SIAM J. OPTIM.
Vol. 30, No. 3, pp. 2251–2271

OPERATOR SPLITTING PERFORMANCE ESTIMATION: TIGHT CONTRACTION FACTORS AND OPTIMAL PARAMETER SELECTION*

ERNEST K. RYU[†], ADRIEN B. TAYLOR[‡], CAROLINA BERGELING[§], AND PONTUS GISELSSON[§]

Abstract. We propose a methodology for studying the performance of common splitting methods through semidefinite programming. We prove tightness of the methodology and demonstrate its value by presenting two applications of it. First, we use the methodology as a tool for computer-assisted proofs to prove tight analytical contraction factors for Douglas–Rachford splitting that are likely too complicated for a human to find bare-handed. Second, we use the methodology as an algorithmic tool to computationally select the optimal splitting method parameters by solving a series of semidefinite programs.

Key words. computer-aided analyses, first-order methods, rates of convergence, monotone operators, splitting methods

AMS subject classifications. 47H05, 47H09, 68Q25, 90C22, 90C25, 90C60

DOI. 10.1137/19M1304854

1. Introduction.

Consider the fixed-point iteration in a real Hilbert space $\mathcal{H}$,

$$z^{k+1} = Tz^k,$$

where $T\colon \mathcal{H} \to \mathcal{H}$. We say $\rho < 1$ is a contraction factor of T if

$$\|Tx - Ty\| \le \rho \|x - y\|$$

for all $x, y \in \mathcal{H}$. We ask the following question: Given a set of assumptions, what is the best (tight) contraction factor one can prove? In this work, we present the operator splitting performance estimation problem (OSPEP), a methodology for studying contraction factors of forward-backward splitting (FBS), Douglas–Rachford splitting (DRS), and Davis–Yin splitting (DYS).

First, we present the OSPEP problem, the infinite-dimensional nonconvex optimization problem of finding the best (smallest) contraction factor given a set of assumptions on the operators. Following the technique of Drori and Teboulle [20], we reformulate the problem into a finite-dimensional convex semidefinite program (SDP). We then establish tightness (exactness) of this reformulation with interpolation conditions.

---

*Received by the editors December 6, 2019; accepted for publication (in revised form) April 29, 2020; published electronically August 13, 2020.
https://doi.org/10.1137/19M1304854

Funding: The first author was partially supported by NSF grant DMS-1720237 and ONR grant N000141712162. The second author was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant 724063). The fourth author was supported by the Swedish Foundation for Strategic Research and the Swedish Research Council.

[†]Department of Mathematical Sciences, Seoul National University, Seoul, 08826, Korea (ernestryu@snu.ac.kr).

[‡]INRIA, Département d'informatique de l'ENS, École normale supérieure, CNRS, PSL Research University, 75012 Paris, France (adrien.taylor@inria.fr).

[§]Department of Automatic Control, Lund University, Lund, 22100, Sweden (carolina.bergeling@control.lth.se, pontus.giselsson@control.lth.se).

Next, we demonstrate the value of OSPEP through two uses. First, we use OSPEP as a tool for computer-assisted proofs to prove tight analytic contraction factors for DRS. The results are tight in that they have exact matching lower bounds. The proofs are computer-assisted in that their discoveries were assisted by a computer, but their verifications do not require a computer. Second, we use OSPEP as an algorithmic tool to automatically select the optimal splitting method parameters.

The tightness guarantee and flexibility of OSPEP make it a powerful tool. Due to tightness, OSPEP can provide both positive and negative results. The flexibility allows users to pick and choose assumptions from a set of standard assumptions.

1.1. Organization and contribution. Section 2 presents operator interpolation, later used in section 3 to establish tightness. Section 3 presents the OSPEP methodology, an exact transformation of the problem of finding the best contraction factor into a convex SDP, and provides tightness guarantees. Section 4 presents tight analytic contraction factors for DRS under assumptions considered in [25, 53] using OSPEP as a tool for computer-assisted proofs. Section 5 presents an automatic parameter selection method using OSPEP as an algorithmic tool. Section 6 concludes the paper.

The long computer-assisted proofs are deferred to the supplementary materials (SM) sections in the arXiv version of this paper [68].

The main contribution of this work is twofold. The first part is an analysis of the performance of monotone splitting methods using SDPs *with tightness guarantees*. The overall formulation generally follows from the technique of Drori and Teboulle [20] and the prior work discussed in section 1.2. The tightness, established with the operator interpolation results of section 2, is a novel theoretical contribution. The second part consists of the techniques of sections 4 and 5, an illustration of how to use the proposed methodology. Although we do consider the results of sections 4 and 5 to be interesting and valuable, we view the technique, rather than the result, to be the major contribution of the second part.

To the best of our knowledge, the major and minor contributions of this work are novel in the following sense. The tightness of section 3 is new. The technique of section 4 is the first use of computer-assisted proofs to obtain provably tight rates for monotone operator splitting methods. The tight results of section 4 improve upon the prior results of [25, 53]. The technique of section 5 is the first use of automatic parameter selection that is optimal with respect to the algorithm and assumptions.

1.2. Prior work. FBS was first stated in the operator theoretic language in [7, 55]. The projected gradient method presented in [28, 44] served as a precursor to FBS. Peaceman–Rachford splitting (PRS) was first presented in [56, 34, 47], and DRS was first presented in [18, 47]. DYS was first presented in [16]. The forward-Douglas–Rachford splitting of Raguet, Fadili, and Peyré [61, 60] and Briceño-Arias [5] served as a precursor to DYS.

What we call interpolation in this work is also called extension. The maximal monotone extension theorem, which we later state as Fact 2.1, is well known, and it follows from a standard application of Zorn's lemma. Reich [62], Bauschke [1], Reich and Simons [63], Bauschke, Wang, and Yao [3, 4, 84], and Crouzeix, Anaya, and Sosa [13, 12, 11] have studied more concrete and constructive extension theorems for maximal monotone, nonexpansive, and firmly nonexpansive operators using tools from monotone operator theory.

Contraction factors and linear convergence for first-order methods have been a subject of intense study. Surprisingly, many of the published contraction factors are not tight. For FBS, Mercier, [51, p. 25], Tseng [78], Chen and Rockafellar [8], and Bauschke and Combettes [2, section 26.5] proved linear rates of convergence but did not provide exact matching lower bounds. Taylor, Hendrickx, and Glineur showed tight contraction factors and provided exact matching lower bounds [75]. For DRS, Lions and Mercier [47] and Davis and Yin [15] proved linear rates of convergence but did not provide exact matching lower bounds. Giselsson and Boyd [26, 27], Giselsson [24, 25], and Moursi and Vandenberghe [53] proved linear rates of convergence and provided exact matching lower bounds for certain cases. Alternating direction method of multipliers (ADMM) is a splitting method closely related to DRS. Deng and Yin [17], Giselsson and Boyd [26, 27], Nishihara et al. [54], França and Bento [21], Hong and Luo [33], Han, Sun, and Zhang [32], and Chen et al. [9] proved linear rates of convergence for ADMM. Matching lower bounds are provided only in [27]. Further, [24] provides matching lower bounds to the rates in [26]. Ghadimi et al. [22, 23] and Teixeira et al. [76, 77] proved linear rates of convergence and provided matching lower bounds for ADMM applied to quadratic problems. For DYS, Davis and Yin [16], Yan [85], Pedregosa and Gidel [59], and Pedregosa, Fatras, and Casotto [58] proved linear rates of convergence but did not provide exact matching lower bounds. Pedregosa [57] analyzed sublinear convergence, but not contraction factors.

Analyzing convex optimization algorithms by formulating the analysis as an SDP has been a rapidly growing area of research in the past five years. Past work analyzed convex optimization algorithms, and, to the best of our knowledge, analysis of the performance of monotone operator splitting methods with SDPs or any form of computer-assisted proof is new. (After the initial version of this paper appeared on arXiv [68], several papers citing our work followed up on our results and used SDPs to analyze other monotone operator splitting methods [29, 30, 31, 70, 83, 69].) Drori and Teboulle [20] and Taylor, Hendrickx, and Glineur [72, 74] presented the performance estimation problem (PEP) methodology. Our work generally follows the techniques presented by Drori and Teboulle [20] while contributing by establishing tightness. Lieder [45] applied the PEP approach to analyze the Halpern iteration without an a priori guarantee of tightness. Lessard, Recht, and Packard [43] leveraged techniques from control theory and used integral quadratic constraints (IQC) for finding Lyapunov functions for analyzing convex optimization algorithms. The IQC and PEP approaches were recently linked by Taylor, Van Scoy, and Lessard [71]. Finally, Nishihara et al. [54] and França and Bento [21] used IQC to analyze ADMM.

Finally, both IQC and PEP approaches allowed for the design of new methods for particular problem settings. For example, the optimized gradient method by Kim and Fessler [35, 36, 37, 38, 39, 40] (the first numerical version was given by Drori and Teboulle [20]) was developed using PEPs and enjoys the best possible worst-case guarantee on the final objective function accuracy after a fixed number of iterations, as shown by Drori [19]. On the other hand, the IQC framework was used by Van Scoy, Freeman, and Lynch [82] for developing the triple momentum method, which is the first-order method with the fastest known convergence rate for minimizing a smooth strongly convex function.

1.3. Preliminaries. We now quickly review standard results and set up the notation. We follow standard notation [66, 2]. Write $\mathcal{H}$ for a real Hilbert space equipped with a (symmetric) inner product $\langle \cdot, \cdot \rangle$. Write $\mathbb{S}^n_+$ for the set of $n \times n$ symmetric positive semidefinite matrices. Write $M \succeq 0$ if and only if $M \in \mathbb{S}^n_+$.

We say A is an operator on $\mathcal{H}$ and write $A \colon \mathcal{H} \rightrightarrows \mathcal{H}$ if A maps a point in $\mathcal{H}$ to a subset of $\mathcal{H}$. Thus $A(x) \subset \mathcal{H}$ for all $x \in \mathcal{H}$. For simplicity, we also write $Ax = A(x)$.

Write $I\colon \mathcal{H}\to\mathcal{H}$ for the identity operator. We say $A\colon \mathcal{H}\rightrightarrows\mathcal{H}$ is monotone if

$$\langle Ax - Ay, x - y\rangle \geq 0$$

for all $x, y \in \mathcal{H}$. To clarify, the inequality means $\langle u - v, x - y\rangle \geq 0$ for all $u \in Ax$ and $v \in Ay$. We say $A\colon \mathcal{H}\rightrightarrows\mathcal{H}$ is μ-strongly monotone if

$$\langle Ax - Ay, x - y\rangle \geq \mu\|x - y\|^2,$$

where $\mu \in (0,\infty)$. We say a single-valued operator $A\colon \mathcal{H}\to\mathcal{H}$ is β-cocoercive if

$$\langle Ax - Ay, x - y\rangle \geq \beta\|Ax - Ay\|^2,$$

where $\beta \in (0,\infty)$. We say a single-valued operator $A\colon \mathcal{H}\to\mathcal{H}$ is L-Lipschitz if

$$\|Ax - Ay\| \leq L\|x - y\|,$$

where $L \in (0,\infty)$. A monotone operator is maximal if it cannot be properly extended to another monotone operator. The resolvent of an operator A is $J_{\alpha A} = (I + \alpha A)^{-1}$, where $\alpha > 0$. We say a single-valued operator $T\colon \mathcal{H}\to\mathcal{H}$ is contractive if it is ρ-Lipschitz with $\rho < 1$. We say $x^\star$ is a fixed point of T if $x^\star = Tx^\star$.

