Convergence Analyses of Davis–Yin Splitting via Scaled Relative Graphs II: Convex Optimization Problems

Soheun Yi[a], Ernest K. Ryu[a,*]

[a]*Seoul National University, Department of mathematical sciences, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Korea*

Abstract

The prior work of [arXiv:2207.04015, 2022] used scaled relative graphs (SRG) to analyze the convergence of Davis–Yin splitting (DYS) iterations on monotone inclusion problems. In this work, we use this machinery to analyze DYS iterations on convex optimization problems and obtain state-of-the-art linear convergence rates.

1. Introduction

Consider the problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \quad f(x)+g(x)+h(x), \tag{1}$$

where $\mathcal{H}$ is a Hilbert space, f, g, and h are convex, closed, and proper functions, and h is differentiable with L-Lipschitz continuous gradients. The Davis–Yin splitting (DYS) [1] solves this problem by performing the fixed-point iteration with

$$\mathbb{T} = \mathbb{I} - \mathrm{Prox}_{\alpha g} + \mathrm{Prox}_{\alpha f}(2\mathrm{Prox}_{\alpha g} - \mathbb{I} - \alpha\nabla h\mathrm{Prox}_{\alpha g}), \tag{2}$$

where $\alpha > 0$, $\mathrm{Prox}_{\alpha f}$ and $\mathrm{Prox}_{\alpha g}$ are the proximal operators with respect to αf and αg, and $\mathbb{I}$ is the identity mapping. DYS has been used as a building

*Corresponding author

Email addresses: lsdluis@snu.ac.kr (Soheun Yi), ernestryu@snu.ac.kr (Ernest K. Ryu)

block for various algorithms for a diverse range of optimization problems [2, 3, 4, 5, 6, 7].

Much prior work has been dedicated to analyzing the convergence rate of DYS iterations [1, 8, 9, 10, 11, 12, 13]. Recently, Lee, Yi, and Ryu [14] leveraged the recently introduced scaled relative graphs (SRG) [15] to obtain tighter analyses. However, the focus of [14] was on DYS applied to general class monotone operators, rather than the narrower class of subdifferential operators.

In this paper, we use the SRG theory of [14] to analyze the linear convergence rates of DYS applied to convex optimization problems and obtain state-of-the-art rates.

1.1. Prior works

For explaining and inducing many convex optimization algorithms, splitting methods for monotone inclusion problems has been a potent tool[16, 17, 18]. Renowned examples of this methodology include forward-backward splitting (FBS) [19, 20], Douglas–Rachford splitting (DRS) [21, 22, 23], and alternating directions method of multipliers (ADMM) [24], which have been widely used in application regimes. While FBS and DRS are concerned with the sum of two monotone operators, Davis and Yin proposed a new splitting method for the sum of three monotone operators [1], thereby uniting the aforementioned methods. This has come up with a variety of applications [2, 3, 4, 5, 6, 7] and many variants, including stochastic DYS [25, 26, 27, 28, 29], inexact DYS [30], adaptive DYS [31], inertial DYS [32], and primal-dual DYS[33].

Compared to the substantial amount of these studies on DYS, there is not much literature on linear convergence analysis of the DYS iteration. One approach is to formulate SDPs that numerically find the tight contraction factors of DYS: Ryu, Taylor, Bergeling, and Giselsson [11] and Wang, Fazlyab, Chen, and Preciado [12] carried out this approach using the performance estimation problem (PEP) and integral quadratic constraint (IQC), respectively. However, this approach does not give an analytical expression of the contraction factors. There is far less literature that gives an analytical expression of the contraction factors. Two of them are respectively done by Davis and Yin [1], and Condat and Richtarik [13] via inequality analysis. On the other hand, Lee, Yi, and Ryu [14] made a different approach utilizing scaled relative graphs.

This novel tool, the scaled relative graphs (SRG) [15], renders a new approach to analyzing the behavior of multi-valued operators (in particular, including nonlinear operators) by mapping them to the extended complex plane. This theory was further studied by Huang, Ryu, and Yin [34], where they identified the SRG of normal matrices. Furthermore, Pates leveraged the Toeplitz–Hausdorff theorem to identify SRGs of linear operators [35]. This approach has also been used for analyzing nonlinear operators. To prove the tightness of the averaged coefficient of the composition of averaged operators [36] and the DYS operator [1], Huang, Ryu, and Yin performed analyses of SRGs of those operators [37]. Moreover, there is certain literature on applying SRG to control theory. SRGs have been leveraged by Chaffey, Forni, and Rodolphe to examine input-output properties of feedback systems [38, 39]. Chaffey and Sepulchre have further found its application to characterize behaviors of a given model by leveraging it as an experimental tool [40, 41, 42].

1.2. Preliminaries

Multi-valued operators.. In general, we follow notations regarding multi-valued operators as in [16, 18]. Write $\mathcal{H}$ for a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. To represent that $\mathbb{A}$ is a multi-valued operator defined on $\mathcal{H}$, write $\mathbb{A}\colon \mathcal{H} \rightrightarrows \mathcal{H}$, and define its domain as $\operatorname{dom} \mathbb{A} = \{x \in \mathcal{H} \,|\, \mathbb{A}x \neq \emptyset\}$. We say $\mathbb{A}$ is single-valued if all outputs of $\mathbb{A}$ are singletons or the empty set, and identify $\mathbb{A}$ with the function from $\operatorname{dom} \mathbb{A}$ to $\mathcal{H}$. Define the graph of an operator $\mathbb{A}$ as

$$\operatorname{graph}(\mathbb{A}) = \{(x, u) \in \mathcal{H} \times \mathcal{H} \,|\, u \in \mathbb{A}x\}.$$

We do not distinguish $\mathbb{A}$ and $\operatorname{graph}(\mathbb{A})$ for the sake of notational simplicity. For instance, it is valid to write $(x, u) \in \mathbb{A}$ to mean $u \in \mathbb{A}x$. Define the inverse of $\mathbb{A}$ as

$$\mathbb{A}^{-1} = \{(u, x) \,|\, (x, u) \in \mathbb{A}\},$$

scalar multiplication with an operator as

$$\alpha\mathbb{A} = \{(x, \alpha u) \,|\, (x, u) \in \mathbb{A}\},$$

the identity operator as

$$\mathbb{I} = \{(x, x) \,|\, x \in \mathcal{H}\},$$

and

$$\mathbb{I} + \alpha\mathbb{A} = \{(x, x + \alpha u) \mid (x, u) \in \mathbb{A}\}$$

for any $\alpha \in \mathbb{R}$. Define the resolvent of $\mathbb{A}$ with stepsize $\alpha > 0$ as

$$\mathbb{J}_{\alpha\mathbb{A}} = (\mathbb{I} + \alpha\mathbb{A})^{-1}.$$

Note that $\mathbb{J}_{\alpha\mathbb{A}}$ is a single-valued operator if $\mathbb{A}$ is monotone, or equivalently if $\langle x - y, u - v\rangle \geq 0$ for all (x, u), $(y, v) \in \mathbb{A}$. Define addition and composition of operators $\mathbb{A}\colon \mathcal{H} \rightrightarrows \mathcal{H}$ and $\mathbb{B}\colon \mathcal{H} \rightrightarrows \mathcal{H}$ as

$$\begin{aligned}\mathbb{A} + \mathbb{B} &= \{(x, u + v) \mid (x, u) \in \mathbb{A}, (x, v) \in \mathbb{B}\}, \\ \mathbb{A}\mathbb{B} &= \{(x, s) \mid \exists\, u \text{ such that } (x, u) \in \mathbb{B}, (u, s) \in \mathbb{A}\}.\end{aligned}$$

We call $\mathcal{A}$ a class of operators if it is a set of operators. For any real scalar $\alpha \in \mathbb{R}$, define

$$\alpha\mathcal{A} = \{\alpha\mathbb{A} \mid \mathbb{A} \in \mathcal{A}\}$$

and

$$\mathbb{I} + \alpha\mathcal{A} = \{\mathbb{I} + \alpha\mathbb{A} \mid \mathbb{A} \in \mathcal{A}\}.$$

Define

$$\mathcal{A}^{-1} = \{\mathbb{A}^{-1} \mid \mathbb{A} \in \mathcal{A}\}$$

and $\mathbb{J}_{\alpha\mathcal{A}} = (\mathbb{I} + \alpha\mathcal{A})^{-1}$ for $\alpha > 0$.

