



**HAL**  
open science

# Operator Splitting Performance Estimation: Tight Contraction Factors and Optimal Parameter Selection

Ernest Ryu, Adrien Taylor, Carolina Bergeling, Pontus Giselsson

► **To cite this version:**

Ernest Ryu, Adrien Taylor, Carolina Bergeling, Pontus Giselsson. Operator Splitting Performance Estimation: Tight Contraction Factors and Optimal Parameter Selection. *SIAM Journal on Optimization*, 2020, 30 (3), pp.2251-2271. 10.1137/19M1304854 . hal-02956361

**HAL Id: hal-02956361**

**<https://inria.hal.science/hal-02956361>**

Submitted on 7 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 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 analytic 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.

**1. Introduction.** Consider the fixed-point iteration in a real Hilbert space  $\mathcal{H}$

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

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

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

for all  $x, y \in \mathcal{H}$ . We ask the 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 non-convex 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.

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 make OSPEP 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

\*Department of Mathematical Sciences, Seoul National University, Seoul, Korea [ernestyu@snu.ac.kr](mailto:ernestyu@snu.ac.kr)

†INRIA, Département d’informatique de l’ENS, École normale supérieure, CNRS, PSL Research University, Paris, France [adrien.taylor@inria.fr](mailto:adrien.taylor@inria.fr)

‡Department of Automatic Control, Lund University, Lund, Sweden [carolina.bergeling@control.lth.se](mailto:carolina.bergeling@control.lth.se), [pontusg@control.lth.se](mailto:pontusg@control.lth.se)

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 main contribution of this work is twofold. The first is analyzing 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 Sections 2, is a novel theoretical contribution. The second contribution is 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 second major contribution.

The major and minor contributions of this work are, to the best of our knowledge, 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]. Forward-Douglas–Rachford splitting of Raguet, Fadili, Peyré, and Brineño-Arias [61, 5, 60] 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 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, 83], and Crouzeix and Anaya [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 [77], 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 [74]. 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. 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. [75, 76] proved linear rates

of convergence and provided matching lower bounds for ADMM applied to quadratic problems. For DYS, Davis and Yin [16], Yan [84], 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 5 years. Past work analyzed convex optimization algorithms, and, to the best of our knowledge, analyzing 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 was made public on arXiv, several papers citing our work followed up on our results and used SDPs to analyze other monotone operator splitting methods [29, 30, 31, 69, 82, 68].) Drori and Teboulle [20] and Taylor, Hendrickx, and Glineur [71, 73] 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 [70]. Finally, Nishihara et al. [54] and França and Bento [21] used IQC to analyze ADMM.

Finally, both IQC and PEP approaches allowed designing new methods for particular problem settings. For example, the optimized gradient method by Kim and Fessler [35, 36, 37, 38, 39, 40] (first numerical version 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 iteration, as showed by Drori [19]. On the other hand, the IQC framework was used by Van Scoy et al. [81] for developing the triple momentum method, 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  $A: \mathcal{H} \rightrightarrows \mathcal{H}$  if  $A$  maps a point in  $\mathcal{H}$  to a subset of  $\mathcal{H}$ . So  $A(x) \subset \mathcal{H}$  for all  $x \in \mathcal{H}$ . For simplicity, we also write  $Ax = A(x)$ . Write  $I: \mathcal{H} \rightarrow \mathcal{H}$  for the identity operator. We say  $A: \mathcal{H} \rightrightarrows \mathcal{H}$  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  $A: \mathcal{H} \rightrightarrows \mathcal{H}$  is  $\mu$ -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  $A: \mathcal{H} \rightarrow \mathcal{H}$  is  $\beta$ -cocoercive if

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

where  $\beta \in (0, \infty)$ . We say a single-valued operator  $A: \mathcal{H} \rightarrow \mathcal{H}$  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  $T: \mathcal{H} \rightarrow \mathcal{H}$  is contractive if it is  $\rho$ -Lipschitz with  $\rho < 1$ . We say  $x^*$  is a fixed point of  $T$  if  $x^* = Tx^*$ .

Davis–Yin splitting (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

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

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.

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

Table 1: Operator classes for which we analyze interpolation. The parameters  $\mu$ ,  $L$ , and  $\beta$  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  $L > 0$ .

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

FACT 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

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

PROPOSITION 1. Let  $\mu \in (0, \infty)$ . Then  $\{(x_i, q_i)\}_{i \in \mathcal{I}}$  is  $\mathcal{M}_\mu$ -interpolable if and only if

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

*Proof.* With Fact 1, the proof follows from a sequence of equivalences:

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

PROPOSITION 2. *Let  $\beta \in (0, \infty)$ . Then  $\{(x_i, q_i)\}_{i \in \mathcal{I}}$  is  $\mathcal{C}_\beta$ -interpolable if and only if*

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

*Proof.* With Proposition 1, the proof follows from a sequence of equivalences:

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

FACT 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*

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

Fact 2 is a special case of the Kirszbraun–Valentine theorem [41, 79, 80]. 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 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*

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

*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  $\{x \mid Qx = 0\}$  is convex [2, Proposition 23.39]. This implies  $Q(1/2, 0) = (0, 0)$ , which is a contradiction.  $\square$

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 4. *Assume  $0 < \mu$ ,  $\mu \leq L < \infty$ , and  $\mu \leq 1/\beta < \infty$ . 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*

$$(2.1) \quad \begin{aligned} \langle q_1 - q_2, x_1 - x_2 \rangle &\geq \mu \|x_1 - x_2\|^2 \\ \langle q_1 - q_2, x_1 - x_2 \rangle &\geq \beta \|q_1 - q_2\|^2 \\ \|q_1 - q_2\|^2 &\leq L^2 \|x_1 - x_2\|^2. \end{aligned}$$

*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,  $\dim \mathcal{H} \geq 2$ .

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  $Q : \mathcal{H} \rightarrow \mathcal{H}$  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  $x \neq 0$ . If  $q = \gamma x$  for some  $\gamma \in \mathbb{R}$ , then the operator  $Q : \mathcal{H} \rightarrow \mathcal{H}$  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,

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

along with an associated bounded linear operator  $A : \mathcal{H} \rightarrow \mathcal{H}$  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  $\text{span}\{e_1, e_2\}$ , define

$$\begin{aligned} Ae_1 &= \frac{\langle q, e_1 \rangle}{\|x\|} e_1 + \frac{\sqrt{\|q\|^2 - (\langle e_1, q \rangle)^2}}{\|x\|} e_2, \\ Ae_2 &= -\frac{\sqrt{\|q\|^2 - (\langle e_1, q \rangle)^2}}{\|x\|} e_1 + \frac{\langle q, e_1 \rangle}{\|x\|} e_2. \end{aligned}$$

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

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

With direct computations, we can verify that  $M$  satisfies

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

This implies  $A : \mathcal{H} \rightarrow \mathcal{H}$  is  $L$ -Lipschitz,  $\mu$ -strongly monotone, and  $\beta$ -cocoercive. Finally, the affine operator  $Q : \mathcal{H} \rightarrow \mathcal{H}$  defined as

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

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

Proposition 4 presents conditions for interpolation with 3 classes. Interpolation conditions with 2 of these classes, such as  $(\mathcal{C}_\beta \cap \mathcal{L}_L)$ ,  $(\mathcal{M}_\mu \cap \mathcal{C}_\beta)$ ,  $(\mathcal{M}_\mu \cap \mathcal{L}_L)$ ,  $(\mathcal{M} \cap \mathcal{L}_L)$ , and are of the same form and follow from the a very similar (identical) proof.

**3. Operator splitting performance estimation problems.** Consider the *operator splitting performance estimation problem* (OSPEP)

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

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 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. (So 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,  $\rho$  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 non-convex 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

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

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

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

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

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

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, in Section SM1.

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

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

For operator  $B$ , we have

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

For operator  $C$ , we have

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

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

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

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 non-primed 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

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

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:

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

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

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

*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

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

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

Write

$$\begin{aligned} M_I &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & M_O &= \begin{pmatrix} 1 & \theta & -\theta & 0 \\ \theta & \theta^2 & -\theta^2 & 0 \\ -\theta & -\theta^2 & \theta^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\ M_\mu^A &= \begin{pmatrix} 0 & -1/2 & 0 & 0 \\ -1/2 & -1 - \alpha\mu & 1 & -1/2 \\ 0 & 1 & 0 & 0 \\ 0 & -1/2 & 0 & 0 \end{pmatrix}, & M_\beta^C &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & -\beta_C/\alpha \end{pmatrix}, \\ M_\beta^B &= \begin{pmatrix} -\beta/\alpha & 0 & 1/2 + \beta/\alpha & 0 \\ 0 & 0 & 0 & 0 \\ 1/2 + \beta/\alpha & 0 & -1 - \beta/\alpha & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & M_L^B &= \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & -1 + \alpha^2 L^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

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

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

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, in Section SM1.

**THEOREM 3.2.** *The OSPEP (3.1) and the SDP of Section SM1 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 another. Given an optimal  $G^*$  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

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

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

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

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)

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

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, in 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 SM1 and SM2. Furthermore, strong duality holds if each class  $\mathcal{Q}_1$ ,  $\mathcal{Q}_2$ , and  $\mathcal{Q}_3$  is a non-degenerate intersection of classes of Table 1.*

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

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

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

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

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

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

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

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

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 non-degenerate 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 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

**3.4. Further remarks.** With analogous steps, the OSPEPs for FBS and DRS can be written as smaller  $3 \times 3$  SDPs. Using the smaller SDP is preferred, as formulating these cases into larger  $4 \times 4$  SDPs, as a special case of the  $4 \times 4$  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 twice or more per iteration, the OSPEP loses the tightness guarantee. Such methods include the extragradient method [42], FBF [78], PDFP [10], Extragradient-Based Alternating Direction Method for Convex Minimization [46], FBHF [6], FRB [50], 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, in 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*

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

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 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*

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

*Proof.* Plug  $\theta = 1$  into Theorem [4.1](#) and simplify. We omit the details.  $\square$

**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*

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

with

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

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

(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 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*

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

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

**4.1. Proof outline.** The discovery of these proofs relied heavily on a 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  $G^* \in \mathbb{R}^{3 \times 3}$  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 non-convex 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  $S^* \in \mathbb{R}^{3 \times 3}$  with rank 1 or 2, which is reasonable due to complementary slackness. We then chose  $\rho^2$  and the other dual variables so that  $S^*$  would have rank 1 or 2. Finally, we minimized the contraction factor  $\rho^2$  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 proof of Theorems 4.1 and 4.3, are deferred to supplementary materials, to 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 readers verify the calculations of Sections SM3 and SM4 with code that performs symbolic manipulations. If a reader is willing to trust the CAS's symbolic manipulations, the proofs are not 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 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 4 cases do not match. This implies, Theorem 5.6 of [25] is tight when  $\theta \geq 2 \frac{\mu\beta + \mu + \beta}{2\mu\beta + \mu + \beta}$  but not in the other cases.

The first contraction factor of Corollary 4.4, but not the second and third, matches the contraction factor of Theorem 5.2 of [53] which instead 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  $[B \in \mathcal{M}_\mu \text{ and } A \in \mathcal{C}_\beta]$  and  $[B \in \mathcal{M}_\mu \text{ and } A \in \mathcal{M} \cap \mathcal{L}_L]$ , 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  $\alpha$  and  $\theta$  minimizing the contraction factor 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 furthermore 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 [73]. 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.

Properties for $A$	Properties for $B$	Reference	Tight
$\partial f, f$ : str. cvx & smooth	$\partial g$	[26, 27]	Y
$\partial f, f$ : str. cvx	$\partial g, g$ : 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: Prior results on contraction factors of Douglas–Rachford splitting.

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  $\alpha$  and  $\theta$  that minimizes 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

$$\rho_{\star}^2(\alpha, \theta) = \left( \begin{array}{l} \text{maximize} \\ \text{subject to} \end{array} \frac{\|T(z; A, B, C, \alpha, \theta) - T(z'; A, B, C, \alpha, \theta)\|^2}{\|z - z'\|^2} \right)$$

$A \in \mathcal{Q}_1, B \in \mathcal{Q}_2, C \in \mathcal{Q}_3$   
 $z, z' \in \mathcal{H}, z \neq z'$

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  $\alpha$  and  $\theta$ . 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 that attain the infimum, if they exist.



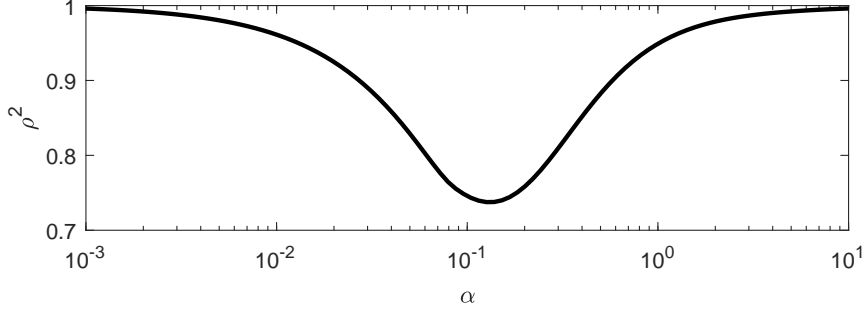


Fig. 1: Plot of  $\rho_*^2(\alpha)$  under the assumptions  $A \in \mathcal{M}_\mu$ ,  $B \in \mathcal{C}_\beta \cap \mathcal{L}_L$ , and  $C \in \mathcal{C}_{\beta_C}$  with  $\mu = 1$ ,  $\beta = 0.01$ ,  $L = 5$ , and  $\beta_C = 9$ . The optimal parameters are  $\alpha_* \approx 0.131$  and  $\theta_*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

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 non-degenerate. So strong duality holds by Theorem 3.3, and we use the dual OSPEP (3.7) to write

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

where  $\rho^2$ ,  $\lambda_\mu^A$ ,  $\lambda_\beta^B$ ,  $\lambda_L^B$ , and  $\lambda_\beta^C$  are the optimization variables and  $S$  is as in (3.8). Note that

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

is the Schur complement of

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

Therefore  $S \succeq 0$  if and only if  $\tilde{S} \succeq 0$ . We use  $\tilde{S}$  as it depends on  $\theta$  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

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

where  $\rho^2$ ,  $\lambda_\mu^A$ ,  $\lambda_\beta^B$ ,  $\lambda_L^B$ ,  $\lambda_\beta^C$ , and  $\theta$  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 non-convex in  $\alpha$ , 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 or why  $\rho_\star^2(\alpha)$  is always unimodal.

To minimize the apparently continuous univariate unimodal function, we use Matlab’s 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  $\alpha$ . 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.

**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 was to prove tight analytic contraction factors for DRS and the second was to provide a method for automatic optimal parameter selection.

*Code.* With this paper, we release the following code: Matlab script implementing OSPEP for FBS, DRS, and DYS; Matlab script used to plot the figures of Section 5; and Mathematica script to help readers verify the algebra of Section SM3. The code uses YALMIP [48] and Mosek [52] and is available at <https://github.com/AdrienTaylor/OperatorSplittingPerformanceEstimation>.

For splitting methods applied to convex functions, one can use the Matlab toolbox PESTO [72], available at <https://github.com/AdrienTaylor/Performance-Estimation-Toolbox>.

**Acknowledgements.** 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 the other participants. Among others, we thank Laurent Lessard for insightful discussions on the topics of DRS and computer-assisted proofs. Ernest Ryu was supported in part by NSF grant DMS-1720237 and ONR grant N000141712162. Adrien Taylor was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement 724063). Pontus Giselsson was supported by the Swedish Foundation for Strategic Research and the Swedish Research Council.

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*, Springer New York, 2nd ed., 2017.

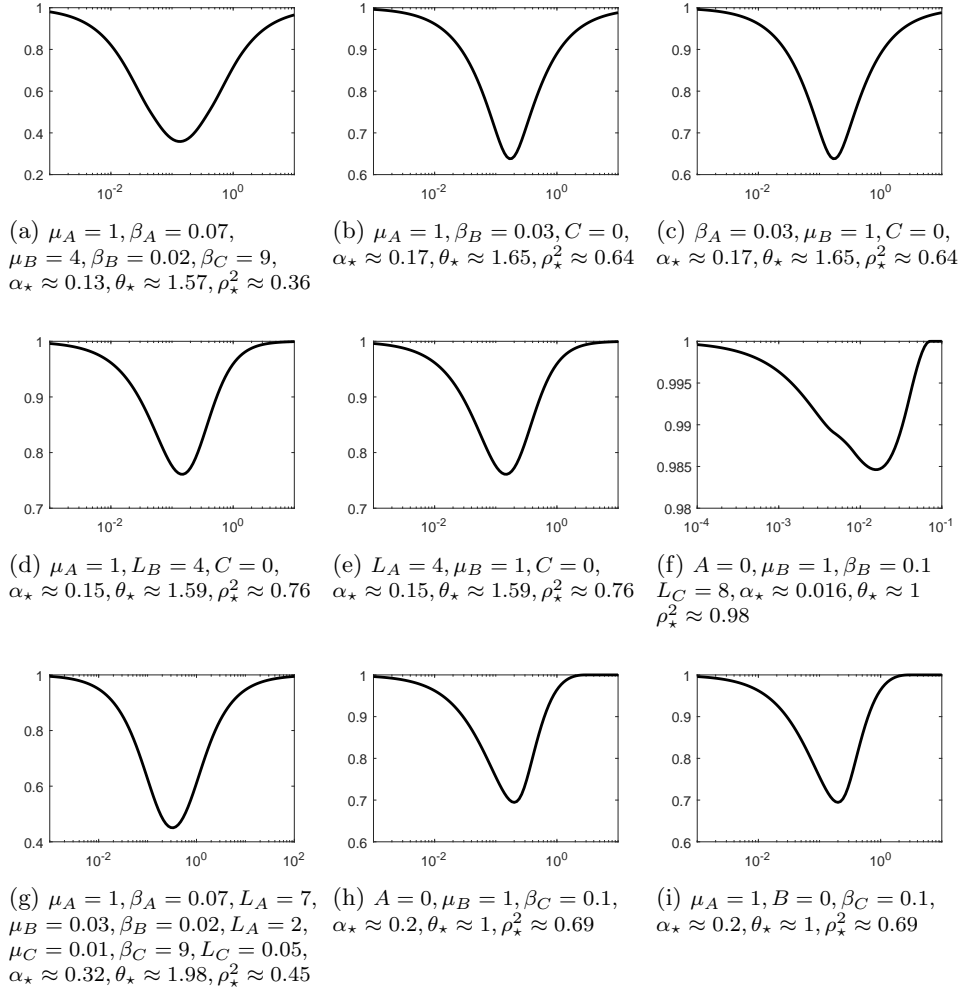


Fig. 2: Plots of  $\rho_*^2(\alpha)$  under various assumptions. The plots are unimodal in all cases. All operator classes are subsets of  $\mathcal{M}$ , and only the parameters used in the intersection are specified. For example, subfigure (e) uses the classes  $Q_1 = \mathcal{M} \cap \mathcal{L}_{L_A}$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

- [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 Journal on Optimization*, 28 (2018), pp. 2839–2871.
- [7] 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 (1977), pp. 159–164.

