

# Tight Global Linear Convergence Rate Bounds for Operator Splitting Methods

Goran Banjac and Paul J. Goulart

**Abstract**—In this paper we establish necessary and sufficient conditions for global linear convergence rate bounds in operator splitting methods for a general class of convex optimization problems where the associated fixed-point operator is strongly quasicontractive. We also provide a tight bound on the achievable convergence rate. Most existing results establishing global linear convergence in such methods require restrictive assumptions regarding strong convexity and smoothness of the constituent functions in the optimization problem. However, there are several examples in the literature showing that linear convergence is possible even when these properties do not hold. We provide a unifying analysis method for establishing global linear convergence based on linear regularity and show that many existing results are special cases of our approach. Moreover, we propose a novel linearly convergent splitting method for linear programming.

## I. INTRODUCTION

We consider convex optimization problems in the form

$$\text{minimize } f(x) + g(x) \quad (1)$$

where  $x \in \mathcal{H}$  is the optimization variable,  $\mathcal{H}$  is a real Hilbert space, and  $f$  and  $g$  are convex functions. Such problems are often referred to as *composite convex* optimization problems, and arise in such disparate areas as signal processing [2], [3], optimal control [4], [5] and machine learning [6]. There exists a variety of methods for solving such problems, known collectively as *operator splitting methods*, which encompasses techniques such as the proximal gradient method (PGM), Peaceman-Rachford splitting (PRS), Douglas-Rachford splitting (DRS), and the alternating direction method of multipliers (ADMM) [7], [8].

Over the last decade convergence rate analysis of operator splitting methods has attracted significant interest in the control community. This is mainly motivated by the need to provide a computation complexity certificate for optimal control algorithms [9]. The authors in [4] established linear convergence of the accelerated PGM for solving a restricted class of input-constrained model predictive control (MPC) problems provided that the constraint set is simple enough. Linear convergence of the same algorithm with an adaptive restart technique was shown in [10], [11]. The ADMM for solving quadratic programs was shown to converge linearly

provided that the constraint matrix has full row-rank [12]. The authors in [13] extended this result to a more general class of problems having a strongly convex and smooth objective function, while the tightness of the obtained convergence rate was shown in [14]. The authors in [15] established linear convergence of a distributed version of the ADMM under similar assumptions.

Many practical problems can be formulated in the form (1) where  $f$  is a smooth function and  $g$  is nonsmooth. Examples include constrained optimization of a smooth function  $f$  when  $g$  is the indicator function of a closed convex set, or regularized optimization when  $g$  is a regularization term. In proximal methods a nonsmooth function is tackled through its proximal operator, whose evaluation requires the solution of a small ancillary convex optimization problem. These subproblems often admit closed-form solutions [3], which makes proximal methods effective in practical applications. Even when the proximal operator of  $g$  is not easy to evaluate, the original problem can sometimes be reformulated such that the new problem has more favorable structure. Such a reformulation is often done by forming and solving the dual of the original problem [16].

First-order methods can sometimes converge very slowly, and in particular converge sublinearly in many problems of practical interest. Many of the results on linear convergence of first-order methods require strong convexity and smoothness of a function involved in the optimization problem, e.g. if the problem can be written in the form (1) with  $f$  strongly convex and smooth and  $g$  as a function whose proximal operator is easy to evaluate [17]. Real-world problems rarely have the aforementioned structure, so current analysis techniques can not ensure linear convergence of first-order optimization methods. This situation has motivated many researchers to look for alternative algorithms and analytical methods that could ensure linear convergence. The authors in [18]–[20] address local or asymptotic linear convergence under a variety of assumptions. Linear convergence of a particular proximal method for linear programming (LP) is shown in [21]. The work in [22] establishes linear convergence of feasible descent methods by proving a global error bound for a particular class of non-strongly convex problems, and the authors in [23], [24] show that the same argument can be used to prove linear convergence of some methods for solving optimization problems of the type arising in MPC. Related, but a more general property than the error bound [25], is a quadratic growth property that was studied in [25]–[28], and used to establish linear convergence of various first-order methods.

Linear convergence of some projection methods, which are a

The authors are with the Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK. Email: {goran.banjac,paul.goulart}@eng.ox.ac.uk

This work was supported by the European Commission research project FP7-PEOPLE-2013-ITN under grant agreement no. 607957 [Training in Embedded Optimization and Predictive Control (TEMPO)]. Part of this material has been presented at the 55<sup>th</sup> IEEE Conference on Decision and Control [1].

special class of proximal methods for solving convex feasibility and best-approximation problems, has also been established in certain special cases. In the case of two affine subspaces, the linear convergence rate of the alternating projection method (APM) and the DRS method is characterized in terms of the Friedrichs angle between the subspaces [29], [30]. The authors in [31] show that the linear convergence rate of the APM applied to two convex polyhedra is characterized by the smallest nonzero Friedrichs angle between faces of the polyhedra. The authors in [32] have identified linear regularity as a sufficient condition for global linear convergence of the fixed-point iteration of an averaged nonexpansive operator and provided a linear convergence rate. In our earlier work in [1] we showed that linear regularity is also a necessary condition for global linear convergence and provided a better rate of convergence compared to [32].

### Contributions of this work

In this paper we establish necessary and sufficient conditions for global linear convergence rate bounds of the fixed-point iteration of a strongly quasicontractive operator in Hilbert spaces and provide a tight convergence rate bound. Parts of this work were presented in preliminary form in [1], which we here extend in several ways: 1) The results herein apply to strongly quasicontractive operators, a wider class of operators (than e.g. contractive or averaged nonexpansive operators) that arise in algorithms for convex optimization; 2) The results herein are shown to hold for general real Hilbert spaces; 3) We derive a linear convergence rate of the generalized Douglas-Rachford splitting for subspaces; 4) We prove tightness of the global linear convergence rate bound established in [1]; and 5) We demonstrate that a variety of existing results establishing linear convergence are special cases of our analysis.

### Structure of the paper

The paper is organized as follows. In Section II we show how the convergence of a given optimization algorithm is related to the properties of a particular operator associated to the algorithm, and we provide a global linear convergence rate bound. The relationship between the error bound and the quadratic growth property of the function and the linear regularity of the operator related to a particular method is discussed in Section III. In Section IV we establish linear convergence of some projection methods using our framework and show tightness of the convergence rate bound established in Section II. A particular linearly convergent method for LP is derived in Section V, while Section VI concludes the paper.

### Notation

Let  $\mathbb{N}$  denote the set of nonnegative integers,  $\mathbb{R}$  the set of real numbers,  $\mathbb{R}_+$  the set of nonnegative real numbers,  $\tilde{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$  the extended real line,  $\mathbb{R}^n$  the  $n$ -dimensional real space, and  $\mathcal{H}$  a real Hilbert space equipped with inner product  $\langle \cdot, \cdot \rangle$ , the induced norm  $\| \cdot \|$  and the identity operator  $\text{Id}$ . The class of closed, proper and convex functions  $f : \mathcal{H} \rightarrow \tilde{\mathbb{R}}$  is denoted by  $\Gamma_0(\mathcal{H})$ . For  $f \in \Gamma_0(\mathcal{H})$ , we

denote by  $\partial f(x)$  its subdifferential at  $x$ . If  $f$  is differentiable, then  $\nabla f(x)$  denotes its gradient at  $x$ . For  $x \in \mathbb{R}^n$  and  $y \in \mathbb{R}^n$ ,  $(x, y) := [x^T y^T]^T$  denotes their vertical concatenation. The set of fixed-points of an operator  $T : \mathcal{H} \rightarrow \mathcal{H}$  is defined as  $\text{Fix } T := \{x \in \mathcal{H} \mid x = Tx\}$ . The distance of a vector  $x$  to a nonempty closed convex set  $\mathcal{C} \subseteq \mathcal{H}$  is  $\text{dist}_{\mathcal{C}}(x) := \min \{\|x - y\| \mid y \in \mathcal{C}\}$ , and the distance between two nonempty sets,  $\mathcal{C}_1$  and  $\mathcal{C}_2$ , is  $\text{dist}(\mathcal{C}_1, \mathcal{C}_2) := \inf \{\|x - y\| \mid x \in \mathcal{C}_1, y \in \mathcal{C}_2\}$ . Projection of a vector  $x$  onto  $\mathcal{C}$  is  $\mathbb{P}_{\mathcal{C}}(x) := \{y \in \mathcal{C} \mid \|y - x\| = \text{dist}_{\mathcal{C}}(x)\}$ . The affine hull of a set  $\mathcal{C}$  is denoted by  $\text{aff}(\mathcal{C})$ , and the subspace obtainable via translation of  $\text{aff}(\mathcal{C})$  is denoted by  $\text{aff}_0(\mathcal{C})$ , i.e.  $\text{aff}_0(\mathcal{C}) := \text{aff}(\mathcal{C}) - c$  for some  $c \in \text{aff}(\mathcal{C})$ . We denote the relative interior of  $\mathcal{C}$  by  $\text{relint } \mathcal{C}$ . A face  $P_x$  of a polyhedron  $P$  is defined as a nonempty minimizer of a linear function over  $P$ , i.e.  $P_x := \text{argmin}_{p \in P} \langle x, p \rangle$ . The set of all faces of  $P$  is denoted by  $\mathcal{F}_P$ .

## II. LINEAR CONVERGENCE VIA LINEAR REGULARITY

We will analyze the convergence properties of general first-order iterative optimization algorithms for problems in the form (1). Let  $\mathcal{D}$  be a nonempty subset of  $\mathcal{H}$ . We define an operator  $T : \mathcal{D} \rightarrow \mathcal{D}$  such that the iterations computed by the algorithm from some initial point  $x^0 \in \mathcal{D}$  are equivalent to the iteration

$$x^{k+1} = Tx^k, \quad \forall k \in \mathbb{N}, \quad (2)$$

for all  $x^k \in \mathcal{D}$ . The iteration (2) is referred to as a *Picard* or *fixed-point iteration* of operator  $T$ . We make the following assumption throughout:

*Assumption 1:* The fixed-point set of  $T$  is nonempty.

Under mild assumptions on operator  $T$ , it can be shown that a bounded sequence  $\{x^k\}$  generated by (2) converges weakly to a point in  $\text{Fix } T$  [33, Thm. 5.13]. The set of optimizers  $\mathcal{X}^*$  of the optimization problem at hand is usually closely related to the fixed-point set of the related operator  $T$ . It is often the case that either  $\mathcal{X}^* \equiv \text{Fix } T$  (e.g. when  $T$  represents the PGM [8]), or it is easy to reconstruct a point  $x^* \in \mathcal{X}^*$  from a point  $\bar{x} \in \text{Fix } T$  (e.g. when  $T$  represents the DRS [8]). If this is the case, then solving an optimization problem can be reformulated as a problem of finding a fixed-point of a suitably selected operator.

We are particularly interested in showing that the operator  $T$  satisfies the condition

$$\text{dist}_{\text{Fix } T}(Tx) \leq \beta \text{dist}_{\text{Fix } T}(x) \quad (3)$$

for all  $x \in \mathcal{D}$ , and for some  $\beta \in [0, 1)$  independent of the choice of  $x$ . Note that this is different from local (asymptotic) linear convergence that requires only that inequality (3) holds when  $x$  is close enough to  $\text{Fix } T$ . The constant  $\beta$  is called the *convergence factor* and determines the linear convergence rate.

In the sequel we will introduce some definitions and results from operator theory that we require in order to analyze the convergence of a fixed-point iteration; we refer the reader to [33] for a comprehensive review.

