

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Tam, MK

Title:

Frugal and decentralised resolvent splittings defined by nonexpansive operators

Date:

2024-09-01

Citation:

Tam, M. K. (2024). Frugal and decentralised resolvent splittings defined by nonexpansive operators. Optimization Letters, 18 (7), pp.1541-1559. <https://doi.org/10.1007/s11590-023-02064-y>.

Persistent Link:

<https://hdl.handle.net/11343/345414>

License:

[cc-by](#)



# Frugal and decentralised resolvent splittings defined by nonexpansive operators

Matthew K. Tam<sup>1</sup>

Received: 20 November 2022 / Accepted: 11 September 2023  
© The Author(s) 2023

## Abstract

Frugal resolvent splittings are a class of fixed point algorithms for finding a zero in the sum of the sum of finitely many set-valued monotone operators, where the fixed point operator uses only vector addition, scalar multiplication and the resolvent of each monotone operator once per iteration. In the literature, the convergence analyses of these schemes are performed in an inefficient, algorithm-by-algorithm basis. In this work, we address this by developing a general framework for frugal resolvent splitting which simultaneously covers and extends several important schemes in the literature. The framework also yields a new resolvent splitting algorithm which is suitable for decentralised implementation on regular networks.

**Keywords** Maximal monotonicity · Resolvent splitting · Decentralised algorithms

## 1 Introduction and preliminaries

Throughout this work,  $\mathcal{H}$  denotes a real Hilbert space with inner-product  $\langle \cdot, \cdot \rangle$  and induced norm  $\| \cdot \|$ . We are interested in solving the *monotone inclusion* given by

$$\text{find } x \in \left( \sum_{i=1}^n F_i \right)^{-1} (0) \subseteq \mathcal{H}, \quad (1)$$

where  $F_1, \dots, F_n : \mathcal{H} \rightrightarrows \mathcal{H}$  are (set-valued) *maximally monotone* operators [4, 18]. Recall that  $F_i$  is said to be maximally monotone if it is *monotone* in the sense

$$\langle x - \bar{x}, y - \bar{y} \rangle \geq 0 \quad \forall y \in F_i(x), \forall \bar{y} \in F_i(\bar{x})$$

---

✉ Matthew K. Tam  
matthew.tam@unimelb.edu.au

<sup>1</sup> School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia

and its graph contains no proper monotone extensions. The abstraction given by (1) covers many fundamental problems arising in mathematical optimisation involving objectives with finite-sum structure. Most notable are the following two types of examples:

**Example 1** (nonsmooth minimisation) Let  $f_1, \dots, f_n : \mathcal{H} \rightarrow (-\infty, +\infty]$  be proper, lsc, convex functions. Consider the finite-sum nonsmooth minimisation problem given by

$$\min_{x \in \mathcal{H}} \sum_{i=1}^n f_i(x). \quad (2)$$

Assuming sufficient regularity so that the sum-rule holds, the first-order optimality condition for (2) is given by the inclusion (1) with  $F_i = \partial f_i$ . Note that here “ $\partial$ ” denotes the *subdifferential* in the sense of convex analysis.

**Example 2** (nonsmooth min–max) Let  $\phi_1, \dots, \phi_n : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow [-\infty, +\infty]$  be proper, saddle-functions [16]. That is,  $\phi_i$  is convex-concave and there exists a point  $(u, v) \in \mathcal{H}_1 \times \mathcal{H}_2$  such that  $\phi_i(\cdot, v) < +\infty$  and  $\phi_i(u, \cdot) > -\infty$ . Consider the finite-sum nonsmooth minmax problem given by

$$\min_{u \in \mathcal{H}_1} \max_{v \in \mathcal{H}_2} \sum_{i=1}^n \phi_i(u, v) \quad (3)$$

Assuming sufficient regularity so that the sum-rule holds, the first-order optimality condition for (3) in the sense of *saddle-points* [16] is given by the inclusion (1) with  $\mathcal{H} := \mathcal{H}_1 \times \mathcal{H}_2$ ,  $x = (u, v)$  and  $F_i := \begin{pmatrix} \partial_u \phi_i \\ \partial_v (-\phi_i) \end{pmatrix}$ . Here “ $\partial_u$ ” (resp.  $\partial_v$ ) denotes the *partial subdifferential* with respect to the variable  $u$  (resp.  $v$ ).

**Frugal resolvent splittings.** In this work, we consider *frugal resolvent splittings* for solving (1). Roughly speaking, a fixed point algorithm, defined by an operator  $T : \mathcal{H}^d \rightarrow \mathcal{H}^d$ , for solving (1) is a *resolvent splitting* if it can be described using only vector addition, scalar multiplication, and the resolvents  $J_{F_1} := (\text{Id} + F_1)^{-1}, \dots, J_{F_n} := (\text{Id} + F_n)^{-1}$ . A resolvent splitting is *frugal* if it uses each of the resolvents only once per iteration. For details and extensions of these notions, see [1, 14, 15, 17].

In recent times, there has been interest in *minimal* frugal resolvent splittings. As a consequence of [14, Theorem 3.3], these are precisely the frugal resolvent splittings such that the operator  $T : \mathcal{H}^d \rightarrow \mathcal{H}^d$  has  $d = n - 1$ . In the case  $n = 2$ , the *Douglas–Rachford algorithm* is a minimal frugal resolvent splitting which is defined by  $T : \mathcal{H} \rightarrow \mathcal{H}$  where

$$T(z) = z + \gamma (J_{F_2} (2J_{F_1}(z) - z) - J_{F_1}(z)).$$

For  $n = 3$ , *Ryu's splitting algorithm* [2, 17] is a minimal resolvent splitting which is defined by  $T : \mathcal{H}^2 \rightarrow \mathcal{H}^2$  where

$$T(\mathbf{z}) = \mathbf{z} + \gamma \begin{pmatrix} J_{F_3}(J_{F_1}(z_1) - z_1 + J_{F_2}(J_{F_1}(z_1) + z_2) - z_2) - J_{F_1}(z_1) \\ J_{F_3}(J_{F_1}(z_1) - z_1 + J_{F_2}(J_{F_1}(z_1) + z_2) - z_2) - J_{F_2}(J_{F_1}(z_1) + z_2) \end{pmatrix}. \quad (4)$$

As observed in [14, Remark 4.7], attempts to extend Ryu's splitting algorithm to  $n > 3$  operators have been so far unsuccessful. To address this for  $n \geq 3$ , Malitsky & Tam [14] introduced the minimal frugal resolvent splitting defined by  $T_A : \mathcal{H}^{n-1} \rightarrow \mathcal{H}^{n-1}$  where

$$T(\mathbf{z}) := \mathbf{z} + \gamma \begin{pmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_n - x_{n-1} \end{pmatrix} \text{ where } \begin{cases} x_1 = J_{F_1}(z_1), \\ x_i = J_{F_i}(z_i + x_{i-1} - z_{i-1}) \quad \forall i \in \{2, \dots, n-1\}, \\ x_n = J_{F_n}(x_1 + x_{n-1} - z_{n-1}). \end{cases}$$

Despite many similarities in the convergence proofs of the above algorithms (e.g., all of the underlying operators are nonexpansive), their convergence analyses are performed in an inefficient, algorithm-by-algorithm basis in the literature. In this work, we address this by providing a general framework for convergence of (not necessarily minimal) frugal resolvent splittings which simultaneously covers all of the aforementioned algorithm. This framework reduces nonexpansiveness of the underlying operator  $T$  to a question of negative definiteness of a certain matrix defined by algorithm coefficients. In addition to significantly simplifying the proofs of [14, 17], the framework also suggests a convergent extension of Ryu's splitting algorithm for  $n > 3$  operators. More precisely, it can be used to show that the operator  $T : \mathcal{H}^{n-1} \rightarrow \mathcal{H}^{n-1}$  is nonexpansive where

$$T(\mathbf{z}) = \mathbf{z} + \gamma \sqrt{\frac{2}{n-1}} \begin{pmatrix} x_n - x_1 \\ x_n - x_2 \\ \vdots \\ x_n - x_{n-1} \end{pmatrix} \text{ where } \begin{cases} x_i = J_{F_i} \left( \sqrt{\frac{2}{n-1}} z_i + \frac{2}{n-1} \sum_{j=1}^{i-1} x_j \right) \quad \forall i \in \{1, \dots, n-1\} \\ x_n = J_{F_n} \left( \frac{2}{n-1} \sum_{j=1}^{n-1} x_i - \sqrt{\frac{2}{n-1}} \sum_{i=1}^{n-1} z_i \right). \end{cases} \quad (5)$$

Note that (5) coincides with (4) when  $n = 3$ .