DYS encodes solutions to

$$\underset{x\in\mathcal{H}}{\text{find}} \quad 0 \in (A + B + C)x,$$

where A, B, and C are maximal monotone and C is single-valued, as fixed points of

$$T(z; A, B, C, \alpha, \theta) = z - \theta J_{\alpha B}z + \theta J_{\alpha A}(2J_{\alpha B} - I - \alpha C J_{\alpha B})z, \tag{1.1}$$

where $\alpha > 0$ and $\theta \neq 0$. FBS and DRS are special cases of DYS: when $C = 0$, DYS reduces to DRS, and when $B = 0$, DYS reduces to FBS. Therefore, our analysis on DYS directly applies to FBS and DRS.

2. Operator interpolation.

Let $\mathcal{Q}$ be a class of operators, and let $\mathcal{I}$ be an arbitrary index set. We say a set of duplets $\{(x_i, q_i)\}_{i\in\mathcal{I}}$, where $x_i, q_i \in \mathcal{H}$ for all $i \in \mathcal{I}$, is $\mathcal{Q}$-*interpolable* if there is an operator $Q \in \mathcal{Q}$ such that $q_i \in Qx_i$ for all $i \in \mathcal{I}$. In this case, we call Q an *interpolation* of $\{(x_i, q_i)\}_{i\in\mathcal{I}}$. In this section, we present conditions that characterize when a set of duplets is interpolable with respect to the class of operators listed in Table 1 and their intersections.

Table 1

Operator classes for which we analyze interpolation. The parameters μ, L, and β are in $(0,\infty)$. Note that $\mathcal{M}_\mu \subset \mathcal{M}$ for any $\mu > 0$, $\mathcal{C}_\beta \subset \mathcal{M}$ for any $\beta > 0$, but $\mathcal{L}_L \not\subset \mathcal{M}$ for any $L > 0$.

Class	Description
$\mathcal{M}$	maximal monotone operators
$\mathcal{M}_\mu$	μ-strongly monotone maximal monotone operators
$\mathcal{L}_L$	L-Lipschitz operators
$\mathcal{C}_\beta$	β-cocoercive operators

2.1. Interpolation with one class. We now present interpolation results for the classes $\mathcal{M}$, $\mathcal{M}_\mu$, $\mathcal{L}_L$, and C_β.

FACT 2.1 (maximal monotone extension theorem [2, Theorem 20.21]). *$\{(x_i, q_i)\}_{i\in\mathcal{I}}$ is $\mathcal{M}$-interpolable if and only if*

$$\langle q_i - q_j, x_i - x_j\rangle \geq 0 \qquad \forall i, j \in \mathcal{I}.$$

PROPOSITION 2.1. *Let $\mu \in (0, \infty)$. Then $\{(x_i, q_i)\}_{i\in\mathcal{I}}$ is M_μ-interpolable if and only if*

$$\langle q_i - q_j, x_i - x_j\rangle \geq \mu\|x_i - x_j\|^2 \qquad \forall i, j \in \mathcal{I}.$$

Proof. With Fact 2.1, the proof follows from a sequence of equivalences:

$$\begin{aligned}
&\forall i, j \in \mathcal{I},\ \langle q_i - q_j, x_i - x_j\rangle \geq \mu\|x_i - x_j\|^2 \\
&\qquad\Leftrightarrow\quad \forall i, j \in \mathcal{I}, \langle (q_i - \mu x_i) - (q_j - \mu x_j), x_i - x_j\rangle \geq 0 \\
&\qquad\Leftrightarrow\quad \exists R \in \mathcal{M}, \forall i \in \mathcal{I}, (q_i - \mu x_i) \in Rx_i \\
&\qquad\Leftrightarrow\quad \exists Q \in \mathcal{M}_\mu, Q = R + \mu I,\ \forall i \in \mathcal{I},\ q_i \in Qx_i. \qquad \square
\end{aligned}$$

PROPOSITION 2.2. *Let $\beta \in (0, \infty)$. Then $\{(x_i, q_i)\}_{i\in\mathcal{I}}$ is $\mathcal{C}_\beta$-interpolable if and only if*

$$\langle q_i - q_j, x_i - x_j\rangle \geq \beta\|q_i - q_j\|^2 \qquad \forall i, j \in \mathcal{I}.$$

Proof. With Proposition 2.1, the proof follows from a sequence of equivalences:

$$\begin{aligned}
&\forall i, j \in \mathcal{I},\ \langle q_i - q_j, x_i - x_j\rangle \geq \beta\|q_i - q_j\|^2 \\
&\qquad\Leftrightarrow\quad \exists R \in \mathcal{M}_\beta, \forall i \in \mathcal{I},\ x_i \in Rq_i \\
&\qquad\Leftrightarrow\quad \exists Q \in \mathcal{C}_\beta, Q = R^{-1},\ \forall i \in \mathcal{I},\ q_i \in Qx_i. \qquad \square
\end{aligned}$$

FACT 2.2 (Kirszbraun–Valentine theorem). *Let $L \in (0, \infty)$. Then $\{(x_i, q_i)\}_{i\in\mathcal{I}}$ is $\mathcal{L}_L$-interpolable if and only if*

$$\|q_i - q_j\|^2 \leq L^2\|x_i - x_j\|^2 \qquad \forall i, j \in \mathcal{I}.$$

Fact 2.2 is a special case of the Kirszbraun–Valentine theorem [41, 80, 81]. A direct proof follows from similar arguments.

2.2. Failure of interpolation with intersection of classes. When considering interpolation with intersections of classes such as $\mathcal{M}\cap\mathcal{L}_L$, one might naively expect results as simple as those of section 2.1. Contrary to this expectation, interpolation can fail.

PROPOSITION 2.3. *$\{(x_i, q_i)\}_{i\in\mathcal{I}}$ may not be $(\mathcal{M} \cap \mathcal{L}_L)$-interpolable for $L \in (0, \infty)$ even if*

$$\|q_i - q_j\|^2 \leq L^2\|x_i - x_j\|^2, \qquad \langle q_i - q_j, x_i - x_j\rangle \geq 0 \qquad \forall i, j \in \mathcal{I}.$$

Proof. Consider the following example in $\mathbb{R}^2$:

$$S = \left\{ \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \left(\begin{bmatrix} 1/2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ L/2 \end{bmatrix} \right) \right\}.$$

These points satisfy the inequalities. However, there is no Lipschitz and maximal monotone operator interpolating these points. Assume for contradiction that $Q \in (\mathcal{M} \cap \mathcal{L}_L)$ is an interpolation of these points. Since Q is Lipschitz, it is single-valued. Since Q is maximal monotone, the set $\{x \,|\, Qx = 0\}$ is convex [2, Proposition 23.39]. This implies $Q(1/2, 0) = (0, 0)$, which is a contradiction. ∎

The subtlety is that the counterexample has two separate interpolations in $\mathcal{M}$ and $\mathcal{L}_L$ but does not have an interpolation in $\mathcal{M} \cap \mathcal{L}_L$. Interpolation with respect to $\mathcal{M}_\mu \cap \mathcal{L}_L$, $\mathcal{C}_\beta \cap \mathcal{L}_L$, and $\mathcal{M}_\mu \cap \mathcal{C}_\beta$ can fail in a similar manner.

2.3. Two-point interpolation. We now present conditions for two-point interpolation, i.e., interpolation when $|\mathcal{I}| = 2$. In this case, interpolation conditions become simple, and the difficulty discussed in section 2.2 disappears. Although the setup $|\mathcal{I}| = 2$ may seem restrictive, it is sufficient for what we need in later sections.

PROPOSITION 2.4. *Assume $0 < \mu$, $\mu \le L < \infty$, and $\mu \le 1/\beta < \infty$. Then $\{(x_1, q_1), (x_2, q_2)\}$ is $(\mathcal{M}_\mu \cap \mathcal{C}_\beta \cap \mathcal{L}_L)$-interpolable if and only if*

$$\begin{aligned} \langle q_1 - q_2, x_1 - x_2 \rangle &\ge \mu \|x_1 - x_2\|^2, \\ \langle q_1 - q_2, x_1 - x_2 \rangle &\ge \beta \|q_1 - q_2\|^2, \\ \|q_1 - q_2\|^2 &\le L^2 \|x_1 - x_2\|^2. \end{aligned} \tag{2.1}$$

Proof. If the points are $(\mathcal{M}_\mu \cap \mathcal{L}_L \cap \mathcal{C}_\beta)$-interpolable, then (2.1) holds by definition. Assume (2.1) holds. When $\dim \mathcal{H} = 1$, the result is trivial, so we assume, without loss of generality, $\dim \mathcal{H} \ge 2$.

Define $q = q_1 - q_2$ and $x = x_1 - x_2$. If $x = 0$, then $\beta > 0$ or $L > 0$ implies $q = 0$, and the operator $Q : \mathcal{H} \to \mathcal{H}$ defined as

$$Q(y) = \mu(y - x_1) + q_1$$

interpolates $\{(x_1, q_1), (x_2, q_2)\}$ and $Q \in \mathcal{M}_\mu \cap \mathcal{L}_L \cap \mathcal{C}_\beta$. Assume $x \ne 0$. If $q = \gamma x$ for some $\gamma \in \mathbb{R}$, then the operator $Q : \mathcal{H} \to \mathcal{H}$ defined as

$$Q(y) = \gamma(y - x_1) + q_1$$

interpolates $\{(x_1, q_1), (x_2, q_2)\}$ and $Q \in \mathcal{M}_\mu \cap \mathcal{L}_L \cap \mathcal{C}_\beta$. Assume q is linearly independent from x. Define the orthonormal vectors,

$$e_1 = \frac{1}{\|x\|} x, \qquad e_2 = \frac{1}{\sqrt{\|q\|^2 - (\langle e_1, q \rangle)^2}} (q - \langle e_1, q \rangle e_1),$$

along with an associated bounded linear operator $A : \mathcal{H} \to \mathcal{H}$ such that

$$A|_{\{e_1, e_2\}^\perp} = \mu I,$$

where $\{e_1, e_2\}^\perp \subset \mathcal{H}$ is the subspace orthogonal to e_1 and e_2, and I is the identity mapping on $\{e_1, e_2\}^\perp$. On $\operatorname{span}\{e_1, e_2\}$, define

$$Ae_1 = \frac{\langle q, e_1 \rangle}{\|x\|} e_1 + \frac{\sqrt{\|q\|^2 - (\langle e_1, q \rangle)^2}}{\|x\|} e_2,$$

$$Ae_2 = -\frac{\sqrt{\|q\|^2 - (\langle e_1, q \rangle)^2}}{\|x\|} e_1 + \frac{\langle q, e_1 \rangle}{\|x\|} e_2.$$