*Definition 1 ([33], [34]):* Let  $\mathcal{D}$  be a nonempty subset of  $\mathcal{H}$  and let  $T : \mathcal{D} \rightarrow \mathcal{H}$ . Then  $T$  is

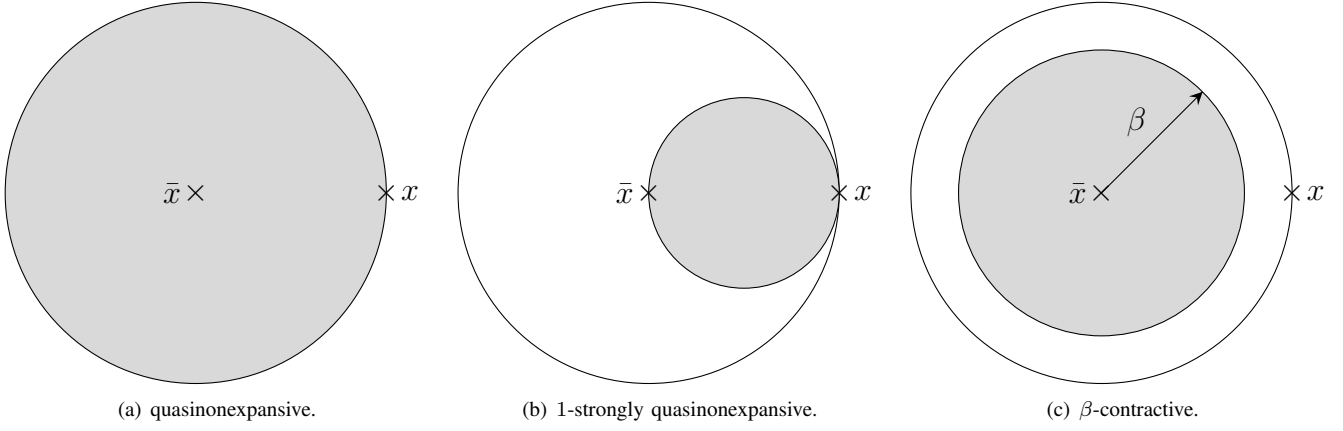


Fig. 1. Graphical representation of some operator properties. The point/set  $Tx$  lies somewhere inside the gray-shaded area. We assume that  $x$  is on the unit circle with the center at  $\bar{x} \in \text{Fix } T$ .

- (i) *nonexpansive* if  $(\forall x \in \mathcal{D})(\forall y \in \mathcal{D})$

$$\|Tx - Ty\| \leq \|x - y\|, \quad (4a)$$

- (ii)  $\alpha$ -*averaged* with  $\alpha \in (0, 1)$  if there exists a nonexpansive operator  $R : \mathcal{D} \rightarrow \mathcal{H}$  such that

$$T = (1 - \alpha)\text{Id} + \alpha R,$$

- (iii) *quasinonexpansive (QNE)* if  $(\forall x \in \mathcal{D})(\forall y \in \text{Fix } T)$

$$\|Tx - y\| \leq \|x - y\|, \quad (4b)$$

- (iv)  $\rho$ -*strongly quasinonexpansive ( $\rho$ -SQNE)* with  $\rho > 0$  if  $(\forall x \in \mathcal{D})(\forall y \in \text{Fix } T)$

$$\|Tx - y\|^2 \leq \|x - y\|^2 - \rho \|x - Tx\|^2, \quad (4c)$$

- (v)  $\beta$ -*contractive* with  $\beta \in [0, 1)$  if  $(\forall x \in \mathcal{D})(\forall y \in \mathcal{D})$

$$\|Tx - Ty\| \leq \beta \|x - y\|. \quad (4d)$$

It can be shown that  $T$  is  $\alpha$ -averaged if and only if it satisfies  $(\forall x \in \mathcal{D})(\forall y \in \mathcal{D})$

$$\|Tx - Ty\|^2 + \frac{1 - \alpha}{\alpha} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2 \quad (5)$$

(see [33, Prop. 4.25]); an operator satisfying (5) for  $\alpha = \frac{1}{2}$  is called *firmly nonexpansive*. The same proof can be used to show that  $T$  is  $\rho$ -SQNE if and only if there exists a QNE operator  $R : \mathcal{D} \rightarrow \mathcal{H}$  such that

$$T = \left(1 - \frac{1}{1 + \rho}\right) \text{Id} + \frac{1}{1 + \rho} R. \quad (6)$$

Note that every nonexpansive operator is also QNE, and every  $\alpha$ -averaged operator is  $\frac{1-\alpha}{\alpha}$ -SQNE. In particular, all the operators in Definition 1 are QNE for which the following fact applies:

*Fact 1 ([34]):* If  $T : \mathcal{H} \rightarrow \mathcal{H}$  is QNE, then  $\text{Fix } T$  is closed and convex.

Operators arising from the iteration of a particular optimization method often enjoy at least some subset of the properties described in Definition 1. Figure 1 illustrates these properties and highlights that the distance of the iterates of a fixed-point iteration to any  $\bar{x} \in \text{Fix } T$  is nonincreasing if any of the

properties above holds. In the case of a  $\rho$ -SQNE operator that satisfies the so called demi-closedness principle (for more details see [34]), the sequence  $\{x^k\}$  converges weakly to a fixed-point [35, Cor. 2]. Moreover, in the case of a contractive operator convergence is linear and the fixed-point iterates satisfy the inequality (3). Without additional assumptions the same can not be claimed for a SQNE operator. Note that if  $T$  is used as a fixed-point operator, it needs to map  $\mathcal{D}$  to  $\mathcal{D}$  as noted just before (2).

In the sequel we will identify the essential additional property of an SQNE operator to ensure that a sequence generated by the related fixed-point iteration satisfies (3).

*Remark 1:* If (3) holds and  $\text{Fix } T$  is a singleton, then  $\{x^k\}$  converges linearly with convergence factor  $\beta$ . If  $\text{Fix } T$  is not a singleton, condition (3) itself does not imply that the sequence is linearly convergent. However, if  $T$  is QNE then the concept of Fejér monotonicity can be used to show that (3) implies linear convergence of  $\{x^k\}$  [33, Thm. 5.12]. Therefore, instead of claiming that (3) holds, we will sometimes state that  $\{x^k\}$  admits global linear convergence because the underlying operator is usually QNE.

#### A. Linear regularity

The linear convergence results we present will largely exploit the concept of *linear regularity* of an operator.

*Definition 2 ([32]):* Let  $T : \mathcal{D} \rightarrow \mathcal{H}$  and suppose that Assumption 1 holds. We say that  $T$  is *linearly regular* with constant  $\kappa \geq 0$  if  $(\forall x \in \mathcal{D})$

$$\text{dist}_{\text{Fix } T}(x) \leq \kappa \|x - Tx\|. \quad (7)$$

If an operator is linearly regular, it means that the distance between successive iterates  $(x^k, x^{k+1})$  of the related fixed-point iteration is lower-bounded by  $\frac{1}{\kappa} \text{dist}_{\text{Fix } T}(x^k)$ . Figure 2 illustrates that if the linear regularity constant of an SQNE operator  $T$  is bounded, i.e.  $\kappa < \infty$ , then  $Tx$  will lie inside the circle with radius  $\beta < 1$ . This observation leads to our main result:

*Theorem 1:* Let  $T : \mathcal{D} \rightarrow \mathcal{H}$  be a  $\rho$ -SQNE operator and suppose that Assumption 1 holds. Then the inequality

$$\text{dist}_{\text{Fix } T}(Tx) \leq \beta \text{dist}_{\text{Fix } T}(x)$$

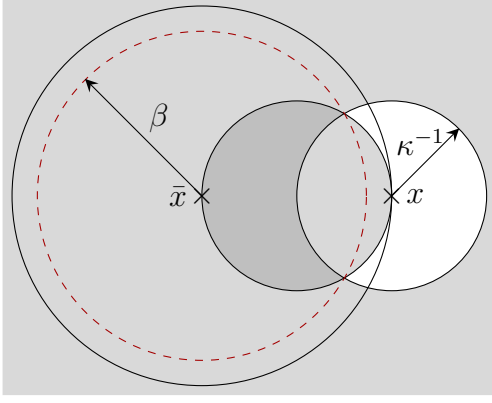


Fig. 2. Linear convergence of a SQNE and linearly regular operator holds if the linear regularity constant  $\kappa$  is bounded from above. In that case  $Tx$  lies inside the circle with a radius  $\beta < 1$ . We assume that  $x$  is on the unit circle with the center at  $\bar{x} \in \text{Fix } T$ .

holds for some constant  $\beta \in [0, 1)$  and  $(\forall x \in \mathcal{D})$  if and only if  $T$  is linearly regular. If the linear regularity constant of such an operator  $T$  is  $\kappa$ , then

$$\beta = \sqrt{1 - \frac{\rho}{\kappa^2}} \in [0, 1). \quad (8)$$

*Proof:* See Appendix A. ■

Theorem 1 states that linear regularity, or a lack thereof, is the essential determinant of global linear convergence (in the sense of (3)) for a sequence generated by fixed-point iteration of an SQNE operator. We provide an example in Section IV of an operator for which our rate bound in (8) is tight.

Previous results in [32] have established linear regularity as a sufficient condition for (3) when  $T$  is  $\alpha$ -averaged, albeit with the weaker convergence rate bound

$$\beta = \sqrt{\frac{\rho^{-1}\kappa^2}{1 + \rho^{-1}\kappa^2}} \quad (9)$$

where  $\rho = \frac{1-\alpha}{\alpha}$ . It is easy to show that our bound in Theorem 1 is strictly better than (9). A related result appears in [21], which established linear convergence of a particular first-order method for linear programming based on the weaker condition

$$\text{dist}_{\text{Fix } T}(Tx) \leq \tilde{\kappa} \|x - Tx\|, \quad (10)$$

for some  $\rho$ -SQNE operator  $T$  and some  $\tilde{\kappa} > 0$ , i.e. by lower bounding  $\|x - Tx\|$  via  $\text{dist}_{\text{Fix } T}(Tx)$  rather than  $\text{dist}_{\text{Fix } T}(x)$ . In this case one can show that condition (3) still holds with rate  $\beta$  as in (9) but with  $\kappa$  replaced by  $\tilde{\kappa}$ . By virtue of Theorem 1 any such operator must be linearly regular, so that (10) is also both necessary and sufficient for linear convergence of a SQNE operator.

*Proposition 1:* Let  $T : \mathcal{D} \rightarrow \mathcal{H}$  be a  $\rho$ -SQNE operator. Then its linear regularity constant  $\kappa$  satisfies

$$\kappa \geq \frac{1 + \rho}{2}.$$

*Proof:* See Appendix B. ■

### B. Improving the convergence factor

It follows from (6) that it is possible to obtain an SQNE operator from QNE via a suitable transformation. It turns out that if  $T_1$  is  $\rho_1$ -SQNE, we can construct a  $\rho_2$ -SQNE operator  $T_2$  via an additional similar transformation, assuming that  $\rho_1 > 0$  and  $\rho_2 > 0$ . We next show how to select  $\rho_2$  in order to obtain the smallest convergence factor of the resulting fixed-point iteration:

*Proposition 2:* Let  $\rho_1 > 0$  and  $\rho_2 > 0$ . Suppose  $T_1 : \mathcal{D} \rightarrow \mathcal{H}$  is  $\rho_1$ -SQNE, with a linear regularity constant  $\kappa_1 > 0$ , and let  $T_2 := \left(1 - \frac{1+\rho_1}{1+\rho_2}\right) \text{Id} + \frac{1+\rho_1}{1+\rho_2} T_1$ . Then the following hold

- (i)  $T_2$  is  $\rho_2$ -SQNE with linear regularity constant  $\kappa_2 = \frac{1+\rho_2}{1+\rho_1} \kappa_1$ , and this estimate is tight.
- (ii) The smallest convergence factor is attained for  $\rho_2 = 1$ .

