Convergence Analysis of Douglas-Rachford Splitting Method for “Strongly + Weakly” Convex Programming

Guo, Ke; Han, Deren; Yuan, Xiaoming

Published in:
SIAM Journal on Numerical Analysis

DOI:
10.1137/16M1078604

Published: 06/07/2017

SIAM J. NUMER. ANAL.
Vol. 55, No. 4, pp. 1549–1577

CONVERGENCE ANALYSIS OF DOUGLAS–RACHFORD SPLITTING METHOD FOR "STRONGLY + WEAKLY" CONVEX PROGRAMMING*

KE GUO[†], DEREN HAN[†], AND XIAOMING YUAN[‡]

Abstract. We consider the convergence of the Douglas–Rachford splitting method (DRSM) for minimizing the sum of a strongly convex function and a weakly convex function; this setting has various applications, especially in some sparsity-driven scenarios with the purpose of avoiding biased estimates which usually occur when convex penalties are used. Though the convergence of the DRSM has been well studied for the case where both functions are convex, its results for some nonconvex-function-involved cases, including the "strongly + weakly" convex case, are still in their infancy. In this paper, we prove the convergence of the DRSM for the "strongly + weakly" convex setting under relatively mild assumptions compared with some existing work in the literature. Moreover, we establish the rate of asymptotic regularity and the local linear convergence rate in the asymptotical sense under some regularity conditions.

Key words. Douglas–Rachford splitting method, weakly convex penalty, Fejér monotone, convergence, convergence rate, rate of asymptotic regularity

AMS subject classifications. 47H05, 49J52, 65K10, 90C25

DOI. 10.1137/16M1078604

1. Introduction.

We consider the minimization problem

$$\min_{x\in\mathcal{R}^n} f(x)+g(x), \tag{1.1}$$

where $f:\mathcal{R}^n\to\mathcal{R}\cup\{+\infty\}$ and $g:\mathcal{R}^n\to\mathcal{R}\cup\{+\infty\}$ are proper and lower semicontinuous, f is strongly convex with constant $\beta>0$ (see Definition 2.6), and g is weakly convex with constant $\omega>0$ (see Definition 2.7). We further assume that $\beta>\omega$ (to be elucidated), and under this assumption the solution set of (1.1) is a singleton. A particular application fitting this model is where $f(x)$ represents a data-fidelity term (e.g., a quadratic function) and $g(x)$ is a sparsity-driven penalty term whose weak convexity can often reduce bias in nonzero estimates (which is a serious problem for various convex penalty terms); see, e.g., [20, 31, 71] for some applications in sparse signal recovery applications. In particular, as proved recently in [49], weakly convex penalties include the well-known smoothly clipped absolute deviation (SCAD) penalty in [44], the minimax concave plus (MCP) penalty in [73], and the smoothed surrogate (i.e., $\sum_{i=1}^n(|x_i|+\epsilon)^p$ with $\epsilon>0$) of the widely considered ℓ_p regularization (i.e., $\|x\|_p^p:=\sum_{i=1}^n|x_i|^p$ with $0<p<1$) in the literature (see, e.g., [32]). One more example is the joint denoising and sharpening image recovery problem in [62] in which the

*Received by the editors June 6, 2016; accepted for publication (in revised form) January 30, 2017; published electronically July 6, 2017.
http://www.siam.org/journals/sinum/55-4/M107860.html
Funding: The first author was supported by a project funded by PAPD of Jiangsu Higher Education Institutions and the Natural Science Foundation of China (grant 11371015). The second author was supported by a project funded by PAPD of Jiangsu Higher Education Institutions and the Natural Science Foundation of China (grants 11625105, 11371197, and 11431002). The third author was supported by the General Research Fund from Hong Kong Research Grants Council: 12313516.
[†]School of Mathematical Sciences and Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing, Jiangsu 210023, China (keguo2014@126.com, handeren@njnu.edu.cn).
[‡]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China (xmyuan@hkbu.edu.hk).

weakly convex function is used for the design of energy functions that describe some desired effects more accurately than purely convex ones. In addition, as well analyzed in the literature (see, e.g., [1, 19, 20, 21, 30]), an interesting theoretical property is that the proximal operators of weakly convex functions are general enough to produce all separable monotone threshold functions. This property indeed plays a crucial role in constructing a penalty function associated with a given monotone function; see, e.g., [20].

We focus on the case where both f and g have their own properties/features; instead of considering the sum of f and g as a whole, we treat them separately in our algorithmic design so that a splitting algorithm capable of taking full advantage of f and g's properties/features can be effectively developed. This philosophy urges the particular consideration of the weakly convex term $g(x)$ in (1.1) because, as we shall show soon, the nonexpansiveness of some operator involved in the convergence analysis no longer holds, and thus new techniques should be developed for analyzing the convergence of splitting algorithms for the "strongly + weakly" convex setting of (1.1). The Moreau envelope function and proximal operator turn out to be fundamental for introducing algorithms for (1.1) and conducting convergence analysis. Thus, let us first recall their definitions. For a proper lower semicontinuous function $h : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ and a parameter $\nu > 0$, the Moreau envelope function $e_\nu h$ and proximal operator $\mathrm{prox}_{\nu h}$ are defined, respectively, as

$$e_\nu h(x) := \inf_{y \in \mathcal{R}^n} \left\{ h(y) + \frac{1}{2\nu}\|y - x\|^2 \right\} \tag{1.2}$$

and

$$\mathrm{prox}_{\nu h}(x) := \arg\min_{y \in \mathcal{R}^n} \left\{ h(y) + \frac{1}{2\nu}\|y - x\|^2 \right\}. \tag{1.3}$$

If $\inf_{y\in\mathcal{R}^n} h(y) > -\infty$, then for every $\nu \in (0, +\infty)$, the set $\mathrm{prox}_{\nu h}(x)$ is nonempty and compact, and $\mathrm{prox}_{\nu h}(x)$ is single-valued if h is further assumed to be convex; see, e.g., [69, Theorems 1.25 and 2.26]. More precisely, our philosophy of algorithmic design is based on the prerequisite that it is the individual proximal operators of f and g, rather than that of $f + g$, that should be considered.

A natural idea is to consider the Douglas–Rachford splitting method (DRSM), which traces back to [40, 60] and has been well studied from various perspectives in the literature; see, e.g., [13, 14, 41, 45, 50]. Extending the DRSM to the "strongly + weakly" convex problem (1.1) yields the iterative scheme

$$z_{k+1} = \widetilde{T}_{DR}(z_k) := ((1 - \alpha)I + \alpha R_{\lambda f} R_{\lambda g})(z_k), \tag{1.4}$$

where $\alpha \in (0, 1)$ is a parameter, I is the identity operator, the proximal parameter $\lambda > 0$ should be judiciously chosen (to be discussed), and $R_{\lambda f} := 2\mathrm{prox}_{\lambda f} - I$ and $R_{\lambda g} := 2\mathrm{prox}_{\lambda g} - I$ are the reflection operators of f and g, respectively. Here we call $\widetilde{T}_{DR}$ defined in (1.4) a Douglas–Rachford operator. Note that the original DRSM in [40, 60] is a special case of (1.4) with $\alpha = \frac{1}{2}$, but its extension to a general $\alpha \in (0, 1)$ has also been studied in, e.g., [33, 48, 50]. For the case where both f and g are convex, the convergence of the DRSM has been extensively studied; see, e.g., [40, 41, 60]. Indeed, its convergence is an immediate conclusion of the convergence result of the well-known Krasnoselskiĭ–Mann theorem [54] if we regard the scheme (1.4) as a convex combination of the nonexpansive operator $R_{\lambda f} R_{\lambda g}$ and the identity operator. We refer the reader to, e.g., [6, 35, 42, 43, 47, 52, 64, 67] for more general and deeper convergence studies on the Krasnoselskiĭ–Mann iteration in the nonexpansive operator context. In particular, the recent work [58] provides a comprehensive convergence rate study of inexact versions of the Krasnoselskiĭ–Mann iterations for nonexpansive operators.

Compared with the rich literature for the case where both f and g are convex, research on the convergence of the DRSM for the "strongly + weakly" convex problem (1.1), however, is still in its infancy. The main difficulty is that even if f is strongly convex, the operator $R_{\lambda f}R_{\lambda g}$ is not necessarily nonexpansive (see Remark 4.2) because $R_{\lambda g}$ is expansive when g is weakly convex [21, Lemma 3]. Thus, some existing results depending on the nonexpansiveness of an operator (including the mentioned works based on the Krasnoselskiĭ–Mann theorem) are not applicable; hence the "strongly + weakly" convex case of (1.1) is more challenging than the "convex + convex" case when discussing the convergence of the DRSM scheme (1.4).

The first effort seems to be the work [21], in which the convergence of (1.4) was established under the additional conditions that f is second-order differentiable, its gradient ∇f is Lipschitz continuous with constant $\sigma > 0$, $\beta = \omega$, $0 < \lambda \leq \frac{1}{\sqrt{\sigma\beta}}$, and $\alpha \in (0, 1)$. Essentially, these additional assumptions ensure the contraction property of the operator $R_{\lambda f}$ and the nonexpansiveness of the operator $R_{\lambda f}R_{\lambda g}$, because the analysis in [21] still follows the framework of the classical Krasnoselskiĭ–Mann theorem to prove the convergence of (1.4).

Our purpose in this paper is to study the convergence of the DRSM scheme (1.4) for the "strongly + weakly" convex problem (1.1) without any differentiability assumption on the strongly convex function f. In contrast to the technique in [21] based on the nonexpansiveness of the operators $R_{\lambda f}R_{\lambda g}$, our analysis is based on the Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ generated by (1.4) with respect to the fixed point set of $\widetilde{T}_{DR}$ defined in (1.4), and it does not require the nonexpansiveness of the operator $R_{\lambda f}R_{\lambda g}$. Here, for an operator $M : \mathcal{R}^n \to \mathcal{R}^n$, the fixed point set of M is defined as $\mathrm{Fix}(M) := \{z \in \mathcal{R}^n : M(z) = z\}$. Thus, differentiability assumptions on f in [21] can be removed in our analysis. Meanwhile, without any differentiability assumption on f, we alternatively require the condition $\beta > \omega$, which is slightly stronger than the condition $\beta = \omega$ in [21], to establish the Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ (see Theorem 4.3). This alternative assumption seems to be satisfied by some applications, e.g., some signal and image processing applications delineated in [21, 53, 62]. For the sake of succinctness, let us elaborate only on the joint image denoising and sharpening problem in [62]. The model of this problem can be written as

$$\min_{x\in\mathcal{R}^n} \quad \frac{c}{2}\|x-b\|^2 + \iota_{[0,1]}(x) + \|\nabla x\|_{2,1} - \frac{\nu}{2}\|\nabla x\|_{2,2}^2, \tag{1.5}$$

where $x \in \mathcal{R}^n$ is the vector representation of a digital image; $b \in \mathcal{R}^n$ is an observed image; $\nabla := (\nabla_1, \nabla_2) : \mathcal{R}^n \to \mathcal{R}^n \times \mathcal{R}^n$ is a discrete gradient operator with $\nabla_1 : \mathcal{R}^n \to \mathcal{R}^n$ and $\nabla_2 : \mathcal{R}^n \to \mathcal{R}^n$ the standard finite difference with periodic boundary conditions in the horizontal and vertical directions, respectively; $\|\nabla x\|_{2,1}$ is the total-variational regularization term (see [70]) to preserve sharp edges; $-\frac{\nu}{2}\|\nabla x\|_{2,2}^2$ is a sharpening/edge enhancement term aiming at removing a blur if the blur is assumed to follow a diffusion process; and

$$\iota_{[0,1]}(x) := \begin{cases} 0, & 0 \leq x \leq 1, \\ \infty & \text{else.} \end{cases}$$

The definitions of $\|\cdot\|_{2,1} : \mathcal{R}^n \times \mathcal{R}^n \to \mathcal{R}$ and $\|\cdot\|_{2,2} : \mathcal{R}^n \times \mathcal{R}^n \to \mathcal{R}$ are given, respectively, by

$$\|y\|_{2,1} := \sum_{i,j=1}^{n} \sqrt{|y^1_{i,j}|^2 + |y^2_{i,j}|^2} \quad \forall y = (y^1, y^2) \in \mathcal{R}^n \times \mathcal{R}^n$$

and

$$\|y\|_{2,2} := \sqrt{\sum_{i,j=1}^{n} |y^1_{i,j}|^2 + |y^2_{i,j}|^2} \quad \forall y = (y^1, y^2) \in \mathcal{R}^n \times \mathcal{R}^n.$$

We refer the reader to [63] for more details and applications of the model (1.5) and its useful variants in image processing. Clearly, the specific application (1.5) corresponds to the special case of the generic model (1.1) with $f(x) := \frac{c}{2}\|x - b\|^2 + \iota_{[0,1]}(x)$ and $g(x) := \|\nabla x\|_{2,1} - \frac{\nu}{2}\|\nabla x\|^2_{2,2}$. Obviously, f is not differentiable, while it is strongly convex with constant c, and g is a weakly convex function with constant $\nu\|\nabla\|^2$. As mentioned in [62], $\nu = 0.7\frac{c}{\|\nabla\|^2}$ is set to better sharpen the edges and lead to an increased contrast. Then the assumption $c > \nu\|\nabla\|^2$, i.e., $\beta > \omega$ for model (1.1), is satisfied. Technically, the condition $\beta > \omega$ suffices to ensure the Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ (see (4.12)) so that the convergence of (1.4) can be established; it is also useful to determine the range of the proximal parameter λ (see (3.4)).

Furthermore, since g is weakly convex, there is a fundamental difference between $R_{\lambda f}$ and $R_{\lambda g}$: $R_{\lambda f}$ is nonexpansive, while $R_{\lambda g}$ is expansive. Thus, the following scheme that swaps the composition of $R_{\lambda f}$ and $R_{\lambda g}$ in (1.4) is also considered in [21]:

$$z_{k+1} = \widehat{T}_{DR}(z_k) := ((1-\alpha)I + \alpha R_{\lambda g} R_{\lambda f})(z_k). \tag{1.6}$$

The operator $\widehat{T}_{DR}$ given in (1.6) is also called a Douglas–Rachford operator. Although we conduct a detailed convergence analysis mainly for (1.4), the convergence of (1.6) will also be briefly mentioned for completeness. Indeed, based on the Fejér monotonicity technique, the schemes (1.4) and (1.6) share significant similarities in their convergence analysis. We also refer the reader to, e.g., [4] for an intensive study of the difference when the operators in the forward-backward splitting method are swapped.