Note that this definition satisfies $Ax = q$. Finally, define M to be a 2×2 matrix isomorphic to $A|_{\mathrm{span}\{e_1,e_2\}}$, i.e.,

$$A|_{\mathrm{span}\{e_1,e_2\}} \cong \underbrace{\frac{1}{\|x\|}\begin{bmatrix} \langle q, e_1\rangle & -\sqrt{\|q\|^2 - (\langle e_1, q\rangle)^2} \\ \sqrt{\|q\|^2 - (\langle e_1, q\rangle)^2} & \langle q, e_1\rangle \end{bmatrix}}_{=M} \in \mathbb{R}^{2\times 2}.$$

With direct computations, we can verify that M satisfies

$$\begin{aligned} L^2 &\geq \lambda_{\max}(M^TM) = \frac{\|q\|^2}{\|x\|^2}, \\ \mu &\leq \lambda_{\min}((1/2)(M + M^T)) = \frac{\langle q, x\rangle}{\|x\|^2}, \\ \beta &\leq \lambda_{\min}((1/2)(M^{-1} + M^{-T})) = \frac{\langle q, x\rangle}{\|q\|^2}. \end{aligned}$$

This implies that $A : \mathcal{H} \to \mathcal{H}$ is L-Lipschitz, μ-strongly monotone, and β-cocoercive. Finally, the affine operator $Q : \mathcal{H} \to \mathcal{H}$ defined as

$$Q(y) = A(y - x_1) + q_1$$

interpolates $\{(x_1, q_1), (x_2, q_2)\}$ and $Q \in \mathcal{M}_\mu \cap \mathcal{L}_L \cap \mathcal{C}_\beta$. ∎

Proposition 2.4 presents conditions for interpolation with three classes. Interpolation conditions with two of these classes, such as $(\mathcal{C}_\beta \cap \mathcal{L}_L)$, $(\mathcal{M}_\mu \cap \mathcal{C}_\beta)$, $(\mathcal{M}_\mu \cap \mathcal{L}_L)$, $(\mathcal{M} \cap \mathcal{L}_L)$, are of the same form and follow from a very similar (almost identical) proof.

3. Operator splitting performance estimation problems.

Consider the *operator splitting performance estimation problem* (OSPEP)

$$(3.1) \qquad \begin{array}{ll} \text{maximize} & \dfrac{\|T(z; A, B, C, \alpha, \theta) - T(z'; A, B, C, \alpha, \theta)\|^2}{\|z - z'\|^2} \\ \text{subject to} & A \in \mathcal{Q}_1,\ B \in \mathcal{Q}_2,\ C \in \mathcal{Q}_3, \\ & z, z' \in \mathcal{H},\ z \neq z', \end{array}$$

where z, z', A, B, and C are the optimization variables. T is the DYS operator defined in (1.1). The scalars $\alpha > 0$ and $\theta > 0$ and the classes $\mathcal{Q}_1$, $\mathcal{Q}_2$, and $\mathcal{Q}_3$ are problem data. Assume that each class $\mathcal{Q}_1$, $\mathcal{Q}_2$, and $\mathcal{Q}_3$ is a single operator class of Table 1 or is an intersection of classes of Table 1. (Thus the reader can freely pick the assumptions; the minimal assumptions are that $\mathcal{Q}_1$, $\mathcal{Q}_2$, and $\mathcal{Q}_3$ are monotone.)

By definition, ρ is a valid contraction factor if and only if

$$\rho^2 \geq \sup_{\substack{A\in\mathcal{Q}_1, B\in\mathcal{Q}_2, C\in\mathcal{Q}_3, \\ z,z'\in\mathcal{H},\ z\neq z'}} \frac{\|T(z; A, B, C, \alpha, \theta) - T(z'; A, B, C, \alpha, \theta)\|^2}{\|z - z'\|^2}.$$

Therefore, the OSPEP, by definition, computes the square of the best contraction factor of T given the assumptions on A, B, and C encoded as the classes $\mathcal{Q}_1$, $\mathcal{Q}_2$, and $\mathcal{Q}_3$. In fact, we say a contraction factor (established through a proof) is *tight* if it is equal to the square root of the optimal value of (3.1). A contraction factor that is not tight can be improved with a better proof without any further assumptions.

At first sight, (3.1) seems difficult to solve, as it is posed as an infinite-dimensional nonconvex optimization problem. In this section, we present a reformulation of (3.1) into a (finite-dimensional convex) SDP. This reformulation is exact; it performs no relaxations or approximations, and the optimal value of the SDP coincides with that of (3.1).

3.1. Convex formulation of OSPEP. We now formulate (3.1) into a (finite-dimensional) convex SDP through a series of equivalent transformations. First, we write (3.1) more explicitly as

$$\begin{array}{ll} \text{maximize} & \dfrac{\|z-\theta(z_B-z_A)-z'+\theta(z'_B-z'_A)\|^2}{\|z-z'\|^2} \\ \text{subject to} & A\in\mathcal{Q}_1,\ B\in\mathcal{Q}_2,\ C\in\mathcal{Q}_3, \\ & z_B=J_{\alpha B}z, \\ & z_C=\alpha C z_B, \\ & z_A=J_{\alpha A}(2z_B-z-z_C), \\ & z'_B=J_{\alpha B}z', \\ & z'_C=\alpha C z'_B, \\ & z'_A=J_{\alpha A}(2z'_B-z'-z'_C), \\ & z,z'\in\mathcal{H},\ z\neq z', \end{array} \tag{3.2}$$

where $z, z' \in \mathcal{H}$, A, B, and C are the optimization variables.

3.1.1. Homogeneity. We say a class of operators $\mathcal{Q}$ is *homogeneous* if

$$A\in\mathcal{Q} \quad\Leftrightarrow\quad (\gamma^{-1}I)A(\gamma I)\in\mathcal{Q}$$

for all $\gamma > 0$. All operator classes of Table 1 are homogeneous. Since $\mathcal{Q}_1$, $\mathcal{Q}_2$, and $\mathcal{Q}_3$ are homogeneous, we can use the change of variables $z \mapsto \gamma^{-1}z$, $z' \mapsto \gamma^{-1}z'$, $A \mapsto (\gamma^{-1}I)A(\gamma I)$, $B \mapsto (\gamma^{-1}I)B(\gamma I)$, and $C \mapsto (\gamma^{-1}I)C(\gamma I)$, where $\gamma = \|z - z'\|$, to equivalently reformulate (3.2) into

$$\begin{array}{ll} \text{maximize} & \|z-\theta(z_B-z_A)-z'+\theta(z'_B-z'_A)\|^2 \\ \text{subject to} & A\in\mathcal{Q}_1,\ B\in\mathcal{Q}_2,\ C\in\mathcal{Q}_3, \\ & z_B=J_{\alpha B}z, \\ & z_C=\alpha C z_B, \\ & z_A=J_{\alpha A}(2z_B-z-z_C), \\ & z'_B=J_{\alpha B}z', \\ & z'_C=\alpha C z'_B, \\ & z'_A=J_{\alpha A}(2z'_B-z'-z'_C), \\ & \|z-z'\|^2=1, \end{array} \tag{3.3}$$

where $z, z' \in \mathcal{H}$, A, B, and C are the optimization variables.

3.1.2. Operator interpolation. For simplicity of exposition, we limit the generality and reformulate the convex SDP under the following operator classes:

- $A \in \mathcal{Q}_1 = \mathcal{M}_\mu$—$\mu$-strongly maximal monotone,
- $B \in \mathcal{Q}_2 = \mathcal{C}_\beta \cap \mathcal{L}_L$—$\beta$-cocoercive and L-Lipschitz,
- $C \in \mathcal{Q}_3 = \mathcal{C}_{\beta_C}$—$\beta_C$-cocoercive.

To clarify, the same analysis can be done in the general setup, and *we can freely pick and choose the assumptions.* The general result is shown in the supplementary materials [68, section SM1].

We use the interpolation results from section 2. For operator A, we have

$$\begin{aligned}
&\exists A \in \mathcal{M}_\mu \text{ such that } z_A = J_{\alpha A}(2z_B - z - z_C),\ z'_A = J_{\alpha A}(2z'_B - z' - z'_C)\\
&\Leftrightarrow \{(z_A, \alpha^{-1}(2z_B - z - z_C - z_A)), (z'_A, \alpha^{-1}(2z'_B - z' - z'_C - z'_A))\} \text{ is } \mathcal{M}_\mu\text{-interpolable}\\
&\Leftrightarrow \langle z_A - z'_A, 2z_B - z - z_C - (2z'_B - z' - z'_C)\rangle \geq (1 + \alpha\mu)\|z_A - z'_A\|^2.
\end{aligned}$$

For operator B, we have

$$\begin{aligned}
&\exists B \in \mathcal{C}_\beta \cap \mathcal{L}_L \text{ such that } z_B = J_{\alpha B} z,\ z'_B = J_{\alpha B} z'\\
&\quad\Leftrightarrow \quad \{(z_B, \alpha^{-1}(z - z_B)), (z'_B, \alpha^{-1}(z' - z'_B))\} \text{ is } \mathcal{C}_\beta\text{-interpolable},\\
&\qquad\quad\ \{(z_B, \alpha^{-1}(z - z_B)), (z'_B, \alpha^{-1}(z' - z'_B))\} \text{ is } \mathcal{L}_L\text{-interpolable}\\
&\quad\Leftrightarrow \quad \langle z - z' - z_B + z'_B, z_B - z'_B\rangle \geq (\beta/\alpha)\|z - z' - z_B + z'_B\|^2,\\
&\qquad\quad\ \alpha^2 L^2 \|z_B - z'_B\|^2 \geq \|z - z' - z_B + z'_B\|^2.
\end{aligned}$$

For operator C, we have

$$\begin{aligned}
&\exists C \in \mathcal{C}_{\beta_C} \text{ such that } z_C = \alpha C z_B,\ z'_C = \alpha C z'_B\\
&\quad\Leftrightarrow \quad \{(z_B, \alpha^{-1} z_C), (z'_B, \alpha^{-1} z'_C)\} \text{ is } \mathcal{C}_{\beta_C}\text{-interpolable}\\
&\quad\Leftrightarrow \quad \langle z_B - z'_B, z_C - z'_C\rangle \geq (\beta_C/\alpha)\|z_C - z'_C\|^2.
\end{aligned}$$

Now we can drop the explicit dependence on the operators A, B, and C and reformulate (3.3) into

$$\begin{array}{ll}
\text{maximize} & \|z - \theta(z_B - z_A) - z' + \theta(z'_B - z'_A)\|^2\\
\text{subject to} & \langle z_A - z'_A, 2z_B - z - z_C - (2z'_B - z' - z'_C)\rangle \geq (1 + \alpha\mu)\|z_A - z'_A\|^2,\\
& \langle z - z' - z_B + z'_B, z_B - z'_B\rangle \geq (\beta/\alpha)\|z - z - z_B + z'_B\|^2,\\
& \alpha^2 L^2\|z_B - z'_B\|^2 \geq \|z - z' - z_B + z'_B\|^2,\\
& \langle z_B - z'_B, z_C - z'_C\rangle \geq (\beta_C/\alpha)\|z_C - z'_C\|^2,\\
& \|z - z'\|^2 = 1,
\end{array}$$

where $z, z', z_A, z'_A, z_B, z'_B, z_C, z'_C \in \mathcal{H}$ are the optimization variables. Since the variables only appear as differences between the primed and nonprimed variables, we can perform a change of variables $z - z' \mapsto z$, $z_A - z'_A \mapsto z_A$, $z_B - z'_B \mapsto z_B$, and $z_C - z'_C \mapsto z_C$ to get

$$\begin{array}{ll}
\text{maximize} & \|z - \theta(z_B - z_A)\|^2\\
\text{subject to} & \langle z_A, 2z_B - z - z_C\rangle \geq (1 + \alpha\mu)\|z_A\|^2,\\
& \langle z - z_B, z_B\rangle \geq (\beta/\alpha)\|z - z_B\|^2,\\
& \alpha^2 L^2 \|z_B\|^2 \geq \|z - z_B\|^2,\\
& \langle z_B, z_C\rangle \geq (\beta_C/\alpha)\|z_C\|^2,\\
& \|z\|^2 = 1,
\end{array} \tag{3.4}$$

where $z, z_A, z_B, z_C \in \mathcal{H}$ are the optimization variables.