*Proof:* See Appendix C. ■

In other words, for any SQNE operator  $T$  achieving linear convergence, the worst-case convergence factor can be minimized by a simple transformation to make it 1-SQNE.

We will next show that a variety of existing results establishing linear convergence for different problem classes can be seen as special cases of the approach introduced in Theorem 1.

### III. LINEAR REGULARITY COMING FROM THE ERROR BOUND AND THE QUADRATIC GROWTH PROPERTY

We first introduce some properties of a convex function that will be used in the sequel.

*Definition 3 ([13]):* Let  $f \in \Gamma_0(\mathcal{H})$ . Then  $f$  is

- (i)  $L$ -smooth if it is differentiable and  $\forall (x, y) \in \mathcal{H}$

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \|x - y\|^2,$$

- (ii)  $\mu$ -strongly convex if  $\forall (x, y) \in \mathcal{H}$  and  $\forall u \in \partial f(y)$

$$f(x) \geq f(y) + \langle u, x - y \rangle + \frac{\mu}{2} \|x - y\|^2.$$

Consider the following convex optimization problem

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && x \in \mathcal{C} \end{aligned} \quad (11)$$

where  $f(x) = h(Ex) + \langle b, x \rangle$ ,  $h$  is  $\mu$ -strongly convex and  $L$ -smooth, and  $\mathcal{C}$  is a nonempty closed polyhedral set whose projection operator can be evaluated efficiently. Note that this problem is a special case of problem (1) since it can be reformulated as

$$\text{minimize} \quad f(x) + \mathcal{I}_{\mathcal{C}}(x)$$

where  $\mathcal{I}_{\mathcal{C}}$  is the indicator function of  $\mathcal{C}$  defined as

$$\mathcal{I}_{\mathcal{C}}(x) = \begin{cases} 0 & x \in \mathcal{C} \\ \infty & \text{otherwise} \end{cases}$$

with  $\mathcal{I}_{\mathcal{C}} \in \Gamma_0(\mathcal{H})$  [33].

The problem (11) arises in many practical applications such as the dual forms of the MPC problem [16] and the L1-loss linear support vector machine (SVM) [22]. Smoothness of  $h$  implies smoothness of  $f$ . If  $E$  has full column rank, then  $f$  is strongly convex [13] and there consequently exists some linearly convergent proximal method for solving (11).

However, in most real-world problems  $E$  does *not* have full column rank, which results in a function  $f$  that is convex but not strongly so.

The authors in [18] were among the first to establish local linear convergence of a first-order method for solving a non-strongly convex problem related to (11). Their analysis is based on a so-called local error bound, with a global error bound established only recently [22]. This global error bound was then used to establish global linear convergence of feasible descent methods for solving (11). This important result is used in [23], [24] to establish linear convergence of some methods for solving optimization problems of the type arising in MPC. The authors in [25] showed that the so-called quadratic growth property is more general than the error bound, and can be also used for establishing linear convergence of some first-order methods. The relationship between error bound and quadratic growth properties is discussed in [25]–[28].

However, the definition of linear convergence in the aforementioned papers does not follow the usual one, but instead refers to linear convergence of the suboptimality gap  $\{f(x^k) - f^*\}$ , where  $f^*$  is the optimal value of problem (11), rather than linear convergence of the sequence  $\{x^k\}$ . We will show that both the global error bound and the quadratic growth property imply linear regularity of fixed-point operators related to the same class of methods, which allows us to prove linear convergence of the sequence  $\{x^k\}$  when the underlying operator is SQNE.

#### A. Error bound

**Definition 4 ([22]):** An optimization problem of the form (11) admits a *global error bound* with constant  $\tau \geq 0$  if  $(\forall x \in \mathcal{C})$

$$\text{dist}_{\mathcal{X}^*}(x) \leq \tau \|\nabla^+ f(x)\|, \quad (12)$$

where  $\mathcal{X}^*$  is the nonempty set of optimizers of the given optimization problem and  $\nabla^+ f(x) := x - \mathbb{P}_{\mathcal{C}}(x - \nabla f(x))$ .

**Definition 5 ([22]):** A sequence  $\{x^k\}$  is said to be generated by a *feasible descent method* if for every  $k$  it satisfies

$$x^{k+1} = \mathbb{P}_{\mathcal{C}}(x^k - \omega^k \nabla f(x^k) + e^k), \quad (13a)$$

$$\|e^k\| \leq \beta \|x^k - x^{k+1}\|, \quad (13b)$$

$$f(x^k) - f(x^{k+1}) \geq \gamma \|x^k - x^{k+1}\|^2, \quad (13c)$$

where  $\omega^k > 0$ ,  $\beta \geq 0$ , and  $\gamma > 0$ .

The inequality in (13c) is often referred to as a *sufficient-decrease condition* [36]. Conditions (12) and (13a)–(13b) imply the inequality

$$\text{dist}_{\mathcal{X}^*}(x^k) \leq \tau \frac{1 + \beta}{\underline{\omega}} \|x^k - x^{k+1}\|,$$

where  $\underline{\omega} := \min(1, \inf_k \omega^k)$  (see eq. (46) in [22]). Since  $\mathcal{X}^*$  is the fixed-point set of feasible descent methods, this inequality means that fixed-point operators associated to these methods enjoy the essential linear regularity property identified in Theorem 1 as long as  $\underline{\omega} > 0$ .

#### B. Quadratic growth property

**Definition 6 ([25]):** An optimization problem of the form (11) admits a *quadratic growth property* with constant  $a > 0$  if  $(\forall x \in \mathcal{C})$

$$\frac{a}{2} \text{dist}_{\mathcal{X}^*}^2(x) \leq f(x) - f^*. \quad (14)$$

**Definition 7 ([27]):** A sequence  $\{x^k\}$  is said to be generated by a *subgradient descent method* if for every  $k$  it satisfies (13c) and there exists  $w^{k+1} \in \nabla f(x^{k+1}) + N_{\mathcal{C}}(x^{k+1})$  such that

$$\|w^{k+1}\| \leq b \|x^k - x^{k+1}\|, \quad (15)$$

where  $N_{\mathcal{C}}(x)$  denotes the normal cone of  $\mathcal{C}$  at  $x$ , and  $b > 0$ . The inequality in (15) is referred to as a *relative error condition*. It is easy to show that conditions (14) and (15) imply linear regularity of the related operator. If we denote by  $\phi = f + \mathcal{I}_{\mathcal{C}}$ , then the right-hand side of (14) can be written as  $\phi(x) - \phi^*$  where  $\phi^* = f^*$ . The quadratic growth property and convexity of  $\phi$ , together with the relative error condition (15) imply that for  $x^{k+1} \notin \mathcal{X}^*$  and  $\bar{x}^{k+1} := \mathbb{P}_{\mathcal{X}^*}(x^{k+1})$

$$\begin{aligned} \frac{a}{2} \text{dist}_{\mathcal{X}^*}^2(x^{k+1}) &\leq \phi(x^{k+1}) - \phi^* \\ &\leq \langle w^{k+1}, x^{k+1} - \bar{x}^{k+1} \rangle \\ &\leq \|w^{k+1}\| \|x^{k+1} - \bar{x}^{k+1}\| \\ &= \|w^{k+1}\| \text{dist}_{\mathcal{X}^*}(x^{k+1}) \\ &\leq b \|x^k - x^{k+1}\| \text{dist}_{\mathcal{X}^*}(x^{k+1}). \end{aligned}$$

Note that since  $\phi$  is closed, so is  $\mathcal{X}^*$  and thus the projection onto  $\mathcal{X}^*$  is well defined. Dividing both sides of the above inequality by  $\frac{a}{2} \text{dist}_{\mathcal{X}^*}(x^{k+1}) > 0$ , we obtain

$$\text{dist}_{\mathcal{X}^*}(x^{k+1}) \leq \tilde{\kappa} \|x^k - x^{k+1}\|,$$

where  $\tilde{\kappa} = \frac{2b}{a}$ . Using Lemma 4 (in Appendix A) it can be shown that the last inequality implies that the fixed-point operator associated to a subgradient descent method is linearly regular with constant  $\kappa = \tilde{\kappa} + 1$ .

#### C. Linear convergence

In order to establish linear convergence of the sequence  $\{x^k\}$  satisfying conditions from either of the previous two subsections, it then only remains to show that the operator related to the method is SQNE. For instance, if  $\beta$  in (13b) is zero then the method reduces to the projected gradient method. We next show that the projected gradient method with a constant step size  $\omega$  is averaged, and thus SQNE.

**Lemma 1:** Let  $T$  be the fixed-point operator associated with the *projected gradient method* for solving an optimization problem of the form (11),

$$T = \mathbb{P}_{\mathcal{C}} \circ (\text{Id} - \omega \nabla f),$$

where  $f$  is  $L$ -smooth, and  $\omega \in (0, \frac{2}{L})$  is the step size. Then  $T$  is  $\frac{1}{2 - \omega L/2}$ -averaged.

**Proof:** By [33, Prop. 4.8], operator  $\mathbb{P}_{\mathcal{C}}$  is firmly non-expansive, and by [8],  $(\text{Id} - \omega \nabla f)$  is  $\frac{\omega L}{2}$ -averaged for  $\omega \in (0, \frac{2}{L})$ . By [37, Prop. 2.4], the composition of  $\alpha_1$ -averaged and  $\alpha_2$ -averaged operator is averaged with parameter  $\alpha =$

$\frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2}$ . The statement follows directly by setting  $\alpha_1 = 1/2$  and  $\alpha_2 = \omega L/2$  in the expression for  $\alpha$ . ■

Note that it is not necessary that  $\beta = 0$  for a feasible descent method to be averaged. For instance, the *proximal minimization algorithm* is a method from this class since it can be represented in the form (13) with  $\beta = \omega L$ , where  $\omega$  is a parameter of the proximal operator [18, Prop. 3.3]. It is also firmly nonexpansive since its iteration is defined via the proximal operator of a closed convex function [33, Prop. 12.27].

*Remark 2:* In Lemma 1 we showed that the fixed-point operator associated with PGM with a constant step size is averaged. It is easy to show that PGM with time-varying step size is also averaged if  $\bar{\omega} < 2/L$ , where  $\bar{\omega} := \sup_k \omega^k$ . However, the fixed-point operator associated with this method would then be time-varying, which is a more general setup than iteration (2).

*Remark 3:* The authors in [22] and [27] do not assume that the fixed-point operators associated to the feasible descent and subgradient descent methods are averaged, but rather make use of a sufficient decrease condition (13c). Note that (13c) looks very similar to (31) (which holds for every SQNE operator), with the difference that the progress of the method towards optimality is measured via suboptimality of the iterates, rather than via the distance of the iterates to the set of optimizers.

#### IV. LINEAR REGULARITY IN METHODS FOR SOLVING FEASIBILITY PROBLEMS

Let  $A$  and  $B$  be two nonempty closed convex subsets of  $\mathcal{H}$  such that  $A \cap B \neq \emptyset$ . The *convex feasibility problem* is to find a point in the set  $A \cap B$ , i.e.

$$\begin{aligned} \text{find} \quad & x \\ \text{subject to} \quad & x \in A \\ & x \in B. \end{aligned}$$

Note that this problem is a special class of problem (1) since it can be reformulated as

$$\text{minimize} \quad \mathcal{I}_A(x) + \mathcal{I}_B(x) \quad (16)$$

where  $\mathcal{I}_A$  and  $\mathcal{I}_B$  are the indicator functions of sets  $A$  and  $B$ , respectively.

