text{Fix } T := \{ x : T(x) = \{ x \} \} \). It is clear that \( \text{Fix } T \subseteq \text{Fix } T \).

In what follows, we shall refer to operators of the form (2) which satisfy the aforementioned properties as \textit{union paracontracting}.

\textbf{Definition 2 (Union paracontracting)} An operator \( T : X \rightrightarrows X \) is said to be \textit{union paracontracting} if there exist a finite index set, \( I \), a collection of single-valued paracontracting operators, \( \{T_i\}_{i \in I} \), and an active selector, \( \phi \), satisfying (P1)–(P2) such that \( T \) is of the form specified by (2).

In this work, we study convergence properties of the fixed-point iteration of union paracontracting operators. More precisely, given an initial point, \( x_0 \in X \), we consider a sequence \( (x_n)_{n \in \mathbb{N}} \) generated by iterating the operator \( T \) in the sense that

\[
x_{n+1} \in T(x_n) \quad \forall n \in \mathbb{N}.
\]

Equivalently, for each \( n \in \mathbb{N} \), there is an index \( i_n \in \phi(x_n) \) such that

\[
x_{n+1} = T_{i_n}(x_n).
\]

The following result is a general convergence criterion for sequences conforming the structure of the iteration defined by (4).
Theorem 1 (Convergence with admissible control [5, Theorem 1]) Let \( \{T_j\}_{j \in J} \) be a collection of paracontracting operators on \( X \) where the index set \( J \) is assumed to be finite. Let \( \{j_n\}_{n \in \mathbb{N}} \subseteq J \) be an admissible control (i.e., each element in \( J \) appears infinitely often in the sequence) and let \( z_0 \in X \) be given. Then the sequence defined by

\[
z_{n+1} = T_{j_n}(z_n) \quad \forall n \in \mathbb{N},
\]

converges if and only if the collection of operators, \( \{T_j\}_{j \in J} \), have a common fixed point. Moreover, in this case \( \lim_{n \to \infty} z_n = z \in \bigcap_{j \in J} \text{Fix } T_j \).

Given the sequence \( (x_n)_{n \in \mathbb{N}} \) in (4) defined by applying the operators \( (T_{i_n})_{n \in \mathbb{N}} \), the admissible set for \( (x_n)_{n \in \mathbb{N}} \) is the (finite) set \( I^* \subseteq I \) such that \( i^* \in I^* \) appears infinitely often in \( (i_n)_{n \in \mathbb{N}} \). Note that \( I^* \) is always nonempty by virtue of the finiteness of \( I \) and, moreover, the definition of \( I^* \) ensures the existence of an \( n_0 \in \mathbb{N} \) such that \( (i_n)_{n=n_0}^\infty \subseteq I^* \) is an admissible control.

2 The main result: local convergence

In this section, we prove an abstract result concerning local convergence of iterations based on union contracting operators. We begin with the following proposition which applies, in particular, to active selectors. In what follows, \( B(x^*; \delta) \) denotes the closed ball centered at \( x^* \) with radius \( \delta > 0 \).

Proposition 1 Suppose \( \varphi : X \rightrightarrows M \) for a finite set \( M \), and let \( x^* \in X \). Then \( \varphi \) is outer semicontinuous at \( x^* \) if and only if there exists \( \delta > 0 \) such that

\[
\varphi(x) \subseteq \varphi(x^*) \quad \forall x \in B(x^*; \delta).
\]

Proof \((\Rightarrow): \) Suppose, by way of a contradiction, that \( \varphi \) is osc at \( x^* \) but there exists a sequence \( (x_n)_{n \in \mathbb{N}} \) with \( x_n \in B(x^*; 1/n) \) such that \( \varphi(x_n) \not\subseteq \varphi(x^*) \). Then there is a sequence, \( (m_n)_{n \in \mathbb{N}} \subseteq M \), such that

\[
m_n \in \varphi(x_n) \setminus \varphi(x^*) \quad \forall n \in \mathbb{N}.
\]

Since \( M \) is finite, by applying the pigeonhole principle, there exists a subsequence \( (m_{k_n})_{n \in \mathbb{N}} \) such that \( m_{k_n} = m \in M \) for all \( n \geq n_0 \). Altogether, we deduce that

\[
m \in \limsup \varphi(x) \subseteq \varphi(x^*),
\]

where the last inclusion follows from the definition of osc. But then (6) contradicts (5), and thus completes the first half of the proof.

