A relaxed version of Ryu's three-operator splitting method for structured nonconvex optimization

Jan Harold Alcantara and Felipe Atenas

Abstract In this work, we propose a modification of Ryu's splitting algorithm for minimizing the sum of three functions, where two of them are convex with Lipschitz continuous gradients, and the third is an arbitrary proper closed function that is not necessarily convex. The modification is essential to facilitate the convergence analysis, particularly in establishing a sufficient descent property for an associated envelope function. This envelope, tailored to the proposed method, is an extension of the well-known Moreau envelope. Notably, the original Ryu splitting algorithm is recovered as a limiting case of our proposal. The results show that the descent property holds as long as the stepsizes remain sufficiently small. Leveraging this result, we prove global subsequential convergence to critical points of the nonconvex objective.

1 Introduction

Modern optimization applications often exhibit a structured form that is well-suited to the application of decomposition techniques. Operator splitting methods have emerged as powerful tools for efficiently solving complex optimization problems. Instances of operator splitting methods include the forward-backward splitting method [17] and the Douglas-Rachford splitting method [13,15]. These algorithmic schemes have been pivotal in the development of techniques for image reconstruction, signal processing, and machine learning [8–11, 14].

Jan Harold Alcantara
Center for Advanced Intelligence Project, RIKEN, 1-4-1 Nihombashi, Chuo City, Tokyo, 103-0027, Japan e-mail: janharold.alcantara@riken.jp

Felipe Atenas
School of Mathematics and Statistics, The University of Melbourne, 813 Swanston St, Parkville, Melbourne, VIC, 3052, Australia e-mail: felipe.atenas@unimelb.edu.au

The aforementioned methods were originally designed to solve maximal monotone problems with a two-block structure. In the context of optimization problems, it translates to the sum of two convex functions. Several generalizations to three-block problems have been proposed in the literature, such as the Davis-Yin operator splitting method [12], Ryu's three-operator splitting method [19], and the Malitsky-Tam operator splitting method [16] (which is also applicable to n-block problems). All these methods, in the context of optimization, have provable convergence guarantees for convex optimization problems.

The literature is relatively scarce for methods that solve nonconvex optimization problems for the sum of three functions. Notable contributions in this direction include the Davis-Yin splitting to nonconvex problems analyzed in [7], the extension of Davis-Yin to four-operator splitting investigated in [1], and an extension of Douglas-Rachford splitting examined in [2] (that also works for n-operators). In this work, we propose a modification of Ryu's splitting method to solve nonconvex three-block optimization problems with a specific structure.

Consider the optimization problem

$$\min_{x\in\mathbb{R}^d} \varphi(x) := f_1(x) + f_2(x) + f_3(x), \tag{1}$$

where $f_i : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ for $i = 1,2,3$. Given $z_1^0, z_2^0 \in \mathbb{R}^d$, $\gamma > 0$ and $\lambda > 0$, the (relaxed) *Ryu's three-operator splitting method* [19] is given by the following iteration: for $k \geq 1$,

$$\begin{cases} x_1^k = \operatorname{prox}_{\gamma f_1}(z_1^k) \\ x_2^k = \operatorname{prox}_{\gamma f_2}(z_2^k + x_1^k) \\ x_3^k \in \operatorname{prox}_{\gamma f_3}\left(x_1^k - z_1^k + x_2^k - z_2^k\right) \\ \begin{pmatrix} z_1^{k+1} \\ z_2^{k+1} \end{pmatrix} = \begin{pmatrix} z_1^k \\ z_2^k \end{pmatrix} + \lambda \begin{pmatrix} x_3^k - x_1^k \\ x_3^k - x_2^k \end{pmatrix}, \end{cases} \tag{2}$$

where $\operatorname{prox}_{\gamma f}$ denotes the *proximal mapping* of γf defined in (4) below.

When the functions f_1, f_2 and f_3 in problem (1) are convex, [19, Theorem 4] states the convergence of the method to minimizers of problem (1). Our objective in this work is to investigate whether the convergence guarantees can be extended to a nonconvex setting.

In our convergence analysis, we adopt the "envelope technique", a strategy that has been employed in the analysis of several splitting algorithms for nonconvex problems [1–3, 20, 21]. A key challenge, however, lies in the fact that the standard form of Ryu's splitting algorithm is not directly amenable to this type of analysis. To address this, we consider the following relaxed variant of Ryu's method: given $z_1^0, z_2^0 \in \mathbb{R}^d$, $\alpha > 0$, $\gamma > 0$ and $\lambda > 0$, we define the iterates for all $k \geq 1$ as follows:

$$\begin{cases} x_1^k = \operatorname{prox}_{\gamma f_1}(z_1^k) \\ x_2^k = \operatorname{prox}_{\frac{\gamma}{\alpha} f_2}(\frac{z_2^k}{\alpha} + x_1^k) \\ x_3^k \in \operatorname{prox}_{\gamma f_3}\left(x_1^k - z_1^k + x_2^k - z_2^k\right) \\ \begin{pmatrix} z_1^{k+1} \\ z_2^{k+1} \end{pmatrix} = \begin{pmatrix} z_1^k \\ z_2^k \end{pmatrix} + \lambda \begin{pmatrix} x_3^k - x_1^k \\ x_3^k - x_2^k \end{pmatrix}. \end{cases} \tag{3}$$

In particular, we modify the f_2-proximal step in (2) to obtain (3). When $\alpha = 1$, this reduces to Ryu's splitting algorithm.

Throughout this paper, we adopt the following blanket assumptions.

Assumption 1 (Blanket assumption) *Suppose the following conditions are satisfied.*

(a) For $i = 1,2$, the function $f_i : \mathbb{R}^d \to \mathbb{R}$ is a convex L_i-smooth function, that is, ∇f_i is globally Lipschitz continuous with modulus $L_i > 0$.
(b) The function $f_3 : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is proper and lower semicontinuous (lsc for short).
(c) Problem (1) has a nonempty set of solutions.

Remark 1 (On the simplified nonconvex setting). The convexity in Assumption 1(a) can be relaxed. However, we impose it in our setting to simplify the analysis. Similar results can be obtained when f_1 and f_2 are merely L_1- and L_2-smooth, respectively, by employing a strategy similar to that in [7, 21].

This paper is organized as follows. In Section 2, we introduce the notation and some variational analysis results we use throughout this paper. In Section 3, we define the merit function we use as the foundation of our analysis − the Ryu's three-operator splitting envelope − and establish key properties relevant to the convergence analysis of the proposed method. Section 4 is dedicated to investigating the convergence properties of the method (3). In particular, we show that the defined envelope satisfies a sufficient decrease condition, and then we exploit this property to prove that all cluster points of the generated sequence are critical points of the objective function of problem (1). Finally, in Section 5 we comment on some ongoing works and future research directions.

2 Preliminaries and notation

Throughout this paper, $\langle \cdot, \cdot \rangle$ denotes an inner product in $\mathbb{R}^d$, and $\| \cdot \|$ its induced norm. We shall make use of the following technical result.

