Swapping objectives accelerates Davis–Yin splitting

Edward Duc Hien Nguyen* Jaewook J. Suh† Xin Jiang‡ Shiqian Ma†

June 30, 2025

Abstract

In this work, we investigate the application of Davis–Yin splitting (DYS) to convex optimization problems and demonstrate that swapping the roles of the two nonsmooth convex functions can result in a faster convergence rate. Such a swap typically yields a different sequence of iterates, but its impact on convergence behavior has been largely understudied or often overlooked. We address this gap by establishing best-known convergence rates for DYS and its swapped counterpart, using the primal–dual gap function as the performance metric. Our results indicate that variants of the Douglas–Rachford splitting algorithm (a special case of DYS) share the same worst-case rate, whereas the convergence rates of the two DYS variants differ. This discrepancy is further illustrated through concrete examples.

1 Introduction

We study proximal splitting methods for convex optimization problems in the following form

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) + g(x) + h(x), \tag{1.1}$$

where f, g, and h are closed convex proper (CCP) functions and h is differentiable. This problem covers a wide variety of applications in machine learning, signal and image processing, operations research, control, and other fields [7,9,11,23,28]. A well-known method for solving problem (1.1) is the Davis–Yin splitting (DYS) algorithm [14, Algorithm 1]:

$$\begin{aligned} w^{k+1} &= \mathbf{prox}_{\alpha g}(y^k) \\ x^{k+1} &= \mathbf{prox}_{\alpha f}(2w^{k+1} - y^k - \alpha\nabla h(w^{k+1})) \\ y^{k+1} &= y^k + x^{k+1} - w^{k+1}, \end{aligned} \tag{DYS}$$

where $\alpha > 0$ is a stepsize. (DYS) recovers several classical methods when parts of problem (1.1) vanish. For example, it reduces to the Douglas–Rachford splitting (DRS) algorithm [15,21,25] when $h = 0$, and to the forward–backward splitting (FBS) algorithm [21,24] when either f or g vanishes.

One may observe that f and g play symmetric roles in problem (1.1), and thus swapping them does not alter the problem at all. Yet this symmetry does not carry over to the algorithmic level.

*Department of Electrical and Computer Engineering, Rice University. Email: en18@rice.edu.

†Department of Computational Applied Mathematics and Operations Research, Rice University. Email: {jacksuh, shiqian.ma}@rice.edu.

‡School of Operations Research and Information Engineering, Cornell University. Email: xjiang@cornell.edu.

Swapping f and g in (DYS) generally leads to a different sequence of variables and thus a different algorithm [17,33]. This distinction is not merely theoretical: in practical applications, f and g often represent structurally different components [7,9], such as a data-fitting term and a regularizer, making it important to clearly specify which function plays which role. While the non-equivalence due to update order is well understood, its effect on convergence rate remains underexplored—particularly because, for commonly used performance measures such as the primal–dual gap function, the analysis does not directly extend to the swapped algorithm.

In view of this subtle yet critical distinction, we analyze the convergence rate of the four algorithms: DYS, DRS, and their swapped versions, with the primal–dual gap function as the performance measure. Perhaps surprisingly, the swapped version of (DYS) achieves a faster worst-case ergodic rate of convergence, underscoring the tangible impact of update order on algorithmic performance.

Contributions. The contributions of this paper are summarized as follows.

- We provide novel convergence analyses of DYS, DRS, and their swapped versions, by deriving equalities that explicitly reveal the critical inequalities used in the proof. These equalities not only imply immediately the ergodic convergence rates for the primal–dual gap function but also provide guidance for constructing worst-case examples for each algorithm.
- Our analyses reveal that the swapped version of DYS converges faster than the original one, whereas both variants of DRS share the same worst-case rate. This appears to be the first result that formally distinguishes between the two variants of splitting methods based on their convergence rates.
- The tightness of the established rates for the two DRS variants is confirmed through worst-case examples. The discrepancy between the two DYS variants is also demonstrated via concrete examples.

Outline. The rest of the paper is organized as follows. Section 2 reviews fundamental concepts from convex analysis and summarizes existing convergence results for DYS and DRS. In Section 3, we analyze the convergence of two DRS variants and establish the tightness of the results via worst-case examples. Similarly, Section 4 presents analyses of two DYS variants, demonstrating a difference in their convergence rates via concrete examples. Finally, Section 5 concludes the paper.

2 Background material and prior work

In this section, we review fundamental concepts from convex optimization and introduce the notation used throughout the paper. We also present two variants of DRS and DYS, along with their known convergence results. Although both DRS and DYS were originally proposed to solve monotone inclusion problems, we focus in this paper on the seemingly more restrictive setting of convex optimization. This choice is motivated by the fact that the discrepancy in convergence rates between the two variants of DYS arises specifically within this narrower setting. In particular, the illustrative examples presented later involve only convex functions.

Throughout the paper, we use the notation $\langle x, y\rangle = x^T y$ for the standard inner product of vectors x and y, and $\|x\| = \langle x, x\rangle^{1/2}$ for the Euclidean norm of a vector x. Also, we denote $\mathbb{N}$ (resp., $\mathbb{N}_+$) as the set of nonnegative (resp., positive) integers, *i.e.*, $\mathbb{N} \coloneqq \{0, 1, 2, \ldots\}$ and $\mathbb{N}_+ \coloneqq \{1, 2, \ldots\}$.

2.1 Basic concepts and notation in convex optimization

We follow the standard definitions in convex optimization; see, *e.g.*, [2, 5, 22, 27, 28]. We denote the subdifferential of a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ as ∂f, defined by $\partial f(x) := \{v \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle v, y - x\rangle \text{ for all } y \in \mathbb{R}^n\}$. An element of $\partial f(x)$ is called a subgradient of f at x, and when its choice is unambiguous, we use the shorthand $\widetilde{\nabla} f(x)$, following the notation introduced in [3]: $\widetilde{\nabla} f(x) \in \partial f(x)$. With this, the inequality in the definition of ∂f becomes

$$f(x) - f(y) + \langle \widetilde{\nabla} f(x), y - x\rangle \leq 0 \quad \text{for all } y \in \mathbf{dom}\, f. \tag{2.1}$$

The notation $\widetilde{\nabla} f$ is particularly useful in studying the proximal operator of a CCP function f:

$$\mathbf{prox}_f(x) := \underset{y\in\mathbb{R}^n}{\operatorname{argmin}}\{f(y) + \tfrac{1}{2}\|y - x\|^2\}.$$

We often denote $\widetilde{\nabla} f(\mathbf{prox}_{\alpha f}(x)) := \frac{1}{\alpha}(\mathbf{prox}_{\alpha f}(x) - x) \in \partial f(\mathbf{prox}_{\alpha f}(x))$, and then the proximal operator of αf (with $\alpha > 0$) can be written as:

$$\mathbf{prox}_{\alpha f}(x) = x - \alpha \widetilde{\nabla} f(\mathbf{prox}_{\alpha f}(x)). \tag{2.2}$$

Since $\mathbf{prox}_{\alpha f}$ is well defined when f is CCP, $\widetilde{\nabla} f(\mathbf{prox}_{\alpha f}(x))$ is uniquely determined.

For any function f, its Fenchel conjugate is defined as $f^*(y) := \sup_x \{\langle y, x\rangle - f(x)\}$. When f is CCP, it holds that

$$f^*(y) = \langle y, x\rangle - f(x) \qquad \Longleftrightarrow \qquad y \in \partial f(x) \qquad \Longleftrightarrow \qquad x \in \partial f^*(y), \tag{2.3}$$

which is called Fenchel's identity, and the equality

$$\mathbf{prox}_{\alpha f}(x) + \alpha\, \mathbf{prox}_{\alpha^{-1} f^*}(x/\alpha) = x \tag{2.4}$$

holds for all x and all $\alpha > 0$, which is known as Moreau identity.

We say f is L-smooth if it is differentiable and its gradient is L-Lipschitz continuous. When f is L-smooth and convex, it satisfies the inequality (see, *e.g.*, [22, Theorem 2.1.5])

$$f(x) - f(y) + \langle \nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \leq 0 \quad \text{for all } x, y \in \mathbf{dom}\, f. \tag{2.5}$$

Dual problem and optimality conditions. The dual of problem (1.1) is

$$\underset{u\in\mathbb{R}^n}{\text{maximize}} \quad -(f + h)^*(-u) - g^*(u), \tag{2.6}$$

where the conjugate $(f + h)^*$ is the infimal convolution of f^* and h^*:

$$(f + h)^*(u) = \inf_w \{f^*(w) + h^*(u - w)\}.$$

The primal–dual optimality conditions for (1.1) and (2.6) are

$$0 \in \partial f(x) + \nabla h(x) + u, \qquad 0 \in \partial g^*(u) - x. \tag{2.7}$$

Throughout the paper, we assume (2.7) is solvable, so strong duality holds for (1.1) and (2.6).

We will refer to the convex–concave function

$$\mathcal{L}(x,u) = f(x) + h(x) + \langle u, x\rangle - g^*(u) \tag{2.8}$$

as the *Lagrangian* of (1.1). We follow the convention that $\mathcal{L}(x,u) = +\infty$ if $x \notin \mathbf{dom}(f+h)$ and $\mathcal{L}(x,u) = -\infty$ if $x \in \mathbf{dom}(f+h)$ and $u \notin \mathbf{dom}\, g^*$. The objectives in (1.1) and the dual problem (2.6) can be respectivley expressed as

$$\sup_u \mathcal{L}(x,u) = f(x) + g(x) + h(x), \qquad \inf_x \mathcal{L}(x,u) = -(f+h)^*(-u) - g^*(u).$$

A solution $(x^\star, u^\star)$ of the optimality conditions (2.7) forms a saddle point of $\mathcal{L}$:

$$\inf_x \sup_u \mathcal{L}(x,u) = \sup_u \mathcal{L}(x^\star,u) = \inf_x \mathcal{L}(x,u^\star) = \sup_u \inf_x \mathcal{L}(x,u).$$

Then, it holds that

$$\mathcal{L}(x^\star,u) \le \mathcal{L}(x^\star,u^\star) \le \mathcal{L}(x,u^\star)$$

for all $x \in \mathbf{dom}\, f$ and $u \in \mathbf{dom}\, g^*$. In particular, $\mathcal{L}(x^\star,u^\star)$ is the optimal value of (1.1) and (2.6), and the pair of primal–dual problems is equivalent to the saddle point problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \underset{u\in\mathbb{R}^n}{\text{maximize}} \quad \mathcal{L}(x,u). \tag{2.9}$$

Discussion on performance measures. Our analysis focuses on the convergence of the algorithms applied to solving the saddle point problem (2.9) and studies the *primal–dual gap function*

$$\mathcal{L}(\overline{x}^K,u) - \mathcal{L}(x,\overline{u}^K) = f(\overline{x}^K) + h(\overline{x}^K) + \langle u, \overline{x}^K\rangle - g^*(u) - \big(f(x) + h(x) + \langle \overline{u}^K, x\rangle - g^*(\overline{u}^K)\big). \tag{2.10}$$

More precisely, the performance measure we choose is

$$\sup_{\substack{x\in\mathbf{dom}\, f\\ u\in\mathbf{dom}\, g^*}} \left\{\mathcal{L}(\overline{x}^K,u) - \mathcal{L}(x,\overline{u}^K)\right\}, \tag{2.11}$$

which has been used extensively in the analysis of primal–dual splitting methods; see, *e.g.*, [6, 8, 10, 20, 29, 31]. One shall note that the quantity $\mathcal{L}(\overline{x}^K,u^\star) - \mathcal{L}(x^\star,\overline{u}^K)$ is not a valid performance measure: while this quantity is 0 when $(\overline{x}^K,\overline{u}^K)$ is a saddle point, the converse is not necessarily true [6]. So the pointwise supremum over (x,u) is necessary in (2.11). Unlike gradient-based methods, the choice of the performance measure in DRS and DYS is somewhat arbitrary and often dictated by the analysis techniques employed. It is difficult to determine which metric is most natural or informative, as each comes with its own advantages and limitations. For example, the objective gap $|f(x^K) + g(w^K) - f(x^\star) - g(x^\star)|$, used in the seminal work [13], does not capture the fact that both x^K and w^K in DRS converge to a primal solution $x^\star$. As a result, this measure alone does not guarantee the convergence of the iterates. By contrast, the objective value $f(x^K) + g(x^K)$ can be $+\infty$ as x^K may not be in the domain of g. So the use of $|f(x^K) + g(x^K) - f(x^\star) - g(x^\star)|$ is valid if, *e.g.*, g is additionally locally Lipschitz [13].

2.2 Douglas–Rachford splitting algorithms

In this section, we present two variants of DRS for solving the problem (1.1) in the special case where $h = 0$. Although these algorithms can be viewed as special cases of DYS, we present them explicitly to highlight that the discrepancy in convergence rates between the two variants of DYS arises from the presence of the smooth function h.

DRS-gf. The Douglas–Rachford splitting algorithm [15, 18, 21, 25] for solving (1.1) with $h = 0$ is

$$w^{k+1} = \mathbf{prox}_{\alpha g}(y^k) \tag{2.12a}$$
$$x^{k+1} = \mathbf{prox}_{\alpha f}(2w^{k+1} - y^k) \tag{2.12b}$$
$$y^{k+1} = y^k + x^{k+1} - w^{k+1}, \tag{2.12c}$$

where $\alpha > 0$ is a stepsize. Then, we define $u^{k+1} = \mathbf{prox}_{\alpha^{-1}g^*}(\frac{1}{\alpha}y^k)$ and eliminate w^{k+1} and y^{k+1} to obtain

$$\begin{aligned} u^{k+1} &= \mathbf{prox}_{\alpha^{-1}g^*}(u^k + \tfrac{1}{\alpha}x^k) \\ x^{k+1} &= \mathbf{prox}_{\alpha f}(x^k - \alpha(2u^{k+1} - u^k)), \end{aligned} \tag{DRS-gf}$$

which we will refer to as DRS-gf, as it calls g first and then f. The presented form (DRS-gf) solves the saddle point problem (2.9) with $h = 0$ and has been studied in, *e.g.*, [11, Eq. (5.18)] and [20, §3]. Its convergence has been analyzed under various regularity conditions and here we focus on the most basic setting where both f and g are CCP functions. In this case, convergence of the iterates follows readily, as (2.12) can be interpreted as an instance of the proximal point method [18]. Specifically, the iterates (x^k, w^k, y^k) generated by (2.12) converge to $(x^\star, x^\star, x^\star + \alpha u^\star)$ and the iterates (x^k, u^k) generated by (DRS-gf) converge to $(x^\star, u^\star)$. In addition, (DRS-gf) converges at a sublinear rate. More precisely, the non-ergodic sequence in (2.12) exhibits the following rate [13, Theorem 4], [19]:

$$|f(x^K) + g(w^K) - f(x^\star) - g(x^\star)| = o(1/\sqrt{K+1}), \tag{2.13}$$

while the ergodic sequence $(\bar{x}^K, \bar{w}^K)$ converges at a faster $\mathcal{O}(1/(K+1))$ rate [13, Theorem 3]:

$$|f(\bar{x}^K) + g(\bar{w}^K) - f(x^\star) - g(x^\star)| = \mathcal{O}(1/(K+1)), \tag{2.14}$$

where given $K \in \mathbb{N}_+$, we define $\bar{z}^K := \frac{1}{K}\sum_{k=1}^K z^k$ for $z \in \{x, w, u\}$. These rates are shown to be tight up to a constant [13]. Moreover, the $\mathcal{O}(1/(K+1))$ ergodic rate remains valid when the primal–dual gap function (2.10) is used [4, 8, 20].