There is a huge literature on the DRSM schemes. Here we briefly review some other works that are relevant to this paper in some sense, while we do not go into detail for succinctness. First, in [21] a modified DRSM scheme that does not require the differentiability of f is considered, while it is shown that its numerical performance is not comparable with that of (1.4) or (1.6). Thus, we do not discuss that modified scheme. In [15], a more elaborated analysis of the DRSM with different orders of the operators is discussed in the setting of two maximal monotone operators. We also refer the reader to [10] for a comprehensive study of the application of the DRSM to the convex feasibility problem which corresponds to the model (1.1) with indicator functions of some closed convex sets in its objective. The DRSM has also been studied in some nonconvex optimization settings, which usually require some regularity assumptions (e.g., (ϵ, δ)-regularity, super-regularity, affine-hull regularity, strong regularity of sets) to study its convergence; see, e.g., [11, 12, 15, 16, 17, 51, 66]. Particularly, in [56] the DRSM is applied to a feasibility reformulation of the minimization of the sum of a proper lower semicontinuous function (not necessarily convex) and a weakly convex function whose gradient is required to be Lipschitz continuous, and both functions are required to satisfy the Kurdyka–Łojasiewicz inequality (see [3]). It is worthwhile to mention that the strong convexity implies the Kurdyka–Łojasiewicz inequality; see [3, 23]. Under the assumption $\beta > \omega$, the whole objective function in (1.1) is strongly convex and thus satisfies the Kurdyka–Łojasiewicz inequality. But our assumption on g is weaker than that in [56] because we do not make any differentiability assumption on g.

Another splitting method closely related to the DRSM is the Peaceman–Rachford splitting method (PRSM) dating back to [60, 65]. Its iterative scheme for model (1.1) reads as

$$z_{k+1} = T_{PR}(z_k) := R_{\lambda f} R_{\lambda g}(z_k),$$

which differs from the DRSM scheme in that α should take the value of 1 in (1.4). Despite its empirically faster convergence for some cases (see, e.g., [46]), the PRSM is inferior to the DRSM in the sense that it is not necessarily convergent even when both f and g are convex; see, e.g., [36] for a counterexample. Some interesting analysis for the convergence of PRSM under stronger conditions can be found in [57].

The rest of this paper is organized as follows. In section 2, we recall some definitions and known results for further analysis. Then we specify the restriction of the proximal parameter λ in our discussion in section 3. The convergence of the DRSM (1.4) is proved in section 4, and the rate of asymptotic regularity is derived in section 5 for the Douglas–Rachford operators defined in (1.4) and (1.6), respectively. Then, the local linear convergence rate of the DRSM (1.4) is established in section 6 under some regularity conditions. Finally, we draw some conclusions in section 7.

2. Preliminaries.

In this section, we recall some definitions and known results that will be used in our analysis. Throughout, $\mathrm{ri}(\mathrm{dom}\ f) \cap \mathrm{ri}(\mathrm{dom}\ g) \neq \emptyset$[1] is assumed, where "ri" is the relative interior of a set and "dom" is the effective domain of a function (see Definition 2.5). As shown in [68], this property guarantees the sum rule for subdifferential operations.

DEFINITION 2.1 ([9, Definition 5.1]). *Let C be a nonempty subset of $\mathcal{R}^n$, and let $\{z_k\}_{k\in N}$ be a sequence in $\mathcal{R}^n$. Then $\{z_k\}_{k\in N}$ is Fejér monotone with respect to C if*

$$\|z_{k+1} - z\| \leq \|z_k - z\| \quad \forall z \in C, \quad \forall k \in N.$$

DEFINITION 2.2 ([9, Definition 4.1]). *Let D be a nonempty subset of $\mathcal{R}^n$, and let $M : D \to \mathcal{R}^n$. Then M is nonexpansive if it is Lipschitz continuous with constant 1, i.e.,*

$$\|M(x) - M(y)\| \leq \|x - y\| \quad \forall x, y \in D.$$

LEMMA 2.1 ([54] (Krasnoselskiĭ–Mann theorem)). *Let D be a nonempty closed convex subset of $\mathcal{R}^n$, and let $M : D \to D$ be a nonexpansive operator such that $\mathrm{Fix}(M) \neq \emptyset$. Let $z_0 \in D$, and set $z_{k+1} := (1-\alpha)z_k + \alpha M(z_k)$ for $0 < \alpha < 1$. Then, $\{z_k\}_{k\in N}$ converges to a point in $\mathrm{Fix}(M)$.*

[1] In fact, this assumption is strongly related to the metric subregularity of the associated fixed point operators. For more details see, e.g., [18]. Also, in the context of feasibility problems where both f and g are indicator functions, this condition implies a type of transversality/regularity of the intersection widely used in the analysis for projection-type algorithms; see, e.g., [51, 66].

LEMMA 2.2 ([9, Proposition 4.2]). *Let D be a nonempty subset of $\mathcal{R}^n$, and let $M : D \to \mathcal{R}^n$. Then the following results are equivalent:*

(i) *M is firmly nonexpansive, i.e.,*

$$\langle M(x) - M(y), x - y\rangle \geq \|M(x) - M(y)\|^2 \quad \forall x, y \in D.$$

(ii) *$2M - I$ is nonexpansive.*

DEFINITION 2.3 ([9, Definition 4.23]). *Let D be a nonempty subset of $\mathcal{R}^n$, let $M : D \to \mathcal{R}^n$ be nonexpansive, and let $\alpha \in (0, 1)$. Then M is α-averaged if there exists a nonexpansive operator $Q : D \to \mathcal{R}^n$ such that $M = (1 - \alpha)I + \alpha D$.*

Remark 2.1. Based on Lemma 2.2 and Definition 2.3, we know M is $\frac{1}{2}$-averaged if and only if M is firmly nonexpansive.

DEFINITION 2.4 ([69, Definition 12.53]). *A mapping $T : \mathcal{R}^n \rightrightarrows \mathcal{R}^n$ is strongly monotone if there exists $\beta > 0$ such that*

$$\langle v_1 - v_0, x_1 - x_0\rangle \geq \beta\|x_1 - x_0\|^2 \quad \forall v_0 \in T(x_0),\ v_1 \in T(x_1).$$

DEFINITION 2.5 ([68, section 4]). *Given a function $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$, the effective domain and the epigraph of f are defined by*

$$\text{dom } f := \{x \in \mathcal{R}^n \mid f(x) < +\infty\}, \quad \text{epi } f := \{(x, \alpha) \in \mathcal{R}^n \times \mathcal{R} : f(x) \leq \alpha\}.$$

We say that the function f is proper (respectively, lower semicontinuous) if dom f (respectively, epi f) is nonempty (respectively, closed).

DEFINITION 2.6 ([69, Definition 12.58]). *A function $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ is strongly convex with constant $\beta > 0$ if for any $x, y \in \mathcal{R}^n$ and for any $\theta \in (0, 1)$, we have*

$$f((1 - \theta)x + \theta y) \leq (1 - \theta)f(x) + \theta f(y) - \frac{\beta\theta(1 - \theta)}{2}\|x - y\|^2.$$

Moreover, if the above inequality holds for $\beta = 0$, then we call f a convex function.

LEMMA 2.3 ([9, Proposition 12.27]). *Let $h : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function, and let $\nu > 0$; then the proximal operator $\text{prox}_{\nu h}$ given in* (1.3) *is firmly nonexpansive.*

DEFINITION 2.7 ([22, Definition 10]). *A proper lower semicontinuous function $g : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ is called weakly convex (or semiconvex) if for some $\omega > 0$, the function*

$$x \longmapsto g(x) + \frac{\omega}{2}\|x\|^2$$

is convex.

Remark 2.2. It is well known that the set of semiconvex functions contains several important classes of (nonsmooth) functions as special cases, for example, φ-convex functions [37] and primal-lower-nice functions [61]. Moreover, any twice continuously differentiable function with a bounded second-order derivative is semiconvex; see, e.g., [22]. We refer the reader to, e.g., [22, 28] for more properties of semiconvex functions. In [72], the semiconvexity is also called hypoconvexity (see [72, Definition 3.10]), and the proximal operator of a hypoconvex function is well studied therein.

Definition 2.8 ([69, Definition 8.3]). *Consider a function* $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ *and a point* $\bar{x}$ *with* $f(\bar{x})$ *finite.*

(i) *The regular subdifferential of* f *at* $\bar{x}$*, written* $\hat{\partial} f(\bar{x})$*, is the set of vectors* $x^* \in \mathcal{R}^n$ *that satisfy*

$$\lim_{y \neq \bar{x}} \inf_{y \to \bar{x}} \frac{f(y) - f(\bar{x}) - \langle x^*, y - \bar{x}\rangle}{\|y - \bar{x}\|} \geq 0.$$

(ii) *The subdifferential of* f *at* $\bar{x}$*, written* $\partial f(\bar{x})$*, is defined as follows:*

$$\partial f(\bar{x}) := \{x^* \in \mathcal{R}^n : \exists x_k \to \bar{x}, f(x_k) \to f(\bar{x}), x_k^* \in \hat{\partial} f(x_k), \textit{with } x_k^* \to x^*\}.$$

Remark 2.3. It follows from Definition 2.8 that the following conclusions hold (see, e.g., [69]).

(i). If $h : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ is a proper function and $f : \mathcal{R}^n \to \mathcal{R}$ is continuously differentiable, then $\partial(f + h)(x) = \nabla f(x) + \partial h(x)$ for any $x \in \text{dom } h$.

(ii). For any proper convex function $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ and for any $\bar{x} \in \text{dom } f$, the subdifferential of f is defined as $\bar{\partial} f(\bar{x}) := \{v \in \mathcal{R}^n |\ f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle \text{ for all } x\}$. For a convex function f, we have $\partial f(\bar{x}) = \hat{\partial} f(\bar{x}) = \bar{\partial} f(\bar{x})$ for any $\bar{x} \in \text{dom } f$. Moreover, for any proper convex functions $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ and $g : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$, if $\text{ri}(\text{dom } f) \cap \text{ri}(\text{dom } g) \neq \emptyset$, then $\bar{\partial}(f + g)(x) = \bar{\partial} f(x) + \bar{\partial} g(x)$ for any x.

Lemma 2.4 ([69, Theorem 12.17]). *A proper lower semicontinuous function* $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ *is convex if and only if* $\bar{\partial} f$ *is monotone, in which case* $\bar{\partial} f$ *is maximal monotone.*

Lemma 2.5 ([9, Proposition 23.7] and [69, Theorem 12.12]). *Let* $T : \mathcal{R}^n \rightrightarrows \mathcal{R}^n$ *be a mapping, and let* $\lambda > 0$*. Then* T *is monotone if and only if* $(I + \lambda T)^{-1}$ *is firmly nonexpansive. Moreover,* T *is maximal monotone if and only if* $\text{dom } (I + \lambda T)^{-1} = \mathcal{R}^n$*. In that case,* $(I + \lambda T)^{-1}$ *is a single-valued mapping from* $\mathcal{R}^n$ *into itself.*

Definition 2.9 ([69, Definition 1.23]). *A function* $h : \mathcal{R}^n \to \mathcal{R}^n \cup \{+\infty\}$ *is prox-bounded if there exists* $\nu > 0$ *such that* $e_\nu h(x) > -\infty$ *for some* $x \in \mathcal{R}^n$*. The supremum of the set of all such* ν *is the threshold* ν_h *of prox-boundedness for* h*.*

Lemma 2.6 ([72, Proposition 3.13]). *Let* $g : \mathcal{R}^n \to \mathcal{R}^n \cup \{+\infty\}$ *be proper, lower semicontinuous, and prox-bounded with threshold* $\nu_g \in (0, +\infty]$*. Then the following assertions are equivalent:*

(i) *For all* $\nu_g > \nu > 0$*,* $\text{prox}_{\nu g}$ *is single-valued on* $\mathcal{R}^n$*.*

(ii) *For all* $\nu_g > \nu > 0$*,* $\text{prox}_{\nu g}$ *is Lipschitz on* $\mathcal{R}^n$*.*

(iii) $g + \frac{1}{2\nu_g}\|\cdot\|^2$ *is convex. (Here, for* $\nu_g = +\infty$*,* $g + \frac{1}{2\nu_g}\|\cdot\|^2$ *reduces to* g*.)*

To close this section, we list the following elementary identity that will be frequently used in our later analysis: for all vector $a, b \in \mathcal{R}^n$ and $\alpha \in \mathcal{R}$, we have

$$\|(1 - \alpha)a + \alpha b\|^2 = (1 - \alpha)\|a\|^2 + \alpha\|b\|^2 - \alpha(1 - \alpha)\|a - b\|^2. \tag{2.1}$$

3. Restriction on $\boldsymbol{\lambda}$.

As mentioned, the proximal parameter λ should be chosen judiciously in order to discuss the convergence of the DRSM scheme (1.4). In this section, we delineate our restriction on this parameter for further discussion; the reason for restricting $\beta > \omega$ for model (1.1) will also be made clear.

Recall that g is assumed to be weakly convex with constant $\omega > 0$. We first show that g is prox-bounded with threshold $\frac{1}{\omega}$. Indeed, since

$$\tilde{g}(x) := g(x) + \frac{\omega}{2}\|x\|^2 \tag{3.1}$$

is convex, if $1-\lambda\omega \neq 0$, then for any $x \in \mathcal{R}^n$, we have

$$\begin{aligned} e_\lambda g(x) &= \inf_{y\in\mathcal{R}^n} \left\{ g(y) + \frac{1}{2\lambda}\|y-x\|^2 \right\} \\ &= \inf_{y\in\mathcal{R}^n} \left\{ \tilde{g}(y) + \frac{1}{2\lambda}\|y-x\|^2 - \frac{\omega}{2}\|y\|^2 \right\} \\ &= \inf_{y\in\mathcal{R}^n} \left\{ \tilde{g}(y) + \frac{1-\lambda\omega}{2\lambda}\|y - \frac{1}{1-\lambda\omega}x\|^2 \right\} - \frac{1}{2\lambda(1-\lambda\omega)}\|x\|^2. \end{aligned} \tag{3.2}$$

It is easy to see that for any $x \in \mathcal{R}^n$, $e_\lambda g(x) > -\infty$ if and only if $\frac{1-\lambda\omega}{2\lambda} > 0$, i.e., $0 < \lambda < \frac{1}{\omega}$. Thus, it follows from Definition 2.9 that g is prox-bounded with threshold $\frac{1}{\omega}$.

Then, according to Lemma 2.6, for $0 < \lambda < \frac{1}{\omega}$, $\text{prox}_{\lambda g}$ is Lipschitz continuous and single-valued on $\mathcal{R}^n$. Since

$$\text{prox}_{\lambda g}(x) = \arg\min_{y\in\mathcal{R}^n} \left\{ g(y) + \frac{1}{2\lambda}\|y-x\|^2 \right\}, \tag{3.3}$$

its optimality condition is

$$x \in (I + \lambda\partial g)(\text{prox}_{\lambda g}(x)).$$

This means $\text{prox}_{\lambda g}(x) \in (I+\lambda\partial g)^{-1}(x)$. Next, we show that $(I+\lambda\partial g)^{-1}(x)$ is single-valued everywhere, and thus we indeed have $\text{prox}_{\lambda g}(x) = (I+\lambda\partial g)^{-1}(x)$. To see this, note that if $1-\lambda\omega \neq 0$, then we have

$$(I+\lambda\partial g)^{-1} = ((1-\lambda\omega)I + \lambda\bar{\partial}\tilde{g})^{-1} \left(I + \frac{\lambda}{1-\lambda\omega}\bar{\partial}\tilde{g} \right)^{-1} \left(\frac{1}{1-\lambda\omega} \right) I.$$

Again, since $\tilde{g}$ is proper lower semicontinuous convex, it follows from Lemma 2.4 that $\bar{\partial}\tilde{g}$ is maximal monotone. For $0 < \lambda < \frac{1}{\omega}$, it follows from Lemma 2.5 that $(I + \frac{\lambda}{1-\lambda\omega}\bar{\partial}\tilde{g})^{-1}$ is single-valued everywhere; so is $(I+\lambda\partial g)^{-1}$. Therefore, we have

$$\text{prox}_{\lambda g}(x) = (I+\lambda\partial g)^{-1}(x).$$

Overall, we need to choose $0 < \lambda < \frac{1}{\omega}$ so as to ensure that $\text{prox}_{\lambda g}$ is Lipschitz continuous and single-valued.