3.1.3. Grammian representation. The optimization problem (3.4) and all other operator classes in section 2 are specified through inner products and squared norms. This structure allows us to rewrite the problem with a Grammian representation:

$$G = \begin{pmatrix} \|z\|^2 & \langle z, z_A\rangle & \langle z, z_B\rangle & \langle z, z_C\rangle\\ \langle z, z_A\rangle & \|z_A\|^2 & \langle z_A, z_B\rangle & \langle z_A, z_C\rangle\\ \langle z, z_B\rangle & \langle z_A, z_B\rangle & \|z_B\|^2 & \langle z_B, z_C\rangle\\ \langle z, z_C\rangle & \langle z_A, z_C\rangle & \langle z_B, z_C\rangle & \|z_C\|^2 \end{pmatrix}. \tag{3.5}$$

LEMMA 3.1. *If* $\dim \mathcal{H} \geq 4$, *then*

$$G \in \mathbb{S}^4_+ \quad \Leftrightarrow \quad \exists z, z_A, z_B, z_C \in \mathcal{H} \text{ such that } G = \text{expression of } (3.5).$$

Proof. ($\Leftarrow$) For any $z, z_A, z_B, z_C \in \mathcal{H}$, G is positive semidefinite since

$$x^T G x = \|x_1 z + x_2 z_A + x_3 z_B + x_4 z_C\|^2 \geq 0$$

for any $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$.

($\Rightarrow$) Let $LL^T = G$ be a Cholesky factorization of G. Write

$$L = \begin{bmatrix} \tilde{z}^T \\ \tilde{z}_A^T \\ \tilde{z}_B^T \\ \tilde{z}_C^T \end{bmatrix},$$

where $\tilde{z}, \tilde{z}_A, \tilde{z}_B, \tilde{z}_C \in \mathbb{R}^4$. We can find orthonormal vectors $e_1, e_2, e_3, e_4 \in \mathcal{H}$ since $\dim \mathcal{H} \geq 4$. Define

$$z = \tilde{z}_1 e_1 + \tilde{z}_2 e_2 + \tilde{z}_3 e_3 + \tilde{z}_4 e_4, \qquad z_A = (\tilde{z}_A)_1 e_1 + (\tilde{z}_A)_2 e_2 + (\tilde{z}_A)_3 e_3 + (\tilde{z}_A)_4 e_4.$$

Define $z_B, z_C \in \mathcal{H}$ similarly. Then G is as given by (3.5) with the constructed $z, z_A, z_B, z_C \in \mathcal{H}$. ∎

Write

$$M_I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad M_O = \begin{pmatrix} 1 & \theta & -\theta & 0 \\ \theta & \theta^2 & -\theta^2 & 0 \\ -\theta & -\theta^2 & \theta^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$M_\mu^A = \begin{pmatrix} 0 & -1/2 & 0 & 0 \\ -1/2 & -1-\alpha\mu & 1 & -1/2 \\ 0 & 1 & 0 & 0 \\ 0 & -1/2 & 0 & 0 \end{pmatrix}, \qquad M_\beta^C = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & -\beta_C/\alpha \end{pmatrix},$$

$$M_\beta^B = \begin{pmatrix} -\beta/\alpha & 0 & 1/2+\beta/\alpha & 0 \\ 0 & 0 & 0 & 0 \\ 1/2+\beta/\alpha & 0 & -1-\beta/\alpha & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad M_L^B = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & -1+\alpha^2 L^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

When $\dim \mathcal{H} \geq 4$, we can use Lemma 3.1 to reformulate (3.4) into the equivalent SDP

$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}(M_O G) \\ \text{subject to} & \mathrm{Tr}(M_\mu^A G) \geq 0, \\ & \mathrm{Tr}(M_\beta^B G) \geq 0, \\ & \mathrm{Tr}(M_L^B G) \geq 0, \\ & \mathrm{Tr}(M_\beta^C G) \geq 0, \\ & \mathrm{Tr}(M_I G) = 1, \\ & G \succeq 0, \end{array} \tag{3.6}$$

where $G \in \mathbb{S}^4_+$ is the optimization variable. Since (3.6) is a finite-dimensional convex SDP, we can solve it efficiently with standard solvers.

These equivalent reformulations prove Theorem 3.2 for this special case. The general case follows from analogous steps, and we show the fully general SDP in the supplementary materials [68, section SM1].

THEOREM 3.2. *The OSPEP* (3.1) *and the SDP of section SM*1 *in* [68] *are equivalent if* $\dim \mathcal{H} \geq 4$ *and* $\mathcal{Q}_1 = \mathcal{M}_{\mu_A} \cap \mathcal{C}_{\beta_A} \cap \mathcal{L}_{L_A}$, $\mathcal{Q}_2 = \mathcal{M}_{\mu_B} \cap \mathcal{C}_{\beta_B} \cap \mathcal{L}_{L_B}$, *and* $\mathcal{Q}_3 = \mathcal{M}_{\mu_C} \cap \mathcal{C}_{\beta_C} \cap \mathcal{L}_{L_C}$.

To clarify, Theorem 3.2 states that the optimal values of the two problems are equal and that a solution from one problem can be transformed into a solution of the other. Given an optimal $G^\star$ of the SDP, we can take its Cholesky factorization as in Lemma 3.1 to get $z, z_A, z_B, z_C \in \mathcal{H}$ and obtain evaluations of the worst-case operators

$$A(z_A) \ni \alpha^{-1}(2z_B - z - z_C - z_A), \quad A(0) \ni 0, \quad \text{where } A \in \mathcal{Q}_1$$
$$B(z_B) \ni \alpha^{-1}(z - z_B), \quad B(0) \ni 0, \quad \text{where } B \in \mathcal{Q}_2$$
$$C(z_B) \ni \alpha^{-1} z_C, \quad C(0) \ni 0, \quad \text{where } C \in \mathcal{Q}_3.$$

3.2. Dual OSPEP. The SDP (3.6) has a dual,

$$\begin{array}{ll} \text{minimize} & \rho^2 \\ \text{subject to} & \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C \geq 0, \\ & S(\rho^2, \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C, \theta, \alpha) \succeq 0, \end{array} \tag{3.7}$$

where $\rho^2, \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C \in \mathbb{R}$ are the optimization variables, and

(3.8)
$$S(\rho^2, \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C, \theta, \alpha) = -M_O - \lambda_\mu^A M_\mu^A - \lambda_\beta^B M_\beta^B - \lambda_L^B M_L^B - \lambda_\beta^C M_\beta^C + \rho^2 M_I$$
$$= \begin{pmatrix} \rho^2 + \frac{\lambda_\beta^B \beta}{\alpha} + \lambda_L^B - 1 & \frac{\lambda_\mu^A}{2} - \theta & -\lambda_\beta^B(\frac{1}{2} + \frac{\beta}{\alpha}) - \lambda_L^B + \theta & 0 \\ \frac{\lambda_\mu^A}{2} - \theta & \lambda_\mu^A(1 + \alpha\mu) - \theta^2 & -\lambda_\mu^A + \theta^2 & \frac{\lambda_\mu^A}{2} \\ -\lambda_\beta^B(\frac{1}{2} + \frac{\beta}{\alpha}) - \lambda_L^B + \theta & -\lambda_\mu^A + \theta^2 & \lambda_\beta^B(\frac{\beta}{\alpha} - 1) + \lambda_L^B(1 - \alpha^2 L^2) - \theta^2 & -\frac{\lambda_\beta^C}{2} \\ 0 & \frac{\lambda_\mu^A}{2} & -\frac{\lambda_\beta^C}{2} & \frac{\lambda_\beta^C \beta_C}{\alpha} \end{pmatrix}.$$

We call (3.7) the *dual OSPEP*. In contrast, we call the OSPEP (3.1) and, equivalently, (3.6), the *primal OSPEP*. Again, this special case illustrates the overall approach. We show the fully general dual OSPEP in the supplementary materials [68, section SM2].

To ensure strong duality between the primal and dual OSPEPs, we enforce Slater's constraint qualification with the following notion of degeneracy. We say the intersections $\mathcal{C}_\beta \cap \mathcal{L}_L$, $\mathcal{M}_\mu \cap \mathcal{C}_\beta$, $\mathcal{M}_\mu \cap \mathcal{L}_L$, and $\mathcal{M}_\mu \cap \mathcal{C}_\beta \cap \mathcal{L}_L$ are, respectively, *degenerate* if $\mathcal{C}_{\beta+\varepsilon} \cap \mathcal{L}_{L-\varepsilon} = \emptyset$, $\mathcal{M}_{\mu+\varepsilon} \cap \mathcal{C}_{\beta+\varepsilon} = \emptyset$, $\mathcal{M}_{\mu+\varepsilon} \cap \mathcal{L}_{L-\varepsilon} = \emptyset$, and $\mathcal{M}_{\mu+\varepsilon} \cap \mathcal{C}_{\beta+\varepsilon} \cap \mathcal{L}_{L-\varepsilon} = \emptyset$ for all $\varepsilon > 0$. For example, $\mathcal{M}_3 \cap \mathcal{L}_3 = \{3I\}$ is a degenerate intersection.

THEOREM 3.3. *Weak duality holds between the primal and dual OSPEPs of sections SM*1 *and SM*2 *of* [68]. *Furthermore, strong duality holds if each class* $\mathcal{Q}_1$, $\mathcal{Q}_2$, *and* $\mathcal{Q}_3$ *is a nondegenerate intersection of classes of Table* 1.

Proof. Weak duality follows from the fact that the SDP of [68, section SM2] is the Lagrange dual of the SDP of [68, section SM1]. To establish strong duality, we show that the nondegeneracy assumption leads to Slater's constraint qualification [65] for the primal OSPEP.