\((\Leftarrow): \) Consider two arbitrary sequences, \( (x_n)_{n \in \mathbb{N}} \) and \( (i_n)_{n \in \mathbb{N}} \), such that \( x_n \to x^* \) and \( i_n \to i \) with \( i_n \in \varphi(x_n) \). In order to deduce that \( \varphi \) is osc at \( x^* \), we must show that \( i \in \varphi(x^*) \). To this end, since \( M \) is finite, there exists an index \( n_0 \in \mathbb{N} \) such that
\[ i_n = i \] for all \( n \geq n_0 \). By assumption, there is a \( \delta > 0 \) such that \( \varphi(x) \subseteq \varphi(x^*) \) for all \( x \in B(x^*; \delta) \) and, consequently, there is an index \( n_1 \geq n_0 \) such that \( \varphi(x_n) \subseteq \varphi(x^*) \) for all \( n \geq n_1 \). Altogether,
\[ i = i_n \in \varphi(x_n) \subseteq \varphi(x^*) \quad \forall n \geq n_1. \]

Thus we have that \( i \notin \varphi(x^*) \) and so the proof is complete. \( \square \)

We are now ready to prove our main result concerning local convergence of the algorithm defined by (3). In what follows, we adopt the convention that \( \sup \emptyset = -\infty \).

**Theorem 2** (Local convergence around strong fixed points) Suppose \( T : X \rightrightarrows X \) is union paracontracting with \( x^* \in \text{Fix } T \), and define
\[ r := \sup \{ \delta > 0 : \varphi(x) \subseteq \varphi(x^*) \text{ for all } x \in B(x^*; \delta) \}. \tag{7} \]
Then \( r > 0 \) and, for any \( \epsilon \in (0, r) \), it holds that \( \|y - x^*\| \leq \|x - x^*\| \) whenever \( x \in B(x^*; \epsilon) \) and \( y \in T(x) \). Furthermore, if the initial point \( x_0 \) is contained in \( B(x^*; \epsilon) \) and \( x_{n+1} \in T(x_n) \) for all \( n \in \mathbb{N} \), then the sequence \( (x_n)_{n \in \mathbb{N}} \) converges to a point \( \bar{x} \in \text{Fix } T \cap B(x^*; \epsilon) \).

**Proof** Since \( x^* \in \text{Fix } T \), it follows from Definition 2 that \( \varphi(x^*) \subseteq I \) is nonempty with
\[ x^* \in \bigcap_{i \in \varphi(x^*)} \text{Fix } T_i \neq \emptyset. \]
As \( \varphi \) is osc at \( x^* \) (Property (P2)), Proposition 1 implies the existence of a \( \delta > 0 \) such that \( \varphi(x) \subseteq \varphi(x^*) \) for all \( x \in B(x^*; \delta) \). Consequently, the constant \( r \), as defined in (7), is not less than \( \delta \) and thus, in particular, strictly greater than zero.

Let \( \epsilon \in (0, r) \). Since \( \{T_i\}_{i \in I} \) is a collection of paracontracting operators, we have that \( \|T_i(x) - x^*\| \leq \|x - x^*\| \) whenever \( i \in \varphi(x^*) \) which implies that
\[ \|y - x^*\| \leq \|x - x^*\|, \]
whenever \( x \in B(x^*; \epsilon) \) and \( y \in T(x) \). This completes the proof of the first two assertions.

Furthermore, suppose \( x_0 \in B(x^*; \epsilon) \) and \( x_{n+1} = T_{i_n}(x_n) \in T(x_n) \) with \( i_n \in \varphi(x_n) \) for all \( n \in \mathbb{N} \). Then it follows that \( x_n \in B(x^*; \epsilon) \) for all \( n \in \mathbb{N} \) and, thus together with (P1), that the set of admissible indices for \( (x_n)_{n \in \mathbb{N}} \), denoted \( I^* \), is a nonempty subset of \( \varphi(x^*) \). Consequently,
\[ x^* \in \bigcap_{i \in \varphi(x^*)} \text{Fix } T_i \subseteq \bigcap_{i \in I^*} \text{Fix } T_i \tag{8} \]
and, by the definition of \( I^* \), there exists an \( n_0 \in \mathbb{N} \) such that \( (i_n)_{n=n_0}^\infty \subseteq I^* \) is an admissible control. Noting (8) and applying Theorem 1 (with \( z_0 = x_{n_0} \) and \( J = I^* \))
gives that $x_n \to \bar{x} \in \text{Fix } T_{i^*}$ for some $i^* \in I^*$. The final claim regarding the limit, $\bar{x}$, follows by noting that $\text{Fix } T_{i^*} \subseteq \text{Fix } T$ and that $B(x^*; \epsilon)$ is closed.