DRS-fg. Since (2.12) is not symmetric in f and g, exchanging f and g yields a different algorithm

$$\begin{aligned} x^{k+1} &= \mathbf{prox}_{\alpha f}(y^k) \\ w^{k+1} &= \mathbf{prox}_{\alpha g}(2x^{k+1} - y^k) \\ y^{k+1} &= y^k + w^{k+1} - x^{k+1}. \end{aligned}$$

Letting $u^{k+1} = \alpha^{-1}(2x^{k+1} - y^k - w^{k+1})$ and applying Moreau identity (2.4), we can eliminate w^{k+1} and y^{k+1} and obtain an equivalent algorithm

$$\begin{aligned} x^{k+1} &= \mathbf{prox}_{\alpha f}(x^k - \alpha u^k) \\ u^{k+1} &= \mathbf{prox}_{\alpha^{-1}g^*}(u^k + \tfrac{1}{\alpha}(2x^{k+1} - x^k)), \end{aligned} \tag{DRS-fg}$$

which we will refer to as DRS-fg.

In general, (DRS-gf) and (DRS-fg) generate different iterates [33] and are thus considered *not equivalent*. Nevertheless, the convergence of (DRS-fg) iterates can be established via similar arguments, as (DRS-fg) can also be interpreted as an instance of the proximal point method with a different splitting strategy [17,18]. Moreover, since the objective gap is invariant in the order of f and g, the convergence rate results (2.13) and (2.14) carry over directly to (DRS-fg). In contrast, the primal–dual gap function (2.10) treats f and g asymmetrically, so the convergence guarantees for (DRS-gf) established in [4, 8] do not automatically extend to (DRS-fg). As in this paper we adopt the primal–dual gap function (2.10) as the performance measure, our analysis for (DRS-fg) must proceed separately, even though both algorithms ultimately exhibit the same worst-case rate.

2.3 Davis–Yin splitting algorithms

DYS-gf. The iterations (DYS) were first presented in [14] and are now referred to as the Davis–Yin splitting algorithm in the literature. Again, we introduce $u^{k+1} = \mathbf{prox}_{\alpha^{-1}g^*}(\frac{1}{\alpha}y^k)$ and apply Moreau identity (2.4) to eliminate w^{k+1} and y^{k+1}:

$$\begin{aligned} u^{k+1} &= \mathbf{prox}_{\alpha^{-1}g^*}(u^k + \tfrac{1}{\alpha}x^k) \\ x^{k+1} &= \mathbf{prox}_{\alpha f}(x^k - \alpha(2u^{k+1} - u^k) - \alpha\nabla h(x^k + \alpha(u^k - u^{k+1}))), \end{aligned} \tag{DYS-gf}$$

which we will refer to as DYS-gf. This form solves the saddle point problem (2.9) and when $h = 0$, (DYS-gf) reduces to (DRS-gf). DYS was originally introduced as a fixed-point iteration scheme for solving monotone inclusion problems, so its convergence rate is often measured by the fixed-point residual; see, *e.g.*, [14]. As in our discussion of DRS, we focus on the convex optimization setting under minimal assumptions, where f and g are merely CCP functions. While convergence of the DYS iterates is well established [14, §4.1], the convergence rate of (DYS-gf) (under the general convex setting) has received less attention. In existing works [26,29,32,37], DYS is typically viewed as a primal–dual splitting method, with the primal–dual gap function (2.10) used as the performance measure, and an $\mathcal{O}(1/K)$ ergodic convergence rate is typically established. Yet the tightness of these rates is not discussed. Finally, we note that numerous variants of DYS have been proposed, including stochastic DYS [34–37], inexact DYS [38], adaptive DYS [26], and inertial DYS [12]. Some of these consider settings different from ours and hence not discussed in detail.

DYS-fg. As for (DRS-gf), we switch the role of f and g in (DYS) and obtain a different algorithm

$$\begin{aligned} x^{k+1} &= \mathbf{prox}_{\alpha f}(x^k - \alpha(u^k + \nabla h(x^k))) \\ u^{k+1} &= \mathbf{prox}_{\alpha^{-1}g^*}(u^k + \tfrac{1}{\alpha}(2x^{k+1} - x^k + \alpha\nabla h(x^k) - \alpha\nabla h(x^{k+1}))), \end{aligned} \tag{DYS-fg}$$

which we will refer to as DYS-fg. When $h = 0$, (DYS-fg) reduces to (DRS-fg). This form (DYS-fg) was first presented as a special case of the PD3O algorithm [31]; see also [11, Eq. (8.6)] and [20, §3]. As noted earlier, (DRS-gf) and (DRS-fg) are not equivalent; nor are (DYS-gf) and (DYS-fg). Consequently, the results in [14, 26, 32, 37] for (DYS-gf) do not directly extend to (DYS-fg), as they adopt performance measures that treat f and g asymmetrically. An $\mathcal{O}(1/K)$ ergodic rate for (DYS-fg) is established in [31, Theorem 2] (using a slightly different gap function, which starts at iteration 0 rather than 1) and [20, Theorem 4], though the tightness of these rates is not addressed.

2.4 Preview of the established convergence rates

Despite all the aforementioned literature on proximal splitting methods, direct analyses of the swapped algorithms, (DRS-fg) and (DYS-fg), remain underdeveloped, and worse-case rate comparisons with their original counterparts, (DRS-gf) and (DYS-gf), have received limited attention.

Table 1: An overview of convergence rates and lower bound examples for the four algorithms. All results are derived under the unified performance measure (2.11) and the initial distance (2.15).

	Convergence rate		Lower bound example	
DRS-gf	$D/(K+1)$	(Theorem 3.2)	$D/(K+1)$	(Theorem 3.4)
DRS-fg	$D/(K+1)$	(Theorem 3.6)	$D/(K+1)$	(Theorem 3.8)
DYS-gf	D/K	(Theorem 4.2)	slower than $D/(K+1)$	(Theorem 4.3)
DYS-fg	$D/(K+1)$	(Theorem 4.5)	$D/(K+1)$	(Corollary 4.6)

To address this gap and better understand the impact of swapping the two nonsmooth functions f and g, we establish the best-known convergence rates for all four algorithms and provide concrete examples for illustration; see Table 1 for a concise overview. Recall that our analysis is based on the unified performance measure (2.11), and for fair comparison, we adopt a unified initial distance

$$D_0(x,u) := \frac{1}{\alpha}\|x^0 - x\|^2 + \alpha\|u^0 - u\|^2. \tag{2.15}$$

Accordingly, a rate $D/(K+1)$ in Table 1 means that the metric (2.10) is smaller than or equal to $D_0(x,u)/(K+1)$ for all $x, u \in \mathbb{R}^n$, a convention we will use throughout the paper. For (DRS-gf), (DRS-fg), and (DYS-fg), we construct concrete examples that attain the stated rate, thereby confirming their tightness. For (DYS-gf), we present an instance whose rate is strictly slower than $D/(K+1)$, the tight worst-case rate for (DYS-fg). This demonstrates that the worst-case rate of (DYS-gf) is indeed slower than that of (DYS-fg), implying that swapping f and g in (DYS) can lead to a faster convergence.

3 Convergence analysis of two variants of DRS

In this section, we analyze the convergence of (DRS-gf) and (DRS-fg) using the primal–dual gap function (2.10) as the performance measure. To demonstrate the tightness of our results, we construct worst-case examples for which the two variants of DRS converge at *exactly* the established rates. Although the worst-case examples differ only by a sign, the analyses of (DRS-gf) and (DRS-fg) must be carried out separately. This separation also facilitates a clear comparison with the analyses of (DYS-gf) and (DYS-fg) in Section 4, whose rates, perhaps surprisingly, do not coincide. Note that $h \equiv 0$ in this section.

3.1 DRS-gf: worst-case rate and its tightness

We begin our analysis of (DRS-gf) with another equivalent reformulation

$$u^{k+1} = u^k + \tfrac{1}{\alpha}x^k - \tfrac{1}{\alpha}\widetilde{\nabla} g^*(u^{k+1}) \tag{3.1a}$$

$$x^{k+1} = x^k - \alpha(2u^{k+1} - u^k) - \alpha\widetilde{\nabla} f(x^{k+1}), \tag{3.1b}$$

where we used (2.2). For later use, we also introduce the notation

$$p^k := \widetilde{\nabla} g^*(u^k) = \alpha u^{k-1} + x^{k-1} - \alpha u^k, \tag{3.2}$$

which implies from (2.3) that $u^{k+1} \in \partial g(p^{k+1})$. We will see how the use of $\widetilde{\nabla}$ in reformulating (DRS-gf), inspired by [3,14], facilitates the analysis of (DRS-gf).

Similarly, for the ergodic iterate $\overline{u}^K$ and an arbitrary $u \in \mathbb{R}^n$, we denote their corresponding subgradients of g^* as $\overline{p}^K$ and p, respectively:

$$\overline{p}^K \in \partial g^*(\overline{u}^K), \qquad p \in \partial g^*(u). \tag{3.3}$$

One shall be aware that $\overline{p}^K$ is a subgradient of g^* at $\overline{u}^K$ rather than the average of $\{p^k\}_{k=1}^K$. It then follows from (2.3) that $\overline{u}^K \in \partial g(\overline{p}^K)$, $u \in \partial g(p)$, and

$$g^*(\overline{u}^K) = \langle \overline{u}^K, \overline{p}^K \rangle - g(\overline{p}^K), \qquad g^*(u) = \langle u, p \rangle - g(p).$$

The introduction of $\overline{p}^K$ and p helps reformulate the primal–dual gap function (2.10) (with $h = 0$):

$$\mathcal{L}(\overline{x}^K, u) - \mathcal{L}(x, \overline{u}^K) = f(\overline{x}^K) + \langle u, \overline{x}^K \rangle - \langle u, p \rangle + g(p) - \big(f(x) + \langle \overline{u}^K, x \rangle - \langle \overline{u}^K, \overline{p}^K \rangle + g(\overline{p}^K) \big). \tag{3.4}$$

We now derive an equality that will play an important role in obtaining the tight convergence rate of (DRS-gf). Its development is motivated by a computer-aided analysis framework known as the performance estimation problem (PEP) [16,30].

Proposition 3.1. *Suppose f and g are CCP functions, and $\{(x^k, u^k)\}_{k \in \mathbb{N}}$ is generated by* (DRS-gf) *with stepsize $\alpha > 0$ and initial points (x^0, u^0). Denote p^k as in* (3.2)*, and $\overline{p}^K$, p as in* (3.3)*. For $K \in \mathbb{N}_+$, define the ergodic iterates*

$$\overline{x}^K := \frac{1}{K} \sum_{k=1}^{K} x^k, \qquad \overline{u}^K := \frac{1}{K} \sum_{k=1}^{K} u^k. \tag{3.5}$$

Then, for all $K \in \mathbb{N}_+$ and all $x, u \in \mathbb{R}^n$, the equality

$$\mathcal{L}(\overline{x}^K, u) - \mathcal{L}(x, \overline{u}^K) - \frac{D_0(x, u)}{K+1} = I_f + I_g - \mathcal{S}_1 - \mathcal{S}_2 \tag{3.6}$$

holds, where $D_0(x, u)$ is defined in (2.15)*,*

$$v^k := \widetilde{\nabla} f(x^k) + u^k \tag{3.7}$$

and

$$\begin{aligned}
I_f &:= \frac{1}{K} \sum_{k=1}^{K} \Big(f(\overline{x}^K) - f(x^k) + \Big\langle \widetilde{\nabla} f(\overline{x}^K), x^k - \overline{x}^K \Big\rangle \Big) + \frac{1}{K} \sum_{k=1}^{K} \Big(f(x^k) - f(x) + \Big\langle \widetilde{\nabla} f(x^k), x - x^k \Big\rangle \Big) \\
I_g &:= \frac{1}{K} \sum_{k=1}^{K} \Big(g(p^k) - g(\overline{p}^K) + \Big\langle u^k, \overline{p}^K - p^k \Big\rangle \Big) + \frac{1}{K} \sum_{k=1}^{K} \Big(g(p) - g(p^k) + \Big\langle u, p^k - p \Big\rangle \Big) \\
\mathcal{S}_1 &:= \frac{1}{\alpha(K+1)} \left(\left\| x^0 - x - \frac{\alpha(K+1)}{2K} \sum_{k=1}^{K} v^k \right\|^2 + \alpha^2 \left\| u^0 - u - \frac{K+1}{2K} \sum_{k=1}^{K} v^k \right\|^2 \right) \\
\mathcal{S}_2 &:= \frac{\alpha}{2K^2} \sum_{k=1}^{K} \sum_{l=1}^{k-1} \left\| v^k - v^l \right\|^2 .
\end{aligned} \tag{3.8}$$

Before proving Proposition 3.1, we present its two immediate implications. More specifically, Proposition 3.1 is used to establish an ergodic convergence rate of (DRS-gf) (see Theorem 3.2) and also helpful in building a worst-case example (see Corollary 3.3).

Theorem 3.2 (Convergence of (DRS-gf)). *Suppose f and g are CCP functions, and $\{(x^k, u^k)\}_{k\in\mathbb{N}}$ is generated by (DRS-gf) with stepsize $\alpha > 0$ and initial points (x^0, u^0). Then, for all $K \in \mathbb{N}_+$, the ergodic iterates $(\bar{x}^K, \bar{u}^K)$ defined in (3.5) satisfy*

$$\mathcal{L}(\bar{x}^K, u) - \mathcal{L}(x, \bar{u}^K) \le \frac{D_0(x,u)}{K+1} \tag{3.9}$$

for all $x \in \mathbf{dom}\, f$ and all $u \in \mathbf{dom}\, g^*$, where $D_0(x,u)$ is defined in (2.15).

Proof. It follows from (2.1) and the convexity of f and g that I_f and I_g defined in (3.8) are nonpositive. Moreover, the two quantities $\mathcal{S}_1$ and $\mathcal{S}_2$ are nonnegative since they are sums of squared terms. The desired conclusion (3.9) then follows directly from Proposition 3.1. □

Moreover, the inequality (3.9) holds with equality if and only if the four quantities defined in (3.8) are all zero. This result is formalized in Corollary 3.3 and will be used to build a worst-case example that demonstrates the tightness of the convergence rate (3.9).