**Decentralised algorithms.** Another focus of this work is on *distributed algorithms* [6] for solving (1) under the following conditions:

- Each operator  $F_i$  in (1) is known by single node in a graph  $G = (V, E)$  with  $|V| = n$ . More precisely,  $F_i$  is known by the  $i$ th node.
- Each node has its own local variables, and can compute standard arithmetic operations as well as the *resolvent* of its monotone operator  $F_i$  given by  $J_{F_i} := (\text{Id} + F_i)^{-1}$  (see [4, Section 23]).
- Two nodes in  $G$  can communicate the results of their computations directly with one another only if they are connected by an edge in  $G$ .
- The algorithm is “decentralised” in the sense that it does not necessarily require the computation of a global sum across all local variables of the nodes.

Stated at the abstraction of monotone inclusions, the classical way to solve (1) (without distributed considerations) involves using the *Douglas–Rachford algorithm* applied to a reformulation in the product-space  $\mathcal{H}^n$  (see, for instance, [14, Example 2.7] or [4, Proposition 26.12]). Given an initial point  $\mathbf{z}^0 = (z_1^0, \dots, z_n^0) \in \mathcal{H}^n$  and a constant  $\gamma \in (0, 2)$ , this approach gives rise to the iteration given by

$$\begin{cases} x^k = \frac{1}{n} \sum_{j=1}^n z_j^k \\ z_i^{k+1} = z_i^k + \gamma (J_{F_i}(2x^k - z_i^k) - x^k) \quad \forall i \in \{1, \dots, n\}. \end{cases} \quad (6)$$

In terms of distributed computing, this iteration requires the global sum across the local variables  $z_1^k, \dots, z_n^k$  to compute the aggregate  $x^k$  and so it does not satisfy requirements (c) or (d) above. On the other hand, the  $z_i$ -update in (6) does satisfy properties (a) and (b), assuming it is computed by node  $i$ . For recent variations on (6), the reader is referred to [8, 12].

In order to describe other algorithms in the literature, it is convenient to introduce some notation. Given a matrix  $W = (w_{ij}) \in \mathbb{R}^{p \times l}$ , we denote  $\mathbf{W} := W \otimes \text{Id}$  where  $\text{Id}$  denotes the identity operator on  $\mathcal{H}$  and  $\otimes$  denotes the Kronecker product. In other words,  $\mathbf{W}$  can be identified with the bounded linear operator from  $\mathcal{H}^l$  to  $\mathcal{H}^p$  given by

$$\mathbf{W} = W \otimes \text{Id} = \begin{bmatrix} w_{11} \text{Id} & w_{12} \text{Id} & \dots & w_{1l} \text{Id} \\ w_{21} \text{Id} & w_{22} \text{Id} & \dots & w_{2l} \text{Id} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p1} \text{Id} & w_{p2} \text{Id} & \dots & w_{pl} \text{Id} \end{bmatrix}.$$

Given maximally monotone operators  $F_1, \dots, F_n : \mathcal{H} \rightrightarrows \mathcal{H}$ , we denote the maximally monotone operator  $\mathbf{F} : \mathcal{H}^n \rightarrow \mathcal{H}^n$  by  $\mathbf{F}(\mathbf{x}) := (F_1(x_1), \dots, F_n(x_n))$  for all  $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{H}^n$ . Consequently, its resolvent  $J_{\mathbf{F}} : \mathcal{H}^n \rightarrow \mathcal{H}^n$  is given by  $J_{\mathbf{F}} = (J_{F_1}, \dots, J_{F_n})$ . Recall also that, when  $F_i = \partial f_i$  where  $f_i$  is proper, lsc and convex, then the resolvent coincides with *proximity operator* [4, Section 24] given by

$$J_{F_i} = \text{prox}_{f_i} = \arg \min_{y \in \mathcal{H}} \left\{ f_i(y) + \frac{1}{2} \|\cdot - y\|^2 \right\}.$$

Now, in the special case of Example 1 where  $J_{\partial f_i} = \text{prox}_{f_i}$ , the *proximal exact first-order algorithm (P-EXTRA)* [20] can be used to solve (1). Given  $\mathbf{x}^0 \in \mathcal{H}^n$  and initialising with  $\mathbf{y}^0 = \mathbf{W}\mathbf{x}^0$  and  $\mathbf{x}^1 = \text{prox}_{af}(\mathbf{y}^0)$ , this method iterates according to

$$\begin{cases} \mathbf{y}^k = \mathbf{W}\mathbf{x}^k + \mathbf{y}^{k-1} - \widetilde{\mathbf{W}}\mathbf{x}^{k-1} \\ \mathbf{x}^{k+1} = \text{prox}_{af}(\mathbf{y}^k). \end{cases} \quad (7)$$

where  $W, \widetilde{W} \in \mathbb{R}^{n \times n}$  are “mixing matrices” satisfying (c) that describe the communication between nodes. For specific examples, the reader is referred to [20,

Assumption 1]. This method satisfies conditions (a)–(d) in the minimisation setting with  $x_i, y_i, z_i$  being the local variables of the  $i$ th node.

Let  $L \in \mathbb{R}^{n \times n}$  denote the *graph Laplacian* of  $G = (V, E)$  [10] where  $|V| = n$ . For the general monotone inclusion (1), the *primal dual hybrid gradient (PDHG) algorithm* [9, 11] can be used. After a change of variables (see [14, Section 5.1]), the PDHG can be expressed as

$$\begin{cases} \mathbf{x}^{k+1} = J_{\tau \mathbf{F}}(\mathbf{x}^k - \tau \mathbf{v}^k) \\ \mathbf{v}^{k+1} = \mathbf{v}^k + \sigma L(2\mathbf{x}^{k+1} - \mathbf{x}^k) \end{cases} \quad (8)$$

where the stepsizes are required to satisfy  $\tau \sigma \|L\| \leq 1$ . This method satisfies conditions (a)–(d) with  $x_i, v_i$  being the variables local to the  $i$ th node.

Using our framework for convergence of frugal resolvent splittings, this work derives new fixed point algorithms for (1) which are suitable to decentralised implementation, as outlined in conditions (a)–(d). Given  $\gamma \in (0, 1)$  and  $\tau := \frac{n}{|E|}$ , the most similar realisation of our approach to (8) (see Example 7) is given by

$$\begin{cases} x_i^k = J_{F_i} \left( v_i^k + \tau \sum_{j=1}^{i-1} A_{ij} x_j^k \right) \quad \forall i \in \{1, \dots, n\} \\ v_i^{k+1} = v_i^k - \gamma \tau \sum_{j=1}^n L_{ij} x_j^k \end{cases} \quad (9)$$

where  $A \in \mathbb{R}^{n \times n}$  denotes the adjacency matrix of the graph  $G$ . Note that this method satisfies conditions (a)–(d) with  $x_i, v_i$  be the variable local to the  $i$ th node. By letting  $N \in \mathbb{R}^{n \times n}$  denote the lower triangular part of  $A$ , the iteration (9) can be compactly expressed as

$$\begin{cases} \mathbf{x}^k = J_{\mathbf{F}}(\mathbf{v}^k + \tau \mathbf{N} \mathbf{x}^k) \\ \mathbf{v}^{k+1} = \mathbf{v}^k - \gamma \tau \mathbf{L} \mathbf{x}^k. \end{cases} \quad (10)$$

Here we note that  $\mathbf{x}^k$  is well-defined due to the fact that  $N$  is lower triangular with zeros on the diagonal. A feature of (10), which is not present in (7) or (8), is that its  $x_i$ -updates in (10) are performed sequentially within each iteration. Furthermore, the update (10) requires no knowledge of past iterates.