Remark 1 (Necessity of strong fixed points) In general, the assumption that $x^*$ is a strong fixed point of $T$ cannot be relaxed to a mere fixed point. For an example, the reader is referred to [4, Remark 2].

Remark 2 (Limit points need only be weak fixed points) Although the reference point $x^*$ is assumed to be a strong fixed point, in general the limit point $\bar{x}$ of the sequence $(x_n)_{n \in \mathbb{N}}$ need only be a weak fixed point. An example of this behaviour follows.

Consider the union paracontracting operator $T : \mathbb{R} \rightrightarrows \mathbb{R}$ defined by

$$T(x) := \{T_1(x), T_2(x)\}$$

where $T_1(x) = 0$ and $T_2(x) = x$ (i.e., the active selector is given by $\varphi(x) = \{1, 2\}$ for all $x \in \mathbb{R}$). Then $\text{Fix } T = \{0\}$, $\text{Fix } \text{Fix } T = \mathbb{R}$ and, for any initial point $x_0 \in \mathbb{R} \setminus \{0\}$, the sequence $(x_n)_{n \in \mathbb{N}}$ with $x_n = x_0$ for all $n \in \mathbb{N}$ satisfies

$$x_{n+1} = T_2(x_n) \in T(x_n) \quad \forall n \in \mathbb{N}.$$ 

This constant sequence converges to $\bar{x} = x_0 \in \mathbb{R} \setminus \{0\} = \text{Fix } T \setminus \text{Fix } T$.

Remark 3 (The Douglas–Rachford algorithm) Theorem 2 is a generalisation of [4, Theorem 1] which considers the case in which $T$ is a specific operator; the Douglas–Rachford operator for two sets, $A$ and $B$, which can be realised as finite unions of closed convex sets $\{A_i\}_{i \in I}$ and $\{B_j\}_{j \in J}$. That is, $A := \bigcup_{i \in I} A_i$, $B := \bigcup_{j \in J} B_j$ and

$$T = \frac{I + R_B \circ R_A}{2},$$

where $P_A(x) := \arg \min_{a \in A} \|x - a\|$ denotes the projector onto $A$ and $R_A := 2P_A - I$. In this case, the function $\phi : X \rightrightarrows I \times J$ can be defined as

$$\phi(x) := \{(i, j) \in I \times J : P_{A_i}(x) \in P_A(x), P_{B_j}(R_{A_i}(x)) \in P_B(R_{A_i}(x))\},$$

and we therefore have that $T(x) = \{T_{ij}(x) : i \in \phi(x)\}$ where $T_{ij} : X \to X$ is defined by

$$T_{ij} = \frac{I + R_{B_j} \circ R_{A_i}}{2} \quad \forall (i, j) \in I \times J,$$

is firmly nonexpansive by [2, Propositions 4.8 & 4.21] and hence, in particular, paracontracting. It is straightforward to see that $\phi$ satisfies (P1). By combining Proposition 1 and [4, Theorem 1(2)], we deduce that $\phi$ also satisfies (P2). We therefore recover [4, Theorem 1] within the framework of Theorem 2.

A consequence of Theorem 2 is the following result concerning global convergence if there exists a strong fixed point such that every operator is active.

\[ \text{Springer} \]
Corollary 1 (Global convergence) Suppose $T : X \rightrightarrows X$ is union paracontracting and there exists a strong fixed point $x^* \in \text{Fix } T$ such that $\phi(x^*) = I$. For any $x_0 \in X$, select $x_{n+1} \in T(x_n)$ for all $n \in \mathbb{N}$. Then the sequence $(x_n)_{n \in \mathbb{N}}$ converges to a point $\bar{x} \in \text{Fix } T$.