Corollary 3.3. *Let $x \neq x^0$ and $u \neq u^0$. Under the same setting as in Theorem 3.2, the inequality (3.9) holds with equality if and only if the four quantities I_f, I_g, $\mathcal{S}_1$, and $\mathcal{S}_2$ are all zero.*

Proof. Recall from the proof of Theorem 3.2 that I_f and I_g are nonpositive and $\mathcal{S}_1$ and $\mathcal{S}_2$ are nonnegative. This implies that $I_f + I_g - \mathcal{S}_1 - \mathcal{S}_2 = 0$ if and only if each term is zero. So, the desired conclusion follows directly from Theorem 3.2. □

Now, we prove Proposition 3.1.

Proof of Proposition 3.1. It follows from (3.4) that the left-hand side of (3.6) equals

$$\text{LHS} = f(\bar{x}^K) + \langle u, \bar{x}^K\rangle + g(p) - \langle u, p\rangle - \big(f(x) + \langle \bar{u}^K, x\rangle + g(\bar{p}^K) - \langle \bar{u}^K, \bar{p}^K\rangle\big) - \frac{D_0(x,u)}{K+1} \tag{3.10}$$

To establish the identity (3.6), we simplify the four terms in (3.8) one by one. For I_f, we observe that

$$\frac{1}{K}\sum_{k=1}^{K}\left\langle \widetilde{\nabla} f(\bar{x}^K), x^k - \bar{x}^K\right\rangle = \left\langle \widetilde{\nabla} f(\bar{x}^K), \frac{1}{K}\sum_{k=1}^{K} x^k - \bar{x}^K\right\rangle = 0.$$

Note that the $f(x^k)$ terms cancel out, so I_f simplifies to

$$I_f = f(\bar{x}^K) - f(x) + \left\langle \frac{1}{K}\sum_{k=1}^{K}\widetilde{\nabla} f(x^k), x\right\rangle - \frac{1}{K}\sum_{k=1}^{K}\left\langle \widetilde{\nabla} f(x^k), x^k\right\rangle, \tag{3.11}$$

in which the first two terms appear in the LHS in (3.10).

For I_g, we apply $\frac{1}{K}\sum_{k=1}^{K}\left\langle u^k, \bar{p}^K\right\rangle = \left\langle \bar{u}^K, \bar{p}^K\right\rangle$ and obtain

$$I_g = g(p) - g(\bar{p}^K) - \langle u, p\rangle + \left\langle \bar{u}^K, \bar{p}^K\right\rangle + \left\langle u, \frac{1}{K}\sum_{k=1}^{K} p^k\right\rangle - \frac{1}{K}\sum_{k=1}^{K}\left\langle u^k, p^k\right\rangle. \tag{3.12}$$

Then it follows from (3.1b), (3.2), and (3.7) that

$$p^k = x^k + \alpha\left(\widetilde{\nabla} f(x^k) + u^k\right) = x^k + \alpha v^k. \tag{3.13}$$

Substituting (3.13) back in (3.12) eliminates $\{p_k\}_{k=1}^K$ and thus I_g becomes

$$I_g = g(p) - g(\bar{p}^K) - \langle u, p\rangle + \langle \bar{u}^K, \bar{p}^K\rangle + \langle u, \bar{x}^K\rangle + \alpha\left\langle u, \frac{1}{K}\sum_{k=1}^K v^k\right\rangle - \frac{1}{K}\sum_{k=1}^K\left\langle u^k, x^k + \alpha v^k\right\rangle. \tag{3.14}$$

Note that all terms appear in the LHS in (3.10) except the last two terms.

Similarly, for $\mathcal{S}_1$, we expand and re-organize the squares and obtain

$$\mathcal{S}_1 = \frac{D_0(x,u)}{K+1} - \left\langle x^0 + \alpha u^0 - x - \alpha u, \frac{1}{K}\sum_{k=1}^K v^k\right\rangle + \frac{\alpha(K+1)}{2K^2}\left\|\sum_{k=1}^K v^k\right\|^2.$$

Then, it is straightforward to verify that

$$\begin{aligned}
&(I_f + I_g - \mathcal{S}_1) - \text{LHS} \\
&= \left\langle \frac{1}{K}\sum_{k=1}^K \widetilde{\nabla} f(x^k), x\right\rangle - \frac{1}{K}\sum_{k=1}^K\left\langle \widetilde{\nabla} f(x^k), x^k\right\rangle + \alpha\left\langle u, \frac{1}{K}\sum_{k=1}^K v^k\right\rangle - \frac{1}{K}\sum_{k=1}^K\left\langle u^k, x^k + \alpha v^k\right\rangle \\
&\quad + \left\langle x^0 + \alpha u^0 - x - \alpha u, \frac{1}{K}\sum_{k=1}^K v^k\right\rangle - \frac{\alpha(K+1)}{2K^2}\left\|\sum_{k=1}^K v^k\right\|^2 + \langle x, \bar{u}^K\rangle \\
&= -\frac{1}{K}\sum_{k=1}^K\left\langle \widetilde{\nabla} f(x^k), x^k\right\rangle - \frac{1}{K}\sum_{k=1}^K\left\langle u^k, x^k + \alpha v^k\right\rangle + \frac{1}{K}\sum_{k=1}^K\left\langle x^0 + \alpha u^0, v^k\right\rangle - \frac{\alpha(K+1)}{2K^2}\left\|\sum_{k=1}^K v^k\right\|^2,
\end{aligned} \tag{3.15}$$

where in the second equality we cancel out all the inner product terms involving x or u. Then, it follows from (3.7) that the first two terms on the right-hand side of (3.15) simplifies to

$$\begin{aligned}
-\frac{1}{K}\sum_{k=1}^K\left\langle \widetilde{\nabla} f(x^k), x^k\right\rangle - \frac{1}{K}\sum_{k=1}^K\left\langle u^k, x^k + \alpha v^k\right\rangle &= -\frac{1}{K}\sum_{k=1}^K\left\langle \widetilde{\nabla} f(x^k) + u^k, x^k\right\rangle - \frac{1}{K}\sum_{k=1}^K\left\langle v^k, \alpha u^k\right\rangle \\
&= -\frac{1}{K}\sum_{k=1}^K\left\langle v^k, x^k + \alpha u^k\right\rangle.
\end{aligned}$$

Substituting this into (3.15) and reorganizing, we obtain

$$(I_f + I_g - \mathcal{S}_1) - \text{LHS} = \frac{1}{K}\sum_{k=1}^K\left\langle v^k, x^0 + \alpha u^0 - x^k - \alpha u^k\right\rangle - \frac{\alpha(K+1)}{2K^2}\left\|\sum_{k=1}^K v^k\right\|^2. \tag{3.16}$$

Now we eliminate x^k. Applying (3.1b) recursively, we obtain

$$x^k = x^{k-1} - \alpha(u^k - u^{k-1}) - \alpha v^k = \cdots = x^0 + \alpha u^0 - \alpha u^k - \alpha\sum_{l=1}^k v^l.$$

Substituting it into (3.16) and using (3.7), we obtain

$$
\begin{aligned}
&(I_f + I_g - \mathcal{S}_1) - \text{LHS} \\
&= \frac{\alpha}{K}\sum_{k=1}^{K}\left\langle v^k, \sum_{l=1}^{k} v^l \right\rangle - \frac{\alpha(K+1)}{2K^2}\left\|\sum_{k=1}^{K} v^k\right\|^2 \\
&= \frac{\alpha}{K}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\langle v^k, v^l\right\rangle + \frac{\alpha}{K}\sum_{k=1}^{K}\left\|v^k\right\|^2 - \frac{\alpha(K+1)}{2K^2}\left(\sum_{k=1}^{K}\left\|v^k\right\|^2 + 2\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\langle v^k, v^l\right\rangle\right) \\
&= \frac{\alpha}{2K^2}\left(-2\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\langle v^k, v^l\right\rangle + (K-1)\sum_{k=1}^{K}\left\|v^k\right\|^2\right) \\
&= \frac{\alpha}{2K^2}\left(-2\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\langle v^k, v^l\right\rangle + \sum_{k=1}^{K}\sum_{l=1}^{k-1}\left(\left\|v^k\right\|^2 + \left\|v^l\right\|^2\right)\right) \\
&= \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\|v^k - v^l\right\|^2 = \mathcal{S}_2.
\end{aligned} \tag{3.17}
$$

The second-to-last equality follows from $(K-1)\sum_{k=1}^{K}\|v^k\|^2 = \sum_{k=1}^{K}\sum_{l=1}^{k-1}(\|v^k\|^2 + \|v^l\|^2)$, which can be verified by comparing the coefficients of each $\left\|v^i\right\|^2$ terms. Therefore, $I_f + I_g - \mathcal{S}_1 - \mathcal{S}_2 = \text{LHS}$, which is our desired conclusion. □

The tightness of (3.9) is now verified using a worst-case example motivated by Corollary 3.3.

Theorem 3.4 (Worst-case example for (DRS-gf)). *Under the same setting as in Theorem 3.2, for any $K \in \mathbb{N}_+$ and any $\alpha > 0$, there exist CCP functions f and g and points $x^0, u^0, \tilde{x}, \tilde{u} \in \mathbb{R}^n$ such that $\alpha D_0(\tilde{x}, \tilde{u}) = 1$ where $D_0(\tilde{x}, \tilde{u})$ is defined in (2.15) and*

$$
\mathcal{L}(\bar{x}^K, \tilde{u}) - \mathcal{L}(\tilde{x}, \bar{u}^K) = \frac{D_0(\tilde{x}, \tilde{u})}{K+1}.
$$

Proof. Fix $K \in \mathbb{N}_+$ and $\alpha > 0$. Let $e_0 \in \mathbb{R}^n$ denote an arbitrary unit vector; that is, a vector with one entry equal to one and all others equal to zero. Define $x^0 = e_0/\sqrt{2} \in \mathbb{R}^n$, $u^0 = x^0/\alpha \in \mathbb{R}^n$, and $\tilde{x} = \tilde{u} = 0 \in \mathbb{R}^n$. Then, the initial condition holds: $\alpha D_0(\tilde{x}, \tilde{u}) = \|x^0 - \tilde{x}\|^2 + \alpha^2\|u^0 - \tilde{u}\|^2 = 1$. Let

$$
f(x) = \frac{\sqrt{2}}{\alpha(K+1)}\|x\|, \qquad g(x) = 0,
$$

(so $g^*(y) = \delta_{\{0\}}(y)$). Under this setup, (DRS-gf) generates the iterates

$$
u^k = \begin{cases} \frac{1}{\alpha}x^0, & k = 0 \\ 0, & k \geq 1, \end{cases} \qquad x^{k+1} = \begin{cases} \mathbf{prox}_{\alpha f}(2x^0), & k = 0 \\ \mathbf{prox}_{\alpha f}(x^k), & k \geq 1. \end{cases} \tag{3.18}
$$

The x-iteration is simply the proximal point method starting at $2x^0$. Then, from the definition of f, we have

$$
\mathbf{prox}_{\alpha f}(y) = \begin{cases} \left(\|y\| - \frac{\sqrt{2}}{K+1}\right)\frac{y}{\|y\|}, & \text{if } \|y\| \geq \frac{\sqrt{2}}{K+1} \\ 0, & \text{otherwise.} \end{cases}
$$

So, with $x^0 = e_0/\sqrt{2}$, we show that

$$x^k = \sqrt{2}\left(1 - \frac{k}{K+1}\right)e_0, \qquad k = 1, \ldots, K$$

by induction.

(i) When $k = 1$, it follows from $\frac{\sqrt{2}}{K+1} \le \sqrt{2} = \|2x^0\|$ that

$$x^1 = \left(\|2x^0\| - \frac{\sqrt{2}}{K+1}\right)\frac{2x^0}{\|2x^0\|} = \sqrt{2}\left(1 - \frac{1}{K+1}\right)e_0.$$

(ii) Assume that the induction hypothesis is true for $k = m \le K - 1$. Then, by the induction hypothesis, we have

$$\frac{\sqrt{2}}{K+1} \le \sqrt{2}\left(1 - \frac{m}{K+1}\right) = \|x^m\|.$$

Thus,

$$x^{m+1} = \left(\|x^m\| - \frac{\sqrt{2}}{K+1}\right)\frac{x^m}{\|x^m\|} = \left(\sqrt{2}\left(1 - \frac{m}{K+1}\right) - \frac{\sqrt{2}}{K+1}\right)e_0 = \sqrt{2}\left(1 - \frac{m+1}{K+1}\right)e_0,$$

so we can conclude with the desired result.

Finally, invoking Corollary 3.3, it remains to prove that the four quantities $(I_f, I_g, \mathcal{S}_1, \mathcal{S}_2)$ in (3.8) are zero when $x = \tilde{x}$ and $u = \tilde{u}$. Observe that the points $x^1, x^2, \ldots, x^K, \overline{x}^K, \tilde{x}$ lie in a line and that $\partial f(x^k)$ is a singleton for all $k = 1, 2, \ldots, K$. It implies that $I_f = 0$. Next, we see that $I_g = 0$ since $g = 0$, $\tilde{u} = 0$, and $u^k = 0$ for all $k = 1, \ldots, K$. In addition, since $v^k = \widetilde{\nabla} f(x^k) + u^k = \widetilde{\nabla} f(x^k) = \frac{\sqrt{2}}{\alpha(K+1)}e_0$ is a constant vector for all $k = 1, \ldots, K$, we obtain $\mathcal{S}_2 = 0$. Finally, from

$$\frac{\alpha(K+1)}{2K}\sum_{k=1}^{K} v^k = \frac{\alpha(K+1)}{2K}\sum_{k=1}^{K}\frac{\sqrt{2}}{\alpha(K+1)}e_0 = \frac{1}{\sqrt{2}}e_0 = x^0 - \tilde{x} = \alpha(u^0 - \tilde{u}),$$

we obtain $\mathcal{S}_1 = 0$. This completes the proof. □

3.2 DRS-fg: worst-case rate and its tightness

We now proceed with the convergence analysis of (DRS-fg). Following the same approach as in the beginning of Section 3.1, we reformulate (DRS-fg) as

$$x^{k+1} = x^k - \alpha u^k - \alpha\widetilde{\nabla} f(x^{k+1}) \tag{3.19a}$$

$$u^{k+1} = u^k + \tfrac{1}{\alpha}(2x^{k+1} - x^k) - \tfrac{1}{\alpha}\widetilde{\nabla} g^*(u^{k+1}), \tag{3.19b}$$

and similarly we define

$$p^{k+1} := \widetilde{\nabla} g^*(u^{k+1}) = x^{k+1} + (x^{k+1} - x^k) + \alpha(u^k - u^{k+1}). \tag{3.20}$$

Similarly to (DRS-gf), we introduce an equality that enables us to derive the tight convergence rate of (DRS-fg).