- [8] G. H.-G. CHEN AND R. T. ROCKAFELLAR, *Convergence rates in forward-backward splitting*, SIAM Journal on Optimization, 7 (1997), pp. 421–444.
- [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*, Mathematical Programming, (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 and Applications, 2016 (2016), p. 54.
- [11] J.-P. CROUZEIX AND E. O. ANAYA, *Maximality is nothing but continuity*, Journal of Convex Analysis, 17 (2010), pp. 521–534.
- [12] J.-P. CROUZEIX AND E. O. ANAYA, *Monotone and maximal monotone affine subspaces*, Operations Research Letters, 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*, Applied Mathematics & Optimization, (2019).
- [15] D. DAVIS AND W. YIN, *Faster convergence rates of relaxed Reaceman–Rachford and ADMM under regularity assumptions*, Mathematics of Operations Research, 42 (2017), pp. 783–805.
- [16] D. DAVIS AND W. YIN, *A three-operator splitting scheme and its optimization applications*, Set-Valued and Variational Analysis, 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*, Journal of Scientific Computing, 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*, Transactions of the American Mathematical Society, 82 (1956), pp. 421–439.
- [19] Y. DRORI, *The exact information-based complexity of smooth convex minimization*, Journal of Complexity, 39 (2017), pp. 1–16.
- [20] Y. DRORI AND M. TEBoulLE, *Performance of first-order methods for smooth convex minimization: a novel approach*, Mathematical Programming, 145 (2014), pp. 451–482.
- [21] G. FRANÇA AND J. BENTO, *An explicit rate bound for over-relaxed ADMM*, in Information Theory (ISIT), 2016 IEEE International Symposium on, 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 Proceedings Volumes, 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 Transactions on Automatic Control, 60 (2015), pp. 644–658.
- [24] P. GISELSSON, *Tight linear convergence rate bounds for Douglas-Rachford splitting and ADMM*, in Proceedings of 54th Conference on Decision and Control, Osaka, Japan, Dec 2015.
- [25] P. GISELSSON, *Tight global linear convergence rate bounds for DouglasRachford splitting*, Journal of Fixed Point Theory and Applications, 19 (2017), pp. 2241–2270.
- [26] P. GISELSSON AND S. BOYD, *Diagonal scaling in Douglas-Rachford splitting and ADMM*, in 53rd IEEE Conference on Decision and Control, Los Angeles, CA, Dec. 2014, pp. 5033–5039.
- [27] P. GISELSSON AND S. BOYD, *Linear convergence and metric selection for Douglas-Rachford splitting and ADMM*, IEEE Transactions on Automatic Control, 62 (2017), pp. 532–544.
- [28] A. A. GOLDSTEIN, *Convex programming in Hilbert space*, Bulletin of the American Mathematical Society, 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*, arXiv preprint arXiv: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*, arXiv preprint arXiv:1905.04537, (2019).
- [31] G. GU AND J. YANG, *Optimal nonergodic sublinear convergence rate of proximal point algorithm for maximal monotone inclusion problems*, arXiv preprint arXiv: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*, Mathematics of Operations Research, 43 (2018), pp. 622–637.
- [33] M. HONG AND Z.-Q. LUO, *On the linear convergence of the alternating direction method of multipliers*, Mathematical Programming, 162 (2017), pp. 165–199.
- [34] R. B. KELLOGG, *A nonlinear alternating direction method*, Mathematics of Computation, 23 (1969), pp. 23–27.
- [35] D. KIM AND J. A. FESSLER, *Optimized first-order methods for smooth convex minimization*,

- Mathematical programming, 159 (2016), pp. 81–107.
- [36] D. KIM AND J. A. FESSLER, *On the convergence analysis of the optimized gradient method*, Journal of Optimization Theory and Applications, 172 (2017), pp. 187–205.
  - [37] D. KIM AND J. A. FESSLER, *Adaptive restart of the optimized gradient method for convex optimization*, Journal of Optimization Theory and Applications, 178 (2018), pp. 240–263.
  - [38] D. KIM AND J. A. FESSLER, *Another look at the fast iterative shrinkage/thresholding algorithm (FISTA)*, SIAM Journal on Optimization, 28 (2018), pp. 223–250.
  - [39] D. KIM AND J. A. FESSLER, *Generalizing the optimized gradient method for smooth convex minimization*, SIAM Journal on Optimization, 28 (2018), pp. 1920–1950.
  - [40] D. KIM AND J. A. FESSLER, *Optimizing the efficiency of first-order methods for decreasing the gradient of smooth convex functions*, arXiv preprint arXiv:1803.06600, (2018).
  - [41] M. KIRSZBRAUN, *Über die zusammenziehende und Lipschitzsche transformationen*, Fundamenta Mathematicae, 22 (1934), pp. 77–108.
  - [42] G. M. KORPELEVICH, *The extragradient method for finding saddle points and other problems*, Ekonomika i Matematicheskie Metody, 12 (1976), pp. 747–756.
  - [43] L. LESSARD, B. RECHT, AND A. PACKARD, *Analysis and design of optimization algorithms via integral quadratic constraints*, SIAM Journal on Optimization, 26 (2016), pp. 57–95.
  - [44] E. S. LEVITIN AND B. T. POLYAK, *Constrained minimization methods*, Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, 6 (1966), pp. 787–823.
  - [45] F. LIEDER, *On the convergence rate of the Halpern-iteration*, Optimization Online preprint:2017-11-6336, (2017).
  - [46] T. LIN, S. MA, AND S. ZHANG, *An extragradient-based alternating direction method for convex minimization*, Foundations of Computational Mathematics, 17 (2017), pp. 35–59.
  - [47] P. L. LIONS AND B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, 16 (1979), pp. 964–979.
  - [48] J. LÖFBERG, *Yalmip : A toolbox for modeling and optimization in matlab*, in In Proceedings of the CACSD Conference, Taipei, Taiwan, 2004.
  - [49] Y. MALITSKY, *Golden ratio algorithms for variational inequalities*, Mathematical Programming, (2019).
  - [50] Y. MALITSKY AND M. K. TAM, *A forward-backward splitting method for monotone inclusions without cocoercivity*, arXiv preprint arXiv: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*, Journal of Optimization Theory and Applications, 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, vol. 37 of Proceedings of Machine Learning Research, 2015, pp. 343–352.
  - [55] G. B. PASSTY, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, Journal of Mathematical Analysis and Applications, 72 (1979), pp. 383–390.
  - [56] D. W. PEACEMAN AND H. H. RACHFORD, *The numerical solution of parabolic and elliptic differential equations*, Journal of the Society for Industrial and Applied Mathematics, 3 (1955), pp. 28–41.
  - [57] F. PEDREGOSA, *On the convergence rate of the three operator splitting scheme*, arXiv preprint arXiv:1610.07830, (2016).
  - [58] F. PEDREGOSA, K. FATRAS, AND M. CASOTTO, *Proximal splitting meets variance reduction*, 2019.
  - [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., vol. 80 of Proceedings of Machine Learning Research, PMLR, 10–15 Jul 2018, pp. 4085–4094.
  - [60] H. RAGUET, *A note on the forward–Douglas–Rachford splitting for monotone inclusion and convex optimization*, Optimization Letters, (2018).
  - [61] H. RAGUET, J. FADILI, AND G. PEYRÉ, *A generalized forward-backward splitting*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 1199–1226.
  - [62] S. REICH, *Extension problems for accretive sets in Banach spaces*, Journal of Functional Analysis, 26 (1977), pp. 378–395.
  - [63] S. REICH AND S. SIMONS, *Fenchel duality, Fitzpatrick functions and the Kirszbraun–Valentine extension theorem*, Proceedings of the American Mathematical Society, 133 (2005), pp. 2657–2660.

- [64] J. RIEGER AND M. K. TAM, *Backward-forward-reflected-backward splitting for three operator monotone inclusions*, arXiv:2001.07327, (2020).
- [65] R. ROCKAFELLAR, *Conjugate Duality and Optimization*, Society for Industrial and Applied Mathematics, 1974.
- [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 2D Euclidean geometry*, arXiv preprint arXiv:1902.09788, (2019).
- [68] E. K. RYU AND B. C. VÛ, *Finding the forward-Douglas–Rachford-forward method*, Journal of Optimization Theory and Applications, (2019).
- [69] 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 Letters, 4 (2020), pp. 199–204.
- [70] 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, vol. 80, PMLR, 2018, pp. 4897–4906.
- [71] A. B. TAYLOR, J. M. HENDRICKX, AND F. GLINEUR, *Exact worst-case performance of first-order methods for composite convex optimization*, SIAM Journal on Optimization, 27 (2017), pp. 1283–1313.
- [72] A. B. TAYLOR, J. M. HENDRICKX, AND F. GLINEUR, *Performance estimation toolbox (PESTO): automated worst-case analysis of first-order optimization methods*, in 2017 IEEE 56th Annual Conference on Decision and Control (CDC), IEEE, 2017, pp. 1278–1283.
- [73] A. B. TAYLOR, J. M. HENDRICKX, AND F. GLINEUR, *Smooth strongly convex interpolation and exact worst-case performance of first-order methods*, Mathematical Programming, 161 (2017), pp. 307–345.
- [74] A. B. TAYLOR, J. M. HENDRICKX, AND F. GLINEUR, *Exact worst-case convergence rates of the proximal gradient method for composite convex minimization*, Journal of Optimization Theory and Applications, 178 (2018), pp. 455–476.
- [75] A. TEIXEIRA, E. GHADIMI, I. SHAMES, H. SANDBERG, AND M. JOHANSSON, *Optimal scaling of the ADMM algorithm for distributed quadratic programming*, in 52nd IEEE Conference on Decision and Control, 2013, pp. 6868–6873.
- [76] A. TEIXEIRA, E. GHADIMI, I. SHAMES, H. SANDBERG, AND M. JOHANSSON, *The ADMM algorithm for distributed quadratic problems: Parameter selection and constraint preconditioning*, IEEE Transactions on Signal Processing, 64 (2016), pp. 290–305.
- [77] P. TSENG, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM Journal on Control and Optimization, 29 (1991), pp. 119–138.
- [78] P. TSENG, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM Journal on Control and Optimization, 38 (2000), pp. 431–446.
- [79] 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.
- [80] F. A. VALENTINE, *A Lipschitz condition preserving extension for a vector function*, American Journal of Mathematics, 67 (1945), pp. 83–93.
- [81] 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 Letters, 2 (2018), pp. 49–54.
- [82] H. WANG, M. FAZLYAB, S. CHEN, AND V. M. PRECIADO, *Robust convergence analysis of three-operator splitting*, arXiv preprint arXiv:1910.04229, (2019).
- [83] X. WANG AND L. YAO, *Maximally monotone linear subspace extensions of monotone subspaces: explicit constructions and characterizations*, Mathematical Programming, 139 (2013), pp. 327–352.
- [84] M. YAN, *A new primal-dual algorithm for minimizing the sum of three functions with a linear operator*, Journal of Scientific Computing, 76 (2018), pp. 1698–1717.

## Appendix

**SM1. Full primal OSPEP.** We state the full primal OSPEP with the operator classes  $\mathcal{Q}_1 = \mathcal{M}_{\mu_A} \cap \mathcal{C}_{\beta_A} \cap \mathcal{L}_{L_A}$ ,  $\mathcal{Q}_2 = \mathcal{M}_{\mu_B} \cap \mathcal{C}_{\beta_B} \cap \mathcal{L}_{L_B}$ , and  $\mathcal{Q}_3 = \mathcal{M}_{\mu_C} \cap \mathcal{C}_{\beta_C} \cap \mathcal{L}_{L_C}$ . The primal OSPEP with fewer assumptions will be of an analogous form with fewer

constraints.

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

where  $G \in \mathbb{S}_+^4$  is the optimization variable and

$$\begin{aligned}
M_I &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & 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 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & -\alpha\mu_A - 1 & 1 & -\frac{1}{2} \\ 0 & 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \end{pmatrix}, & M_\beta^A &= \begin{pmatrix} -\frac{\beta_A}{\alpha} & -\frac{\beta_A}{\alpha} - \frac{1}{2} & \frac{2\beta_A}{\alpha} & -\frac{\beta_A}{\alpha} \\ -\frac{\beta_A}{\alpha} - \frac{1}{2} & -\frac{\beta_A}{\alpha} - 1 & \frac{2\beta_A}{\alpha} + 1 & -\frac{\beta_A}{\alpha} - \frac{1}{2} \\ \frac{2\beta_A}{\alpha} & \frac{2\beta_A}{\alpha} + 1 & -\frac{4\beta_A}{\alpha} & \frac{2\beta_A}{\alpha} \\ -\frac{\beta_A}{\alpha} & -\frac{\beta_A}{\alpha} - \frac{1}{2} & \frac{2\beta_A}{\alpha} & -\frac{\beta_A}{\alpha} \end{pmatrix}, \\
M_L^A &= \begin{pmatrix} -1 & -1 & 2 & -1 \\ -1 & \alpha^2 L_A^2 - 1 & 2 & -1 \\ 2 & 2 & -4 & 2 \\ -1 & -1 & 2 & -1 \end{pmatrix}, & M_\mu^B &= \begin{pmatrix} 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & -\alpha\mu_B - 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
M_\beta^B &= \begin{pmatrix} -\frac{\beta_B}{\alpha} & 0 & \frac{\beta_B}{\alpha} + \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ \frac{\beta_B}{\alpha} + \frac{1}{2} & 0 & -\frac{\beta_B}{\alpha} - 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, & M_L^B &= \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & \alpha^2 L_B^2 - 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \\
M_\mu^C &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\alpha\mu_C & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}, & M_\beta^C &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & -\frac{\beta_C}{\alpha} \end{pmatrix}, & M_L^C &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha^2 L_C^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.
\end{aligned}$$

The objective  $\text{Tr}(M_O G)$  corresponds to  $\|z - \theta(z_B - z_A) - z' + \theta(z'_B - z'_A)\|^2$ . The equality constraint  $\text{Tr}(M_I G) = 1$  corresponds to  $\|z - z'\|^2 = 1$ . The other 9 inequality constraints correspond to the three assumptions on the three operators. In particular,  $\text{Tr}(M_\mu^A G) \geq 0$ ,  $\text{Tr}(M_\beta^A G) \geq 0$ , and  $\text{Tr}(M_L^A G) \geq 0$  respectively correspond to the  $\mu$ -strong monotonicity,  $\beta$ -cocoercivity, and  $L$ -Lipschitz continuity assumptions on  $A$  respectively. The assumptions on  $B$  and  $C$  have analogous correspondences.

**SM2. Full dual OSPEP.** We state the full dual OSPEP with the same operator classes as in Section SM1. The dual OSPEP with fewer assumptions will be of an analogous form with fewer  $\lambda$ -variables.