Lemma 1. *For all $a, b, c, d \in \mathbb{R}^d$, it holds*

$$\|a-b\|^2 - \|a-c\|^2 = -\|b-c\|^2 + 2\langle c-b, a-b\rangle.$$

A function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is called proper when its domain, the set $\mathrm{dom}(\varphi) = \{x \in \mathbb{R}^d : \varphi(x) < +\infty\}$, is nonempty. We say φ is lsc if at any $x \in \mathrm{dom}(\varphi)$, $f(x) \leq \liminf_{y \to x} f(y)$, and it is convex if for all $x, y \in \mathrm{dom}(\varphi)$, and all $\beta \in (0,1)$, $\varphi(\beta x + (1-\beta)y) \leq \beta\varphi(x) + (1-\beta)\varphi(y)$. A function $\Phi : \mathbb{R}^d \to \mathbb{R}$ is said to be L-smooth if it is differentiable and its gradient is L-Lipschitz continuous, that is, for all $x, y \in \mathbb{R}^d$, $\|\nabla\Phi(x) - \nabla\Phi(y)\| \leq L\|x-y\|$. In the next result, we recall some propeties of L-smooth functions that will be important in our analysis (see, for instance, [6, Theorem 5.8]).

Lemma 2. *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ is L-smooth. Then,*

$$(\forall x, y \in \mathbb{R}^d)\ |f(x) - f(y) - \langle \nabla f(y), x - y\rangle| \leq \frac{L}{2}\|y - x\|^2.$$

If, in addiiton, f is convex, then

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2.$$

For a proper function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, and $x \in \mathrm{dom}(\varphi)$, we denote by $\hat{\partial}\varphi(x)$ the Fréchet (or regular) subdifferential of φ at x, defined as

$$\hat{\partial}\varphi(x) = \left\{ v \in \mathbb{R}^d : \liminf_{y \to x} \frac{\varphi(y) - \varphi(x) - \langle v, y - x\rangle}{\|y - x\|} \geq 0 \right\}.$$

The limiting (or general) subdifferential of φ at x, denoted $\partial\varphi(x)$, is defined as

$$\partial\varphi(x) = \limsup_{y \xrightarrow[\varphi]{} x} \hat{\partial}\varphi(x),$$

where $y \xrightarrow[\varphi]{} x$ denotes convergence in the attentive sense, that is, $y \to x$ and $\varphi(y) \to \varphi(x)$. When φ is smooth, then $\hat{\partial}\varphi(x) = \partial\varphi(x) = \{\nabla\varphi(x)\}$. If φ is proper lsc convex, then the Fréchet and limiting subdifferentials coincide with the subdifferential of convex analysis, namely,

$$\{v \in \mathbb{R}^d : (\forall y \in \mathbb{R}^d)\ \varphi(y) \geq \varphi(x) + \langle v, y - x\rangle\}.$$

A set of points of particular interest is the zeros of the subdifferential operator. We say $\bar{x} \in \mathbb{R}^d$ is a critical point of φ if $0 \in \partial\varphi(\bar{x})$. If φ is convex, critical points are exactly the global minimizers of φ.

Remark 2 (Critical points of the sum). Under Assumption 1 and for φ defined in (1), in view of [18, Exercise 8.8(c)], $\bar{x}$ is a critical point of φ if

$$0 \in \nabla f_1(\bar{x}) + \nabla f_2(\bar{x}) + \partial f_3(\bar{x}).$$

Given a function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, a parameter $\gamma > 0$, and a point $z \in \mathbb{R}^d$, the proximal operator of $\gamma\varphi$ at x is defined as

$$\operatorname{prox}_{\gamma\varphi}(z) := \operatorname*{argmin}_{y\in\mathbb{R}^d} \ \varphi(y) + \frac{1}{2\gamma}\|y-z\|^2. \tag{4}$$

The associated optimal value function is known as the Moreau envelope, defined as follows:

$$\varphi_\gamma^{\texttt{Moreau}}(z) := \min_{y\in\mathbb{R}^d} \ \varphi(y) + \frac{1}{2\gamma}\|y-z\|^2.$$

We say φ is prox-bounded if, for some $\gamma > 0$, $\varphi(\cdot) + \frac{1}{2\gamma}\|\cdot\|^2$ is bounded from below. The supremum $\gamma_\varphi > 0$ of such parameters γ is called the threshold of prox-boundedness. If φ is a proper lsc prox-bounded function with threshold $\gamma_\varphi > 0$, then for any $\gamma \in (0, \gamma_\varphi)$, $\operatorname{prox}_{\gamma\varphi}$ is nonempty and compact-valued, and $\varphi_{\gamma\varphi}^{\texttt{Moreau}}$ is finite-valued [18, Theorem 1.25]. In particular, if φ is a proper lsc convex function, then $\gamma_\varphi = +\infty$, and $\operatorname{prox}_{\gamma\varphi}$ is a single-valued mapping [18, Theorem 12.12, Theorem 12.17].

Remark 3 (On the well-definedness of proximal operations). Under Assumption 1, $\operatorname{prox}_{\gamma f_1}$ and $\operatorname{prox}_{\gamma f_2}$ are single-valued since f_1 and f_2 are proper lsc convex, while $\operatorname{prox}_{\gamma f_3}$ is well-defined for $\gamma < \frac{1}{L_1+L_2}$, as f_3 is prox-bounded with threshold at least $\frac{1}{L_1+L_2}$ from [20, Remark 3.1].

3 An envelope for Ryu's splitting method

Our convergence analysis for (3) under Assumption 1 builds on the approach in [1, 2, 20], which utilizes an envelope function, akin to the Moreau envelope, well-suited to the corresponding iterative method. We begin our analysis by motivating such merit function.

Proposition 1. *Suppose that Assumption 1 holds, and let $\alpha > 0$ and $(z_1^k, z_2^k) \in \mathbb{R}^d \times \mathbb{R}^d$. Then, x_3^k defined in (3) solves the following minimization problem*

$$\min_{y\in\mathbb{R}^d} \left\{ f_3(y) + \sum_{i=1}^{2} \left[f_i(x_i^k) + \langle y - x_i^k, \nabla f_i(x_i^k)\rangle + \frac{1}{2\gamma_i}\|y - x_i^k\|^2 \right] \right\},$$

where $\gamma_1 := \frac{\gamma}{\alpha}$ and $\gamma_2 := \frac{\gamma}{1-\alpha}$.[1]

Proof. From the x_3-update in (3), we have

$$x_3^k \in \operatorname*{argmin}_{y\in\mathbb{R}^d} \left\{ f_3(y) + \frac{1}{2\gamma}\|y - (x_1^k - z_1^k + x_2^k - z_2^k)\|^2 \right\}. \tag{5}$$

---

[1] We adopt the convention that $\frac{c}{0} = \infty$ and $\frac{d}{\infty} = 0$ for any $c > 0$ and $d \in \mathbb{R}$. Hence, when $\alpha = 1$, $\gamma_2 = \infty$ and the term $\frac{1}{2\gamma_i}\|y - x_i^k\|^2$ vanishes when $i = 2$.