Proposition 3.5. *Suppose f and g are CCP functions, and $\{(x^k, u^k)\}_{k\in\mathbb{N}}$ is generated by (DRS-fg) with stepsize $\alpha > 0$ and initial points (x^0, u^0). Denote $\bar{x}^K, \bar{u}^K, \bar{p}^K, p$ as in Proposition 3.1, and p^k as in (3.20). Then, for all $K \in \mathbb{N}_+$ and all $x, u \in \mathbb{R}^n$, the equality*

$$\mathcal{L}(\bar{x}^K, u) - \mathcal{L}(x, \bar{u}^K) - \frac{D_0(x,u)}{K+1} = I_f + I_g - \mathcal{S}_1 - \mathcal{S}_2 \tag{3.21}$$

holds, where $D_0(x, u)$ is defined in (2.15), v^k is defined in (3.7), $(I_f, I_g, \mathcal{S}_2)$ are defined in (3.8), and $\mathcal{S}_1$ is redefined as

$$\mathcal{S}_1 := \frac{1}{\alpha(K+1)} \left(\left\| x^0 - x - \frac{\alpha(K+1)}{2K} \sum_{k=1}^{K} v^k \right\|^2 + \alpha^2 \left\| u^0 - u + \frac{K+1}{2K} \sum_{k=1}^{K} v^k \right\|^2 \right). \tag{3.22}$$

We point out that (3.22) differs from $\mathcal{S}_1$ in Proposition 3.1 by only a sign in the second term. This similarity suggests that the proof of Proposition 3.5 closely parallels that of Proposition 3.1. Before presenting the proof, we emphasize that Proposition 3.5 also serves as the key to establishing the tight convergence rate of (DRS-fg).

Theorem 3.6 (Convergence of (DRS-fg))**.** *Suppose f and g are CCP functions, and $\{(x^k, u^k)\}_{k\in\mathbb{N}}$ is generated by (DRS-fg) with stepsize $\alpha > 0$ and initial points (x^0, u^0). Then, for all $K \in \mathbb{N}_+$, the ergodic iterates $(\bar{x}^K, \bar{u}^K)$ defined in (3.5) satisfy*

$$\mathcal{L}(\bar{x}^K, u) - \mathcal{L}(x, \bar{u}^K) \leq \frac{D_0(x,u)}{K+1} \tag{3.23}$$

for all $x \in \mathbf{dom}\, f$ and all $u \in \mathbf{dom}\, g^*$, where $D_0(x, u)$ is defined in (2.15).

Proof. It follows from (2.1) and the convexity of f and g that the two quantities I_f and I_g defined in (3.22) are nonpositive. Moreover, the two quantities $\mathcal{S}_1$ and $\mathcal{S}_2$ are nonnegative since they are sums of square terms. So, the desired conclusion (3.23) follows directly from Proposition 3.5. □

A similar convergence result for the primal–dual hybrid gradient (PDHG) method, which includes (DRS-fg) as a special case, was established in [4] for the primal–dual gap function, but using a different initial distance. However, the tightness of the convergence result in [4] was not proved.

Proposition 3.5 also provides an explicit if-and-only-if condition that guides the construction of a worst-case example. We will leverage this corollary in the proof of Theorem 3.8.

Corollary 3.7. *Let $x \neq x^0$ and $u \neq u^0$. Under the same setting as in Theorem 3.6, the inequality (3.23) holds with equality if and only if the four quantities I_f, I_g, $\mathcal{S}_1$, and $\mathcal{S}_2$ are all zero.*

Proof. Recall from the proof of Theorem 3.6 that I_f and I_g are nonpositive and $\mathcal{S}_1$ and $\mathcal{S}_2$ are nonnegative. This implies that $I_f + I_g - \mathcal{S}_1 - \mathcal{S}_2 = 0$ if and only if each term is zero. So, the desired conclusion follows directly from Proposition 3.5. □

Now, we prove Proposition 3.5.

Proof of Proposition 3.5. Denote the left-hand side of (3.21) as LHS, as in (3.10). The first step in organizing I_f and I_g can be done in the same way as in (3.11) and (3.12). Now we eliminate the

$\{p^k\}_{k=1}^K$ terms. We obtain from (3.19a) that $x^k - x^{k-1} = -\alpha u^{k-1} - \alpha\widetilde{\nabla} f(x^k)$. Substituting this into (3.20) and recalling the definition of v^k in (3.7) yields

$$p^k = x^k - \alpha(\widetilde{\nabla} f(x^k) + u^k) = x^k - \alpha v^k. \tag{3.24}$$

Note that the sign of v^k is flipped compared to (3.13). Then, eliminating p^k in I_g yields

$$I_g = g(p) - g(\bar{p}^K) - \langle u, p\rangle + \left\langle \bar{u}^K, \bar{p}^K \right\rangle + \left\langle u, \bar{x}^K \right\rangle - \alpha \left\langle u, \frac{1}{K}\sum_{k=1}^K v^k \right\rangle - \frac{1}{K}\sum_{k=1}^K \left\langle u^k, x^k - \alpha v^k \right\rangle,$$

where all terms, except the last two, appear in the LHS. Then, proceeding with a similar calculation to that used to obtain (3.15), but being careful with the signs of u, u^0, and v^k, we obtain:

$$\begin{aligned}&(I_f + I_g - \mathcal{S}_1) - \text{LHS}\\ &= -\frac{1}{K}\sum_{k=1}^K \left\langle \widetilde{\nabla} f(x^k), x^k \right\rangle - \frac{1}{K}\sum_{k=1}^K \left\langle u^k, x^k - \alpha v^k \right\rangle + \frac{1}{K}\sum_{k=1}^K \left\langle x^0 - \alpha u^0, v^k \right\rangle - \frac{\alpha(K+1)}{2K^2}\left\| \sum_{k=1}^K v^k \right\|^2.\end{aligned} \tag{3.25}$$

Then, it follows from the definition of v^k (3.7) that the first two terms on the right-hand side of (3.25) simplifies to

$$\begin{aligned}-\frac{1}{K}\sum_{k=1}^K \left\langle \widetilde{\nabla} f(x^k), x^k \right\rangle - \frac{1}{K}\sum_{k=1}^K \left\langle u^k, x^k - \alpha v^k \right\rangle &= -\frac{1}{K}\sum_{k=1}^K \left\langle \widetilde{\nabla} f(x^k) + u^k, x^k \right\rangle + \frac{1}{K}\sum_{k=1}^K \left\langle v^k, \alpha u^k \right\rangle\\ &= \frac{1}{K}\sum_{k=1}^K \left\langle v^k, -x^k + \alpha u^k \right\rangle.\end{aligned}$$

Substituting this into (3.25) gives

$$(I_f + I_g - \mathcal{S}_1) - \text{LHS} = \frac{1}{K}\sum_{k=1}^K \left\langle v^k, x^0 - \alpha u^0 - x^k + \alpha u^k \right\rangle - \frac{\alpha(K+1)}{2K^2}\left\| \sum_{k=1}^K v^k \right\|^2. \tag{3.26}$$

Now we eliminate x^k. Applying (3.19a) recursively, we obtain

$$x^k = \cdots = x^0 - \alpha\sum_{l=0}^{k-1}\left(u^l + \widetilde{\nabla} f(x^{l+1})\right) = x^0 - \alpha u^0 - \alpha\widetilde{\nabla} f(x^k) - \alpha\sum_{l=1}^{k-1} v^l.$$

Substituting it into (3.26) and using (3.7), and proceeding with the same calculation as in (3.17), we obtain

$$(I_f + I_g - \mathcal{S}_1) - \text{LHS} = \frac{\alpha}{K}\sum_{k=1}^K \left\langle v^k, \sum_{l=1}^k v^l \right\rangle - \frac{\alpha(K+1)}{2K^2}\left\| \sum_{k=1}^K v^k \right\|^2 = \mathcal{S}_2.$$

Therefore, $I_f + I_g - \mathcal{S}_1 - \mathcal{S}_2 = \text{LHS}$, which is our desired conclusion. □

The tightness of (3.23) is now verified using a worst-case example motivated by Corollary 3.7.

Theorem 3.8 (Worst-case example for (DRS-fg)). *Under the same setting as in Theorem 3.6, for any $K \in \mathbb{N}_+$ and any $\alpha > 0$, there exist CCP functions f and g and points $x^0, u^0, \tilde{x}, \tilde{u} \in \mathbb{R}^n$ such that $\alpha D_0(\tilde{x}, \tilde{u}) = 1$ where $D_0(\tilde{x}, \tilde{u})$ is defined in (2.15) and*

$$\mathcal{L}(\bar{x}^K, \tilde{u}) - \mathcal{L}(\tilde{x}, \bar{u}^K) = \frac{D_0(\tilde{x}, \tilde{u})}{K+1}.$$

Proof. Fix $K \in \mathbb{N}_+$ and $\alpha > 0$. Let $e_0 \in \mathbb{R}^n$ denote an arbitrary unit vector; that is, a vector with one entry equal to one and all others equal to zero. Define $x^0 = e_0/\sqrt{2} \in \mathbb{R}^n$, $u^0 = -x^0/\alpha \in \mathbb{R}^n$, and $\tilde{x} = \tilde{u} = 0 \in \mathbb{R}^n$. Then, the initial condition holds: $\alpha D_0(\tilde{x}, \tilde{u}) = \|x^0 - \tilde{x}\|^2 + \alpha^2 \|u^0 - \tilde{u}\|^2 = 1$. Let

$$f(x) = \frac{\sqrt{2}}{\alpha(K+1)} \|x\|, \qquad g(x) = 0,$$

(so $g^*(y) = \delta_{\{0\}}(y)$). Under this setup, (DRS-fg) generates the iterates

$$u^k = \begin{cases} -\frac{1}{\alpha} x^0, & k = 0 \\ 0, & k \geq 1, \end{cases} \qquad x^{k+1} = \begin{cases} \mathbf{prox}_{\alpha f}(2x^0), & k = 0 \\ \mathbf{prox}_{\alpha f}(x^k), & k \geq 1. \end{cases}$$

Comparison with (3.18) reveals that (DRS-fg) generates the same sequence $\{(x^k, u^k)\}$ despite a different initial point u^0. Since the x-iterates are exactly the same as in the proof of Theorem 3.4, the remainder of the proof readily extends from that of Theorem 3.4. □

As discussed earlier, (DRS-fg) and (DRS-gf) are not equivalent, in the sense that they generally produce different sequences of iterates. However, as shown in Theorem 3.4 and Theorem 3.8, their tight convergence rates are identical (under the general convex setting). Moreover, the corresponding worst-case examples are nearly the same, differing only in the sign of the initial point u^0.

4 Convergence analysis of two variants of DYS

Section 3 derives an ergodic $D/(K+1)$ rate of convergence for both variants of DRS. Yet, the known ergodic rate for both (DYS-gf) and (DYS-fg) is $\mathcal{O}(1/K)$ [20, 29, 31], of which the tightness is not addressed. So in this section, we investigate the convergence rates of both DYS variants. Interestingly, (DYS-gf) (originally proposed in [14]) has a slower rate than its special case (DRS-gf), whereas the swapped version (DYS-fg) restores the $D/(K+1)$ rate as in (DRS-fg).

Again, our analysis uses the primal–dual gap function (2.10), which, from the definition of p and $\bar{p}^K$ in (3.3), can be reformulated as

$$\begin{aligned} \mathcal{L}(\bar{x}^K, u) - \mathcal{L}(x, \bar{u}^K) = f(\bar{x}^K) + h(\bar{x}^K) + \langle u, \bar{x}^K \rangle - \langle u, p \rangle + g(p) \\ - \big(f(x) + h(x) + \langle \bar{u}^K, x \rangle - \langle \bar{u}^K, \bar{p}^K \rangle + g(\bar{p}^K)\big). \end{aligned} \tag{4.1}$$

4.1 Analysis of DYS-gf

In parallel to Section 3.1, we reformulate (DYS-gf) as

$$u^{k+1} = u^k + \tfrac{1}{\alpha} x^k - \tfrac{1}{\alpha} \widetilde{\nabla} g^*(u^{k+1}) \tag{4.2a}$$

$$p^{k+1} := \widetilde{\nabla} g^*(u^{k+1}) = \alpha u^k + x^k - \alpha u^{k+1} \tag{4.2b}$$

$$x^{k+1} = x^k - \alpha(2u^{k+1} - u^k) - \alpha\nabla h(p^{k+1}) - \alpha\widetilde{\nabla} f(x^{k+1}) \tag{4.2c}$$
$$= p^{k+1} - \alpha u^{k+1} - \alpha\nabla h(p^{k+1}) - \alpha\widetilde{\nabla} f(x^{k+1}). \tag{4.2d}$$

Not surprisingly, this iteration reduces to (3.1) when $h = 0$.

We now prove the core equality that provides the convergence proof. Remarkably, Proposition 4.1 does not reduce to Proposition 3.1 when $h = 0$.

Proposition 4.1. *Suppose f, g, and h are CCP functions and h is L-smooth (with $L > 0$). Suppose also that $\{(x^k, u^k)\}_{k\in\mathbb{N}}$ is generated by (DYS-gf) with stepsize $\alpha = \frac{1}{L}$ and initial points (x^0, u^0). Denote p^k as in (3.2), $\bar{p}^K$, p as in (3.3), and $(\bar{x}^K, \bar{u}^K)$ as in (3.5). Then, for all $K \in \mathbb{N}_+$ and all $x, u \in \mathbb{R}^n$, the following equality holds*

$$\mathcal{L}(\bar{x}^K, u) - \mathcal{L}(x, \bar{u}^K) - \frac{D_0(x,u)}{K} = I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1 - \mathcal{S}_2, \tag{4.3}$$

where $D_0(x, u)$ is defined in (2.15),

$$v^k := \widetilde{\nabla} f(x^k) + u^k + \nabla h(p^k), \tag{4.4}$$

(I_f, I_g) are defined in (3.8), and

$$\begin{aligned}
I_h &:= \frac{1}{K}\sum_{k=1}^{K}\left(h(\bar{x}^K) - h(p^k) + \left\langle \nabla h(\bar{x}^K), p^k - \bar{x}^K\right\rangle + \frac{\alpha}{2}\left\|\nabla h(\bar{x}^K) - \nabla h(p^k)\right\|^2\right)\\
&\quad + \frac{1}{K}\sum_{k=1}^{K}\left(h(p^k) - h(x) + \left\langle \nabla h(p^k), x - p^k\right\rangle + \frac{\alpha}{2}\left\|\nabla h(x) - \nabla h(p^k)\right\|^2\right)\\
\mathcal{S}_h &:= \frac{\alpha}{2}\left\|\nabla h(\bar{x}^K) + \frac{1}{K}\sum_{k=1}^{K}\left(\widetilde{\nabla} f(x^k) + u^k\right)\right\|^2 + \frac{\alpha}{2}\left\|\nabla h(x) - \frac{1}{K}\sum_{k=1}^{K}\nabla h(p^k)\right\|^2\\
\mathcal{S}_1 &:= \frac{1}{\alpha K}\left(\left\|x^0 - x - \frac{\alpha}{2}\sum_{k=1}^{K} v^k\right\|^2 + \alpha^2\left\|u^0 - u - \frac{1}{2}\sum_{k=1}^{K} v^k\right\|^2\right)\\
\mathcal{S}_2 &:= \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left(\left\|v^k - \nabla h(p^k) - (v^l - \nabla h(p^l))\right\|^2 + \left\|\nabla h(p^k) - \nabla h(p^l)\right\|^2\right).
\end{aligned}$$

In Proposition 4.1, the stepsize is set to $\alpha = \frac{1}{L}$ for two main reasons. First, this choice simplifies presentation of the analysis. The quantities $(I_h, \mathcal{S}_h, \mathcal{S}_1, \mathcal{S}_2)$ are already intricate owing to the presence of the smooth term h; allowing for a broader range of stepsizes would further complicate the presentation and proofs with limited additional insight. Second, unlike DRS, the stepsize α in (DYS-gf) must be upper bounded by a function of L, and the exact admissible range remains unclear. A recent paper [1] explores ways to enlarge this range, but the question is still open. Given these considerations, we believe the simplified setting $\alpha = \frac{1}{L}$ is sufficient for the purpose of this paper.