Obviously, the condition

$$0 < \lambda < \frac{(1-\alpha)(\beta-\omega)}{\beta\omega} \tag{3.4}$$

suffices to ensure $\frac{\lambda}{1-\lambda\omega} > 0$. We thus assume (3.4) throughout our discussion. Technically, this restriction ensures the Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ (as shown in Theorems 4.3 and 4.5) and the monotonicity of both sequences $\{\|\widetilde{e}(z_k,\lambda)\|\}_{k\in N}$ (see (5.13)) and $\{\|\widehat{e}(z_k,\lambda)\|\}_{k\in N}$ (see (5.30)). With these monotonicity results, the convergence of the DRSM schemes (1.4) and (1.6) and the rate of asymptotic regularity for $\widetilde{T}_{DR}$ and $\widehat{T}_{DR}$ can be established accordingly.

Remark 3.1. When $\beta = \omega = 0$, the model (1.1) reduces to the regular "convex + convex" setting, and our restriction (3.4) for λ coincides with the standard assumption $\lambda > 0$ (in the sense of $\frac{0}{0} := +\infty$) in the DRSM literature. Indeed, for some specific applications of the "convex + convex" setting of (1.1), the convergence of the DRSM has been well studied in the literature. We refer the reader to, e.g., the feasibility problem in [11], the basis pursuit problem in [38], and some more generalized problems in [59] for some insightful results about the local linear convergence rate of the DRSM that are independent of λ's value.

4. Convergence analysis. In this section, we prove the convergence of the DRSM schemes (1.4) and (1.6).

4.1. Some preparations. First, notice that the solution set of model (1.1) is a singleton under the assumption $\beta > \omega$. It follows from the assumption $\mathrm{ri}(\mathrm{dom}\ f) \cap \mathrm{ri}(\mathrm{dom}\ g) \neq \emptyset$, and the assertions (i) and (ii) in Remark 2.3, that x^* is the solution of (1.1) if and only if

$$0 \in \bar{\partial}(f+g)(x^*) = \bar{\partial} f(x^*) + \partial g(x^*). \tag{4.1}$$

PROPOSITION 4.1. *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be proper lower semicontinuous strongly convex with constant $\beta > 0$, and let $g : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be proper lower semicontinuous weakly convex with constant $\omega > 0$. Suppose $\beta > \omega$ and that λ satisfies* (3.4). *Then $\mathrm{Fix}(\widetilde{T}_{DR}) \neq \emptyset$, where $\widetilde{T}_{DR}$ is the Douglas–Rachford operator defined in* (1.4). *Moreover, $\mathrm{prox}_{\lambda g}(z^*)$ is the unique solution of* (1.1) *for $z^* \in \mathrm{Fix}(\widetilde{T}_{DR})$.*

Proof. It follows from (4.1) that

$$2x^* \in (I + \lambda\bar{\partial} f)(x^*) + (I + \lambda \partial g)(x^*).$$

Let $z^* \in (I + \lambda\partial g)(x^*)$, i.e., $x^* = (I + \lambda\partial g)^{-1}(z^*) = \mathrm{prox}_{\lambda g}(z^*)$; we have

$$2x^* - z^* \in (I + \lambda\bar{\partial} f)(x^*).$$

Since $R_{\lambda g}(z^*) = 2\mathrm{prox}_{\lambda g}(z^*) - z^* = 2x^* - z^*$ and $\mathrm{prox}_{\lambda f} = (I + \lambda\bar{\partial} f)^{-1}$ is single-valued, we have

$$x^* = \mathrm{prox}_{\lambda f} R_{\lambda g}(z^*).$$

Thus,

$$z^* = 2x^* - R_{\lambda g}(z^*) = 2\mathrm{prox}_{\lambda f} R_{\lambda g}(z^*) - R_{\lambda g}(z^*) = R_{\lambda f} R_{\lambda g}(z^*).$$

This means $\widetilde{T}_{DR}(z^*) = z^*$, i.e., $\mathrm{Fix}(\widetilde{T}_{DR}) \neq \emptyset$. Moreover, for any $z^* \in \mathrm{Fix}(\widetilde{T}_{DR})$, we have

$$z^* = R_{\lambda f} R_{\lambda g}(z^*) = 2\mathrm{prox}_{\lambda f} R_{\lambda g}(z^*) - R_{\lambda g}(z^*). \tag{4.2}$$

Setting $y^* := \mathrm{prox}_{\lambda f} R_{\lambda g}(z^*)$ in (4.2), we get

$$2y^* - z^* = R_{\lambda g}(z^*) = 2\mathrm{prox}_{\lambda g}(z^*) - z^*, \tag{4.3}$$

which means $y^* = \mathrm{prox}_{\lambda g}(z^*)$, i.e.,

$$z^* \in (I + \lambda\partial g)(y^*). \tag{4.4}$$

On the other hand, it follows from the definition of y^* and (4.3) that $y^* = \text{prox}_{\lambda f}(2y^* - z^*)$. This means

$$2y^* - z^* \in (I + \lambda\bar{\partial} f)(y^*). \tag{4.5}$$

Adding (4.4) and (4.5), and recalling $\text{ri}(\text{dom } f) \cap \text{ri}(\text{dom } g) \neq \emptyset$, we know that $y^* = \text{prox}_{\lambda g}(z^*)$ is a solution of problem (1.1). Recall that problem (1.1) has the unique solution because of $\beta > \omega$. The proof is complete. □

To prove the convergence of the DRSM scheme (1.4) or (1.6), we first need to extensively analyze the terms $\|R_{\lambda f}(x) - R_{\lambda f}(y)\|^2$ and $\|R_{\lambda g}(x) - R_{\lambda g}(y)\|^2$ and derive their bounds. The following theorem focuses on $\|R_{\lambda f}(x) - R_{\lambda f}(y)\|^2$, which is an important property for the convergence analysis.

THEOREM 4.1. *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be proper lower semicontinuous strongly convex with constant $\beta > 0$. Then for any $x, y \in \mathcal{R}^n$ and $\lambda > 0$, we have*

$$\|R_{\lambda f}(x) - R_{\lambda f}(y)\|^2 \leq \|x - y\|^2 - 4\beta\lambda\|\text{prox}_{\lambda f}(x) - \text{prox}_{\lambda f}(y)\|^2.$$

Proof. Recall the definition of proximal operator (1.3). We have

$$\text{prox}_{\lambda f}(x) = \arg\min_{y\in\mathcal{R}^n} \left\{ f(y) + \frac{1}{2\lambda}\|y - x\|^2 \right\},$$

and its optimality condition is given by

$$0 \in \bar{\partial} f(\text{prox}_{\lambda f}(x)) + \frac{1}{\lambda}(\text{prox}_{\lambda f}(x) - x),$$

which can be written as

$$\frac{1}{\lambda}(x - \text{prox}_{\lambda f}(x)) \in \bar{\partial} f(\text{prox}_{\lambda f}(x)). \tag{4.6}$$

Recall that $\bar{\partial} f$ is the subdifferential operator of f (see Remark 2.3(ii)). Since f is strongly convex with constant $\beta > 0$, it follows from [69, Exercise 12.59] that $\bar{\partial} f$ is strongly monotone with constant $\beta > 0$. Thus, for any $x, y \in \mathcal{R}^n$, it follows from (4.6) that

$$\left\langle \frac{1}{\lambda}(x - \text{prox}_{\lambda f}(x)) - \frac{1}{\lambda}(y - \text{prox}_{\lambda f}(y)), \text{prox}_{\lambda f}(x) - \text{prox}_{\lambda f}(y) \right\rangle \geq \beta\|\text{prox}_{\lambda f}(x) - \text{prox}_{\lambda f}(y)\|^2.$$

Rearranging terms yields

$$(1 + \beta\lambda)\|\text{prox}_{\lambda f}(x) - \text{prox}_{\lambda f}(y)\|^2 \leq \langle x - y, \text{prox}_{\lambda f}(x) - \text{prox}_{\lambda f}(y)\rangle. \tag{4.7}$$

Note that for any $x, y \in \mathcal{R}^n$, we have

$$\begin{aligned} &\|R_{\lambda f}(x) - R_{\lambda f}(y)\|^2 \\ &= \|x - y\|^2 - 4\langle \text{prox}_{\lambda f}(x) - \text{prox}_{\lambda f}(y), x - y\rangle + 4\|\text{prox}_{\lambda f}(x) - \text{prox}_{\lambda f}(y)\|^2. \end{aligned} \tag{4.8}$$

Substituting (4.7) into (4.8), we prove the assertion. □

Remark 4.1. According to Theorem 4.1, we know $R_{\lambda f}$ is Lipschitz continuous with constant 1. Moreover, if $\beta = 0$, i.e., f is a convex function, then (4.7) reduces to Lemma 2.3.

Next, we estimate $\|R_{\lambda g}(x) - R_{\lambda g}(y)\|^2$ for the weakly convex function g.

THEOREM 4.2. *Let $g : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous weakly convex function with constant $\omega > 0$. Then for any $x, y \in \mathcal{R}^n$ and $0 < \lambda < \frac{1}{\omega}$, we have*

$$\|R_{\lambda g}(x) - R_{\lambda g}(y)\|^2 \le \|x - y\|^2 + 4\omega\lambda \|\text{prox}_{\lambda g}(x) - \text{prox}_{\lambda g}(y)\|^2.$$

Proof. First, we know that $\tilde{g}$ defined in (3.1) is a convex function. Since $0 < \lambda < \frac{1}{\omega}$, by (3.1) we have

$$\begin{aligned} \text{prox}_{\lambda g}(x) &= \arg\min_{y \in \mathcal{R}^n} \left\{ g(y) + \frac{1}{2\lambda}\|y - x\|^2 \right\} \\ &= \arg\min_{y \in \mathcal{R}^n} \left\{ \tilde{g}(y) + \frac{1}{2\lambda}\|y - x\|^2 - \frac{\omega}{2}\|y\|^2 \right\} \\ &= \text{prox}_{\frac{\lambda}{1-\lambda\omega}\tilde{g}}\left(\frac{1}{1 - \lambda\omega} x \right). \end{aligned} \tag{4.9}$$

Then, it follows from (4.9) that

$$R_{\lambda g}(x) = 2\text{prox}_{\lambda g}(x) - x = 2\text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1 - \omega\lambda} x \right) - x.$$

Thus, for any $x, y \in \mathcal{R}^n$, we have

$$\begin{aligned} &\|R_{\lambda g}(x) - R_{\lambda g}(y)\|^2 \\ &= \left\| \left(2\text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} x \right) - x \right) - \left(2\text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} y \right) - y \right) \right\|^2 \\ &= 4\left\| \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} x \right) - \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} y \right) \|^2 + \|x - y \right\|^2 \\ &\quad -4(1-\omega\lambda)\left\langle \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} x \right) - \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} y \right), \frac{1}{1-\omega\lambda} x - \frac{1}{1-\omega\lambda} y \right\rangle \\ &\le 4\left\| \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} x \right) - \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} y \right) \|^2 + \|x - y \right\|^2 \\ &\quad -4(1-\omega\lambda)\left\| \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} x \right) - \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} y \right) \right\|^2 \\ &= \|x - y\|^2 + 4\omega\lambda \left\| \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} x \right) - \text{prox}_{\frac{\lambda}{1-\omega\lambda}\tilde{g}}\left(\frac{1}{1-\omega\lambda} y \right) \right\|^2 \\ &= \|x - y\|^2 + 4\omega\lambda \|\text{prox}_{\lambda g}(x) - \text{prox}_{\lambda g}(y)\|^2, \end{aligned}$$

where the inequality follows from Lemmas 2.2 and 2.3, and the last equality follows from (4.9). The proof is complete. ∎

Some immediate conclusions based on the results in Theorems 4.1 and 4.2 are summarized in the following corollary.

COROLLARY 4.1. *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be proper lower semicontinuous strongly convex with constant $\beta > 0$, and let $g : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous weakly convex function with constant $\omega > 0$. Then for any $x, y \in \mathcal{R}^n$*

and $0 < \lambda < \frac{1}{\omega}$, *we have*

$$\text{(4.10)}\quad \|R_{\lambda f}R_{\lambda g}(x) - R_{\lambda f}R_{\lambda g}(y)\|^2 \\ \le \|x-y\|^2 + 4\omega\lambda\|\text{prox}_{\lambda g}(x) - \text{prox}_{\lambda g}(y)\|^2 - 4\beta\lambda\|\text{prox}_{\lambda f}R_{\lambda g}(x) - \text{prox}_{\lambda f}R_{\lambda g}(y)\|^2$$

and

$$\text{(4.11)}\quad \|R_{\lambda g}R_{\lambda f}(x) - R_{\lambda g}R_{\lambda f}(y)\|^2 \\ \le \|x-y\|^2 - 4\beta\lambda\|\text{prox}_{\lambda f}(x) - \text{prox}_{\lambda f}(y)\|^2 + 4\omega\lambda\|\text{prox}_{\lambda g}R_{\lambda f}(x) - \text{prox}_{\lambda g}R_{\lambda f}(y)\|^2.$$

Proof. For any $x, y \in \mathcal{R}^n$ and $0 < \lambda < \frac{1}{\omega}$, it follows from the assertions in Theorems 4.1 and 4.2 that

$$\|R_{\lambda f}R_{\lambda g}(x) - R_{\lambda f}R_{\lambda g}(y)\|^2 \\ \le \|R_{\lambda g}(x) - R_{\lambda g}(y)\|^2 - 4\beta\lambda\|\text{prox}_{\lambda f}R_{\lambda g}(x) - \text{prox}_{\lambda f}R_{\lambda g}(y)\|^2 \\ \le \|x-y\|^2 + 4\omega\lambda\|\text{prox}_{\lambda g}(x) - \text{prox}_{\lambda g}(y)\|^2 - 4\beta\lambda\|\text{prox}_{\lambda f}R_{\lambda g}(x) - \text{prox}_{\lambda f}R_{\lambda g}(y)\|^2.$$

This proves the assertion (4.11). The proof of (4.12) is similar to that of (4.11) and thus is omitted. ∎

Remark 4.2. The inequalities (4.11) and (4.12) do not ensure whether the operator $R_{\lambda f}R_{\lambda g}$ or $R_{\lambda g}R_{\lambda f}$ is nonexpansive or not. Thus, the classical Krasnoselskiĭ–Mann Theorem (see Lemma 2.1) is not applicable to prove the convergence of the DRSMs (1.4) and (1.6). This further explains the difficulty of proving the convergence of the DRSM in the "strongly + weakly" convex setting (1.1) and urges more sophisticated techniques for this purpose.

4.2. Convergence of (1.4). The following theorem concerns the Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ generated by the DRSM (1.4).