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

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

$$\begin{aligned} S(\rho^2, \lambda_\mu^A, \lambda_\beta^A, \lambda_L^A, \lambda_\mu^B, \lambda_\beta^B, \lambda_L^B, \lambda_\mu^C, \lambda_\beta^C, \lambda_L^C, \theta, \alpha) = & \rho^2 M_I - M_O - \lambda_\mu^A M_\mu^A - \lambda_\beta^A M_\beta^A - \lambda_L^A M_L^A \\ & - \lambda_\mu^B M_\mu^B - \lambda_\beta^B M_\beta^B - \lambda_L^B M_L^B \\ & - \lambda_\mu^C M_\mu^C - \lambda_\beta^C M_\beta^C - \lambda_L^C M_L^C \end{aligned}$$

is symmetric. The matrix can also explicitly be written as

$$S(\rho^2, \lambda_\mu^A, \lambda_\beta^A, \lambda_L^A, \lambda_\mu^B, \lambda_\beta^B, \lambda_L^B, \lambda_\mu^C, \lambda_\beta^C, \lambda_L^C, \theta, \alpha) = \begin{pmatrix} S_{1,1} & S_{2,1} & S_{3,1} & S_{4,1} \\ S_{2,1} & S_{2,2} & S_{3,2} & S_{4,2} \\ S_{3,1} & S_{3,2} & S_{3,3} & S_{4,3} \\ S_{4,1} & S_{4,2} & S_{4,3} & S_{4,4} \end{pmatrix}$$

with

$$\begin{aligned} S_{1,1} &= \rho^2 - 1 + \frac{\beta_A}{\alpha} \lambda_\beta^A + \frac{\beta_B}{\alpha} \lambda_\beta^B + \lambda_L^A + \lambda_L^B, \\ S_{2,1} &= \frac{1}{2}(2\frac{\beta_A}{\alpha} \lambda_\beta^A - 2\theta + \lambda_\beta^A + 2\lambda_L^A + \lambda_\mu^A), \\ S_{3,1} &= -2\frac{\beta_A}{\alpha} \lambda_\beta^A - \frac{\beta_B}{\alpha} \lambda_\beta^B + \theta - 2\lambda_L^A - \frac{\lambda_\beta^B}{2} - \lambda_L^B - \frac{\lambda_\mu^B}{2}, \\ S_{4,1} &= \frac{\beta_A}{\alpha} \lambda_\beta^A + \lambda_L^A, \\ S_{2,2} &= \frac{\beta_A}{\alpha} \lambda_\beta^A - \theta^2 + \lambda_\beta^A + \lambda_L^A + \lambda_\mu^A \alpha \mu_A + \lambda_\mu^A - \lambda_L^A \alpha^2 L_A^2, \\ S_{3,2} &= -(2\frac{\beta_A}{\alpha} + 1)\lambda_\beta^A + \theta^2 - 2\lambda_L^A - \lambda_\mu^A, \\ S_{4,2} &= \frac{1}{2}(2\frac{\beta_A}{\alpha} \lambda_\beta^A + \lambda_\beta^A + 2\lambda_L^A + \lambda_\mu^A), \\ S_{3,3} &= 4\frac{\beta_A}{\alpha} \lambda_\beta^A + \frac{\beta_B}{\alpha} \lambda_\beta^B - \theta^2 + 4\lambda_L^A + \lambda_\beta^B + \lambda_L^B + \lambda_\mu^B \alpha \mu_B + \lambda_\mu^B + \lambda_\mu^C \alpha \mu_C - \lambda_L^B \alpha^2 L_B^2 - \lambda_L^C \alpha^2 L_C^2, \\ S_{4,3} &= \frac{1}{2}(-4\frac{\beta_A}{\alpha} \lambda_\beta^A - 4\lambda_L^A - \lambda_\beta^C - \lambda_\mu^C), \\ S_{4,4} &= \frac{\beta_A}{\alpha} \lambda_\beta^A + \frac{\beta_C}{\alpha} \lambda_\beta^C + \lambda_L^A + \lambda_L^C. \end{aligned}$$

**SM3. Proofs of results in Section 4.** We now prove Theorems 4.1 and 4.3. The approach is to provide an upper bound and a lower bound for each case (5 cases for Theorem 4.1 and 3 cases for Theorem 4.3). Since the upper and lower bounds match, weak duality tells us that the bounds are optimal, i.e., the contraction factors are tight.

In the language of the SDPs, the upper and lower bounds correspond to primal and dual feasible points, and their optimality is certified since they match (0 duality gap). Note that the strong duality result of Theorem 3.3 guarantees the existence of lower bounds matching the optimal upper bounds. Here, we explicitly provide lower bounds to certify the upper bounds are indeed optimal.

The proofs rely on inequalities that we assert by saying ‘‘It is possible to verify that ...’’ Whenever we do so, we provide a rigorous (and arduous) verification separately in Section SM4. We make this separation because the verifications are purely algebraic and do not illuminate the main proof. As an alternative means of verification, we provide code that uses symbolic manipulation to verify the inequalities.



**SM3.1. Proof of Theorem 4.1.** Define

$$\begin{aligned} R_{(a)} &= \left\{ (\mu, \beta, \theta) \mid \mu\beta - \mu + \beta < 0, \theta \leq 2 \frac{(\beta+1)(\mu-\beta-\mu\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, \mu > 0, \beta > 0, \theta \in (0, 2) \right\} \\ R_{(b)} &= \left\{ (\mu, \beta, \theta) \mid \mu\beta - \mu - \beta > 0, \theta \leq 2 \frac{\mu^2+\beta^2+\mu\beta+\mu+\beta-\mu^2\beta^2}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2}, \mu > 0, \beta > 0, \theta \in (0, 2) \right\} \\ R_{(c)} &= \left\{ (\mu, \beta, \theta) \mid \theta \geq 2 \frac{\mu\beta+\mu+\beta}{2\mu\beta+\mu+\beta}, \mu > 0, \beta > 0, \theta \in (0, 2) \right\} \\ R_{(d)} &= \left\{ (\mu, \beta, \theta) \mid \mu\beta + \mu - \beta < 0, \theta \leq 2 \frac{(\mu+1)(\beta-\mu-\mu\beta)}{\beta+\mu\beta-\mu-\mu^2-2\mu^2\beta}, \mu > 0, \beta > 0, \theta \in (0, 2) \right\} \\ R_{(e)} &= \left\{ (\mu, \beta, \theta) \mid \mu > 0, \beta > 0, \theta \in (0, 2) \right\} \setminus R_{(a)} \setminus R_{(b)} \setminus R_{(c)} \setminus R_{(d)} \end{aligned}$$

which correspond to the 5 cases of Theorem 4.1.

**SM3.1.1. Upper bounds.** By weak duality between the primal and dual OS-PEP,  $\rho$  is a valid contraction factor if there exists  $\rho$ ,  $\lambda_\mu^A \geq 0$ , and  $\lambda_\beta^B \geq 0$  such that

$$S = \begin{pmatrix} \rho^2 + \beta\lambda_\beta^B - 1 & -\theta + \frac{\lambda_\mu^A}{2} & \theta - (\frac{1}{2} + \beta)\lambda_\beta^B \\ -\theta + \frac{\lambda_\mu^A}{2} & -\theta^2 + (1 + \mu)\lambda_\mu^A & \theta^2 - \lambda_\mu^A \\ \theta - (\frac{1}{2} + \beta)\lambda_\beta^B & \theta^2 - \lambda_\mu^A & -\theta^2 + (1 + \beta)\lambda_\beta^B \end{pmatrix} \succeq 0.$$

For each of the 5 cases, we establish an upper bound by providing values for  $\rho$ ,  $\lambda_\mu^A \geq 0$ , and  $\lambda_\beta^B \geq 0$  such that  $S \succeq 0$ . We establish  $S \succeq 0$  with a sum-of-squares factorization

$$(SM3.1) \quad \text{Tr}(SG(z, z_A, z_B)) = K_1 \|m_1 z_A + m_2 z_B + m_3 z\|^2 + K_2 \|m_4 z_B + m_5 z\|^2,$$

for some  $m_1, m_2, m_3, m_4, m_5 \in \mathbb{R}$  and  $K_1, K_2 \geq 0$ , where

$$(SM3.2) \quad G(z, z_A, z_B) = \begin{pmatrix} \|z\|^2 & \langle z, z_A \rangle & \langle z, z_B \rangle \\ \langle z, z_A \rangle & \|z_A\|^2 & \langle z_A, z_B \rangle \\ \langle z, z_B \rangle & \langle z_A, z_B \rangle & \|z_B\|^2 \end{pmatrix} \in \mathbb{S}_+^3$$

for  $z, z_A, z_B \in \mathcal{H}$ . By arguments similar to that of Lemma 3.1,  $G(z, z_A, z_B) \in \mathbb{S}_+^3$  can be any  $3 \times 3$  positive semidefinite matrix. Therefore

$$\text{Tr}(SG(z, z_A, z_B)) \geq 0, \forall z, z_A, z_B \in \mathcal{H} \quad \Leftrightarrow \quad \text{Tr}(SM) \geq 0, \forall M \succeq 0 \quad \Leftrightarrow \quad S \succeq 0,$$

i.e., the sum-of-squares factorization proves  $S \succeq 0$ . (We only need 2 terms in the sum-of-squares factorization, because it turns out that the optimal  $S$  has rank at most 2.)

*Case (a).* When  $(\mu, \beta, \theta) \in R_{(a)}$ , we use

$$\rho^2 = \left(1 - \theta \frac{\beta}{\beta+1}\right)^2, \quad \lambda_\mu^A = 2\theta \frac{1+\beta}{1-\beta} \left(1 - \theta \frac{\beta}{\beta+1}\right), \quad \lambda_\beta^B = 2\theta \left(1 - \theta \frac{\beta}{\beta+1}\right).$$

This gives us the sum-of-square factorization (SM3.1) with

$$\begin{aligned} m_1 &= -1, \quad m_2 = \frac{(2-\theta)(\beta+1)}{(2-\theta)(\beta+1)+2\mu(1+\beta-\theta\beta)}, \quad m_3 = -\frac{(2-\theta)\beta}{(2-\theta)(\beta+1)+2\mu(1+\beta-\theta\beta)}, \\ m_4 &= -\frac{\beta+1}{\beta}, m_5 = 1, \quad K_1 = \theta \frac{(2-\theta)(\beta+1)+2\mu(1+\beta-\theta\beta)}{1-\beta}, \quad K_2 = 2\beta^2 \theta \frac{1+\beta-\theta\beta}{(1-\beta)(\beta+1)^2} \frac{(\beta-1)\mu(2\beta(\theta-1)+\theta-2)-(2-\theta)\beta(\beta+1)}{(2-\theta)(\beta+1)+2\mu(1+\beta-\theta\beta)}. \end{aligned}$$

It is possible to verify that there is no division by 0 in the definitions and that  $\lambda_\mu^A, \lambda_\beta^B, K_1, K_2 \geq 0$  when  $(\mu, \beta, \theta) \in R_{(a)}$ .

Case (b). When  $(\mu, \beta, \theta) \in R_{(b)}$ , we use

$$\rho^2 = \left(1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}\right)^2, \quad \lambda_\mu^A = 2\theta \frac{\beta+1}{\beta-1} \left(1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}\right), \quad \lambda_\beta^B = 2\theta \frac{\mu-1}{\mu+1} \left(1 - \theta \frac{1+\mu\beta}{(\mu+1)(\beta+1)}\right).$$

This gives us the sum-of-square factorization (SM3.1) with

$$\begin{aligned} m_1 &= -1, & m_2 &= \frac{2}{\mu+1} - \frac{(2-\theta)(\beta+1)}{(2-\theta)(\beta+1)+2\mu(1+\beta-\theta\beta)}, & m_3 &= -\frac{1}{\mu+1} + \frac{(2-\theta)\beta}{(2-\theta)(\beta+1)+2\mu(1+\beta-\theta\beta)}, \\ m_4 &= -\frac{\beta+1}{\beta}, & m_5 &= 1, & K_1 &= \theta \frac{(2-\theta)(\beta+1)+2\mu(1+\beta-\theta\beta)}{\beta-1}, \\ K_2 &= 2\theta\beta^2 \frac{(\beta+1)(\mu+1)-\theta(1+\mu\beta)}{(\mu+1)^2(\beta-1)(\beta+1)^2} \frac{\mu\beta^2(-2\theta\mu+\theta+2\mu)+\theta\beta^2+(\theta-2)\mu(\mu+1)+\beta(\theta\mu^2+\theta-2\mu-2)-2\beta^2}{(2-\theta)(\beta+1)+2\mu(1+\beta-\theta\beta)}. \end{aligned}$$

It is possible to verify that there is no division by 0 in the definitions and that  $\lambda_\mu^A, \lambda_\beta^B, K_1, K_2 \geq 0$  when  $(\mu, \beta, \theta) \in R_{(b)}$ .

Case (c). When  $(\mu, \beta, \theta) \in R_{(c)}$ , we use

$$\rho^2 = (\theta - 1)^2, \quad \lambda_\mu^A = 2\theta(\theta - 1), \quad \lambda_\beta^B = 2\theta(\theta - 1)$$

This gives us the sum-of-square factorization (SM3.1) with

$$\begin{aligned} m_1 &= -1, & m_2 &= -\frac{2-\theta}{2(\theta-1)\mu+\theta-2}, & m_3 &= \frac{2-\theta}{2(\theta-1)\mu+\theta-2}, & m_4 &= -1, & m_5 &= 1, \\ K_1 &= \theta(2(\theta-1)\mu + \theta - 2), & K_2 &= 2(\theta-1)\theta \frac{\theta\beta+\theta\mu(1+2\beta)-2(\mu+\beta+\mu\beta)}{2(\theta-1)\mu+\theta-2}. \end{aligned}$$

It is possible to verify that there is no division by 0 in the definitions and that  $\lambda_\mu^A, \lambda_\beta^B, K_1, K_2 \geq 0$  when  $(\mu, \beta, \theta) \in R_{(c)}$ .

Case (d). When  $(\mu, \beta, \theta) \in R_{(d)}$ , we use

$$\rho^2 = \left(1 - \theta \frac{\mu}{\mu+1}\right)^2, \quad \lambda_\mu^A = 2\theta \left(1 - \theta \frac{\mu}{\mu+1}\right), \quad \lambda_\beta^B = 2\theta \frac{1-\mu}{1+\mu} \left(1 - \theta \frac{\mu}{\mu+1}\right).$$

This gives us the sum-of-square factorization (SM3.1) with

$$\begin{aligned} m_1 &= -1, & m_2 &= \frac{2-\theta}{2(\theta-1)\mu+\theta-2} + \frac{2}{\mu+1}, & m_3 &= -\frac{\theta\mu}{(\mu+1)(2(\theta-1)\mu+\theta-2)}, & m_4 &= -1, & m_5 &= 1, \\ K_1 &= -\theta(2(\theta-1)\mu + \theta - 2), & K_2 &= 2\theta((\theta-1)\mu - 1) \frac{-\theta(\beta-\mu^2(1+2\beta)-\mu(1-\beta))-2(\mu+1)(\mu\beta+\mu-\beta)}{(\mu+1)^2(2(\theta-1)\mu+\theta-2)}. \end{aligned}$$

It is possible to verify that there is no division by 0 in the definitions and that  $\lambda_\mu^A, \lambda_\beta^B, K_1, K_2 \geq 0$  when  $(\mu, \beta, \theta) \in R_{(d)}$ .

Case (e). When  $(\mu, \beta, \theta) \in R_{(e)}$ , we use

$$\begin{aligned} \rho^2 &= \frac{2-\theta}{4} \frac{((2-\theta)\mu(\beta+1)+\theta\beta(1-\mu))((2-\theta)\beta(\mu+1)+\theta\mu(1-\beta))}{\mu\beta(2\mu\beta(1-\theta)+(2-\theta)(\mu+\beta+1))}, & \lambda_\mu^A &= \theta \frac{(2-\theta)\mu(\beta+1)+\theta\beta(1-\mu)}{\beta}, \\ \lambda_\beta^B &= \frac{\theta(2-\theta)}{\beta} \frac{(2-\theta)\mu(\beta+1)+\theta\beta(1-\mu)}{2\mu\beta(1-\theta)+(2-\theta)(\mu+\beta+1)}. \end{aligned}$$

This gives us the sum-of-square factorization (SM3.1) with

$$\begin{aligned} m_1 &= -2\mu \frac{2(\theta-1)\mu\beta+\theta\beta+(\theta-2)(\mu+1)-2\beta}{\beta}, & m_2 &= 2\mu \frac{(\beta+1)(\theta-2)+\beta\theta}{\beta}, & m_3 &= (\theta-2) - \mu \frac{(\beta+1)(\theta-2)+\beta\theta}{\beta} \\ m_4 &= 0, & m_5 &= 0, & K_1 &= \frac{\beta}{4\mu} \frac{\theta}{2\mu\beta(1-\theta)+(2-\theta)(\mu+\beta+1)}, & K_2 &= 0. \end{aligned}$$

It is possible to verify that there is no division by 0 in the definitions and that  $\rho^2, \lambda_\mu^A, \lambda_\beta^B, K_1, K_2 \geq 0$  when  $(\mu, \beta, \theta) \in R_{(e)}$ .