Proof of Proposition 4.1. It follows from (4.2d) and (4.4) that

$$p^k = x^k + \alpha(\widetilde{\nabla} f(x^k) + u^k + \nabla h(p^k)) = x^k + \alpha v^k. \tag{4.5}$$

Recalling (4.1), the left-hand side of (4.3) is

$$\begin{aligned}\text{LHS} &= f(\bar{x}^K) + h(\bar{x}^K) + \langle u, \bar{x}^K\rangle + g(p) - \langle u, p\rangle - \left(f(x) + h(x) + \langle \bar{u}^K, x\rangle + g(\bar{p}^K) - \langle \bar{u}^K, \bar{p}^K\rangle\right) \\ &\quad - \frac{1}{\alpha K}\left(\|x^0 - x\|^2 + \alpha^2\|u^0 - u\|^2\right).\end{aligned}$$

With the same argument as in Proposition 3.1, I_f and I_g simplify to

$$\begin{aligned}I_f &= f(\bar{x}^K) - f(x) + \left\langle x, \frac{1}{K}\sum_{k=1}^K \widetilde{\nabla} f(x^k)\right\rangle - \frac{1}{K}\sum_{k=1}^K \left\langle \widetilde{\nabla} f(x^k), x^k\right\rangle \\ I_g &= g(p) - g(\bar{p}^K) - \langle u, p\rangle + \left\langle \bar{u}^K, \bar{p}^K\right\rangle + \left\langle u, \bar{x}^K\right\rangle + \alpha\left\langle u, \frac{1}{K}\sum_{k=1}^K v^k\right\rangle - \frac{1}{K}\sum_{k=1}^K\left\langle u^k, x^k + \alpha v^k\right\rangle.\end{aligned} \tag{4.6}$$

To regroup some of the terms, we define

$$\begin{aligned}\tilde{I}_h &= I_h - \frac{\alpha}{2K}\sum_{k=1}^K \left\|\nabla h(\bar{x}^K) - \nabla h(p^k)\right\|^2 - \frac{\alpha}{2K}\sum_{k=1}^K\left\|\nabla h(x) - \nabla h(p^k)\right\|^2 \\ \tilde{\mathcal{S}}_h &= \mathcal{S}_h - \frac{\alpha}{2K}\sum_{k=1}^K \left\|\nabla h(\bar{x}^K) - \nabla h(p^k)\right\|^2 - \frac{\alpha}{2K}\sum_{k=1}^K\left\|\nabla h(x) - \nabla h(p^k)\right\|^2.\end{aligned}$$

We can easily verify that $\tilde{I}_h - \tilde{\mathcal{S}}_h = I_h - \mathcal{S}_h$. Next, we simplify $\tilde{I}_h$, $\tilde{\mathcal{S}}_h$, and $\mathcal{S}_1$ one by one.

For $\tilde{I}_h$, it follows from (4.5) and $\bar{x}^K = \frac{1}{K}\sum_{k=1}^K x^k$ that

$$\frac{1}{K}\sum_{k=1}^K\left\langle \nabla h(\bar{x}^K), p^k - \bar{x}^K\right\rangle = \frac{1}{K}\sum_{k=1}^K\left\langle \nabla h(\bar{x}^K), x^k + \alpha v^k - \bar{x}^K\right\rangle = \frac{\alpha}{K}\sum_{k=1}^K\left\langle \nabla h(\bar{x}^K), v^k\right\rangle.$$

Applying (4.5), we can verify that $\tilde{I}_h$ simplifies to

$$\tilde{I}_h = h(\bar{x}^K) - h(x) + \frac{\alpha}{K}\sum_{k=1}^K\left\langle \nabla h(\bar{x}^K), v^k\right\rangle + \frac{1}{K}\sum_{k=1}^K\left\langle \nabla h(p^k), x - x^k - \alpha v^k\right\rangle, \tag{4.7}$$

where the first two terms appear in the LHS.

Similarly, we have for $\tilde{\mathcal{S}}_h$ that

$$\begin{aligned}\tilde{\mathcal{S}}_h &= -\frac{\alpha}{2K}\sum_{k=1}^K\left\|\nabla h(x) - \nabla h(p^k)\right\|^2 + \frac{\alpha}{2}\left\|\nabla h(x) - \frac{1}{K}\sum_{k=1}^K \nabla h(p^k)\right\|^2 \\ &\quad - \frac{\alpha}{2K}\sum_{k=1}^K\left\|\nabla h(\bar{x}^K) - \nabla h(p^k)\right\|^2 + \frac{\alpha}{2}\left\|\nabla h(\bar{x}^K) + \frac{1}{K}\sum_{k=1}^K\left(\widetilde{\nabla} f(x^k) + u^k\right)\right\|^2 \\ &= -\frac{\alpha}{2K}\sum_{k=1}^K\left\|\nabla h(p^k)\right\|^2 + \frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^K\nabla h(p^k)\right\|^2 \\ &\quad - \frac{\alpha}{2K}\sum_{k=1}^K\left\|\nabla h(\bar{x}^K) - \nabla h(p^k)\right\|^2 + \frac{\alpha}{2}\left\|\nabla h(\bar{x}^K) + \frac{1}{K}\sum_{k=1}^K\left(\widetilde{\nabla} f(x^k) + u^k\right)\right\|^2.\end{aligned} \tag{4.8}$$

The last two terms on the right-hand side of (4.8) can be further simplified to

$$
\begin{aligned}
& -\frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(\bar{x}^K)-\nabla h(p^k)\right\|^2+\frac{\alpha}{2}\left\|\nabla h(\bar{x}^K)+\frac{1}{K}\sum_{k=1}^{K}\left(\widetilde{\nabla} f(x^k)+u^k\right)\right\|^2 \\
&= -\frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2+\frac{\alpha}{K}\sum_{k=1}^{K}\left\langle\nabla h(\bar{x}^K),\nabla h(p^k)+\widetilde{\nabla} f(x^k)+u^k\right\rangle+\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\left(\widetilde{\nabla} f(x^k)+u^k\right)\right\|^2 \\
&= -\frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2+\frac{\alpha}{K}\sum_{k=1}^{K}\left\langle\nabla h(\bar{x}^K),v^k\right\rangle+\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\left(\widetilde{\nabla} f(x^k)+u^k\right)\right\|^2.
\end{aligned}
$$

Substituting it back to (4.8) gives

$$
\tilde{\mathcal{S}}_h=-\frac{\alpha}{K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2+\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\left(\widetilde{\nabla} f(x^k)+u^k\right)\right\|^2+\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\nabla h(p^k)\right\|^2+\frac{\alpha}{K}\sum_{k=1}^{K}\left\langle\nabla h(\bar{x}^K),v^k\right\rangle. \tag{4.9}
$$

Next, $\mathcal{S}_1$ can be simplified to

$$
\mathcal{S}_1=\frac{1}{\alpha K}\left(\|x^0-x\|^2+\alpha^2\|u^0-u\|^2\right)-\left\langle x^0+\alpha u^0-x-\alpha u,\frac{1}{K}\sum_{k=1}^{K}v^k\right\rangle+\frac{\alpha}{2K}\left\|\sum_{k=1}^{K}v^k\right\|^2. \tag{4.10}
$$

Combining (4.7), (4.9) and (4.10) yields

$$
\begin{aligned}
&(I_f+I_g+I_h-\mathcal{S}_h-\mathcal{S}_1)-\text{LHS} \\
&= (I_f+I_g+\tilde{I}_h-\tilde{\mathcal{S}}_h-\mathcal{S}_1)-\text{LHS} \\
&= \left\langle x,\frac{1}{K}\sum_{k=1}^{K}\widetilde{\nabla} f(x^k)\right\rangle-\frac{1}{K}\sum_{k=1}^{K}\left\langle\widetilde{\nabla} f(x^k),x^k\right\rangle+\alpha\left\langle u,\frac{1}{K}\sum_{k=1}^{K}v^k\right\rangle-\frac{1}{K}\sum_{k=1}^{K}\left\langle u^k,x^k+\alpha v^k\right\rangle \\
&\quad+\frac{1}{K}\sum_{k=1}^{K}\left\langle\nabla h(p^k),x-x^k-\alpha v^k\right\rangle+\frac{\alpha}{K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2-\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\left(\widetilde{\nabla} f(x^k)+u^k\right)\right\|^2 \\
&\quad-\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\nabla h(p^k)\right\|^2+\left\langle x^0+\alpha u^0-x-\alpha u,\frac{1}{K}\sum_{k=1}^{K}v^k\right\rangle-\frac{\alpha}{2K}\left\|\sum_{k=1}^{K}v^k\right\|^2+\langle\bar{u}^K,x\rangle \\
&= -\frac{1}{K}\sum_{k=1}^{K}\left\langle\widetilde{\nabla} f(x^k),x^k\right\rangle-\frac{1}{K}\sum_{k=1}^{K}\left\langle u^k+\nabla h(p^k),x^k+\alpha v^k\right\rangle+\frac{\alpha}{K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2-\frac{\alpha}{2K}\left\|\sum_{k=1}^{K}v^k\right\|^2 \\
&\quad-\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\left(\widetilde{\nabla} f(x^k)+u^k\right)\right\|^2-\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\nabla h(p^k)\right\|^2+\left\langle x^0+\alpha u^0,\frac{1}{K}\sum_{k=1}^{K}v^k\right\rangle.
\end{aligned} \tag{4.11}
$$

Note that in the last equality, inner product terms with x and u are canceled out by using (4.4). Again, using (4.4), the first two terms on the right-hand side of (4.11) become

$$
-\frac{1}{K}\sum_{k=1}^{K}\left\langle\widetilde{\nabla} f(x^k),x^k\right\rangle-\frac{1}{K}\sum_{k=1}^{K}\left\langle u^k+\nabla h(p^k),x^k+\alpha v^k\right\rangle
$$

$$= -\frac{1}{K}\sum_{k=1}^{K}\left\langle \widetilde{\nabla} f(x^k) + u^k + \nabla h(p^k), x^k\right\rangle - \frac{\alpha}{K}\sum_{k=1}^{K}\left\langle v^k, u^k + \nabla h(p^k)\right\rangle$$

$$= \frac{1}{K}\sum_{k=1}^{K}\left\langle v^k, -x^k - \alpha\left(u^k + \nabla h(p^k)\right)\right\rangle. \tag{4.12}$$

Substituting it back to (4.11) and reorganizing, we obtain

$$\begin{aligned}&(I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1) - \text{LHS}\\ &= \frac{1}{K}\sum_{k=1}^{K}\left\langle v^k, x^0 + \alpha u^0 - x^k - \alpha\left(u^k + \nabla h(p^k)\right)\right\rangle + \frac{\alpha}{K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2\\ &\quad - \frac{\alpha}{2K^2}\left\|\sum_{k=1}^{K}\left(v^k - \nabla h(p^k)\right)\right\|^2 - \frac{\alpha}{2K^2}\left\|\sum_{k=1}^{K}\nabla h(p^k)\right\|^2 - \frac{\alpha}{2K}\left\|\sum_{k=1}^{K}v^k\right\|^2.\end{aligned} \tag{4.13}$$

Now, we eliminate x^k in the first term on the right-hand side of (4.13). Applying (4.2c) and (4.4) recursively gives

$$x^k = x^{k-1} - \alpha\left(u^k - u^{k-1}\right) - \alpha v^k = \cdots = x^0 + \alpha u^0 - \alpha u^k - \alpha\sum_{l=1}^{k}v^l.$$

Substituting it back to (4.13) gives

$$\begin{aligned}&(I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1) - \text{LHS}\\ &= \frac{\alpha}{K}\sum_{k=1}^{K}\sum_{l=1}^{k}\left\langle v^k, v^l\right\rangle - \frac{\alpha}{K}\sum_{k=1}^{K}\left\langle v^k, \nabla h(p^k)\right\rangle - \frac{\alpha}{2K^2}\left\|\sum_{k=1}^{K}\left(v^k - \nabla h(p^k)\right)\right\|^2 - \frac{\alpha}{2K}\left\|\sum_{k=1}^{K}v^k\right\|^2\\ &\quad + \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2 + \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2 - \frac{\alpha}{2K^2}\left\|\sum_{k=1}^{K}\nabla h(p^k)\right\|^2.\end{aligned} \tag{4.14}$$

Finally, observe that for any $\{a^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$, we have

$$\sum_{k=1}^{K}\left\|a^k\right\|^2 - \frac{1}{K}\left\|\sum_{k=1}^{K}a^k\right\|^2 = \frac{1}{K}\sum_{k=1}^{K}(K-1)\left\|a^k\right\|^2 - \frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{k-1}2\left\langle a^k, a^l\right\rangle = \frac{1}{K}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\|a^k - a^l\right\|^2, \tag{4.15}$$

and thus

$$\frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2 - \frac{\alpha}{2K^2}\left\|\sum_{k=1}^{K}\nabla h(p^k)\right\|^2 = \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\|\nabla h(p^k) - \nabla h(p^l)\right\|^2.$$

Next, with

$$\frac{\alpha}{K}\sum_{k=1}^{K}\sum_{l=1}^{k}\left\langle v^k, v^l\right\rangle - \frac{\alpha}{2K}\left\|\sum_{k=1}^{K}v^k\right\|^2 = \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|v^k\right\|^2,$$

the first five terms on the right-hand side of (4.14) become

$$\begin{aligned}
&\frac{\alpha}{K}\sum_{k=1}^{K}\sum_{l=1}^{k}\left\langle v^k, v^l\right\rangle - \frac{\alpha}{K}\sum_{k=1}^{K}\left\langle v^k, \nabla h(p^k)\right\rangle - \frac{\alpha}{2K^2}\left\|\sum_{k=1}^{K}\left(v^k - \nabla h(p^k)\right)\right\|^2 - \frac{\alpha}{2K}\left\|\sum_{k=1}^{K} v^k\right\|^2 \\
&\quad + \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2 \\
&= \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|v^k\right\|^2 - \frac{\alpha}{K}\sum_{k=1}^{K}\left\langle v^k, \nabla h(p^k)\right\rangle + \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(p^k)\right\|^2 - \frac{\alpha}{2K^2}\left\|\sum_{k=1}^{K}\left(v^k - \nabla h(p^k)\right)\right\|^2 \\
&= \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|v^k - \nabla h(p^k)\right\|^2 - \frac{\alpha}{2K^2}\left\|\sum_{k=1}^{K}\left(v^k - \nabla h(p^k)\right)\right\|^2 \\
&= \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\|v^k - \nabla h(p^k) - (v^l - \nabla h(p^l))\right\|^2,
\end{aligned}$$

where the last equation follows from (4.15). Finally, combining with (4.14) yields

$$\begin{aligned}
&(I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1) - \mathrm{LHS} \\
&= \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left(\left\|v^k - \nabla h(p^k) - (v^l - \nabla h(p^l))\right\|^2 + \left\|\nabla h(p^k) - \nabla h(p^l)\right\|^2\right) = \mathcal{S}_2.
\end{aligned}$$

Therefore, LHS $= I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1 - \mathcal{S}_2$, and we conclude the desired result. □

An ergodic D/K rate of convergence for (DYS-gf) follows immediately from Proposition 4.1.