## 1.1 Structure of this work

The remainder of the work is structured as follows. In Sect. 2, we develop an abstract framework for convergence of frugal resolving splitting algorithms as fixed point iterations, which extend ideas from [14], and analyse its convergence properties. More precisely, the main results in Sect. 2 establish averaged nonexpansivity of the algorithmic operator (Lemma 2) and weak convergence of its iterates (Theorem 1). Then, in Sect. 3, we provide various concrete realisation of the abstract framework. In particular, this includes a previously unknown extension of Ryu's method [17]

to finitely many operators (Example 6), the method proposed in [14] (Example 5), as well as the new decentralised scheme (10) when the communication graph  $G$  is regular (Example 7). Numerical examples to demonstrated the proposed framework are reported in Sect. 4.

## 2 A framework for resolvent splitting

In this section, we consider an operator  $T : \mathcal{H}^m \rightarrow \mathcal{H}^m$  of the form

$$T(\mathbf{z}) := \mathbf{z} + \gamma \mathbf{M}\mathbf{x}, \quad \mathbf{y} = \mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x}, \quad \mathbf{x} = J_{\mathbf{F}}(\mathbf{y}), \quad (11)$$

where  $M \in \mathbb{R}^{m \times n}$ ,  $S \in \mathbb{R}^{n \times m}$  and  $N \in \mathbb{R}^{n \times n}$  are coefficient matrices. Moreover, we assume that the resolvents are evaluated in the order  $J_{F_1}, \dots, J_{F_n}$ , which means that the matrix  $N$  is necessarily lower triangular with zeros on the diagonal. Altogether, we require the following assumptions on the coefficient matrices.

**Assumption 1** The coefficient matrices  $M \in \mathbb{R}^{m \times n}$ ,  $S \in \mathbb{R}^{n \times m}$  and  $N \in \mathbb{R}^{n \times n}$  satisfy the following properties.

- (a)  $\ker M \subseteq \mathbb{R}e$  where  $e = (1, \dots, 1)^\top \in \mathbb{R}^n$ .
- (b)  $\sum_{i,j=1}^n N_{ij} = n$  and  $N$  is lower triangular with zeros on the diagonal.
- (c)  $S^* = -M$ .
- (d)  $M^*M + N + N^* - 2\text{Id} \leq 0$ .

**Remark 1** When Assumption 1 holds, properties of the Kronecker product imply that

$$\ker \mathbf{M} = \Delta := \{\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{H}^n : x_1 = \dots = x_n\}, \quad (12)$$

$S^* = -M$  and  $M^*M + N + N^* - 2\text{Id} \leq 0$ . To see (12), we note the following chain of implications

$$\begin{aligned} (x_1, \dots, x_n) \in \ker \mathbf{M} &\implies (\langle x_1, v \rangle, \dots, \langle x_n, v \rangle) \in \ker M \quad \forall v \in \mathcal{H} \\ &\implies \langle x_1, v \rangle = \dots = \langle x_n, v \rangle \quad \forall v \in \mathcal{H} \\ &\implies x_1 = \dots = x_n \\ &\implies (x_1, \dots, x_n) \in \ker \mathbf{M}. \end{aligned}$$

**Lemma 1** (Fixed points and zeros) Suppose Assumptions 1(a)–(c) holds, and denote

$$\Omega := \{(\mathbf{z}, x) \in \mathcal{H}^m \times \mathcal{H} : \mathbf{x} = J_{\mathbf{F}}(\mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x}) \text{ where } \mathbf{x} = (x, \dots, x) \in \Delta\}.$$

Then the following assertions hold.

- (a) If  $\mathbf{z} \in \text{Fix } T$ , then there exists  $x \in \mathcal{H}$  such that  $(\mathbf{z}, x) \in \Omega$ .

- (b) If  $x \in \text{zer} \left( \sum_{i=1}^n F_i \right)$ , then there exists  $\mathbf{z} \in \mathcal{H}^m$  such that  $(\mathbf{z}, x) \in \Omega$ .  
 (c) If  $(\mathbf{z}, x) \in \Omega$ , then  $\mathbf{z} \in \text{Fix } T$  and  $x \in \text{zer} \left( \sum_{i=1}^n F_i \right)$ .

Consequently,

$$\text{Fix } T \neq \emptyset \iff \Omega \neq \emptyset \iff \text{zer} \left( \sum_{i=1}^n F_i \right) \neq \emptyset.$$

**Proof** (a): Let  $\mathbf{z} \in \text{Fix } T$ . Set  $\mathbf{y} := \mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x}$  and  $\mathbf{x} := J_{\mathbf{F}}(\mathbf{y})$ . Since  $\mathbf{z} \in \text{Fix } T$ , we have  $\mathbf{x} \in \ker M$  and so Assumption 1(a) implies  $\mathbf{x} \in \Delta$ .

(b): Let  $x \in \text{zer} \left( \sum_{i=1}^n F_i \right)$  and set  $\mathbf{x} := (x, \dots, x) \in \Delta$ . Then there exists  $\mathbf{v} = (v_1, \dots, v_n) \in \mathbf{F}(\mathbf{x})$  such that  $\sum_{i=1}^n v_i = 0$ . Define  $\mathbf{y} = (y_1, \dots, y_n) \in \mathcal{H}^n$  according to  $\mathbf{y} := \mathbf{v} + \mathbf{x}$ , so that  $\mathbf{x} = J_{\mathbf{F}}(\mathbf{y})$ . To complete the proof, we must show there exists  $\mathbf{z} \in \mathcal{H}^m$  such that  $\mathbf{y} = \mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x}$ , which is equivalent to the  $\mathbf{y} - \mathbf{N}\mathbf{x} \in \text{range } \mathbf{S}$ . To this end, first note that Assumption 1(a) & (c) imply

$$\text{range } \mathbf{S} = (\ker \mathbf{S}^*)^\perp = (\ker \mathbf{M})^\perp \supseteq \Delta^\perp = \left\{ (x_1, \dots, x_n) \in \mathcal{H}^n : \sum_{i=1}^n x_i = 0 \right\}.$$

With the help of Assumption 1(b), we see that

$$\sum_{i=1}^n (\mathbf{y} - \mathbf{N}\mathbf{x})_i = \sum_{i=1}^n y_i - \sum_{i,j=1}^n N_{ij}x = \left( \sum_{i=1}^n v_i + nx \right) - nx = 0,$$

Hence  $\mathbf{y} - \mathbf{N}\mathbf{x} \in \text{range } \mathbf{S}$ , as required.

(c): Let  $(\mathbf{z}, x) \in \Omega$  and set  $\mathbf{y} := \mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x}$  where  $\mathbf{x} = (x, \dots, x) \in \Delta$ . Since  $\mathbf{x} = J_{\mathbf{F}}(\mathbf{y}) \in \ker \mathbf{M}$ , we have  $\mathbf{z} \in \text{Fix } T$  by Assumption 1(a). Denote  $\mathbf{e} = (1, \dots, 1)^\top \in \mathbb{R}^n$ . Then  $\mathbf{M}\mathbf{e} = 0$  by Assumption 1(a). Since  $\mathbf{F}(\mathbf{x}) \ni \mathbf{y} - \mathbf{x} = -\mathbf{M}^*\mathbf{z} + \mathbf{N}\mathbf{x} - \mathbf{x}$ , Assumption 1(b) gives

$$\sum_{i=1}^n F_i(x) \ni -((\mathbf{e}^\top \mathbf{M}^*) \otimes \text{Id})\mathbf{z} + \sum_{i,j=1}^n N_{ij}x - nx = -((\mathbf{M}\mathbf{e})^\top \otimes \text{Id})\mathbf{z} + nx - nx = 0,$$

This shows that  $x \in \text{zer} \left( \sum_{i=1}^n F_i \right)$  and so completes the proof.  $\square$