THEOREM 4.3. *Let* $\{z_k\}_{k\in N}$ *be the sequence generated by the DRSM* (1.4) *with* $\beta > \omega$ *and* λ *satisfying* (3.4). *Then there exists* $\eta_1 > 0$ *and* $\eta_2 > 0$ *such that for any* $k \in N$, *it holds that*

$$\text{(4.12)}\quad \|z_{k+1}-z^*\|^2 \le \|z_k - z^*\|^2 - \eta_1\|z_{k+1}-z_k\|^2 - \eta_2\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2 \;\forall z^* \in \text{Fix}(\widetilde{T}_{DR}).$$

Proof. Since $z^* \in \text{Fix}(\widetilde{T}_{DR})$, we have $z^* = \widetilde{T}_{DR}(z^*)$ and also $z^* = R_{\lambda f}R_{\lambda g}(z^*)$. It thus follows from (1.4) that for any $k \in N$,

$$\begin{aligned} \|z_{k+1} - z^*\|^2 &= \|(1-\alpha)(z_k - z^*) + \alpha(R_{\lambda f}R_{\lambda g}(z_k) - R_{\lambda f}R_{\lambda g}(z^*))\|^2 \\ &= (1-\alpha)\|z_k - z^*\|^2 + \alpha\|R_{\lambda f}R_{\lambda g}(z_k) - R_{\lambda f}R_{\lambda g}(z^*)\|^2 - \alpha(1-\alpha)\|R_{\lambda f}R_{\lambda g}(z_k) - z_k\|^2 \\ &\le (1-\alpha)\|z_k - z^*\|^2 + \alpha\|z_k - z^*\|^2 - \alpha(1-\alpha)\|R_{\lambda f}R_{\lambda g}(z_k) - z_k\|^2 \\ &\quad + 4\alpha\omega\lambda\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2 - 4\alpha\beta\lambda\|\text{prox}_{\lambda f}R_{\lambda g}(z_k) - \text{prox}_{\lambda f}R_{\lambda g}(z^*)\|^2 \\ &= \|z_k - z^*\|^2 - \alpha(1-\alpha)\|R_{\lambda f}R_{\lambda g}(z_k) - z_k\|^2 + 4\alpha\omega\lambda\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2 \\ &\quad - 4\alpha\beta\lambda\|\text{prox}_{\lambda f}R_{\lambda g}(z_k) - \text{prox}_{\lambda f}R_{\lambda g}(z^*)\|^2, \end{aligned} \tag{4.13}$$

where the second equality is obtained by (2.1) applied to the case where $a := z_k - z^*$ and $b := R_{\lambda f}R_{\lambda g}(z_k) - R_{\lambda f}R_{\lambda g}(z^*)$, and the inequality follows from (4.11). Recall $z^* = R_{\lambda f}R_{\lambda g}(z^*)$. Let

$$y := R_{\lambda g}(z^*) = 2\text{prox}_{\lambda g}(z^*) - z^*.$$

Then, we have

$$z^* = R_{\lambda f}(y) = 2\text{prox}_{\lambda f}(y) - y,$$

and hence

$$\text{prox}_{\lambda g}(z^*) = \text{prox}_{\lambda f}(y) = \text{prox}_{\lambda f} R_{\lambda g}(z^*). \tag{4.14}$$

From (1.4) we know that

$$\begin{aligned} z_{k+1} &= (1-\alpha)z_k + \alpha R_{\lambda f}(2\text{prox}_{\lambda g}(z_k) - z_k) \\ &= (1-\alpha)z_k + \alpha(2\text{prox}_{\lambda f}(2\text{prox}_{\lambda g}(z_k) - z_k) - (2\text{prox}_{\lambda g}(z_k) - z_k)) \\ &= z_k + 2\alpha\text{prox}_{\lambda f} R_{\lambda g}(z_k) - 2\alpha\text{prox}_{\lambda g}(z_k); \end{aligned}$$

then it holds that

$$\text{prox}_{\lambda f} R_{\lambda g}(z_k) = \text{prox}_{\lambda g}(z_k) + \frac{1}{2\alpha}(z_{k+1} - z_k). \tag{4.15}$$

Moreover, (1.4) can be written as

$$R_{\lambda f} R_{\lambda g}(z_k) = \frac{1}{\alpha}(z_{k+1} - z_k) + z_k. \tag{4.16}$$

Substituting (4.14), (4.15), and (4.16) into (4.13), we have

$$\begin{aligned} \|z_{k+1} - z^*\|^2 \le\ & \|z_k - z^*\|^2 - \alpha(1-\alpha)\|\frac{1}{\alpha}(z_{k+1} - z_k)\|^2 + 4\alpha\omega\lambda\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2 \\ & -4\alpha\beta\lambda\|(\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)) + \frac{1}{2\alpha}(z_{k+1} - z_k)\|^2 \\ =\ & \|z_k - z^*\|^2 - \left(\frac{1-\alpha}{\alpha} + \frac{\beta\lambda}{\alpha}\right)\|z_{k+1} - z_k\|^2 - 4\alpha\lambda(\beta - \omega)\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2 \\ & -4\beta\lambda\langle \text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*), z_{k+1} - z_k\rangle. \end{aligned}$$

Using the condition (3.4), we obtain

$$0 < \frac{\beta}{2\alpha(\beta - \omega)} < \frac{1 - \alpha + \beta\lambda}{2\alpha\beta\lambda}.$$

Thus, for an arbitrarily fixed $\xi_1 \in (\frac{\beta}{2\alpha(\beta-\omega)}, \frac{1-\alpha+\beta\lambda}{2\alpha\beta\lambda})$, it follows from the Cauchy–Schwarz inequality that

$$\langle \text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*), z_{k+1} - z_k\rangle \ge -\frac{1}{2\xi_1}\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2 - \frac{\xi_1}{2}\|z_{k+1} - z_k\|^2. \tag{4.17}$$

Substituting (4.17) into (4.17), we get

$$\begin{aligned} \|z_{k+1} - z^*\|^2 \le \|z_k - z^*\|^2 &- \left(\frac{1-\alpha}{\alpha} + \frac{\beta\lambda}{\alpha} - 2\beta\lambda\xi_1\right)\|z_{k+1} - z_k\|^2 \\ &- \left(4\alpha\lambda(\beta - \omega) - \frac{2\beta\lambda}{\xi_1}\right)\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2. \end{aligned}$$

Therefore, the assertion (4.12) is proved with

$$\eta_1 := \frac{1-\alpha}{\alpha} + \frac{\beta\lambda}{\alpha} - 2\beta\lambda\xi_1 \quad \text{and} \quad \eta_2 := 4\alpha\lambda(\beta - \omega) - \frac{2\beta\lambda}{\xi_1}, \tag{4.18}$$

whose positiveness is guaranteed by the range of ξ_1. ∎

Remark 4.3. The Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ established in (4.12) indicates the difficulty of analyzing the convergence for the DRSM scheme (1.4) in the "strongly + weakly" convex setting of the model (1.1). Indeed, for the case where both f and g are convex, the Fejér monotonicity of the sequence generated by the DRSM scheme (1.4) can be easily established (see, e.g., [50, Lemma 2.3]), because for this case the corresponding operator $R_{\lambda f}R_{\lambda g}$ is nonexpansive. Moreover, for this case, as recently analyzed in [58, Lemma 6], the operator $R_{\lambda f}$ or $R_{\lambda g}$ can even be evaluated approximately, and accordingly the so-called quasi-Fejér monotonicity[2] can be established for the corresponding sequence. Differently, for the "strongly + weakly" convex setting of the model (1.1) under discussion, the operator $R_{\lambda f}R_{\lambda g}$ is not necessarily nonexpansive because of the weak convexity of g, and thus more sophisticated analysis is required for proving the Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ generated by even the exact version of the DRSM (1.4).

Now we prove the convergence of the DRSM (1.4). This is the main result of this paper.

THEOREM 4.4. *Let $\{z_k\}_{k\in N}$ be the sequence generated by the DRSM* (1.4) *with $\beta > \omega$ and λ satisfying* (3.4). *Then $\{z_k\}_{k\in N}$ converges to a fixed point of $\widetilde{T}_{DR}$. Moreover, $\{\text{prox}_{\lambda g}(z_k)\}_{k\in N}$ converges to the unique solution of model* (1.1).

Proof. First, it follows from Proposition 4.1 that $\text{Fix}(\widetilde{T}_{DR}) \neq \emptyset$. Let $z^* \in \text{Fix}(\widetilde{T}_{DR})$. Then it follows from Theorem 4.3 that the sequence $\{z_k\}_{k\in N}$ is Fejér monotone with respect to $\text{Fix}(\widetilde{T}_{DR})$. That is, we have

$$\|z_{k+1} - z^*\| \le \|z_k - z^*\| \quad \forall k \in N. \tag{4.19}$$

This means $\{\|z_k - z^*\|\}_{k\in N}$ is monotonically nonincreasing and $\{z_k\}_{k\in N}$ is bounded (indeed, $\{\|z_k - z^*\|\}_{k\in N}$ is convergent). Again, it follows from Theorem 4.3 that

$$\eta_1\|z_{k+1} - z_k\|^2 + \eta_2\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2 \le \|z_k - z^*\|^2 - \|z_{k+1} - z^*\|^2 \quad \forall k \in N.$$

Adding the above inequality from $k = 0$ to $k = m$, we have

$$\sum_{k=0}^{m}\{\eta_1\|z_{k+1}-z_k\|^2+\eta_2\|\text{prox}_{\lambda g}(z_k)-\text{prox}_{\lambda g}(z^*)\|^2\} \le \|z_0-z^*\|^2-\|z_{m+1}-z^*\|^2 \le \|z_0-z^*\|^2.$$

Thus, letting $m \to +\infty$, we get

$$\sum_{k=0}^{+\infty}\{\eta_1\|z_{k+1} - z_k\|^2 + \eta_2\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\|^2\} \le \|z_0 - z^*\|^2 < +\infty,$$

and hence

$$\lim_{k\to+\infty}\|z_{k+1} - z_k\| = 0 \quad \text{and} \quad \lim_{k\to+\infty}\|\text{prox}_{\lambda g}(z_k) - \text{prox}_{\lambda g}(z^*)\| = 0. \tag{4.20}$$

Moreover, the boundedness of $\{z_k\}_{k\in N}$ indicates that it has at least one cluster point. Let $\bar{z}$ be an arbitrary cluster point of $\{z_k\}_{k\in N}$, and let $\{z_{k_j}\}_{j\in N}$ be the subsequence

[2]According to [34, Definition 1.1], let C be a nonempty subset of $\mathcal{R}^n$, and let $\{z_k\}_{k\in N}$ be a sequence in $\mathcal{R}^n$. Then $\{z_k\}_{k\in N}$ is quasi-Fejér monotone with respect to C if there exists a nonnegative sequence $\{\epsilon_k\}_{k\in N}$ with $\sum_{k=0}^{\infty}\epsilon_k < +\infty$ such that $\|z_{k+1} - z\|^2 \le \|z_k - z\|^2 + \epsilon_k$ $\forall z \in C$, $\forall k \in N$.

converging to $\bar{z}$. By Theorem 4.2 and the Lipschitz continuity of $\text{prox}_{\lambda g}$, we further have $R_{\lambda g}(z_{k_j}) \to R_{\lambda g}(\bar{z})$. Thus, because of the Lipschitz continuity of $R_{\lambda f}$, we have $R_{\lambda f}R_{\lambda g}(z_{k_j}) \to R_{\lambda f}R_{\lambda g}(\bar{z})$. Since $\|z_{k+1} - z_k\| \to 0$, we know that $z_{k_j+1} \to \bar{z}$. Therefore, taking limits in (1.4) along the subsequence $\{z_{k_j}\}_{j\in N}$, we have $\bar{z} = R_{\lambda f}R_{\lambda g}(\bar{z})$, which implies $\bar{z} = \widetilde{T}_{DR}(\bar{z})$. This means that any cluster point of $\{z_k\}_{k\in N}$ is a fixed point of $\widetilde{T}_{DR}$. Thus, we can replace z^* by $\bar{z}$ in (4.19) and (4.20). Then we derive that the sequence $\{\|z_k - \bar{z}\|\}_{k\in N}$ is convergent. Since $z_{k_j} - \bar{z} \to 0$, the Fejér monotonicity of $\{z_k\}_{k\in N}$ implies that $z_k \to \bar{z}$. This means $\{z_k\}_{k\in N}$ converges to a fixed point of $\widetilde{T}_{DR}$. Furthermore, by Proposition 4.1 we know that $\text{prox}_{\lambda g}(\bar{z})$ is the unique solution of (1.1). Because of (4.20), we know $\text{prox}_{\lambda g}(z_k) \to \text{prox}_{\lambda g}(\bar{z})$. The proof is complete. ∎

4.3. Convergence of (1.6). Next, we briefly present the convergence for the DRSM (1.6). Its analysis follows the same roadmap as that for (1.4), with a main difference shown in the following proposition.

PROPOSITION 4.2. *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be proper lower semicontinuous strongly convex with constant $\beta > 0$, and let $g : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be proper lower semicontinuous weakly convex with constant $\omega > 0$. Suppose $\beta > \omega$ and that λ satisfies* (3.4). *Then $\text{Fix}(\widehat{T}_{DR}) \neq \emptyset$, where $\widehat{T}_{DR}$ is the Douglas–Rachford operator defined in* (1.6). *Moreover, $\text{prox}_{\lambda f}(z^*)$ is the unique solution of* (1.1) *for $z^* \in \text{Fix}(\widehat{T}_{DR})$.*

Proof. From (4.1), we know that x^* is the solution of (1.1) if and only if

$$2x^* \in (I + \lambda\bar{\partial}f)(x^*) + (I + \lambda\partial g)(x^*).$$

Letting $z^* \in (I + \lambda\bar{\partial}f)(x^*)$, i.e., $x^* = (I + \lambda\bar{\partial}f)^{-1}(z^*) = \text{prox}_{\lambda f}(z^*)$, we have

$$2x^* - z^* \in (I + \lambda\partial g)(x^*).$$

Moreover, as discussed in section 3, we know that $\text{prox}_{\lambda g}(x) = (I + \lambda\partial g)^{-1}(x)$ is single-valued. Since $R_{\lambda f}(z^*) = 2\text{prox}_{\lambda f}(z^*) - z^* = 2x^* - z^*$, we have

$$x^* = \text{prox}_{\lambda g}R_{\lambda f}(z^*).$$

Thus,

$$z^* = 2x^* - R_{\lambda f}(z^*) = 2\text{prox}_{\lambda g}R_{\lambda f}(z^*) - R_{\lambda f}(z^*) = R_{\lambda g}R_{\lambda f}(z^*).$$

This means that $\widehat{T}_{DR}(z^*) = z^*$, i.e., $\text{Fix}(\widehat{T}_{DR}) \neq \emptyset$. The rest of the proof is similar to that of the second part of Proposition 4.1 and hence is omitted. ∎

The following theorem concerns the Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ generated by the DRSM (1.6). The proof is analogous to that of Theorem 4.3.

