

Global Linear Convergence in Operator Splitting Methods

Goran Banjac and Paul J. Goulart

Abstract—We establish necessary and sufficient conditions for linear convergence of operator splitting methods for a general class of convex optimization problems where the associated fixed-point operator is averaged. Most existing results establishing 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 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 $x \in \mathbb{R}^n$ is the optimization variable 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 [1], [2], optimal control [3], [4] and machine learning [5]. There exist 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) [6], [7].

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 an 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 [2], 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 [8].

Over the last decade first-order methods have attracted renewed interest due to their tractability in large-scale optimization and to the fact that computed solutions need not

be very accurate in many applications [8]. However, 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 [3], [9], [10], [7].

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 [11], [12], [13] 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 [14]. The work in [15] establishes linear convergence of feasible descent methods for solving a particular class of non-strongly convex problems, and the authors in [16], [17] use a similar approach to prove linear convergence of some methods for solving optimization problems of the type arising in model predictive control (MPC). Linear convergence of some projection methods, which are a 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 [18], [19]. The authors in [20] 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.

In this paper we establish necessary and sufficient conditions for linear convergence of a broad class of first-order optimization methods. These results provide a unifying framework for the analysis of linear convergence encompassing all of the aforementioned results as special cases, and identify linear regularity as the essential feature of averaged non-contractive operators to ensure linear convergence.

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. In Section III we establish linear convergence of the APM for solving a feasibility problem involving two convex polyhedra using our framework. A particular linearly convergent method for LP is derived in IV. Section V concludes the paper.

Notation: We denote by \mathbb{R}^n the n -dimensional real space, the associated Euclidean norm by $\|\cdot\|$ and the

The authors are with the 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)].

extended real line by $\tilde{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$. The class of closed, proper and convex functions $f : \mathbb{R}^n \rightarrow \tilde{\mathbb{R}}$ is denoted by $\Gamma_0(\mathbb{R}^n)$. The notation $(x, y, z) := [x^\top \ y^\top \ z^\top]^\top$ is used to denote stacked real column vectors. The set of fixed-points of an operator $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined as $\text{fix}T := \{x \in \mathbb{R}^n \mid x = Tx\}$. The distance of a vector x from a set C is $d_C(x) := \inf \{\|x - y\| \mid y \in C\}$, and the distance between two sets, C_1 and C_2 , is $d(C_1, C_2) := \inf \{\|x - y\| \mid x \in C_1, y \in C_2\}$. The Euclidean projection of a vector x onto a set C is $\mathbb{P}_C(x) := \{y \in C \mid \|y - x\| = d_C(x)\}$. The affine hull of a set C is denoted by $\text{aff}(C)$, and the subspace obtainable via translation of $\text{aff}(C)$ is denoted by $\text{aff}_0(C)$, i.e. $\text{aff}_0(C) := \text{aff}(C) - c$ for some $c \in \text{aff}(C)$. A face P_x of a polyhedron $P \subseteq \mathbb{R}^n$ is defined as a nonempty minimizer of a linear function over P , i.e. $P_x := \text{argmin}_{p \in P} x^\top p$. The set of all faces of a polyhedron P is denoted by \mathcal{F}_P , i.e. $\mathcal{F}_P := \{P_x \mid x \in \mathbb{R}^n\}$.

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). We define an operator T such that the iterations computed by the algorithm are equivalent to the iteration

$$x^{k+1} = Tx^k \quad (2)$$

for all $x^k \in \mathcal{X} \subseteq \mathbb{R}^n$. The iteration (2) is referred to as *Picard* or *fixed-point iteration* of an operator T .

Under mild assumptions on such an operator T , it can be shown that a convergent sequence $\{x^k\}$ generated by (2) converges to a point in $\text{fix}T$ [21, Theorem 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 [7]), 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 [7]). 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

$$d_{\text{fix}T}(Tx) \leq \beta d_{\text{fix}T}(x) \quad (3)$$

for all $x \in \mathcal{X}$, and for some $\beta \in [0, 1)$. The constant β is called the *convergence factor* and determines the 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 [21] for a comprehensive review.

Definition 1 ([21]). *Let \mathcal{X} be a nonempty subset of \mathbb{R}^n and let $T : \mathcal{X} \rightarrow \mathbb{R}^n$. Then T is*

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

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

(ii) *α -averaged with $\alpha \in (0, 1)$ if there exists a nonexpansive operator $R : \mathcal{X} \rightarrow \mathbb{R}^n$ such that $T = (1 - \alpha)\text{Id} + \alpha R$,*

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

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

It can be shown that every α -averaged operator T satisfies

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

(see [21, Proposition 4.25]). An operator satisfying (5) for $\alpha = \frac{1}{2}$ is called *firmly nonexpansive*, and hence, an operator is firmly nonexpansive if and only if it is $\frac{1}{2}$ -averaged.