Recall that an operator  $S$  is  $\alpha$ -averaged nonexpansive if  $\alpha \in (0, 1)$  and, for all  $x, y$ , we have

$$\|S(x) - S(y)\|^2 + \frac{1-\alpha}{\alpha} \|(\text{Id} - S)(x) - (\text{Id} - S)(y)\|^2 \leq \|x - y\|^2.$$

**Lemma 2** (Nonexpansivity) *Let  $F_1, \dots, F_n : \mathcal{H} \rightrightarrows \mathcal{H}$  be maximally monotone. Suppose Assumptions 1(c) & (d) hold, and let  $\gamma > 0$ . Then, for all  $\mathbf{z}, \bar{\mathbf{z}} \in \mathcal{H}^m$ , we have*

$$\begin{aligned} & \|T(\mathbf{z}) - T(\bar{\mathbf{z}})\|^2 + \frac{1-\gamma}{\gamma} \|(\text{Id} - T)(\mathbf{z}) - (\text{Id} - T)(\bar{\mathbf{z}})\|^2 \\ & + \gamma \langle \mathbf{x} - \bar{\mathbf{x}}, [2\text{Id} - \mathbf{M}^*\mathbf{M} - \mathbf{N} - \mathbf{N}^*](\mathbf{x} - \bar{\mathbf{x}}) \rangle \leq \|\mathbf{z} - \bar{\mathbf{z}}\|^2. \end{aligned} \quad (13)$$

where  $\mathbf{x} = J_{\mathbf{F}}(\mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x})$  and  $\bar{\mathbf{x}} = J_{\mathbf{F}}(\mathbf{S}\bar{\mathbf{z}} + \mathbf{N}\bar{\mathbf{x}})$ . In particular,  $T$  is  $\gamma$ -averaged nonexpansive whenever  $\gamma \in (0, 1)$ .

**Proof** Let  $\mathbf{z}, \bar{\mathbf{z}} \in \mathcal{H}^m$ . Set  $\mathbf{x} := J_{\mathbf{F}}(\mathbf{y})$  where  $\mathbf{y} := \mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x}$  and  $\bar{\mathbf{x}} := J_{\mathbf{F}}(\bar{\mathbf{y}})$  where  $\bar{\mathbf{y}} := \mathbf{S}\bar{\mathbf{z}} + \mathbf{N}\bar{\mathbf{x}}$ . Since  $\mathbf{y} - \mathbf{x} \in \mathbf{F}(\mathbf{x})$  and  $\bar{\mathbf{y}} - \bar{\mathbf{x}} \in \mathbf{F}(\bar{\mathbf{x}})$ , monotonicity of  $\mathbf{F} = (F_1, \dots, F_n)$  gives

$$\begin{aligned} 0 & \leq \langle \mathbf{x} - \bar{\mathbf{x}}, (\mathbf{y} - \mathbf{x}) - (\bar{\mathbf{y}} - \bar{\mathbf{x}}) \rangle \\ & = \langle \mathbf{x} - \bar{\mathbf{x}}, (\mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x} - \mathbf{x}) - (\mathbf{S}\bar{\mathbf{z}} + \mathbf{N}\bar{\mathbf{x}} - \bar{\mathbf{x}}) \rangle \\ & = \langle \mathbf{S}^*\mathbf{x} - \mathbf{S}^*\bar{\mathbf{x}}, \mathbf{z} - \bar{\mathbf{z}} \rangle + \langle \mathbf{x} - \bar{\mathbf{x}}, (\mathbf{N} - \text{Id})\mathbf{x} - (\mathbf{N} - \text{Id})\bar{\mathbf{x}} \rangle. \end{aligned} \quad (14)$$

By Assumption 1(c),  $\mathbf{S}^* = -\mathbf{M}$  and so the first term in (14) can be expressed as

$$\begin{aligned} & \langle \mathbf{S}^*\mathbf{x} - \mathbf{S}^*\bar{\mathbf{x}}, \mathbf{z} - \bar{\mathbf{z}} \rangle \\ & = \langle (-\mathbf{M}\mathbf{x}) - (-\mathbf{M}\bar{\mathbf{x}}), \mathbf{z} - \bar{\mathbf{z}} \rangle \\ & = \frac{1}{\gamma} \langle (\text{Id} - T)(\mathbf{z}) - (\text{Id} - T)(\bar{\mathbf{z}}), \mathbf{z} - \bar{\mathbf{z}} \rangle \\ & = \frac{1}{2\gamma} (\|\mathbf{z} - \bar{\mathbf{z}}\|^2 + \|(\text{Id} - T)(\mathbf{z}) - (\text{Id} - T)(\bar{\mathbf{z}})\|^2 - \|T(\mathbf{z}) - T(\bar{\mathbf{z}})\|^2). \end{aligned} \quad (15)$$

The second term in (14) can be expressed as

$$\begin{aligned} & \langle \mathbf{x} - \bar{\mathbf{x}}, (\mathbf{N} - \text{Id})\mathbf{x} - (\mathbf{N} - \text{Id})\bar{\mathbf{x}} \rangle \\ & = \frac{1}{2} \langle \mathbf{x} - \bar{\mathbf{x}}, [\mathbf{M}^*\mathbf{M} + 2\mathbf{N} - 2\text{Id}](\mathbf{x} - \bar{\mathbf{x}}) \rangle - \frac{1}{2} \|\mathbf{M}\mathbf{x} - \mathbf{M}\bar{\mathbf{x}}\|^2 \\ & = \frac{1}{2} \langle \mathbf{x} - \bar{\mathbf{x}}, [\mathbf{M}^*\mathbf{M} + \mathbf{N} + \mathbf{N}^* - 2\text{Id}](\mathbf{x} - \bar{\mathbf{x}}) \rangle \\ & \quad - \frac{1}{2\gamma^2} \|(\text{Id} - T)(\mathbf{z}) - (\text{Id} - T)(\bar{\mathbf{z}})\|^2. \end{aligned} \quad (16)$$

Substituting (15) and (16) into (14), followed by multiplying by  $2\gamma$ , gives (13). In particular, if Assumption 1(d) holds, then the inner-product on the LHS of (13) is non-negative and hence  $T$  is  $\gamma$ -averaged nonexpansive whenever  $\gamma \in (0, 1)$ .  $\square$

**Remark 2** As we will explain in Sect. 3, the setting considered here includes Ryu's method [17] and the resolvent splitting with minimal lifting from [14] as special cases. Consequently, Lemma 2 provides a very concise proof covering [17, Section 4.2] and [14, Lemma 4.3] (each of which took several pages of algebraic manipulation).

The following theorem is our main result regarding convergence of the fixed point iteration corresponding to  $T$  in (11).

**Theorem 1** Let  $F_1, \dots, F_n : \mathcal{H} \rightrightarrows \mathcal{H}$  be maximally monotone operators with  $\text{zer}(\sum_{i=1}^n F_i) \neq \emptyset$ . Suppose Assumption 1 holds and let  $\gamma \in (0, 1)$ . Let  $\mathbf{z}^0 \in \mathcal{H}$  and define the sequences  $(\mathbf{z}^k)$  and  $(\mathbf{x}^k)$  according to

$$\mathbf{z}^{k+1} = T(\mathbf{z}^k) = \mathbf{z}^k + \gamma \mathbf{M} \mathbf{x}^k, \quad \mathbf{x}^k = J_{\mathbf{F}}(\mathbf{S} \mathbf{z}^k + \mathbf{N} \mathbf{x}^k). \quad (17)$$

Then the following assertions hold.

- (a) As  $k \rightarrow \infty$ , we have  $\mathbf{z}^k - \mathbf{z}^{k+1} \rightarrow 0$  and  $\sum_{i=1}^n s_i x_i^k \rightarrow 0$  for all  $s \in \mathbb{R}^n$  with  $\sum_{i=1}^n s_i = 0$ .
- (b) The sequence  $(\mathbf{z}^k)$  converges weakly to a point  $\bar{\mathbf{z}} \in \text{Fix } T$ .
- (c) The sequence  $(\mathbf{x}^k)$  converges weakly to a point  $(\bar{x}, \dots, \bar{x}) \in \mathcal{H}^n$  such that  $\bar{x} \in \text{zer}(\sum_{i=1}^n F_i)$ .