Meanwhile, the first-order optimality conditions of the x_1-update and the x_2-update yield, respectively,

$$z_1^k = \gamma \nabla f_1(x_1^k) + x_1^k \tag{6}$$

$$z_2^k = \gamma \nabla f_2(x_2^k) + \alpha x_2^k - \alpha x_1^k. \tag{7}$$

Hence,

$$\begin{aligned} x_1^k - z_1^k + x_2^k - z_2^k &= x_1^k - \left(\gamma \nabla f_1(x_1^k) + x_1^k\right) + x_2^k - \left(\gamma \nabla f_2(x_2^k) + \alpha x_2^k - \alpha x_1^k\right) \\ &= \alpha x_1^k - \gamma \nabla f_1(x_1^k) + (1-\alpha) x_2^k - \gamma \nabla f_2(x_2^k) \\ &= \alpha \left(x_1^k - \frac{\gamma}{\alpha} \nabla f_1(x_1^k)\right) + (1-\alpha) \left(x_2^k - \frac{\gamma}{1-\alpha} \nabla f_2(x_2^k)\right), \end{aligned}$$

which, combined with (5), yields

$$x_3^k \in \operatorname*{argmin}_{y \in \mathbb{R}^d} \left\{ f_3(y) + \frac{1}{2\gamma} \left\| y - \left(\sum_{i=1}^2 \alpha_i (x_i^k - \gamma_i \nabla f_i(x_i^k) \right) \right\|^2 \right\} \tag{8}$$

for $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$. Following the calculations in [2, Theorem 5.5], by expanding the squared norm, dropping the constant term $\left\| \sum_{i=1}^2 \alpha_i (x_i^k - \gamma_i \nabla f_i(x_i^k)) \right\|^2$, and adding the constant term $\sum_{i=1}^2 f_i(x_i^k)$, we get the desired result. □

In view of Proposition 1, we define the *Relaxed Ryu envelope* (RRE), for all $(z_1, z_2) \in \mathbb{R}^d \times \mathbb{R}^d$, by

$$\varphi_\gamma^{\mathtt{Ryu}}(z_1, z_2) := \min_{y \in \mathbb{R}^d} \left\{ f_3(y) + \sum_{i=1}^2 \left[f_i(x_i) + \langle y - x_i, \nabla f_i(x_i) \rangle + \frac{1}{2\gamma_i} \|y - x_i\|^2 \right] \right\}. \tag{9}$$

Meanwhile, the corresponding set-valued iteration operator associated with (3) is given by

$$T_\gamma^{\mathtt{Ryu}} : (z_1, z_2) \mapsto (z_1 + \lambda(x_3 - x_1), z_2 + \lambda(x_3 - x_2)),$$

where

$$\begin{cases} x_1 = \operatorname{prox}_{\gamma f_1}(z_1) \\ x_2 = \operatorname{prox}_{\frac{\gamma}{\alpha} f_2}(\frac{z_2}{\alpha} + x_1) \\ x_3 \in \operatorname{prox}_{\gamma f_3}\left(x_1 - z_1 + x_2 - z_2\right). \end{cases} \tag{10}$$

In view of Remark 3, note that $T_\gamma^{\mathtt{Ryu}}$ is nonempty- and compact-valued for any $(z_1, z_2) \in \mathbb{R}^d \times \mathbb{R}^d$ provided that $\gamma < \frac{1}{L_1 + L_2}$.

Remark 4 (Relationship between the Moreau envelope and the RRE). Observe that from (8), we have

$$\varphi_\gamma^{\texttt{Ryu}}(z_1,z_2) = \varphi_{\gamma f_3}^{\texttt{Moreau}}\left(\sum_{i=1}^{2}\alpha_i(x_i-\gamma_i\nabla f_i(x_i))\right) - \left\|\sum_{i=1}^{2}\alpha_i(x_i-\gamma_i\nabla f_i(x_i))\right\|^2 + \sum_{i=1}^{2} f_i(x_i).$$

We now establish some properties of the envelope function. The following result states that the RRE inherits continuity properties of the Moreau envelope (cf. [21, Proposition 4.2] and [20, Proposition 3.2]).

Proposition 2 (Continuity of RRE). *The RRE is a real-valued and locally Lipschitz continuous function.*

Proof. Since $\mathrm{prox}_{\gamma f_1}$ and $\mathrm{prox}_{\frac{\gamma}{\alpha} f_2}$ are nonexpansive [6, Theorem 6.42(b)], then the maps $z_1 \mapsto x_1$ and $(z_1,z_2)\mapsto x_2$ defined in (10) are (globally) Lipschitz continuous. Furthermore, from Assumption 1, ∇f_i is (globally) Lipschitz continuous, for $i=1,2$. Then, as the Moreau envelope is locally Lipschitz continuous [18, Example 10.32], the conclusion follows from Remark 4. $\square$

Next, we show some sandwich-type bounds relating the RRE and the original objective function in problem (1) (cf. [21, Proposition 4.3] and [20, Proposition 3.3]).

Proposition 3. *Suppose Assumption 1 holds, $\gamma \in (0, \frac{1}{L_1+L_2})$, and $\lambda,\alpha > 0$. Then, given $(z_1,z_2)\in\mathbb{R}^d\times\mathbb{R}^d$, consider $(x_1,x_2,x_3)\in\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d$ defined by (10). Then,*

(i) For $\gamma_1=\frac{\gamma}{\alpha}$ and $\gamma_2=\frac{\gamma}{1-\alpha}$,

$$\varphi_\gamma^{Ryu}(z_1,z_2) \le \min\left\{\varphi(x_1)+\frac{1}{2}\left(L_2+\frac{1}{\gamma_2}\right)\|x_1-x_2\|^2, \varphi(x_2)+\frac{1}{2}\left(L_1+\frac{1}{\gamma_1}\right)\|x_1-x_2\|^2\right\}.$$

(ii) $\varphi_\gamma^{Ryu}(z_1,z_2) \ge \varphi(x_3)+\frac{\alpha-\gamma L_1}{2\gamma}\|x_3-x_1\|^2+\frac{(1-\alpha)-\gamma L_2}{2\gamma}\|x_3-x_2\|^2$.

(iii) If $\alpha\in(0,1)$ and $\gamma\le\min\left\{\frac{\alpha}{L_1},\frac{1-\alpha}{L_2}\right\}$, then $\varphi_\gamma^{Ryu}(z_1,z_2)\ge\varphi(x_3)$.

Proof. Take $y=x_1$ in (9) and apply Lemma 2 to $f=f_2$ to obtain

$$\begin{aligned}\varphi_\gamma^{\texttt{Ryu}}(z_1,z_2) &\le f_3(x_1)+f_1(x_1)+f_2(x_2)+\langle\nabla f_2(x_2),x_1-x_2\rangle+\tfrac{1}{2\gamma_2}\|x_1-x_2\|^2\\ &\le \varphi(x_1)+\frac{1}{2}\left(L_2+\tfrac{1}{\gamma_2}\right)\|x_1-x_2\|^2.\end{aligned}$$

Similarly, taking $y=x_2$ in (9) and applying Lemma 2 to $f=f_1$ yields

$$\varphi_\gamma^{\texttt{Ryu}}(z_1,z_2) \le \varphi(x_2)+\frac{1}{2}\left(L_1+\tfrac{1}{\gamma_1}\right)\|x_2-x_1\|^2.$$

From these, we immediately get (i). Moreover, since $y = x_3$ minimizes the problem in (9),

$$\begin{aligned}\varphi_\gamma^{\mathtt{Ryu}}(z_1,z_2) &= f_3(x_3) + \sum_{i=1}^{2}\left[f_i(x_i) + \langle \nabla f_i(x_i), x_3 - x_i\rangle + \frac{1}{2\gamma_i}\|x_3 - x_i\|^2\right] \\ &\geq f_3(x_3) + \sum_{i=1}^{2}\left[f_i(x_3) - \frac{L_i}{2}\|x_3 - x_i\|^2 + \frac{1}{2\gamma_i}\|x_3 - x_i\|^2\right] \\ &= \varphi(x_3) + \tfrac{\alpha - \gamma L_1}{2\gamma}\|x_3 - x_1\|^2 + \tfrac{(1-\alpha)-\gamma L_2}{2\gamma}\|x_3 - x_2\|^2,\end{aligned}$$

where the inequality holds by Lemma 2 and the last equality holds by plugging in $\gamma_1 = \frac{\gamma}{\alpha}$ and $\gamma_2 = \frac{\gamma}{1-\alpha}$. This completes the proof of (ii). Finally, part (iii) immediately follows from (ii). □

Note that the iteration in (3) is designed to find a fixed point of the relaxed Ryu splitting operator $T_\gamma^{\mathtt{Ryu}}$, that is, a point in the set

$$\operatorname{Fix} T_\gamma^{\mathtt{Ryu}} := \left\{(z_1,z_2) \in \mathbb{R}^d \times \mathbb{R}^d : (z_1,z_2) \in T_\gamma^{\mathtt{Ryu}}(z_1,z_2)\right\}.$$

In the next proposition, we establish a connection between such fixed points and the notion of criticality, as commonly used in optimization.