Operators arising from the iteration of a particular optimization method often enjoy at least some subset of properties described in Definition 1. Figure 1 illustrates these properties and highlights that the distance from $\text{fix}T$ of the iterates of a fixed-point iteration is nonincreasing if any of the properties above holds. In the case of α -averaged and β -contractive operators the sequence converges to a fixed-point [22]. Moreover, in the latter case the fixed-point iterates satisfy the inequality (3).

Note that without additional assumptions the same can not be claimed for an averaged operator. In the sequel we will identify the essential additional property of an averaged 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. If $\text{fix}T$ is not a singleton, then condition (3) itself does not imply linear convergence of the sequence. However, if T is nonexpansive then (3) implies linear convergence of $\{x^k\}$ [21, Theorem 5.12]. Therefore, instead of claiming that (3) holds, we will sometimes state that $\{x^k\}$ converges linearly because the underlying operator is usually averaged, hence nonexpansive.*

A. Linear regularity

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

Definition 2 ([22]). *Let $T : \mathcal{X} \rightarrow \mathbb{R}^n$ be such that $\text{fix}T \neq \emptyset$. We say that T is linearly regular with constant $\kappa \geq 0$ if $(\forall x \in \mathcal{X})$*

$$d_{\text{fix}T}(x) \leq \kappa \|x - Tx\|. \quad (6)$$

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 $\frac{1}{\kappa} d_{\text{fix}T}(x^k)$. Figure 2 illustrates that if the linear regularity constant of an averaged 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 $T : \mathcal{X} \rightarrow \mathbb{R}^n$ be an α -averaged operator and suppose that $\text{fix}T \neq \emptyset$. Then the inequality*

$$d_{\text{fix}T}(Tx) \leq \beta d_{\text{fix}T}(x)$$

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

$$\beta = \sqrt{1 - \frac{1-\alpha}{\alpha\kappa^2}} \in [0, 1). \quad (7)$$

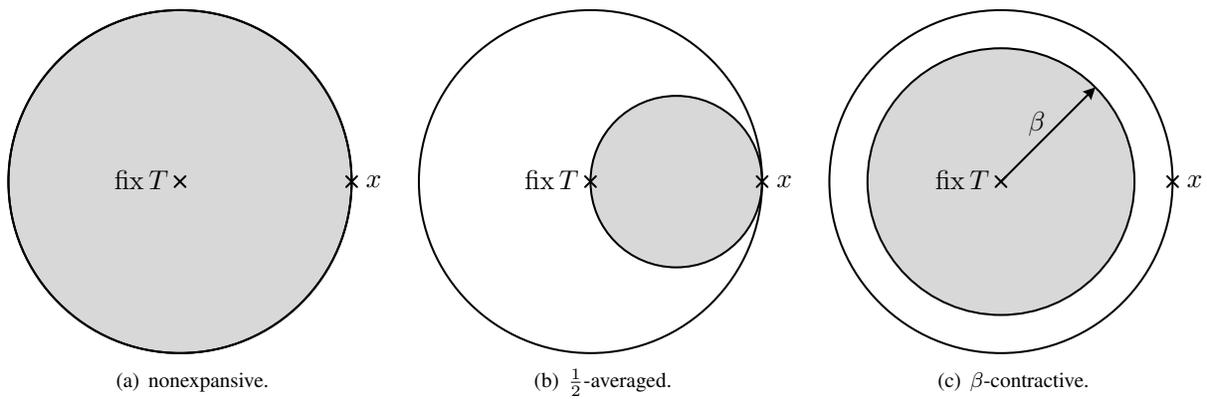


Fig. 1. Graphical representation of some operator properties. The point/set Tx lies somewhere inside the grey shaded area. We assume that x is on the unit circle with the center at $\text{fix } T$.

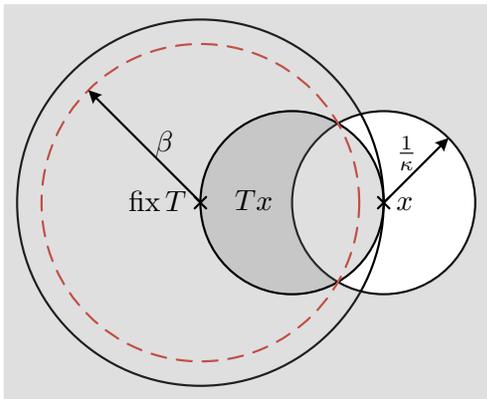


Fig. 2. Linear convergence of an averaged and linearly regular operator holds if the linear regularity constant κ is bounded from above. In that case Tx lies inside the circle with a radius $\beta < 1$. We assume that x is on the unit circle with the center at $\text{fix } T$.