Subdifferential operators.. Unless otherwise stated, functions defined on $\mathcal{H}$ are extended real-valued, which means

$$f\colon \mathcal{H} \to \mathbb{R} \cup \{\pm\infty\}.$$

For a function f, we define the subdifferential operator ∂f via

$$\partial f(x) = \{g \in \mathcal{H} \mid f(y) \geq f(x) + \langle g, y - x\rangle, \forall y \in \mathcal{H}\}$$

(we allow $\infty \geq \infty$ and $-\infty \geq -\infty$). In some cases, the subdifferential operator ∂f is a single-valued operator. Then, we write $\nabla f = \partial f$.

Proximal operators.. We call f a CCP function if it is a convex, closed, and proper [18, 16]. For a CCP function $f\colon \mathcal{H} \to \mathbb{R} \cup \{\pm\infty\}$ and $\alpha > 0$, we define the proximal operator with respect to αf as

$$\mathrm{Prox}_{\alpha f}(x) = \operatorname*{argmin}_{y \in \mathcal{H}} \left\{ \alpha f(y) + \frac{1}{2}\|x - y\|^2 \right\}.$$

Then, $\mathbb{J}_{\alpha\partial f} = \mathrm{Prox}_{\alpha f}$.

Class of functions and subdifferential operators.. Define $f\colon \mathcal{H} \to \mathbb{R} \cup \{\pm\infty\}$ being μ-strongly convex (for $\mu \in (0,\infty)$) and L-smooth (for $L \in (0,\infty)$) as they are defined in [43]. Write

$$\mathcal{F}_{\mu,L} = \{f \mid f \text{ is } \mu\text{-strongly convex, } L\text{-smooth, and CCP.}\}\,.$$

for collection of functions that are μ-strongly convex and L-smooth at the same time. For notational simplicity, we extend $\mathcal{F}_{\mu,L}$ to allow $\mu = 0$ or $L = \infty$ by defining

$$\begin{aligned}
\mathcal{F}_{0,L} &= \{f \mid f \text{ is } L\text{-smooth and CCP.}\}\,,\\
\mathcal{F}_{\mu,\infty} &= \{f \mid f \text{ is } \mu\text{-strongly convex and CCP.}\}\,,\\
\mathcal{F}_{0,\infty} &= \{f \mid f \text{ is CCP.}\}\,.
\end{aligned}$$

for μ, $L \in (0,\infty)$.

Subdifferential operators of any functions in $\mathcal{F}_{\mu,L}$ are denoted

$$\partial\mathcal{F}_{\mu,L} = \{\partial f \mid f \in \mathcal{F}_{\mu,L}\}\,.$$

Complex set notations.. Denote $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, and define $0^{-1} = \infty$ and $\infty^{-1} = 0$ in $\overline{\mathbb{C}}$. For $A \subset \overline{\mathbb{C}}$ and $\alpha \in \mathbb{C}$, define

$$\alpha A = \{\alpha z \mid z \in A\}\,, \quad \alpha + A = \{\alpha + z \mid z \in A\}\,, \quad A^{-1} = \left\{z^{-1} \mid z \in A\right\}.$$

For $A \subseteq \mathbb{C}$, define the boundary of A

$$\partial A = \overline{A} \setminus \mathrm{int} A.$$

We clarify that the usage of ∂ operator is different when it is applied to a function or a complex set; the former is the subdifferential operator, and the latter is the boundary operator. For circles and disks on the complex plane, write

$$\mathrm{Circ}(z,r) = \{w \in \mathbb{C} \mid |w-z| = r\}, \qquad \mathrm{Disk}(z,r) = \{w \in \mathbb{C} \mid |w-z| \leq r\}$$

for $z \in \mathbb{C}$ and $r \in (0,\infty)$. Note, $\mathrm{Circ}(z,r) = \partial\mathrm{Disk}(z,r)$.

Scaled relative graphs *[15].*. Define the SRG of an operator $\mathbb{A}\colon \mathcal{H} \rightrightarrows \mathcal{H}$ as

$$\mathcal{G}(\mathbb{A}) = \left\{ \frac{\|u-v\|}{\|x-y\|} \exp\left[\pm i\angle(u-v, x-y)\right] \,\middle|\, u \in \mathbb{A}x,\, v \in \mathbb{A}y,\, x \neq y \right\}$$
$$\Big(\cup \{\infty\} \text{ if } \mathbb{A} \text{ is not single-valued} \Big).$$

where the angle between $x \in \mathcal{H}$ and $y \in \mathcal{H}$ is defined as

$$\angle(x, y) = \begin{cases} \arccos\left(\frac{\langle x, y\rangle}{\|x\|\|y\|}\right) & \text{if } x \neq 0,\, y \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Note, SRG is a subset of $\overline{\mathbb{C}}$. Define the SRG of a class of operators $\mathcal{A}$ as

$$\mathcal{G}(\mathcal{A}) = \bigcup_{\mathbb{A} \in \mathcal{A}} \mathcal{G}(\mathbb{A}).$$

Say $\mathcal{A}$ is *SRG-full* if

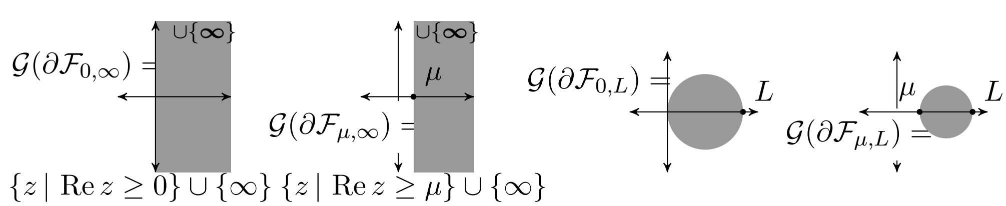
$$\mathbb{A} \in \mathcal{A} \quad \Leftrightarrow \quad \mathcal{G}(\mathbb{A}) \subseteq \mathcal{G}(\mathcal{A}).$$

Fact 1 (Theorem 4, 5 [15])**.** *If $\mathcal{A}$ is a class of operators, then*

$$\mathcal{G}(\alpha\mathcal{A}) = \alpha\mathcal{G}(\mathcal{A}), \quad \mathcal{G}(\mathbb{I} + \mathcal{A}) = 1 + \mathcal{G}(\mathcal{A}), \quad \mathcal{G}(\mathcal{A}^{-1}) = \mathcal{G}(\mathcal{A})^{-1}.$$

where α is a nonzero real number. If $\mathcal{A}$ is furthermore SRG-full, then $\alpha\mathcal{A}$, $\mathbb{I}+\mathcal{A}$, and $\mathcal{A}^{-1}$ are SRG-full.