THEOREM 4.5. *Let $\{z_k\}_{k\in N}$ be the sequence generated by the DRSM* (1.6) *with $\beta > \omega$ and λ satisfying* (3.4). *Then there exists $\eta_3 > 0$ and $\eta_4 > 0$ such that for any $k \in N$, it holds that*

$$\|z_{k+1}-z^*\|^2 \le \|z_k-z^*\|^2 - \eta_3\|z_{k+1}-z_k\|^2 - \eta_4\|\text{prox}_{\lambda f}(z_k) - \text{prox}_{\lambda f}(z^*)\|^2 \;\forall z^* \in \text{Fix}(\widehat{T}_{DR}). \tag{4.21}$$

Proof. Since $z^* \in \text{Fix}(\widehat{T}_{DR})$, we have $z^* = \widehat{T}_{DR}(z^*)$ and also $z^* = R_{\lambda g}R_{\lambda f}(z^*)$.

Then it follows from (1.6) that for any $k \in N$,

$$
\begin{aligned}
\|z_{k+1} - z^*\|^2 &= \|(1-\alpha)(z_k - z^*) + \alpha(R_{\lambda g}R_{\lambda f}(z_k) - R_{\lambda g}R_{\lambda f}(z^*))\|^2 \\
&= (1-\alpha)\|z_k - z^*\|^2 + \alpha\|R_{\lambda g}R_{\lambda f}(z_k) - R_{\lambda g}R_{\lambda f}(z^*)\|^2 - \alpha(1-\alpha)\|R_{\lambda g}R_{\lambda f}(z_k) - z_k\|^2 \\
&\le (1-\alpha)\|z_k - z^*\|^2 + \alpha\|z_k - z^*\|^2 - \alpha(1-\alpha)\|R_{\lambda g}R_{\lambda f}(z_k) - z_k\|^2 \\
&\quad + 4\alpha\omega\lambda\|\mathrm{prox}_{\lambda g}R_{\lambda f}(z_k) - \mathrm{prox}_{\lambda g}R_{\lambda f}(z^*)\|^2 - 4\alpha\beta\lambda\|\mathrm{prox}_{\lambda f}(z_k) - \mathrm{prox}_{\lambda f}(z^*)\|^2 \\
&= \|z_k - z^*\|^2 - \alpha(1-\alpha)\|R_{\lambda g}R_{\lambda f}(z_k) - z_k\|^2 - 4\alpha\beta\lambda\|\mathrm{prox}_{\lambda f}(z_k) - \mathrm{prox}_{\lambda f}(z^*)\|^2 \\
&\quad + 4\alpha\omega\lambda\|\mathrm{prox}_{\lambda g}R_{\lambda f}(z_k) - \mathrm{prox}_{\lambda g}R_{\lambda f}(z^*)\|^2, \qquad (4.22)
\end{aligned}
$$

where the second equality is obtained by (2.1) applied to the case where $a := z_k - z^*$ and $b := R_{\lambda g}R_{\lambda f}(z_k) - R_{\lambda g}R_{\lambda f}(z^*)$, and the inequality follows from (4.12). Recall that $z^* = R_{\lambda g}R_{\lambda f}(z^*)$. Let $y := R_{\lambda f}(z^*)$; then we have

$$z^* = R_{\lambda g}(y) = 2\mathrm{prox}_{\lambda g}(y) - y$$

and

$$y = 2\mathrm{prox}_{\lambda f}(z^*) - z^*.$$

Then we have

$$\mathrm{prox}_{\lambda f}(z^*) = \mathrm{prox}_{\lambda g}(y) = \mathrm{prox}_{\lambda g}R_{\lambda f}(z^*). \qquad (4.23)$$

From (1.6) we know that

$$
\begin{aligned}
z_{k+1} &= (1-\alpha)z_k + \alpha R_{\lambda g}(2\mathrm{prox}_{\lambda f}(z_k) - z_k) \\
&= (1-\alpha)z_k + \alpha(2\mathrm{prox}_{\lambda g}(2\mathrm{prox}_{\lambda f}(z_k) - z_k) - (2\mathrm{prox}_{\lambda f}(z_k) - z_k)) \\
&= z_k + 2\alpha\mathrm{prox}_{\lambda g}R_{\lambda f}(z_k) - 2\alpha\mathrm{prox}_{\lambda f}(z_k).
\end{aligned}
$$

Therefore, we have

$$\mathrm{prox}_{\lambda g}R_{\lambda f}(z_k) = \mathrm{prox}_{\lambda f}(z_k) + \frac{1}{2\alpha}(z_{k+1} - z_k). \qquad (4.24)$$

Moreover, (1.6) can be written as

$$R_{\lambda g}R_{\lambda f}(z_k) = \frac{1}{\alpha}(z_{k+1} - z_k) + z_k. \qquad (4.25)$$

Thus, substituting (4.23), (4.24), and (4.25) into (4.22), for any $k \in N$, we obtain

$$
\begin{aligned}
\|z_{k+1} - z^*\|^2 &\le \|z_k - z^*\|^2 - \alpha(1-\alpha)\|\frac{1}{\alpha}(z_{k+1} - z_k)\|^2 - 4\alpha\beta\lambda\|\mathrm{prox}_{\lambda f}(z_k) - \mathrm{prox}_{\lambda f}(z^*)\|^2 \\
&\quad + 4\alpha\omega\lambda\|\mathrm{prox}_{\lambda f}(z_k) - \mathrm{prox}_{\lambda f}(z^*) + \frac{1}{2\alpha}(z_{k+1} - z_k)\|^2 \\
&= \|z_k - z^*\|^2 - \left(\frac{1-\alpha}{\alpha} - \frac{\omega\lambda}{\alpha}\right)\|z_{k+1} - z_k\|^2 - 4\alpha\lambda(\beta-\omega)\|\mathrm{prox}_{\lambda f}(z_k) - \mathrm{prox}_{\lambda f}(z^*)\|^2 \\
&\quad + 4\omega\lambda\langle \mathrm{prox}_{\lambda f}(z_k) - \mathrm{prox}_{\lambda f}(z^*), z_{k+1} - z_k\rangle. \qquad (4.26)
\end{aligned}
$$

Because of the condition (3.4), we have

$$0 < \frac{\omega}{2\alpha(\beta-\omega)} < \frac{1-\alpha-\omega\lambda}{2\alpha\lambda\omega}.$$

Thus, for an arbitrarily fixed $\xi_2 \in (\frac{\omega}{2\alpha(\beta-\omega)}, \frac{1-\alpha-\omega\lambda}{2\alpha\lambda\omega})$, it follows from the Cauchy–Schwarz inequality that

$$\langle \mathrm{prox}_{\lambda f}(z_k)-\mathrm{prox}_{\lambda f}(z^*), z_{k+1}-z_k\rangle \leq \frac{1}{2\xi_2}\|\mathrm{prox}_{\lambda f}(z_k)-\mathrm{prox}_{\lambda f}(z^*)\|^2+\frac{\xi_2}{2}\|z_{k+1}-z_k\|^2. \tag{4.27}$$

Substituting (4.27) into (4.26), we get

$$\|z_{k+1}-z^*\|^2 \leq \|z_k - z^*\|^2 - \left(\frac{1-\alpha}{\alpha} - \frac{\omega\lambda}{\alpha} - 2\omega\lambda\xi_2\right)\|z_{k+1}-z_k\|^2 - \left(4\alpha\lambda(\beta-\omega) - \frac{2\omega\lambda}{\xi_2}\right)\|\mathrm{prox}_{\lambda f}(z_k) - \mathrm{prox}_{\lambda f}(z^*)\|^2.$$

Therefore, the assertion (4.21) is proved with

$$\eta_3 := \frac{1-\alpha}{\alpha} - \frac{\omega\lambda}{\alpha} - 2\omega\lambda\xi_2 \quad \text{and} \quad \eta_4 := 4\alpha\lambda(\beta-\omega) - \frac{2\omega\lambda}{\xi_2},$$

whose positiveness is guaranteed by the range of ξ_2. The proof is complete. ∎

Based on Theorem 4.5, we can prove the convergence of the DRSM (1.6). The proof is similar to that of Theorem 4.4 and thus is omitted.

THEOREM 4.6. *Let $\{z_k\}_{k\in N}$ be the sequence generated by the DRSM* (1.6) *with $\beta > \omega$ and λ satisfying* (3.4). *Then $\{z_k\}_{k\in N}$ converges to a fixed point of $\widehat{T}_{DR}$. Moreover, $\{\mathrm{prox}_{\lambda f}(z_k)\}_{k\in N}$ converges to the unique solution of model* (1.1).

5. Rates of asymptotic regularity.

In this section, we establish the rates of asymptotic regularity for the Douglas–Rachford operators $\widetilde{T}_{DR}$ and $\widehat{T}_{DR}$ defined in (1.4) and (1.6), respectively. Let C be a closed convex subset of $\mathcal{R}^n$, let $M : C \to C$ be an operator, and let the sequence $\{z_k\}_{k\in N}$ be generated by

$$z_{k+1} := (1-\alpha_k)z_k + \alpha_k M(z_k),$$

with $z_0 \in C$ and $\{\alpha_k\}_{k\in N} \subset [0, 1]$. According to [24], if $\|z_k - M(z_k)\| \to 0$, then M is said to be asymptotically regular. We refer the reader to, e.g., [5, 6, 7, 8, 24, 25, 26, 27, 29, 35, 58] for intensive studies on the rate of asymptotic regularity in various contexts of nonexpansive operators. Note that our analysis essentially differs from these works in that the operators $R_{\lambda f}R_{\lambda g}$ and $R_{\lambda g}R_{\lambda f}$ defined in (1.4) and (1.6) are not necessarily nonexpansive.

5.1. Rate of asymptotic regularity of $\widetilde{T}_{DR}$.

In this subsection, we establish the rate of asymptotic regularity for the Douglas–Rachford operator $\widetilde{T}_{DR}$ defined in (1.4). For any $z_0 \in \mathcal{R}^n$, it then follows from (1.4) that

$$z_k - \widetilde{T}_{DR}(z_k) = z_k - z_{k+1} = \alpha(z_k - R_{\lambda f}R_{\lambda g}(z_k)).$$

If we define

$$\widetilde{e}(z, \lambda) := z - R_{\lambda f}R_{\lambda g}(z), \tag{5.1}$$

then we have

$$z_k - \widetilde{T}_{DR}(z_k) = \alpha\widetilde{e}(z_k, \lambda). \tag{5.2}$$

Also, using the notation (5.1), the DRSM scheme (1.4) can be rewritten as

$$z_{k+1} = z_k - \alpha\widetilde{e}(z_k, \lambda) = (1-\alpha)z_k + \alpha R_{\lambda f} R_{\lambda g}(z_k), \quad \alpha \in (0,1). \tag{5.3}$$

Hence, it is reasonable to use $\|z_k - R_{\lambda f} R_{\lambda g}(z_k)\|$, i.e., $\|\widetilde{e}(z_k, \lambda)\|$, to measure the rate of asymptotic regularity of $\widetilde{T}_{DR}$. We first investigate the monotonicity of $\{\|\widetilde{e}(z_k, \lambda)\|\}_{k\in N}$ in the following lemma.

LEMMA 5.1. *Let $\{z_k\}_{k\in N}$ be the sequence generated by the DRSM* (1.4) *with $\beta > \omega$ and λ satisfying* (3.4)*. Then for any $k \in N$, we have*

$$\|\widetilde{e}(z_{k+1}, \lambda)\|^2 \le \|\widetilde{e}(z_k, \lambda)\|^2. \tag{5.4}$$

Proof. For any $k \in N$, taking $x := z_{k+1}$ and $y := z_k$ in (4.11), we have

$$\begin{aligned} &\|R_{\lambda f} R_{\lambda g}(z_{k+1}) - R_{\lambda f} R_{\lambda g}(z_k)\|^2 \\ &\le \|z_{k+1} - z_k\|^2 + 4\omega\lambda \|\mathrm{prox}_{\lambda g}(z_{k+1}) - \mathrm{prox}_{\lambda g}(z_k)\|^2 \\ &\quad - 4\beta\lambda \|\mathrm{prox}_{\lambda f} R_{\lambda g}(z_{k+1}) - \mathrm{prox}_{\lambda f} R_{\lambda g}(z_k)\|^2. \end{aligned} \tag{5.5}$$

Then, it follows from (5.1) that

$$\begin{aligned} \|\widetilde{e}(z_{k+1}, \lambda)\|^2 &= \|\widetilde{e}(z_{k+1}, \lambda) - \widetilde{e}(z_k, \lambda)\|^2 + 2\langle \widetilde{e}(z_k, \lambda), \widetilde{e}(z_{k+1}, \lambda) - \widetilde{e}(z_k, \lambda)\rangle + \|\widetilde{e}(z_k, \lambda)\|^2 \\ &= \|(z_{k+1} - R_{\lambda f} R_{\lambda g}(z_{k+1})) - (z_k - R_{\lambda f} R_{\lambda g}(z_k))\|^2 \\ &\quad + 2\langle \widetilde{e}(z_k, \lambda), \widetilde{e}(z_{k+1}, \lambda) - \widetilde{e}(z_k, \lambda)\rangle + \|\widetilde{e}(z_k, \lambda)\|^2 \\ &= \|z_{k+1} - z_k\|^2 + \|R_{\lambda f} R_{\lambda g}(z_{k+1}) - R_{\lambda f} R_{\lambda g}(z_k)\|^2 + \|\widetilde{e}(z_k, \lambda)\|^2 \\ &\quad + 2\langle \widetilde{e}(z_k, \lambda), \widetilde{e}(z_{k+1}, \lambda) - \widetilde{e}(z_k, \lambda)\rangle - 2\langle z_{k+1} - z_k, R_{\lambda f} R_{\lambda g}(z_{k+1}) - R_{\lambda f} R_{\lambda g}(z_k)\rangle \\ &\le 2\|z_{k+1} - z_k\|^2 + 4\omega\lambda \|\mathrm{prox}_{\lambda g}(z_{k+1}) - \mathrm{prox}_{\lambda g}(z_k)\|^2 + \|\widetilde{e}(z_k, \lambda)\|^2 \\ &\quad - 4\beta\lambda \|\mathrm{prox}_{\lambda f} R_{\lambda g}(z_{k+1}) - \mathrm{prox}_{\lambda f} R_{\lambda g}(z_k)\|^2 + 2\langle \widetilde{e}(z_k, \lambda), \widetilde{e}(z_{k+1}, \lambda) - \widetilde{e}(z_k, \lambda)\rangle \\ &\quad - 2\langle z_{k+1} - z_k, R_{\lambda f} R_{\lambda g}(z_{k+1}) - R_{\lambda f} R_{\lambda g}(z_k)\rangle, \end{aligned} \tag{5.6}$$

where the inequality follows from (5.5). Moreover, it follows from (4.15), (5.1), and (5.3) that

$$\mathrm{prox}_{\lambda f} R_{\lambda g}(z_k) = \mathrm{prox}_{\lambda g}(z_k) - \frac{1}{2}\widetilde{e}(z_k, \lambda), \quad R_{\lambda f} R_{\lambda g}(z_k) = z_k - \widetilde{e}(z_k, \lambda). \tag{5.7}$$