REMARK 1 (Constructing a classical proof with a dual solution). *Given  $\rho^2$ ,  $\lambda_\mu^A \geq 0$ , and  $\lambda_\beta^B \geq 0$  such that  $S \succeq 0$ , one can construct a classical proof establishing  $\rho^2$  as a valid contraction factor without relying on the OSPEP methodology. With  $G$  defined as in (SM3.2), we have*

$$\mathrm{Tr}(SG(z, z_A, z_B)) = \rho^2 \|z\|^2 - \|z - \theta(z_B - z_A)\|^2 - \lambda_\mu^A \left( \langle \Delta A, z_A \rangle - \mu \|z_A\|^2 \right) - \lambda_\beta^B \left( \langle \Delta B, z_B \rangle - \beta \|\Delta B\|^2 \right),$$

with  $\Delta A = 2z_b - z - z_A$  and  $\Delta B = z - z_B$ . The sum-of-square factorization gives us

$$\begin{aligned} \rho^2 \|z\|^2 - \|z - \theta(z_B - z_A)\|^2 - \lambda_\mu^A \left( \langle \Delta A, z_A \rangle - \mu \|z_A\|^2 \right) - \lambda_\beta^B \left( \langle \Delta B, z_B \rangle - \beta \|\Delta B\|^2 \right) \\ = K_1 \|m_1 z_A + m_2 z_B + m_3 z\|^2 + K_2 \|m_4 z_B + m_5 z\|^2. \end{aligned}$$

Reorganizing, we get

$$\begin{aligned} \|z - \theta(z_B - z_A)\|^2 = \rho^2 \|z\|^2 - \lambda_\mu^A \left( \langle \Delta A, z_A \rangle - \mu \|z_A\|^2 \right) - \lambda_\beta^B \left( \langle \Delta B, z_B \rangle - \beta \|\Delta B\|^2 \right) \\ - (\text{sum of squares}). \end{aligned}$$

Now revert the change of variables of Section 3.1 by substituting  $z \mapsto z - z'$ ,  $z - \theta(z_B - z_A) \mapsto T^{\mathrm{DRS}}(z) - T^{\mathrm{DRS}}(z')$ ,  $z_B \mapsto J_B z - J_B z'$ , and  $z_A \mapsto J_A(2J_B z - z) - J_A(2J_B z' - z')$  to get a classical proof of the form

$$\begin{aligned} \|T^{\mathrm{DRS}}(z; A, B, 1, \theta) - T^{\mathrm{DRS}}(z'; A, B, 1, \theta)\|^2 = \rho^2 \|z - z'\|^2 \\ - \lambda_\mu^A \left( \langle \Delta A, J_A(2J_B z - z) - J_A(2J_B z' - z') \rangle - \mu \|J_A(2J_B z - z) - J_A(2J_B z' - z')\|^2 \right) \\ - \lambda_\beta^B \left( \langle \Delta B, J_B z - J_B z' \rangle - \beta \|\Delta B\|^2 \right) \\ - (\text{sum of squares}) \end{aligned}$$

where now  $\Delta A = 2J_B z - z - J_A(2J_B z - z) - 2J_B z' + z' + J_A(2J_B z' - z')$  and  $\Delta B = z - J_B z - z' + J_B z'$ . Since  $A$  is  $\mu$ -strong monotone, we have

$$\langle \Delta A, J_A(2J_B z - z) - J_A(2J_B z' - z') \rangle - \mu \|J_A(2J_B z - z) - J_A(2J_B z' - z')\|^2 \geq 0.$$

Since  $B$  is  $\beta$ -cocoercive, we have

$$\langle \Delta B, J_B z - J_B z' \rangle - \beta \|\Delta B\|^2 \geq 0.$$

Since  $\lambda_\mu^A \geq 0$  and  $\lambda_\beta^B \geq 0$ , we have a valid proof establishing

$$\|T^{\mathrm{DRS}}(z; A, B, 1, \theta) - T^{\mathrm{DRS}}(z'; A, B, 1, \theta)\|^2 \leq \rho^2 \|z - z'\|^2.$$

**SM3.1.2. Lower bounds.** We now show that for the five cases, there are operators  $A \in \mathcal{M}_\mu$  and  $B \in \mathcal{C}_\beta$  and inputs  $z_1, z_2 \in \mathcal{H}$  such that

$$\|Tz_1 - Tz_2\| \geq \rho \|z_1 - z_2\|,$$

where  $T = I - \theta J_B + \theta J_A(2J_B - I)$  and  $\rho$  is given by Theorem 4.1. We construct the lower bounds for  $\mathbb{R}^2$ , since the construction can be embedded into the higher dimensional space  $\mathcal{H}$ .

*Case (a).*  $A = N_{\{0\}}$ ,  $B = \frac{1}{\beta}I$ , and  $T = \left(1 - \theta \frac{\beta}{1+\beta}\right)I$ . This construction provides the lower bound  $\rho = |1 - \theta \frac{\beta}{\beta+1}|$ , and it is valid when  $(\mu, \beta, \theta) \in R_{(a)}$ . (In fact, it is always valid.)

*Case (b).*  $A = \mu I$ ,  $B = \frac{1}{\beta}I$ , and  $T = \left(1 - \theta \frac{1+\beta\mu}{(\beta+1)(\mu+1)}\right)I$ . This construction provides the lower bound  $\rho = |1 - \theta \frac{1+\beta\mu}{(\mu+1)(\beta+1)}|$ , and it is valid when  $(\mu, \beta, \theta) \in R_{(b)}$ . (In fact, it is always valid.)

*Case (c).*  $A = N_{\{0\}}$ ,  $B = 0$ , and  $T = (1 - \theta)I$ . This construction provides the lower bound  $\rho = |1 - \theta|$ , and it is valid when  $(\mu, \beta, \theta) \in R_{(c)}$ . (In fact, it is always valid.)

*Case (d).*  $A = \mu I$ ,  $B = 0$ , and  $T = \left(1 - \theta \frac{\mu}{\mu+1}\right)I$ . This construction provides the lower bound  $\rho = |1 - \theta \frac{\mu}{\mu+1}|$ , and it is valid when  $(\mu, \beta, \theta) \in R_{(d)}$ . (In fact, it is always valid.)

*Case (e).* Define

$$K = \frac{((2-\theta)\mu(\mu+1)+\beta(\mu-1)(2-\theta+2\mu(1-\theta)))((2-\theta)\mu+\beta(2(1-\theta)\mu-\theta+2))}{\beta^2(\theta-2)\mu(2\beta(\theta-2)-\theta-2)+\beta^2(2\beta+1)(\theta-2)^2+(2\beta-1)\mu^3(2\beta(\theta-1)+\theta-2)^2-(2\beta-1)\mu^2(2\beta-\theta+2)(2\beta(\theta-1)+\theta-2)}.$$

It is possible to verify that there is no division by 0 in the definition of  $K$  and that  $0 < K < 1/\beta^2$  when  $(\mu, \beta, \theta) \in R_{(e)}$ . Let

$$A = \begin{pmatrix} \mu & -a \\ a & \mu \end{pmatrix}, \quad B = \begin{pmatrix} \frac{\beta K}{\sqrt{K - K^2 \beta^2}} & -\sqrt{K - K^2 \beta^2} \\ \sqrt{K - K^2 \beta^2} & \beta K \end{pmatrix},$$

where

$$a = \frac{2\theta\mu + \theta - 2\beta\theta K\mu - \theta K + 2K\mu + 2K - 2\mu - 2 + \sqrt{4(\theta-2)^2(\mu+1)^2(K - \beta^2 K^2) + ((\theta-2)(K-1) - 2\mu(\theta - \beta\theta K + K - 1))^2}}{2(\theta-2)\sqrt{K - \beta^2 K^2}}.$$

Since  $\frac{1}{2}(A + A^T)$  has two eigenvalues equal to  $\mu$  and  $\frac{1}{2}(B^{-T} + B^{-1})$  has two eigenvalues equal to  $\beta$ , we have  $A \in \mathcal{M}_\mu$  and  $B \in \mathcal{C}_\beta$ . Then

$$T = \begin{pmatrix} T_1 & -T_2 \\ T_2 & T_1 \end{pmatrix}$$

with

$$T_1 = \frac{-\theta(a^2(\beta K + 1) + (\mu + 1)(K(\beta\mu + \beta + 1) + \mu)) + (a^2 + (\mu + 1)^2)(2\beta K + K + 1) - 2a\theta\sqrt{K - \beta^2 K^2}}{(a^2 + (\mu + 1)^2)(2\beta K + K + 1)},$$

$$T_2 = \frac{\theta(a^2\sqrt{K - \beta^2 K^2} + a(K - 1) + (\mu^2 - 1)\sqrt{K - \beta^2 K^2})}{(a^2 + (\mu + 1)^2)(2\beta K + K + 1)},$$

and

$$T^T T = \begin{pmatrix} T_1^2 + T_2^2 & 0 \\ 0 & T_1^2 + T_2^2 \end{pmatrix}.$$

So this construction provides the lower bound  $\rho^2 = T_1^2 + T_2^2$ . Under the assumption  $\theta < 2 \frac{\beta\mu + \beta + \mu}{2\beta\mu + \beta + \mu}$  (i.e., when  $(\mu, \beta, \theta) \notin R_{(c)}$ ), the lower bound simplifies to

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

and it is valid when  $(\mu, \beta, \theta) \in R_{(e)}$ .

**SM3.2. Proof of Theorem 4.3.** Define

$$\begin{aligned} R_{(a)} &= \left\{ (\mu, L, \theta) \mid \mu \frac{-(2(\theta-1)\mu+\theta-2)+L^2(\theta-2(1+\mu))}{\sqrt{(2(\theta-1)\mu+\theta-2)^2+L^2(\theta-2(\mu+1))^2}} \leq \sqrt{L^2+1}, \mu > 0, L > 0, \theta \in (0, 2) \right\} \\ R_{(b)} &= \left\{ (\mu, L, \theta) \mid L < 1, \mu > \frac{L^2+1}{(L-1)^2}, \theta \leq \frac{2(\mu+1)(L+1)(\mu+\mu L^2-L^2-2\mu L-1)}{2\mu^2-\mu+\mu L^3-L^3-3\mu L^2-L^2-2\mu^2 L-\mu L-L-1}, \mu > 0, L > 0, \theta \in (0, 2) \right\} \\ R_{(c)} &= \left\{ (\mu, L, \theta) \mid \mu > 0, L > 0, \theta \in (0, 2) \right\} \setminus R_{(a)} \setminus R_{(b)} \end{aligned}$$

which correspond to the 3 cases of Theorem 4.3.

**SM3.2.1. Upper bounds.** The approach is similar to that of Section SM3.1.1. By weak duality between the primal and dual OSPEP,  $\rho$  is a valid contraction factor if there exists  $\rho, \lambda_\mu^A \geq 0, \lambda_L^B \geq 0,$  and  $\lambda_\mu^B \geq 0$  such that

$$S = \begin{pmatrix} \rho^2 + \lambda_L^B - 1 & \frac{\lambda_\mu^A}{2} - \theta & \theta - \lambda_L^B - \frac{\lambda_\mu^B}{2} \\ \frac{\lambda_\mu^A}{2} - \theta & -\theta^2 + \lambda_\mu^A + \lambda_\mu^A \mu & \theta^2 - \lambda_\mu^A \\ \theta - \lambda_L^B - \frac{\lambda_\mu^B}{2} & \theta^2 - \lambda_\mu^A & -\lambda_L^B L^2 - \theta^2 + \lambda_L^B + \lambda_\mu^B \end{pmatrix} \succeq 0.$$

We establish  $S \succeq 0$  with a sum-of-squares factorization

$$(SM3.3) \quad \text{Tr}(SG(z, z_A, z_B)) = K_1 \|m_1 z_A + m_2 z_B + m_3 z\|^2 + K_2 \|m_4 z_B + m_5 z\|^2,$$

for some  $m_1, m_2, m_3, m_4, m_5 \in \mathbb{R}$  and  $K_1, K_2 \geq 0$ , where  $G(z, z_A, z_B)$  is as defined in (SM3.2).

REMARK 2. *As before, the dual matrix  $S$  satisfies*

$$\text{Tr}(SG(z, z_A, z_B)) = \rho^2 \|z\|^2 - \|z_+\|^2 - \lambda_\mu^A (\langle \Delta A, z_A \rangle - \mu \|z_A\|^2) - \lambda_\mu^B \langle \Delta B, z_B \rangle - \lambda_L^B (L^2 \|z_B\|^2 - \|\Delta B\|^2),$$

with  $\Delta A = 2z_b - z - z_A$  and  $\Delta B = z - z_B$ . One can use this to construct a classical proof establishing  $\rho^2$  as a valid contraction factor without relying on the OSPEP methodology, given  $\rho, \lambda_\mu^A \geq 0, \lambda_L^B \geq 0,$  and  $\lambda_\mu^B \geq 0$  such that  $S \succeq 0$ .

Case (a). When  $(\mu, L, \theta) \in R_{(a)}$ , we define

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

which satisfies  $C > 0$  for all values of  $L, \mu, \theta > 0$ , and we use

$$\begin{aligned} \rho^2 &= \left( \frac{\theta+C}{2(\mu+1)} \right)^2, \quad \lambda_\mu^A = \frac{\theta(\theta+C)}{(\mu+1)}, \quad \lambda_L^B = \frac{(2-\theta)\theta\mu}{(\mu+1)(L^2+1)} \frac{\theta+C}{C}, \\ \lambda_\mu^B &= \frac{\theta(\theta+C)}{(\mu+1)^2 C} \left( C + \mu \frac{(2(\theta-1)\mu+\theta-2)-L^2(\theta-2(\mu+1))}{L^2+1} \right). \end{aligned}$$

This gives us the sum-of-square factorization (SM3.3) with

$$\begin{aligned} K_1 &= \theta C, \quad m_1 = -1, \quad m_2 = \frac{C-\theta\mu}{C(1+\mu)}, \quad m_3 = \frac{2(\mu+1)-(C+\theta)}{2C(\mu+1)}, \\ K_2 &= 0, \quad m_4 = 0, \quad m_5 = 0. \end{aligned}$$

When  $\mu, L > 0$  and  $\theta \in (0, 2)$ , we have  $\lambda_\mu^A, \lambda_L^B, K_1, K_2 \geq 0$ . Furthermore,  $\lambda_\mu^B \geq 0$  when

$$C \geq \mu \frac{-(2(\theta-1)\mu+\theta-2)+L^2(\theta-2(\mu+1))}{L^2+1},$$

which is immediately equivalent to the main condition defining  $R_{(a)}$ .

Case (b). When  $(\mu, L, \theta) \in R_{(b)}$ , we use

$$\rho^2 = \left(1 - \theta \frac{L+\mu}{(\mu+1)(L+1)}\right)^2$$

$$\lambda_\mu^A = 2\theta \frac{1+L}{1-L} \left(1 - \theta \frac{\mu+L}{(\mu+1)(L+1)}\right), \quad \lambda_L^B = \frac{\theta}{L} \frac{\mu-1}{\mu+1} \left(1 - \theta \frac{\mu+L}{(\mu+1)(L+1)}\right), \quad \lambda_\mu^B = 0.$$

This gives us the sum-of-square factorization (SM3.3) with

$$K_1 = \theta \frac{2(\mu+1)(L+1) - \theta(2\mu+L+1)}{1-L},$$

$$K_2 = \theta \frac{(L+1)(\mu+1) - \theta(L+\mu)}{(\mu+1)^2(1-L)L(L+1)^2} \frac{2(\mu+1)(L+1)(\mu(1-L)^2 - (L^2+1)) + \theta(\mu(1+L+3L^2-L^3) + (1+L+L^2+L^3) + 2\mu^2(L-1))}{2(\mu+1)(L+1) - \theta(2\mu+L+1)}$$

$$m_1 = -1, \quad m_2 = \frac{2}{\mu+1} - \frac{(2-\theta)(L+1)}{2(\mu+1)(L+1) - \theta(2\mu+L+1)},$$

$$m_3 = \frac{1}{1+\mu} \frac{\theta(\mu+L) - 2L(\mu+1)}{2(\mu+1)(L+1) - \theta(2\mu+L+1)}, \quad m_4 = -(1+L), \quad m_5 = 1.$$

It is possible to verify that there is no division by 0 in the definitions and that  $\lambda_\mu^A, \lambda_L^B, \lambda_\mu^B, K_1, K_2 \geq 0$  when  $(\mu, L, \theta) \in R_{(b)}$ .

Case (c). When  $(\mu, L, \theta) \in R_{(c)}$ , we use

$$\rho^2 = \frac{(2-\theta)}{4\mu(L^2+1)} \frac{(\theta(L^2+1) - 2\mu(\theta+L^2-1))(\theta(1+2\mu+L^2) - 2(\mu+1)(L^2+1))}{2\mu(\theta+L^2-1) - (2-\theta)(1-L^2)}$$

$$\lambda_\mu^A = \theta \left( \theta - \frac{2\mu(\theta+L^2-1)}{L^2+1} \right), \quad \lambda_L^B = \frac{(2-\theta)\theta}{(L^2+1)} \frac{\theta(L^2+1) - 2\mu(\theta+L^2-1)}{(2-\theta)(1-L^2) - 2\mu(\theta+L^2-1)}, \quad \lambda_\mu^B = 0.$$

This gives us the sum-of-square factorization (SM3.3) with

$$K_1 = \frac{\theta}{4\mu(L^2+1)((2-\theta)(1-L^2) - 2\mu(\theta+L^2-1))}, \quad m_1 = 4\mu^2(1-L^2-\theta) + 2\mu(2-\theta)(1-L^2),$$

$$m_2 = 4\mu(L^2+\theta-1), \quad m_3 = 2\mu(1-L^2-\theta) - (2-\theta)(L^2+1),$$

$$K_2 = 0, \quad m_4 = 0, \quad m_5 = 0.$$

It is possible to verify that there is no division by 0 in the definitions and that  $\rho^2, \lambda_\mu^A, \lambda_L^B, \lambda_\mu^B, K_1, K_2 \geq 0$  when  $(\mu, L, \theta) \in R_{(c)}$ .

**SM3.2.2. Lower bounds.** We now show that for the three cases, there are operators  $A \in \mathcal{M}_\mu$  and  $B \in \mathcal{L}_L \cap \mathcal{M}$  and inputs  $z_1, z_2 \in \mathcal{H}$  such that

$$\|Tz_1 - Tz_2\| \geq \rho \|z_1 - z_2\|,$$

where  $T = I - \theta J_B + \theta J_A(2J_B - I)$  and  $\rho$  is given by Theorem 4.3. We construct the lower bounds for  $\mathbb{R}^2$ , since the construction can be embedded into the higher dimensional space  $\mathcal{H}$ .

Case (a). Let

$$A = \mu \text{Id} + N_{\{0\} \times \mathbb{R}}, \quad B = L \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then

$$J_A = \frac{1}{\mu+1} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad J_B = \frac{1}{L^2+1} \begin{bmatrix} 1 & -L \\ L & 1 \end{bmatrix},$$

and

$$T = \begin{bmatrix} 1 - \frac{\theta}{L^2+1} & \frac{L\theta}{L^2+1} \\ -\frac{L\theta(\mu-1)}{(L^2+1)(\mu+1)} & \frac{(-\theta+\mu+1)L^2 - \theta\mu+1}{(L^2+1)(\mu+1)} \end{bmatrix}.$$

The eigenvalues of  $T^T T$  are given by

$$\frac{\left(\theta \pm \sqrt{\frac{(2(\theta-1)\mu + \theta - 2)^2 + L^2(\theta - 2(\mu+1))^2}{L^2 + 1}}\right)^2}{4(\mu+1)^2}.$$

This construction provides the lower bound

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

and it is valid when  $(\mu, L, \theta) \in R_{(a)}$ . (In fact, it is always valid.) This construction was inspired by Example 5.3 of [53].