Proposition 4. *Suppose Assumption 1 holds, and let* $\gamma \in (0, \frac{1}{L_1+L_2})$ *and* $\alpha, \lambda > 0$. *Then,* $(\bar z_1, \bar z_2) \in \operatorname{Fix} T_\gamma^{Ryu}$ *if and only if* $\bar x := \bar x_1 = \bar x_2 = \bar x_3$, *where*

$$\begin{cases} \bar x_1 = \operatorname{prox}_{\gamma f_1}(\bar z_1) \\ \bar x_2 = \operatorname{prox}_{\frac{\gamma}{\alpha} f_2}(\frac{\bar z_2}{\alpha} + \bar x_1) \\ \bar x_3 \in \operatorname{prox}_{\gamma f_3}\big(\bar x_1 - \bar z_1 + \bar x_2 - \bar z_2\big). \end{cases} \tag{11}$$

Furthermore, such $\bar x$ *is a critical point of* (1), *and* $\varphi(\bar x) = \varphi_\gamma^{Ryu}(\bar z_1, \bar z_2)$. *In particular, if* $\alpha \in (0,1)$ *and* $\gamma \leq \min\left\{\frac{\alpha}{L_1}, \frac{1-\alpha}{L_2}\right\}$, *then*

$$\min \varphi = \min \varphi_\gamma^{Ryu}.$$

Proof. It is straightforward from the definition of $T_\gamma^{\mathtt{Ryu}}$ that $(\bar z_1, \bar z_2) \in \operatorname{Fix} T_\gamma^{\mathtt{Ryu}}$ is equivalent to having $\bar x := \bar x_1 = \bar x_2 = \bar x_3$ satisfying the conditions in (11). Hence, it suffices to prove that such $\bar x$ is, in this case, always a critical point of φ. Evaluating the first-order optimality conditions of the x_3-step, namely,

$$0 \in \gamma \partial f_3(x_3) + x_3 - x_1 + z_1 - x_2 + z_2 \tag{12}$$

at $(x_1,x_2,x_3) = (\bar x, \bar x, \bar x)$ and $z_i = \bar z_i$, for $i = 1,2$, yields

$$0 \in \gamma \partial f_3(\bar x) + \bar z_1 - \bar x + \bar z_2.$$

Adding this inclusion with (6) and (7) using the substitutions $x_1^k \leftarrow \bar{x}$ and $x_2^k \leftarrow \bar{x}$, we obtain

$$0 \in \gamma\big(\nabla f_1(\bar{x}) + \nabla f_2(\bar{x}) + \partial f_3(\bar{x})\big),$$

that is, $0 \in \partial\varphi(\bar{x})$ by Remark 2.

Furthermore, given $(\bar{z}_1, \bar{z}_2) \in \operatorname{Fix} T_\gamma^{\mathtt{Ryu}}$, Proposition 3(i)&(ii) imply that $\varphi(\bar{x}) = \varphi_\gamma^{\mathtt{Ryu}}(\bar{z}_1, \bar{z}_2)$. Finally, in view of Proposition 3(iii), for any $(z_1, z_2) \in \mathbb{R}^d \times \mathbb{R}^d$, and $\bar{x} \in \operatorname{argmin} \varphi$,

$$\varphi_\gamma^{\mathtt{Ryu}}(\bar{z}_1, \bar{z}_2) = \varphi(\bar{x}) \le \varphi(x_3) \le \varphi_\gamma^{\mathtt{Ryu}}(z_1, z_2).$$

Hence, $(\bar{z}_1, \bar{z}_2) \in \operatorname{argmin} \varphi_\gamma^{\mathtt{Ryu}}$ and $\min \varphi = \min \varphi_\gamma^{\mathtt{Ryu}}$. □

At this point, we have built the necessary tools regarding the RRE for the analysis of convergence of the proposed relaxed Ryu splitting method. In the next section, we show subsequential convergence of the method under standard assumptions in the literature.

4 Convergence of modified Ryu's three-operator splitting method via envelopes

To establish the (subsequential) convergence of the iterative method in (3), we follow the approach introduced in [20] for the Douglas–Rachford splitting method. Specifically, the core argument relies on a sufficient decrease property satisfied by the RRE.

4.1 Sufficient decrease property for RRE

We first prove three technical lemmas that we will use in the main result of this section. We use the following notation: for $i = 1, 2$,

$$\begin{aligned} \Delta x_i^k &= x_i^{k+1} - x_i^k, \\ \Delta g_i^k &= \nabla f_i(x_i^{k+1}) - \nabla f_i(x_i^k), \\ \Delta z_i^k &= z_i^{k+1} - z_i^k. \end{aligned} \tag{13}$$

Lemma 3. *Under Assumption 1, the sequences* $(x_i^k)_k$ *for* $i = 1, 2, 3$*, and* $(z_i^k)_k$ *for* $i = 1, 2$ *generated by* (3) *satisfy:*

$$\|x_3^k - x_1^k\|^2 - \|x_3^k - x_1^{k+1}\|^2 = \left(\frac{2}{\lambda} - 1\right)\|\Delta x_1^k\|^2 + \frac{2\gamma}{\lambda}\langle \Delta x_1^k, \Delta g_1^k\rangle \tag{14}$$

and

$$\|x_3^k - x_2^k\|^2 - \|x_3^k - x_2^{k+1}\|^2 = \left(\frac{2\alpha}{\lambda} - 1\right)\|\Delta x_2^k\|^2 + \frac{2\gamma}{\lambda}\langle \Delta x_2^k, \Delta g_2^k\rangle - \frac{2\alpha}{\lambda}\langle \Delta x_2^k, \Delta x_1^k\rangle. \tag{15}$$

Proof. Using the notations (13), we have from (6) and (7) that

$$\Delta z_1^k = \gamma \Delta g_1^k + \Delta x_1^k \tag{16}$$
$$\Delta z_2^k = \gamma \Delta g_2^k + \alpha \Delta x_2^k - \alpha \Delta x_1^k. \tag{17}$$

On the other hand, Lemma 1 and the (z_1, z_2)-update in (3) yield

$$\|x_3^k - x_i^k\|^2 - \|x_3^k - x_i^{k+1}\|^2 = -\|\Delta x_i^k\|^2 + 2\langle \Delta x_i^k, x_3^k - x_i^k\rangle = -\|\Delta x_i^k\|^2 + \frac{2}{\lambda}\langle \Delta x_i^k, \Delta z_i^k\rangle.$$