Projection methods are a special class of proximal methods for solving such problems and date to von Neumann's *alternating projection method* [38], given by the following iteration

$$a^{k+1} = \mathbb{P}_A(b^k), \quad (17a)$$

$$b^{k+1} = \mathbb{P}_B(a^{k+1}). \quad (17b)$$

It is generally assumed that projection operators for both sets,  $\mathbb{P}_A$  and  $\mathbb{P}_B$ , can be evaluated efficiently. *Douglas-Rachford splitting* is a related proximal method that has received increasing attention due to its generally observed good practical performance. In the case of the feasibility problem (16), DRS is defined as a fixed-point iteration of the Douglas-Rachford operator

$$T_{DR} := \mathbb{P}_B(2\mathbb{P}_A - \text{Id}) + \text{Id} - \mathbb{P}_A, \quad (18)$$

or more explicitly

$$x^{k+1} = \mathbb{P}_A(z^k), \quad (19a)$$

$$y^{k+1} = \mathbb{P}_B(2x^{k+1} - z^k), \quad (19b)$$

$$z^{k+1} = z^k + y^{k+1} - x^{k+1}. \quad (19c)$$

In the case when  $A$  and  $B$  are closed subspaces and  $A + B$  is closed, both methods converge linearly and the convergence rate is characterized in terms of the *Friedrichs angle* between the subspaces [29], [30]. The Friedrichs angle is a generalization of the angle between subspaces in higher dimensions:

*Definition 8 ([30]):* Suppose that  $U$  and  $V$  are closed subspaces of  $\mathcal{H}$ . *Cosine of the Friedrichs angle* between  $U$  and  $V$  is

$$\begin{aligned} c_F(U, V) := \sup \{ \langle u, v \rangle \mid u \in U \cap (U \cap V)^\perp, \|u\| \leq 1, \\ v \in V \cap (U \cap V)^\perp, \|v\| \leq 1 \}. \end{aligned} \quad (20)$$

If  $U + V$  is closed, which is true for any finite dimensional space  $\mathcal{H}$ , then  $c_F(U, V) < 1$  [30, Fact 2.3(i)].

We will next show that the linear convergence in such cases can be recovered as a special case of Theorem 1 by establishing that the underlying fixed-point operators are both linearly regular and SQNE.

#### A. Alternating projection method

The authors in [31] show that the APM for two convex polyhedra in  $\mathbb{R}^n$ ,  $A$  and  $B$ , converges linearly and that the convergence rate is characterized via the Friedrichs angle between the faces of the two polyhedra. In the sequel we will briefly present this result in the case when  $A \cap B \neq \emptyset$ . Note that the original result is established without this assumption.

*Theorem 2 ([31]):* Let  $A$  and  $B$  be closed convex polyhedra in  $\mathbb{R}^n$ ,  $A \cap B \neq \emptyset$ , and  $b^0 \in B$ . Then the sequences  $\{a^k\}$  and  $\{b^k\}$  generated by the APM converge linearly towards some point in  $A \cap B$ , so that

$$\begin{aligned} \text{dist}_{A \cap B}(a^{k+1}) &\leq \beta \text{dist}_{A \cap B}(a^k), \\ \text{dist}_{A \cap B}(b^{k+1}) &\leq \beta \text{dist}_{A \cap B}(b^k), \end{aligned}$$

where the convergence factor is given by

$$\beta = \max_{\substack{A_x \in \mathcal{F}_A \\ B_y \in \mathcal{F}_B}} c_F^2(\text{aff}_0(A_x), \text{aff}_0(B_y)) < 1. \quad (21)$$

We will show that this result can be seen as a special case of Theorem 1. We can represent iteration (17) via a double iteration of the following fixed-point operator defined on  $\mathcal{D} = A \cup B$

$$Tx = \begin{cases} \mathbb{P}_A(x) & x \in B \\ \mathbb{P}_B(x) & x \in A \end{cases}. \quad (22)$$

Since  $\mathbb{P}_A$  and  $\mathbb{P}_B$  are firmly nonexpansive operators [33, Prop. 4.8], for every  $x \in \mathcal{D}$  and  $y \in A \cap B$  we have

$$\|Tx - y\|^2 \leq \|x - y\|^2 - \|x - Tx\|^2,$$

which means that  $T$  is 1-SQNE with  $\text{Fix } T = A \cap B$ . Also, it can be shown that this operator is linearly regular with constant

$$\kappa = \left( 1 - \max_{\substack{A_x \in \mathcal{F}_A \\ B_y \in \mathcal{F}_B}} c_F^2(\text{aff}_0(A_x), \text{aff}_0(B_y)) \right)^{-1/2}$$

which holds for all  $x \in \mathcal{D}$  (see [31, Prop. 4 and Cor. 5]). According to Theorem 1 these two properties ensure global linear convergence with convergence factor  $\sqrt{\beta}$ , where  $\beta$  is given in (21). Double iteration of a fixed-point operator given in (22), which is equivalent to iteration (17), consequently has convergence factor  $\beta$ . Therefore, convergence factor given in Theorem 2 can be seen as a special case of Theorem 1. Observe that, although linear regularity does not hold for the whole space  $\mathbb{R}^n$ , by restricting  $b^0 \in B$ , linear convergence of the generated sequence is ensured since then the sequences  $\{a^k\}$  and  $\{b^k\}$  are in  $\mathcal{D}$ . Note that the linear convergence of the APM for two subspaces is a special case of the above result.

### B. Douglas-Rachford splitting

The authors in [30] show that the linear convergence rate of the DRS for two subspaces can also be characterized in terms of the Friedrichs angle.

*Proposition 3 ([30]):* Suppose that  $A$  and  $B$  are closed subspaces of  $\mathcal{H}$  such that  $A + B$  is closed and let  $z \in \mathcal{H}$ . Then as  $n \rightarrow \infty$

$$\begin{aligned} T_{DR}^n z &\rightarrow \mathbb{P}_{\text{Fix } T_{DR}}(z), \\ \mathbb{P}_A(T_{DR}^n z) &\rightarrow \mathbb{P}_{A \cap B}(z), \\ \mathbb{P}_B(T_{DR}^n z) &\rightarrow \mathbb{P}_{A \cap B}(z). \end{aligned}$$

The convergence is linear with rate  $c_F(A, B) < 1$ .

We will show that the convergence rate in Proposition 3 is again a special case of Theorem 1 by quantifying the linear regularity constant for the Douglas-Rachford operator in the case of a feasibility problem involving two subspaces.

*Proposition 4:* Under the assumptions of Proposition 3 the linear regularity constant of  $T_{DR} : \mathcal{H} \rightarrow \mathcal{H}$  is

$$\kappa_{DR} = (1 - c_F^2(A, B))^{-1/2}.$$

*Proof:* See Appendix D. ■

*Corollary 1:* Under the assumptions of Proposition 3, for all  $z \in \mathcal{H}$

$$\text{dist}_{\text{Fix } T_{DR}}(T_{DR} z) \leq c_F(A, B) \text{dist}_{\text{Fix } T_{DR}}(z).$$

*Proof:* The Douglas-Rachford operator in (18) is  $\frac{1}{2}$ -averaged [13], and thus 1-SQNE. The result follows directly from Theorem 1 by setting  $\rho = 1$  and  $\kappa = \kappa_{DR}$  from Proposition 4. ■

We will now provide an example showing that global linear convergence (in the sense of (3)) of the DRS for two subspaces can not be extended to the case of two convex polyhedra as is the case for the APM because the associated operator is not linearly regular.

*Example 1:* Let  $A = \varepsilon > 0$ , and  $B = \mathbb{R}_+$ , and  $z^0 = t \gg \varepsilon$  the initial point of the DRS algorithm. The first iteration of the DRS is

$$\begin{aligned} x^1 &= \mathbb{P}_A(z^0) = \varepsilon, \\ y^1 &= \mathbb{P}_B(2x^1 - z^0) = 0, \\ z^1 &= z^0 + y^1 - x^1 = t - \varepsilon. \end{aligned}$$

Sequence  $\{z^k\}$  converges to  $\varepsilon \in \text{Fix } T_{DR}$ . Note that the convergence rate in the first iteration is

$$\frac{\text{dist}_{\text{Fix } T_{DR}}(z^1)}{\text{dist}_{\text{Fix } T_{DR}}(z^0)} = \frac{|t - 2\varepsilon|}{|t - \varepsilon|}.$$

By taking  $t$  to be arbitrarily large, the convergence rate becomes arbitrarily close to 1 and, as long as the  $z$ -iterates are far enough from  $\varepsilon$ , the sequence converges with a constant step size. Note however that the auxiliary sequence  $\{x^k\}$  converges to a fixed-point in one iteration.

The absence of a global linear convergence rate arises from the absence of linear regularity of the Douglas-Rachford operator, which is, according to Theorem 1, essential for a global linear convergence rate. According to (7), the linear regularity constant is an upper bound to the following ratio

$$\kappa \geq \frac{\text{dist}_{\text{Fix } T_{DR}}(z^0)}{\|z^0 - z^1\|} = \frac{|t - \varepsilon|}{|\varepsilon|},$$

and by taking  $t$  be arbitrarily large w.r.t.  $\varepsilon$ , it is clear that such an upper bound is not finite.

### C. Generalized Douglas-Rachford splitting

Generalized DRS is a relaxed version of DRS, and is defined via the following fixed-point operator

$$T_{GDR} := (1 - 2\alpha) \text{Id} + 2\alpha T_{DR}, \quad (23)$$

where  $\alpha \in (0, 1)$ . Since  $T_{DR}$  is  $\frac{1}{2}$ -averaged, according to [1, Prop. 1],  $T_{GDR}$  is  $\alpha$ -averaged and thus  $\frac{1-\alpha}{\alpha}$ -SQNE.

We next derive a linear convergence rate of the generalized DRS for two subspaces. We will use Theorem 1 to provide an upper bound on the convergence factor. We first derive a linear regularity constant of  $T_{GDR}$  in the following lemma.

*Lemma 2:* Suppose that  $A$  and  $B$  are closed subspaces of  $\mathcal{H}$  such that  $A + B$  is closed. Then the linear regularity constant of  $T_{GDR} : \mathcal{H} \rightarrow \mathcal{H}$  is

$$\kappa_{GDR} = \frac{1}{2\alpha} (1 - c_F^2(A, B))^{-1/2}.$$

*Proof:* The result follows directly from [1, Prop. 1(i)], Proposition 4 and the fact that  $T_{DR}$  is  $\frac{1}{2}$ -averaged. ■

*Corollary 2:* Under the assumptions of Lemma 2,

$$\sup_{z \notin \text{Fix } T_{GDR}} \frac{\text{dist}_{\text{Fix } T_{GDR}}(T_{GDR} z)}{\text{dist}_{\text{Fix } T_{GDR}}(z)} \leq \beta_{GDR}, \quad (24)$$

where

$$\beta_{GDR} := \sqrt{c_F^2(A, B) + (1 - 2\alpha)^2(1 - c_F^2(A, B))}. \quad (25)$$

*Proof:* The generalized Douglas-Rachford operator in (23) is  $\frac{1-\alpha}{\alpha}$ -SQNE. Choosing  $\rho = \frac{1-\alpha}{\alpha}$  and  $\kappa = \kappa_{GDR}$  from Lemma 2, Theorem 1 implies that for all  $z \in \mathcal{H}$

$$\text{dist}_{\text{Fix } T_{DR}}(T_{DR} z) \leq \beta_{GDR} \text{dist}_{\text{Fix } T_{DR}}(z),$$

from which (24) follows directly. ■

We now derive a lower bound on the linear convergence rate in the following lemma.