Thus, by means of (5.7) we have

$$\begin{aligned} &\|\mathrm{prox}_{\lambda f} R_{\lambda g}(z_{k+1}) - \mathrm{prox}_{\lambda f} R_{\lambda g}(z_k)\|^2 \\ &= \|\mathrm{prox}_{\lambda g}(z_{k+1}) - \mathrm{prox}_{\lambda g}(z_k)\|^2 + \left\|\frac{1}{2}\widetilde{e}(z_{k+1}, \lambda) - \frac{1}{2}\widetilde{e}(z_k, \lambda)\right\|^2 \\ &\quad - \langle \mathrm{prox}_{\lambda g}(z_{k+1}) - \mathrm{prox}_{\lambda g}(z_k), \widetilde{e}(z_{k+1}, \lambda) - \widetilde{e}(z_k, \lambda)\rangle \end{aligned} \tag{5.8}$$

and

$$\begin{aligned} &\langle z_{k+1} - z_k, R_{\lambda f} R_{\lambda g}(z_{k+1}) - R_{\lambda f} R_{\lambda g}(z_k)\rangle \\ &= \|z_{k+1} - z_k\|^2 - \langle z_{k+1} - z_k, \widetilde{e}(z_{k+1}, \lambda) - \widetilde{e}(z_k, \lambda)\rangle. \end{aligned} \tag{5.9}$$

Substituting (5.8) and (5.9) into (5.6), we get

$$\begin{aligned}\|\widetilde{e}(z_{k+1},\lambda)\|^2 \le\ & 2\|z_{k+1}-z_k\|^2+(4\omega\lambda-4\beta\lambda)\|\mathrm{prox}_{\lambda g}(z_{k+1})-\mathrm{prox}_{\lambda g}(z_k)\|^2+\|\widetilde{e}(z_k,\lambda)\|^2\\ & -\beta\lambda\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2+2\langle\widetilde{e}(z_k,\lambda),\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\rangle\\ & +4\beta\lambda\langle\mathrm{prox}_{\lambda g}(z_{k+1})-\mathrm{prox}_{\lambda g}(z_k),\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\rangle\\ & -2\|z_{k+1}-z_k\|^2+2\langle z_{k+1}-z_k,\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\rangle\\ =\ & (4\omega\lambda-4\beta\lambda)\|\mathrm{prox}_{\lambda g}(z_{k+1})-\mathrm{prox}_{\lambda g}(z_k)\|^2+\|\widetilde{e}(z_k,\lambda)\|^2\\ & -\beta\lambda\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2+(2-2\alpha)\langle\widetilde{e}(z_k,\lambda),\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\rangle\\ & +4\beta\lambda\langle\mathrm{prox}_{\lambda g}(z_{k+1})-\mathrm{prox}_{\lambda g}(z_k),\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\rangle, \end{aligned} \tag{5.10}$$

where the equality follows from the fact $z_{k+1}-z_k=-\alpha\widetilde{e}(z_k,\lambda)$. It follows from the Cauchy–Schwarz inequality that

$$\begin{aligned}&4\beta\lambda\langle\mathrm{prox}_{\lambda g}(z_{k+1})-\mathrm{prox}_{\lambda g}(z_k),\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\rangle\\ &\quad\le(4\beta\lambda-4\omega\lambda)\|\mathrm{prox}_{\lambda g}(z_{k+1})-\mathrm{prox}_{\lambda g}(z_k)\|^2+\frac{\beta^2\lambda}{\beta-\omega}\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2.\end{aligned} \tag{5.11}$$

Then, substituting (5.12) into (5.10), we obtain

$$\begin{aligned}\|\widetilde{e}(z_{k+1},\lambda)\|^2 \le\ & \|\widetilde{e}(z_k,\lambda)\|^2-\beta\lambda\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2+\frac{\beta^2\lambda}{\beta-\omega}\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2\\ & +(2-2\alpha)\langle\widetilde{e}(z_k,\lambda),\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\rangle\\ =\ & \|\widetilde{e}(z_k,\lambda)\|^2-\beta\lambda\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2+\frac{\beta^2\lambda}{\beta-\omega}\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2\\ & +(1-\alpha)(\|\widetilde{e}(z_{k+1},\lambda)\|^2-\|\widetilde{e}(z_k,\lambda)\|^2-\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2),\end{aligned}$$

where the equality follows from the fact that

$$\langle a,b-a\rangle=\frac{1}{2}(\|b\|^2-\|a\|^2-\|b-a\|^2) \tag{5.12}$$

for a and b with the same dimension. Hence, rearranging the terms, we know that

$$\alpha\|\widetilde{e}(z_{k+1},\lambda)\|^2\le\alpha\|\widetilde{e}(z_k,\lambda)\|^2+\left(\frac{\beta\lambda\omega}{\beta-\omega}+(\alpha-1)\right)\|\widetilde{e}(z_{k+1},\lambda)-\widetilde{e}(z_k,\lambda)\|^2. \tag{5.13}$$

Because of the condition (3.4), we know that $\frac{\beta\lambda\omega}{\beta-\omega}+\alpha-1<0$. The proof is complete. ∎

Remark 5.1. Lemma 5.1 shows another difficulty of analyzing the convergence for the DRSM scheme (1.6) in the "strongly + weakly" convex setting of the model (1.1). Indeed, if both f and g are convex in (1.1), the monotonicity of the corresponding sequence $\{\|\widetilde{e}(z_k,\lambda)\|^2\}_{k\in N}$ can be easily established because of the nonexpansiveness of the corresponding operator $R_{\lambda f}R_{\lambda g}$; see details in [50, Lemma 2.4]. Moreover, even when the operator $R_{\lambda f}$ or $R_{\lambda g}$ is evaluated approximately, as analyzed in [58, Lemma 7] and [58, Theorem 1], the quasi-Fejér monotonicity of the corresponding sequence and hence the rate of asymptotic regularity can be established. For the "strongly + weakly" convex setting of the model (1.1) under discussion, however, the operator $R_{\lambda f}R_{\lambda g}$ is not necessarily nonexpansive, and it requires more technical analysis for proving the monotonicity of the sequence $\{\|\widetilde{e}(z_k,\lambda)\|^2\}_{k\in N}$.

Now, we establish the rate of asymptotic regularity for $\widetilde{T}_{DR}$ in the following theorem.

THEOREM 5.1. *Let* $\{z_k\}_{k\in N}$ *be the sequence generated by the DRSM* (1.4) *with* $\beta > \omega$ *and* λ *satisfying* (3.4). *Then there exists* $\eta_1 > 0$ *such that*

$$\|\widetilde{e}(z_k,\lambda)\|^2 \le \frac{d^2(z_0, \mathrm{Fix}(\widetilde{T}_{DR}))}{(k+1)\eta_1\alpha^2} \quad \forall k \in N. \tag{5.14}$$

Moreover, it holds that

$$\|\widetilde{e}(z_k,\lambda)\|^2 = o(1/k) \quad \text{as } k \to \infty. \tag{5.15}$$

That is, the rate of asymptotic regularity for $\widetilde{T}_{DR}$ *is* $o(1/\sqrt{k})$.

Proof. It follows from Proposition 4.1 that $\mathrm{Fix}(\widetilde{T}_{DR}) \neq \emptyset$. Let z^* be a fixed point of $\widetilde{T}_{DR}$. Then it follows from (5.3) and Theorem 4.3 that there exists $\eta_1 > 0$ such that

$$\|z_{k+1} - z^*\|^2 \le \|z_k - z^*\|^2 - \eta_1\alpha^2\|\widetilde{e}(z_k,\lambda)\|^2 \quad \forall k \in N. \tag{5.16}$$

Thus, we have

$$\sum_{k=0}^{\infty} \eta_1\alpha^2\|\widetilde{e}(z_k,\lambda)\|^2 \le \|z_0 - z^*\|^2.$$

Since z^* is an arbitrary fixed point of $\widetilde{T}_{DR}$, we then have

$$\sum_{k=0}^{\infty} \eta_1\alpha^2\|\widetilde{e}(z_k,\lambda)\|^2 \le d^2(z_0, \mathrm{Fix}(\widetilde{T}_{DR})). \tag{5.17}$$

Moreover, it follows from (5.4) that

$$(k+1)\|\widetilde{e}(z_k,\lambda)\|^2 \le \sum_{i=0}^{k} \|\widetilde{e}(z_i,\lambda)\|^2. \tag{5.18}$$

Therefore, the assertion (5.14) follows from (5.17) and (5.18). Also, according to (5.17), we have

$$\sum_{i=\lfloor \frac{k}{2} \rfloor}^{k} \|\widetilde{e}(z_i,\lambda)\|^2 \to 0 \quad \text{as } k \to \infty,$$

where $\lfloor \frac{k}{2} \rfloor$ is the largest integer which is not greater than the number $\frac{k}{2}$. Again from (5.4), we have

$$\frac{k}{2}\|\widetilde{e}(z_k,\lambda)\|^2 \le \sum_{i=\lfloor \frac{k}{2} \rfloor}^{k} \|\widetilde{e}(z_i,\lambda)\|^2 \to 0 \quad \text{as } k \to \infty, \tag{5.19}$$

which directly implies the assertion (5.15). Thus, combining (5.15) with (5.2), we show the $o(1/\sqrt{k})$ rate of asymptotic regularity for $\widetilde{T}_{DR}$ defined in (1.4). The proof is complete. ∎

Remark 5.2. By (4.9) and Lemma 2.3, we can deduce that $\text{prox}_{\lambda g}$ is Lipschitz continuous with constant $\frac{1}{1-\lambda\omega}$. Then it follows from (5.3) and (5.14) that

$$\begin{aligned}\|\text{prox}_{\lambda g}(z_{k+1}) - \text{prox}_{\lambda g}(z_k)\|^2 &\leq \frac{1}{(1-\lambda\omega)^2}\|z_{k+1} - z_k\|^2 \\ &\leq \frac{d^2(z_0, \text{Fix}(\widetilde{T}_{DR}))}{(1-\lambda\omega)^2(k+1)\eta_1} \quad \forall k \in N. \end{aligned} \tag{5.20}$$

Recall that Theorem 4.4 also shows the convergence of $\{\text{prox}_{\lambda g}(z_k)\}_{k\in N}$ to the unique solution of (1.1). Thus, (5.20) also implies the $o(1/\sqrt{k})$ rate of asymptotic regularity for $\text{prox}_{\lambda g}$.

5.2. Rate of asymptotic regularity of $\widehat{T}_{DR}$. In this subsection, we established the rate of asymptotic regularity for the Douglas–Rachford operator $\widehat{T}_{DR}$ defined in (1.6). Similarly to the last subsection, for any $z_0 \in \mathcal{R}^n$, we have

$$z_k - \widehat{T}_{DR}(z_k) = z_k - z_{k+1} = \alpha(z_k - R_{\lambda g}R_{\lambda f}(z_k)).$$

If we define

$$\widehat{e}(z, \lambda) := z - R_{\lambda g}R_{\lambda f}(z), \tag{5.21}$$

then it holds that

$$z_k - \widehat{T}_{DR}(z_k) = \alpha\widehat{e}(z_k, \lambda).$$

Also, using the notation (5.21), the DRSM scheme (1.6) can be rewritten as

$$z_{k+1} = z_k - \alpha\widehat{e}(z_k, \lambda) = (1-\alpha)z_k + \alpha R_{\lambda g}R_{\lambda f}(z_k), \quad \alpha \in (0, 1). \tag{5.22}$$

Hence, it is reasonable to use $\|z_k - R_{\lambda g}R_{\lambda f}(z_k)\|$, i.e., $\|\widehat{e}(z_k, \lambda)\|$, to measure the rate of asymptotic regularity of $\widehat{T}_{DR}$. We first investigate the monotonicity of $\{\|\widehat{e}(z_k, \lambda)\|\}_{k\in N}$ in the following lemma.

LEMMA 5.2. *Let $\{z_k\}_{k\in N}$ be the sequence generated by the DRSM* (1.6) *with $\beta > \omega$ and λ satisfying* (3.4). *Then for any $k \in N$, we have*

$$\|\widehat{e}(z_{k+1}, \lambda)\|^2 \leq \|\widehat{e}(z_k, \lambda)\|^2.$$

Proof. For any $k \in N$, taking $x := z_{k+1}$ and $y := z_k$ in (4.12), we have

$$\begin{aligned}&\|R_{\lambda g}R_{\lambda f}(z_{k+1}) - R_{\lambda g}R_{\lambda f}(z_k)\|^2 \\ &\quad \leq \|z_{k+1} - z_k\|^2 - 4\beta\lambda\|\text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k)\|^2 \\ &\qquad + 4\omega\lambda\|\text{prox}_{\lambda g}R_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda g}R_{\lambda f}(z_k)\|^2. \end{aligned} \tag{5.23}$$

Thus, it holds that

$$\begin{aligned}\|\tilde{e}(z_{k+1}, \lambda)\|^2 &= \|\tilde{e}(z_{k+1}, \lambda) - \tilde{e}(z_k, \lambda)\|^2 + 2\langle\tilde{e}(z_{k+1}, \lambda), \tilde{e}(z_{k+1}, \lambda) - \tilde{e}(z_k, \lambda)\rangle + \|\tilde{e}(z_k, \lambda)\|^2 \\ &= \|(z_{k+1} - R_{\lambda g}R_{\lambda f}(z_{k+1})) - (z_k - R_{\lambda g}R_{\lambda f}(z_k))\|^2 \\ &\quad + 2\langle\widehat{e}(z_k, \lambda), \widehat{e}(z_{k+1}, \lambda) - \widehat{e}(z_k, \lambda)\rangle + \|\widehat{e}(z_k, \lambda)\|^2 \\ &= \|z_{k+1} - z_k\|^2 + \|R_{\lambda g}R_{\lambda f}(z_{k+1}) - R_{\lambda g}R_{\lambda f}(z_k)\|^2 + \|\widehat{e}(z_k, \lambda)\|^2 \\ &\quad + 2\langle\widehat{e}(z_k, \lambda), \widehat{e}(z_{k+1}, \lambda) - \widehat{e}(z_k, \lambda)\rangle - 2\langle z_{k+1} - z_k, R_{\lambda g}R_{\lambda f}(z_{k+1}) - R_{\lambda g}R_{\lambda f}(z_k)\rangle \\ &\leq 2\|z_{k+1} - z_k\|^2 - 4\beta\lambda\|\text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k)\|^2 + \|\widehat{e}(z_k, \lambda)\|^2 \\ &\quad + 4\omega\lambda\|\text{prox}_{\lambda g}R_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda g}R_{\lambda f}(z_k)\|^2 + 2\langle\widehat{e}(z_k, \lambda), \widehat{e}(z_{k+1}, \lambda) - \widehat{e}(z_k, \lambda)\rangle \\ &\quad - 2\langle z_{k+1} - z_k, R_{\lambda g}R_{\lambda f}(z_{k+1}) - R_{\lambda g}R_{\lambda f}(z_k)\rangle, \end{aligned} \tag{5.24}$$

where the inequality follows from (5.23). Moreover, it follows from (4.24), (5.21), and (5.22) that

$$\text{prox}_{\lambda g} R_{\lambda f}(z_k) = \text{prox}_{\lambda f}(z_k) - \frac{1}{2}\widehat{e}(z_k,\lambda) \quad \text{and} \quad R_{\lambda g} R_{\lambda f}(z_k) = z_k - \widehat{e}(z_k,\lambda). \tag{5.25}$$