Together with (16) and (17), we get (14) and (15), respectively. □

Lemma 4. *Under Assumption 1, the sequences $(x_i^k)_k$ for $i = 1,2,3$, and $(z_i^k)_k$ for $i = 1,2$ generated by (2) satisfy:*

$$\begin{aligned}\sum_{i=1}^{2} f_i(x_i^k) - f_i(x_i^{k+1}) - \langle \nabla f_i(x_i^{k+1}), x_3^k - x_i^{k+1}\rangle + \langle \nabla f_i(x_i^k), x_3^k - x_i^k\rangle \\ \geq \sum_{i=1}^{2}\left(\frac{1}{2L_i} - \frac{\gamma}{\lambda}\right)\|\Delta g_i^k\|^2 - \frac{1}{\lambda}\langle \Delta g_1^k, \Delta x_1^k\rangle \\ - \frac{\alpha}{\lambda}\langle \Delta g_2^k, \Delta x_2^k\rangle + \frac{\alpha}{\lambda}\langle \Delta g_2^k, \Delta x_1^k\rangle.\end{aligned}$$

Proof. For $i = 1,2$, we have from Lemma 2 and the (z_1, z_2)-update rule in (3) that

$$\begin{aligned} & f_i(x_i^k) - f_i(x_i^{k+1}) - \langle \nabla f_i(x_i^{k+1}), x_3^k - x_i^{k+1}\rangle + \langle \nabla f_i(x_i^k), x_3^k - x_i^k\rangle \\ =& f_i(x_i^k) - f_i(x_i^{k+1}) - \langle \nabla f_i(x_i^{k+1}), x_i^k - x_i^{k+1}\rangle - \langle \Delta g_i^k, x_3^k - x_i^k\rangle \\ \geq& \frac{1}{2L_i}\|\Delta g_i^k\|^2 - \langle \Delta g_i^k, x_3^k - x_i^k\rangle \\ =& \frac{1}{2L_i}\|\Delta g_i^k\|^2 - \frac{1}{\lambda}\langle \Delta g_i^k, \Delta z_i^k\rangle, \end{aligned} \tag{18}$$

Meanwhile, we have from (16) that

$$\langle \Delta g_1^k, \Delta z_1^k\rangle = \gamma\|\Delta g_1^k\|^2 + \langle \Delta g_1^k, \Delta x_1^k\rangle. \tag{19}$$

On the other hand, (17) yields

$$\langle \Delta g_2^k, \Delta z_2^k\rangle = \gamma\|\Delta g_2^k\|^2 + \alpha\langle \Delta g_2^k, \Delta x_2^k\rangle - \alpha\langle \Delta g_2^k, \Delta x_1^k\rangle. \tag{20}$$

Combining (18), (19) and (20) gives the desired inequality. □

Lemma 5. *Let $\lambda \in (0,2)$ and let $\underline{\alpha} := \frac{2\lambda - 3 + \sqrt{9 - 4\lambda}}{2}$. Then the following holds:*

(i) $\underline{\alpha} \in \left(\frac{\lambda}{2}, 1\right)$.

(ii) The interval $\left(\dfrac{\alpha}{2\alpha-\lambda}, \dfrac{2-\lambda}{1-\alpha}\right)$ *is nonempty for any* $\alpha \in (\underline{\alpha}, 1)$.

(iii) For $\varepsilon_1, \varepsilon_2 > 0$ *and* $\alpha \in (\underline{\alpha}, 1)$, *define the constants*

$$\begin{aligned} \bar{\gamma}_1 &:= \frac{\lambda}{2L_2} - \frac{\alpha}{2\varepsilon_2}, \\ \bar{\gamma}_2 &:= \frac{\alpha(2-\lambda-(1-\alpha)\varepsilon_1)}{\alpha\varepsilon_2 + 2(1-\alpha)L_1}, \\ \bar{\gamma}_3 &:= \frac{(1-\alpha)(\varepsilon_1(2\alpha-\lambda)-\alpha)}{2\alpha L_2 \varepsilon_1}, \end{aligned} \tag{21}$$

and the intervals

$$I_1 := \left(\frac{\alpha}{2\alpha-\lambda}, \frac{2-\lambda}{1-\alpha}\right) \quad \text{and} \quad I_2 := \left(\frac{\alpha L_2}{\lambda}, +\infty\right). \tag{22}$$

If $\varepsilon_j \in I_j$ *for* $j = 1, 2$, *then* $\bar{\gamma}_i$ *is strictly positive for* $i = 1, 2, 3$.

Proof. These results follow from straightforward calculations. □

With these lemmas in place, we are now ready to present the first main result of this paper. We establish the sufficient descent property of the RRE, provided that the stepsize is chosen sufficiently small. In particular, we restrict the stepsize γ in the interval

$$\Gamma := (0, \min\{\bar{\gamma}_0, \bar{\gamma}_1\}] \cap \left(0, \min\left\{\bar{\gamma}_2, \bar{\gamma}_3, \frac{1}{L_1+L_2}\right\}\right), \tag{23}$$

where $\bar{\gamma}_0 := \frac{\lambda}{2L_1}$, and$\bar{\gamma}_i$ is given by (21) for $i = 1, 2, 3$, with ε_j taken from I_j given in (22) for $j = 1, 2$.

Theorem 1 (RRE sufficient descent). *Suppose that Assumption 1 holds. Let* $\alpha \in (\underline{\alpha}, 1)$ *where* $\underline{\alpha} := \frac{2\lambda-3+\sqrt{9-4\lambda}}{2}$, *and* $\lambda \in (0, 2)$. *Let* $\gamma \in \Gamma$, *where* Γ *is given in* (23). *For the sequences* $(z_i^k)_k$ *for* $i = 1, 2$ *generated by* (3), *there exists* $M = M(\gamma) > 0$ *such that for all* $k \geq 1$,

$$\varphi_\gamma^{Ryu}(z_1^k, z_2^k) \geq \varphi_\gamma^{Ryu}(z_1^{k+1}, z_2^{k+1}) + M(\|z_1^{k+1} - z_1^k\|^2 + \|z_2^{k+1} - z_2^k\|^2).$$

Proof. From the definition of the relaxed Ryu envelope, we have

$$\begin{aligned} &\varphi_\gamma^{\mathtt{Ryu}}(z_1^{k+1}, z_2^{k+1}) \\ &\leq f_3(x_3^k) + \textstyle\sum_{i=1}^2 \left(f_i(x_i^{k+1}) + \langle \nabla f_i(x_i^{k+1}), x_3^k - x_i^{k+1} \rangle + \dfrac{1}{2\gamma_i} \|x_3^k - x_i^{k+1}\|^2 \right). \end{aligned}$$

Thus, together with Lemma 1, we have

$$
\begin{aligned}
&\varphi_\gamma^{\texttt{Ryu}}(z_1^k,z_2^k)-\varphi_\gamma^{\texttt{Ryu}}(z_1^{k+1},z_2^{k+1})\\
&\geq \sum_{i=1}^{2}\left(f_i(x_i^k)-f_i(x_i^{k+1})+\langle\nabla f_i(x_i^k),x_3^k-x_i^k\rangle-\langle\nabla f_i(x_i^{k+1}),x_3^k-x_i^{k+1}\rangle\right)\\
&+\sum_{i=1}^{2}\frac{1}{2\gamma_i}\left(\|x_3^k-x_i^k\|^2-\|x_3^k-x_i^{k+1}\|^2\right)
\end{aligned}
$$