*Lemma 3:* Under the assumptions of Lemma 2,

$$\sup_{z \notin \text{Fix } T_{GDR}} \frac{\text{dist}_{\text{Fix } T_{GDR}}(T_{GDR} z)}{\text{dist}_{\text{Fix } T_{GDR}}(z)} \geq \beta_{GDR},$$

where  $\beta_{GDR}$  is given in (25).

*Proof:* See Appendix E. ■

We can now state the following theorem that substantially improves our result from Theorem 1. Note that an upper bound is said to be *tight* or *the least upper bound* if no smaller value is an upper bound.

**Theorem 3:** The convergence rate bound in Theorem 1 for a  $\rho$ -SQNE and  $\kappa$ -linearly regular operator is tight for all admissible values of  $\rho$  and  $\kappa$ .

*Proof:* Under the assumptions of Lemma 2, we have the following equality

$$\sup_{z \notin \text{Fix } T_{GDR}} \frac{\text{dist}_{\text{Fix } T_{GDR}}(T_{GDR}z)}{\text{dist}_{\text{Fix } T_{GDR}}(z)} = \beta_{GDR},$$

which follows directly from Corollary 2 and Lemma 3. Since the upper bound in (24) is obtained from (8), this means that the convergence rate bound in Theorem 1 is tight.

Since  $A + B$  is closed, according to [30, Fact 2.3(i)],

$$0 \leq c_F(A, B) < 1,$$

and thus  $\kappa_{GDR}$  from Lemma 2 satisfies

$$\kappa_{GDR} \geq \frac{1}{2\alpha} = \frac{1+\rho}{2},$$

where  $\rho = (1 - \alpha)/\alpha$  is the constant of strong quasinonexpansiveness of operator  $T_{GDR}$ . Note that the range of values of  $\kappa_{GDR}$  covers all admissible values of linear regularity constant given in Proposition 1. By changing  $\alpha \in (0, 1)$  and the angle between subspaces, we can produce  $\rho$ -SQNE and  $\kappa$ -linearly regular operator  $T_{GDR}$  with arbitrary  $\rho > 0$  and  $\kappa \geq (1 + \rho)/2$ . ■

Note that the results in Corollary 2 and Lemma 3 provide linear convergence rate of the generalized DRS for subspaces, and thus extend those in [30] where the linear convergence rate is established for  $\alpha = 1/2$ . Observe from (25) that the smallest value of  $\beta_{GDR}$  is obtained for  $\alpha = 1/2$  which is consistent with the result in Proposition 2.

**Remark 4:** Tightness of the bound in Theorem 1 is proved by providing an example with an averaged operator  $T_{GDR}$  for which an upper bound on the worst-case convergence factor coincides with a lower bound. This means that the linear convergence rate bound for the class of  $\alpha$ -averaged operators given in [1, Thm. 1] is also tight.

## V. LINEAR PROGRAMMING

This section first appeared in the conference publication [1]. We include it here since it provides linear convergence guarantees for first-order methods in LPs, which is an open problem of direct interest to the control community.

Most existing results on linear convergence for optimization problems arising in model predictive control assume a strongly convex quadratic objective function and linear system dynamics, resulting in a quadratic programming problem. However, there exist well-known applications of predictive control for which a *linear* program arises, including problems based on  $\ell_1$ -norm minimization [39], [40] and robust min-max predictive control [41], [42].

Linear convergence of a particular first-order method for linear programming was shown in [21], with an assumption that the problem is feasible with a bounded objective value. In this section we will show that this result is a special case of Theorem 1. The underlying fixed-point operator can be shown to be SQNE (see Lemma 8 in [21]) and that inequality (10) holds (see the proof of Theorem 4 in [21]). As we note at the end of Section II-A this implies linear regularity of the operator. These two properties taken together imply linear convergence of a sequence generated by this method by virtue of Theorem 1.

In this section we propose a linearly convergent method for linear programming that does not assume feasibility of the problem or boundedness of the objective value. We first introduce a reformulation of the original problem that is used in the proposed method.

### A. HSD embedding

Consider the following primal-dual pair of the convex conic optimization problem

$$\begin{aligned} \min. \quad & c^T x & \max. \quad & -b^T y \\ \text{s. t.} \quad & Ax + s = b & \text{s. t.} \quad & -A^T y + r = c \\ & (x, s) \in \mathbb{R}^n \times \mathcal{K} & & (r, y) \in \{0\}^n \times \mathcal{K}^* \end{aligned} \quad (26)$$

where  $x \in \mathbb{R}^n$  and  $s \in \mathbb{R}^m$  (with  $n \leq m$ ) are the primal variables,  $r \in \mathbb{R}^n$  and  $y \in \mathbb{R}^m$  are the dual variables,  $\mathcal{K}$  is a nonempty closed convex cone and  $\mathcal{K}^*$  is its dual cone. The problem data are  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ , and  $c \in \mathbb{R}^n$ . In the case of a linear program  $\mathcal{K} = \mathcal{K}^* = \mathbb{R}_+^m$ .

The *homogeneous self-dual (HSD) embedding* has been widely used with interior-point methods. The authors in [43] proposed solving such an embedding with a first-order method instead. The HSD embedding is a formulation that encodes the primal-dual pair of optimization problems into the convex feasibility problem

$$\begin{aligned} \text{find} \quad & (u, v) \\ \text{subject to} \quad & v = Qu \\ & (u, v) \in \mathcal{C} \times \mathcal{C}^*, \end{aligned} \quad (27)$$

where  $u, v, Q, \mathcal{C}$  and  $\mathcal{C}^*$  are defined as

$$u := \begin{bmatrix} x \\ y \\ \tau \end{bmatrix}, \quad v := \begin{bmatrix} r \\ s \\ \kappa \end{bmatrix}, \quad Q := \begin{bmatrix} 0 & A^T & c \\ -A & 0 & b \\ -c^T & -b^T & 0 \end{bmatrix},$$

$$\mathcal{C} := \mathbb{R}^n \times \mathcal{K}^* \times \mathbb{R}_+, \quad \mathcal{C}^* := \{0\}^n \times \mathcal{K} \times \mathbb{R}_+.$$

The objective of the above optimization problem is to find a point  $(u, v)$  that satisfies both the subspace and the conic constraint in (27). Any solution of the self-dual embedding  $(x, y, \tau, r, s, \kappa)$  falls into one of the following three cases:

1. If  $\tau > 0$  and  $\kappa = 0$ , then  $(x/\tau, y/\tau, s/\tau)$  is a primal-dual solution of (26).
2. If  $\tau = 0$  and  $\kappa > 0$ , then either primal or dual problem is infeasible. The case  $b^T y < 0$  is a certificate for primal infeasibility, and the case  $c^T x < 0$  is a certificate for dual infeasibility.



3. If  $\tau = \kappa = 0$ , then nothing can be concluded about the solution of (26).

The problem (27) is referred to as *homogeneous* because the feasible set is a convex cone, hence any nonnegative scaling of a solution is also in the solution set. The authors in [43] showed that, by an appropriate selection of the initial point  $(u^0, v^0)$ , any convergent method whose associated fixed-point operator is nonexpansive will not converge to zero if a nonzero solution  $(u^*, v^*)$  exists. The appropriate initial point is any point satisfying  $(u^0, v^0)^T(u^*, v^*) > 0$ . Since  $(u^*, v^*)$  lies on the cone  $\mathcal{C} \times \mathcal{C}^*$ , it is sufficient for  $(u^0, v^0)$  to be contained in the (relative) interior of  $\mathcal{C}^* \times \mathcal{C}$ . In the case of a linear program we have

$$\begin{aligned} (u^0, v^0) &\in \text{relint}(\mathcal{C}^* \times \mathcal{C}) \\ &= \{0\}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++} \times \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}. \end{aligned}$$

### B. APM for solving linear programs in the HSD embedding form

The authors in [43] proposed solving conic optimization problems in HSD embedding form using ADMM. In this paper we propose solving a linear program in the HSD embedding form using the APM. Since in the case of LP the cone  $\mathcal{C} \times \mathcal{C}^*$  is polyhedral, we can apply Theorem 2 to show that the sequence of iterates generated by the method is linearly convergent.

*Corollary 3:* The sequence generated by the alternating projection method for solving linear program in the form (27) converges linearly.

The proposed method is as follows

$$(u_A^{k+1}, v_B^{k+1}) = \mathbb{P}_{Qu=v}(u_B^k, v_B^k), \quad (28a)$$

$$(u_B^{k+1}, v_B^{k+1}) = \mathbb{P}_{\mathcal{C} \times \mathcal{C}^*}(u_A^{k+1}, v_A^{k+1}). \quad (28b)$$

If the initial point  $(u_B^0, v_B^0)$  is selected as described in the previous subsection, then the sequence generated by the method will converge to a nonzero fixed-point. It should be noted that the projection (28a) requires solving a linear system involving a matrix  $(I + Q^T Q)$ . Step (28a) of the method can be computed as

$$\begin{aligned} u_A^{k+1} &= (I + Q^T Q)^{-1} (u_B^k - Qv_B^k), \\ v_A^{k+1} &= Qu_A^{k+1}. \end{aligned}$$

Since  $(I + Q^T Q)$  does not change throughout the iterations, it can be factored once and the factors are then used in cheaper back-solve operations in the following iterations [6]. Projection onto the cone in (28b) is trivial and separable componentwise. Therefore, all the operations except matrix factorization in the first iteration are basic arithmetic operations. Moreover, the method can be implemented very efficiently on computation platforms that enable a high degree of parallelism.

*Remark 5:* It is argued in [44, Sec. 3.7.3.1] and [45, App. F.1] that the linear convergence of the DRS applied to the HSD embedding of a linear program holds and that this result was shown in [32]. However, in [32, Thm. 4.4] linear convergence of the DRS was shown only for transversal sets (i.e. for sets  $A$  and  $B$  for which  $\text{relint } A \cap \text{relint } B \neq \emptyset$ ), which does

not hold for the HSD form of a linear program. Moreover, even for transversal sets, only local linear convergence can be established (see Example 1).

## VI. CONCLUSION

In this paper we provide necessary and sufficient conditions for global linear convergence of an optimization method whose associated fixed-point operator is strongly quasicontractive. We also provide a tight bound on the achievable convergence rate. The proposed framework is based on properties of such a fixed-point operator. We show that many published results on linear convergence can be viewed as special cases of the proposed framework. Also, we propose a novel linearly convergent method for linear programming which not only solves feasible problems, but also detects infeasibility and unboundedness of a solution.

## APPENDIX A PROOF OF THEOREM 1

We first provide a lemma that will be useful for proving the main theorem.

*Lemma 4:* Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two nonempty closed subsets of  $\mathcal{H}$ . Then for any  $z \in \mathcal{H}$ , the following *generalized triangle inequality* holds

$$\text{dist}(\mathcal{X}, \mathcal{Y}) \leq \text{dist}_{\mathcal{X}}(z) + \text{dist}_{\mathcal{Y}}(z). \quad (29)$$

*Proof:* Suppose that there exists some  $z_1 \in \mathcal{H}$  for which (29) does not hold. Then we have

$$\begin{aligned} \text{dist}(\mathcal{X}, \mathcal{Y}) &> \text{dist}_{\mathcal{X}}(z_1) + \text{dist}_{\mathcal{Y}}(z_1) \\ &= \|x_1 - z_1\| + \|y_1 - z_1\| \\ &\geq \|x_1 - y_1\| \end{aligned}$$

where  $x_1 \in \mathcal{X}$  and  $y_1 \in \mathcal{Y}$  are projections of  $z_1$  onto  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, and the last inequality follows from the triangle inequality. The above inequality means that the distance between  $x_1$  and  $y_1$  is strictly smaller than  $\text{dist}(\mathcal{X}, \mathcal{Y})$ , which is a contradiction. This concludes the proof. ■

The proof presented here is similar to that in our earlier work in [1] and is included here for the reader's convenience.