Theorem 4.2 (Convergence of (DYS-gf))**.** *Suppose f, g, and h are CCP functions and h is L-smooth (with $L > 0$). Suppose $\{(x^k, u^k)\}_{k\in\mathbb{N}}$ is generated by (DYS-gf) with stepsize $\alpha = \frac{1}{L}$ and initial points (x^0, u^0). Then, for all $K \in \mathbb{N}_+$, the ergodic iterates $(\overline{x}^K, \overline{u}^K)$ defined in (3.5) satisfy*

$$\mathcal{L}(\overline{x}^K, u) - \mathcal{L}(x, \overline{u}^K) \leq \frac{D_0(x,u)}{K} \tag{4.16}$$

for all $x \in \mathbf{dom}\, f$ and all $u \in \mathbf{dom}\, g^*$, where $D_0(x,u)$ is defined in (2.15). Moreover, (4.16) holds with equality if and only if $I_f, I_g, I_h, \mathcal{S}_h, \mathcal{S}_1$ and $\mathcal{S}_2$ are all zero.

Proof. I_f and I_g are nonpositive since f and g are convex. It then follows from the convexity and L-smoothness of h that I_h is nonpositive. Moreover, $\mathcal{S}_h$, $\mathcal{S}_1$ and $\mathcal{S}_2$ are nonnegative as they are sum of squares. Hence, we conclude (4.16) from Proposition 4.1. The second conclusion is an immediate consequence of Proposition 4.1 and the facts $I_f, I_g, I_h \leq 0$ and $\mathcal{S}_h, \mathcal{S}_1, \mathcal{S}_2 \geq 0$. □

The presented rate for (DYS-gf) is slower than that for its special case (DRS-gf) (see Theorem 3.2). In the next subsection, we showcase a simple example for which (DYS-gf) is slower than $D/(K+1)$ and examine how the smooth term h slows down convergence.

4.2 DYS-gf fails to achieve a $D/(K+1)$ rate

One may notice the difference in the rate (4.16) for (DYS-gf) and that (3.9) for (DRS-gf). It is natural to ask whether (DYS-gf) can attain the $D/(K+1)$ rate as its special case (DRS-gf). This section provides a negative answer by presenting a simple example in which (DYS-gf) converges more slowerly. Although the example does not match the exact D/K rate established in Theorem 4.2, it effectively demonstrates that the rate of (DYS-gf) can indeed be worse than $D/(K+1)$.

Theorem 4.3 (Bad example for (DYS-gf))**.** *Under the same setting as in Theorem 4.2, for any $K \in \mathbb{N}_+$ and any $\alpha > 0$, there exist CCP functions f and g, $\frac{1}{\alpha}$-smooth convex function h, and points $x^0, u^0, \tilde{x}, \tilde{u} \in \mathbb{R}^n$ such that $\alpha D_0(\tilde{x}, \tilde{u}) = 1$ where $D_0(\tilde{x}, \tilde{u})$ is defined in (2.15) and*

$$\mathcal{L}(\overline{x}^K, \tilde{u}) - \mathcal{L}(\tilde{x}, \overline{u}^K) > \frac{D_0(\tilde{x}, \tilde{u})}{K+1}.$$

Proof. Fix $K \in \mathbb{N}_+$ and $\alpha > 0$. Again, define $x^0 = e_0/\sqrt{2}$, $u^0 = x^0/\alpha$, and $\tilde{x} = \tilde{u} = 0$. (Recall $e_0 \in \mathbb{R}^n$ denotes an arbitrary unit vector, so the example holds for any dimension n.) It is straightforward to check that $\alpha D_0(\tilde{x}, \tilde{u}) = \|x^0 - \tilde{x}\|^2 + \alpha^2 \|u^0 - \tilde{u}\|^2 = 1$. We also denote

$$\eta_K := \frac{\sqrt{2}(K-1)}{K^2}.$$

Consider

$$f(x) = \frac{(K-1)\eta_K}{\alpha}\|x\|, \qquad g^*(y) = \eta_K \|y\|, \qquad h(x) = \frac{1}{2\alpha}\|x\|^2.$$

Thus, we have

$$\mathbf{prox}_{\alpha f}(x) = \begin{cases} (\|x\| - (K-1)\eta_K)\frac{x}{\|x\|}, & \text{if } \|x\| \geq (K-1)\eta_K \\ 0, & \text{otherwise} \end{cases}$$
$$\mathbf{prox}_{\alpha^{-1} g^*}(y) = \begin{cases} \left(\|y\| - \frac{\eta_K}{\alpha}\right)\frac{y}{\|y\|}, & \text{if } \|y\| \geq \frac{\eta_K}{\alpha} \\ 0, & \text{otherwise.} \end{cases}$$

Then, we show by induction that

$$x^k = \begin{cases} -(\sqrt{2} - K\eta_K)e_0, & \text{if } k = 1 \\ 0, & \text{if } k = 2, 3, \ldots, K, \end{cases} \qquad u^k = \begin{cases} \frac{1}{\alpha}(\sqrt{2} - \eta_K)e_0, & \text{if } k = 1 \\ \frac{1}{\alpha}(K-k)\eta_K e_0, & \text{if } k = 2, 3, \ldots, K. \end{cases} \tag{4.17}$$

(i) When $k = 1$, recalling the definition (DYS-gf) and (4.2b), it is straightforward to verify that

$$\begin{aligned} u^1 &= \mathbf{prox}_{\alpha^{-1} g^*}(u^0 + \tfrac{1}{\alpha}x^0) = \mathbf{prox}_{\alpha^{-1} g^*}(\tfrac{\sqrt{2}}{\alpha}e_0) = \tfrac{1}{\alpha}\big(\sqrt{2} - \eta_K\big)e_0 \\ p^1 &= \alpha(u^0 + \tfrac{1}{\alpha}x^0) - \alpha u^1 = \eta_K e_0 \\ x^1 &= \mathbf{prox}_{\alpha f}(p^1 - \alpha \nabla h(p^1) - \alpha u^1) = \mathbf{prox}_{\alpha f}(-(\sqrt{2} - \eta_K)e_0) = -(\sqrt{2} - K\eta_K)e_0, \end{aligned}$$

where we used the fact $\nabla h(x) = \frac{1}{\alpha}x$ for all $x \in \mathbb{R}^n$, $-\frac{-\alpha u^1}{\|-\alpha u^1\|} = e_0$, and

$$\sqrt{2} - \eta_K = \left(1 - \frac{K-1}{K^2}\right)\sqrt{2} = \frac{K^2 - K + 1}{K^2}\sqrt{2} > \frac{(K-1)^2}{K^2}\sqrt{2} = (K-1)\eta_K.$$

Similarly, when $k = 2$, we have

$$u^2 = \mathbf{prox}_{\alpha^{-1}g^*}(u^1 + \tfrac{1}{\alpha}x^1) = \mathbf{prox}_{\alpha^{-1}g^*}(\tfrac{1}{\alpha}(K-1)\eta_K e_0) = \tfrac{1}{\alpha}(K-2)\eta_K e_0$$
$$p^2 = \alpha(u^1 + \tfrac{1}{\alpha}x^1) - \alpha u^2 = \eta_K e_0$$
$$x^2 = \mathbf{prox}_{\alpha f}(p^2 - \alpha\nabla h(p^2) - \alpha u^2) = \mathbf{prox}_{\alpha f}(-\alpha u^2) = 0,$$

where the last equality follows from the fact $\|-\alpha u^2\| = (K-2)\eta_K < (K-1)\eta_K$.

(ii) Assume that the induction hypothesis is true for $k = m \le K-1$. Then, by the induction hypothesis, we have

$$\|u^m + \tfrac{1}{\alpha}x^m\| \ge \tfrac{\eta_K}{\alpha}, \qquad \|-\alpha u^m\| \le (K-1)\eta_K.$$

Thus, we have $x^{m+1} = 0$ and

$$u^{m+1} = \mathbf{prox}_{\alpha^{-1}g^*}(u^m + \tfrac{1}{\alpha}x^m) = \tfrac{1}{\alpha}(K-m-1)\eta_K e_0.$$

So the expression (4.17) holds for $k = m+1 \le K$.

From (4.17), the ergodic sequence can be written as

$$\begin{aligned} \bar{x}^K &= \frac{1}{K}x^1 = -\frac{1}{K}\left(\sqrt{2} - K\eta_K\right)e_0 = -\frac{1}{K}\left(1 - \frac{K-1}{K}\right)\sqrt{2}e_0 = -\frac{\sqrt{2}}{K^2}e_0, \\ \bar{u}^K &= \frac{1}{\alpha K}\left(\left(\sqrt{2} - \eta_K\right) + \sum_{k=2}^{K}\eta_K(K-k)\right)e_0 = \frac{\sqrt{2}(K^2-2K+3)}{2\alpha K^2}e_0, \end{aligned} \tag{4.18}$$

where the last equality follows from

$$\begin{aligned} \sqrt{2} - \eta_K + \sum_{k=2}^{K}(K-k)\eta_K &= \sqrt{2} + \left(-1 + \sum_{k=1}^{K-2}k\right)\eta_K = \sqrt{2} + \left(-1 + \frac{(K-1)(K-2)}{2}\right)\frac{K-1}{K^2}\sqrt{2} \\ &= \frac{(K-3)(K-1)}{2K}\sqrt{2} + \sqrt{2} = \frac{\sqrt{2}(K^2-2K+3)}{2K}. \end{aligned}$$

Finally, substituting (4.18) and $\tilde{x} = \tilde{u} = 0$ into the performance metric (2.10) yields

$$\begin{aligned} \mathcal{L}(\bar{x}^K, \tilde{u}) - \mathcal{L}(\tilde{x}, \bar{u}^K) &= f(\bar{x}^K) + h(\bar{x}^K) + \langle \tilde{u}, \bar{x}^K\rangle - g^*(\tilde{u}) - (f(\tilde{x}) + h(\tilde{x}) + \langle \bar{u}^K, \tilde{x}\rangle - g^*(\bar{u}^K)) \\ &= f(\bar{x}^K) + h(\bar{x}^K) + g^*(\bar{u}^K) \\ &= \frac{(K-1)\eta_K}{\alpha}\left\|-\frac{1}{K^2}\sqrt{2}e_0\right\| + \frac{1}{2\alpha}\left\|-\frac{\sqrt{2}}{K^2}e_0\right\|^2 + \eta_K\left\|\frac{\sqrt{2}}{2\alpha K^2}\left(K^2-2K+3\right)e_0\right\| \\ &= \frac{1}{\alpha}\frac{(K-1)^2}{K^2}\sqrt{2}\left(\frac{\sqrt{2}}{K^2}\right) + \frac{1}{2\alpha}\left(\frac{\sqrt{2}}{K^2}\right)^2 + \frac{K-1}{K^2}\frac{1}{\alpha K^2}\left(K^2-2K+3\right) \\ &= \frac{1}{\alpha K^4}\left(2(K^2-2K+1) + 1 + (K-1)\left(K^2-2K+3\right)\right) \\ &= \frac{1}{\alpha K^4}\left(2K^2 - 4K + 2 + 1 + K^3 - 3K^2 + 5K - 3\right) = \frac{K^2-K+1}{\alpha K^3}. \end{aligned}$$

Finally, the desired result follows from

$$\frac{K^2-K+1}{K^3} > \frac{1}{K+1} \qquad \Longleftrightarrow \qquad K^3 + 1 = (K+1)(K^2-K+1) > K^3. \qquad \square$$

Note that the rate $\frac{K^2-K+1}{\alpha K^3}$ equals $\frac{1}{\alpha K}$ when $K=1$; that is, the bound in Theorem 4.2 is tight at $K=1$. We conjecture that the rate in Theorem 4.2 is tight in general, but leave a formal proof as an open question for future work.

4.3 DYS-fg: restoring the $D/(K+1)$ rate by swapping f and g

We proceed to analyze (DYS-fg). As before, we reformulate (DYS-fg) as

$$x^k = x^{k-1} - \alpha(u^{k-1} + \nabla h(x^{k-1})) - \alpha\widetilde{\nabla} f(x^k) \tag{4.19a}$$

$$p^k = x^k + (x^k - x^{k-1}) + \alpha((\nabla h(x^{k-1}) + u^{k-1}) - (\nabla h(x^k) + u^k)) \in \partial g^*(u^k). \tag{4.19b}$$

We now establish the tight convergence rate of (DYS-fg). As in the case of (DRS-fg), we prove an equality that immediately yields the convergence of (DYS-fg). Although Proposition 4.4 reduces to Proposition 3.5 when $h=0$, its proof is more involved because it must account for the additional smooth function h.