Using Lemmas 3 and 4, we obtain

$$
\begin{aligned}
&\varphi_\gamma^{\texttt{Ryu}}(z_1^k,z_2^k)-\varphi_\gamma^{\texttt{Ryu}}(z_1^{k+1},z_2^{k+1})\\
&\geq\left(\frac{1}{2L_1}-\frac{\gamma}{\lambda}\right)\left\|\Delta g_1^k\right\|^2+\left(\frac{1}{2L_2}-\frac{\gamma}{\lambda}\right)\left\|\Delta g_2^k\right\|^2\\
&+\left(\frac{\alpha}{\gamma\lambda}-\frac{\alpha}{2\gamma}\right)\left\|\Delta x_1^k\right\|^2+\left(\frac{\alpha(1-\alpha)}{\gamma\lambda}-\frac{1-\alpha}{2\gamma}\right)\left\|\Delta x_2^k\right\|^2\\
&+\frac{\alpha-1}{\lambda}\left\langle\Delta x_1^k,\Delta g_1^k\right\rangle+\frac{1-2\alpha}{\lambda}\left\langle\Delta x_2^k,\Delta g_2^k\right\rangle\\
&-\frac{(1-\alpha)\alpha}{\gamma\lambda}\left\langle\Delta x_2^k,\Delta x_1^k\right\rangle+\frac{\alpha}{\lambda}\left\langle\Delta x_1^k,\Delta g_2^k\right\rangle.
\end{aligned}
$$

Since $\alpha<1$, then

$$
\begin{aligned}
&\frac{\alpha-1}{\lambda}\left\langle\Delta x_1^k,\Delta g_1^k\right\rangle\geq\frac{(\alpha-1)L_1}{\lambda}\left\|\Delta x_1^k\right\|^2,\\
&\frac{1-2\alpha}{\lambda}\left\langle\Delta x_2^k,\Delta g_2^k\right\rangle=\frac{1-\alpha}{\lambda}\left\langle\Delta x_2^k,\Delta g_2^k\right\rangle-\frac{\alpha}{\lambda}\left\langle\Delta x_2^k,\Delta g_2^k\right\rangle\geq 0-\frac{\alpha L_2}{\lambda}\left\|\Delta x_2^k\right\|^2,
\end{aligned}
$$

where the first inequality holds by L_1-smoothness of f_1, while the second inequality holds by the convexity of f_2 and L_2-smoothness of f_2. Moreover, by Young's inequality, we have

$$
\begin{aligned}
-\frac{(1-\alpha)\alpha}{\gamma\lambda}\left\langle\Delta x_2^k,\Delta x_1^k\right\rangle&\geq-\frac{(1-\alpha)\alpha}{\gamma\lambda}\left(\frac{\left\|\Delta x_2^k\right\|^2}{2\varepsilon_1}+\frac{\varepsilon_1\left\|\Delta x_1^k\right\|^2}{2}\right),\\
\frac{\alpha}{\lambda}\left\langle\Delta x_1^k,\Delta g_2^k\right\rangle\geq-\frac{\alpha}{\lambda}\left\|\Delta x_1^k\right\|\left\|\Delta g_2^k\right\|&\geq-\frac{\alpha}{\lambda}\left(\frac{\left\|\Delta g_2^k\right\|^2}{2\varepsilon_2}+\frac{\varepsilon_2\left\|\Delta x_1^k\right\|^2}{2}\right),
\end{aligned}
$$

where $\varepsilon_1,\varepsilon_2>0$ are arbitrary. With these, we obtain

$$
\begin{aligned}
&\varphi_\gamma^{\texttt{Ryu}}(z_1^k,z_2^k)-\varphi_\gamma^{\texttt{Ryu}}(z_1^{k+1},z_2^{k+1})\\
&\geq C_0\left\|\Delta g_1^k\right\|^2+C_1\left\|\Delta g_2^k\right\|^2+C_2\left\|\Delta x_1^k\right\|^2+C_3\left\|\Delta x_2^k\right\|^2,
\end{aligned}
\tag{24}
$$

where the constants C_i for $i=0,1,2,3$ are given by

$$
\begin{aligned}
C_0 &= \frac{1}{2L_1} - \frac{\gamma}{\lambda} \\
C_1 &= \frac{1}{2L_2} - \frac{\gamma}{\lambda} - \frac{\alpha}{2\lambda\varepsilon_2} \\
C_2 &= \frac{\alpha}{\gamma\lambda} - \frac{\alpha}{2\gamma} + \frac{(\alpha-1)L_1}{\lambda} - \frac{(1-\alpha)\alpha\varepsilon_1}{2\gamma\lambda} - \frac{\alpha\varepsilon_2}{2\lambda} \\
C_3 &= \frac{\alpha(1-\alpha)}{\gamma\lambda} - \frac{1-\alpha}{2\gamma} - \frac{\alpha L_2}{\lambda} - \frac{(1-\alpha)\alpha}{2\gamma\lambda\varepsilon_1}.
\end{aligned}
$$

Choosing $\varepsilon_j \in I_j$ for $j = 1,2$, it is not difficult to compute that $C_0, C_1 \geq 0$ since $\gamma \leq \min\{\bar{\gamma}_0, \bar{\gamma}_1\}$. In addition, $C_2, C_3 > 0$ since $\gamma < \min\{\bar{\gamma}_2, \bar{\gamma}_3\}$. Hence, for the given stepsize γ, we obtain from (24) that

$$
\varphi_\gamma^{\texttt{Ryu}}(z_1^k, z_2^k) - \varphi_\gamma^{\texttt{Ryu}}(z_1^{k+1}, z_2^{k+1}) \geq C_2 \left\|\Delta x_1^k\right\|^2 + C_3 \left\|\Delta x_2^k\right\|^2. \tag{25}
$$

Meanwhile, from (16), (17) and L_i-smoothness of f_i for $i = 1,2$, it follows

$$
\|\Delta z_1^k\|^2 \leq (1 + L_1\gamma)^2 \|\Delta x_1^k\|^2
$$

$$
\|\Delta z_2^k\|^2 \leq 2(\alpha + \gamma L_2)^2 \|\Delta x_2^k\|^2 + 2\alpha^2 \|\Delta x_1^k\|^2.
$$

Defining

$$
C_4 = \min\{C_2, C_3\}, \quad C_5 = \max\{2\alpha^2 + \left(1 + L_1\gamma\right)^2, 2(\alpha + L_2\gamma)^2\},
$$

we obtain from (25) that

$$
\begin{aligned}
\varphi_\gamma^{\texttt{Ryu}}(z^k) - \varphi_\gamma^{\texttt{Ryu}}(z^{k+1}) &\geq C_4(\|\Delta x_1^k\|^2 + \|\Delta x_2^k\|^2) \\
&\geq \frac{C_4}{C_5}(\|\Delta z_1^k\|^2 + \|\Delta z_2^k\|^2).
\end{aligned}
$$

The proof concludes by setting $M = C_4/C_5$. □

Theorem 1 suggests that our modification of Ryu's three-operator splitting method behaves like a descent method for the suitably defined RRE. Hence, one could expect this method to converge. In the next section, we formalize this idea.

4.2 Convergence properties of the modified Ryu's algorithm

The convergence analysis of the method in (3) relies on a particular relationship between the RRE and a Lagrangian associated with a reformulation of problem (1), which we proceed to define. By duplicating variables, problem (1) can be reformulated as follows

$$
\min_{x_1, x_2, x_3 \in \mathbb{R}^d} f_1(x_1) + f_2(x_2) + f_3(x_3) \quad \text{s.t. } x_1 = x_3,\ x_2 = x_3.
$$

We define the following Lagrangian associated with this problem reformulation:

$$\mathscr{L}_{\beta_1,\beta_2}(x_1,x_2,x_3,\mu_1,\mu_2) = f_3(x_3) + \sum_{i=1}^{2}\left(f_i(x_i) + \langle \mu_i, x_i - x_3\rangle + \frac{\beta_i}{2}\|x_i - x_3\|^2\right),$$

where, for $i = 1,2$, $\mu_i \in \mathbb{R}^d$ is a Lagrangian multiplier associated with the constraint $x_i = x_3$. The next result states the aforementioned relationship between the RRE and the augmented Lagrangian.