Fact 2 (Proposition 2 [15])**.** *Let $0 < \mu < L < \infty$. Then*



DYS operators.. Let

$$\mathbb{T}_{\mathbb{A},\mathbb{B},\mathbb{C},\alpha,\lambda} = \mathbb{I} - \lambda \mathbb{J}_{\alpha\mathbb{B}} + \lambda \mathbb{J}_{\alpha\mathbb{A}}(2\mathbb{J}_{\alpha\mathbb{B}} - \mathbb{I} - \alpha\mathbb{C}\mathbb{J}_{\alpha\mathbb{B}})$$

be the DYS operator for operators $\mathbb{A}\colon \mathcal{H} \rightrightarrows \mathcal{H}$, $\mathbb{B}\colon \mathcal{H} \rightrightarrows \mathcal{H}$, and $\mathbb{C}\colon \mathcal{H} \rightrightarrows \mathcal{H}$ with stepsize $\alpha \in (0,\infty)$ and averaging parameter $\lambda \in (0,\infty)$. In this paper, we usually take $\mathbb{A} = \partial f$, $\mathbb{B} = \partial g$, and $\mathbb{C} = \nabla h$ for some CCP functions f, g, and h defined on $\mathcal{H}$, to obtain

$$\mathbb{T}_{\partial f,\partial g,\nabla h,\alpha,\lambda} = \mathbb{I} - \lambda \mathrm{Prox}_{\alpha g} + \lambda \mathrm{Prox}_{\alpha f}(2\mathrm{Prox}_{\alpha g} - \mathbb{I} - \alpha \nabla h \mathrm{Prox}_{\alpha g})$$

what we call the *subdifferential DYS operator*.

Let

$$\mathbb{T}_{\mathcal{A},\mathcal{B},\mathcal{C},\alpha,\lambda} = \{\mathbb{T}_{\mathbb{A},\mathbb{B},\mathbb{C},\alpha,\lambda} \mid \mathbb{A} \in \mathcal{A}, \mathbb{B} \in \mathcal{B}, \mathbb{C} \in \mathcal{C}\}$$

be the class of DYS operators for operator classes $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ with $\alpha, \lambda \in (0,\infty)$. Define

$$\begin{aligned}\zeta_{\mathrm{DYS}}(z_A, z_B, z_C; \alpha, \lambda) &= 1 - \lambda z_B + \lambda z_A(2z_B - 1 - \alpha z_C z_B)\\ &= 1 - \lambda z_A - \lambda z_B + \lambda(2 - \alpha z_C) z_A z_B,\end{aligned}$$

which exhibits symmetry with respect to z_A and z_B, and

$$\mathcal{Z}^{\mathrm{DYS}}_{\mathcal{A},\mathcal{B},\mathcal{C},\alpha,\lambda} = \{\zeta_{\mathrm{DYS}}(z_A, z_B, z_C; \alpha, \lambda) \mid z_A \in \mathcal{G}(\mathbb{J}_{\alpha\mathcal{A}}), z_B \in \mathcal{G}(\mathbb{J}_{\alpha\mathcal{B}}), z_C \in \mathcal{G}(\mathcal{C})\}$$

for operator classes $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ with $\alpha, \lambda \in (0,\infty)$.

Identifying the tight Lipschitz coefficient via SRG.. Say a subset of $\overline{\mathbb{C}}$ is a *generalized disk* if it is a disk, $\{z \mid \operatorname{Re} z \geq a\} \cup \{\infty\}$, or $\{z \mid \operatorname{Re} z \leq a\} \cup \{\infty\}$ for a real number a. The following is the key fact for calculating the Lipschitz coefficients of the DYS operators via SRG.

Fact 3 (Corollary 1 of [14]). *Let $\alpha, \lambda > 0$. Let $\mathcal{A}$ and $\mathcal{B}$ be SRG-full classes of monotone operators where $\mathcal{G}(\mathbb{I} + \alpha\mathcal{A})$ forms a generalized disk. Let $\mathcal{C}$ be an SRG-full class of single-valued operators. Assume $\mathcal{G}(\mathcal{A})$, $\mathcal{G}(\mathcal{B})$, and $\mathcal{G}(\mathcal{C})$ are nonempty. Then,*

$$\sup_{\substack{\mathbb{T} \in \mathbb{T}_{\mathcal{A},\mathcal{B},\mathcal{C},\alpha,\lambda}\\ x,y \in \operatorname{dom}\mathbb{T}, x \neq y}} \frac{\|\mathbb{T}x - \mathbb{T}y\|}{\|x - y\|} = \sup_{z \in \mathcal{Z}^{DYS}_{\mathcal{A},\mathcal{B},\mathcal{C},\alpha,\lambda}} |z|.$$

Indeed, the original version of Fact 3 is consistent with $\mathcal{G}(\mathbb{I} + \alpha\mathcal{A})$ having a more general property, namely an arc property. We can calculate bounds for the right-hand-side of the equality in Fact 3 efficiently by using the following fact.

Fact 4 (Lemma 1 of [14])**.** *Let $f : \mathbb{C}^3 \to \mathbb{C}$ be a polynomial of three complex variables. Let A, B, and C be compact subsets of $\mathbb{C}$. Then,*

$$\max_{\substack{z_A \in A, z_B \in B, \\ z_C \in C}} |f(z_A, z_B, z_C)| = \max_{\substack{z_A \in \partial A, z_B \in \partial B, \\ z_C \in \partial C}} |f(z_A, z_B, z_C)|.$$

2. Lipschitz factors of subdifferential DYS

We present Lipschitz factors of subdifferential DYS operators. To the best of our knowledge, the following results are the best linear convergence rates of the DYS iteration compared to Theorem 9 of [13] and Theorems 3, 4, and 5 of [14]. To clarify, our rates are not slower than the prior rates in all cases and faster in most cases.

Theorem 1. *Let $f \in \mathcal{F}_{\mu_f, L_f}$, $g \in \mathcal{F}_{\mu_g, L_g}$, and $h \in \mathcal{F}_{\mu_h, L_h}$, where*

$$0 \le \mu_f < L_f \le \infty, \quad 0 \le \mu_g < L_g \le \infty, \quad 0 \le \mu_h < L_h < \infty.$$

Let $\lambda > 0$ be an averaging parameter and $\alpha > 0$ be a step size. Throughout this theorem, define $r/\infty = 0$ for any real number r. Write

$$C_f = \frac{1}{2}\left(\frac{1}{1+\alpha\mu_f} + \frac{1}{1+\alpha L_f}\right), \quad C_g = \frac{1}{2}\left(\frac{1}{1+\alpha\mu_g} + \frac{1}{1+\alpha L_g}\right),$$
$$R_f = \frac{1}{2}\left(\frac{1}{1+\alpha\mu_f} - \frac{1}{1+\alpha L_f}\right), \quad R_g = \frac{1}{2}\left(\frac{1}{1+\alpha\mu_g} - \frac{1}{1+\alpha L_g}\right),$$
$$d = \max\{|2 - \lambda - \alpha\mu_h|, |2 - \lambda - \alpha L_h|\}.$$

If $\lambda < 1/C_f$, then $\mathbb{T}_{\partial f, \partial g, \nabla h, \alpha, \lambda}$ is ρ_f-Lipschitz, where

$$\rho_f^2 = \left(1 - \lambda\frac{C_f^2 - R_f^2}{C_f}\right) \max\left\{ \left(1 - \frac{\lambda}{1+\alpha\mu_g}\right)^2 + \frac{\lambda d^2}{1/C_f - \lambda}\left(\frac{1}{1+\alpha\mu_g}\right)^2, \right.$$
$$\left. \left(1 - \frac{\lambda}{1+\alpha L_g}\right)^2 + \frac{\lambda d^2}{1/C_f - \lambda}\left(\frac{1}{1+\alpha L_g}\right)^2 \right\}.$$