**Proof** (a) & (b): Since  $\text{zer}(\sum_{i=1}^n F_i) \neq \emptyset$ , Lemma 1 implies  $\text{Fix } T \neq \emptyset$ . By Lemma 2, the operator  $T$  is averaged nonexpansive and so it follows from [4, Proposition 5.16] that  $\mathbf{z}^k - \mathbf{z}^{k+1} \rightarrow 0$  and  $(\mathbf{z}^k)$  converges weakly to a point  $\bar{\mathbf{z}} \in \text{Fix } T$ . Next, let  $s \in \{s \in \mathbb{R}^n : \sum_{i=1}^n s_i = 0\} = \Delta^\perp \subseteq (\ker M)^\perp = \text{range } M^* = \text{range } S$ . Then, there exists  $v \in \mathbb{R}^m$  such that  $s = Sv = -M^*v$  and hence

$$\begin{aligned} -\sum_{i=1}^n s_i x_i^k &= -(s^\top \otimes \text{Id}) \mathbf{x}^k = ((v^\top M) \otimes \text{Id}) \mathbf{x}^k \\ &= (v^\top \otimes \text{Id}) \mathbf{M} \mathbf{x}^k = \frac{1}{\gamma} (v^\top \otimes \text{Id}) (\mathbf{z}^{k+1} - \mathbf{z}^k) \rightarrow 0. \end{aligned}$$

(c): Denote  $\mathbf{y}^k = \mathbf{S} \mathbf{z}^k + \mathbf{N} \mathbf{x}^k$ . We claim that the sequence  $(\mathbf{x}^k)$  is bounded. To see this, let  $S_1 \in \mathbb{R}^{1 \times n}$  denote the first row of the matrix  $S$  and let  $N_1 = (0, \dots, 0) \in \mathbb{R}^{1 \times n}$  denote the first row of the matrix  $N$  (which is zero by Assumption 1(b)). Thus, by definition, we have

$$x_1^k = J_{F_1}(y_1^k) = J_{F_1}(S_1 \mathbf{z}^k + N_1 \mathbf{x}^k) = J_{F_1}(S_1 \mathbf{z}^k).$$

By Lemma 1,  $(\bar{\mathbf{z}}, \bar{x}) \in \Omega$  where  $\bar{x} = J_{F_1}(S_1 \bar{\mathbf{z}})$ . Thus, by nonexpansivity of resolvents [4, Proposition 23.8] and boundedness of  $(\mathbf{z}^k)$ , it then follows that  $(x_1^k)$  is also bounded. By (a), it follows that  $(\mathbf{x}^k)$  is bounded, as claimed.

Next, using the identity  $\mathbf{x}^k = J_{\mathbf{F}}(\mathbf{y}^k)$ , followed by the identity  $\mathbf{y}^k = \mathbf{S} \mathbf{z}^k + \mathbf{N} \mathbf{x}^k$  and the fact that  $\text{range } \mathbf{S} \subseteq \Delta^\perp$ , we deduce that

$$S \left( \begin{pmatrix} y_1^k - x_1^k \\ \vdots \\ y_{n-1}^k - x_{n-1}^k \\ x_n^k \end{pmatrix} \right) \ni \begin{pmatrix} x_1^k - x_n^k \\ \vdots \\ x_{n-1}^k - x_n^k \\ \sum_{i=1}^n y_i^k - \sum_{i=1}^n x_i^k \end{pmatrix} = \begin{pmatrix} x_1^k - x_n^k \\ \vdots \\ x_{n-1}^k - x_n^k \\ \sum_{i=1}^n (\sum_{j=1}^{n-1} N_{ij} - 1) x_i^k \end{pmatrix} \quad (18)$$

where  $S : \mathcal{H}^n \rightrightarrows \mathcal{H}^n$  denotes the operator given by

$$\mathcal{S} := \begin{pmatrix} (F_1)^{-1} \\ \vdots \\ (F_{n-1})^{-1} \\ F_n \end{pmatrix} + \begin{bmatrix} 0 & \dots & 0 & -\text{Id} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & -\text{Id} \\ \text{Id} & \dots & \text{Id} & 0 \end{bmatrix}.$$

Note that  $\mathcal{S}$  is maximally monotone operator as the sum of the two maximally monotone operators, the latter having full domain [4, Corollary 24.4(i)]. Consequently, its graph is sequentially closed in the weak-strong topology [4, Proposition 20.32]. Note also that the RHS of (18) converges strongly to zero as a consequence of (a) and Assumption 1(b).

Let  $\mathbf{x} \in \mathcal{H}^n$  be an arbitrary weak cluster point of the sequence  $(\mathbf{x}^k)$ . As a consequence of (a), it follows that  $\mathbf{x} = (x, \dots, x)$  for some  $x \in \mathcal{H}$ . Denote  $\mathbf{y} = \mathbf{S}\bar{\mathbf{z}} + \mathbf{N}\mathbf{x}$ . Taking the limit in (18) along a subsequence of  $(\mathbf{x}^k)$  which converges weakly to  $\mathbf{x}$  gives

$$\mathcal{S} \left( \begin{pmatrix} y_1 - x \\ \vdots \\ y_{n-1} - x \\ x \end{pmatrix} \right) \ni \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \implies \begin{cases} F_i(x) \ni y_i - x & \forall i \in \{1, \dots, n-1\} \\ F_n(x) \ni (n-1)x - \sum_{i=1}^{n-1} y_i. \end{cases}$$

Altogether, this shows that  $x = J_{F_1}(y_1) = J_{F_1}(S_1\bar{\mathbf{z}}) = \bar{x}$ . In other words,  $\mathbf{x} = (\bar{x}, \dots, \bar{x})$  is the unique weak sequential cluster point of the bounded sequence  $(\mathbf{x}^k)$ . We therefore deduce that  $(\mathbf{x}^k)$  converges weakly to  $\mathbf{x}$ , which completes the proof.  $\square$

**Remark 3** A closer look at the iteration (17) shows that, although convergence is analysed with respect to the sequence  $(\mathbf{z}^k)$ , this sequence does not need to be directly computed. Indeed, using the fact that  $\mathbf{S} = -\mathbf{M}^*$  and making the change of variables  $\mathbf{v}^k = \mathbf{S}\mathbf{z}^k$ , we may write (17) as

$$\mathbf{v}^{k+1} = \mathbf{v}^k - \gamma \mathbf{M}^* \mathbf{M} \mathbf{x}^k, \quad \mathbf{x}^k = J_{\mathbf{F}}(\mathbf{v}^k + \mathbf{N} \mathbf{x}^k).$$

In the case where  $m > n$ , this transformation also has the advantage of replacing a variable  $\mathbf{z}^k \in \mathcal{H}^m$  in the larger space with a variable  $\mathbf{v}^k \in \mathcal{H}^n$  in the smaller space.

### 3 Realisations of the framework

In what follows, we provide several realisations of Theorem 1 for solving the monotone inclusion (1). Examples 3, 4 and 5 are already known in the literature. However, Examples 6 and 7 are new.

