SPLITTING THE FORWARD-BACKWARD ALGORITHM: A FULL CHARACTERIZATION

, A. ÅKERMAN, E. CHENCHENE, P. GISELSSON, AND E. NALDI

Introduction. Splitting algorithms are fundamental for nonsmooth optimization. They define a class of iterative methods that allow to solve highly structured problems by decomposing them into simpler components, each handled through efficient and easily computable operations. In this paper, we consider monotone inclusion problems of the form:

$$\text{Find } x \in \mathcal{H} \text{ such that:} \quad 0 \in \sum_{i=1}^{n} A_i(x) + \sum_{i=1}^{m} C_i(x), \tag{0.1}$$

where $\mathcal{H}$ is a real Hilbert space, each operator $A_i : \mathcal{H} \to 2^{\mathcal{H}}$ is maximal monotone, and $C_i : \mathcal{H} \to \mathcal{H}$ is $\frac{1}{\beta_i}$-cocoercive for some $\beta_i \in \mathbb{R}_+$. If the resolvent $J_{\gamma A} := (\mathrm{Id} + \gamma A)^{-1}$ with $A := A_1 + \cdots + A_n$ can be computed efficiently for any given step size $\gamma > 0$, one can address (0.1) through the *forward-backward* algorithm [7] which, from a starting point $x^0 \in \mathcal{H}$, iterates:

$$x^{k+1} = J_{\gamma A}(x^k - \gamma C(x^k)), \quad \text{for all } k \in \mathbb{N}, \tag{0.2}$$

where $C := C_1 + \cdots + C_m$. The sequence generated by (0.2) is well known to converge weakly to a solution to (0.1), if one exists and if $\gamma < \frac{2}{\beta}$. However, in practice the operator $J_{\gamma A}$ can rarely be computed efficiently. The classical alternative if $n = 2$ and $m = 0$ is the *Douglas–Rachford* splitting (DRS) algorithm [7] and if $m = 1$, the *Davis–Yin* method [6, 10], that, given $z^0 \in \mathcal{H}$, iterates:

$$\begin{cases} x_1^{k+1} = J_{\gamma A_1}(z^k), \\ x_2^{k+1} = J_{\gamma A_2}(2x_1^{k+1} - \gamma C(x_1^{k+1}) - z^k), \\ z^{k+1} = z^k + (x_2^{k+1} - x_1^{k+1}), \end{cases} \quad \text{for all } k \in \mathbb{N}. \tag{0.3}$$

which reduces to Douglas-Rachford splitting if $C = 0$, and to (0.2) if $A_1 = 0$.

General methods for solving (0.1) often use lifting tricks, applying a simpler method, like the proximal point method, or the Davis-Yin method, to a reformulation in a higher dimensional product-space, usually $\mathcal{H}^{n-1}$ or $\mathcal{H}^n$, see [10], the graph-DRS algorithmic framework [4] [4] for $m = 0$, and the graph-forward-backward algorithm [2] for $m = n - 1$. The need to introduce $n - 1$ (or more) variables is not a coincidence. A result by Ryu [11], later extended in [8, 9], demonstrated that these types of *frugal splitting methods* cannot solve arbitrary instances of (0.1) by storing fewer than $n - 1$ variables between iterations. For $n > 2$, the landscape of such methods becomes richer, giving rise to a variety of structurally diverse algorithms that have attracted significant attention in recent years [8, 9, 1, 3, 5]. These can all be understood as fixed-point iterations with respect to *averaged* operators, which lead us to the core question:

Can we characterize all averaged frugal splitting methods with minimal lifting to solve (0.1)*?*

To be more precise, we consider fixed-point iterations with respect to $T : \mathcal{H}^{n-1} \to \mathcal{H}^{n-1}$ to solve arbitrary instances of (0.1), with the following properties:

- (P1) *Splitting*: Each evaluation of T is constructed solely from resolvent evaluations of A_i, direct evaluations of C_i, and arbitrary linear combinations of their inputs and outputs.
- (P2) *Frugality*: Each operator is evaluated exactly once per evaluation of T, either directly for C_i or through a resolvent for A_i.
- (P3) *Minimal lifting*: The method only needs to store $n - 1$ variables between iterations.
- (P4) *Fixed-point encoding*: Fixed points of T correspond to solutions of (0.1) and vice versa.
- (P5) *Averaged nonexpansive*: The operator T is θ-averaged on $\mathcal{H}^{n-1}$, $\theta \in (0, 1)$, equipped with the product norm.