Since the intersections are nondegenerate, there is a small $\varepsilon > 0$ and A, B, and C such that

$$A \in \mathcal{M}_{\mu_A+\varepsilon} \cap \mathcal{C}_{\beta_A+\varepsilon} \cap \mathcal{L}_{L_A-\varepsilon},$$
$$B \in \mathcal{M}_{\mu_B+\varepsilon} \cap \mathcal{C}_{\beta_B+\varepsilon} \cap \mathcal{L}_{L_B-\varepsilon},$$
$$C \in \mathcal{M}_{\mu_C+\varepsilon} \cap \mathcal{C}_{\beta_C+\varepsilon} \cap \mathcal{L}_{L_C-\varepsilon}.$$

With any inputs $z, z' \in \mathcal{H}$ such that $z \neq z'$, we can follow the arguments of section 3.1 and construct a G matrix as defined in (3.5). This G satisfies

$$\mathrm{Tr}(M_\mu^A G) > 0, \ldots, \quad \mathrm{Tr}(M_L^C G) > 0, \quad \mathrm{Tr}(M_I G) = 1, \quad G \succeq 0.$$

Define $G_\delta = (1-\delta)G + \delta I$. There exists a small $\delta > 0$ such that

$$\mathrm{Tr}(M_\mu^A G_\delta) > 0, \ldots, \quad \mathrm{Tr}(M_L^C G_\delta) > 0, \quad \mathrm{Tr}(M_I G_\delta) = 1, \quad G_\delta \succ 0.$$

Note that the equality constraint $\mathrm{Tr}(M_I G_\delta) = 1$ holds since $\mathrm{Tr}(M_I) = 1$. Since G_δ is a strictly feasible point, Slater's condition gives us strong duality. ∎

More generally, the strong duality argument of Theorem 3.3 applies if each $\mathcal{Q}_1$, $\mathcal{Q}_2$, and $\mathcal{Q}_3$ is a single operator class of Table 1 or is a nondegenerate intersection of those classes.

3.3. Primal and dual interpretations and computer-assisted proofs. A feasible point of the primal OSPEP provides a lower bound on any contraction factor, as it corresponds to operator instances that exhibit a contraction corresponding to the objective value. An optimal point of the primal OSPEP corresponds to the worst-case operators. A feasible point of the dual OSPEP provides an upper bound as it corresponds to a proof of a contraction factor. A convergence proof in optimization is a nonnegative combination of known valid inequalities. The nonnegative variables of the dual OSPEP correspond to weights of such a nonnegative combination, and the objective value is the contraction factor that the nonnegative combination of inequalities (i.e., the proof) proves.

We can use the OSPEP methodology as a tool for computer-assisted proofs. Given the operator classes, we can choose specific numerical values for the parameters, such as the strong convexity and cocoercivity parameters, and numerically solve the SDP. We do this for many parameter values, observe the pattern of primal and dual solutions, and guess the analytical, parameterized solution to the SDPs. To put it differently, the SDP solver provides a valid and optimal proof for a given choice of parameters, and we use this to infer the general proof for all parameter choices.

3.4. Further remarks. With analogous steps, the OSPEPs for FBS and DRS can be written as smaller 3×3 SDPs. Using the smaller SDP is preferred, as formulating these cases into larger 4×4 SDPs, as a special case of the 4×4 SDP for DYS, can lead to numerical difficulties.

The tightness of the OSPEP methodology relies on the two-point interpolation results of section 2, which we can use because the operators A, B, and C are evaluated *once* per iteration. (To analyze the contraction factor, we consider a single evaluation of the operator at two distinct points, which leads to two evaluations of each operator.) For splitting methods without this property, methods that access one of the operators two or more times per iteration, the OSPEP loses the tightness guarantee. Such methods include the extragradient method [42], FBF [79], PDFP [10], extragradient-based alternating direction method for convex minimization [46], FBHF [6], FRB [50], the Golden ratio algorithm [49], shadow-Douglas–Rachford [14], and BFRB/BRFB [64]. Nevertheless, the OSPEP is applicable for analyzing these types of methods and, in particular, can be used to find the convergence proofs presented in these references.

4. Tight analytic contraction factors for DRS. In this section, we present tight analytic contraction factors for DRS under two sets of assumptions considered in [25, 53]. The primary purpose of this section is to demonstrate the strength of the

OSPEP methodology through proving results that are likely too complicated for a human to find bare-handed. The proofs are computer-assisted in that their discoveries were assisted by a computer, but their verifications do not require a computer.

The results below are presented for $\alpha = 1$. The general rate for $\alpha > 0$ follows from the scaling $\mu \mapsto \alpha\mu$, $\beta \mapsto \beta/\alpha$, and $L \mapsto \alpha L$. The proofs are presented in the supplementary materials [68, section SM3].

THEOREM 4.1. *Let $A \in \mathcal{M}_\mu$ and $B \in \mathcal{C}_\beta$ with $\mu, \beta > 0$, and assume* $\dim \mathcal{H} \geq 3$. *The* tight *contraction factor of the DRS operator $I - \theta J_B + \theta J_A(2J_B - I)$ for $\theta \in (0, 2)$ is*

$$\rho = \begin{cases} |1 - \theta\frac{\beta}{\beta+1}| & \text{if } \mu\beta - \mu + \beta < 0 \text{ and } \theta \leq 2\frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \\ |1 - \theta\frac{1+\mu\beta}{(\mu+1)(\beta+1)}| & \text{if } \mu\beta - \mu - \beta > 0 \text{ and } \theta \leq 2\frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \\ |1-\theta| & \text{if } \theta \geq 2\frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}, \\ |1 - \theta\frac{\mu}{\mu+1}| & \text{if } \mu\beta + \mu - \beta < 0 \text{ and } \theta \leq 2\frac{(\mu+1)(\beta-\mu-\mu\beta)}{\beta+\mu\beta-\mu-\mu^2-2\mu^2\beta}, \\ \rho_5 & \text{otherwise,} \end{cases}$$

with

$$\rho_5 = \frac{\sqrt{2-\theta}}{2}\sqrt{\frac{((2-\theta)\mu(\beta+1)+\theta\beta(1-\mu))\,((2-\theta)\beta(\mu+1)+\theta\mu(1-\beta))}{\mu\beta(2\mu\beta(1-\theta)+(2-\theta)(\mu+\beta+1))}}.$$

(In the first, second, and fourth cases, the former parts of the conditions ensure that there is no division by 0 in the latter parts. We show this in [68, section SM4.1.1, case (a) part (ii), case (b) part (ii), and case (d) part (ii)].)

COROLLARY 4.2. *Let $A \in \mathcal{M}_\mu$ and $B \in \mathcal{C}_\beta$ with $\mu, \beta > 0$, and assume* $\dim \mathcal{H} \geq 3$. *The* tight *contraction factor of the DRS operator $I - J_B + J_A(2J_B - I)$ is*

$$\rho = \begin{cases} |1 - \frac{\beta}{\beta+1}| & \text{if } \beta^2+\mu\beta+\beta-\mu \leq 0, \\ |1 - \frac{1+\mu\beta}{(\mu+1)(\beta+1)}| & \text{if } \mu\beta - \mu - \beta \geq 1, \\ |1 - \frac{\mu}{\mu+1}| & \text{if } \mu^2 + \mu\beta + \mu - \beta \leq 0, \\ \frac{1}{2}\frac{\beta+\mu}{\sqrt{\beta\mu(\beta+\mu+1)}} & \text{otherwise.} \end{cases}$$

Proof. Plug $\theta = 1$ into Theorem 4.1 and simplify. We omit the details. ☐

THEOREM 4.3. *Let $A \in \mathcal{M}_\mu$ and $B \in \mathcal{M} \cap \mathcal{L}_L$ with $\mu, L > 0$, and assume* $\dim \mathcal{H} \geq 3$. *The* tight *contraction factor of the DRS operator $I - \theta J_B + \theta J_A(2J_B - I)$ for $\theta \in (0, 2)$ is*

$$\rho = \begin{cases} \frac{\theta + \sqrt{\frac{(2(\theta-1)\mu+\theta-2)^2+L^2(\theta-2(\mu+1))^2}{L^2+1}}}{2(\mu+1)} & \text{if (a)}, \\ |1 - \theta\frac{L+\mu}{(\mu+1)(L+1)}| & \text{if (b)}, \\ \sqrt{\frac{(2-\theta)}{4\mu(L^2+1)}\frac{\left(\theta(L^2+1)-2\mu(\theta+L^2-1)\right)\left(\theta\left(1+2\mu+L^2\right)-2(\mu+1)\left(L^2+1\right)\right)}{2\mu(\theta+L^2-1)-(2-\theta)(1-L^2)}} & \text{otherwise,} \end{cases}$$

with

(a) $\mu\frac{-(2(\theta-1)\mu+\theta-2)+L^2(\theta-2(1+\mu))}{\sqrt{(2(\theta-1)\mu+\theta-2)^2+L^2(\theta-2(\mu+1))^2}} \leq \sqrt{L^2+1}$,

(b) $L < 1$, $\mu > \frac{L^2+1}{(L-1)^2}$, *and* $\theta \leq \frac{2(\mu+1)(L+1)(\mu+\mu L^2-L^2-2\mu L-1)}{2\mu^2-\mu+\mu L^3-L^3-3\mu L^2-L^2-2\mu^2 L-\mu L-L-1}$.

(In case (b), the former part of the condition ensures that there is no division by 0 in the latter part. We show this in [68, section SM4.2.1, case (b) part (ii)].)

Corollary 4.4. *Let $A \in \mathcal{M}_\mu$ and $B \in \mathcal{M} \cap \mathcal{L}_L$ with $\mu, L > 0$, and assume $\dim \mathcal{H} \geq 3$. The* tight *contraction factor of the DRS operator $I - J_B + J_A(2J_B - I)$ is*

$$\rho = \begin{cases} \frac{1+\sqrt{\frac{(1-2(\mu+1))^2L^2+1}{L^2+1}}}{2(1+\mu)} & \text{if } (\mu-1)(2\mu+1)^2L^2 \geq 2\mu^2 - 2\sqrt{2}\sqrt{\mu+1}\mu + \mu + 1 \text{ or } \mu \leq 1, \\ \frac{1+\mu L}{(1+\mu)(1+L)} & \text{if } L \leq \frac{2\mu^2(L-1)L^2+\mu(1-2L)-1}{(\mu+1)(L^2+L+1)} \text{ and } L < 1, \\ \sqrt{\frac{(2\mu L^2+L^2+1)(2\mu L^2-L^2-1)}{4\mu(L^2+1)(2\mu L^2+L^2-1)}} & \text{otherwise.} \end{cases}$$

Proof. Plug $\theta = 1$ into Theorem 4.3 and simplify. We omit the details. ∎

4.1. Proof outline. The discovery of these proofs relied heavily on the computer algebra system (CAS) *Mathematica*. When symbolically solving the primal problem, we conjectured that the worst-case operators would exist in $\mathbb{R}^2$. This is equivalent to conjecturing that the solution $G^\star \in \mathbb{R}^{3\times 3}$ has rank 2 or less, which is reasonable due to complementary slackness. We then formulated the problem of finding this 2-dimensional worst case as a nonconvex quadratic program, rather than an SDP, formulated the KKT system, and solved the stationary points using the CAS. When symbolically solving the dual problem, we conjectured that the optimal solution would correspond to $S^\star \in \mathbb{R}^{3\times 3}$ with rank 1 or 2, which is reasonable due to complementary slackness. We then chose ρ^2 and the other dual variables so that $S^\star$ would have rank 1 or 2. Finally, we minimized the contraction factor ρ^2 under those rank conditions to obtain the optimum. These two approaches gave us analytic expressions for optimal primal and dual SDP solutions. To verify the solutions, we formulated them into primal and dual feasible points and verified that their optimal values are equal for all parameter choices.