**Example 3** (Douglas–Rachford splitting) Douglas–Rachford splitting [5, 7, 13, 21] for (1) with  $n = 2$  is defined by the operator  $T : \mathcal{H} \rightarrow \mathcal{H}$  given by

$$T(z) = z + \gamma(x_2 - x_1) \text{ where } \begin{cases} x_1 = J_{F_1}(z) \\ x_2 = J_{F_2}(2x_1 - z). \end{cases}$$

Thus, the corresponding coefficient matrices  $M \in \mathbb{R}^{1 \times 2}$ ,  $S \in \mathbb{R}^{2 \times 1}$  and  $N \in \mathbb{R}^{2 \times 2}$  are given by

$$M = -S^* = \begin{bmatrix} -1 & 1 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}.$$

Here, we have

$$M^*M + N + N^* - 2\text{Id} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \leq 0.$$

**Example 4** (Ryu splitting for  $n = 3$ ) Ryu's splitting algorithm [2, 17] for (1) with  $n = 3$  is defined by the operator  $T : \mathcal{H}^2 \rightarrow \mathcal{H}^2$  given by

$$T(\mathbf{z}) = \mathbf{z} + \gamma \begin{pmatrix} x_3 - x_1 \\ x_3 - x_2 \end{pmatrix} \text{ where } \begin{cases} x_1 = J_{F_1}(z_1) \\ x_2 = J_{F_2}(z_2 + x_1) \\ x_3 = J_{F_3}(x_1 - z_1 + x_2 - z_2). \end{cases}$$

Thus, the corresponding coefficient matrices  $M \in \mathbb{R}^{2 \times 3}$ ,  $S \in \mathbb{R}^{3 \times 2}$  and  $N \in \mathbb{R}^{3 \times 3}$  are given by

$$M = -S^* = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}. \quad (19)$$

It is straightforward to verify that Assumption 1(a)–(c) holds. To verify Assumption 1(d), observe that

$$M^*M + N + N^* - 2\text{Id} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \leq 0.$$

**Example 5** (Resolvent splitting with minimal lifting) The resolvent splitting algorithm considered in [3, 14] for (1) with  $n \geq 2$  is defined by the operator  $T : \mathcal{H}^{n-1} \rightarrow \mathcal{H}^{n-1}$  given by

$$T(\mathbf{z}) = \mathbf{z} + \gamma \begin{pmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_n - x_{n-1} \end{pmatrix} \text{ where } \begin{cases} x_1 = J_{F_1}(z_1), \\ x_i = J_{F_i}(z_i + x_{i-1} - z_{i-1}) \quad \forall i \in \{2, \dots, n-1\} \\ x_n = J_{F_n}(x_1 + x_{n-1} - z_{n-1}). \end{cases}$$

Thus, the coefficient matrices  $M \in \mathbb{R}^{(n-1) \times n}$ ,  $S \in \mathbb{R}^{n \times (n-1)}$  and  $N \in \mathbb{R}^{n \times n}$  are given by

$$M = -S^* = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

It is straightforward to verify that Assumption 1(a)–(c) holds. To verify Assumption 1(d), observe that

$$M^*M + N + N^* - 2\text{Id} = \begin{bmatrix} -1 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & -1 \end{bmatrix} \leq 0. \quad (20)$$

**Example 6** (An extension of Ryu splitting for  $n \geq 3$  operators) As explained in [14, Remark 4.7], the naïve extension of Ryu splitting to  $n > 3$  operators fails to converge. However, the more general perspective offered by our framework suggests an alternative.

Indeed, for (1) with  $n \geq 2$ , consider the operator  $T : \mathcal{H}^{n-1} \rightarrow \mathcal{H}^{n-1}$  given by

$$T(\mathbf{z}) = \mathbf{z} + \gamma \sqrt{\frac{2}{n-1}} \begin{pmatrix} x_n - x_1 \\ x_n - x_2 \\ \vdots \\ x_n - x_{n-1} \end{pmatrix}$$

where  $x_1, \dots, x_n \in \mathcal{H}$  are given by

$$\begin{cases} x_i = J_{F_i} \left( \sqrt{\frac{2}{n-1}} z_i + \frac{2}{n-1} \sum_{j=1}^{i-1} x_j \right) & \forall i \in \{1, \dots, n-1\} \\ x_n = J_{F_n} \left( \frac{2}{n-1} \sum_{j=1}^{n-1} x_j - \sqrt{\frac{2}{n-1}} \sum_{i=1}^{n-1} z_i \right). \end{cases}$$

This has coefficient matrices  $M \in \mathbb{R}^{(n-1) \times n}$ ,  $S \in \mathbb{R}^{n \times (n-1)}$  and  $L \in \mathbb{R}^{n \times n}$  are given by

$$M = -S^* = \sqrt{\frac{2}{n-1}} \begin{bmatrix} -1 & 0 & \dots & 0 & 1 \\ 0 & -1 & \dots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}, \quad N = \frac{2}{n-1} \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}.$$

Note that, when  $n = 3$ , these matrices are precisely those given in (19) of Example 4. Consequently, the setting considered in this example provides a one parameter family that coincides with Ryu splitting in the special case when  $n = 3$ .

It is straightforward to verify that Assumption 1(a)–(c) holds. To verify Assumption 1(d), observe that

$$M^*M + N + N^* - 2\text{Id} = \frac{2}{n-1} \begin{bmatrix} 2-n & 1 & \dots & 1 & 0 \\ 1 & 2-n & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & \dots & 2-n & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

This matrix is diagonally dominated with non-positive diagonal entries, hence it is negative definite.

Before our next example, we first recall some notational from graph theory (see, for instance, [10]). Consider an (undirected) graph  $G = (V, E)$  with vertex set  $V = \{v_1, \dots, v_n\}$  and edge set  $E = \{e_1, \dots, e_m\}$ . The *degree matrix* of  $G$  is the diagonal matrix  $D \in \mathbb{R}^{n \times n}$  such that  $D_{ii}$  is equal to the degree of  $v_i$ . The *adjacency matrix* of  $G$  is the matrix  $A \in \{0, 1\}^{n \times n}$  such that  $A_{ij} = 1$  if only if  $v_i$  and  $v_j$  are adjacent. The *graph Laplacian* is the matrix  $L \in \mathbb{R}^{|V| \times |V|}$  given by  $L = D - A$ . An *orientation* of an undirected graph is a directed graph obtained by assigning a direction to each edge in the original graph. The *incidence matrix* of a directed graph  $G = (V, E)$  is the matrix  $B \in \{0, \pm 1\}^{|V| \times |E|}$  given by

$$B_{ij} = \begin{cases} -1 & \text{if } e_j \text{ leaves vertex } v_i, \\ +1 & \text{if } e_j \text{ enters vertex } v_i, \\ 0 & \text{otherwise.} \end{cases} \quad (21)$$

Let  $B_\sigma$  denote an oriented incidence matrix of an undirected graph  $G$ , that is, an incidence matrix for any orientation of  $G$ . Then its graph Laplacian can also satisfies  $\mathcal{L} = B_\sigma B_\sigma^T$ .

**Example 7** (*d*-regular networks) Consider a simple, connected *d*-regular graph  $G = (V, E)$  with  $|V| = n$ . As a consequence of the Handshake Lemma,  $G$  necessarily has  $|E| = \frac{nd}{2}$  edges. Let  $B_\sigma$  be an oriented incidence matrix for  $G$ . Consider the coefficient matrices given by

$$-S^* = M = \sqrt{\frac{2}{d}} B_\sigma^T \in \mathbb{R}^{|E| \times |V|}$$

and  $N \in \mathbb{R}^{|V| \times |V|}$  being the lower triangular matrix (with zero diagonal) satisfying  $N + N^* = \frac{2}{d}A$ . We claim that these matrices satisfy Assumptions 1. Indeed, we have:

- (a) By (21), we have  $\ker M = \ker B_\sigma^T = \mathbb{R}e$  where  $e = (1, \dots, 1)^T \in \mathbb{R}^{|E|}$ .
- (b) Since  $\sum_{i,j=1}^n A_{ij} = 2|E|$ , it follows that  $\sum_{i,j=1}^n N_{ij} = \frac{1}{d} \sum_{i,j=1}^n A_{ij} = \frac{2|E|}{d} = n$ .
- (c)  $S^* = -M$  holds by definition.