*Case (b).*  $A = \mu I$ ,  $B = LI$ , and  $T = \left(1 - \theta \frac{L+\mu}{(L+1)(\mu+1)}\right)$ . This construction provides the lower bound  $\rho = |1 - \theta \frac{L+\mu}{(L+1)(\mu+1)}|$ , and it is valid when  $(\mu, L, \theta) \in R_{(b)}$ . (In fact, it is always valid.)

*Case (c).* Define

$$K = \frac{L^2+1}{2L} \frac{(\mu-1)(L^2(\theta-2(\mu+1)) - (2(\theta-1)\mu + \theta - 2)^2 - 4(\theta-2)^2(\mu+1)L^2)}{4\mu^2(\theta+L^2-1)((1-\theta)(L^2-\mu) + L^2\mu-1) + (\theta-2)^2(\mu+1)(L^2+1)^2}.$$

It is possible to verify that there is no division by 0 in the definition of  $K$  and that  $0 \leq K \leq 1$  when  $(\mu, L, \theta) \in R_{(c)}$ . Let

$$A = \mu \text{Id} + N_{\mathbb{R} \times \{0\}}, \quad B = L \begin{pmatrix} K & -\sqrt{1-K^2} \\ \sqrt{1-K^2} & K \end{pmatrix}.$$

Since the eigenvalues of  $B^T B$  and  $\frac{1}{2}(B^T + B)$  are respectively both equal to  $L^2$  and  $KL$ , we have  $B \in \mathcal{M} \cap \mathcal{L}_L$ . Then

$$J_A = \frac{1}{\mu+1} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad J_B = \begin{pmatrix} \frac{KL+1}{L^2+2KL+1} & \frac{\sqrt{1-K^2}L}{L^2+2KL+1} \\ -\frac{\sqrt{1-K^2}L}{L^2+2KL+1} & \frac{KL+1}{L^2+2KL+1} \end{pmatrix},$$

and

$$T = \begin{pmatrix} 1 - \frac{KL\theta + \theta}{L^2+2KL+1} & -\frac{\sqrt{1-K^2}L\theta}{L^2+2KL+1} \\ \frac{\sqrt{1-K^2}L\theta(\mu-1)}{(L^2+2KL+1)(\mu+1)} & \frac{(-\theta+\mu+1)L^2 - K(\theta-2)(\mu+1)L - \theta\mu + \mu+1}{(L^2+2KL+1)(\mu+1)} \end{pmatrix}.$$

The eigenvalues of  $T^T T$  are

$$\frac{T_1 \pm \theta \sqrt{T_2}}{2(\mu+1)^2(2KL+L^2+1)}$$

with

$$T_1 = \theta^2 (2K\mu L + 2\mu(\mu+1) + L^2 + 1) - 2\theta(\mu+1)(2K(\mu+1)L + 2\mu + L^2 + 1) + 2(\mu+1)^2(2KL + L^2 + 1),$$

$$T_2 = (-2KL + L^2 + 1)((2(\theta-1)\mu + \theta - 2)^2 + 2KL(\theta - 2(\mu+1))(2(\theta-1)\mu + \theta - 2) + L^2(\theta - 2\mu - 2)^2).$$

This construction provides the lower bound

$$\rho^2 = \frac{T_1 + \theta \sqrt{T_2}}{2(\mu+1)^2(2KL+L^2+1)}.$$

Under the assumption

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

(i.e., when  $(\mu, L, \theta) \notin R_{(a)}$ ), the lower bound simplifies to

$$\rho = \sqrt{\frac{(2-\theta)}{4\mu(L^2+1)} \frac{(\theta(L^2+1)-2\mu(\theta+L^2-1))(\theta(1+2\mu+L^2)-2(\mu+1)(L^2+1))}{2\mu(\theta+L^2-1)-(2-\theta)(1-L^2)}},$$

and it is valid when  $(\mu, L, \theta) \in R_{(c)}$ .

**SM4. Algebraic verification of inequalities.** We now provide the algebraic verifications of the inequalities asserted in Section SM3. The proofs of this section are based on elementary arguments and arduous algebra. To help readers follow and verify the basic but tedious computation, we provide code that uses symbolic manipulation to verify the steps.

**SM4.1. Inequalities for Theorem 4.1.**

**SM4.1.1. Upper bounds.** It remains to show that there is no division by 0 and  $\lambda_\mu^A, \lambda_\beta^B, K_1, K_2 \geq 0$  in each case.

*Case (a).* Assume  $0 < \mu, 0 < \beta, 0 < \theta < 2$ , and

$$\begin{aligned} \text{(a1)} \quad & \mu - \mu\beta - \beta > 0 \\ \text{(a2)} \quad & \theta \leq 2 \frac{(\beta+1)(\mu-\mu\beta-\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2} \quad (\text{no division by 0 implied, see (ii) below}) \end{aligned}$$

Then:

- (i) From (a1) it is direct to note that  $1 - \beta > \beta\mu^{-1} > 0$  and hence also  $\beta < 1$ .
- (ii) As the numerator of (a2) is positive (from  $\beta > 0$  and (a1)) and as  $\theta > 0$ , the denominator of (a2) is positive, i.e.,  $\mu + \mu\beta - \beta - \beta^2 - 2\mu\beta^2 \geq 0$ . To prove strict positivity, assume for contradiction that  $\mu + \mu\beta - \beta - \beta^2 - 2\mu\beta^2 = 0$ . This implies

$$\beta(\beta + 2(\beta - 1)\mu) = \mu - \mu\beta - \beta \stackrel{\text{(a1)}}{>} 0,$$

and hence

$$0 < \beta + 2(\beta - 1)\mu \stackrel{\text{(a1)}}{<} \mu - \mu\beta + 2(\beta - 1)\mu = \mu(\beta - 1) \stackrel{\text{(i)}}{<} 0,$$

which is a contradiction. Therefore, we conclude

$$\mu + \mu\beta - \beta - \beta^2 - 2\mu\beta^2 > 0.$$

- (iii) Multiply both sides of (a2) by  $\beta$ , reorganize, and use (a1) to get

$$1 + \beta - \theta\beta \stackrel{\text{(a2)}}{\geq} \frac{(1-\beta^2)(\mu-\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2} \stackrel{\text{(a1)}}{>} \frac{\mu\beta(1-\beta^2)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2} > 0.$$

- (iv) Multiply both sides of (a2) by the denominator of (a2) (which is positive by (ii)) and reorganize to get

$$(\beta - 1)\mu(2\beta(\theta - 1) + \theta - 2) - (2 - \theta)\beta(\beta + 1) \geq 0.$$

- (v) Multiply both sides of (a2) by  $\beta/(1 + \beta)$  and reorganize to get

$$1 - \theta \frac{\beta}{\beta+1} \geq \frac{(1-\beta)(\mu-\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2} > 0.$$

(v) shows  $\lambda_\mu^A$  and  $\lambda_\beta^B$  are nonnegative. (i) shows the denominator of  $K_1$  is positive. (iii) shows the numerator of  $K_1$  is nonnegative. (i) and (iii) show the denominator of  $K_2$  is positive. (iii) and (iv) show the numerator of  $K_2$  is nonnegative. (iii) shows the denominator of  $m_2$  and  $m_3$  are positive.



Case (b). Assume  $0 < \mu$ ,  $0 < \beta$ ,  $0 < \theta < 2$ , and

$$(b1) \quad \mu\beta - \mu - \beta > 0$$

$$(b2) \quad \theta \leq \frac{2(\mu+1)(\mu+\beta+\beta^2-\mu\beta^2)}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2} \quad (\text{no division by 0 implied, see (ii) below})$$

Then:

- (i) From (b1), we have  $\mu\beta > \mu + \beta$ . Therefore  $\mu > 1 + \mu/\beta > 1$  and  $\beta > 1 + \beta/\mu > 1$ .
- (ii) The numerator of (b2) is negative since  $\mu + \beta + \beta^2 - \mu\beta^2 = (1 + \beta)(\mu - \mu\beta + \beta) \stackrel{(b1)}{<} 0$ . Since  $\theta > 0$ , the denominator of (b2) is nonpositive. To prove strict negativity, we view the denominator of (b2) as a quadratic function of  $\mu$ :

$$\begin{aligned} \phi_\beta(\mu) &= \mu^2 + \beta^2 + \mu^2\beta + \mu\beta^2 + \mu + \beta - 2\mu^2\beta^2 \\ &= \underbrace{(1 + \beta - 2\beta^2)}_{<0 \text{ by (i)}} \mu^2 + (1 + \beta^2)\mu + \beta + \beta^2. \end{aligned}$$

This quadratic is nonnegative only between its roots and

$$\phi_\beta(0) = \beta + \beta^2 > 0, \quad \phi_\beta(\beta/(\beta - 1)) = -\frac{\beta^2}{\beta - 1} < 0.$$

Therefore  $\phi_\beta(\mu) < 0$  for any  $\mu > \beta/(\beta - 1)$ , which holds by (b1). Therefore we conclude denominator of (b2) is strictly negative.

- (iii) Multiply both sides of (b2) by  $(1 + \beta + 2\mu\beta)$  and reorganize to get

$$(2 - \theta)(\beta + 1) + 2\mu(1 + \beta - \theta\beta) \geq -\frac{2\mu^2(\mu+1)(\beta^2-1)}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2} > 0.$$

The latter inequality follows from (i) and (ii).

- (iv) Multiply both sides of (b2) by  $(1 + \mu\beta)$  and reorganize to get

$$(\beta + 1)(\mu + 1) - \theta(1 + \mu\beta) \geq -\frac{(\mu^2-1)(\beta^2-1)(\mu+\beta)}{\mu^2+\beta^2+\mu^2\beta+\mu\beta^2+\mu+\beta-2\mu^2\beta^2} > 0.$$

The latter inequality follows from (i) and (ii).

- (v) Multiply both sides of (b2) by  $-(\mu^2 + \beta^2 + \mu^2\beta + \mu\beta^2 + \mu + \beta - 2\mu^2\beta^2)$  (which is positive by (ii)) and reorganize to get

$$\mu\beta^2(-2\theta\mu + \theta + 2\mu) + \theta\beta^2 + (\theta - 2)\mu(\mu + 1) + \beta(\theta\mu^2 + \theta - 2\mu - 2) - 2\beta^2 \geq 0.$$

(i) and (iv) show  $\lambda_\mu^A$  and  $\lambda_\beta^B$  are nonnegative. (i) shows the denominator of  $K_1$  is positive. (iii) shows the numerator of  $K_1$  is nonnegative. (i) and (iii) show the denominator of  $K_2$  is positive. (iv) and (v) show the numerator of  $K_2$  is nonnegative. (iii) shows the denominators of  $m_2$  and  $m_3$  are positive.

Case (c). Assume  $0 < \mu$ ,  $0 < \beta$ ,  $0 < \theta < 2$ , and

$$(c1) \quad \theta \geq 2\frac{\mu+\beta+\mu\beta}{\mu+\beta+2\mu\beta}$$

Then:

- (i) Multiply both sides of (c1) by  $(1 + 2\mu)$  and reorganize to get

$$2(\theta - 1)\mu + \theta - 2 \geq \frac{2\mu^2}{\mu + \beta + 2\mu\beta} > 0.$$

(ii) Reorganize (c1) to get

$$\theta \geq 2 \frac{\mu + \beta + \mu\beta}{\mu + \beta + 2\mu\beta} = 2 \frac{1}{1 + \frac{\mu\beta}{\mu + \beta + \mu\beta}} > 1.$$

(iii) Multiply both sides of (c1) by  $(\mu + \beta + 2\mu\beta)$  to get  $\theta\beta + \theta\mu(1 + 2\beta) - 2(\mu + \beta + \mu\beta) \geq 0$ .

(ii) shows  $\lambda_\mu^A$  and  $\lambda_\beta^B$  are nonnegative. (i) shows  $K_1$  is nonnegative. (i) shows the denominator of  $K_2$  is positive. (ii) and (iii) show the numerator of  $K_2$  is nonnegative. (i) shows the denominators of  $m_2$  and  $m_3$  are positive.

*Case (d).* Assume  $0 < \mu$ ,  $0 < \beta$ ,  $0 < \theta < 2$ , and

$$(d1) \quad \mu + \mu\beta - \beta < 0$$

$$(d2) \quad \theta \leq \frac{2(\mu+1)(\beta-\mu-\mu\beta)}{\beta+\beta\mu-\mu-\mu^2-2\beta\mu^2} \quad (\text{no division by 0 implied, see (ii) below})$$

Then:

(i) From (d1) it is direct to note that  $\mu < \frac{\beta}{\beta+1} < 1$  and  $\beta > \frac{\mu}{1-\mu} > \mu$ .

(ii) As the numerator of (d2) is positive (from  $\mu > 0$  and (d1)) and as  $\theta > 0$ , the denominator of (d2) is nonnegative. To prove strict positivity of the of the denominator, assume for contradiction that  $\beta + \beta\mu - \mu - \mu^2 - 2\beta\mu^2 = 0$ . This implies that

$$-2\beta\mu^2 + 2\beta\mu - \mu^2 = \beta\mu - \beta + \mu \stackrel{(d1)}{<} 0,$$

and hence

$$0 > 2\beta(1 - \mu) - \mu \stackrel{(i)}{>} \mu > 0,$$

which is a contradiction. Therefore we conclude

$$\beta + \beta\mu - \mu - \mu^2 - 2\beta\mu^2 > 0.$$

(iii) Multiply both sides of (d2) by  $(1 + 2\mu)$  and reorganize to get

$$2(\theta - 1)\mu + \theta - 2 \leq -\frac{2\mu^2(\mu+1)}{\beta+\beta\mu-\mu-\mu^2-2\beta\mu^2} < 0.$$

(iv) Multiply both sides of (d2) by  $\mu$  and reorganize to get

$$(\theta - 1)\mu - 1 \leq -\frac{(1-\mu^2)(\beta-\mu)}{\beta+\beta\mu-\mu-\mu^2-2\beta\mu^2} < 0,$$

where the latter inequality follows from (i) and (ii).

(v) Multiply both sides of (d2) by the denominator of (d2) (which is positive by (ii)) and reorganize to get

$$-\theta(\beta - \mu^2(1 + 2\beta) - \mu(1 - \beta)) - 2(\mu + 1)(\mu\beta + \mu - \beta) \geq 0.$$

(vi) Multiply both sides of (d2) by  $\mu/(1 + \mu)$  and reorganize to get

$$1 - \theta \frac{\mu}{\mu+1} \geq \frac{(1-\mu)(\beta-\mu)}{\beta+\beta\mu-\mu-\mu^2-2\beta\mu^2} > 0,$$

where the latter inequality follows from (i) and (ii).

(vi) shows  $\lambda_\mu^A$  is nonnegative. (i) and (vi) show  $\lambda_\beta^B$  is nonnegative. (iii) shows  $K_1$  is nonnegative. (iv) and (v) show the numerator of  $K_2$  is nonpositive. (iii) shows the denominator of  $K_2$  is negative. (iii) shows the denominators of  $m_2$  and  $m_3$  are negative.

Case (e). Assume  $0 < \mu$ ,  $0 < \beta$ , and  $0 < \theta < 2$ . First, we show that if  $(\mu, \beta, \theta) \in R_{(e)}$ , then

$$(e1) \quad \theta < \frac{2(\mu+1)(1+\beta)}{2\mu\beta+\mu+\beta+1}$$

If (e1) does not hold, i.e.,

$$\theta \geq 2 \frac{\mu+\beta+\mu\beta+1}{\mu+\beta+2\mu\beta+1},$$

then

$$\begin{aligned} \theta(\mu + \beta + 2\mu\beta) - 2(\mu + \beta + \mu\beta) &\geq 2 \frac{\mu+\beta+\mu\beta+1}{\mu+\beta+2\mu\beta+1} (\mu + \beta + 2\mu\beta) - 2(\mu + \beta + \mu\beta) \\ &= \frac{2\mu\beta}{\mu+\beta+2\mu\beta+1} > 0, \end{aligned}$$

which implies (c1). So

$$(\mu, \beta, \theta) \in R_{(e)} \quad \Rightarrow \quad (\mu, \beta, \theta) \notin R_{(c)} \quad \Rightarrow \quad (e1).$$

(i) We have

$$(2 - \theta)\mu(\beta + 1) + \theta\beta(1 - \mu) = \theta(\beta - (2\beta + 1)\mu) + 2(\beta + 1)\mu > 0,$$

because either  $\beta - (2\beta + 1)\mu \geq 0$  and the inequality immediately follows or  $\beta - (2\beta + 1)\mu < 0$  and we use (e1) to get

$$\begin{aligned} \theta(\beta - (2\beta + 1)\mu) + 2(\beta + 1)\mu &> \frac{2(\mu+1)(1+\beta)}{2\mu\beta+\mu+\beta+1} (\beta - (2\beta + 1)\mu) + 2(\beta + 1)\mu \\ &= \frac{2\beta(\beta+1)}{\mu+\beta+2\mu\beta+1} > 0. \end{aligned}$$

(ii) We have  $2\mu\beta(1 - \theta) + (2 - \theta)(\mu + \beta + 1) > 0$ , because

$$\begin{aligned} 2\mu\beta(1 - \theta) + (2 - \theta)(\mu + \beta + 1) &= -\theta(\mu + \beta + 2\mu\beta + 1) + 2(\mu + 1)(\beta + 1) \\ &> -\frac{2(\mu+1)(\beta+1)}{\mu+\beta+2\mu\beta+1} (\mu + \beta + 2\mu\beta + 1) + 2(\mu + 1)(\beta + 1) = 0, \end{aligned}$$

where the inequality follows from plugging in (e1).