Lemma 6. *Suppose Assumption 1 holds. Let $\gamma \in (0, \frac{1}{L_1+L_2})$, and $\lambda, \alpha > 0$. Denoting*

$$\xi^k = \left(x_1^k, x_2^k, x_3^k, \gamma^{-1}(x_1^k - z_1^k), \gamma^{-1}(\alpha(x_2^k - x_1^k) - z_2^k)\right),$$

then

$$(\forall k \geq 1)\ \varphi_\gamma^{Ryu}(z_1^k, z_2^k) = \mathscr{L}_{\frac{1}{\gamma_1},\frac{1}{\gamma_2}}(\xi^k).$$

Proof. From (6)–(7), $\nabla f_1(x_1^k) = \gamma^{-1}(z_1^k - x_1^k)$ and $\nabla f_2(x_2^k) = \gamma^{-1}(z_2^k - \alpha(x_2^k - x_1^k))$. Substituting these identities into (9) yields the result.

We are now ready to state the second main result of this paper. We follow the approach taken in [3, 20].

Theorem 2 (Subsequential convergence of nonconvex Ryu's three-operator splitting). *Suppose that Assumption 1 holds. Let $\alpha \in (\underline{\alpha}, 1)$ where $\underline{\alpha} := \frac{2\lambda - 3 + \sqrt{9-4\lambda}}{2}$, and $\lambda \in (0,2)$. Let $\gamma \in \Gamma$ such that $\gamma \leq \min\left\{\frac{\alpha}{L_1}, \frac{1-\alpha}{L_2}\right\}$, where Γ is given in (23). Then, for any sequence $((x_1^k, x_2^k, x_3^k), (z_1^k, z_2^k))_k$ generated by algorithm (3), then*

(i) $x_i^{k+1} - x_i^k \to 0$ for $i = 1,2,3$, and $z_i^{k+1} - z_i^k \to 0$ and $x_3^k - x_i^k \to 0$ for $i = 1,2$.

If, in addition, $((x_1^k, x_2^k, x_3^k), (z_1^k, z_2^k))_k$ is bounded, then

(ii) For any cluster point x^* of $(x_3^k)_k$, $\varphi_\gamma(z^k) \to \varphi(x^*)$, and $f_1(x_1^k) + f_2(x_2^k) + f_3(x_3^k) \to \varphi(x^*)$.

(iii) All cluster points of the sequences $(x_i^k)_k$, for $i = 1,2,3$, coincide and are critical points of problem (1).

Proof. From the assumption, $\min \varphi > -\infty$, then Proposition 4 implies $(\varphi_\gamma^{\mathrm{Ryu}}(z_1^k, z_2^k))_k$ is a bounded non-increasing sequence, thus convergent to some $\varphi_\gamma^\star \in \mathbb{R}$. As a consequence, for $i = 1,2$, Theorem 1 yields $z_i^{k+1} - z_i^k \to 0$, and the z_i^k-update rule in (3) gives $x_3^k - x_i^k \to 0$. Since f_1 is convex, then $\mathrm{prox}_{\gamma f_1}$ is nonexpansive. In particular,

$$\|x_1^{k+1} - x_1^k\| \leq \|z_1^{k+1} - z_1^k\|$$

so that $x_1^{k+1} - x_1^k \to 0$. Likewise, $\mathrm{prox}_{\gamma f_2}$ is nonexpansive, yielding

$$\|x_2^{k+1} - x_2^k\| \leq \frac{1}{\alpha}\|z_2^{k+1} - z_2^k\| + \|x_1^{k+1} - x_1^k\|,$$

and thus $x_2^{k+1} - x_2^k \to 0$. Moreover, as

$$x_3^{k+1} - x_3^k = x_3^{k+1} - x_2^{k+1} + x_2^{k+1} - x_2^k + x_2^k - x_3^k,$$

then $x_3^{k+1} - x_3^k \to 0$ as well. This proves (i). Next, observe that $x_3^k - x_1^k \to 0$ and $x_3^k - x_2^k \to 0$ imply that the sequences $(x_1^k)_k$, $(x_2^k)_k$ and $(x_3^k)_k$ have the same cluster points. Let x^* be a cluster point of $(x_3^k)_k$ and (z_1^*, z_2^*) be a cluster point of $(z_1^k, z_2^k)_k$, and let $x_i^{k_j} \to x^*$ for $i = 1,2,3$, and $z_i^{k_j} \to z_i^*$ for $i = 1,2$. Note that from continuity of the proximal operator, $x^* = \mathrm{prox}_{\gamma f_1}(z_1^*)$ and $x^* = \mathrm{prox}_{\frac{\gamma}{\alpha} f_2}(\frac{z_2^*}{\alpha} + x^*)$. Then

$$\begin{aligned}
\varphi(x^*) &\le \liminf_j \varphi(x_3^{k_j}) && (\varphi \text{ is lsc}) \\
&\le \limsup_j \varphi(x_3^{k_j}) \\
&\le \limsup_j \varphi_\gamma^{\texttt{Ryu}}(z_1^{k_j}, z_2^{k_j}) && (\text{Proposition 3(ii)}) \\
&= \varphi_\gamma^{\texttt{Ryu}}(z_1^*, z_2^*) && (\text{Proposition 2}) \\
&\le \varphi(x^*) && (\text{Proposition 3(i)}).
\end{aligned}$$

Hence, $\varphi_\gamma^\star = \lim_k \varphi_\gamma^{\texttt{Ryu}}(z_1^k, z_2^k) = \lim_j \varphi(x_3^{k_j}) = \varphi(x^*)$. Furthermore, from Lemma 6, $\lim_k \mathscr{L}_{\frac{1}{\gamma_1},\frac{1}{\gamma_2}}(\xi^k) = \varphi(x^*)$. Since $(x_1^k)_k$, $(x_2^k)_k$, $(z_1^k)_k$ and $(z_2^k)_k$ are bounded, then part (i) implies $\lim_k \mathscr{L}_{\frac{1}{\gamma_1},\frac{1}{\gamma_2}}(\xi^k) = \lim_k f_1(x_1^k) + f_2(x_2^k) + f_3(x_3^k)$, from where (ii) follows. Earlier in the proof we already showed the first part of item (iii). It remains to prove that x^* is a critical point. A simple computation shows that for any $g_3 \in \partial f_3(x_3)$,

$$\begin{pmatrix}
\nabla f_1(x_1) + \mu_1 + \dfrac{1}{\gamma_1}(x_1 - x_3) \\
\nabla f_2(x_2) + \mu_2 + \dfrac{1}{\gamma_2}(x_2 - x_3) \\
g_3 - \left(\mu_1 + \mu_2 + \frac{1}{\gamma_1}(x_3 - x_1) + \frac{1}{\gamma_2}(x_3 - x_2)\right)
\end{pmatrix}
\in \hat{\partial}_{(x_1,x_2,x_3)} \mathscr{L}_{\frac{1}{\gamma_1},\frac{1}{\gamma_2}}(x_1, x_2, x_3, \mu_1, \mu_2).$$