(d) Since  $G$  is  $d$ -regular,  $L = B_\sigma B_\sigma^T = d \text{Id} - A$ . Hence

$$M^*M + N + N^* - 2 \text{Id} = \frac{2}{d}(d \text{Id} - A) + \frac{2}{d}A - 2 \text{Id} = 0 \leq 0.$$

Altogether, we have that the operator  $T : \mathcal{H}^{|E|} \rightarrow \mathcal{H}^{|E|}$  given by

$$T(\mathbf{z}) = \mathbf{z} + \gamma \mathbf{M}\mathbf{x}, \quad \mathbf{x} = J_{\mathbf{F}}(\mathbf{S}\mathbf{z} + \mathbf{N}\mathbf{x}) \quad (22)$$

satisfies the conditions of Sect. 2. It is interesting to interpret the variables in the context of the above example. Note that  $\mathbf{z} \in \mathcal{H}^{|E|}$  and  $\mathbf{x} \in \mathcal{H}^{|V|}$ , so that  $\mathbf{z}$  represents edges of  $G$  and  $\mathbf{x}$  represents vertices of  $G$ . The matrix  $M$  describes information flow from vertices to their incident edges. Similarly, the matrix  $S$  describes information flow from edges to their adjacent vertices. Finally,  $N$  describes a direct information flow between adjacent vertices.

Since  $|E|$  is potentially large, it is often better to avoid using the operator  $T$  directly. Using the observation outlined in Remark 3 and the fact that  $SM = -M^*M = -L$ , (22) can be rewritten in terms of the operator  $\tilde{T} : \mathcal{H}^{|V|} \rightarrow \mathcal{H}^{|V|}$  given by

$$\tilde{T}(\mathbf{v}) = \mathbf{v} - \gamma \frac{2}{d} \mathbf{L}\mathbf{x}, \quad \mathbf{x} = J_{\mathbf{F}}(\mathbf{v} + \mathbf{N}\mathbf{x}), \quad (23)$$

This is suited for a distributed implementation with node  $i$  responsible for computing  $v_i$  and  $x_i$ . The corresponding iteration for (23) is given explicitly in (9) (or (10)).

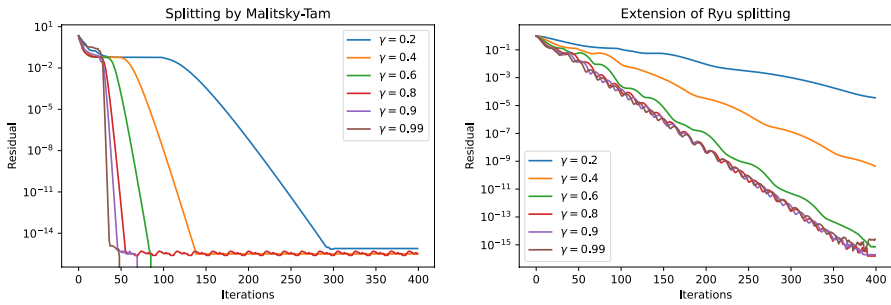
Note that the setup used in this example does not directly generalise to arbitrary network for two reasons. Firstly, (b) requires the number of edges in  $G$  to be exactly  $|E| = \frac{nd}{2}$ , and secondly, the computation in (d) used the fact that the degree matrix is given by  $D = d \text{Id}$ .

## 4 Numerical experiments

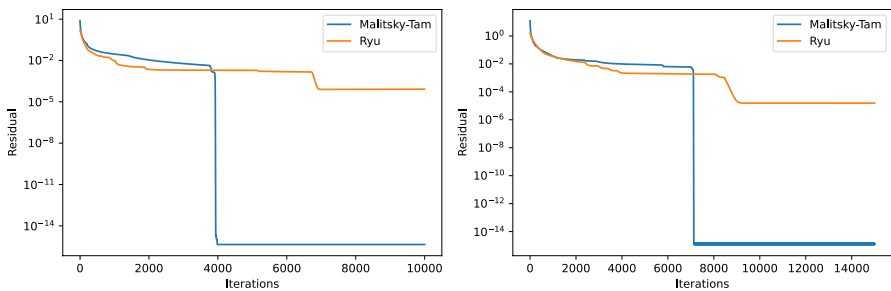
In this section, we demonstrate the algorithms presented in Sect. 3 on a toy problem from statistics. Given a finite dataset  $C := \{c_1, \dots, c_n\} \subseteq \mathbb{R}$ , we consider finding a point  $x \in \mathbb{R}$  which minimises dispersion to  $C$  as measured by the  $\ell_p$ -norm ( $p \geq 1$ ). That is, we consider

$$\min_{x \in \mathbb{R}} \left( \frac{1}{n} \sum_{i=1}^n |x - c_i|^p \right)^{1/p}. \quad (24)$$

The optimal value of (24) represents a measure of *dispersion* for the dataset  $C$ , and solutions of (24) represent measures of *central tendency*. Important examples of measures of dispersion/central tendency pairs within this framework include: *average absolute deviation/median* when  $p = 1$ , *standard deviation/mean* when  $p = 2$ , and *maximum deviation/midrange* when  $p = +\infty$  (which is interpreted via the limit for  $p \rightarrow +\infty$  in (24)).



**Fig. 1** The effect of varying the parameter  $\gamma \in (0, 1)$  for the splitting algorithm by Malitsky–Tam (left) and the proposed extension of Ryu’s splitting algorithm (right)



**Fig. 2** Results with  $\gamma = 0.99$  for problems with  $n = 100$  (left) and  $n = 250$  (right)

For the purpose of this demonstration, we focus on the setting where  $p = 1$ , which was previously considered in [14]. In this case, the sub-differential sum rule [4, Corollary 16.48] implies that solutions to (24) can be characterised by the monotone inclusion

$$0 \in \sum_{i=1}^n A_i(x) \subseteq \mathbb{R} \text{ where } A_i := \partial |\cdot - c_i|.$$

In all our experiments, problem instances are obtained by generating a vector  $c \in \mathbb{R}^n$  entrywise through sampling the standard normal distribution.

#### 4.1 Splitting algorithms with minimal lifting

Our first experiment is a comparison between frugal resolvent splitting algorithms for (14) with *minimal lifting*. That is, algorithms for which the underlying fixed point operator is defined on the space  $\mathcal{H}^{n-1}$ . For further details on minimal lifting, the reader is referred to [14, 17]. In particular, we compare the “resolvent splitting with minimal lifting” due to Malitsky & Tam [14] (see also Example 5) with our proposed extension of Ryu’s splitting algorithm [2, 17] from Example 6.

Figure 1 shows the effect of changing the parameter  $\gamma \in (0, 1)$  on the decay of the fixed point residual, as measured by the quantity  $\|(\text{Id} - T)(\mathbf{z}^k)\|/\gamma$  where  $(\mathbf{z}^k)$  denotes sequence generated by the algorithm, for a problem with  $n = 10$ . In both cases, the initial point  $\mathbf{z}^0$  was taken as the vector of all zeros. These results suggest that both algorithms perform better with larger values of  $\gamma$  for this scenario.

Figure 2 shows results for datasets of size  $n = 100$  and  $n = 250$ . For both algorithms,  $\gamma = 0.99$  was used as was best in Fig. 1. Figure 2 suggest that the splitting algorithm by Malitsky–Tam converges faster, although the proposed extension of Ryu’s algorithm is better when only a relatively fewer number of iteration can be computed.