(iii) We have

$$(2 - \theta)(1 + \mu)\beta - \theta\mu(\beta - 1) = 2(\mu + 1)\beta + \theta(\mu(1 - 2\beta) - \beta) > 0,$$

because either  $\mu(1 - 2\beta) - \beta \geq 0$  and the inequality immediately follows or  $\mu(1 - 2\beta) - \beta < 0$  and we use

$$\begin{aligned} 2(\mu + 1)\beta + \theta(\mu(1 - 2\beta) - \beta) &> 2(\mu + 1)\beta + \frac{2(\mu+1)(1+\beta)}{2\mu\beta+\mu+\beta+1} (\mu(1 - 2\beta) - \beta) \\ &= \frac{2\mu(\mu+1)}{2\mu\beta+\mu+\beta+1} > 0. \end{aligned}$$

(ii) shows that the denominator of  $\rho^2$  is positive. (i) and (iii) show that the numerator of  $\rho^2$  is nonnegative. (i) shows that  $\lambda_\mu^A$  is nonnegative. (i) shows that the numerator of  $\lambda_\beta^B$  is nonnegative. (ii) shows that the denominator of  $\lambda_\beta^B$  is positive. (ii) shows that the denominator of  $K_1$  is positive. Hence we arrive to  $\rho, \lambda_\mu^A, \lambda_\beta^B, K_1, K_2 \geq 0$  in the region of interest.

**SM4.1.2. Lower bounds.** It remains to show that  $(\mu, \beta, \theta) \in R_{(e)} \Rightarrow 0 < K < 1/\beta^2$ . In what follows, we first show  $0 < K$  in part I, and we then show  $K < 1/\beta^2$  in part II.

Before we proceed, let us introduce the symbol  $\neg$  that denotes the logical negation, and point out the following elementary fact. If  $f(\theta) = a\theta + b$  is an affine function of  $\theta$ , then either

$$f(\theta) < \max\{f(\theta_{\min}), f(\theta_{\max})\} \quad \text{for all } \theta \in (\theta_{\min}, \theta_{\max})$$

or

$$f(\theta) = f(\theta_{\min}) = f(\theta_{\max}).$$

The two cases respectively correspond to  $a \neq 0$  and  $a = 0$ .

*Part I.* We now prove  $0 < K$ . Define

$$\begin{aligned} a_1 &= -\theta(2\beta\mu + \beta + \mu) + 2(\beta\mu + \beta + \mu) \\ a_2 &= \theta(-(2\beta + 1)\mu^2 + (\beta - 1)\mu + \beta) + 2(\mu + 1)(\beta(\mu - 1) + \mu) \end{aligned}$$

and

$$\begin{aligned} K_{\text{den}} &= \theta^2 (8\beta^3\mu^3 + 2\beta^3\mu + 2\beta^3 + 4\beta^2\mu^3 + 4\beta^2\mu^2 - \beta^2\mu + \beta^2 - 2\beta\mu^3 - \mu^3 - \mu^2) \\ &\quad - 4\theta (4\beta^3\mu^3 + 2\beta^3\mu^2 + 2\beta^3\mu + 2\beta^3 + 4\beta^2\mu^3 + 3\beta^2\mu^2 + \beta^2 - \beta\mu^3 - \mu^3 - \mu^2) \\ &\quad + 4(\mu + 1) (2\beta^3\mu^2 + 2\beta^3 + 3\beta^2\mu^2 + \beta^2 - \mu^2) \end{aligned}$$

so that

$$K = \frac{a_1 a_2}{K_{\text{den}}}.$$

We have  $a_1 > 0$  since  $(\mu, \beta, \theta) \notin R_{(c)}$ , i.e.,  $\theta < 2\frac{\beta\mu + \beta + \mu}{2\beta\mu + \beta + \mu}$ .

We now show  $a_2 > 0$ . Since  $(\mu, \beta, \theta) \notin R_{(d)}$ , we have 2 cases:

- (i) Assume  $\neg(\text{d1})$ . Since  $a_2$  is an affine function of  $\theta$  and  $\theta \in \left(0, 2\frac{\beta\mu + \beta + \mu}{2\beta\mu + \beta + \mu}\right)$ , either

$$a_2 > \min \left\{ a_2|_{\theta=0}, a_2|_{\theta=2\frac{\beta\mu + \beta + \mu}{2\beta\mu + \beta + \mu}} \right\} = \min \left\{ 2(\mu + 1)(\beta\mu - \beta + \mu), \frac{4\beta\mu^2}{2\beta\mu + \beta + \mu} \right\} \geq 0,$$

where the first term is nonnegative by  $\neg(\text{d1})$ , or

$$a_2 = 2(\mu + 1)(\beta\mu - \beta + \mu) = \frac{4\beta\mu^2}{2\beta\mu + \beta + \mu} > 0.$$

In both cases,  $a_2 > 0$ .

- (ii) Assume  $(\text{d1})$  and  $\neg(\text{d2})$ . In Section SM4.1.1 case (d) part (ii), we proved that  $(\text{d1})$  implies  $-2\beta\mu^2 + \beta\mu + \beta - \mu^2 - \mu > 0$ . Since  $a_2$  is an affine function of  $\theta$  and  $\theta \in \left(\frac{2(\mu+1)(\beta(-\mu)+\beta-\mu)}{-2\beta\mu^2+\beta\mu+\beta-\mu^2-\mu}, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ , either

$$a_2 > \min \left\{ a_2|_{\theta=\frac{2(\mu+1)(\beta(-\mu)+\beta-\mu)}{-2\beta\mu^2+\beta\mu+\beta-\mu^2-\mu}}, a_2|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} \right\} = \min \left\{ 0, \frac{4\beta\mu^2}{2\beta\mu + \beta + \mu} \right\} = 0,$$

or

$$a_2 = 0 = \frac{4\beta\mu^2}{2\beta\mu + \beta + \mu}.$$

The latter case is impossible and we conclude  $a_2 > 0$ .

We now show  $K_{\text{den}} > 0$ . We will use the following elementary fact. Let  $f(\theta) = a\theta^2 + b\theta + c$  be a quadratic function. If  $a \leq 0$ ,  $f(\theta_{\min}) > 0$ , and  $f(\theta_{\max}) > 0$ , then  $f(\theta) > 0$  for all  $\theta \in (\theta_{\min}, \theta_{\max})$ . If  $a > 0$ ,  $f(\theta_{\min}) > 0$ ,  $f(\theta_{\max}) > 0$ , and  $f'(\theta_{\max}) < 0$ , then  $f(\theta) > 0$  for all  $\theta \in (\theta_{\min}, \theta_{\max})$ . Consider the 2 cases:

- (i) Assume  $\neg(\mathbf{a1})$ . Under this case,  $\theta \in \left(0, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ . We view  $K_{\text{den}}$  as a quadratic function of  $\theta$ . First, note

$$K_{\text{den}}|_{\theta=0} = 4(\mu+1)((2\beta+1)\beta^2 + (\beta+1)^2(2\beta-1)\mu^2)$$

and define  $\phi_\mu(\beta) = (2\beta+1)\beta^2 + (\beta+1)^2(2\beta-1)\mu^2$ . Since

$$\begin{aligned}\phi_\mu\left(\frac{\mu}{\mu+1}\right) &= \frac{4\mu^5}{(\mu+1)^3} > 0 \\ \frac{d\phi_\mu}{d\beta}(\beta) &= 2\beta(3(\beta+1)\mu^2 + 3\beta + 1) > 0\end{aligned}$$

and  $\beta \geq \frac{\mu}{\mu+1}$  by  $\neg(\mathbf{a1})$ , we have  $\phi_\mu(\beta) > 0$  and  $K_{\text{den}}|_{\theta=0} > 0$ . We have

$$K_{\text{den}}|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} = \frac{16\beta^3\mu^2(\beta+\mu)}{(2\beta\mu+\beta+\mu)^2} > 0.$$

Finally, note

$$\frac{dK_{\text{den}}}{d\theta}|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} = -\frac{4\beta\mu(4\beta^3\mu + 2\beta^3 + 4\beta^2\mu^2 + 2\beta^2 + \beta\mu - \mu^2)}{2\beta\mu + \beta + \mu}$$

and define  $\phi_\mu(\beta) = (4\beta^3\mu + 2\beta^3 + 4\beta^2\mu^2 + 2\beta^2 + \beta\mu - \mu^2)$ . Since

$$\begin{aligned}\phi_\mu\left(\frac{\mu}{\mu+1}\right) &= \frac{\mu^2(3\mu(\mu+1)^2+2)}{(\mu+1)^3} > 0 \\ \frac{d\phi_\mu}{d\beta}(\beta) &= 2\beta(\beta(6\mu+3) + 4\mu^2 + 2) + \mu > 0\end{aligned}$$

and  $\beta \geq \frac{\mu}{\mu+1}$  by  $\neg(\mathbf{a1})$ , we have  $\phi_\mu(\beta) > 0$  and  $\frac{dK_{\text{den}}}{d\theta}|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} < 0$ .

Therefore, the quadratic function satisfies  $K_{\text{den}} > 0$  for  $\theta \in \left(0, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ .

- (ii) Assume  $(\mathbf{a1})$  and  $\neg(\mathbf{a2})$ . Under this case,  $\theta \in \left(\frac{2(\beta+1)(\mu-\mu\beta-\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ , and, in Section [SM4.1.1](#) case (a) part (ii), we proved that  $(\mathbf{a1})$  implies  $\mu + \mu\beta - \beta - \beta^2 - 2\mu\beta^2 > 0$ .

We first show  $\mu > \beta$ . The fact that  $\theta$  is in the interval implies

$$\begin{aligned}\frac{2(\beta+1)(\mu-\mu\beta-\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2} &< 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu} \\ \Leftrightarrow \frac{4\beta^2\mu}{\beta^2(2\mu+1) + \beta(1-\mu) - \mu} &= (2\beta\mu + \beta + \mu)\frac{2(\beta+1)(\mu-\mu\beta-\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2} - 2(\beta\mu + \beta + \mu) < 0 \\ \Leftrightarrow \beta^2(2\mu+1) + \beta(1-\mu) - \mu &< 0.\end{aligned}$$

Define

$$\phi_\mu(\beta) = \beta^2 \underbrace{(2\mu+1)}_{>0} + \beta(1-\mu) - \mu.$$

Since the coefficient of the quadratic term is positive,  $\phi_\mu(0) = -\mu < 0$ , and  $\phi_\mu(\mu) = 2\mu^3 > 0$ , we conclude that  $\phi_\mu(\beta) < 0$  is only possible when  $\beta < \mu$ , i.e.,  $\phi_\mu(\beta) < 0 \Rightarrow \mu > \beta$ .

Now we view  $K_{\text{den}}$  as a quadratic function of  $\theta$ . We have

$$K_{\text{den}} \Big|_{\theta = \frac{2(\beta+1)(\mu-\mu\beta-\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}} = \frac{16\beta^5\mu^2(\mu-\beta)}{(\mu-\beta(2\beta\mu+\beta-\mu+1))^2} > 0 \quad (\text{since } \mu > \beta)$$

$$(SM4.1) \quad K_{\text{den}} \Big|_{\theta = 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} = \frac{16\beta^3\mu^2(\beta+\mu)}{(2\beta\mu+\beta+\mu)^2} > 0.$$

Define

$$\phi(\beta, \mu) = 8\beta^3\mu^3 + 2\beta^3\mu + 2\beta^3 + 4\beta^2\mu^3 + 4\beta^2\mu^2 - \beta^2\mu + \beta^2 - 2\beta\mu^3 - \mu^3 - \mu^2,$$

which is the coefficient for the quadratic term of  $K_{\text{den}}$ . If  $\phi(\beta, \mu) \leq 0$  (i.e., the curvature  $K_{\text{den}}$  of is nonpositive), then the two inequalities (SM4.1) implies

$$K_{\text{den}} > 0 \text{ for } \theta \in \left( \frac{2(\beta+1)(\mu-\mu\beta-\beta)}{\mu+\mu\beta-\beta-\beta^2-2\mu\beta^2}, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu} \right).$$

Now assume  $\phi(\beta, \mu) > 0$ . In this case, we have

$$\begin{aligned} \frac{dK_{\text{den}}}{d\theta} \Big|_{\theta = 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} &= -\frac{4\beta\mu(\beta^3(4\mu+2)+\beta^2(4\mu^2+2)+\beta\mu-\mu^2)}{2\beta\mu+\beta+\mu} \\ &< -\frac{4\beta\mu}{2\beta\mu+\beta+\mu} (\beta^3 4\mu + \beta^2 4\mu^2 - \mu^2) \\ &= -\frac{4\beta\mu^2}{2\beta\mu+\beta+\mu} (4\beta^2(\beta + \mu) - \mu) \end{aligned}$$

If  $4\beta^2(\beta + \mu) - \mu > 0$ , then  $\frac{dK_{\text{den}}}{d\theta} \Big|_{\theta = 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} < 0$ , and we conclude that

$K_{\text{den}} > 0$ .

We now show  $4\beta^2(\beta + \mu) - \mu > 0$ . Assume for contradiction that  $4\beta^2(\beta + \mu) - \mu \leq 0$ , or equivalently,

$$4\beta^2(\beta + \mu) - \mu \leq 0 \quad \Leftrightarrow \quad 4\beta^3 \leq \mu(1 - 4\beta^2),$$

which implies  $\beta < 1/2$  and hence

$$4\beta^2(\beta + \mu) - \mu \leq 0 \quad \Leftrightarrow \quad 4\beta^3 \leq \mu(1 - 4\beta^2) \quad \Leftrightarrow \quad \mu \geq \frac{4\beta^3}{1-4\beta^2},$$

and also

$$\frac{d\phi}{d\mu}(\beta, \mu) = (2\beta - 1) (\beta^2 + 4\beta\mu + 3(2\beta\mu + \mu)^2 + 2\mu) < 0.$$

We have assumed for contradiction that  $\mu \geq \frac{4\beta^3}{1-4\beta^2}$ , and on the other hand, (a1) and  $\beta < 1/2$  imply  $\mu > \frac{\beta}{1-\beta}$ . If

$$\phi\left(\beta, \max\left\{\frac{4\beta^3}{1-4\beta^2}, \frac{\beta}{1-\beta}\right\}\right) < 0,$$

then  $\phi(\beta, \mu) < 0$ , the contradiction forces us to conclude  $4\beta^2(\beta + \mu) - \mu > 0$ .

(A) If  $\frac{4\beta^3}{1-4\beta^2} \geq \frac{\beta}{1-\beta} \Leftrightarrow -4\beta^3 + 8\beta^2 - 1 \geq 0$ , then

$$\phi\left(\beta, \frac{4\beta^3}{1-4\beta^2}\right) = -\frac{\beta^2(64\beta^7 - 16\beta^5 - 16\beta^4 + 4\beta^3 + 8\beta^2 - 1)}{(2\beta-1)^2(2\beta+1)} < 0,$$

since

$$\begin{aligned} 64\beta^7 - 16\beta^5 - 16\beta^4 + 4\beta^3 + 8\beta^2 - 1 &\geq 64\beta^7 - 16\beta^5 - 16\beta^4 + 8\beta^3 \quad (\text{since } -4\beta^3 + 8\beta^2 - 1 \geq 0) \\ &= 8\beta^3(2\beta - 1)^2(2\beta^2 + 2\beta + 1) \\ &> 0 \quad (\text{using } \beta < \frac{1}{2}). \end{aligned}$$

(B) If  $\frac{4\beta^3}{1-4\beta^2} < \frac{\beta}{1-\beta} \Leftrightarrow -4\beta^3 + 8\beta^2 - 1 < 0$ , then

$$\phi\left(\beta, \frac{\beta}{1-\beta}\right) = \frac{\beta^3(8\beta^3+3\beta-2)}{(1-\beta)^3} < 0,$$

since

$$-4\beta^3 + 8\beta^2 - 1 < 0 \text{ and } \beta \in (0, 1/2) \Rightarrow 8\beta^3 + 3\beta - 2 < 0.$$

This final point can be verified with simple plotting or with the following algebraic argument. Define  $\chi(\beta) = -4\beta^3 + 8\beta^2 - 1$  and  $\psi(\beta) = 8\beta^3 + 3\beta - 2$ . Since  $\chi'(\beta) = 4(4-3\beta)\beta$  and  $\psi'(\beta) = 3(8\beta^2 + 1)$ , the functions  $\chi$  and  $\psi$  are increasing on  $\beta \in (0, 1/2)$ . We have  $\chi(0) < 0$  and  $\psi(0) < 0$ . Furthermore,

$$\chi\left(\frac{\sqrt[3]{2}(3\sqrt{2}+4)^{2/3}-2^{2/3}}{4\sqrt[3]{3\sqrt{2}+4}}\right) = \frac{1}{4}\left(\frac{\sqrt[3]{2}(48\sqrt{2}+67)}{(3\sqrt{2}+4)^{5/3}} + \frac{\sqrt[3]{2}(62\sqrt{2}+87)}{(3\sqrt{2}+4)^{4/3}} - 16\right) \approx 0.207156 > 0$$

and

$$\psi\left(\frac{\sqrt[3]{2}(3\sqrt{2}+4)^{2/3}-2^{2/3}}{4\sqrt[3]{3\sqrt{2}+4}}\right) = 0.$$

So  $\chi$  and  $\psi$  are increasing functions and  $\chi$  hits its first positive root before  $\psi$ . Therefore  $\chi(\beta) < 0 \Rightarrow \psi(\beta) < 0$  on  $\beta \in (0, \frac{1}{2})$ .