To show that linear regularity is a necessary property of an operator  $T$  for (3) to hold, we combine the generalized triangle inequality (29) ( $\mathcal{X} = \text{Fix } T$ ,  $\mathcal{Y} = \{x\}$ ,  $z = Tx$ )

$$\text{dist}_{\text{Fix } T}(x) \leq \text{dist}_{\text{Fix } T}(Tx) + \|x - Tx\|$$

with (3) to produce

$$(1 - \beta) \text{dist}_{\text{Fix } T}(x) \leq \text{dist}_{\text{Fix } T}(x) - \text{dist}_{\text{Fix } T}(Tx) \leq \|x - Tx\|,$$

which implies

$$\text{dist}_{\text{Fix } T}(x) \leq \frac{1}{1 - \beta} \|x - Tx\|.$$

This proves that  $T$  is linearly regular.

The authors in [32, Lem. 3.8] showed that linear regularity is a sufficient condition for (3) when  $T$  is  $\alpha$ -averaged. However, we derive this result with a better convergence factor, and thus we repeat some arguments from [32] for the sake of completeness.

Suppose that  $T$  is linearly regular with constant  $\kappa$ , that is

$$\frac{1}{\kappa^2} \text{dist}_{\text{Fix } T}^2(x) \leq \|x - Tx\|^2. \quad (30)$$

Note that since  $T$  is SQNE, according to Fact 1,  $\text{Fix } T$  is closed and convex and thus the projection  $\mathbb{P}_{\text{Fix } T}$  is well defined. By taking  $y = \mathbb{P}_{\text{Fix } T}(x)$ , (4c) implies

$$\begin{aligned} & \text{dist}_{\text{Fix } T}^2(Tx) + \rho \|x - Tx\|^2 \\ & \leq \|Tx - \mathbb{P}_{\text{Fix } T}(x)\|^2 + \rho \|x - Tx\|^2 \\ & \leq \text{dist}_{\text{Fix } T}^2(x), \end{aligned} \quad (31)$$

where the first inequality exploits properties of the distance between a point and a set, i.e.  $\text{dist}_{\text{Fix } T}(Tx) \leq \|Tx - \mathbb{P}_{\text{Fix } T}(x)\|$ . Combining (30) and (31) implies the following

$$\frac{1}{\kappa^2} \text{dist}_{\text{Fix } T}^2(x) \leq \frac{1}{\rho} (\text{dist}_{\text{Fix } T}^2(x) - \text{dist}_{\text{Fix } T}^2(Tx)), \quad (32)$$

hence

$$\text{dist}_{\text{Fix } T}(Tx) \leq \sqrt{1 - \frac{\rho}{\kappa^2}} \text{dist}_{\text{Fix } T}(x). \quad (33)$$

Note that (32) implies

$$\frac{1}{\kappa^2} \text{dist}_{\text{Fix } T}^2(x) \leq \frac{1}{\rho} \text{dist}_{\text{Fix } T}^2(x),$$

which means that  $\kappa^2 \geq \rho$ . Since  $\rho > 0$  and  $\kappa > 0$ , we have

$$0 \leq \sqrt{1 - \frac{\rho}{\kappa^2}} < 1,$$

so (33) satisfies condition (3).

## APPENDIX B PROOF OF PROPOSITION 1

Since  $T$  is  $\rho$ -SQNE, then according to (6)

$$R = (1 + \rho)T - \rho \text{Id}$$

is QNE, which implies that for every  $x \in \mathcal{D}$  and  $y = \mathbb{P}_{\text{Fix } T}(x)$

$$\begin{aligned} \|x - y\| & \geq \|Rx - y\| = \|(1 + \rho)Tx - \rho x - y\| \\ & = \|(1 + \rho)(Tx - x) + (x - y)\| \\ & \geq (1 + \rho)\|Tx - x\| - \|x - y\|, \end{aligned}$$

where the last line follows from the triangle inequality. Rearranging the terms gives

$$2\|x - y\| \geq (1 + \rho)\|Tx - x\| \geq \frac{1 + \rho}{\kappa} \|x - y\|,$$

which implies that  $\kappa \geq (1 + \rho)/2$ .

## APPENDIX C PROOF OF PROPOSITION 2

The proof presented here is similar to that in our earlier work in [1] and is included here for the reader's convenience.

*Proof of (i):*

Since  $T_1$  is  $\rho_1$ -SQNE, then according to (6)

$$R := (1 + \rho_1)T_1 - \rho_1 \text{Id}$$

is QNE, which itself implies that

$$\begin{aligned} T_2 & := \left(1 - \frac{1 + \rho_1}{1 + \rho_2}\right) \text{Id} + \frac{1 + \rho_1}{1 + \rho_2} T_1 \\ & = \left(1 - \frac{1}{1 + \rho_2}\right) \text{Id} + \frac{1}{1 + \rho_2} R \end{aligned} \quad (34)$$

is  $\rho_2$ -SQNE. Observe that  $T_1$  and  $T_2$  have the same fixed-point sets, i.e.  $\text{Fix } T_1 \equiv \text{Fix } T_2$ . Indeed, if  $T_1 x = x$  then according to (34) we have  $T_2 x = x$ , and vice versa. Since  $T_1$  is  $\kappa_1$ -linearly regular, i.e.

$$\text{dist}_{\text{Fix } T_1}(x) \leq \kappa_1 \|x - T_1 x\|,$$

we have

$$\begin{aligned} \|x - T_2 x\| & = \frac{1 + \rho_1}{1 + \rho_2} \|x - T_1 x\| \\ & \geq \frac{1 + \rho_1}{1 + \rho_2} \frac{1}{\kappa_1} \text{dist}_{\text{Fix } T_2}(x), \end{aligned}$$

which proves that  $T_2$  is linearly regular with constant  $\kappa_2 := \frac{1 + \rho_2}{1 + \rho_1} \kappa_1$ .

Suppose now that  $\kappa_1$  is a tight estimate of the linear regularity constant, i.e. there exists some  $\bar{x}$  such that

$$\text{dist}_{\text{Fix } T_1}(\bar{x}) = \kappa_1 \|\bar{x} - T_1 \bar{x}\|.$$

Then we have

$$\|\bar{x} - T_2 \bar{x}\| = \frac{1}{\kappa_2} \text{dist}_{\text{Fix } T_2}(\bar{x}),$$

which means that  $\kappa_2$  is a tight estimate of the linear regularity constant of  $T_2$ .

*Proof of (ii):*

The convergence factor of  $T_2$  as a function of  $\rho_2$  is

$$\beta_2(\rho_2) = \sqrt{1 - \frac{\rho_2}{\kappa_2^2}} = \sqrt{1 - \frac{\rho_2}{(1 + \rho_2)^2} \left(\frac{1 + \rho_1}{\kappa_1}\right)^2}.$$

This function is minimized when  $\rho_2/(1 + \rho_2)^2$  is maximized, which implies that the smallest convergence factor is attained for  $\rho_2 = 1$ .

## APPENDIX D PROOF OF PROPOSITION 4

The Douglas-Rachford operator  $T_{DR}$  is denoted by  $T$  in the sequel. Note from [30, Prop. 3.6(i)] that  $z^* := \mathbb{P}_{\text{Fix } T}(z^k)$  is in general not equal to  $x^* := \mathbb{P}_{A \cap B}(z^k)$ , and that they coincide only when  $z^k \in A + B$ . Since  $T$  is an averaged operator, the sequence  $\{z^k\}$  is Fejér monotone with respect to  $\text{Fix } T$ , and according to [30, Prop. 3.6(i)]  $\text{Fix } T$  is a closed subspace. This means that  $\mathbb{P}_{\text{Fix } T}(z^k) = \mathbb{P}_{\text{Fix } T}(z^0)$  for all  $k \in \mathbb{N}$  [33, Prop. 5.9], i.e.  $z^*$  does not depend on the iteration  $k$ . Also, from  $\mathbb{P}_{A \cap B} = \mathbb{P}_{A \cap B} \mathbb{P}_{\text{Fix } T}$  [30, Prop. 3.6(v)], it follows that  $x^* = \mathbb{P}_{A \cap B}(z^*)$ , meaning that  $x^*$  does not depend on the iteration  $k$  either.

Let  $z_P^k := \mathbb{P}_{A+B}(z^k)$ . We first provide a lemma that will be helpful in proving the main result.

*Lemma 5:* Let  $A$  and  $B$  be two closed subspaces of  $\mathcal{H}$  such that  $A + B$  is closed. Then the following hold for the iterates of the Douglas-Rachford splitting

- (i)  $z^k - z^* = z_P^k - x^*$ .
- (ii)  $x^{k+1} = \mathbb{P}_A(z_P^k)$  and  $y^{k+1} = \mathbb{P}_B(2x^{k+1} - z_P^k)$ .
- (iii)  $\|x^{k+1} - \mathbb{P}_B(x^{k+1})\|^2 \geq (1 - c_F^2(A, B)) \|x^{k+1} - x^*\|^2$ .
- (iv)  $\|y^{k+1} - \mathbb{P}_B(x^{k+1})\|^2 \geq (1 - c_F^2(A, B)) \|x^{k+1} - z_P^k\|^2$ .

*Proof of (i):*

By [30, Prop. 3.6(ii)] we have

$$z^* = x^* + \mathbb{P}_{A^\perp \cap B^\perp}(z^k),$$

which combined with the relation

$$(A^\perp \cap B^\perp)^\perp = A + B, \quad (35)$$

implies

$$z^k - \mathbb{P}_{A+B}(z^k) = \mathbb{P}_{(A+B)^\perp}(z^k) = z^* - x^*.$$

*Proof of (ii):*

From the linearity of the projection onto a subspace, it follows

$$\begin{aligned} x^{k+1} &= \mathbb{P}_A(z^k) = \mathbb{P}_A(\overbrace{z^k - \mathbb{P}_{A+B}(z^k)}^{\in (A+B)^\perp}) + \mathbb{P}_A(\mathbb{P}_{A+B}(z^k)) \\ &= \mathbb{P}_A(\mathbb{P}_{A+B}(z^k)), \end{aligned}$$

and similarly

$$\begin{aligned} y^{k+1} &= \mathbb{P}_B(2x^{k+1} - z^k) \\ &= \mathbb{P}_B(2x^{k+1} - \mathbb{P}_{A+B}(z^k)) - \mathbb{P}_B(\underbrace{z^k - \mathbb{P}_{A+B}(z^k)}_{\in (A+B)^\perp}) \\ &= \mathbb{P}_B(2x^{k+1} - \mathbb{P}_{A+B}(z^k)). \end{aligned}$$

*Proof of (iii):*

We first show that the following hold

$$\begin{aligned} x^{k+1} - x^* &\in A \cap (A \cap B)^\perp, \\ \mathbb{P}_B(x^{k+1}) - x^* &\in B \cap (A \cap B)^\perp. \end{aligned}$$