## 4.2 Distributed splitting on regular networks

Our second experiment is a comparison between decentralised resolvent splitting algorithms for (14) on  $d$ -regular networks. In particular, we compare the algorithm proposed in Example 7 with *PDHG* [9, 11] (as discussed in [14, Section 5.1]) and

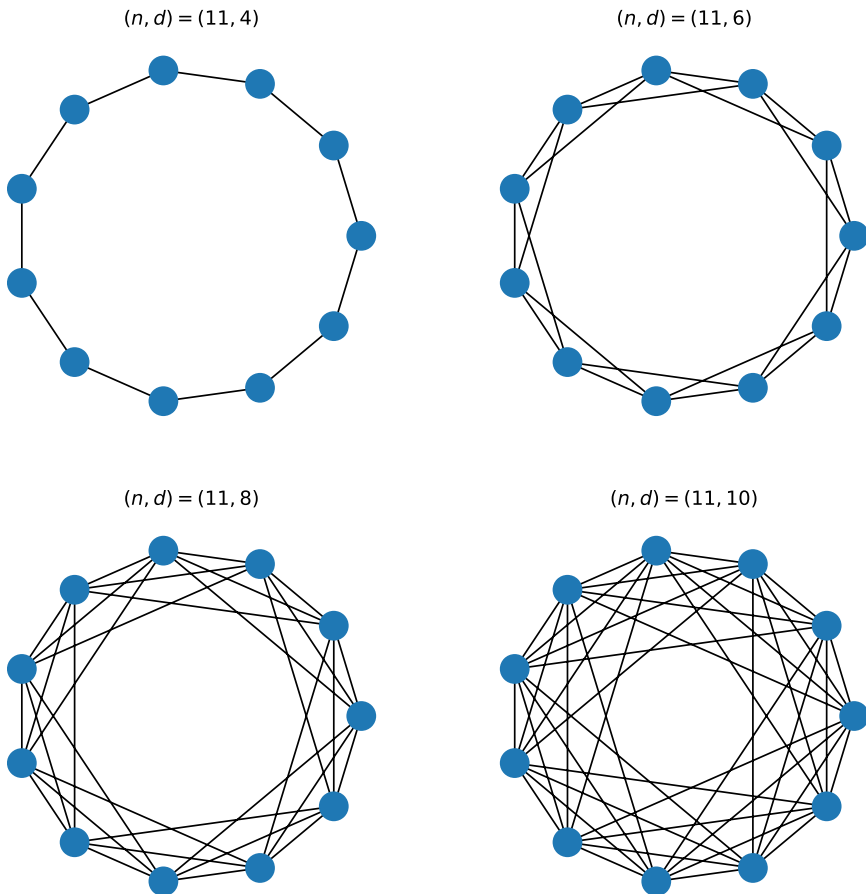
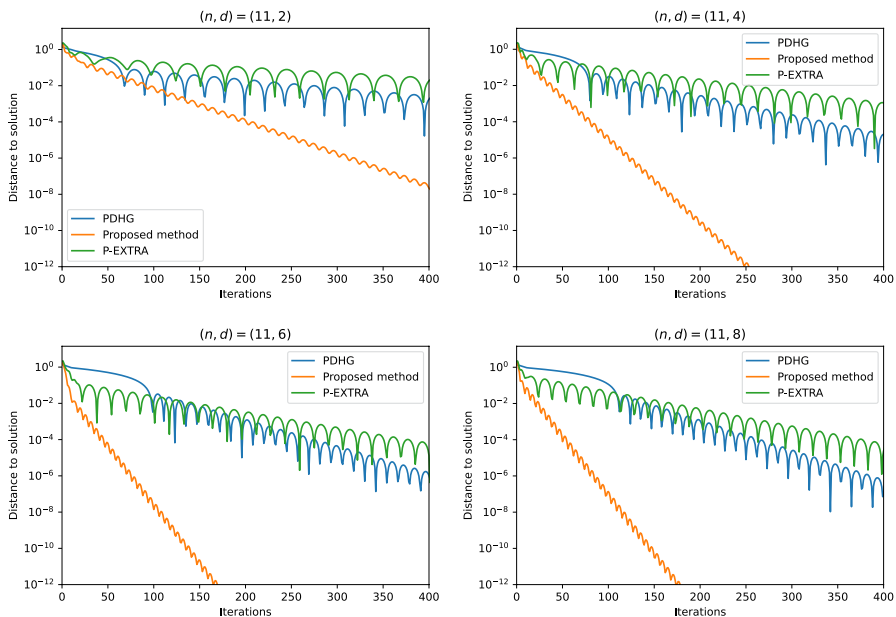


Fig. 3 The four regular networks used in the experiments

*P-EXTRA* [20]. Our experiments are performed using four  $d$ -regular networks with  $n = 11$  vertices and  $d \in \{2, 4, 6, 8\}$ , as shown in Fig. 3.

Parameters in the proposed method and PDHG were chosen by trial and error for best performance; for the proposed method,  $\gamma = 0.5$  was used, and for the PDHG,  $(\tau, \sigma) = (\frac{1}{10\|L\|^{1/2}}, \frac{10}{\|L\|^{1/2}})$  was used. For P-EXTRA, the mixing matrix  $W$  was taken to be *Laplacian-based constant edge weight matrix*, as suggested in [19, Section 2.4].

Figure 4 reports the distance of the current iterate to the solution for the three methods, on each of the networks in Fig. 3. The figure suggests favourable performance of the proposed method. However, this might be expected as the proposed method is more specialised — it can only be used on regular networks, whereas the other methods can be applied to arbitrary connected networks. Overall, all methods were observed to converge faster as the network connectivity increased, as expected.



**Fig. 4** Results for the four regular networks shown in Fig. 3

**Acknowledgements** The author would like to thank the associate editor, anonymous referee and Yura Malitsky for their comments which improved the manuscript. MKT is supported in part by Australian Research Council grants DE200100063 and DP230101749.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Aragón-Artacho, F.J., Boţ, R.I., Torregrosa-Belén, D.: A primal-dual splitting algorithm for composite monotone inclusions with minimal lifting. *Numer. Algorithms* **93**, 103 (2022)
2. Aragón Artacho, F.J., Campoy, R., Tam, M.K.: Strengthened splitting methods for computing resolvents. *Comput. Optim. Appl.* **80**(2), 549–585 (2021)
3. Aragón-Artacho, F.J., Malitsky, Y., Tam, M.K., Torregrosa-Belén, D.: Distributed forward-backward methods for ring networks. *Comput. Optim. Appl.* pp. 1–26 (2022)
4. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd edn. CMS Books in Mathematics, Springer International Publishing (2017)
5. Bauschke, H.H., Moursi, W.M.: On the Douglas–Rachford algorithm. *Math. Program.* **164**(1), 263–284 (2017)
6. Bertsekas, D., Tsitsiklis, J.: *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific (2015)
7. Búi, M.N., Combettes, P.L.: The Douglas–Rachford algorithm converges only weakly. *SIAM J. Control Optim.* **58**(2), 1118–1120 (2020)
8. Campoy, R.: A product space reformulation with reduced dimension for splitting algorithms. *Comput. Optim. Appl.* **83**(1), 319–348 (2022)
9. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011)
10. Chung, F.R.K.: *Lectures on spectral graph theory*. CBMS Lect. Fresno **6**(92), 17–21 (1996)
11. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximal and linear composite terms. *J. Optim. Theory Appl.* **158**(2), 460–479 (2013)
12. Condat, L., Kitahara, D., Contreras, A., Hirabayashi, A.: Proximal splitting algorithms for convex optimization: a tour of recent advances, with new twists. *SIAM Rev.* p. to appear (2022)
13. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **55**(1), 293–318 (1992)
14. Malitsky, Y., Tam, M.K.: Resolvent splitting for sums of monotone operators with minimal lifting. *Math. Program.* **201**, 231 (2022)
15. Morin, M., Banert, S., Giselsson, P.: Frugal splitting operators: Representation, minimal lifting and convergence. *arXiv preprint arXiv:2206.11177* (2022)
16. Rockafellar, R.T.: Monotone operators associated with saddle-functions and minimax problems. *J. Nonlinear Funct. Anal.* **18**, 397–407 (1970)
17. Ryu, E.K.: Uniqueness of DRS as the 2 operator resolvent-splitting and impossibility of 3 operator resolvent-splitting. *Math. Program.* **182**(1), 233–273 (2020)
18. Ryu, E.K., Yin, W.: *Large-Scale Convex Optimization*. Cambridge University Press (2022)
19. Shi, W., Ling, Q., Wu, G., Yin, W.: EXTRA: an exact first-order algorithm for decentralized consensus optimization. *SIAM J. Optim.* **25**(2), 944–966 (2015)

20. Shi, W., Ling, Q., Wu, G., Yin, W.: A proximal gradient algorithm for decentralized composite optimization. *IEEE Trans. Signal Process.* **63**(22), 6013–6023 (2015)
21. Svaiter, B.F.: On weak convergence of the Douglas–Rachford method. *SIAM J. Control. Optim.* **49**(1), 280–287 (2011)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.