Proof Since $\phi(x^*) = I$ and $\text{range}(\phi) \subseteq I$, the constant $r > 0$ in (7) becomes

$$r = \sup \{ \delta > 0 : \phi(x) \subseteq I \text{ for all } x \in B(x^*; \delta) \} = +\infty.$$ 

The result then follows from Theorem 2. \hfill \Box

3 An application: sparsity constrained minimisation

In this section, we give an application of Theorem 2 in sparsity constrained minimisation. In particular, we consider minimisation problems (with $X = \mathbb{R}^k$) of the form

$$\min_{x \in X} \{ f(x) : \|x\|_0 \leq s \}, \quad (9)$$

where $f : X \to \mathbb{R}$ is convex and continuously differentiable with Lipschitz continuous gradient, $\nabla f$, having Lipschitz constant $L > 0$, $\|x\|_0$ denotes the $\ell_0$-functional which counts the number of non-zero entries in a vector $x$, and $s \in \{0, 1, \ldots, k - 1\}$ is an a priori estimate on the sparsity of $x$ (typically $s \ll k$).

Denote $A := \{ x \in X : \|x\|_0 \leq s \}$. A forward–backward algorithm (or a projected gradient algorithm) [2, (27.26)] for (9) can now be compactly described as

$$x_{n+1} \in T(x_n) := P_A(x_n - \gamma \nabla f(x_n)) \quad \forall n \in \mathbb{N}, \quad (10)$$

where the constant $\gamma$ is selected such that $\gamma \in (0, 2/L)$. This algorithm also appears, for instance, in [6] under the name fast gradient hard thresholding pursuit (FGraHTP).

In order to investigate the properties of the forward–backward operator for (9), we introduce some further notation. Denote $I := \{ I \in 2^{\{1, 2, \ldots, k\}} : |I| = s \}$. Then the set $A$ can be expressed as the union of nonempty subspaces, precisely

$$A = \bigcup_{I \in I} A_I \text{ where } A_I := \{ x \in X : x_i \neq 0 \text{ only if } i \in I \}.$$ 

Consequently its projector has the form

$$P_A(x) = \{ P_{A_I}(x) : I \in \phi(x) \},$$

where $\phi(x) := \{ I \in I : \|x - P_{A_I}(x)\| = \min_{J \in I} \|x - P_{A_J}(x)\| \}$ and, for any $I \in I$, the single-valued projector onto $A_I$ is given componentwise by

$$P_{A_I}(x)_i = \begin{cases} x_i & i \in I \\ 0 & i \notin I, \end{cases} \quad \forall i \in \{1, 2, \ldots, k\}.$$ 

(11)
The operator $T$ may then be expressed in terms of single-valued operators as

$$T(x) = \{(P_{A_I} \circ Q)(x) : I \in \varphi(x)\},$$

where we denote $Q := I - \gamma \nabla f$ and $\varphi := (\phi \circ Q)$. Note that by [3, Proposition 3.6] we see that $\varphi(x)$ can be equivalently described as

$$\varphi(x) = \left\{ J \in \mathcal{I} : \min_{i \in J} |Q(x)_i| \geq \max_{i \notin J} |Q(x)_i| \right\}.$$

**Proposition 2** (Forward–backward with a sparsity constraint) The forward–backward operator for $(9)$, $T : X \Rightarrow A$, defined by $(10)$ is union paracontracting.

**Proof** We use the representation (12) to show that $T$ is union paracontracting. First note that, for each $J \in \mathcal{I}$, the argument in [2, Theorem 25.8] shows that the (single-valued) operator $P_{A_J} \circ Q$ is averaged nonexpansive and thus, in particular, also paracontracting. The finiteness of the index set $\mathcal{I}$ together with the definition of $\phi$ imply that $\varphi$ is nonempty valued. To see that $\varphi$ is osc, consider two sequences, $(x_n)_{n \in \mathbb{N}}$ and $(I_n)_{n \in \mathbb{N}}$, such that $x_n \to x \in X$ and $I_n \to I \in \mathcal{I}$ such that $I_n \in \varphi(x_n)$. Since $\mathcal{I}$ is finite, it follows that $I = I_n$ for all sufficiently large $n$ and hence, using the fact that both $P_{A_J}$ and $Q$ are continuous, we deduce that