Symmetrically, if $\lambda < 1/C_g$, then $\mathbb{T}_{\partial f,\partial g,\nabla h,\alpha,\lambda}$ is ρ_g-Lipschitz, where

$$\rho_g^2 = \left(1 - \lambda\frac{C_g^2 - R_g^2}{C_g}\right)\max\left\{\left(1 - \frac{\lambda}{1+\alpha\mu_f}\right)^2 + \frac{\lambda d^2}{1/C_g - \lambda}\left(\frac{1}{1+\alpha\mu_f}\right)^2,\right.$$
$$\left.\left(1 - \frac{\lambda}{1+\alpha L_f}\right)^2 + \frac{\lambda d^2}{1/C_g - \lambda}\left(\frac{1}{1+\alpha L_f}\right)^2\right\}.$$

Theorem 2. *Let f, g, h, μ_f, L_f, μ_g, L_g, μ_h, L_h, λ, and α be the same as in Theorem 1. Additionally, assume $\lambda < 2 - \frac{\alpha(\mu_h+L_h)}{2}$. Define $\infty/\infty^2 = 0$. Write*

$$\nu_f = \min\left\{\frac{2\mu_f+\mu_h}{(1+\alpha\mu_f)^2}, \frac{2L_f+\mu_h}{(1+\alpha L_f)^2}\right\}, \quad \nu_g = \min\left\{\frac{2\mu_g+\mu_h}{(1+\alpha\mu_g)^2}, \frac{2L_g+\mu_h}{(1+\alpha L_g)^2}\right\},$$
$$\theta = \frac{2}{4-\alpha(\mu_h+L_h)}.$$

Then, $\mathbb{T}_{\partial f,\partial g,\nabla h,\alpha,\lambda}$ is ρ-contractive, where

$$\rho^2 = 1 - \lambda\theta + \lambda\sqrt{(\theta - \alpha\nu_f)(\theta - \alpha\nu_g)}.$$

2.1. Proofs of Theorems 1 and 2

To leverage Fact 3, we need to set adequate SRG-full operator classes $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ to apply it. The most natural choice is to set

$$\mathcal{A} = \partial\mathcal{F}_{\mu_f,L_f}, \quad \mathcal{B} = \partial\mathcal{F}_{\mu_g,L_g}, \quad \mathcal{C} = \partial\mathcal{F}_{\mu_h,L_h}.$$

However, this choice is not appropriate since these are not SRG-full classes. To overcome this issue, we introduce the following operator classes:

$$\mathcal{D}_f = \{\mathbb{A}\colon \mathcal{H} \rightrightarrows \mathcal{H} \,|\, \mathcal{G}(\mathbb{A}) \subseteq \mathcal{G}(\partial\mathcal{F}_{\mu_f,L_f})\},$$
$$\mathcal{D}_g = \{\mathbb{B}\colon \mathcal{H} \rightrightarrows \mathcal{H} \,|\, \mathcal{G}(\mathbb{B}) \subseteq \mathcal{G}(\partial\mathcal{F}_{\mu_g,L_g})\},$$
$$\mathcal{D}_h = \{\mathbb{C}\colon \mathcal{H} \rightrightarrows \mathcal{H} \,|\, \mathcal{G}(\mathbb{C}) \subseteq \mathcal{G}(\partial\mathcal{F}_{\mu_h,L_h})\}.$$

To elaborate, we gather all operators that have its SRG within $\mathcal{G}(\partial\mathcal{F}_{\mu_f,L_f})$ to form $\mathcal{D}_f$, and so on. Then, $\mathcal{D}_f$, $\mathcal{D}_g$, and $\mathcal{D}_h$ are SRG-full classes by definition. We now consider $\mathcal{A} = \mathcal{D}_f$, $\mathcal{B} = \mathcal{D}_g$, and $\mathcal{C} = \mathcal{D}_h$ in the following proof.

We mention two elementary facts.

Fact 5. *For $a, b, c, d \in [0, \infty)$,*

$$\left(\sqrt{ab} + \sqrt{cd}\right)^2 \leq (a+c)(b+d).$$

Proof. This inequality is an instance of Cauchy–Schwartz. $\square$

Fact 6. *Let k, l, and r be positive real numbers, and b, c be a real number. For $z \in \mathrm{Circ}(c, r)$,*

$$k|z-b|^2 + l|z|^2$$

is maximized at $z = c - r$ or $z = c + r$.

Proof to Fact 6. Observe that

$$k|z-b|^2 + l|z|^2 = (k+l)\left|z - \frac{kb}{k+l}\right|^2 + \frac{klb^2}{k+l}.$$

and distance from $\frac{kb}{k+l}$ to $z \in \mathrm{Circ}(c, r)$ is maximized at $z = c - r$ if $\frac{kb}{k+l} > c$ and $z = c + r$ otherwise. $\square$

We now prove Theorem 1, 2. We use the similar proof technique introduced in the previous work [14].

Proof to Theorem 1. We first prove the first statement and show the other by the same reasoning. Invoking Fact 3 and Fact 4, it suffices to show that

$$|\zeta_{\mathrm{DYS}}(z_f, z_g, z_h; \alpha, \lambda)|^2 \leq \rho_f^2$$

for

$$\begin{aligned}
z_f &\in \partial\mathcal{G}\left(\mathbb{J}_{\alpha\mathcal{D}_f}\right) = \mathrm{Circ}\left(C_f, R_f\right), \\
z_g &\in \partial\mathcal{G}\left(\mathbb{J}_{\alpha\mathcal{D}_g}\right) = \mathrm{Circ}\left(C_g, R_g\right), \\
z_h &\in \partial\mathcal{G}\left(\mathcal{D}_h\right) = \mathrm{Circ}\left(\frac{L_h + \mu_h}{2}, \frac{L_h - \mu_h}{2}\right)
\end{aligned}$$

(refer to Fact 2 and Fact 1 to see why $\partial\mathcal{G}\left(\mathbb{J}_{\alpha\mathcal{D}_f}\right)$, $\partial\mathcal{G}\left(\mathbb{J}_{\alpha\mathcal{D}_g}\right)$, $\mathcal{G}\left(\mathcal{D}_h\right)$ are formed correspondingly) when $\lambda < 1/C_f$ holds.

Define

$$r = \frac{d}{1/C_f - \lambda}.$$

Then,

$$
\begin{aligned}
|\zeta_{\mathrm{DYS}}(z_f, z_g, z_h; \alpha, \lambda)|^2 &= |1 - \lambda z_f - \lambda z_g + \lambda(2 - \alpha z_h) z_f z_g|^2 \\
&= |(1 - \lambda z_f)(1 - \lambda z_g) + \lambda(2 - \lambda - \alpha z_h) z_f z_g|^2 \\
&\le \left(|(1 - \lambda z_f)(1 - \lambda z_g)| + |\lambda(2 - \lambda - \alpha z_h) z_f z_g|\right)^2 \\
&\overset{(i)}{\le} \left(|(1 - \lambda z_f)(1 - \lambda z_g)| + \lambda d |z_f z_g|\right)^2 \\
&\overset{(ii)}{\le} \left(|1 - \lambda z_f|^2 + \lambda d r^{-1} |z_f|^2\right)\left(|1 - \lambda z_g|^2 + \lambda d r |z_g|^2\right).
\end{aligned} \tag{3}
$$

We have used $|2 - \lambda - \alpha z_h| \le \max\{|2 - \lambda - \alpha\mu_h|, |2 - \lambda - \alpha L_h|\} = d$ in (i) and Fact 5 in (ii).