Thus, by means of (5.25) we have

$$\begin{aligned} &\|\text{prox}_{\lambda g} R_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda g} R_{\lambda f}(z_k)\|^2 \\ &\quad = \|\text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k)\|^2 + \left\|\frac{1}{2}\widehat{e}(z_{k+1},\lambda) - \frac{1}{2}\widehat{e}(z_k,\lambda)\right\|^2 \\ &\qquad - \langle \text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k), \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle \end{aligned} \tag{5.26}$$

and

$$\begin{aligned} &\langle z_{k+1} - z_k, R_{\lambda g} R_{\lambda f}(z_{k+1}) - R_{\lambda g} R_{\lambda f}(z_k)\rangle \\ &\quad = \|z_{k+1} - z_k\|^2 - \langle z_{k+1} - z_k, \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle. \end{aligned} \tag{5.27}$$

Substituting (5.26) and (5.27) into (5.24), we get

$$\begin{aligned} \|\widehat{e}(z_{k+1},\lambda)\|^2 \le\ & 2\|z_{k+1} - z_k\|^2 + (4\omega\lambda - 4\beta\lambda)\|\text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k)\|^2 + \|\widehat{e}(z_k,\lambda)\|^2 \\ & + \omega\lambda\|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2 + 2\langle \widehat{e}(z_k,\lambda), \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle \\ & - 2\|z_{k+1} - z_k\|^2 + 2\langle z_{k+1} - z_k, \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle \\ & - 4\omega\lambda\langle \text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k), \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle \\ =\ & (4\omega\lambda - 4\beta\lambda)\|\text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k)\|^2 + \|\widehat{e}(z_k,\lambda)\|^2 \\ & + \omega\lambda\|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2 + (2 - 2\alpha)\langle \widehat{e}(z_k,\lambda), \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle \\ & - 4\omega\lambda\langle \text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k), \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle, \end{aligned} \tag{5.28}$$

where the equality follows from the fact that $z_{k+1} - z_k = -\alpha\widehat{e}(z_k,\lambda)$. Furthermore, it follows from the Cauchy–Schwarz inequality that

$$\begin{aligned} & - 4\omega\lambda\langle \text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k), \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle \\ &\quad \le (4\beta\lambda - 4\omega\lambda)\|\text{prox}_{\lambda f}(z_{k+1}) - \text{prox}_{\lambda f}(z_k)\|^2 + \frac{\omega^2\lambda}{\beta - \omega}\|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2. \end{aligned} \tag{5.29}$$

Thus, substituting (5.30) into (5.28), we obtain

$$\begin{aligned} \|\widehat{e}(z_{z+1},\lambda)\|^2 \le\ & \|\widehat{e}(z_k,\lambda)\|^2 + \omega\lambda\|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2 + \frac{\omega^2\lambda}{\beta - \omega}\|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2 \\ & + (2 - 2\alpha)\langle \widehat{e}(z_k,\lambda), \widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\rangle \\ =\ & \|\widehat{e}(z_k,\lambda)\|^2 + \omega\lambda\|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2 + \frac{\omega^2\lambda}{\beta - \omega}\|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2 \\ & + (1 - \alpha)(\|\widehat{e}(z_{k+1},\lambda)\|^2 - \|\widehat{e}(z_k,\lambda)\|^2 - \|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2), \end{aligned}$$

where the equality follows from (5.12). Hence, rearranging the terms, we get

$$\alpha\|\widehat{e}(z_{k+1},\lambda)\|^2 \le \alpha\|\widehat{e}(z_k,\lambda)\|^2 + \left(\frac{\beta\lambda\omega}{\beta - \omega} + (\alpha - 1)\right)\|\widehat{e}(z_{k+1},\lambda) - \widehat{e}(z_k,\lambda)\|^2. \tag{5.30}$$

Recall the condition (3.4); we thus have $\frac{\beta\lambda\omega}{\beta-\omega} + \alpha - 1 < 0$. The proof is complete. ◻

Based on Theorem 4.6 and Lemma 5.2, we can establish the rate of asymptotic regularity for $\widehat{T}_{DR}$ in the following theorem. The proof is similar to that of Theorem 5.1 and thus is omitted.

THEOREM 5.2. *Let $\{z_k\}_{k\in N}$ be the sequence generated by the DRSM* (1.6) *with $\beta > \omega$ and λ satisfying* (3.4). *Then there exists $\eta_3 > 0$ such that*

$$\|\widehat{e}(z_k,\lambda)\|^2 \le \frac{d^2(z_0,\mathrm{Fix}(\widehat{T}_{DR}))}{(k+1)\eta_3\alpha^2} \quad \forall k\in N.$$

Moreover, it holds that

$$\|\widehat{e}(z_k,\lambda)\|^2 = o(1/k) \quad \text{as } k\to\infty.$$

That is, the rate of asymptotic regularity for $\widehat{T}_{DR}$ is $o(1/\sqrt{k})$.

5.3. Relationship to some existing works. There is a rich literature focusing on analyzing the rate of asymptotic regularity for the Krasnoselskiĭ–Mann iteration in the nonexpansive operator context; see, e.g., [5, 6, 7, 24, 26, 35, 58].

Since the first explicit estimate of the rate of asymptotic regularity for an averaged operator $\widehat{M}$ was derived in [6], i.e., $\|z_k - \widehat{M}(z_k)\| = O(1/\sqrt{k})$, some refined results have been studied in the literature such as [35, 58]. Let $(X, \|\cdot\|)$ be a normed space, let $C\subset X$ be a convex compact subset, let $M: C\to C$ be a nonexpansive operator, and let

$$z_k := (1-\alpha_k)z_{k-1} + \alpha_k M(z_{k-1}),$$

with $z_0\in C$ and $\{\alpha_k\}_{k\in N}\subset[0,1]$. It was conjectured in [6] whether there exists a universal constant $\varsigma > 0$ such that

$$\|z_k - M(z_k)\| \le \varsigma\cdot\frac{\mathrm{diam}(C)}{\sqrt{\sum_{i=1}^k \alpha_i(1-\alpha_i)}}, \tag{5.31}$$

where $\mathrm{diam}(C)$ is the diameter of C. This conjecture was answered positively in [35] by proving that the bound given in the right-hand side of (5.31) holds in general with $\varsigma = 1/\sqrt{\pi}$ for any sequence $\{\alpha_k\}_{k\in N}\subset[0,1]$ and nonexpansive operator $M: C\to C$. Moreover, when C is unbounded and $\mathrm{Fix}(M)\neq\emptyset$, by virtue of (5.31), the authors of [35] proved the following rate of asymptotic regularity:

$$\|z_k - M(z_k)\| \le 2\varsigma\cdot\frac{d(z_0,\mathrm{Fix}(M))}{\sqrt{\sum_{i=1}^k \alpha_i(1-\alpha_i)}},$$

where $d(z_0,\mathrm{Fix}(M))$ is the distance of the point z_0 to the set $\mathrm{Fix}(M)$.

Recently, the following inexact version of the Krasnoselskiĭ–Mann iteration with errors was considered in [58]:

$$z_{k+1} := z_k + \alpha_k(M(z_k) + \epsilon_k - z_k),$$

where $M:\mathcal{H}\to\mathcal{H}$ is a nonexpansive operator, $\mathcal{H}$ is a real Hilbert space with norm $\|\cdot\|$, $\{\alpha_k\}_{k\in N}\subset[0,1]$, and ϵ_k is the error of approximating $M(z_k)$. The authors proved that the global pointwise and ergodic iteration-complexity bounds (i.e., the rate of asymptotic regularity) are $O(1/\sqrt{k})$ and $O(1/k)$, respectively. More precisely, it is the residual operator $e_1(z) := z - M(z)$ for the nonexpansive operator M that is used to measure the accuracy of an iterate. The nonexpansiveness of the operator M guarantees that the residual operator $e_1(\cdot)$ has the property

$$\langle y_1 - y_2, e_1(y_1) - e_1(y_2)\rangle \ge \frac{1}{2}\|e_1(y_1) - e_1(y_2)\|^2 \quad \text{for any } \ y_1, y_2\in\mathcal{H}.$$

This property plays a role similar to the firm nonexpansiveness of an operator, and it is essential for proving the convergence rate results in [58]. These interesting results can be immediately applied to the convergence rate analysis for inexact versions of the DRSM schemes (1.4) and (1.6) when both f and g are convex in (1.1), but not for the "strongly + weakly" convex case because the operators $R_{\lambda f}R_{\lambda g}$ and $R_{\lambda g}R_{\lambda f}$ given, respectively, in (1.4) and (1.6) are not necessarily nonexpansive.

As mentioned, some other works in the literature require additional stronger conditions to ensure the nonexpansiveness of the operators $R_{\lambda f}R_{\lambda g}$ and $R_{\lambda g}R_{\lambda f}$ and hence the convergence of the DRSM schemes (1.4) and (1.6). One example, as mentioned in section 1, is [21]. In our analysis, we directly study the Fejér monotonicity of the sequence $\{z_k\}_{k\in N}$ generated by the DRSM (1.4) or (1.6), and stronger conditions to ensure the nonexpansiveness of $R_{\lambda f}R_{\lambda g}$ and $R_{\lambda g}R_{\lambda f}$ can thus be avoided.

6. Local linear convergence rate. Under the metric subregularity assumption on model (1.1), we can establish the local linear convergence rate in the asymptotical sense for the DRSM schemes (1.4) and (1.6). According to [39, p. 199], a set-valued mapping $T : \mathcal{R}^n \rightrightarrows \mathcal{R}^n$ is called metrically subregular at z^* for y^* if $y^* \in T(z^*)$ and there exists $\kappa \geq 0$, along with a neighborhood $N(z^*)$ of z^*, such that

$$d(z, T^{-1}(y^*)) \leq \kappa \cdot d(y^*, T(z)) \quad \forall z \in N(z^*). \tag{6.1}$$

The infimum of all κ such that (6.1) holds is the modulus of metric subregularity, denoted by $\text{subreg}(T; z^*|y^*)$. The absence of metric subregularity is signaled by $\text{subreg}(T; z^*|y^*) = +\infty$.

Based on this assumption, we can establish the local linear convergence rate for the DRSM schemes (1.4) and (1.6). Because of their similarity, we present detailed analysis only for (1.4). Recall that the operators $R_{\lambda f}R_{\lambda g}$ and $R_{\lambda g}R_{\lambda f}$ under discussion are not necessarily nonexpansive. Hence, the local linear convergence rates to be established are not covered by the results in [58], which are valid for nonexpansive operators under the metric subregularity assumption of $T' := I - T$ in which T is a nonexpansive operator.

THEOREM 6.1. *Let $\{z_k\}_{k\in N}$ be the sequence generated by the DRSM* (1.4) *with $\beta > \omega$ and λ satisfying* (3.4)*, and let z^* be a fixed point of* $\text{Fix}(\widetilde{T}_{DR})$*. Assume that $\widetilde{e}(\cdot, \lambda) = (I - R_{\lambda f}R_{\lambda g})(\cdot)$ is metrically subregular at z^* for* 0 *with a neighborhood $N(z^*)$ of z^* and modulus $\kappa > 0$. Take sufficiently small $r > 0$ such that $B(z^*, r) \subseteq N(z^*)$. Then for any starting point $z_0 \in B(z^*, r)$ and for any $k \in N$, we have*

$$d(z_{k+1}, \text{Fix}(\widetilde{T}_{DR})) \leq \sqrt{1 - \frac{\eta_1 \alpha^2}{\kappa^2}} \cdot d(z_k, \text{Fix}(\widetilde{T}_{DR})),$$

where η_1 is defined in (4.18)*. That is, the DRSM* (1.4) *converges to* $\text{Fix}(\widetilde{T}_{DR})$ *linearly.*

Proof. For any $k \in N$, it follows from (4.19) that

$$\|z_{k+1} - z^*\| \leq \|z_k - z^*\| \leq \cdots \leq \|z_0 - z^*\| \leq r,$$

which means $\{z_k\}_{k\in N} \subset N(z^*)$. Since $\widetilde{e}(\cdot, \lambda)^{-1}(0) = \text{Fix}(\widetilde{T}_{DR})$ and $\widetilde{e}(\cdot, \lambda)$ is metrically subregular at z^* for 0 with a neighborhood $N(z^*)$, we have

$$d(z, \text{Fix}(\widetilde{T}_{DR})) \leq \kappa \|\widetilde{e}(z, \lambda)\| \quad \forall z \in N(z^*).$$

Thus, we have

$$d(z_k, \text{Fix}(\widetilde{T}_{DR})) \leq \kappa \|\widetilde{e}(z_k, \lambda)\| \quad \forall k \in N. \tag{6.2}$$

Note that $\text{Fix}(\widetilde{T}_{DR})$ is a closed set. Taking $\hat{z}$ as a fixed point of $\widetilde{T}_{DR}$ such that $d(z_k, \text{Fix}(\widetilde{T}_{DR})) = \|z_k - \hat{z}\|$, by (5.16) we know that

$$\begin{aligned} d^2(z_{k+1}, \text{Fix}(\widetilde{T}_{DR})) &\le \|z_{k+1} - \hat{z}\|^2 \\ &\le \|z_k - \hat{z}\|^2 - \eta_1 \alpha^2 \|\widetilde{e}(z_k, \lambda)\|^2 \\ &= d^2(z_k, \text{Fix}(\widetilde{T}_{DR})) - \eta_1 \alpha^2 \|\widetilde{e}(z_k, \lambda)\|^2 \\ &\le d^2(z_k, \text{Fix}(\widetilde{T}_{DR})) - \frac{\eta_1 \alpha^2}{\kappa^2} \cdot d^2(z_k, \text{Fix}(\widetilde{T}_{DR})) \quad \forall k \in N, \end{aligned}$$

where the last inequality follows from (6.2). The proof is complete. ∎

Remark 6.1. The local linear convergence rate can also be established for the DRSM schemes (1.4) and (1.6) under the coercivity condition. According to [51, Lemma 3.1(b)], $T : D \rightrightarrows \mathcal{R}^n$ satisfies the coercivity condition if there exists $\zeta > 0$ such that

$$\|z - z_+\| \ge \zeta \cdot d(z, S) \quad \forall z_+ \in T(z), \quad \forall z \in U,$$

where $D \subset \mathcal{R}^n$, $S \subset \text{Fix}(T)$, and $U \subset D$. The proof is similar to that of Theorem 6.1 and thus is omitted. For the feasibility problem of finding an intersection point of a subspace and a closed super-regular set in a finite dimensional space, the local linear convergence rate of the DRSM scheme has been well studied under the coercivity condition; see, e.g., [51]. We also refer the reader to [2] for the local linear convergence rate of the DRSM in the "convex + convex" case of minimizing two functions under the coercivity condition. Here, our result can be regarded as an extension of these known results to the "strongly + weakly" convex setting of (1.1).

Remark 6.2. The linear convergence proved in Theorem 6.1 is also called weak linear convergence in [55]. Moreover, if it is assumed that $\text{Fix}(\widetilde{T}_{DR})$ is a singleton,[3] i.e., $\text{Fix}(\widetilde{T}_{DR}) = \{z^*\}$, then $\{z_k\}_{k\in N}$ linearly converges to the unique fixed point z^* in norm.