$$\|Q(x) - P_{A_I}(Q(x))\| = \lim_{n \to \infty} \|Q(x_n) - P_{A_I}(Q(x_n))\|$$

$$= \lim_{n \to \infty} \left( \min_{J \in \mathcal{I}} \|Q(x_n) - P_{A_J}(Q(x_n))\| \right)$$

$$= \min_{J \in \mathcal{I}} \left( \lim_{n \to \infty} \|Q(x_n) - P_{A_J}(Q(x_n))\| \right)$$

$$= \min_{J \in \mathcal{I}} \|Q(x) - P_{A_J}(Q(x))\|,$$

which shows that $I \in \varphi(x)$ and we conclude that $\varphi$ is osc. This shows that $T$ is union paracontracting and completes the proof. \(\square\)

For a vector $x \in X$, let $\sigma_t(x)$ denote its $t$-th largest entry in magnitude, so that

$$\sigma_1(x) \geq \sigma_2(x) \geq \cdots \geq \sigma_k(x) \geq 0.$$

The following describes the fixed point sets of the forward–backward operator. Given a subspace $S \subseteq X$, its orthogonal complement, denoted $S^\perp$, is defined as $S^\perp := \{x \in X : \langle x, s \rangle = 0, \ \forall s \in X\}$.

**Proposition 3** (Fixed points and local minima) Let $T : X \Rightarrow A$ denote the forward–backward operator for $(9)$ and $x \in X$. The following assertions hold.

(i) Let $I \in \mathcal{I}$. Then $\text{Fix}(P_{A_I} \circ Q) = \arg \min_{z \in A_I} f(z)$.

(ii) Let $I \in \mathcal{I}$. If $x \in \text{Fix}(P_{A_I} \circ Q)$ then $x \in A_I$ and $\nabla f(x) \in A_I^\perp$. \(\square\)
(iii) \( x \in \text{Fix} \, T \) if and only if there exists an \( I \in \varphi(x) \) such that \( x = \arg \min_{z \in A_I} f(z) \). Moreover, in this case, \( x \) satisfies

\[
\min_{i \in I} |x_i| \geq \gamma \| \nabla f(x) \|_\infty.
\]  

(iv) If \( x \in \text{Fix} \, T \) and \( \|x\|_0 = s \) (resp. \( \|x\|_0 < s \)) then \( x \) is a local (resp. global) minimum of \( (9) \). In particular, every weak fixed point of \( T \) is a local minimum of \( (9) \).

(v) \( x \in \text{Fix} \, T \) if and only if \( x \in \text{Fix} \, T \) and \( \bigcup_{I \in \varphi(x)} \arg \min_{z \in A_I} f(z) \) is a singleton. In particular, if \( x \in \text{Fix} \, T \) then \( x \in \text{Fix} \, T \) as soon as either

\[
\nabla f(x) = 0 \quad \text{or} \quad \sigma_{\gamma}(x) > \gamma \| \nabla f(x) \|_\infty.
\]  

\( \sigma_{\gamma}(x) \) is a subspace its projector, \( P_{A_I} \), is a linear operator. As \( x \in A_I \), we therefore have

\[
x = P_{A_I}(x - \gamma \nabla f(x)) = P_{A_I}(x) - \gamma P_{A_I}(\nabla f(x)) = x - \gamma P_{A_I}(\nabla f(x)),
\]

which implies that \( P_{A_I}(\nabla f(x)) = 0 \). The claim follows. (iii): By the definition of \( T \) and (i), it follows that \( x \in \text{Fix} \, T \) if and only if there exists an \( I \in \varphi(x) \) such that \( x \in \text{Fix}(P_{A_I} \circ Q) = \arg \min_{z \in A_I} f(z) \). To prove (14), combine (ii), (11) and the fact that \( I \in \varphi(x) \) to deduce

\[
\min_{i \in I} |x_i| = \min_{i \in I} |x_i - \gamma \nabla f(x)_i| \geq \max_{i \notin I} |x_i - \gamma \nabla f(x)_i| = \max_{i \notin I} |\gamma \nabla f(x)_i|.
\]