Proposition 4.4. *Suppose f, g, and h are CCP functions and h is L-smooth (with $L>0$). Suppose also that $\{(x^k,u^k)\}_{k\in\mathbb{N}}$ is generated by (DYS-fg) with stepsize $\alpha=\frac{1}{L}$ and initial points (x^0,u^0). Denote $\overline{x}^K, \overline{u}^K, \overline{p}^K, p$ as in Proposition 4.1, and p^k as in (4.19b). Then, for all $K\in\mathbb{N}_+$ and all $x,u\in\mathbb{R}^n$, the following equality holds*

$$\mathcal{L}(\overline{x}^K,u) - \mathcal{L}(x,\overline{u}^K) - \frac{D_0(x,u)}{K+1} = I_f + I_g + I_h - \mathcal{S}_1 - \mathcal{S}_2 - \mathcal{S}_h, \tag{4.20}$$

where $D_0(x,u)$ is defined in (2.15),

$$v^k = \widetilde{\nabla} f(x^k) + u^k + \nabla h(x^k), \tag{4.21}$$

(I_f, I_g) are defined in Proposition 4.1, and

$$\begin{aligned}
I_h :=\ & \frac{1}{K}\sum_{k=1}^{K}\left(h(\overline{x}^K) - h(x^k) + \left\langle \nabla h(\overline{x}^K), x^k - \overline{x}^K\right\rangle + \frac{\alpha}{2}\left\|\nabla h(x^k) - \nabla h(\overline{x}^K)\right\|^2\right)\\
&+ \frac{K-1}{K(K+1)}\sum_{k=1}^{K}\left(h(x^k) - h(x) + \left\langle \nabla h(x^k), x - x^k\right\rangle + \frac{\alpha}{2}\left\|\nabla h(x^k) - \nabla h(x)\right\|^2\right)\\
&+ \frac{2}{K+1}\left(h(x^0) - h(x) + \langle \nabla h(x^0), x - x^0\rangle + \frac{\alpha}{2}\left\|\nabla h(x^0) - \nabla h(x)\right\|^2\right)\\
&+ \frac{2}{K(K+1)}\sum_{k=1}^{K}\left(h(x^k) - h(x^0) + \left\langle \nabla h(x^k), x^0 - x^k\right\rangle + \frac{\alpha}{2}\left\|\nabla h(x^k) - \nabla h(x^0)\right\|^2\right)
\end{aligned}$$

$$\mathcal{S}_h := \frac{\alpha}{2}\left\|\nabla h(\overline{x}^K) - \frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2 + \frac{\alpha(K-1)}{2(K+1)}\left\|\nabla h(x) - \frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2$$

$$\begin{aligned}
\mathcal{S}_1 := \frac{1}{K+1}\Bigg(&\frac{1}{\alpha}\left\|x^0 - x - \frac{\alpha(K+1)}{2K}\sum_{k=1}^{K}v^k - \alpha\left(\nabla h(x^0) - \frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right)\right\|^2\\
&+ \alpha\left\|u^0 - u + \frac{K+1}{2K}\sum_{k=1}^{K}\left(\widetilde{\nabla} f(x^k) + u^k + \nabla h(x^k)\right)\right\|^2 + \frac{\alpha}{K+1}\left\|\nabla h(x^0) - \nabla h(x)\right\|^2\Bigg)
\end{aligned}$$

$$\mathcal{S}_2 := \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left(\left\|v^k - \nabla h(x^k) - (v^l - \nabla h(x^l))\right\|^2 + \left\|\nabla h(x^k) - \nabla h(x^l)\right\|^2\right).$$

Proof. We repeat some arguments done in Proposition 3.5. First, we verify that p^k in (4.19b) can still be written as in (3.24). From (4.19a) we obtain $x^k - x^{k-1} = -\alpha(u^{k-1} + \nabla h(x^{k-1})) - \alpha\widetilde{\nabla} f(x^k)$. Substituting it into (4.19b) yields

$$p^k = x^k - \alpha(\widetilde{\nabla} f(x^k) + u^k + \nabla h(x^k)) = x^k - \alpha v^k.$$

With the same argument of Proposition 3.5, I_f and I_g simplify to

$$I_f = f(\overline{x}^K) - f(x) + \left\langle x, \frac{1}{K}\sum_{k=1}^{K}\widetilde{\nabla} f(x^k)\right\rangle - \frac{1}{K}\sum_{k=1}^{K}\left\langle \widetilde{\nabla} f(x^k), x^k\right\rangle$$

$$I_g = g(p) - g(\overline{p}^K) - \langle u, p\rangle + \langle \overline{u}^K, \overline{p}^K\rangle + \langle u, \overline{x}^K\rangle - \alpha\left\langle u, \frac{1}{K}\sum_{k=1}^{K} v^k\right\rangle - \frac{1}{K}\sum_{k=1}^{K}\left\langle u^k, x^k - \alpha v^k\right\rangle,$$

which are restated here for later reference.

Now, to regroup some of the terms, we define

$$\begin{aligned}
\tilde{I}_h &= I_h - \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(\overline{x}^K) - \nabla h(x^k)\right\|^2 - \frac{\alpha(K-1)}{2K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x) - \nabla h(x^k)\right\|^2 \\
&\quad - \frac{\alpha}{K+1}\left\|\nabla h(x^0) - \nabla h(x)\right\|^2 - \frac{\alpha}{K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x^0) - \nabla h(x^k)\right\|^2 \\
\tilde{\mathcal{S}}_h &= \mathcal{S}_h - \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(\overline{x}^K) - \nabla h(x^k)\right\|^2 - \frac{\alpha(K-1)}{2K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x) - \nabla h(x^k)\right\|^2 \\
\tilde{\mathcal{S}}_1 &= \mathcal{S}_1 - \frac{\alpha}{K+1}\left\|\nabla h(x^0) - \nabla h(x)\right\|^2 - \frac{\alpha}{K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x^0) - \nabla h(x^k)\right\|^2.
\end{aligned}$$

Moreover, the terms $h(x^k)$, $h(x^0)$, and $\langle \nabla h(x^k), x^k - \overline{x}^k\rangle$ cancel out, and thus $\tilde{I}_h$ simplifies to

$$\begin{aligned}
\tilde{I}_h &= h(\overline{x}^K) - h(x) + \frac{2}{K(K+1)}\left\langle x^0, \sum_{k=1}^{K}\nabla h(x^k)\right\rangle - \frac{1}{K}\sum_{k=1}^{K}\left\langle \nabla h(x^k), x^k\right\rangle \\
&\quad + \frac{K-1}{K(K+1)}\sum_{k=1}^{K}\left\langle \nabla h(x^k), x\right\rangle + \frac{2}{K+1}\langle \nabla h(x^0), x - x^0\rangle,
\end{aligned}$$

where the first two terms appear in the left-hand side of (4.1).

For $\tilde{\mathcal{S}}_h$, straightforward calculations show that

$$\tilde{\mathcal{S}}_h = \frac{\alpha}{2}\left\|\nabla h(\overline{x}^K) - \frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2 - \frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(\overline{x}^K) - \nabla h(x^k)\right\|^2$$

$$
\begin{aligned}
&+\frac{\alpha(K-1)}{2(K+1)}\left\|\nabla h(x)-\frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2-\frac{\alpha(K-1)}{2K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x)-\nabla h(x^k)\right\|^2\\
&=-\frac{\alpha}{2K}\sum_{k=1}^{K}\left\|\nabla h(x^k)\right\|^2+\frac{\alpha}{2}\left\|\frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2\\
&+\frac{\alpha(K-1)}{2(K+1)}\left\|\nabla h(x)-\frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2-\frac{\alpha(K-1)}{2K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x)-\nabla h(x^k)\right\|^2\\
&=-\frac{\alpha}{K+1}\sum_{k=1}^{K}\left\|\nabla h(x^k)\right\|^2+\frac{\alpha K}{K+1}\left\|\frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2.
\end{aligned}
$$

The first equality follows from the definition of $\tilde{\mathcal{S}}_h$. In the second equality, we cancel out all the $\nabla h(\bar{x}^K)$ terms, and the last equality cancels out the $\nabla h(x)$ terms.

Next, we move on to $\tilde{\mathcal{S}}_1$. Observe that the sum of the first and the last term of $\tilde{\mathcal{S}}_1$ can be rewritten as:

$$
\begin{aligned}
&\frac{1}{\alpha(K+1)}\left\|x^0-x-\frac{\alpha(K+1)}{2K}\sum_{k=1}^{K}v^k-\alpha\left(\nabla h(x^0)-\frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right)\right\|^2\\
&-\frac{\alpha}{K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x^0)-\nabla h(x^k)\right\|^2\\
&=\frac{1}{\alpha(K+1)}\left\|x^0-x-\alpha\frac{K+1}{2K}\sum_{k=1}^{K}v^k\right\|^2-\frac{2}{K+1}\left\langle x^0-x,\nabla h(x^0)\right\rangle\\
&+\frac{2}{K(K+1)}\left\langle x^0-x,\sum_{k=1}^{K}\nabla h(x^k)\right\rangle+\left\langle\frac{1}{K}\sum_{k=1}^{K}v^k,\alpha\nabla h(x^0)\right\rangle-\frac{\alpha}{K^2}\left\langle\sum_{k=1}^{K}v^k,\sum_{k=1}^{K}\nabla h(x^k)\right\rangle\\
&-\frac{\alpha}{K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x^k)\right\|^2+\frac{\alpha}{K+1}\left\|\frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2\\
&=\frac{1}{\alpha(K+1)}\left\|x^0-x\right\|^2-\left\langle x^0-\alpha\nabla h(x^0)-x,\frac{1}{K}\sum_{k=1}^{K}v^k\right\rangle+\frac{\alpha(K+1)}{4K^2}\left\|\sum_{k=1}^{K}v^k\right\|^2\\
&-\frac{2}{K+1}\left\langle x^0-x,\nabla h(x^0)\right\rangle+\frac{2}{K(K+1)}\left\langle x^0-x,\sum_{k=1}^{K}\nabla h(x^k)\right\rangle-\frac{\alpha}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle v^l,\nabla h(x^k)\right\rangle\\
&-\frac{\alpha}{K(K+1)}\sum_{k=1}^{K}\left\|\nabla h(x^k)\right\|^2+\frac{\alpha}{K+1}\left\|\frac{1}{K}\sum_{k=1}^{K}\nabla h(x^k)\right\|^2.
\end{aligned}
$$

In the first equality, we expand the two squared terms and distribute the cross terms. In the second equality, we expand the first square and gather the inner product terms with $\sum_{k=1}^{K}v^k$. Similarly, the second term in $\tilde{\mathcal{S}}_1$ can be rewritten as

$$
\frac{\alpha}{K+1}\left\|u^0-u+\frac{K+1}{2K}\sum_{k=1}^{K}v^k\right\|^2
$$

$$= \frac{\alpha}{K+1} \left\|u^0 - u\right\|^2 + \alpha \left\langle u^0 - u, \frac{1}{K}\sum_{k=1}^{K} v^k \right\rangle + \frac{\alpha(K+1)}{4K^2} \left\|\sum_{k=1}^{K} v^k\right\|^2 .$$

Now, gathering the observations, we denote by LHS the left-hand side of (4.20) and obtain

$$\begin{aligned}
&(I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1) - \text{LHS} \\
&= \left\langle x, \frac{1}{K}\sum_{k=1}^{K} \widetilde{\nabla} f(x^k) \right\rangle - \frac{1}{K}\sum_{k=1}^{K} \left\langle \widetilde{\nabla} f(x^k), x^k \right\rangle - \alpha \left\langle u, \frac{1}{K}\sum_{k=1}^{K} v^k \right\rangle - \frac{1}{K}\sum_{k=1}^{K} \left\langle u^k, x^k - \alpha v^k \right\rangle \\
&\quad + \frac{2}{K(K+1)} \left\langle x^0, \sum_{k=1}^{K} \nabla h(x^k) \right\rangle + \frac{K-1}{K(K+1)} \sum_{k=1}^{K} \left\langle \nabla h(x^k), x \right\rangle - \frac{1}{K}\sum_{k=1}^{K} \left\langle \nabla h(x^k), x^k \right\rangle \\
&\quad + \frac{\alpha}{K}\sum_{k=1}^{K} \left\|\nabla h(x^k)\right\|^2 - \alpha \left\|\frac{1}{K}\sum_{k=1}^{K} \nabla h(x^k)\right\|^2 + \left\langle x^0 - \alpha\left(u^0 + \nabla h(x^0)\right), \frac{1}{K}\sum_{k=1}^{K} v^k \right\rangle \\
&\quad - \left\langle x - \alpha u, \frac{1}{K}\sum_{k=1}^{K} v^k \right\rangle - \frac{\alpha(K+1)}{2K^2} \left\|\sum_{k=1}^{K} v^k\right\|^2 - \frac{2}{K(K+1)} \left\langle x^0 - x, \sum_{k=1}^{K} \nabla h(x^k) \right\rangle \\
&\quad + \frac{\alpha}{K^2} \sum_{k=1}^{K}\sum_{l=1}^{K} \left\langle \nabla h(x^k), v^l \right\rangle + \langle x, \bar{u}^K \rangle \\
&= -\frac{1}{K}\sum_{k=1}^{K} \left\langle \widetilde{\nabla} f(x^k), x^k \right\rangle - \frac{1}{K}\sum_{k=1}^{K} \left\langle u^k, x^k - \alpha v^k \right\rangle - \frac{1}{K}\sum_{k=1}^{K} \left\langle \nabla h(x^k), x^k \right\rangle \\
&\quad + \frac{\alpha}{K}\sum_{k=1}^{K} \left\|\nabla h(x^k)\right\|^2 - \alpha \left\|\frac{1}{K}\sum_{k=1}^{K} \nabla h(x^k)\right\|^2 + \left\langle x^0 - \alpha\left(u^0 + \nabla h(x^0)\right), \frac{1}{K}\sum_{k=1}^{K} v^k \right\rangle \\
&\quad - \frac{\alpha(K+1)}{2K^2} \left\|\sum_{k=1}^{K} v^k\right\|^2 + \frac{\alpha}{K^2} \sum_{k=1}^{K}\sum_{l=1}^{K} \left\langle \nabla h(x^k), v^l \right\rangle,
\end{aligned} \tag{4.22}$$

In the second equality, the terms $\frac{2}{K+1}\langle \nabla h(x^0), x - x^0\rangle$, $\frac{2}{K(K+1)}\langle x^0, \sum_{k=1}^{K} \nabla h(x^k)\rangle$ and the inner product terms with x and u are canceled out. As a detail for the inner product terms with x, they are canceled out since

$$\begin{aligned}
0 = &\left\langle x, \frac{1}{K}\sum_{k=1}^{K} \widetilde{\nabla} f(x^k) \right\rangle + \frac{K-1}{K(K+1)} \sum_{k=1}^{K} \left\langle \nabla h(x^k), x \right\rangle - \left\langle x, \frac{1}{K}\sum_{k=1}^{K} v^k \right\rangle \\
&+ \frac{2}{K(K+1)} \left\langle x, \sum_{k=1}^{K} \nabla h(x^k) \right\rangle + \left\langle x, \bar{u}^K \right\rangle,
\end{aligned}$$

where we also use the definition of v^k (4.21). Again, it follows from (4.21) that the first three terms on the right-hand side of (4.22) is

$$-\frac{1}{K}\sum_{k=1}^{K} \left\langle \widetilde{\nabla} f(x^k), x^k \right\rangle - \frac{1}{K}\sum_{k=1}^{K} \left\langle u^k, x^k - \alpha v^k \right\rangle - \frac{1}{K}\sum_{k=1}^{K} \left\langle \nabla h(x^k), x^k \right\rangle$$

$$= -\frac{1}{K}\sum_{k=1}^{K}\left\langle \widetilde{\nabla} f(x^k) + u^k + \nabla h(x^k), x^k\right\rangle + \frac{1}{K}\sum_{k=1}^{K}\left\langle v^k, \alpha u^k\right\rangle = \frac{1}{K}\sum_{k=1}^{K}\left\langle v^k, -x^k + \alpha u^k\right\rangle. \quad (4.23)$$

Substituting (4.23) back to (4.22) gives

$$\begin{aligned}
&(I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1) - \text{LHS} \\
&= \frac{1}{K}\sum_{k=1}^{K}\left\langle v^k, x^0 - \alpha\left(u^0 + \nabla h(x^0)\right) - x^k + \alpha u^k\right\rangle + \frac{\alpha}{K}\sum_{k=1}^{K}\left\|\nabla h(x^k)\right\|^2 \\
&\quad - \frac{\alpha}{K^2}\left\|\sum_{k=1}^{K}\nabla h(x^k)\right\|^2 - \frac{\alpha(K+1)}{2K^2}\left\|\sum_{k=1}^{K} v^k\right\|^2 + \frac{\alpha}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k), v^l\right\rangle. \quad (4.24)
\end{aligned}$$