The written proofs of Theorems 4.1 and 4.3 are deferred to the supplementary material [68, sections SM3 and SM4]. The point we wish to make in this section is that the OSPEP is a powerful tool that enables us to prove incredibly complex results. The length and complexity of the proofs demonstrate this point.

The proofs provided on paper are complete and rigorous. However, we help the reader verify the calculations of [68, sections SM3 and SM4] with code that performs symbolic manipulations. If the reader is willing to trust the CAS's symbolic manipulations, the proofs will not be difficult to follow. We also verified the results through the following alternative approach: we finely discretized the parameter space and verified that the upper and lower bounds of [68, section SM3] are valid and that they match up to machine precision. The link to the code is provided in the conclusion.

4.2. Further remarks. The third contraction factor of Theorem 4.1, the factor $|1-\theta|$, matches the contraction factor of Theorem 5.6 of [25]. The contraction factor for the other four cases do not match. This implies that Theorem 5.6 of [25] is tight when $\theta \geq 2\frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}$, but it is not in the other cases.

The first (but neither the second nor third) contraction factor of Corollary 4.4 matches the contraction factor of Theorem 5.2 of [53] which assumes B is a skew symmetric L-Lipschitz linear operator, a stronger assumption than $B \in \mathcal{M} \cap \mathcal{L}$.

One can show that the contraction factors of Theorems 4.1 and 4.3 are symmetric in the assumptions. Specifically, if we swap the assumptions and instead assume [$B \in \mathcal{M}_\mu$ and $A \in \mathcal{C}_\beta$] and [$B \in \mathcal{M}_\mu$ and $A \in \mathcal{M} \cap \mathcal{L}_L$], the contraction factors of Theorems 4.1 and 4.3 remain valid and tight. The proof follows from using the "scaled relative graph" developed in the concurrent work by Ryu, Hannah, and Yin [67, Theorem 7].

The optimal α and θ minimizing the contraction factors of Theorems 4.1 and 4.3 can be computed with the algorithm presented in section 5. However, their analytical expressions seem to be quite complicated.

If we further assume A and B are subdifferential operators of closed convex proper functions, the contraction factors of Theorems 4.1 and 4.3 remain valid, but our proof no longer guarantees tightness; with the additional assumptions, it may be possible to obtain a smaller contraction factor. Such setups can be analyzed with the machinery and interpolation results of [74]. By numerically solving the SDP with the added subdifferential operator assumption, we find that Theorem 4.1 remains tight. For subdifferential operators of convex functions, Lipschitz continuity implies cocoercivity by the Baillon–Haddad theorem, so there is no reason to consider Theorem 4.3. Indeed, numerical solutions of the SDP indicate Theorem 4.3 is not tight in this setup.

TABLE 2
Prior results on contraction factors of Douglas–Rachford splitting.

Properties for A	Properties for B	Reference	Tight
∂f, f: str. cvx & smooth	∂g	[26, 27]	Y
∂f, f: str. cvx	∂g, g: smooth	[25]	N
str. mono. & cocoercive	-	[25]	Y
str. mono. & Lipschitz	-	[25]	Y
str. mono.	cocoercive	[25]	N
str. mono.	Lipschitz	[53]	N

Table 2 lists other commonly considered assumptions providing linear convergence of DRS and the corresponding prior work analyzing them. The results of Theorems 4.1 and 4.3 provide the tight contraction factors for the three cases for which there had not been tight results.

5. Automatic optimal parameter selection. When using FBS, DRS, or DYS, how should one choose the parameters $\alpha > 0$ and $\theta \in (0, 2)$? One option is to find a contraction factor and choose the α and θ that minimize it. However, this may be suboptimal if the contraction factor is not tight or if no known contraction factors fully utilize a given set of assumptions.

In this section, we use the OSPEP to automatically select the optimal algorithm parameters for FBS, DRS, and DYS. Write

$$\rho_\star^2(\alpha,\theta) = \left(\begin{array}{ll} \text{maximize} & \dfrac{\|T(z;A,B,C,\alpha,\theta) - T(z';A,B,C,\alpha,\theta)\|^2}{\|z-z'\|^2} \\ \text{subject to} & A \in \mathcal{Q}_1,\ B \in \mathcal{Q}_2,\ C \in \mathcal{Q}_3 \\ & z, z' \in \mathcal{H},\ z \neq z', \end{array} \right),$$

where z, z', A, B, and C are the optimization variables. This is the tight contraction factor of (3.1), and we make explicit its dependence on α and θ. Define

$$\rho_\star^2 = \inf_{\alpha>0,\, \theta\in(0,2)} \rho_\star^2(\alpha,\theta),$$

and write $\alpha_\star$ and $\theta_\star$ for the optimal parameters (if they exist) that attain the infimum.

Again, for simplicity of exposition, we limit the generality and consider the operator classes $\mathcal{Q}_1 = \mathcal{M}_\mu$, $\mathcal{Q}_2 = \mathcal{C}_\beta \cap \mathcal{L}_L$, and $\mathcal{Q}_3 = \mathcal{C}_{\beta_C}$, as in section 3.1.2. For $\beta \in (0,\infty)$ and $L \in (0,\infty)$, the intersection $\mathcal{C}_\beta \cap \mathcal{L}_L$ is nondegenerate. So strong duality holds by Theorem 3.3, and we use the dual OSPEP (3.7) to write

$$\rho_\star^2(\alpha,\theta) = \left(\begin{array}{ll} \text{minimize} & \rho^2 \\ \text{subject to} & \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C \geq 0 \\ & S(\rho^2, \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C, \theta, \alpha) \succeq 0 \end{array}\right),$$

where ρ^2, λ_μ^A, λ_β^B, λ_L^B, and λ_β^C are the optimization variables, and S is as in (3.8). Note that

$$\begin{aligned} &S(\rho^2, \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C, \theta, \alpha) \\ &= \begin{pmatrix} \rho^2 + \frac{\lambda_\beta^B \beta}{\alpha} + \lambda_L^B - 1 & \frac{\lambda_\mu^A}{2} & -\lambda_\beta^B(\frac{1}{2} + \frac{\beta}{\alpha}) - \lambda_L^B & 0 \\ \frac{\lambda_\mu^A}{2} & \lambda_\mu^A(1+\alpha\mu) & -\lambda_\mu^A & \frac{\lambda_\mu^A}{2} \\ -\lambda_\beta^B(\frac{1}{2} + \frac{\beta}{\alpha}) - \lambda_L^B & -\lambda_\mu^A & \lambda_\beta^B(\frac{\beta}{\alpha} - 1) + \lambda_L^B(1-\alpha^2 L^2) & -\frac{\lambda_\beta^C}{2} \\ 0 & \frac{\lambda_\mu^A}{2} & -\frac{\lambda_\beta^C}{2} & \frac{\lambda_\beta^C \beta_C}{\alpha} \end{pmatrix} \\ &- \begin{pmatrix} 1 \\ \theta \\ -\theta \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ \theta \\ -\theta \\ 0 \end{pmatrix}^T \end{aligned}$$

is the Schur complement of

$$\begin{aligned} &\tilde{S}(\rho^2, \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C, \theta, \alpha) \\ &= \begin{pmatrix} \rho^2 + \frac{\lambda_\beta^B \beta}{\alpha} + \lambda_L^B - 1 & \frac{\lambda_\mu^A}{2} & -\lambda_\beta^B(\frac{1}{2} + \frac{\beta}{\alpha}) - \lambda_L^B & 0 & 1 \\ \frac{\lambda_\mu^A}{2} & \lambda_\mu^A(1+\alpha\mu) & -\lambda_\mu^A & \frac{\lambda_\mu^A}{2} & \theta \\ -\lambda_\beta^B(\frac{1}{2} + \frac{\beta}{\alpha}) - \lambda_L^B & -\lambda_\mu^A & \lambda_\beta^B(\frac{\beta}{\alpha} - 1) + \lambda_L^B(1-\alpha^2 L^2) & -\frac{\lambda_\beta^C}{2} & -\theta \\ 0 & \frac{\lambda_\mu^A}{2} & -\frac{\lambda_\beta^C}{2} & \frac{\lambda_\beta^C \beta_C}{\alpha} & 0 \\ 1 & \theta & -\theta & 0 & 1 \end{pmatrix} \in \mathbb{R}^{5\times 5}. \end{aligned}$$

Therefore $S \succeq 0$ if and only if $\tilde{S} \succeq 0$. We use $\tilde{S}$, as it depends on θ linearly. Define $\rho_\star^2(\alpha) = \inf_{\theta\in(0,2)} \rho_\star^2(\alpha,\theta)$. We evaluate $\rho_\star^2(\alpha)$ by solving the SDP

$$\rho_\star^2(\alpha) = \left(\begin{array}{ll} \text{minimize} & \rho^2 \\ \text{subject to} & \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C \geq 0 \\ & \tilde{S}(\rho^2, \lambda_\mu^A, \lambda_\beta^B, \lambda_L^B, \lambda_\beta^C, \theta, \alpha) \succeq 0 \end{array}\right),$$

where ρ^2, λ_μ^A, λ_β^B, λ_L^B, λ_β^C, and θ are the optimization variables.

It remains to solve

$$\rho_\star^2 = \inf_{\alpha>0} \rho_\star^2(\alpha).$$

The function $\rho_\star^2(\alpha)$ is nonconvex in α, and it does not seem possible to compute $\rho_\star^2$ with a single SDP. However, $\rho^2(\alpha)$ seems to be continuous and unimodal for a wide range of operator classes and parameter choices. Continuity is not surprising. We do not know whether $\rho_\star^2(\alpha)$ is always unimodal; if it is, we do not know why.

To minimize the apparently continuous univariate unimodal function, we use the MATLAB derivative-free optimization (DFO) solver `fminunc`. We provide a routine that evaluates $\rho_\star^2(\alpha)$ by solving an SDP, and the DFO solver calls it to evaluate $\rho_\star^2(\alpha)$ at various values of α. Figure 1 shows an example of the function $\rho_\star^2(\alpha)$, and its minimizer was approximated with this approach. In Figure 2, we plot $\rho_\star^2(\alpha)$ under several assumptions. In all cases, $\rho_\star^2(\alpha)$ is continuous and unimodal.