*Part II.* We now prove  $K < 1/\beta^2$ . Define

$$\begin{aligned} a_3 &= \theta(\beta^2(2\mu^2 - \mu - 1) - \beta(\mu^2 + 1) - \mu(\mu + 1)) + 2(\beta + 1)(\mu + 1)(\beta + \mu - \mu\beta), \\ a_4 &= \theta(\beta^2(2\mu + 1) - \beta\mu + \beta - \mu) - 2(\beta + 1)(\beta\mu + \beta - \mu). \end{aligned}$$

Then  $\beta^2(K - 1/\beta^2)K_{\text{den}} = a_3a_4$ , where we know  $K_{\text{den}} > 0$  from part I. Now verifying  $K < \frac{1}{\beta^2}$  is equivalent to verifying  $a_3a_4 < 0$ .

Given  $0 < \mu$ ,  $0 < \beta$ , and  $0 < \theta < 2$ , we have

$$(\mu, \beta, \theta) \in R_{(a)} \Leftrightarrow \text{(a1) and (a2)}$$

and

$$(\mu, \beta, \theta) \in R_{(b)} \Leftrightarrow \text{(b1) and (b2)},$$

where (a1), (a2), (b1), and (b2) are defined as in Section SM4.1.1. We show

$$(\mu, \beta, \theta) \in R_{(e)} \Rightarrow (\mu, \beta, \theta) \notin R_{(a)} \text{ and } (\mu, \beta, \theta) \notin R_{(b)} \text{ and } (\mu, \beta, \theta) \notin R_{(c)} \Rightarrow a_3a_4 < 0$$

by considering the following 3 cases:

- (i)  $\neg(\text{a1})$  and  $\neg(\text{b1})$  and  $(\mu, \beta, \theta) \notin R_{(c)} \Rightarrow a_3a_4 < 0$
  - (ii)  $\neg(\text{a1})$  and (b1) and  $\neg(\text{b2})$  and  $(\mu, \beta, \theta) \notin R_{(c)} \Rightarrow a_3a_4 < 0$
  - (iii) (a1) and  $\neg(\text{a2})$  and  $(\mu, \beta, \theta) \notin R_{(c)} \Rightarrow a_3a_4 < 0$
- (i) & (ii) Since  $a_4$  is an affine function of  $\theta$  and  $\theta \in \left(0, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ , either

$$a_4 < \max\left\{a_4|_{\theta=0}, a_4|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}}\right\} = \max\left\{-2(\beta+1)(\beta\mu+\beta-\mu), -\frac{4\beta^2\mu}{2\beta\mu+\beta+\mu}\right\} \leq 0,$$

where the first term is nonpositive by the assumption  $\beta\mu + \beta - \mu \geq 0$ , or

$$a_4 = -2(\beta+1)(\beta\mu+\beta-\mu) = -\frac{4\beta^2\mu}{2\beta\mu+\beta+\mu} < 0.$$

In both cases,  $a_4 < 0$ .

(i) Since  $a_3$  is an affine function of  $\theta$  and  $\theta \in \left(0, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ , either

$$a_3 > \min \left\{ a_3|_{\theta=0}, a_3|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} \right\} = \min \left\{ 2(\beta+1)(\mu+1)(\beta+\mu-\beta\mu), \frac{4\beta\mu(\beta+\mu)}{2\beta\mu+\beta+\mu} \right\} \geq 0,$$

where the first term is nonnegative by  $\neg(\text{b2})$ , which is  $\beta\mu - \beta - \mu \leq 0$ ,  
or

$$a_3 = 2(\beta+1)(\mu+1)(\beta+\mu-\beta\mu) = \frac{4\beta\mu(\beta+\mu)}{2\beta\mu+\beta+\mu} > 0.$$

In both cases, we have  $a_3 > 0$ .

(ii) As we had shown in Section SM4.1.1 case (b) part (ii), we have

$$(\text{b1}) \quad \Rightarrow \quad -2\beta^2\mu^2 + \beta^2\mu + \beta^2 + \beta\mu^2 + \beta + \mu^2 + \mu < 0,$$

and we have no division by 0 in considering  $\neg(\text{b2})$ .

Since  $a_3$  is an affine function of  $\theta$  and  $\theta \in \left(\frac{2(\beta^2+\beta\mu+\beta+\mu^2+\mu-\mu^2\beta^2)}{-2\beta^2\mu^2+\beta^2\mu+\beta^2+\beta\mu^2+\beta+\mu^2+\mu}, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ ,  
either

$$a_3 > \min \left\{ a_3|_{\theta=\frac{2(\beta^2+\beta\mu+\beta+\mu^2+\mu-\mu^2\beta^2)}{-2\beta^2\mu^2+\beta^2\mu+\beta^2+\beta\mu^2+\beta+\mu^2+\mu}}, a_3|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} \right\} = \min \left\{ 0, \frac{4\beta\mu(\beta+\mu)}{2\beta\mu+\beta+\mu} \right\} = 0,$$

or

$$a_3 = 0 = \frac{4\beta\mu(\beta+\mu)}{2\beta\mu+\beta+\mu}.$$

The latter case is impossible, and we conclude  $a_3 > 0$ .

(iii) As we had shown in Section SM4.1.1 case (a) part (ii), we have

$$(\text{a1}) \quad \Rightarrow \quad \mu + \beta\mu - \beta - \beta^2 - 2\beta^2\mu > 0,$$

and we have no division by 0 in considering  $\neg(\text{a2})$ .

Since  $a_4$  is an affine function of  $\theta$  and  $\theta \in \left(2\frac{(\beta+1)(\mu-\beta\mu-\beta)}{\mu+\beta\mu-\beta-\beta^2-2\beta^2\mu}, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ ,  
either

$$a_4 < \max \left\{ a_4|_{\theta=2\frac{(\beta+1)(\mu-\beta\mu-\beta)}{\mu+\beta\mu-\beta-\beta^2-2\beta^2\mu}}, a_4|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} \right\} = \max \left\{ 0, -\frac{4\beta^2\mu}{2\beta\mu+\beta+\mu} \right\} = 0.$$

or

$$a_4 = 0 = -\frac{4\beta^2\mu}{2\beta\mu+\beta+\mu}$$

The latter case is impossible, and we conclude  $a_4 < 0$ .

Since  $a_3$  is an affine function of  $\theta$  and  $\theta \in \left(2\frac{(\beta+1)(\mu-\beta\mu-\beta)}{\mu+\beta\mu-\beta-\beta^2-2\beta^2\mu}, 2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}\right)$ ,  
either

$$a_3 > \min \left\{ a_3|_{\theta=2\frac{(\beta+1)(\mu-\beta\mu-\beta)}{\mu+\beta\mu-\beta-\beta^2-2\beta^2\mu}}, a_3|_{\theta=2\frac{\beta\mu+\beta+\mu}{2\beta\mu+\beta+\mu}} \right\} = \min \left\{ \frac{4\beta(1-\beta^2)\mu^2}{\mu+\beta\mu-\beta-\beta^2-2\beta^2\mu}, \frac{4\beta\mu(\beta+\mu)}{2\beta\mu+\beta+\mu} \right\} > 0,$$

where the first term is positive since  $1 > \beta$  follows from (a1), or

$$a_3 = \frac{4\beta(1-\beta^2)\mu^2}{\mu+\beta\mu-\beta-\beta^2-2\beta^2\mu} = \frac{4\beta\mu(\beta+\mu)}{2\beta\mu+\beta+\mu} > 0.$$

In both cases, we have  $a_3 > 0$ .



**SM4.2. Inequalities for Theorem 4.3.**

**SM4.2.1. Upper bounds.** It remains to show that there is no division by 0 and  $\lambda_\mu^A, \lambda_L^B, K_1, K_2 \geq 0$  in each case.

*Case (a).* Since the quantities in Section SM3.2.1 Case (a) are simple, we verify by inspection that there is no division by 0 and  $\lambda_\mu^A, \lambda_L^B, K_1, K_2 \geq 0$ .

*Case (b).* Assume  $0 < \mu, 0 < L, 0 < \theta < 2$ , and

$$(SM4.2) \quad L < 1$$

$$(SM4.3) \quad \mu > \frac{L^2+1}{(L-1)^2}$$

$$(SM4.4) \quad \theta \leq 2(\mu+1)(L+1) \frac{\mu+\mu L^2-L^2-2\mu L-1}{2\mu^2-\mu+\mu L^3-L^3-3\mu L^2-L^2-2\mu^2 L-\mu L-L-1}$$

(no division by 0 implied by (ii) below)

- (i) From (SM4.2) and (SM4.3), it is direct to obtain  $\mu > 1$ .
- (ii) The numerator of (SM4.4) is positive since, by (SM4.3), we have

$$\mu + \mu L^2 - L^2 - 2\mu L - 1 = \mu(1-L)^2 - (L^2+1) > 0.$$

Since  $\theta > 0$ , the denominator of (SM4.4) is nonnegative. To prove strict positivity, we view the denominator of (SM4.4) as a quadratic function of  $\mu$ :

$$\begin{aligned} \phi_{L,\theta}(\mu) &= 2\mu^2 - \mu + \mu L^3 - L^3 - 3\mu L^2 - L^2 - 2\mu^2 L - \mu L - L - 1 \\ &= \underbrace{2(1-L)}_{>0 \text{ by (SM4.2)}} \mu^2 + (L^3 - 3L^2 - L - 1)\mu - (1+L)(1+L^2). \end{aligned}$$

This quadratic is nonpositive only between its roots and

$$\phi_{L,\theta}(0) = -(1+L)(1+L^2) < 0, \quad \phi_{L,\theta}\left(\frac{L^2+1}{(L-1)^2}\right) = 2L \frac{(L^2+1)^2}{(1-L)^3} > 0.$$

Therefore  $\phi_{L,\theta}(\mu) > 0$  for all  $\mu > \frac{L^2+1}{(L-1)^2}$ , which holds by (SM4.3). Therefore we conclude the denominator of (SM4.4) is strictly positive.

- (iii) Multiply both sides of (SM4.4) by  $(\mu+L)/(\mu+1)(L+1)$  and reorganize to get

$$\begin{aligned} 1 - \theta \frac{\mu+L}{(\mu+1)(L+1)} &\geq 1 - 2(\mu+L) \frac{\mu+\mu L^2-L^2-2\mu L-1}{2\mu^2-\mu+\mu L^3-L^3-3\mu L^2-L^2-2\mu^2 L-\mu L-L-1} \\ &= \frac{(\mu-1)(1-L)(1+2\mu L+L^2)}{2\mu^2-\mu+\mu L^3-L^3-3\mu L^2-L^2-2\mu^2 L-\mu L-L-1} \\ &> 0, \end{aligned}$$

where the latter inequality follows from (SM4.2), (i), and (ii).

- (iv) Multiply both sides of (SM4.4) by  $(2\mu+L+1)$  and reorganize to get

$$2(\mu+1)(L+1) - \theta(2\mu+L+1) \geq \frac{4\mu^2(\mu+1)L(1-L^2)}{2\mu^2-\mu+\mu L^3-L^3-3\mu L^2-L^2-2\mu^2 L-\mu L-L-1} > 0,$$

where the latter inequality follows from (SM4.2) and (ii).

- (v) Multiply both sides of (SM4.4) by the denominator of (SM4.4) (which is positive by (ii)) and reorganize to get

$$2(\mu+1)(L+1) (\mu(1-L)^2 - (L^2+1)) + \theta (\mu(1+L+3L^2-L^3) + (1+L+L^2+L^3) + 2\mu^2(L-1)) \geq 0.$$

(vi) Multiply both sides of (SM4.4) by  $L + \mu$  and reorganize to get

$$(L + 1)(\mu + 1) - \theta(L + \mu) \geq \frac{(\mu^2 - 1)(1 - L^2)(1 + L^2 + 2\mu L)}{2\mu^2 - \mu + \mu L^3 - L^3 - 3\mu L^2 - L^2 - 2\mu^2 L - \mu L - L - 1}.$$

Since  $\mu > 1$  by (i),  $L < 1$  by (SM4.2), and the denominator is positive by (ii), we have

$$(L + 1)(\mu + 1) - \theta(L + \mu) > 0.$$

(iv) shows  $\lambda_\mu^A$  is nonnegative. (i) and (iv) show  $\lambda_L^B$  is nonnegative. (v) shows the numerator of  $K_1$  is nonnegative. (v) and (vi) show the numerator of  $K_2$  is nonnegative. (v) shows the denominator of  $K_2$  is positive. (v) shows the denominators of  $m_2$  and  $m_3$  are positive.

*Case (c).* Assume  $(\mu, L, \theta) \in R_{(c)}$ . Then:

(i) We show  $L < 1$ . Since  $(\mu, L, \theta) \notin R_{(a)}$ , we have

$$0 < \sqrt{L^2 + 1} < \mu \frac{-(2(\theta - 1)\mu + \theta - 2) + L^2(\theta - 2(1 + \mu))}{\sqrt{(2(\theta - 1)\mu + \theta - 2)^2 + L^2(\theta - 2(1 + \mu))^2}}.$$

This implies the numerator of the right-most-side is positive:

$$(SM4.5) \quad 0 < -(2(\theta - 1)\mu + \theta - 2) + L^2 \underbrace{(\theta - 2(1 + \mu))}_{<0 \text{ by } \mu > 0, \theta < 2}.$$

Therefore

$$L^2 < \frac{2(\theta - 1)\mu + \theta - 2}{\theta - 2(1 + \mu)} = 1 + \frac{2\theta\mu}{\underbrace{\theta - 2(1 + \mu)}_{<0 \text{ by } \mu > 0, \theta < 2}} < 1,$$

and we conclude  $L < 1$ .

- (ii) We have  $(2 - \theta)(1 - L^2) - 2\mu(\theta + L^2 - 1) > 0$ , since it is a simple reformulation of (SM4.5).  
 (iii) We have  $\theta(L^2 + 1) - 2\mu(\theta + L^2 - 1) > 0$ , because either  $\theta + L^2 - 1 < 0$  and

$$\theta(L^2 + 1) - 2\mu(\theta + L^2 - 1) > 0$$

or  $\theta + L^2 - 1 \geq 0$  and

$$\theta(L^2 + 1) - 2\mu(\theta + L^2 - 1) > \theta(L^2 + 1) - (2 - \theta)(1 - L^2) = 2(\theta + L^2 - 1) \geq 0,$$

where the first inequality follows from (ii).

- (iv) We prove  $\theta(1 + 2\mu + L^2) - 2(\mu + 1)(L^2 + 1) < 0$ . Note that (SM4.5), the main inequality from (i), can be reformulated as

$$\begin{aligned} 0 &< (2 - \theta)(1 - L^2) - 2\mu(\theta + L^2 - 1) \\ &= \theta \underbrace{(-2\mu + L^2 - 1)}_{<0 \text{ by (i)}} - 2(\mu + 1)(L^2 - 1). \end{aligned}$$

Therefore,  $\theta < \frac{2(\mu + 1)(1 - L^2)}{2\mu - L^2 + 1}$  and plugging this into  $\theta$  gives us

$$\theta(1 + 2\mu + L^2) - 2(\mu + 1)(L^2 + 1) < -\frac{8\mu(\mu + 1)L^2}{2\mu + 1 - L^2} < 0.$$

(v) Finally, we show  $\mu > 1$ . (We use this later in Section SM4.2.2.) Since  $(\mu, L, \theta) \notin R_{(a)}$  we have

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

which implies

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

which, in turn, implies

$$\begin{aligned} \sqrt{(2(\theta-1)\mu + (\theta-2))^2} &< \sqrt{(2(\theta-1)\mu + (\theta-2))^2 + L^2(\theta - 2(\mu+1))^2} \\ &< \mu(-2(\theta-1)\mu + \theta - 2 + L^2(\theta - 2(1+\mu))) < \mu(-2(\theta-1)\mu + \theta - 2), \end{aligned}$$

where we used  $\theta < 2 < 2(1+\mu)$  for the last inequality. Finally, using the implied  $\mu(-2(\theta-1)\mu + \theta - 2) > 0$ , we get

$$0 < (-2(\theta-1)\mu + \theta - 2) < \mu(-2(\theta-1)\mu + \theta - 2) \Rightarrow 1 < \mu.$$

(iii) and (iv) shows the numerator of  $\rho^2$  is nonpositive. (ii) shows the denominator of  $\rho^2$  is negative. (iii) shows that  $\lambda_\mu^A$  is nonnegative. (iii) shows that the numerator of  $\lambda_L^B$  is nonnegative. (ii) shows that the denominator of  $\lambda_L^B$  is positive. (ii) shows that the denominator of  $K_1$  is positive.

**SM4.2.2. Lower bounds.** It remains to show that  $(\mu, L, \theta) \in R_{(c)} \Rightarrow 0 \leq K \leq 1$  Note that

$$(\mu, L, \theta) \in R_{(c)} \quad \Rightarrow \quad (\mu, L, \theta) \notin R_{(a)} \text{ and } (\mu, L, \theta) \notin R_{(b)}.$$

Write

$$K = \frac{L^2 + 1}{2L} \frac{K_{\text{num}}}{K_{\text{den}}}.$$

In part I, we show  $K_{\text{den}} > 0$  and  $K_{\text{num}} \geq 0$  using  $(\mu, L, \theta) \notin R_{(a)}$ . In part I, we show  $K \leq 1$  using  $(\mu, L, \theta) \notin R_{(a)}$  and  $(\mu, L, \theta) \notin R_{(b)}$ .

*Part I.* We now show  $K_{\text{den}} > 0$  and  $K_{\text{num}} \geq 0$ .