Recall that $\partial\mathcal{G}\left(\mathbb{J}_{\alpha\mathcal{D}_f}\right) = \mathrm{Circ}(C_f, R_f)$. This renders

$$
|1 - \lambda z_f|^2 + \lambda d r^{-1} |z_f|^2 = \frac{\lambda}{C_f}|z_f - C_f|^2 + 1 - \lambda C_f = 1 - \lambda \frac{C_f^2 - R_f^2}{C_f}. \tag{4}
$$

For the latter term, $z_g = \frac{1}{1+\alpha\mu_g}$ or $z_g = \frac{1}{1+\alpha L_g}$ give the maximum, invoking Fact 6. Therefore,

$$
\begin{aligned}
&|1 - \lambda z_g|^2 + \lambda d r |z_g|^2 \\
&\le \max\Bigg\{ \left(1 - \frac{\lambda}{1+\alpha\mu_g}\right)^2 + \frac{\lambda d^2}{1/C_f - \lambda}\left(\frac{1}{1+\alpha\mu_g}\right)^2, \\
&\qquad \left(1 - \frac{\lambda}{1+\alpha L_g}\right)^2 + \frac{\lambda d^2}{1/C_f - \lambda}\left(\frac{1}{1+\alpha L_g}\right)^2 \Bigg\}.
\end{aligned} \tag{5}
$$

(4) and (5) together give

$$
|\zeta_{\mathrm{DYS}}(z_f, z_g, z_h; \alpha, \lambda)|^2 \le \rho_f^2
$$

which concludes the proof for the first statement. The second statement can be proven in the same reasoning. □

Proof to Theorem 2. By the same reasoning in the proof of Theorem 1, it suffices to show that

$$
|\zeta_{\mathrm{DYS}}(z_f, z_g, z_h; \alpha, \lambda)| \le \rho
$$

for

$$
z_f \in \partial\mathcal{G}\left(\mathbb{J}_{\alpha\mathcal{D}_f}\right), \quad z_g \in \partial\mathcal{G}\left(\mathbb{J}_{\alpha\mathcal{D}_g}\right), \quad z_h \in \partial\mathcal{G}\left(\mathcal{D}_h\right).
$$

Recall that $\theta = \frac{2}{4-\alpha(\mu_h+L_h)}$. Note that

$$\partial\mathcal{G}\left(\mathcal{D}_h\right) = \mathrm{Circ}\left(\frac{L_h+\mu_h}{2}, \frac{L_h-\mu_h}{2}\right)$$

thus

$$|2-\theta^{-1}-\alpha z_h| = \alpha\left|z_h - \frac{L_h+\mu_h}{2}\right| = \alpha\frac{L_h-\mu_h}{2} = 2-\alpha\mu_h-\theta^{-1}. \tag{6}$$

Now, observe

$$\begin{aligned}
&|\zeta_{\mathrm{DYS}}(z_f, z_g, z_h; \alpha, \lambda) - (1-\lambda\theta)|^2 \\
&= \lambda^2|\theta - z_f - z_g + (2-\alpha z_h)z_f z_g|^2 \\
&= \lambda^2|\theta^{-1}(z_f-\theta)(z_g-\theta) + (2-\theta^{-1}-\alpha z_h)z_f z_g|^2 \\
&\le \lambda^2\left(\theta^{-1}|z_f-\theta||z_g-\theta| + |2-\theta^{-1}-\alpha z_h||z_f||z_g|\right)^2 \\
&\overset{(i)}{=} \lambda^2\left(\theta^{-1}|z_f-\theta||z_g-\theta| + (2-\alpha\mu_h-\theta^{-1})|z_f||z_g|\right)^2 \\
&\overset{(ii)}{\le} \lambda^2\left(\theta^{-1}|z_f-\theta|^2 + (2-\alpha\mu_h-\theta^{-1})|z_f|^2\right)\left(\theta^{-1}|z_g-\theta|^2 + (2-\alpha\mu_h-\theta^{-1})|z_g|^2\right).
\end{aligned} \tag{7}$$

Here, (i) follows from (6) and (ii) follows from Fact 5.

Invoking Fact 6,

$$\theta^{-1}|z_f-\theta|^2 + (2-\alpha\mu_h-\theta^{-1})|z_f|^2$$

is maximized at either $z_f = \frac{1}{1+\alpha L_f}$ or $z_f = \frac{1}{1+\alpha\mu_f}$. The first term evaluates to

$$\theta^{-1}|z_f-\theta|^2 + (2-\alpha\mu_h-\theta^{-1})|z_f|^2 = \theta - \alpha\frac{2L_f+\mu_h}{(1+\alpha L_f)^2}$$

when $z_f = \frac{1}{1+\alpha L_f}$ and similarly to

$$\theta^{-1}|z_f-\theta|^2 + (2-\alpha\mu_h-\theta^{-1})|z_f|^2 = \theta - \alpha\frac{2\mu_f+\mu_h}{(1+\alpha\mu_f)^2}$$

when $z_f = \frac{1}{1+\alpha\mu_f}$. Hence,

$$\begin{aligned}
\theta^{-1}|z_f-\theta|^2 + (2-\alpha\mu_h-\theta^{-1})|z_f|^2 &\le \theta - \alpha\min\left\{\frac{2L_f+\mu_h}{(1+\alpha L_f)^2}, \frac{2\mu_f+\mu_h}{(1+\alpha\mu_f)^2}\right\} \\
&= \theta - \alpha\nu_f.
\end{aligned} \tag{8}$$

Similarly, we have

$$\theta^{-1}|z_g-\theta|^2+(2-\alpha\mu_h-\theta^{-1})|z_g|^2 \le \theta-\alpha\nu_g. \tag{9}$$

Combining (7), (8), and (9), we obtain

$$|\zeta_{\mathrm{DYS}}(z_f,z_g,z_h;\alpha,\lambda)| \le 1-\lambda\theta+\lambda\sqrt{(\theta-\alpha\nu_f)\,(\theta-\alpha\nu_g)}.$$

which is the desired bound. □

2.2. Comparison with previous results

We now compare our linear convergence rates with prior results and show that our results are better in general.

2.2.1. Comparison with the result of [13].

Condat and Richtarik [13] showed the following rate.

Fact 7 (Setting $\omega=0$ in Theorem 9 of [13])**.** *Assume $\mu_g>0$ or $\mu_h>0$, and that L_f, $L_h\in(0,\infty)$. Furthermore, suppose that $\alpha\in(0,2/L_h)$. Consider problem (1) where $f\in\mathcal{F}_{0,L_f}$, $g\in\mathcal{F}_{\mu_g,\infty}$, and $h\in\mathcal{F}_{\mu_h,L_h}$. Let $x^\star$ be a solution for (1). The DYS iteration converges with the geometric rate of,*

$$\rho_{old}^2=\max\left\{\frac{(1-\alpha\mu_h)^2}{1+\alpha\mu_g},\frac{(1-\alpha L_h)^2}{1+\alpha\mu_g},\frac{\alpha L_f}{\alpha L_f+2}\right\}.$$

Assuming the same setting of Fact 7, we get the following as a direct result of Theorem 1.