Hence, taking $x_i = x_i^{k_j}$ for $i = 1,2,3$,

$$\mu_1^{k_j} = \gamma^{-1}(x_1^{k_j} - z_1^{k_j}), \text{ and } \mu_2^{k_j} = \gamma^{-1}(\alpha(x_2^{k_j} - x_1^{k_j}) - z_2^{k_j}).$$

Then, in view of (6), (7), and (12), we get

$$\begin{pmatrix} \frac{1}{\gamma_1}(x_1^{k_j} - x_3^{k_j}) \\ \frac{1}{\gamma_2}(x_2^{k_j} - x_3^{k_j}) \\ \frac{2\alpha}{\gamma}(x_1^{k_j} - x_2^{k_j}) + \frac{2}{\gamma}(x_2^{k_j} - x_3^{k_j}) \end{pmatrix} \in \hat{\partial}_{(x_1,x_2,x_3)} \mathscr{L}_{\frac{1}{\gamma_1},\frac{1}{\gamma_2}}(x_1^{k_j}, x_2^{k_j}, x_3^{k_j}, \mu_1^{k_j}, \mu_2^{k_j}).$$

Taking the limit as $j \to \infty$, since $x_i^{k_j} \to x^*$ for $i = 1,2,3$, and $z_i^{k_j} \to z_i^*$ for $i = 1,2$, item (i) then implies

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \in \partial_{(x_1,x_2,x_3)} \mathscr{L}_{\frac{1}{\gamma_1},\frac{1}{\gamma_2}}(x^*, x^*, x^*, \mu_1^*, \mu_2^*),$$

where $\mu_1^* = \gamma^{-1}(x^* - z_1^*)$, and $\mu_2^* = -\gamma^{-1} z_2^*$. In turn, this inclusion is equivalent to the following conditions:

$$\begin{aligned} 0 &= \nabla f_1(x^*) + \gamma^{-1}(x^* - z_1^*) \\ 0 &= \nabla f_2(x^*) - \gamma^{-1} z_2^* \\ 0 &\in \partial f_3(x^*) - \gamma^{-1}(x^* - z_1^* - z_2^*). \end{aligned}$$

Addind these equations yields $0 \in \nabla f_1(x^*) + \nabla f_2(x^*) + \partial f_3(x^*) = \partial \varphi(x^*)$, concluding the proof.

Remark 5 (Boundedness of the generated sequences). Using a similar argument to [21, Theorem 3.4(iii)], we can show that if φ has bounded level sets, then $\varphi_\gamma^{\mathtt{Ryu}}$ also has bounded level sets. Since for appropriately chosen stepsize, $(\varphi_\gamma^{\mathtt{Ryu}}(z_1^k, z_2^k))_k$ is a non-decreasing sequence as shown in Theorem 1, then for all $k \geq 1$, $(z_1^k, z_2^k) \in \{(z_1, z_2) : \varphi_\gamma^{\mathtt{Ryu}}(z_1, z_2) \leq \varphi_\gamma^{\mathtt{Ryu}}(z_1^0, z_2^0)\}$, and thus $(z_1^k, z_2^k)_k$ is bounded. Moreover, boundedness of $(x_1^k, x_2^k)_k$ follows from the continuity properties of $\mathrm{prox}_{\gamma f_1}$ and $\mathrm{prox}_{\frac{\gamma}{\alpha} f_2}$, and boundedness of $(x_3^k)_k$ is a consequence of Theorem 2(i).

Remark 6 (Global convergence). In Theorem 2, we establish subsequential convergence of the relaxed variant of Ryu's three-operator splitting method in a specific nonconvex setting. A natural question is whether the method converges globally. This can be affirmatively addressed by adopting the now standard approach based on the Kurdyka–Łojasiewicz (KL) inequality [5], or alternatively, by leveraging the subdifferential-based error bound technique used in [3] within the unifying framework proposed in [4]. Under the same set of assumptions, one can also obtain local linear convergence rates.

5 Conclusion

By defining a Moreau-type envelope tailored to the algorithmic scheme in (3), the core of the analysis relies on the sufficient decrease property shown in Theorem 1. Observe that our analysis does not cover the limiting case $\alpha = 1$ (corresponding to the original method proposed by Ryu in (2)), as it would make the stepsize interval in (23) empty. Nevertheless, when α is sufficiently close to 1, we can still guarantee global subsequential convergence for sufficiently small stepsizes. A full treatment of the limiting case $\alpha = 1$ requires a more refined analysis, which is the subject of ongoing work by the authors.

Acknowledgments

The authors thank the mathematical research institute MATRIX in Australia where part of this research was performed. JHA's visit at MATRIX was supported in part by the MATRIX-Simons Travel Grant. The research of FA was supported in part by Australian Research Council grant DP230101749.

References

1. J. H. Alcantara, C.-p. Lee, and A. Takeda. A four-operator splitting algorithm for nonconvex and nonsmooth optimization. *arXiv preprint arXiv:2406.16025*, 2024.
2. J. H. Alcantara and A. Takeda. Douglas-Rachford algorithm for nonmonotone multioperator inclusion problems. *arXiv preprint arXiv:2501.02752*, 2025.
3. F. Atenas. Understanding the Douglas–Rachford splitting method through the lenses of Moreau-type envelopes. *Computational Optimization and Applications*, 90:1–30, 2025.
4. F. Atenas, C. Sagastizábal, P. J. Silva, and M. Solodov. A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods. *SIAM Journal on Optimization*, 33(1):89–115, 2023.
5. H. Attouch, J. Bolte, and B. F. Svaiter. Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Mathematical programming*, 137(1):91–129, 2013.
6. A. Beck. *First-Order Methods in Optimization*. SIAM - Society for Industrial and Applied Mathematics, Philadelphia, PA, United States, 2017.
7. F. Bian and X. Zhang. A three-operator splitting algorithm for nonconvex sparsity regularization. *SIAM Journal on Scientific Computing*, 43(4):A2809–A2839, 2021.
8. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine learning*, 3(1):1–122, 2011.
9. J.-F. Cai, S. Osher, and Z. Shen. Split Bregman methods and frame based image restoration. *Multiscale Modeling & Simulation*, 8(2):337–369, 2010.
10. P. L. Combettes and J.-C. Pesquet. A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574, 2007.
11. P. L. Combettes and J.-C. Pesquet. *Proximal Splitting Methods in Signal Processing*, pages 185–212. Springer New York, New York, NY, 2011.

12. D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-valued and variational analysis*, 25:829–858, 2017.
13. J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439, 1956.
14. R. Glowinski, S. Osher, and W. Yin. *Splitting Methods in Communication, Imaging, Science, and Engineering*. Scientific Computation. Springer International Publishing, Cham, 2017.
15. P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.
16. Y. Malitsky and M. K. Tam. Resolvent splitting for sums of monotone operators with minimal lifting. *Mathematical Programming*, 201(1):231–262, 2023.
17. G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383–390, 1979.
18. R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer Verlag Berlin, Berlin, 3rd printing edition, 2009.
19. E. K. Ryu. Uniqueness of DRS as the 2 operator resolvent-splitting and impossibility of 3 operator resolvent-splitting. *Mathematical Programming*, 182:233–273, 2020.
20. A. Themelis and P. Patrinos. Douglas–Rachford splitting and ADMM for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, 2020.
21. A. Themelis, L. Stella, and P. Patrinos. Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone linesearch algorithms. *SIAM Journal on Optimization*, 28(3):2274–2303, 2018.