Proof. See Appendix A. ■

Theorem 1 states that linear regularity, or a lack thereof, is the essential determinant of linear convergence for a sequence generated by fixed-point iteration of an averaged operator. Moreover, it can be shown that the bound in (7) is actually tight [23].

Previous results in [22] have established linear regularity as a sufficient condition for (3), albeit with the weaker convergence rate bound

$$\beta = \sqrt{\frac{\sigma^{-1}\kappa^2}{1+\sigma^{-1}\kappa^2}} \quad (8)$$

where $\sigma := \frac{1-\alpha}{\alpha}$. It is easy to show that our bound in Theorem 1 is strictly better than (8). A related result appears in [14], which established linear convergence of a particular proximal method for linear programming based on the weaker condition

$$d_{\text{fix } T}(Tx) \leq \tilde{\kappa}\|x - Tx\|, \quad (9)$$

for some α -averaged operator T and some $\tilde{\kappa} > 0$, i.e. by lower bounding $\|x - Tx\|$ via $d_{\text{fix } T}(Tx)$ rather than $d_{\text{fix } T}(x)$. In this case one can show that condition (3) still holds with

convergence factor β as in (8) but with κ replaced by $\tilde{\kappa}$. By virtue of Theorem 1 any such operator must be linearly regular, so that (9) is also both necessary and sufficient for linear convergence of an averaged operator.

B. Improving the convergence factor

It follows from Definition 1(ii) that it is possible to obtain an averaged operator from a nonexpansive one via a suitable transformation. It turns out that if T_1 is α_1 -averaged, we can construct an α_2 -averaged operator T_2 via a further similar transformation, assuming that $(\alpha_1, \alpha_2) \in (0, 1)$. We next show how to select α_2 in order to obtain the smallest convergence factor of the resulting fixed-point iteration:

Proposition 1. *Let $\alpha_1, \alpha_2 \in (0, 1)$. Suppose $T_1 : \mathcal{X} \rightarrow \mathbb{R}^n$ is α_1 -averaged, with a linear regularity constant $\kappa_1 > 0$, and let $T_2 := (1 - \frac{\alpha_2}{\alpha_1})\text{Id} + \frac{\alpha_2}{\alpha_1}T_1$. Then the following hold*

- (i) T_2 is α_2 -averaged with linear regularity constant $\kappa_2 = \frac{\alpha_1}{\alpha_2}\kappa_1$.
- (ii) The smallest convergence factor is attained for $\alpha_2 = \frac{1}{2}$.

Proof. See Appendix B. ■

We will next show that a variety of existing results establishing linear convergence in convex feasibility problems can be seen as special cases of the approach introduced in Theorem 1. It can also be shown that linear regularity is very closely related to the error bound property [23], which is often used as an assumption for establishing linear convergence of some first order methods, e.g. in [11], [15], [16].

III. LINEAR CONVERGENCE OF THE ALTERNATING PROJECTION METHOD

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

$$\begin{aligned} &\text{find} && x \\ &\text{subject to} && x \in A \\ &&& x \in B. \end{aligned} \quad (10)$$

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

$$\text{minimize } \mathcal{I}_A(x) + \mathcal{I}_B(x) \quad (11)$$

where \mathcal{I}_A and \mathcal{I}_B are the indicator functions of sets A and B , respectively. The indicator function of a closed convex set $C \subset \mathbb{R}^n$ is defined as

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

with $\mathcal{I}_C \in \Gamma_0(\mathbb{R}^n)$ [21].

Projection methods are a special class of proximal methods for solving feasibility problems and date to von Neumann's *alternating projection method* [24] given by the following iteration

$$a^{k+1} = \mathbb{P}_A(b^k), \quad (12a)$$

$$b^{k+1} = \mathbb{P}_B(a^{k+1}). \quad (12b)$$

It is generally assumed that both set projection operators, \mathbb{P}_A and \mathbb{P}_B , can be evaluated efficiently. It can be shown that the APM is just a special case of the PGM for solving a reformulation of the problem (11) [7].

In the case when A and B are affine subspaces, the APM converges linearly and the convergence rate is characterized in terms of the *Friedrichs angle* between the subspaces [18]. The Friedrichs angle is a generalization of the angle between affine subspaces in higher dimensions:

Definition 3 ([19]). *Suppose that U and V are subspaces in \mathbb{R}^n . Cosine of the Friedrichs angle between U and V is*

$$c_F(U, V) := \sup\{u^\top v \mid u \in U \cap (U \cap V)^\perp, v \in V \cap (U \cap V)^\perp, \|u\| \leq 1, \|v\| \leq 1\}. \quad (13)$$

The authors in [20] showed that this result can be generalized to the case when A and B are two convex polyhedra. Observe that the indicator function of a closed convex set is neither strongly convex nor smooth, and thus the classical analysis of linear convergence based on strong convexity and smoothness of the constituent functions in the optimization problem is not applicable in this case. We will show that 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 averaged.