Corollary 1. *Assume the same conditions as in Fact 7. Then, $\mathbb{T}_{\partial f,\partial g,\nabla h,\alpha,1}$ is ρ_{new}-contractive, where*

$$d=\max\{|1-\alpha\mu_h|,|1-\alpha L_h|\}$$

and

$$\rho_{new}^2=\max\left\{\frac{d^2}{1+2\alpha\mu_g},\frac{1}{(1+\alpha L_f)^2}\left(\alpha^2L_f^2+\frac{d^2}{1+2\alpha\mu_g}\right)\right\}. \tag{10}$$

The following proposition claims that the contraction factor introduced in Corollary 1 is strictly better than that of Fact 7 in general cases.

Proposition 1. *Let μ_g, μ_h, L_f, L_h, α be as in Fact 7. Let ρ_{old} and ρ_{new} be as in Fact 7 and Corollary 1, respectively. Then, $\rho_{new} < \rho_{old}$ if $\mu_g > 0$, and $\rho_{new} = \rho_{old}$ if $\mu_g = 0$.*

Proof to Proposition 1. Denote $d = \max\{|1-\alpha\mu_h|, |1-\alpha L_h|\}$ and recall that

$$\rho_{\text{old}}^2 = \max\left\{\frac{(1-\alpha\mu_h)^2}{1+\alpha\mu_g}, \frac{(1-\alpha L_h)^2}{1+\alpha\mu_g}, \frac{\alpha L_f}{\alpha L_f + 2}\right\} = \max\left\{\frac{d^2}{1+\alpha\mu_g}, \frac{\alpha L_f}{\alpha L_f + 2}\right\}$$
$$\rho_{\text{new}}^2 = \max\left\{\frac{d^2}{1+2\alpha\mu_g}, \frac{1}{(1+\alpha L_f)^2}\left(\alpha^2 L_f^2 + \frac{d^2}{1+2\alpha\mu_g}\right)\right\}.$$

First, consider the case where

$$\frac{d^2}{1+2\alpha\mu_g} \geq \frac{\alpha L_f}{\alpha L_f + 2}.$$

In this case,

$$\frac{d^2}{1+2\alpha\mu_g} \geq \frac{1}{(1+\alpha L_f)^2}\left(\alpha^2 L_f^2 + \frac{d^2}{1+2\alpha\mu_g}\right)$$

holds, and together with the assumption that $\mu_g > 0$,

$$\rho_{\text{new}}^2 = \frac{d^2}{1+2\alpha\mu_g} < \frac{d^2}{1+\alpha\mu_g} \leq \rho_{\text{old}}^2.$$

Otherwise if

$$\frac{d^2}{1+2\alpha\mu_g} < \frac{\alpha L_f}{\alpha L_f + 2},$$

observe that

$$\begin{aligned}
\rho_{\text{new}}^2 &= \frac{1}{(1+\alpha L_f)^2}\left(\alpha^2 L_f^2 + \frac{d^2}{1+2\alpha\mu_g}\right)\\
&< \frac{1}{(1+\alpha L_f)^2}\left(\alpha^2 L_f^2 + \frac{\alpha L_f}{\alpha L_f + 2}\right)\\
&= \frac{\alpha L_f}{\alpha L_f + 2}\\
&\leq \rho_{\text{old}}^2.
\end{aligned}$$

□

2.2.2. Comparison with [14].

Lee, Yi, and Ryu introduced contraction factors in Theorems 3, 4, and 5 of [14]. The following proposition exhibits those contraction factors and claims ours are strictly better in most cases due to stronger assumptions on the operators.

Proposition 2. *Let μ_f, μ_g, L_f, L_g, L_h, α, and λ be as in Theorem 1. Additionally, assume $\alpha L_h < 4$ and $\lambda < 2 - \frac{\alpha L_h}{2}$. Then,*

1. *Assume $f \in \mathcal{F}_{\mu_f, L_f}$, $g \in \mathcal{F}_{0,\infty}$, and $h \in \mathcal{F}_{0,L_h}$. Then, Theorem 3 of [14] implies $\mathbb{T}_{\partial f, \partial g, \nabla h, \alpha, \lambda}$ is ρ_{old}-contractive, where*

$$\rho_{old} = 1 - \frac{2\lambda}{4 - \alpha L_h} + \lambda \sqrt{\frac{2}{4 - \alpha L_h}\left(\frac{2}{4 - \alpha L_h} - \frac{2\alpha\mu_f}{\alpha^2 L_f^2 + 2\alpha\mu_f + 1}\right)}.$$

Meanwhile, Theorem 2 implies $\mathbb{T}_{\partial f, \partial g, \nabla h, \alpha, \lambda}$ is ρ_{new}-contractive, where

$$\rho_{new} = 1 - \frac{2\lambda}{4 - \alpha L_h} + \lambda \sqrt{\frac{2}{4 - \alpha L_h}\left(\frac{2}{4 - \alpha L_h} - \alpha \min\left\{\frac{2\mu_f}{(1 + \alpha\mu_f)^2}, \frac{2L_f}{(1 + \alpha L_f)^2}\right\}\right)}.$$

and $\rho_{new} < \rho_{old}$ if $\mu_f > 0$, and $\rho_{new} = \rho_{old}$ if $\mu_f = 0$.

2. *Assume $f \in \mathcal{F}_{0,L_f}$, $g \in \mathcal{F}_{\mu_g,\infty}$, and $h \in \mathcal{F}_{0,L_h}$. Then, $\mathbb{T}_{\partial f, \partial g, \nabla h, \alpha, \lambda}$ is ρ_{old}-contractive where*

$$\rho_{old} = \sqrt{1 - 2\lambda\alpha \min\left\{\frac{(2-\lambda)\mu_g}{(1 + \alpha^2 L_f^2)(2 - \lambda + 2\alpha\mu_g)}, \frac{(2-\lambda)(\mu_g + L_f) + 2\alpha\mu_g L_f}{(1 + \alpha L_f)^2(2 - \lambda + 2\alpha\mu_g)}\right\}}.$$

Meanwhile, Theorem 1 implies $\mathbb{T}_{\partial f, \partial g, \nabla h, \alpha, \lambda}$ is ρ_{new}-contractive where

$$\rho_{new} = \sqrt{1 - 2\lambda\alpha \min\left\{\frac{(2-\lambda)\mu_g}{2 - \lambda + 2\alpha\mu_g}, \frac{(2-\lambda)(\mu_g + L_f) + 2\alpha\mu_g L_f}{(1 + \alpha L_f)^2(2 - \lambda + 2\alpha\mu_g)}\right\}}.$$

and $\rho_{new} \le \rho_{old}$, and $\rho_{new} < \rho_{old}$ if $(2-\lambda)(1 - 2\alpha\mu_g + \alpha^2 L_f^2) + 2\alpha\mu_g(1 + \alpha^2 L_f^2) > 0$ and $\mu_g > 0$.