We show that the entire class of methods satisfying properties (P1)–(P5) is given by Algorithm 1. This represents the first, yet partial, answer to a challenging open problem proposed by Ryu in [11], on the characterization of all *unconditionally stable* methods. Algorithm 1 can be shown to encompass several methods in the literature, and also allows us to devise new ones with desirable numerical properties. In particular, based on Algorithm 1, we formulate methods that efficiently handle data heterogeneity and admits distributed implementations on general networks. We further study the influence of the four parameterizing matrices M, P and H, K and propose three heuristics to achieve excellent performance in practice.

Contributions.

- We show that Algorithm 1 parametrizes all θ-averaged frugal splitting methods with minimal lifting,
- We propose a special case of Algorithm 1 that can efficiently handle different operators with different cocoercivity constants, and admits distributed implementations on general networks without requiring the knowledge of a global cocoercivity constant.
- We propose heuristic choices of M, P and H, K that make Algorithm 1 particularly efficient compared to other existing special cases.

Algorithm 1: The complete class of methods for solving (0.1) satisfying (P1)–(P5).

Pick: A relaxation parameter $\theta \in (0,1)$, $M \in \mathbb{R}^{n\times(n-1)}$, with $\operatorname{Ran}(\mathbb{1}) = \operatorname{Null}(M^T)$, $P \in \mathbb{R}^{n\times(n-1)}$ with $\operatorname{Ran}(\mathbb{1}) \subset \operatorname{Null}(P^T)$ and $H, K^T \in \mathbb{R}^{n\times m}$, with necessary zero structure, and such that $H^T\mathbb{1} = K\mathbb{1} = \mathbb{1}$

Let: $S := MM^T + PP^T + \frac{1}{2}(H - K^T)\operatorname{diag}(\beta)(H^T - K)$ and $\gamma := 2\operatorname{diag}(S)^{-1}$

Input: $\boldsymbol{z}^0 = (z_1^0, \dots, z_{n-1}^0) \in \mathcal{H}^{n-1}$

for $k = 0, 1, 2, \dots$ **do**

  for $i = 1, 2, \dots, n$ **do**

$$x_i^{k+1} = J_{\gamma_i A_i}\Big(-\gamma_i \sum_{h=1}^{i-1} S_{ih} x_h^{k+1} - \gamma_i \sum_{j=1}^{m} H_{ij} C_j \Big(\sum_{h=1}^{i-1} K_{jh} x_h^{k+1} \Big) + \gamma_i \sum_{j=1}^{n-1} M_{ij} z_j^k \Big)$$

  for $i = 1, 2, \dots, n-1$ **do**

$$z_i^{k+1} = z_i^k - \theta \sum_{i=1}^{n} M_{ij} x_i^{k+1}$$

References

1. Francisco J. Aragón-Artacho, Radu I. Boţ, and David Torregrosa-Belén, *A primal-dual splitting algorithm for composite monotone inclusions with minimal lifting*, Numerical Algorithms **93** (2023), no. 1, 103–130.
2. Francisco J. Aragón-Artacho, Rubén Campoy, and César López-Pastor, *Forward-backward algorithms devised by graphs*, 2024.
3. Francisco J. Aragón-Artacho, Yura Malitsky, Matthew K. Tam, and David Torregrosa-Belén, *Distributed forward-backward methods for ring networks*, Computational Optimization and Applications **86** (2023), no. 3, 845–870.
4. Kristian Bredies, Enis Chenchene, and Emanuele Naldi, *Graph and Distributed Extensions of the Douglas—Rachford Method*, SIAM Journal on Optimization **34** (2024), no. 2, 1569–1594.
5. Rubén Campoy, *A product space reformulation with reduced dimension for splitting algorithms*, Computational Optimization and Applications **83** (2022), no. 1, 319–348.
6. Damek Davis and Wotao Yin, *A three-operator splitting scheme and its optimization applications*, Set-Valued and Variational Analysis **25** (2017), no. 4, 829–858.
7. Pierre-Louis Lions and Bertrand Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis **16** (1979), no. 6, 964–979.
8. Yura Malitsky and Matthew K. Tam, *Resolvent splitting for sums of monotone operators with minimal lifting*, Mathematical Programming **201** (2023), no. 1, 231–262.
9. Martin Morin, Sebastian Banert, and Pontus Giselsson, *Frugal splitting operators: Representation, minimal lifting, and convergence*, SIAM Journal on Optimization **34** (2024), no. 2, 1595–1621.
10. Hugo Raguet, Jalal M. Fadili, and Gabriel Peyré, *A generalized forward-backward splitting*, SIAM Journal on Imaging Sciences **6** (2013), no. 3, 1199–1226.
11. Ernest K. Ryu, *Uniqueness of DRS as the 2 operator resolvent-splitting and impossibility of 3 operator resolvent-splitting*, Mathematical Programming **182** (2020), no. 1, 233–273.