From (19a) and the definition of  $x^*$  it is clear that  $x^{k+1} - x^* \in A$ . Since the projection onto  $A \cap B$  is a linear operation, we have

$$\begin{aligned} \mathbb{P}_{A \cap B}(x^{k+1}) &= \mathbb{P}_{A \cap B}(\mathbb{P}_A(z^k)) \\ &= \mathbb{P}_{A \cap B}(\underbrace{\mathbb{P}_A(z^k) - z^k}_{\in A^\perp}) + \mathbb{P}_{A \cap B}(z^k) \\ &= x^*, \end{aligned}$$

which proves that  $x^{k+1} - x^* = x^{k+1} - \mathbb{P}_{A \cap B}(x^{k+1}) \in (A \cap B)^\perp$ . Similarly, from the definition of  $x^*$  it is clear that  $\mathbb{P}_B(x^{k+1}) - x^* \in B$ . We also have

$$\begin{aligned} \mathbb{P}_{A \cap B}(\mathbb{P}_B(x^{k+1})) &= \mathbb{P}_{A \cap B}(\underbrace{\mathbb{P}_B(x^{k+1}) - x^{k+1}}_{\in B^\perp}) \\ &\quad + \mathbb{P}_{A \cap B}(x^{k+1}) \\ &= x^*, \end{aligned}$$

which proves that

$$\mathbb{P}_B(x^{k+1}) - x^* = \mathbb{P}_B(x^{k+1}) - \mathbb{P}_{A \cap B}(\mathbb{P}_B(x^{k+1})) \in (A \cap B)^\perp.$$

From (20) it follows

$$\begin{aligned} c_F(A, B) &\geq \frac{\overbrace{\langle x^{k+1} - x^*, \mathbb{P}_B(x^{k+1}) - x^* \rangle}^{\in A \cap (A \cap B)^\perp} \overbrace{\langle \mathbb{P}_B(x^{k+1}) - x^*, x^{k+1} - x^* \rangle}^{\in B \cap (A \cap B)^\perp}}{\|x^{k+1} - x^*\| \|\mathbb{P}_B(x^{k+1}) - x^*\|} \\ &= \frac{\|\mathbb{P}_B(x^{k+1}) - x^*\|^2}{\|x^{k+1} - x^*\| \|\mathbb{P}_B(x^{k+1}) - x^*\|} \\ &= \frac{\|\mathbb{P}_B(x^{k+1}) - x^*\|}{\|x^{k+1} - x^*\|}, \end{aligned}$$

so that

$$\begin{aligned} 1 - c_F^2(A, B) &\leq \frac{\|x^{k+1} - x^*\|^2 - \overbrace{\|\mathbb{P}_B(x^{k+1}) - x^*\|^2}^{\in B}}{\|x^{k+1} - x^*\|^2} \\ &= \frac{\overbrace{\|x^{k+1} - \mathbb{P}_B(x^{k+1})\|^2}^{\in B^\perp}}{\|x^{k+1} - x^*\|^2}. \end{aligned}$$

*Proof of (iv):*

We first show that the following hold

$$\begin{aligned} x^{k+1} - z_P^k &\in A^\perp \cap (A^\perp \cap B^\perp)^\perp, \\ \mathbb{P}_{B^\perp}(x^{k+1} - z_P^k) &\in B^\perp \cap (A^\perp \cap B^\perp)^\perp. \end{aligned}$$

From part (ii) of the lemma it is clear that  $x^{k+1} - z_P^k = \mathbb{P}_A(z_P^k) - z_P^k \in A^\perp$ . From the definition of  $x^{k+1}$  and  $z_P^k$  we also have  $x^{k+1} - z_P^k \in A + B$ , which combined with (35) proves that  $x^{k+1} - z_P^k \in (A^\perp \cap B^\perp)^\perp$ . Similarly, it is clear that  $\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k) \in B^\perp$ . Note that  $\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k)$  can be split as follows

$$\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k) = (x^{k+1} - z_P^k) - \mathbb{P}_B(x^{k+1} - z_P^k).$$

Since the first summand of the right hand side is in  $A + B$  and the second one is in  $B$ , their difference is in  $A + B$ , which proves that  $\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k) \in (A^\perp \cap B^\perp)^\perp$ .

From (20) it follows

$$\begin{aligned} c_F(A^\perp, B^\perp) &\geq \frac{\overbrace{\langle x^{k+1} - z_P^k, \mathbb{P}_{B^\perp}(x^{k+1} - z_P^k) \rangle}^{\in A^\perp \cap (A+B)} \overbrace{\langle \mathbb{P}_{B^\perp}(x^{k+1} - z_P^k), x^{k+1} - z_P^k \rangle}^{\in B^\perp \cap (A+B)}}{\|x^{k+1} - z_P^k\| \|\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k)\|} \\ &= \frac{\|\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k)\|^2}{\|x^{k+1} - z_P^k\| \|\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k)\|} \\ &= \frac{\|\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k)\|}{\|x^{k+1} - z_P^k\|}, \end{aligned}$$

so that

$$\begin{aligned} 1 - c_F^2(A^\perp, B^\perp) &\leq \frac{\|x^{k+1} - z_P^k\|^2 - \overbrace{\|\mathbb{P}_{B^\perp}(x^{k+1} - z_P^k)\|^2}^{\in B^\perp}}{\|x^{k+1} - z_P^k\|^2} \\ &= \frac{\overbrace{\|x^{k+1} - \mathbb{P}_{B^\perp}(x^{k+1} - z_P^k)\|^2}^{\in B}}{\|x^{k+1} - z_P^k\|^2}. \end{aligned}$$

By observing that the following holds

$$\begin{aligned}\mathbb{P}_B(x^{k+1} - z_P^k) &= \mathbb{P}_B(2x^{k+1} - z_P^k) - \mathbb{P}_B(x^{k+1}) \\ &= y^{k+1} - \mathbb{P}_B(x^{k+1}),\end{aligned}$$

and that  $c_F(A^\perp, B^\perp) = c_F(A, B)$  [30, Fact 2.3], we have

$$1 - c_F^2(A, B) \leq \frac{\|y^{k+1} - \mathbb{P}_B(x^{k+1})\|^2}{\|x^{k+1} - z_P^k\|^2}.$$

Now we are ready to derive the linear regularity constant. The following holds

$$\begin{aligned}\|z^k - z^*\|^2 &= \|z_P^k - x^*\|^2 \\ &= \underbrace{\|z_P^k - x^{k+1}\|^2}_{\in A^\perp} + \underbrace{\|x^{k+1} - x^*\|^2}_{\in A} \\ &\leq \frac{1}{1 - c_F^2(A, B)} \underbrace{\|y^{k+1} - \mathbb{P}_B(x^{k+1})\|^2}_{\in B} \\ &\quad + \frac{1}{1 - c_F^2(A, B)} \underbrace{\|x^{k+1} - \mathbb{P}_B(x^{k+1})\|^2}_{\in B^\perp} \\ &= \frac{1}{1 - c_F^2(A, B)} \|y^{k+1} - x^{k+1}\|^2 \\ &= \frac{1}{1 - c_F^2(A, B)} \|z^k - z^{k+1}\|^2.\end{aligned}$$

The first equality follows from Lemma 5(i), first inequality from Lemma 5(iii)–(iv), and the last equality follows from (19c). This concludes the proof.

## APPENDIX E PROOF OF LEMMA 3

We first provide the following supporting results. We denote by  $\text{Fix } T := \text{Fix } T_{DR} = \text{Fix } T_{GDR}$ .

*Lemma 6:* Let  $A$  and  $B$  be closed subspaces of  $\mathcal{H}$ . Then

$$c_F(A, B) = \|(\mathbb{P}_B - \mathbb{P}_{A \cap B})(\mathbb{P}_A - \mathbb{P}_{A \cap B})\|.$$

*Proof:* From [30, Fact 2.3] we have the following identity

$$c_F(A, B) = \|\mathbb{P}_B \mathbb{P}_A - \mathbb{P}_{A \cap B}\|.$$

It is now sufficient to show that the right-hand terms of the above identities are equal. We have

$$\begin{aligned}(\mathbb{P}_B - \mathbb{P}_{A \cap B})(\mathbb{P}_A - \mathbb{P}_{A \cap B}) &= \mathbb{P}_B \mathbb{P}_A - \mathbb{P}_B \mathbb{P}_{A \cap B} - \mathbb{P}_{A \cap B} \mathbb{P}_A + \mathbb{P}_{A \cap B}^2 \\ &= \mathbb{P}_B \mathbb{P}_A - \mathbb{P}_{A \cap B} - \mathbb{P}_{A \cap B} + \mathbb{P}_{A \cap B} \\ &= \mathbb{P}_B \mathbb{P}_A - \mathbb{P}_{A \cap B},\end{aligned}$$

where the second equality follows from [30, Eq. (7)].  $\blacksquare$

*Lemma 7:* Let  $A$  and  $B$  be closed subspaces of  $\mathcal{H}$ , and  $T_{DR}$  the operator given in (18). Then

$$c_F(A, B) \leq \sup_{u \in A \setminus \text{Fix } T} \frac{\text{dist}_{A \cap B}(\mathbb{P}_B(u))}{\text{dist}_{A \cap B}(u)}. \quad (36)$$

*Proof:* The characterization of  $c_F(A, B)$  in Lemma 6 implies

$$c_F(A, B) = \sup_{x \neq 0} \frac{\|(\mathbb{P}_B - \mathbb{P}_{A \cap B})(\mathbb{P}_A - \mathbb{P}_{A \cap B})x\|}{\|x\|}. \quad (37)$$

If  $c_F(A, B) = 0$ , then (36) holds trivially. We thus assume that  $c_F(A, B) > 0$ , which implies that for some  $x \neq 0$  the numerator in (37) is nonzero. We can therefore disregard  $x \in (A \cap B) + A^\perp$  from the above supremum since in that case  $(\mathbb{P}_A - \mathbb{P}_{A \cap B})x = 0$  and the numerator in (37) is zero. We now have

$$\begin{aligned}c_F(A, B) &= \sup_{x \notin (A \cap B) + A^\perp} \frac{\|(\mathbb{P}_B - \mathbb{P}_{A \cap B})(\mathbb{P}_A - \mathbb{P}_{A \cap B})x\|}{\|x\|} \\ &\leq \sup_{x \notin (A \cap B) + A^\perp} \frac{\|(\mathbb{P}_B - \mathbb{P}_{A \cap B})(\mathbb{P}_A - \mathbb{P}_{A \cap B})x\|}{\|(\mathbb{P}_A - \mathbb{P}_{A \cap B})x\|} \\ &\leq \sup_{u \in A \setminus \{0\}} \frac{\|(\mathbb{P}_B - \mathbb{P}_{A \cap B})u\|}{\|u\|}\end{aligned}$$

where the second line follows from the nonexpansiveness of projection  $\mathbb{P}_A$  and the fact that  $0 \in A \cap B$ , and the third from the fact that  $(\mathbb{P}_A - \mathbb{P}_{A \cap B})x \in A$ . Similarly, we can disregard  $u \in \text{Fix } T$  from the above supremum since then  $(\mathbb{P}_B - \mathbb{P}_{A \cap B})x = 0$  [30, Prop. 3.6(v)]. We now obtain

$$c_F(A, B) \leq \sup_{u \in A \setminus \text{Fix } T} \frac{\text{dist}_{A \cap B}(\mathbb{P}_B(u))}{\text{dist}_{A \cap B}(u)},$$

where we used the fact that  $0 \in A \cap B$  and  $\mathbb{P}_{A \cap B} = \mathbb{P}_{A \cap B} \mathbb{P}_B$  [46, Thm. 5.14].  $\blacksquare$

We are now in a position to prove Lemma 3. From the definition of the generalized Douglas-Rachford operator (23) we obtain

$$T_{GDR}z^k - z^{k+1} = (1 - 2\alpha)(z^k - z^{k+1}),$$

where  $z^{k+1} = T_{DR}z^k$ .