Now, we eliminate x^k. Applying (4.19a) recursively, we obtain

$$x^k = \cdots = x^0 - \alpha\sum_{l=0}^{k-1}\left(u^l + \nabla h(x^l) + \widetilde{\nabla} f(x^{l+1})\right) = x^0 - \alpha\left(u^0 + \nabla h(x^0)\right) - \alpha\widetilde{\nabla} f(x^k) - \alpha\sum_{l=1}^{k-1} v^l.$$

Substituting it back to (4.24) gives

$$\begin{aligned}
&(I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1) - \text{LHS} \\
&= \frac{\alpha}{K}\sum_{k=1}^{K}\sum_{l=1}^{k}\left\langle v^k, v^l\right\rangle - \frac{\alpha(K+1)}{2K^2}\left\|\sum_{k=1}^{K} v^k\right\|^2 + \frac{\alpha}{K}\sum_{k=1}^{K}\left\|\nabla h(x^k)\right\|^2 \\
&\quad - \frac{\alpha}{K}\sum_{k=1}^{K}\left\langle v^k, \nabla h(x^k)\right\rangle - \frac{\alpha}{K^2}\left\|\sum_{k=1}^{K}\nabla h(x^k)\right\|^2 + \frac{\alpha}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k), v^l\right\rangle. \quad (4.25)
\end{aligned}$$

Finally, recalling (3.17) we know

$$\frac{\alpha}{K}\sum_{k=1}^{K}\sum_{l=1}^{k}\left\langle v^k, v^l\right\rangle - \frac{\alpha(K+1)}{2K^2}\left\|\sum_{k=1}^{K} v^k\right\|^2 = \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\|v^k - v^l\right\|^2,$$

and observe that

$$\begin{aligned}
&\frac{\alpha}{K}\sum_{k=1}^{K}\left\|\nabla h(x^k)\right\|^2 - \frac{\alpha}{K}\sum_{k=1}^{K}\left\langle v^k, \nabla h(x^k)\right\rangle - \frac{\alpha}{K^2}\left\|\sum_{k=1}^{K}\nabla h(x^k)\right\|^2 + \frac{\alpha}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k), v^l\right\rangle \\
&= \frac{\alpha}{K^2}\left(\sum_{k=1}^{K}\sum_{l=1}^{K}\left\|\nabla h(x^k)\right\|^2 - \sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k), v^k\right\rangle - \sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k), \nabla h(x^l)\right\rangle + \sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k), v^l\right\rangle\right) \\
&= \frac{\alpha}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k), \nabla h(x^k) - v^k - \nabla h(x^l) + v^l\right\rangle \\
&= \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k), \nabla h(x^k) - v^k - \nabla h(x^l) + v^l\right\rangle
\end{aligned}$$

$$+\frac{\alpha}{2K^2}\sum_{l=1}^{K}\sum_{k=1}^{K}\left\langle \nabla h(x^l), \nabla h(x^l) - v^l - \nabla h(x^k) + v^k \right\rangle$$

$$= \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\left\langle \nabla h(x^k) - \nabla h(x^l), \nabla h(x^k) - v^k - \nabla h(x^l) + v^l \right\rangle$$

$$= \frac{\alpha}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left\langle \nabla h(x^k) - \nabla h(x^l), \nabla h(x^k) - v^k - \nabla h(x^l) + v^l \right\rangle.$$

Combining with (4.25) gives

$$\begin{aligned}
&(I_f + I_g + I_h - \mathcal{S}_h - \mathcal{S}_1) - \text{LHS} \\
&= \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left(\left\|v^k - v^l\right\|^2 + 2\left\langle \nabla h(x^k) - \nabla h(x^l), \nabla h(x^k) - v^k - \nabla h(x^l) + v^l \right\rangle\right) \\
&= \frac{\alpha}{2K^2}\sum_{k=1}^{K}\sum_{l=1}^{k-1}\left(\left\|v^k - \nabla h(x^k) - (v^l - \nabla h(x^l))\right\|^2 + \left\|\nabla h(x^k) - \nabla h(x^l)\right\|^2\right) = \mathcal{S}_2,
\end{aligned}$$

which is the desired result. □

Theorem 4.5 (Convergence of (DYS-fg))**.** *Suppose f, g, and h are CCP functions and h is L-smooth (with $L > 0$). Suppose $\{(x^k, u^k)\}_{k\in\mathbb{N}}$ is generated by (DYS-fg) with stepsize $\alpha = \frac{1}{L}$ and initial points (x^0, u^0). Then, for all $K \in \mathbb{N}_+$, the ergodic iterates $(\overline{x}^K, \overline{u}^K)$ defined in (3.5) satisfy*

$$\mathcal{L}(\overline{x}^K, u) - \mathcal{L}(x, \overline{u}^K) \leq \frac{D_0(x,u)}{K+1} \tag{4.26}$$

for all $x \in \mathbf{dom}\, f$ and all $u \in \mathbf{dom}\, g^*$, where $D_0(x,u)$ is defined in (2.15).

Proof. It follows from the convexity of f and g that I_f and I_g are nonpositive and from the convexity and the L-smoothness of h that I_h is nonpositive. Moreover, $\mathcal{S}_h$, $\mathcal{S}_1$ and $\mathcal{S}_2$ are nonnegative since they are sum of squares. Therefore, we conclude (4.26) from Proposition 4.4. □

Recall that (DYS-fg) reduces to (DRS-fg) when $h = 0$. So, (DYS-fg) cannot exhibit a faster worst-case rate than (DRS-fg). However, Theorem 4.5 presents the same upper bound for (DYS-fg) as for (DRS-fg). Combined with the tightness of our (DRS-fg) rate, we readily conclude that the $D/(K+1)$ ergodic rate in Theorem 4.5 must be tight for (DYS-fg). This argument is not evident prior to obtaining Theorem 4.5. So we present the tightness of (4.26) as a corollary.

Corollary 4.6 (Worst-case example for (DYS-fg))**.** *Under the same setting as in Theorem 4.5, for any $K \in \mathbb{N}_+$ and any $\alpha > 0$, there exist CCP functions f and g, $\frac{1}{\alpha}$-smooth convex function h, and points $x^0, u^0, \tilde{x}, \tilde{u} \in \mathbb{R}^n$ such that $\alpha D_0(\tilde{x}, \tilde{u}) = 1$ where $D_0(\tilde{x}, \tilde{u})$ is defined in (2.15) and*

$$\mathcal{L}(\overline{x}^K, \tilde{u}) - \mathcal{L}(\tilde{x}, \overline{u}^K) = \frac{D_0(\tilde{x}, \tilde{u})}{K+1}.$$

Proof. Note that Proposition 4.4 reduces to Proposition 3.5 when $h = 0$. Thus, following the same argument as in the proof of Theorem 3.4, we can show that the worst-case example for (DRS-fg) in Theorem 3.4, together with $h = 0$, also serves as a worst-case example for (DYS-fg). □

4.4 Discussion: comparison between the two variants

For (DYS-fg), a worst-case example exists with $h = 0$, and in fact, the same example serves both (DYS-fg) and its special case (DRS-fg). This naturally arises the question of whether (DYS-gf) and (DRS-gf) can also share the same worst-case example, or whether a worst-case instance for (DYS-gf) can be constructed with $h = 0$. The following proposition provides a negative answer: (DYS-gf) restores the $D/(K + 1)$ rate as long as either $g = 0$ or $h = 0$.

Proposition 4.7. *Suppose $\{(x^k, u^k)\}_{k\in\mathbb{N}}$ is generated by (DYS-gf) with stepsize $\alpha > 0$ and initial points (x^0, u^0). Denote $\overline{x}^K$, $\overline{u}^K$ as in (3.5). For any $K \in \mathbb{N}_+$, there do* not *exist CCP functions f and g, $\frac{1}{\alpha}$-smooth convex function h, and points $x^0, u^0, \tilde{x}, \tilde{u} \in \mathbb{R}^n$ such that at least one of g or h vanishes, $\alpha D_0(\tilde{x}, \tilde{u}) = 1$ where $D_0(\tilde{x}, \tilde{u})$ is defined in (2.15), and*

$$\mathcal{L}(\overline{x}^K, \tilde{u}) - \mathcal{L}(\tilde{x}, \overline{u}^K) > \frac{D_0(\tilde{x}, \tilde{u})}{K + 1}.$$

Proof. When $g = 0$, we have $g^* = \delta_{\{0\}}$ and thus $\mathbf{prox}_{\alpha^{-1}g^*}(y) = 0$ for all $y \in \mathbb{R}^n$. Then, the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by (DYS-gf) reduces to a special case of (DYS-fg) with $g = 0$. So, the desired conclusion follows from Theorem 4.5.

When $h = 0$, the sequence $\{(x^k, u^k)\}_{k\in\mathbb{N}}$ generated by (DYS-gf) reduces to (DRS-gf). Therefore we obtain the desired conclusion from Theorem 3.2. □

Proposition 4.7 reveals that, in the worst case, the presence of the smooth term h slows down the convergence of (DYS-gf). In contrast, this phenomenon does not occur in (DYS-fg). This asymmetry highlights a key distinction between the two variants of DYS: while (DYS-fg) remains robust to the inclusion of h, the performance of (DYS-gf) is more sensitive to the smooth term. Consequently, care must be taken in selecting which formulation to use, especially when the smooth term h plays a significant role in the objective.

5 Conclusion

This paper presents novel convergence analyses of the Davis–Yin splitting (DYS) algorithm and its variant obtained by swapping the roles of the two nonsmooth objective functions. While the two functions appear symmetrically in the problem formulation, this symmetry breaks at the algorithmic level. Surprisingly, we show that the swapped DYS algorithm (which we call DYS-fg) achieves a faster $D/(K+1)$ ergodic rate, compared to the standard D/K rate of the original DYS. These results are established under a unified primal–dual gap metric and illustrated via concrete examples.

In contrast, for Douglas–Rachford splitting (DRS), which arises as a special case of DYS when the smooth term vanishes, both the original and swapped versions exhibit the same convergence rates and nearly identical worst-case instances. This contrast highlights how the presence of a smooth term alters the algorithmic behavior under different update orders.

Our findings suggest that the update order is not merely a structural nuance, but one that may affect algorithmic performance. Future work may extend our analysis to broader settings, including a composite extension of (1.1) in which g is replaced by $g \circ A$ for a linear operator A, as well as to DYS and DRS for monotone inclusion problems.

Acknowledgments

The work of S. Ma was supported in part by the National Science Foundation under Grants CCF-2311275 and ECCS-2326591, and in part by the Office of Naval Research under Grant N00014-24-1-2705. The work of J.J. Suh was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, under Grant RS-2024-00410486.

References

[1] F. J. Aragón-Artacho and D. Torregrosa-Belén. A direct proof of convergence of Davis–Yin splitting algorithm allowing larger stepsizes. *Set-Valued and Variational Analysis*, 30(3):1011–1029, 2022.

[2] H. H. Bauschke and P. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Publishing Company, Incorporated, 2nd edition, 2017.

[3] D. P. Bertsekas. Incremental gradient, subgradient, and proximal methods for convex optimization: A survey. *Optimization for Machine Learning*, 2010(1-38):3, 2011.

[4] N. Bousselmi, J. M. Hendrickx, and F. Glineur. Interpolation conditions for linear operators and applications to performance estimation problems. *SIAM Journal on Optimization*, 34(3):3033–3063, 2024.

[5] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[6] A. Chambolle and T. Pock. A first-order primal–dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[7] A. Chambolle and T. Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[8] A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal–dual algorithm. *Mathematical Programming*, 159(1):253–287, 2016.

[9] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications, pages 185–212. Springer New York, 2011.

[10] L. Condat. A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2):460–479, 2013.

[11] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi. Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists. *SIAM Review*, 65(2):375–435, 2023.

[12] F. Cui, Y. Tang, and Y. Yang. An inertial three-operator splitting algorithm with applications to image inpainting. *Applied Set-Valued Analysis and Optimization*, 1:113–134, 2019.

[13] D. Davis and W. Yin. Convergence rate analysis of several splitting schemes. In R. Glowinski, S. J. Osher, and W. Yin, editors, *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 115–163. Springer, 2017.

[14] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-Valued and Variational Analysis*, 25:829–858, 2017.

[15] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82(2):421–439, 1956.

[16] Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145(1-2):451–482, 2014.

[17] J. Eckstein. *Splitting methods for monotone operators with applications to parallel optimization*. PhD thesis, Massachusetts Institute of Technology, 1989.

[18] J. Eckstein and D. Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

[19] B. He and X. Yuan. On the convergence rate of Douglas–Rachford operator splitting method. *Mathematical Programming*, 153(2):715–722, 2015.

[20] X. Jiang and L. Vandenberghe. Bregman three-operator splitting methods. *Journal of Optimization Theory and Applications*, 196(3):936–972, 2023.

[21] P. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[22] Y. Nesterov. *Lectures on Convex Optimization*. Springer, 2018.

[23] N. Parikh and S. P. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):127–239, 2014.

[24] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383–390, 1979.

[25] D. W. Peaceman and H. H. Rachford. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, 3(1):28–41, 1955.

[26] F. Pedregosa and G. Gidel. Adaptive three operator splitting. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 4085–4094, 2018.

[27] R. T. Rockafellar. *Convex Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 1970.

[28] E. K. Ryu and W. Yin. *Large-Scale Convex Optimization: Algorithms and Analyses via Monotone Operators*. Cambridge University Press, 2022.

[29] A. Salim, L. Condat, K. Mishchenko, and P. Richtárik. Dualize, split, randomize: Toward fast nonsmooth optimization algorithms. *Journal of Optimization Theory and Applications*, pages 1–29, 2022.

[30] A. B. Taylor, J. M. Hendrickx, and F. Glineur. Exact worst-case performance of first-order methods for composite convex optimization. *SIAM Journal on Optimization*, 27(3):1283–1313, 2017.

[31] M. Yan. A new primal–dual algorithm for minimizing the sum of three functions with a linear operator. *Journal of Scientific Computing*, 76(3):1698–1717, 2018.

[32] M. Yan and Y. Li. On the improved conditions for some primal–dual algorithms. *Journal of Scientific Computing*, 99(3):74, 2024.

[33] M. Yan and W. Yin. Self equivalence of the alternating direction method of multipliers. In R. Glowinski, S. J. Osher, and W. Yin, editors, *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 165–194. Springer, 2017.

[34] A. Yurtsever, A. Gu, and S. Sra. Three operator splitting with subgradients, stochastic gradients, and adaptive learning rates. In *Neural Information Processing Systems*, 2021.

[35] A. Yurtsever, B. C. Vu, and V. Cevher. Stochastic three-composite convex minimization. In *Neural Information Processing Systems*, 2016.

[36] R. Zhao and V. Cevher. Stochastic three-composite convex minimization with a linear operator. In *International Conference on Artificial Intelligence and Statistics*, 2018.

[37] R. Zhao, W. B. Haskell, and V. Y. F. Tan. An optimal algorithm for stochastic three-composite optimization. In *International Conference on Artificial Intelligence and Statistics*, 2019.

[38] C. Zong, Y. Tang, and Y. J. Cho. Convergence analysis of an inexact three-operator splitting algorithm. *Symmetry*, 10(11):563, 2018.