Remark 6.3. In [18], the linear convergence of the DRSM is established for the specific convex feasibility problem. It is shown in [18] that the corresponding Douglas–Rachford operator is boundedly linearly regular[4] [18, Definition 2.1(ii)] in the convex feasibility scenario. This is a regularity condition weaker than the standard regularity condition, e.g., $\text{ri}(A) \cap \text{ri}(B) \ne \emptyset$ (see, [18, Theorem 4.4]). Moreover, the bounded linear regularity is also related to the metric subregularity, as elaborated in [18, 58].

7. Conclusions.

In this paper, we analyzed the convergence of the well-known Douglas–Rachford splitting method (DRSM) for the minimization of two functions, where one is strongly convex and the other is weakly convex. The convergence of two DRSM schemes was proved, and the rates of asymptotic regularity for the corresponding Douglas–Rachford operators were derived under relatively mild assumptions. Moreover, the local linear convergence rate of two DRSM schemes in the asymptotical sense was also analyzed under some regularity assumptions. Though the strongly convex function dominates the objective function in our discussion, and thus analytically

[3]When $\widetilde{T}_{DR}$ is a contraction mapping in the sense that there exists $0 < \sigma < 1$ such that $d(\widetilde{T}_{DR}(x), \widetilde{T}_{DR}(y)) \le \sigma \cdot d(x, y)\ \forall x, y \in \mathcal{R}^n$, it follows from the Banach contraction mapping principle [29, Theorem 2.1.7] that $\widetilde{T}_{DR}$ has the unique fixed point.

[4]Let $\mathcal{H}$ be a real Hilbert space, and let $M : \mathcal{H} \to \mathcal{H}$ be such that $\text{Fix}(M) \ne \emptyset$. We say that M is boundedly linearly regular with constant $\kappa \ge 0$ if $(\forall \rho > 0)$ $(\exists \kappa \ge 0)$ $(\forall z \in B(0; \rho))$ $d(z, \text{Fix}(M)) \le \kappa \cdot \|z - Mz\|$.

the model under our discussion is strongly convex, the algorithmic design for splitting methods including the DRSM intrinsically requires individual treatment of the proximal operators of two functions in the objective. Thus, proving the convergence of the DRSM for the "strongly + weakly" convex case requires more sophisticated analysis than for the "convex + convex" case, because the reflection operator of the weakly convex function is expansive instead of nonexpansive. Unless some stronger additional assumptions are made, this difficulty seems to exclude the applicability of some existing techniques that essentially rely on the nonexpansiveness property of the involved operators in DRSM schemes.

It is interesting to consider extending our analysis to inexact versions of the DRSM schemes (1.4) and (1.6) which may evaluate the operators $R_{\lambda f}$ and $R_{\lambda g}$ approximately subject to certain inexactness criteria. We expect the proof of the Fejér monotonicity of the corresponding sequence to be more challenging than the exact versions considered in this paper. The recent work [58] offers valuable techniques for the possible extension despite the fact that the analysis therein is mainly for nonexpansive operators. It is also interesting to conduct the convergence analysis for other splitting schemes such as the more challenging Peaceman–Rachford splitting method (PRSM) in the "strongly + weakly" convex setting. Our techniques focusing on the Fejér monotonicity of the sequence instead of the nonexpansiveness of the operators (as elaborated in Remark 4.2) still seem to be crucial to completing these works.

REFERENCES

[1] A. Antoniadis, *Wavelet methods in statistics: Some recent developments and their applications*, Statist. Surv., 1 (2007), pp. 16–55.

[2] T. Aspelmeier, C. Charitha, and D. R. Luke, *Linear convergence of the ADMM/Douglas–Rachford algorithms without strong convexity and application to statistical imaging*, SIAM J. Imaging Sci., 9 (2016), pp. 842–868, https://doi.org/10.1137/15M103580X.

[3] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.

[4] H. Attouch, J. Peypouquet, and P. Redont, *Backward-forward algorithms for structured monotone inclusions in Hilbert spaces*, J. Math. Anal. Appl., to appear, http://doi.org/10.1016/j.jmaa.2016.06.025.

[5] J. B. Baillon and R. E. Bruck, *Optimal rates of asymptotic regularity for averaged nonexpansive mappings*, in Proceedings of the Second International Conference on Fixed Point Theory and Applications, K.-K. Tan, ed., World Scientific, Singapore, 1992, pp. 27–66.

[6] J. B. Baillon and R. E. Bruck, *The rate of asymptotic regularity in $O(1/\sqrt{n})$*, in Theory and Applications of Nonlinear Operators of Accretive and Monotone Types, Lecture Notes in Pure Appl. Math. 178, Marcel Dekker, New York, 1996, pp. 51–81.

[7] J. B. Baillon, R. E. Bruck, and S. Reich, *On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces*, Houston J. Math., 4 (1978), pp. 1–9.

[8] H. H. Bauschke, *The composition of the projections onto closed convex sets in Hilbert space is asymptotically regular*, Proc. Amer. Math. Soc., 131 (2002), pp. 141–146.

[9] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, Berlin, 2011.

[10] H. H. Bauschke, P. L. Combettes, and D. R. Luke, *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces*, J. Approx. Theory, 127 (2004), pp. 178–192.

[11] H. H. Bauschke, J. Y. B. Cruz, T. T. A. Nghia, H. M. Phan, and X. F. Wang, *The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle*, J. Approx. Theory, 185 (2014), pp. 63–79.

[12] H. H. Bauschke, M. N. Dao, D. Noll, and H. M. Phan, *On Slater's condition and finite convergence of the Douglas-Rachford algorithm for solving convex feasibility problems in Euclidean spaces*, J. Global Optim., 65 (2016), pp. 329–349.

[13] H. H. Bauschke, W. L. Hare, and W. M. Moursi, *On the range of the Douglas-Rachford operator*, Math. Oper. Res., 41 (2016), pp. 884–897.

[14] H. H. Bauschke, V. R. Koch, and H. M. Phan, *Stadium norm and Douglas-Rachford splitting: A new approach to road design optimization*, Oper. Res., 64 (2016), pp. 201–218.
[15] H. H. Bauschke and W. M. Moursi, *On the order of the operators in the Douglas-Rachford algorithm*, Optim. Lett., 10 (2016), pp. 447–455.
[16] H. H. Bauschke and W. M. Moursi, *The Douglas-Rachford algorithm for two (not necessarily intersecting) affine subspaces*, SIAM J. Optim., 26 (2016), pp. 968–985, https://doi.org/10.1137/15M1016989.
[17] H. H. Bauschke and D. Noll, *On the local convergence of the Douglas-Rachford algorithm*, Arch. Math., 102 (2014), pp. 589–600.
[18] H. H. Bauschke, D. Noll, and H. M. Phan, *Linear and strong convergence of algorithms involving averaged nonexpansive operators*, J. Math. Anal. Appl., 421 (2015), pp. 1–20.
[19] İ. Bayram, *On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty*, IEEE Trans. Signal Process., 64 (2016), pp. 1597–1608.
[20] İ. Bayram, *Penalty functions derived from monotone mappings*, IEEE Signal Process. Lett., 22 (2015), pp. 265–269.
[21] İ. Bayram and I. W. Selesnick, *The Douglas-Rachford Algorithm for Weakly Convex Penalties*, preprint, http://arxiv.org/abs/1511.03920v1, 2015.
[22] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, *Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity*, Trans. Amer. Math. Soc., 362 (2010), pp. 3319–3363.
[23] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), Ser. A, pp. 459–494.
[24] J. M. Borwein, S. Reich, and I. Shafrir, *Krasnosel'ski-Mann iterations in normed spaces*, Canad. Math. Bull., 35 (1992), pp. 21–28.
[25] F. E. Browder and W. V. Petryshyn, *The solution by iteration of nonlinear functional equations in Banach spaces*, Bull. Amer. Math. Soc., 72 (1966), pp. 571–575.
[26] R. E. Bruck, *Structure of the approximate fixed-point sets of nonexpansive mappings in general Banach spaces*, in Fixed Point Theory and Applications (Marseille, 1989), Pitman Res. Notes Math. 252, Longman Sci. Tech., Harlow, 1991, pp. 91–96.
[27] R. E. Bruck and S. Reich, *Nonexpansive projections and resolvents of accretive operators in Banach spaces*, Houston J. Math., 3 (1977), pp. 459–470.
[28] P. Cannarsa and C. Sinestrari, *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*, Progr. Nonlinear Differential Equations Appl. 58, Birkhäuser Boston, Inc., Boston, MA, 2004.
[29] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Lecture Notes in Math. 2057, Springer, Heidelberg, 2012.
[30] R. Chartrand, *Shrinkage mappings and their induced penalty functions*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, 2014, pp. 1026–1029.
[31] L. M. Chen and Y. T. Gu, *The convergence guarantees of a non-convex approach for sparse recovery using regularized least squares*, IEEE Trans. Signal Process., 62 (2014), pp. 3754–3767.
[32] X. J. Chen and W. J. Zhou, *Convergence of reweighted ℓ_1 minimization algorithms for ℓ_2-ℓ_p minimization*, Comput. Optim. Appl., 59 (2014), pp. 47–61.
[33] P. L. Combettes, *Fejér monotonicity in convex optimization*, in Encyclopedia of Optimization, C. A. Floudas and P. M. Pardalos, eds., Kluwer Academic, Boston, 2001, pp. 106–114.
[34] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms*, in Inherently Parallel Algorithms for Feasibility and Optimization, D. Butnariu, Y. Censor, and S. Reich, eds., Elsevier, New York, 2001, pp. 115–152.
[35] R. Cominetti, J. A. Soto, and J. Vaisman, *On the rate of convergence of Krasnosel'ski-Mann iterations and their connection with sums of Bernoullis*, Isr. J. Math., 199 (2014), pp. 757–772.
[36] E. Corman and X. Yuan, *A generalized proximal point algorithm and its convergence rate*, SIAM J. Optim., 24 (2014), pp. 1614–1638, https://doi.org/10.1137/130940402.
[37] M. Degiovanni, A. Marino, and M. Tosques, *Evolution equations with lack of convexity*, Nonlinear Anal., 9 (1985), pp. 1401–1443.
[38] L. Demanet and X. Zhang, *Eventual linear convergence of the Douglas-Rachford iteration for basis pursuit*, Math. Comput., 85 (2016), pp. 209–238.
[39] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings: A View from Variational Analysis*, Springer, New York, 2009.
[40] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two or three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[41] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[42] M. Edelstein, *A remark on a theorem of M. A. Krasnosel'ski*, Amer. Math. Monthly, 14 (1970), pp. 65–73.

[43] M. Edelstein and R. C. O'Brien, *Nonexpansive mappings, asymptotic regularity and successive approximations*, J. London Math. Soc., 17 (1978), pp. 547–554.

[44] J. Q. Fan and R. Z. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Statist. Assoc., 96 (2001), pp. 1348–1360.

[45] M. Fukushima, *The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem*, Math. Program., 72 (1996), pp. 1–15.

[46] R. Glowinski, T. Kärkkäinen, and K. Majava, *On the convergence of operator-splitting methods*, in Numerical Methods for Scientific Computing. Variational Problems and Applications, E. Heikkola et al., eds., CIMNE, Barcelona, 2003, pp. 67–69.

[47] K. Goebel and W. A. Kirk, *Iteration processes for nonexpansive mappings*, in Topological Methods in Nonlinear Functional Analysis. Contemp. Math. 21, S. P. Singh, S. Thomeier, and B. Watson, eds., Amer. Math. Soc., Providence, RI, 1983, pp. 115–123.

[48] E. G. Gol'shtein and N. V. Tret'yakov, *Modified Lagrangians in convex programming and their generalizations*, Math. Program. Stud., 10 (1979), pp. 86–97.

[49] K. Guo, X. M. Yuan, and S. Z. Zeng, *Convergence Analysis of ISTA and FISTA for "Strongly + Semi" Convex Programming*, manuscript.

[50] B. S. He and X. M. Yuan, *On the convergence rate of the Douglas-Rachford operator splitting method*, Math. Program., 153 (2015), pp. 715–722.

[51] R. Hesse and D. R. Luke, *Nonconvex notions and regularity and convergence of fundamental algorithm for feasibility problems*, SIAM J. Optim., 23 (2013), pp. 2397–2419, https://doi.org/10.1137/120902653.

[52] S. Ishikawa, *Fixed points and iterations of a nonexpansive mapping in a Banach space*, Proc. Amer. Math. Soc., 59 (1976), pp. 65–71.

[53] M. Kowalski, *Thresholding RULES and iterative shrinkage/thresholding algorithm: A convergence study*, in Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 4151–4155.

[54] M. A. Krasnosel'ski, *Two remarks on the method of successive approximations*, Uspehi Mat. Nauk (N.S.), 10 (1955), pp. 123–127 (in Russian).

[55] D. Leventhal, *Metric subregularity and the proximal point method*, J. Math. Anal. Appl., 360 (2009), pp. 681–688.

[56] G. Y. Li and T. K. Pong, *Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems*, Math. Program., 159 (2016), pp. 371–401.

[57] G.Y Li and T. K. Pong, *Peaceman-Rachford Splitting for a Class of Nonconvex Optimization Problems*, preprint, http://arxiv.org/abs/1507.00887v1, 2015.

[58] J. W. Liang, J. Fadili, and G. Peyré, *Convergence rates with inexact non-expansive operators*, Math. Program., 159 (2016), Ser. A, pp. 403–434.

[59] J. W. Liang, M. J. Fadili, and G. Peyré, *Local Convergence Properties of Douglas-Rachford and ADMM*, preprint, http://arxiv.org/abs/1606.02116, 2016.

[60] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979, https://doi.org/10.1137/0716071.

[61] S. Marcellin and L. Thibault, *Evolution problems associated with primal lower nice functions*, J. Convex Anal., 13 (2006), pp. 385–421.

[62] T. Mollenhoff, E. Strekalovskiy, M. Moeller, and D. Cremers, *The primal-dual hybrid gradient method for semiconvex splittings*, SIAM J. Imaging Sci., 8 (2015), pp. 827–857, https://doi.org/10.1137/140976601.

[63] M. Nikolova, *Estimation of binary images by minimizing convex criteria*, in Proceedings of the IEEE International Conference on Image Processing (ICIP), Chicago, vol. 2, 1998, pp. 108–112.

[64] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., 73 (1967), pp. 591–597.

[65] D. W. Peaceman and H. H. Rachford, Jr., *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Ind. Appl. Math., 3 (1955), pp. 28–41, https://doi.org/10.1137/0103003.

[66] H. M. Phan, *Linear convergence of the Douglas-Rachford method for two closed sets*, Optim., 65 (2016), pp. 369–385.

[67] S. Reich and A. J. Zaslavski, *Convergence of Krasnosel'ski-Mann iterations of nonexpansive operators*, Math. Comput. Model., 32 (2000), pp. 1423–1431.

[68] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 2015.
[69] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
[70] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 227–238.
[71] I. W. Selesnick and İ. Bayram, *Signal estimation by maximally sparse convex optimization*, IEEE Trans. Signal Process., 62 (2014), pp. 1078–1092.
[72] X. F. Wang, *On Chebyshev functions and Klee functions*, J. Math. Anal. Appl., 368 (2010), pp. 293–310.
[73] C.-H. Zhang, *Nearly unbiased variable selection under minimax concave penalty*, Ann. Statist., 38 (2010), pp. 894–942.