For  $z^k \in A \setminus \text{Fix } T$  DRS produces the following iterates in the subsequent iteration

$$\begin{aligned}x^{k+1} &= z^k, \\ y^{k+1} &= \mathbb{P}_B(z^k), \\ z^{k+1} &= \mathbb{P}_B(z^k).\end{aligned}$$

Note also that the following inclusions hold

$$\begin{aligned}z^k - z^{k+1} &= z^k - \mathbb{P}_B(z^k) \in B^\perp, \\ z^{k+1} - z^* &= \mathbb{P}_B(z^k) - \mathbb{P}_{A \cap B}(z^k) \in B,\end{aligned}$$

where  $z^* := \mathbb{P}_{\text{Fix } T}(z^k) = \mathbb{P}_{A \cap B}(z^k) = \mathbb{P}_{\text{Fix } T}(T_{GDR}z^k)$  [30, Prop. 3.6], [47, Lem. 3.12]. We now have

$$\begin{aligned}T_{GDR}z^k - z^* &= (T_{GDR}z^k - z^{k+1}) + (z^{k+1} - z^*) \\ &= (1 - 2\alpha) \underbrace{(z^k - z^{k+1})}_{\in B^\perp} + \underbrace{(z^{k+1} - z^*)}_{\in B},\end{aligned}$$

and therefore

$$\|T_{GDR}z^k - z^*\|^2 = (1 - 2\alpha)^2 \|z^k - z^{k+1}\|^2 + \|z^{k+1} - z^*\|^2.$$

Dividing the above equality by  $\|z^k - z^*\|^2 > 0$ , and taking the supremum over  $z^k \in A \setminus \text{Fix } T$ , we obtain

$$\begin{aligned}&\sup_{z^k \in A \setminus \text{Fix } T} \frac{\|T_{GDR}z^k - z^*\|^2}{\|z^k - z^*\|^2} \\ &= \sup_{z^k \in A \setminus \text{Fix } T} \left( (1 - 2\alpha)^2 \frac{\|z^k - z^{k+1}\|^2}{\|z^k - z^*\|^2} + \frac{\|\mathbb{P}_B(z^k) - z^*\|^2}{\|z^k - z^*\|^2} \right) \\ &\geq (1 - 2\alpha)^2 (1 - c_F^2(A, B)) + c_F^2(A, B) \\ &= \beta_{GDR}^2,\end{aligned}$$

where we used  $z^{k+1} = \mathbb{P}_B(z^k)$  in the second line, and the third line follows from Proposition 4 and Lemma 7. Finally, we have

$$\begin{aligned} \sup_{z^k \notin \text{Fix } T} \frac{\|T_{GDR} z^k - z^*\|^2}{\|z^k - z^*\|^2} &\geq \sup_{z^k \in A \setminus \text{Fix } T} \frac{\|T_{GDR} z^k - z^*\|^2}{\|z^k - z^*\|^2} \\ &\geq \beta_{GDR}^2. \end{aligned}$$

## REFERENCES

- [1] G. Banjac and P. J. Goulart, “Global linear convergence in operator splitting methods,” in *IEEE Conference on Decision and Control (CDC)*, 2016, pp. 233–238.
- [2] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [3] P. L. Combettes and J.-C. Pesquet, “Proximal splitting methods in signal processing,” *Springer Optimization and Its Applications*, vol. 49, pp. 185–212, 2011.
- [4] S. Richter, C. N. Jones, and M. Morari, “Computational complexity certification for real-time MPC with input constraints based on the fast gradient method,” *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1391–1403, 2012.
- [5] B. O’Donoghue, G. Stathopoulos, and S. Boyd, “A splitting method for optimal control,” *IEEE Transactions on Control Systems Technology*, vol. 21, no. 6, pp. 2432–2442, 2013.
- [6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [7] N. Parikh and S. Boyd, “Proximal algorithms,” *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [8] E. Ryu and S. Boyd, “Primer on monotone operator methods,” *International Journal of Applied and Computational Mathematics*, vol. 15, no. 1, pp. 3–43, 2016.
- [9] S. Richter, C. N. Jones, and M. Morari, “Certification aspects of the fast gradient method for solving the dual of parametric convex programs,” *Mathematical Methods of Operations Research*, vol. 77, no. 3, pp. 305–321, 2013.
- [10] B. O’Donoghue and E. Candès, “Adaptive restart for accelerated gradient schemes,” *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 715–732, 2015.
- [11] W. Su, S. Boyd, and E. Candès, “A differential equation for modeling Nesterov’s accelerated gradient method: Theory and insights,” *Journal of Machine Learning Research*, vol. 17, pp. 1–43, 2016.
- [12] G. E., 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*, vol. 60, no. 3, pp. 644–658, 2015.
- [13] P. Giselsson and S. Boyd, “Linear convergence and metric selection in Douglas-Rachford splitting and ADMM,” *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 532–544, 2017.
- [14] P. Giselsson, “Tight global linear convergence rate bounds for Douglas-Rachford splitting,” *Journal of Fixed Point Theory and Applications*, pp. 1–30, 2017.
- [15] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, “Explicit convergence rate of a distributed alternating direction method of multipliers,” *IEEE Transactions on Automatic Control*, vol. 61, no. 4, pp. 892–904, 2016.
- [16] S. Richter, “Computational complexity certification of gradient methods for real-time model predictive control,” Ph.D. dissertation, ETH-Zürich, 2012.
- [17] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, ser. Applied optimization. Kluwer Academic Publishers, 2004.
- [18] Z.-Q. Luo and P. Tseng, “Error bounds and convergence analysis of feasible descent methods: a general approach,” *Annals of Operations Research*, vol. 46, no. 1, pp. 157–178, 1993.
- [19] D. Boley, “Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs,” *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2183–2207, 2013.
- [20] H. M. Phan, “Linear convergence of the Douglas-Rachford method for two closed sets,” *Optimization*, vol. 65, no. 2, pp. 369–385, 2016.
- [21] J. Eckstein and D. Bertsekas, “An alternating direction method for linear programming,” available: <http://dspace.mit.edu/handle/1721.1/3197>, 1990.
- [22] P.-W. Wang and C.-J. Lin, “Iteration complexity of feasible descent methods for convex optimization,” *Journal of Machine Learning Research*, vol. 15, pp. 1523–1548, 2014.
- [23] I. Necoara and V. Nedelcu, “On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems,” *Automatica*, vol. 55, pp. 209–216, 2015.
- [24] I. Necoara, “Computational complexity certification for dual gradient method: Application to embedded MPC,” *Systems & Control Letters*, vol. 81, pp. 49–56, 2015.
- [25] I. Necoara, Y. Nesterov, and F. Glineur, “Linear convergence of first order methods for non-strongly convex optimization,” *arXiv preprint arXiv:1504.06298*, 2016.
- [26] H. Zhang, “The restricted strong convexity revisited: analysis of equivalence to error bound and quadratic growth,” *Optimization Letters*, pp. 1–17, 2016.
- [27] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. W. Suter, “From error bounds to the complexity of first-order descent methods for convex functions,” *Mathematical Programming*, pp. 1–37, 2016.
- [28] D. Drusvyatskiy and A. S. Lewis, “Error bounds, quadratic growth, and linear convergence of proximal methods,” *arXiv preprint arXiv:1602.06661*, 2016.
- [29] F. Deutsch and H. Hundal, “The rate of convergence of Dykstra’s cyclic projections algorithm: The polyhedral case,” *Numerical Functional Analysis and Optimization*, vol. 15, no. 5-6, pp. 537–565, 1994.
- [30] H. H. Bauschke, J. Y. Bello Cruz, T. T. A. Nghia, H. M. Phan, and X. Wang, “The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle,” *Journal of Approximation Theory*, vol. 185, pp. 63–79, 2014.
- [31] R. Nishihara, S. Jegelka, and M. I. Jordan, “On the convergence rate of decomposable submodular function minimization,” in *International Conference on Neural Information Processing Systems (NIPS)*, 2014, pp. 640–648.
- [32] H. H. Bauschke, D. Noll, and H. M. Phan, “Linear and strong convergence of algorithms involving averaged nonexpansive operators,” *Journal of Mathematical Analysis and Applications*, vol. 421, no. 1, pp. 1–20, 2015.
- [33] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 1st ed. Springer Publishing Company, Incorporated, 2011.
- [34] A. Cegielski, “Application of quasi-nonexpansive operators to an iterative method for variational inequality,” *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2165–2181, 2015.
- [35] Y. Song and Q. Li, “Successive approximations for quasi-firmly type nonexpansive mappings,” *Mathematical Communications*, vol. 16, no. 1, pp. 251–264, 2011.
- [36] H. Attouch, J. Bolte, and B. F. Svaiter, “Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods,” *Mathematical Programming*, vol. 137, no. 1, pp. 91–129, 2013.
- [37] P. L. Combettes and I. Yamada, “Compositions and convex combinations of averaged nonexpansive operators,” *Journal of Mathematical Analysis and Applications*, vol. 425, no. 1, pp. 55–70, 2015.
- [38] J. von Neumann, *Functional Operators, Vol. II*. Princeton University Press, 1950.
- [39] C. V. Rao and J. B. Rawlings, “Linear programming and model predictive control,” *Journal of Process Control*, vol. 10, no. 2–3, pp. 283–289, 2000.
- [40] A. Bemporad, F. Borrelli, and M. Morari, “Model predictive control based on linear programming – the explicit solution,” *IEEE Transactions on Automatic Control*, vol. 47, no. 12, pp. 1974–1985, 2002.
- [41] —, “Min-max control of constrained uncertain discrete-time linear systems,” *IEEE Transactions on Automatic Control*, vol. 48, no. 9, pp. 1600–1606, 2003.
- [42] E. C. Kerrigan and J. M. Maciejowski, “Feedback min-max model predictive control using a single linear program: robust stability and the explicit solution,” *International Journal of Robust and Nonlinear Control*, vol. 14, no. 4, pp. 395–413, 2004.
- [43] B. O’Donoghue, E. Chu, N. Parikh, and S. Boyd, “Conic optimization via operator splitting and homogeneous self-dual embedding,” *Journal of Optimization Theory and Applications*, vol. 169, no. 3, pp. 1042–1068, 2016.
- [44] D. Davis, “On the design and analysis of operator-splitting schemes,” Ph.D. dissertation, University of California, Los Angeles, 2015.
- [45] D. Davis and W. Yin, “Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions,” *Mathematics of Operations Research*, 2017.

- [46] F. Deutsch, *Best Approximation in Inner Product Spaces*, 1st ed. Springer-Verlag, 2001.
- [47] H. H. Bauschke, F. Deutsch, H. Hundal, and S.-H. Park, "Accelerating the convergence of the method of alternating projections," *Transactions of the American Mathematical Society*, vol. 355, no. 9, pp. 3433–3461, 2003.



**Goran Banjac** received the B.Sc. and M.Sc. degrees, both in electrical engineering, from the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, in 2011 and 2013, respectively. He was working as a research assistant at the Department of Control and Computer Engineering at the same Faculty in 2013 and 2014. He is pursuing his D.Phil. (Ph.D.) degree in control engineering at the University of Oxford under the supervision of Professor Paul J. Goulart.

His research interests include mathematical programming, model predictive control, high-speed and large-scale optimization.



**Paul J. Goulart** received the B.Sc. and M.Sc. degrees in aeronautics and astronautics from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He was selected as a Gates Scholar at the University of Cambridge, Cambridge, U.K., where he received the Ph.D. degree in control engineering in 2007.

From 2007 to 2011, he was a Lecturer in control systems in the Department of Aeronautics, Imperial College London, and from 2011 to 2014 a Senior Researcher in the Automatic Control Laboratory, ETH Zürich. He is currently an Associate Professor in the Department of Engineering Science and Tutorial Fellow, St. Edmund Hall, University of Oxford, Oxford, U.K. His research interests include model predictive control, robust optimization, and control of fluid flows.