3. *Assume $f \in \mathcal{F}_{0,L_f}$, $g \in \mathcal{F}_{0,\infty}$, and $h \in \mathcal{F}_{\mu_h, L_h}$. Then, $\mathbb{T}_{\partial f, \partial g, \nabla h, \alpha, \lambda}$ is ρ_{old}-contractive where*

$$\rho_{old} = \sqrt{1 - 2\lambda\alpha \min\left\{\frac{\mu_h\left(1 - \frac{\alpha L_h}{2(2-\lambda)}\right)}{1 + \alpha^2 L_f^2}, \frac{L_f + \mu_h\left(1 - \frac{\alpha L_h}{2(2-\lambda)}\right)}{(1 + \alpha L_f)^2}\right\}}.$$

Meanwhile, Theorem 1 implies $\mathbb{T}_{\partial f,\partial g,\nabla h,\alpha,\lambda}$ *is* ρ_{new}*-contractive where*

$$\rho_{new} = \sqrt{1 - 2\lambda\alpha \min\left\{\xi, \frac{L_f + \xi}{(1+\alpha L_f)^2}\right\}}.$$

where

$$\xi = \min\left\{\mu_h\left(1 - \frac{\alpha\mu_h}{2(2-\lambda)}\right), L_h\left(1 - \frac{\alpha L_h}{2(2-\lambda)}\right)\right\}.$$

Furthermore, $\rho_{new} < \rho_{old}$ *if* $\mu_h > 0$*, and* $\rho_{new} = \rho_{old}$ *if* $\mu_h = 0$*.*

Proof to Proposition 2-1. To show $\rho_{\text{new}} < \rho_{\text{old}}$ given $\mu_f > 0$, it suffices to show

$$\min\left\{\frac{2\mu_f}{(1+\alpha\mu_f)^2}, \frac{2L_f}{(1+\alpha L_f)^2}\right\} > \frac{2\mu_f}{\alpha^2 L_f^2 + 2\alpha\mu_f + 1},$$

or equivalently

$$\frac{2\mu_f}{(1+\alpha\mu_f)^2} > \frac{2\mu_f}{\alpha^2 L_f^2 + 2\alpha\mu_f + 1}$$
$$\frac{2L_f}{(1+\alpha L_f)^2} > \frac{2\mu_f}{\alpha^2 L_f^2 + 2\alpha\mu_f + 1}$$

which are straightforward from $L_f > \mu_f > 0$. □

Proof to Proposition 2-2. $\rho_{\text{new}} \le \rho_{\text{old}}$ is obvious.

By straightforward calculations, the mentioned conditions imply

$$\frac{(2-\lambda)\mu_g}{(1+\alpha^2 L_f^2)(2-\lambda+2\alpha\mu_g)} < \frac{(2-\lambda)(\mu_g + L_f) + 2\alpha\mu_g L_f}{(1+\alpha L_f)^2(2-\lambda+2\alpha\mu_g)}. \tag{11}$$

Thus,

$$\rho_{\text{old}} = \sqrt{1 - 2\lambda\alpha\frac{(2-\lambda)\mu_g}{(1+\alpha^2 L_f^2)(2-\lambda+2\alpha\mu_g)}}$$
$$\rho_{\text{new}} = \sqrt{1 - 2\lambda\alpha \min\left\{\frac{(2-\lambda)\mu_g}{2-\lambda+2\alpha\mu_g}, \frac{(2-\lambda)(\mu_g + L_f) + 2\alpha\mu_g L_f}{(1+\alpha L_f)^2(2-\lambda+2\alpha\mu_g)}\right\}}$$

holds. Then,

$$\frac{(2-\lambda)\mu_g}{(1+\alpha^2 L_f^2)(2-\lambda+2\alpha\mu_g)} < \frac{(2-\lambda)\mu_g}{2-\lambda+2\alpha\mu_g}$$

together with (11) gives $\rho_{\text{new}} < \rho_{\text{old}}$. □

Proof to Proposition 2-3. First, observe that $0 < \mu_h < L_h$ renders

$$\xi > \mu_h \left(1 - \frac{\alpha L_h}{2(2-\lambda)}\right).$$

Therefore,

$$\xi > \frac{\mu_h \left(1 - \frac{\alpha L_h}{2(2-\lambda)}\right)}{1 + \alpha^2 L_f^2},$$

$$\frac{L_f + \xi}{(1+\alpha L_f)^2} > \frac{L_f + \mu_h \left(1 - \frac{\alpha L_h}{2(2-\lambda)}\right)}{(1+\alpha L_f)^2}$$

which give $\rho_{\text{new}} < \rho_{\text{old}}$. □

3. Discussion and conclusion

The reduction of Fact 3 allows us to obtain the Lipschitz coefficients Theorems 1 and 2 by characterizing the maximum modulus of

$$\mathcal{Z}^{\text{DYS}}_{\mathcal{A},\mathcal{B},\mathcal{C},\alpha,\lambda} = \left\{ \zeta_{\text{DYS}}(z_f, z_g, z_h; \alpha, \lambda) \,\middle|\, z_f \in \mathcal{G}\left(\mathbb{J}_{\alpha \mathcal{D}_f}\right), z_g \in \mathcal{G}\left(\mathbb{J}_{\alpha \mathcal{D}_g}\right), z_h \in \mathcal{G}\left(\mathcal{D}_h\right) \right\},$$

where $\zeta_{\text{DYS}} = 1 - \lambda z_B + \lambda z_A(2z_B - 1 - \alpha z_C z_B)$ is a relatively simple polynomial. This only requires elementary mathematics, and it is considerably easier than directly analyzing

$$\left\{ \frac{\|\mathbb{T}x - \mathbb{T}y\|}{\|x-y\|} \,\middle|\, \mathbb{T} \in \mathbb{T}_{\mathcal{D}_f, \mathcal{D}_g, \mathcal{D}_h, \alpha, \lambda},\ x, y \in \operatorname{dom} \mathbb{T},\ x \neq y \right\}.$$

Furthermore, by obtaining tighter bounds on the set $\mathcal{Z}^{\text{DYS}}_{\mathcal{A},\mathcal{B},\mathcal{C},\alpha,\lambda}$, one can improve upon the contraction factors presented in this work. We leave the tighter analysis to future work.

We point out that the explicit and simple description of $\mathcal{Z}^{\text{DYS}}_{\mathcal{A},\mathcal{B},\mathcal{C},\alpha,\lambda}$ allows one to investigate it in a numerical and computer-assisted manner. Sampling points from $\mathcal{Z}^{\text{DYS}}_{\mathcal{D}_f,\mathcal{D}_g,\mathcal{D}_h,\alpha,\lambda}$ is straightforward, and doing so provides a numerical estimate of the maximum modulus. For example, Figure 1 depicts $\mathcal{Z}^{\text{DYS}}_{\mathcal{D}_f,\mathcal{D}_g,\mathcal{D}_h,\alpha,\lambda}$ with a specific choice of μ_f, μ_g, μ_h, L_f, L_g, L_h, α, and λ. It shows that ρ_g, the contraction factor of Theorem 1, is valid but not tight; the gap between $\mathcal{Z}^{\text{DYS}}_{\mathcal{D}_f,\mathcal{D}_g,\mathcal{D}_h,\alpha,\lambda}$ and $\text{Circ}(0, \rho_g)$ indicates the contraction factor has room for improvement. Interestingly, if we modify the proof of Theorem 1 to choose r in (3) more carefully, we seem to obtain a tight contraction factor in the instance of Figure 1. Specifically, when we numerically minimize ρ as a function of r, we observe that $\mathrm{Circ}(0, \rho)$ touches $\mathcal{Z}^{\mathrm{DYS}}_{\mathcal{D}_f, \mathcal{D}_g, \mathcal{D}_h, \alpha, \lambda}$ in Figure 1 and the contact indicates tightness.