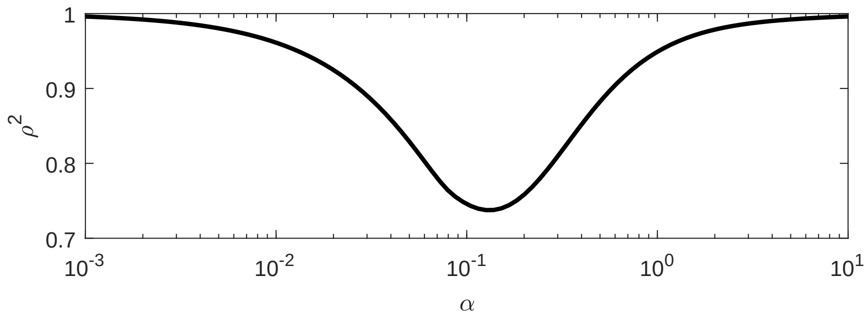
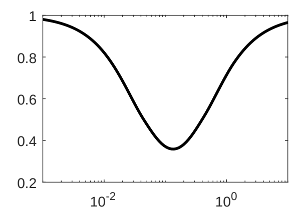
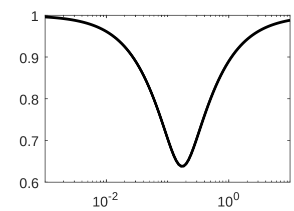


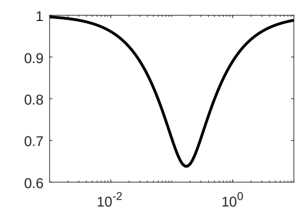
FIG. 1. *Plot of $\rho_\star^2(\alpha)$ under the assumptions $A \in \mathcal{M}_\mu$, $B \in \mathcal{C}_\beta \cap \mathcal{L}_L$, and $C \in \mathcal{C}_{\beta_C}$ with $\mu = 1$, $\beta = 0.01$, $L = 5$, and $\beta_C = 9$. The optimal parameters are $\alpha_\star \approx 0.131$ and $\theta_\star \approx 1.644$, and they produce the optimal contraction factor $\rho_\star^2 \approx 0.737$. We used the MATLAB solver* `fminunc` *for the minimization.*



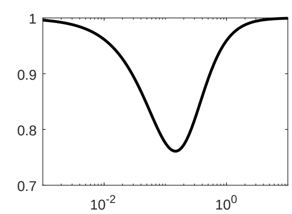
(a) $\mu_A = 1, \beta_A = 0.07,$ $\mu_B = 4, \beta_B = 0.02, \beta_C = 9,$ $\alpha_\star \approx 0.13, \theta_\star \approx 1.57, \rho_\star^2 \approx 0.36$



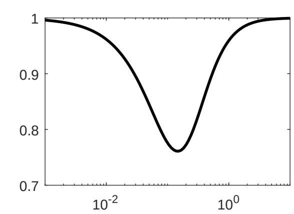
(b) $\mu_A = 1, \beta_B = 0.03, C = 0,$ $\alpha_\star \approx 0.17, \theta_\star \approx 1.65, \rho_\star^2 \approx 0.64$



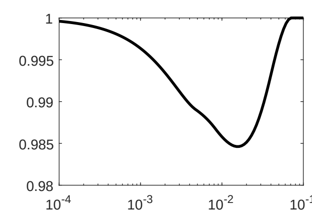
(c) $\beta_A = 0.03, \mu_B = 1, C = 0,$ $\alpha_\star \approx 0.17, \theta_\star \approx 1.65, \rho_\star^2 \approx 0.64$



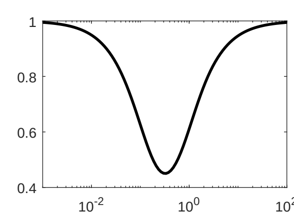
(d) $\mu_A = 1, L_B = 4, C = 0,$ $\alpha_\star \approx 0.15, \theta_\star \approx 1.59, \rho_\star^2 \approx 0.76$



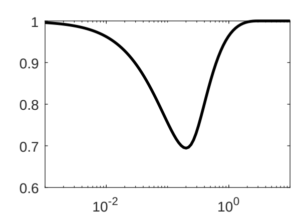
(e) $L_A = 4, \mu_B = 1, C = 0,$ $\alpha_\star \approx 0.15, \theta_\star \approx 1.59, \rho_\star^2 \approx 0.76$



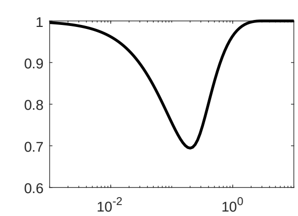
(f) $A = 0, \mu_B = 1, \beta_B = 0.1,$ $L_C = 8, \alpha_\star \approx 0.016, \theta_\star \approx 1,$ $\rho_\star^2 \approx 0.98$



(g) $\mu_A = 1, \beta_A = 0.07, L_A = 7,$ $\mu_B = 0.03, \beta_B = 0.02, L_A = 2,$ $\mu_C = 0.01, \beta_C = 9, L_C = 0.05,$ $\alpha_\star \approx 0.32, \theta_\star \approx 1.98, \rho_\star^2 \approx 0.45$



(h) $A = 0, \mu_B = 1, \beta_C = 0.1,$ $\alpha_\star \approx 0.2, \theta_\star \approx 1, \rho_\star^2 \approx 0.69$



(i) $\mu_A = 1, B = 0, \beta_C = 0.1,$ $\alpha_\star \approx 0.2, \theta_\star \approx 1, \rho_\star^2 \approx 0.69$

FIG. 2. *Plots of $\rho_\star^2(\alpha)$ under various assumptions. The plots are unimodal in all cases. All operator classes are subsets of $\mathcal{M}$, and only the parameters used in the intersection are specified. For example, subfigure* (e) *uses the classes $Q_1 = \mathcal{M} \cap \mathcal{L}_{L_A}$, $Q_2 = \mathcal{M}_{\mu_B}$, and $Q_3 = \{0\}$.*

6. Conclusion. In this work, we presented the OSPEP methodology, proved its tightness, and demonstrated its value by presenting two applications of it. The first application proved tight analytic contraction factors for DRS, and the second provided a method for automatic optimal parameter selection.

Proofs. The proofs discovered with the OSPEP metholodogy are available from the arXiv version of this article [68].

Code. With this paper, we release the following code: MATLAB script for implementing OSPEP for FBS, DRS, and DYS; MATLAB script for plotting Figures 1 and 2 of section 5; and *Mathematica* script for helping the reader verify the algebra of [68, section SM3]. The code uses YALMIP [48] and MOSEK [52] and is available from https://github.com/AdrienTaylor/OperatorSplittingPerformanceEstimation.

For splitting methods applied to convex functions, one can use the MATLAB toolbox PESTO [73], available from https://github.com/AdrienTaylor/Performance-Estimation-Toolbox.

Acknowledgments. Collaborations between the authors started during the LCCC Focus Period on Large-Scale and Distributed Optimization organized by the Automatic Control Department of Lund University. The authors thank the organizers and other participants. The authors thank Laurent Lessard (among others) for insightful discussions on the topics of DRS and computer-assisted proofs.

REFERENCES

[1] H. H. Bauschke, *Fenchel duality, Fitzpatrick functions and the extension of firmly nonexpansive mappings*, Proc. Amer. Math. Soc., 135 (2007), pp. 135–139.

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

[3] H. H. Bauschke and X. Wang, *Firmly Nonexpansive and Kirszbraun–Valentine extensions: A constructive approach via monotone operator theory*, in Nonlinear Analysis and Optimization I: Nonlinear Analysis, American Mathematics Society, 2010, pp. 55–64.

[4] H. H. Bauschke, X. Wang, and L. Yao, *General resolvents for monotone operators: Characterization and extension*, in Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning, and Inverse Problems, Medical Physics Publishing, 2010, pp. 57–74.

[5] L. M. Briceño-Arias, *Forward-Douglas–Rachford splitting and forward-partial inverse method for solving monotone inclusions*, Optimization, 64 (2015), pp. 1239–1261.

[6] L. M. Briceño-Arias and D. Davis, *Forward-backward-half forward algorithm for solving monotone inclusions*, SIAM J. Optim., 28 (2018), pp. 2839–2871, https://doi.org/10.1137/17M1120099.

[7] R. E. Bruck, *On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space*, J. Math. Anal. Appl., 61 (1977), pp. 159–164.

[8] G. H-G. Chen and R. T. Rockafellar, *Convergence rates in forward–backward splitting*, SIAM J. Optim., 7 (1997), pp. 421–444, https://doi.org/10.1137/S1052623495290179.

[9] L. Chen, X. Li, D. Sun, and K.-C. Toh, *On the equivalence of inexact proximal ALM and ADMM for a class of convex composite programming*, Math. Program., 2019; published online August 26, 2019.

[10] P. Chen, J. Huang, and X. Zhang, *A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions*, Fixed Point Theory Appl., 2016 (2016), 54.

[11] J.-P. Crouzeix and E. O. Anaya, *Maximality is nothing but continuity*, J. Convex Anal., 17 (2010), pp. 521–534.

[12] J.-P. Crouzeix and E. O. Anaya, *Monotone and maximal monotone affine subspaces*, Oper. Res. Lett., 38 (2010), pp. 139–142.

[13] J.-P. Crouzeix, E. O. Anaya, and W. Sosa, *A construction of a maximal monotone extension of a monotone map*, ESAIM Proc., 20 (2007), pp. 93–104.

[14] E. R. Csetnek, Y. Malitsky, and M. K. Tam, *Shadow Douglas–Rachford splitting for monotone inclusions*, Appl. Math. Optim., 80 (2019), pp. 665–678.

[15] D. Davis and W. Yin, *Faster convergence rates of relaxed Reaceman–Rachford and ADMM*

under regularity assumptions, Math. Oper. Res., 42 (2017), pp. 783–805.

[16] D. Davis and W. Yin, *A three-operator splitting scheme and its optimization applications*, Set-Valued Var. Anal., 25 (2017), pp. 829–858.

[17] W. Deng and W. Yin, *On the global and linear convergence of the generalized alternating direction method of multipliers*, J. Sci. Comput., 66 (2016), pp. 889–916.

[18] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[19] Y. Drori, *The exact information-based complexity of smooth convex minimization*, J. Complexity, 39 (2017), pp. 1–16.

[20] Y. Drori and M. Teboulle, *Performance of first-order methods for smooth convex minimization: A novel approach*, Math. Program., 145 (2014), pp. 451–482.

[21] G. França and J. Bento, *An explicit rate bound for over-relaxed ADMM*, in Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), IEEE, 2016, pp. 2104–2108.

[22] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, *On the optimal step-size selection for the alternating direction method of multipliers*, IFAC Proc. Vol., 45 (2012), pp. 139–144.

[23] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, *Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems*, IEEE Trans. Automat. Control, 60 (2015), pp. 644–658.

[24] P. Giselsson, *Tight linear convergence rate bounds for Douglas-Rachford splitting and ADMM*, in Proceedings of the 54th Conference on Decision and Control, Osaka, Japan, 2015, pp. 3305–3310.