First, we quickly recall  $(\mu, L, \theta) \notin R_{(a)} \Rightarrow \mu > 1$  and  $(\mu, L, \theta) \notin R_{(a)} \Rightarrow L < 1$  by respectively (v) and (i) of case (c) of Section SM4.2.1.

Next we move on to the main proof. Note

$$\begin{aligned} &(\mu, L, \theta) \notin R_{(a)} \\ \Rightarrow &\frac{\mu(-2(\theta-1)\mu - \theta + L^2(\theta - 2(\mu+1)) + 2)}{\sqrt{(2(\theta-1)\mu + \theta - 2)^2 + L^2(\theta - 2(\mu+1))^2}} > \sqrt{L^2 + 1} \\ \Rightarrow &\mu(-2(\theta-1)\mu - \theta + L^2(\theta - 2(\mu+1)) + 2) > \sqrt{L^2 + 1} \sqrt{(2(\theta-1)\mu + \theta - 2)^2 + L^2(\theta - 2(\mu+1))^2} \\ \Rightarrow &(\mu(-2(\theta-1)\mu - \theta + L^2(\theta - 2(\mu+1)) + 2))^2 > (L^2 + 1)((2(\theta-1)\mu + \theta - 2)^2 + L^2(\theta - 2(\mu+1))^2). \end{aligned}$$

Reorganizing the last inequality gives us

$$(\mu + 1)K_{\text{num}} > 0,$$

which proves  $K_{\text{num}} > 0$ .

Let us view  $K_{\text{den}}$  as a quadratic function of  $\theta$ . From (ii) of case (c) of Section SM4.2.1, we have  $(2-\theta)(1-L^2)-2\mu(\theta+L^2-1) > 0$ , which we can reformulate as  $\theta < \frac{2(\mu+1)(1-L^2)}{1-L^2+2\mu}$  using to  $L < 1$ . Basic computation gives us

$$\begin{aligned} K_{\text{den}} \Big|_{\theta=0} &= 4(\mu+1) \left( \mu^2 (L^2-1)^2 + (L^2+1)^2 \right) > 0, \\ K_{\text{den}} \Big|_{\theta=\frac{2(\mu+1)(1-L^2)}{1-L^2+2\mu}} &= \frac{16\mu^2(\mu+1)L^2(L^2+1)^2}{(1-L^2+2\mu)^2} > 0, \\ \frac{dK_{\text{den}}}{d\theta} \Big|_{\theta=\frac{2(\mu+1)(1-L^2)}{1-L^2+2\mu}} &= -\frac{4\mu(L^2+1)^2(2\mu+L^2+1)}{1-L^2+2\mu} < 0. \end{aligned}$$

If the quadratic term of  $K_{\text{den}}$  is negative,  $K_{\text{den}}$  is positive when  $\theta$  is between the roots, and the interval  $\theta \in \left(0, \frac{2(\mu+1)(1-L^2)}{1-L^2+2\mu}\right)$  lies between the roots. If the quadratic term

of  $K_{\text{den}}$  is nonnegative, then the roots (if they exist) would lie in  $\left(\frac{2(\mu+1)(1-L^2)}{1-L^2+2\mu}, \infty\right)$  because of the sign of the derivative. Therefore we conclude  $K_{\text{den}} > 0$  in both cases.

*Part II.* We now show  $K = \frac{L^2+1}{2L} \frac{K_{\text{num}}}{K_{\text{den}}} \leq 1$ . Since  $K_{\text{den}} > 0$ , we can equivalently show

$$(SM4.6) \quad (L^2+1)K_{\text{num}} - 2LK_{\text{den}} = a_1a_2 \leq 0$$

with

$$\begin{aligned} a_1 &= 2\mu(1-L)(\theta+L^2-1) - (2-\theta)(L+1)(L^2+1) \\ a_2 &= 2\mu^2(1-L)(\theta+L^2-1) + (2-\theta)(L+1)(L^2+1) + \theta\mu((L-1)^3 - 4L) + 4\mu L(L+1). \end{aligned}$$

Remember that in Section SM4.2.1 case (c) part (ii) we had shown  $2\mu(\theta+L^2-1) < (2-\theta)(1-L^2)$ . Plugging this inequality into the first term of  $a_1$  gives us

$$a_1 \leq (2-\theta)(1-L)(1-L^2) - (2-\theta)(L+1)(L^2+1) = -2(2-\theta)L(L+1) < 0.$$

Finally, we show  $a_2 \geq 0$ . Remember that

$$(\mu, L, \theta) \in R_{(b)} \quad \Leftrightarrow \quad (SM4.2) \text{ and } (SM4.3) \text{ and } (SM4.4)$$

so we divide  $(\mu, L, \theta) \notin R_{(b)}$  into the following three cases:

- (i) Case  $\neg(SM4.2)$ . This case corresponds to  $L \geq 1$ , but this cannot happen as  $(\mu, L, \theta) \notin R_{(a)}$  implies  $L < 1$ .
- (ii) Case  $(SM4.2)$  and  $\neg(SM4.3)$ . This case corresponds to  $\mu \leq \frac{L^2+1}{(L-1)^2}$ . Plug  $L^2+1 \geq \mu(L-1)^2$  into the second term of  $a_2$  to get

$$\begin{aligned} a_2 &\geq 2\mu(\mu(1-L)(\theta+L^2-1) + (L^2+1)(1+L-\theta)) \\ &\geq \mu(1-L)(\theta+L^2-1) + (L^2+1)(1+L-\theta), \end{aligned}$$

which we further split in two cases: either  $1+L-\theta \geq 0$  and we use  $L^2+1 \geq \mu(L-1)^2$  again to get

$$a_2 \geq \theta\mu L(1-L) > 0,$$

or  $1 + L - \theta < 0$  and we use  $0 < \theta - L - 1 \leq \theta + L^2 - 1$  and hence from the previous inequality on  $a_2$  and  $\mu > 1$ :

$$\begin{aligned} a_2 &\geq (1 - L)(\theta + L^2 - 1) + (L^2 + 1)(1 + L - \theta) \\ &= (2 - \theta)L(L + 1) \\ &> 0, \end{aligned}$$

where the last inequality follows from  $\theta < 2$ .

- (iii) Case (SM4.2) and (SM4.3) and  $\neg$ (SM4.4). Remember that in Section SM4.2.1 case (b) part (ii), we have shown that (SM4.2) and (SM4.3) implies

$$2\mu^2 - \mu + \mu L^3 - L^3 - 3\mu L^2 - L^2 - 2\mu^2 L - \mu L - L - 1 > 0.$$

So there is no division by 0 in  $\neg$ (SM4.4) and we have

$$\theta(2\mu^2 - \mu + \mu L^3 - L^3 - 3\mu L^2 - L^2 - 2\mu^2 L - \mu L - L - 1) > 2(\mu + 1)(L + 1)(\mu + \mu L^2 - L^2 - 2\mu L - 1).$$

We conclude by noting

$$\begin{aligned} a_2 &= \theta(2\mu^2 - \mu + \mu L^3 - L^3 - 3\mu L^2 - L^2 - 2\mu^2 L - \mu L - L - 1) - 2(\mu + 1)(L + 1)(\mu + \mu L^2 - L^2 - 2\mu L - 1) \\ &\geq 0. \end{aligned}$$

**SM5. Visualization.** In this section, we visualize the cases for Theorems 4.1 and 4.3 and the contraction factors for Corollaries 4.2 and 4.4.

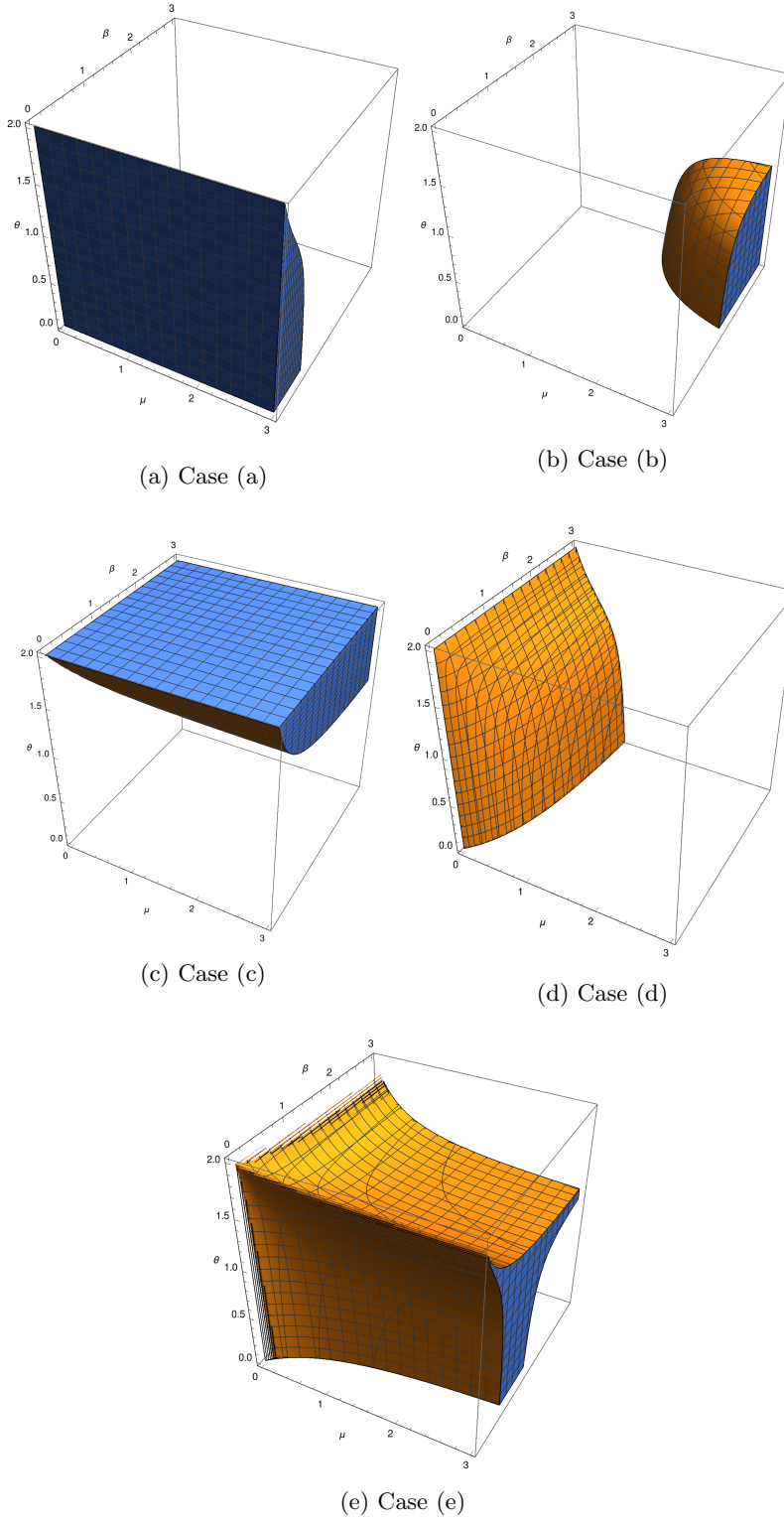


Fig. 3: Parameter regions for the 5 cases of Theorem 4.1 in the  $\mu$ - $\beta$ - $\theta$  plane.

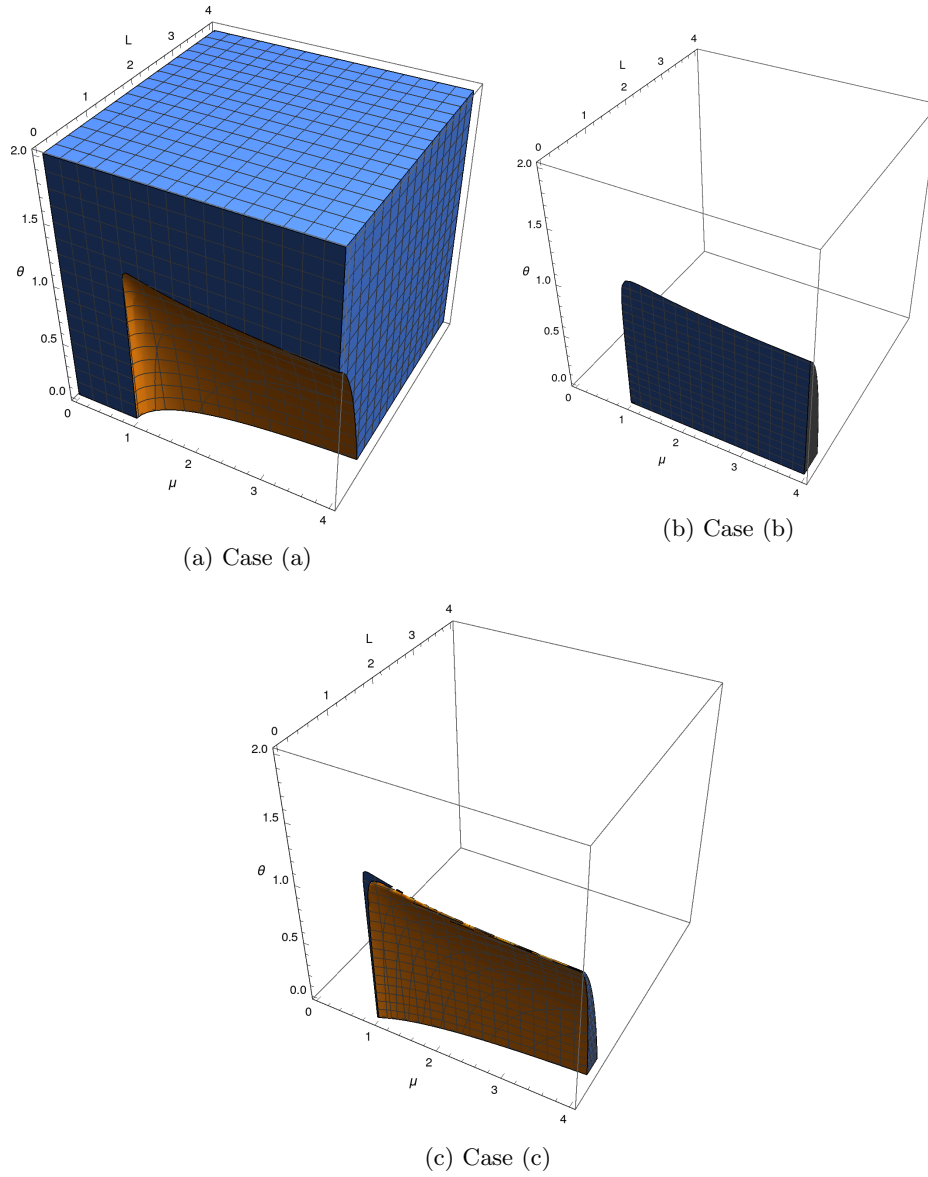


Fig. 4: Parameter regions for the 3 cases of Theorem 4.3 in the  $\mu$ - $L$ - $\theta$  plane.

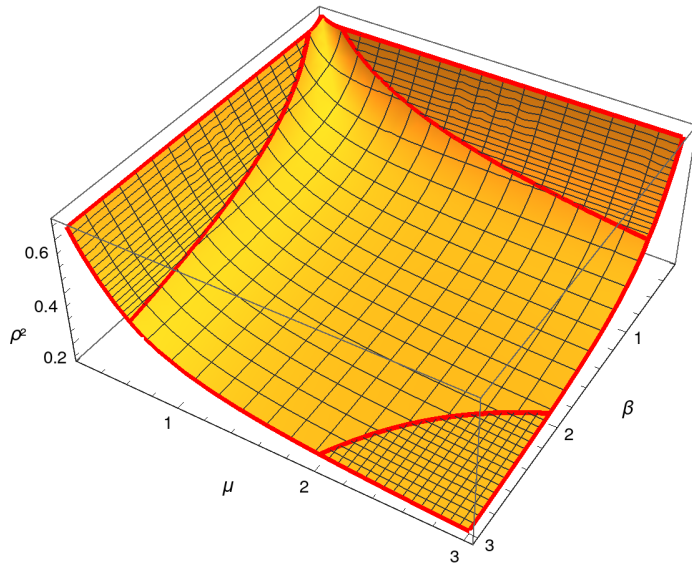


Fig. 5: Contraction factor for Corollary 4.2 (when  $\theta = 1$ ) in the  $\mu$ - $\beta$  plane.

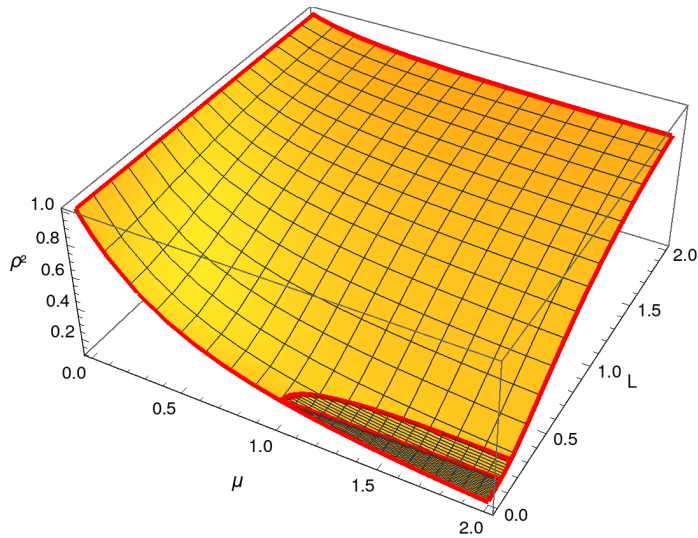


Fig. 6: Contraction factor for Corollary 4.4 (when  $\theta = 1$ ) in the  $\mu$ - $L$  plane.