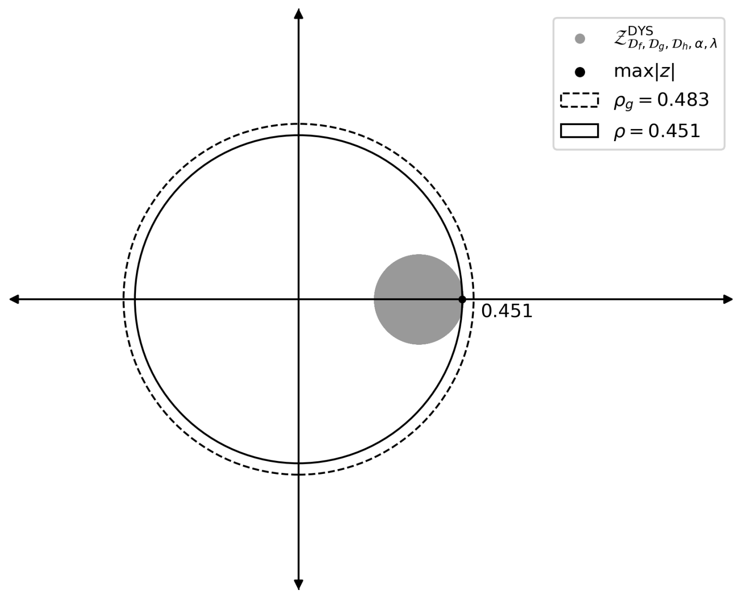


Figure 1: $\mathcal{Z}^{\mathrm{DYS}}_{\mathcal{D}_f, \mathcal{D}_g, \mathcal{D}_h, \alpha, \lambda}$ with $\mathrm{Circ}(0, \rho_f)$, $\mathrm{Circ}(0, \rho_g)$, and $\mathrm{Circ}(0, \rho)$, where $\mu_f = 0.7$, $\mu_g = 2$, $\mu_h = 0.8$, $L_f = 1.5$, $L_g = 3$, $L_h = 1.3$, $\alpha = 0.9$, and $\lambda = 1$.

References

[1] D. Davis, W. Yin, A three-operator splitting scheme and its optimization applications, Set-valued and variational analysis 25 (4) (2017) 829–858.

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

[3] Y. Wang, H. Zhou, S. Zu, W. Mao, Y. Chen, Three-operator proximal splitting scheme for 3-d seismic data reconstruction, IEEE Geoscience and Remote Sensing Letters 14 (10) (2017) 1830–1834.

[4] J. A. Carrillo, K. Craig, L. Wang, C. Wei, Primal dual methods for wasserstein gradient flows, Foundations of Computational Mathematics 22 (2) (2022) 389–443.

[5] D. Van Hieu, L. Van Vy, P. K. Quy, Three-operator splitting algorithm for a class of variational inclusion problems, Bulletin of the Iranian Mathematical Society 46 (4) (2020) 1055–1071.

[6] M. Weylandt, Splitting methods for convex bi-clustering and co-clustering, 2019 IEEE Data Science Workshop (2019) 237–242.

[7] H. Heaton, D. McKenzie, Q. Li, S. W. Fung, S. Osher, W. Yin, Learn to predict equilibria via fixed point networks, arXiv preprint arXiv:2106.00906 (2021).

[8] F. J. Aragón-Artacho, D. Torregrosa-Belén, A direct proof of convergence of Davis–Yin splitting algorithm allowing larger stepsizes, Set-Valued and Variational Analysis 30 (2022) 1–19.

[9] M. N. Dao, H. M. Phan, An adaptive splitting algorithm for the sum of three operators, arXiv:2104.05460 (2021).

[10] F. Pedregosa, On the convergence rate of the three operator splitting scheme, arXiv preprint arXiv:1610.07830 (2016).

[11] E. K. Ryu, A. B. Taylor, C. Bergeling, P. Giselsson, Operator splitting performance estimation: Tight contraction factors and optimal parameter selection, SIAM Journal on Optimization 30 (3) (2020) 2251–2271.

[12] H. Wang, M. Fazlyab, S. Chen, V. M. Preciado, Robust convergence analysis of three-operator splitting, Allerton Conference on Communication, Control, and Computing (2019).

[13] L. Condat, P. Richtárik, RandProx: Primal-Dual Optimization Algorithms with Randomized Proximal Updates (2022). `arXiv:2207.12891`.

[14] J. Lee, S. Yi, E. K. Ryu, Convergence Analyses of Davis-Yin Splitting via Scaled Relative Graphs, arXiv preprint arXiv:2207.04015 (2022). `arXiv:2207.04015`.

[15] E. K. Ryu, R. Hannah, W. Yin, Scaled relative graphs: Nonexpansive operators via 2D Euclidean geometry, Mathematical Programming 194 (1–2) (2022) 569–619.

[16] H. H. Bauschke, P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd Edition, Springer, 2017.

[17] E. K. Ryu, S. Boyd, Primer on monotone operator methods, Appl. Comput. Math 15 (1) (2016) 3–43.

[18] E. K. Ryu, W. Yin, Large-scale convex optimization via monotone operators, Cambridge University Press, 2022.

[19] R. E. Bruck, On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space, Journal of Mathematical Analysis and Applications 61 (1) (1977) 159–164.

[20] 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) (1979) 383–390.

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

[22] J. Douglas, 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) (1956) 421–439.

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

[24] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, Computers & Mathematics with Applications 2 (1) (1976) 17–40.

[25] A. Yurtsever, A. Gu, S. Sra, Three operator splitting with subgradients, stochastic gradients, and adaptive learning rates, NeurIPS (2021).

[26] A. Yurtsever, B. C. Vũ, V. Cevher, Stochastic three-composite convex minimization, NeurIPS (2016).

[27] V. Cevher, B. C. Vũ, A. Yurtsever, Stochastic forward Douglas–Rachford splitting method for monotone inclusions, in: A. R. Pontus Giselsson (Ed.), Large-Scale and Distributed Optimization, Springer, 2018, pp. 149–179.

[28] A. Yurtsever, V. Mangalick, S. Sra, Three Operator Splitting with a Nonconvex Loss Function, ICML (2021).

[29] F. Pedregosa, K. Fatras, M. Casotto, Proximal splitting meets variance reduction, in: AISTATS, 2019.

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

[31] F. Pedregosa, G. Gidel, Adaptive three operator splitting, ICML (2018).

[32] F. Cui, Y. Tang, Y. Yang, An inertial three-operator splitting algorithm with applications to image inpainting, Applied Set-Valued Analysis and Optimization 1 (2019).

[33] A. Salim, L. Condat, K. Mishchenko, P. Richtárik, Dualize, split, randomize: Toward fast nonsmooth optimization algorithms, Journal of Optimization Theory and Applications (2022).

[34] X. Huang, E. K. Ryu, W. Yin, Scaled relative graph of normal matrices, arXiv preprint arXiv:2001.02061 (2019).

[35] R. Pates, The Scaled Relative Graph of a Linear Operator, arXiv preprint arXiv:2106.05650 (2021).

[36] N. Ogura, I. Yamada, Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping, Numerical Functional Analysis and Optimization 23 (1-2) (2002) 113–137.

[37] X. Huang, E. K. Ryu, W. Yin, Tight coefficients of averaged operators via scaled relative graph, Journal of Mathematical Analysis and Applications 490 (1) (2020) 124211.

[38] T. Chaffey, F. Forni, R. Sepulchre, Graphical Nonlinear System Analysis, arXiv preprint arXiv:2107.11272 (2021).

[39] T. Chaffey, F. Forni, R. Sepulchre, Scaled relative graphs for system analysis, IEEE Conference on Decision and Control (2021).

[40] T. Chaffey, R. Sepulchre, Monotone one-port circuits, arXiv preprint arXiv:2111.15407 (2021).

[41] T. Chaffey, A rolled-off passivity theorem, Systems & Control Letters 162 (2022) 105198.

[42] T. Chaffey, A. Padoan, Circuit model reduction with scaled relative graphs, arXiv preprint arXiv:2204.01434 (2022).

[43] Y. Nesterov, Introductory Lectures on Convex Optimization: A Basic Course (2014).