(iv): Suppose \( x \in \text{Fix} \, T \). By (iii), there exists an \( I \in \varphi(x) \) such that

\[
x \in \arg \min_{z \in A_I} f(z) \quad \text{and} \quad \min_{i \in I} |x_i| \geq \gamma \| \nabla f(x) \|_\infty.
\]  

(16)

On one hand, if \( \|x\|_0 < s \) then \( \min_{i \in I} |x_i| = 0 \) and (16) implies \( \nabla f(x) = 0 \). Since \( f \) is convex and continuous, it follows that \( x \) is a global minimum to the unconstrained problem \( \min_{z \in X} f(z) \) (see [2, Proposition 26.1]). Since \( x \in A \), it must also be a global minimum of the constrained problem \( \min_{z \in A} f(z) \). On the other hand, if \( \|x\|_0 = s \), denote \( \delta := \min_{i \in I} |x_i| \). Then [3, Lemma 3.4] implies the inclusion

\[
\{ y \in A : \|x - y\| \leq \delta \} \subseteq A_I.
\]

In other words, there is an open ball centered at \( x \) whose intersection with \( A \) is contained entirely in \( A_I \). Together with (16), this shows that \( x \) is a local minimum of \( \min_{z \in A} f(z) \). (v): The equivalence follows from (i) and the definition of the strong fixed point set. In particular, let \( x \in \text{Fix} \, T \subseteq A \). On one hand, if \( \nabla f(x) = 0 \), then \( T(x) = P_A(x - \gamma \cdot 0) = P_A(x) = \{ x \} \), and hence \( x \in \text{Fix} \, T \). On the other hand, the definition of \( Fix \, T \) ensures the existence of an index set \( I \in \varphi(x) \) such that
$x \in \text{Fix}(P_{A_I} \circ Q)$. (ii) then gives $\nabla f(x) \in A_I^\perp$ which, by combining with (15), yields the inequality

$$
\min_{i \in I} |Q(x)_i| = \min_{i \in I} |x_i| = \sigma_s(x) > \gamma \|\nabla f(x)\|_\infty = \max_{i \notin I} |\gamma \nabla f(x)_i| = \max_{i \notin I} |Q(x)_i|.
$$

(17)

Since the inequality in (17) is strict, (13) together with the definition of $I$ shows that $\{I\} = \varphi(x)$. It then follows that $x \in \text{Fix} T$ which completes the proof. \hfill \square

Altogether, we arrive at the following specialisation of Theorem 2 for (9).

**Corollary 2** (Local convergence for sparsity constrained minimisation) *The conclusions of Theorem 2 apply to the forward–backward operator for (9) defined by (10). In particular, around strong fixed points of $T$, the iteration (10) is locally convergent to a local minimiser of (9).*

**Remark 4** (Restricted isometry/convexity/smoothness properties) Many results in the literature concerning the behaviour of the forward–backward algorithm applied to (9) proceed by assuming a strong hypothesis such as *restricted isometry property* or one of its variants. For instance, [6, Theorems 1 & 2] rely on *restricted strong convexity/smoothness* of the function $f$. Such properties typically provide a guarantee on accuracy of the recovered solutions and not only convergence. Whilst Corollary 2 provides no such guarantee, we emphasis that it also does not require any “restricted” property be assumed, and so is complementary to any such result.

**Acknowledgements** This work was supported by a Postdoctoral Fellowship from the Alexander von Humboldt Foundation. The author wishes to thank Yura Malitsky for suggesting Corollary 1 and the two anonymous referees whose comments have improved the manuscript.

**References**

1. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings, 2nd edn. Springer, New York (2014)
2. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)
3. Bauschke, H.H., Luke, D.R., Phan, H.M., Wang, X.: Restricted normal cones and sparsity optimization with affine constraints. Found. Comput. Math. 14, 63–83 (2014)
4. Bauschke, H.H., Noll, D.: On the local convergence of the Douglas–Rachford algorithm. Arch. Math. 102(6), 589–600 (2014)
5. Elsner, L., Koltracht, I., Neumann, M.: Convergence of sequential and asynchronous nonlinear paracontractions. Numer. Math. 62(1), 305–319 (1992)
6. Yuan, X., Li, P., Zhang, T.: Gradient hard thresholding pursuit for sparsity-constrained optimization, In: Proceedings of the 31st International Conference on Machine Learning (ICML-14), pp. 127–135 (2014)