[25] P. Giselsson, *Tight global linear convergence rate bounds for Douglas–Rachford splitting*, J. Fixed Point Theory Appl., 19 (2017), pp. 2241–2270.

[26] P. Giselsson and S. Boyd, *Diagonal scaling in Douglas-Rachford splitting and ADMM*, in Proceedings of the 53rd IEEE Conference on Decision and Control, Los Angeles, CA, 2014, pp. 5033–5039.

[27] P. Giselsson and S. Boyd, *Linear convergence and metric selection for Douglas-Rachford splitting and ADMM*, IEEE Trans. Automat. Control, 62 (2017), pp. 532–544.

[28] A. A. Goldstein, *Convex programming in Hilbert space*, Bull. Amer. Math. Soc., 70 (1964), pp. 709–710.

[29] G. Gu and J. Yang, *On the Optimal Ergodic Sublinear Convergence Rate of the Relaxed Proximal Point Algorithm for Variational Inequalities*, preprint, https://arxiv.org/abs/1905.06030, 2019.

[30] G. Gu and J. Yang, *On the Optimal Linear Convergence Factor of the Relaxed Proximal Point Algorithm for Monotone Inclusion Problems*, preprint, https://arxiv.org/abs/1905.04537, 2019.

[31] G. Gu and J. Yang, *Optimal Nonergodic Sublinear Convergence Rate of Proximal Point Algorithm for Maximal Monotone Inclusion Problems*, preprint, https://arxiv.org/abs/1904.05495, 2019.

[32] D. Han, D. Sun, and L. Zhang, *Linear rate convergence of the alternating direction method of multipliers for convex composite programming*, Math. Oper. Res., 43 (2018), pp. 622–637.

[33] M. Hong and Z.-Q. Luo, *On the linear convergence of the alternating direction method of multipliers*, Math. Program., 162 (2017), pp. 165–199.

[34] R. B. Kellogg, *A nonlinear alternating direction method*, Math. Comp., 23 (1969), pp. 23–27.

[35] D. Kim and J. A. Fessler, *Optimized first-order methods for smooth convex minimization*, Math. Program., 159 (2016), pp. 81–107.

[36] D. Kim and J. A. Fessler, *On the convergence analysis of the optimized gradient method*, J. Optim. Theory Appl., 172 (2017), pp. 187–205.

[37] D. Kim and J. A. Fessler, *Adaptive restart of the optimized gradient method for convex optimization*, J. Optim. Theory Appl., 178 (2018), pp. 240–263.

[38] D. Kim and J. A. Fessler, *Another look at the fast iterative shrinkage/thresholding algorithm (FISTA)*, SIAM J. Optim., 28 (2018), pp. 223–250, https://doi.org/10.1137/16M108940X.

[39] D. Kim and J. A. Fessler, *Generalizing the optimized gradient method for smooth convex minimization*, SIAM J. Optim., 28 (2018), pp. 1920–1950, https://doi.org/10.1137/17M112124X.

[40] D. Kim and J. A. Fessler, *Optimizing the Efficiency of First-Order Methods for Decreasing the Gradient of Smooth Convex Functions*, preprint, https://arxiv.org/abs/1803.06600, 2018.

[41] M. Kirszbraun, *Über die zusammenziehende und Lipschitzsche transformationen*, Fund. Math., 22 (1934), pp. 77–108.

[42] G. M. Korpelevič, *An extragradient method for finding saddle points and for other problems*,

Èkonom. i Mat. Metody, 12 (1976), pp. 747–756 (in Russian).

[43] L. Lessard, B. Recht, and A. Packard, *Analysis and design of optimization algorithms via integral quadratic constraints*, SIAM J. Optim., 26 (2016), pp. 57–95, https://doi.org/10.1137/15M1009597.

[44] E. S. Levitin and B. T. Polyak, *Constrained minimization methods*, Zh. Vychisl. Mat. Mat. Fiz., 6 (1966), pp. 787–823 (in Russian).

[45] F. Lieder, *On the Convergence Rate of the Halpern-Iteration*, preprint 2017-11-6336, Optimization Online, 2017.

[46] T. Lin, S. Ma, and S. Zhang, *An extragradient-based alternating direction method for convex minimization*, Found. Comput. Math., 17 (2017), pp. 35–59.

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

[48] J. Löfberg, *YALMIP: A toolbox for modeling and optimization in MATLAB*, in Proceedings of the CACSD Conference, Taipei, Taiwan, 2004.

[49] Y. Malitsky, *Golden ratio algorithms for variational inequalities*, Math. Program., 2019; published online July 31, 2019.

[50] Y. Malitsky and M. K. Tam, *A Forward-Backward Splitting Method for Monotone Inclusions without Cocoercivity*, preprint, https://arxiv.org/abs/1808.04162, 2018.

[51] B. Mercier, *Inéquations variationnelles de la mécanique*, Université de Paris-Sud, Département de mathématique, 1980.

[52] MOSEK ApS, *The MOSEK Optimization Toolbox for MATLAB Manual, Version* 8.1, 2017, http://docs.mosek.com/8.1/toolbox/index.html.

[53] W. M. Moursi and L. Vandenberghe, *Douglas–Rachford splitting for the sum of a Lipschitz continuous and a strongly monotone operator*, J. Optim. Theory Appl., 183 (2019), pp. 179–198.

[54] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. Jordan, *A general analysis of the convergence of ADMM*, in Proceedings of the 32nd International Conference on Machine Learning, Proc. Mach. Learn. Res. 37, PMLR, 2015, pp. 343–352.

[55] G. B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Anal. Appl., 72 (1979), pp. 383–390.

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

[57] F. Pedregosa, *On the Convergence Rate of the Three Operator Splitting Scheme*, preprint, https://arxiv.org/abs/1610.07830, 2016.

[58] F. Pedregosa, K. Fatras, and M. Casotto, *Proximal splitting meets variance reduction*, in Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, K. Chaudhuri and M. Sugiyama, eds., Proc. Mach. Learn. Res. 89, PMLR, 2019, pp. 1–10.

[59] F. Pedregosa and G. Gidel, *Adaptive three operator splitting*, in Proceedings of the 35th International Conference on Machine Learning, J. Dy and A. Krause, eds., Proc. Mach. Learn. Res. 80, PMLR, 2018, pp. 4085–4094.

[60] H. Raguet, *A note on the forward-Douglas–Rachford splitting for monotone inclusion and convex optimization*, Optim. Lett., 13 (2019), pp. 717–740.

[61] H. Raguet, J. Fadili, and G. Peyré, *A generalized forward-backward splitting*, SIAM J. Imaging Sci., 6 (2013), pp. 1199–1226, https://doi.org/10.1137/120872802.

[62] S. Reich, *Extension problems for accretive sets in Banach spaces*, J. Funct. Anal., 26 (1977), pp. 378–395.

[63] S. Reich and S. Simons, *Fenchel duality, Fitzpatrick functions and the Kirszbraun–Valentine extension theorem*, Proc. Amer. Math. Soc., 133 (2005), pp. 2657–2660.

[64] J. Rieger and M. K. Tam, *Backward-Forward-Reflected-Backward Splitting for Three Operator Monotone Inclusions*, preprint, https://arxiv.org/abs/2001.07327, 2020.

[65] R. T. Rockafellar, *Conjugate Duality and Optimization*, CBMS-NSF Reg. Conf. Ser. Appl. Math. 16, SIAM, 1974, https://doi.org/10.1137/1.9781611970524.

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

[67] E. K. Ryu, R. Hannah, and W. Yin, *Scaled Relative Graph: Nonexpansive Operators via* 2*D Euclidean Geometry*, preprint, https://arxiv.org/abs/1902.09788, 2019.

[68] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson, *Operator Splitting Performance Estimation: Tight Contraction Factors and Optimal Parameter Selection*, preprint, https://arxiv.org/abs/1812.00146, 2020.

[69] E. K. Ryu and B. C. Vũ, *Finding the forward-Douglas–Rachford-forward method*, J. Optim. Theory Appl., 184 (2020), pp. 858–876.

[70] J. H. Seidman, M. Fazlyab, V. M. Preciado, and G. J. Pappas, *A control-theoretic approach to analysis and parameter selection of Douglas–Rachford splitting*, IEEE Control Systems Lett., 4 (2020), pp. 199–204.

[71] A. Taylor, B. Van Scoy, and L. Lessard, *Lyapunov functions for first-order methods: Tight automated convergence guarantees*, in Proceedings of the 35th International Conference on Machine Learning, Proc. Mach. Learn. Res. 80, PMLR, 2018, pp. 4897–4906.

[72] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Exact worst-case performance of first-order methods for composite convex optimization*, SIAM J. Optim., 27 (2017), pp. 1283–1313, https://doi.org/10.1137/16M108104X.

[73] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Performance estimation toolbox (PESTO): Automated worst-case analysis of first-order optimization methods*, in Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC), IEEE, 2017, pp. 1278–1283.

[74] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Smooth strongly convex interpolation and exact worst-case performance of first-order methods*, Math. Program., 161 (2017), pp. 307–345.

[75] A. B. Taylor, J. M. Hendrickx, and F. Glineur, *Exact worst-case convergence rates of the proximal gradient method for composite convex minimization*, J. Optim. Theory Appl., 178 (2018), pp. 455–476.

[76] A. Teixeira, E. Ghadimi, I. Shames, H. Sandberg, and M. Johansson, *Optimal scaling of the ADMM algorithm for distributed quadratic programming*, in Proceedings of the 52nd IEEE Conference on Decision and Control, IEEE, 2013, pp. 6868–6873.

[77] A. Teixeira, E. Ghadimi, I. Shames, H. Sandberg, and M. Johansson, *The ADMM algorithm for distributed quadratic problems: Parameter selection and constraint preconditioning*, IEEE Trans. Signal Process., 64 (2016), pp. 290–305.

[78] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim., 29 (1991), pp. 119–138, https://doi.org/10.1137/0329006.

[79] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446, https://doi.org/10.1137/S0363012998338806.

[80] F. A. Valentine, *On the extension of a vector function so as to preserve a Lipschitz condition*, Bull. Amer. Math. Soc., 49 (1943), pp. 100–108.

[81] F. A. Valentine, *A Lipschitz condition preserving extension for a vector function*, Amer. J. Math., 67 (1945), pp. 83–93.

[82] B. Van Scoy, R. A. Freeman, and K. M. Lynch, *The fastest known globally convergent first-order method for minimizing strongly convex functions*, IEEE Control Systems Lett., 2 (2018), pp. 49–54.

[83] H. Wang, M. Fazlyab, S. Chen, and V. M. Preciado, *Robust Convergence Analysis of Three-Operator Splitting*, preprint, https://arxiv.org/abs/1910.04229, 2019.

[84] X. Wang and L. Yao, *Maximally monotone linear subspace extensions of monotone subspaces: Explicit constructions and characterizations*, Math. Program., 139 (2013), pp. 327–352.

[85] M. Yan, *A new primal-dual algorithm for minimizing the sum of three functions with a linear operator*, J. Sci. Comput., 76 (2018), pp. 1698–1717.