A. Linear regularity in the APM

The authors in [20] show that the APM for two convex polyhedra, 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, with notation following the one in [25].

The sets of *closest points* between two polyhedra, A and B , are defined as

$$E = \{a \in A \mid d_B(a) = d(A, B)\}, \quad (14a)$$

$$H = \{b \in B \mid d_A(b) = d(B, A)\}. \quad (14b)$$

Let $v = \mathbb{P}_{B-A}(0)$ and note that $H = E + v$. When $A \cap B \neq \emptyset$ it follows that $v = 0$ and $E = H = A \cap B$. Pairs of points $(e, e + v) \in E \times H$ are solutions to the best approximation problem between A and B .

Theorem 2 ([20]). *Let A and B be convex polyhedra and let $b^0 \in B$. Then sequences $\{a^k\}$ and $\{b^k\}$ generated by the alternating projection method converge linearly towards some points in E and H , so that*

$$\begin{aligned} d_E(a^{k+1}) &\leq \beta d_E(a^k), \\ d_H(b^{k+1}) &\leq \beta d_H(b^k), \end{aligned} \quad (15)$$

where the convergence factor is given by

$$\beta = \max_{\substack{A_x \in \mathcal{F}_A \\ B_y \in \mathcal{F}_B}} c_F^2(\text{aff}_0(A_x), \text{aff}_0(B_y)) < 1. \quad (16)$$

We will show that this result is a special case of Theorem 1. We can represent iteration (12) via a double iteration of the following fixed-point operator defined for $\mathcal{X} \equiv A \cup B$

$$Tx = \begin{cases} \mathbb{P}_A(x), & x \in B \\ \mathbb{P}_B(x), & x \in A \end{cases} \quad (17)$$

Clearly, this operator is firmly nonexpansive since it is defined via projection operators [21, Proposition 4.8]. 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 $\mathcal{X} \equiv A \cup B$ (see Proposition 4 and Corollary 5 in [20]). According to Theorem 1 these two properties ensure linear convergence with convergence factor $\sqrt{\beta}$, where β is given in (16). Double iteration of a fixed-point operator given in (17), which is equivalent to iteration (12), has thus convergence factor β . 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 all $x \in \mathbb{R}^n$, 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{X} . Note that the linear convergence of the APM for two subspaces is a special case of the above result.

IV. LINEAR PROGRAMMING

Linear convergence of a particular proximal method for linear programming was shown in [14], with an assumption that the problem is feasible with a bounded objective value. We will show that this result is a special case of Theorem 1. The underlying fixed-point operator can be shown to be averaged (see Lemma 8 in [14]) and that inequality (9) holds (see the proof of Theorem 4 in [14]). 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

$$\begin{aligned} \min. \quad & c^\top x & \max. \quad & -b^\top y \\ \text{s. t.} \quad & Ax + s = b & \text{s. t.} \quad & -A^\top y + r = c \\ & (x, s) \in \mathbb{R}^n \times \mathcal{K} & & (r, y) \in \{0\}^n \times \mathcal{K}^* \end{aligned} \quad (18)$$

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 [26] 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

$$\begin{aligned} \text{find} \quad & (u, v) \\ \text{subject to} \quad & v = Qu \\ & (u, v) \in \mathcal{C} \times \mathcal{C}^*, \end{aligned} \quad (19)$$

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

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

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

The objective of the above optimization problem is to find a point (u, v) that satisfies both the subspace and the conic constraint in (19). 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 (18).
2. If $\tau = 0$ and $\kappa > 0$, then either primal or dual problem is infeasible. The case $b^\top y < 0$ is a certificate for primal infeasibility, and the case $c^\top x < 0$ is a certificate for dual infeasibility.
3. If $\tau = \kappa = 0$, then nothing can be concluded about the solution of (18).

The problem (19) 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 [26] 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 $(u^0, v^0)^\top (u^*, v^*) > 0$. 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 [26] 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 1. *The sequence generated by the alternating projection method for solving linear program in the form (19) converges linearly.*

The proposed method is as follows

$$(u_A^{k+1}, v_A^{k+1}) = \mathbb{P}_{Qu=v}(u_B^k, v_B^k), \quad (20a)$$

$$(u_B^{k+1}, v_B^{k+1}) = \mathbb{P}_{\mathcal{C} \times \mathcal{C}^*}(u_A^{k+1}, v_A^{k+1}). \quad (20b)$$

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 (20a) requires solving a linear system involving a matrix $(I + Q^\top Q)$. Step (20a) of the method can be computed as

$$u_A^{k+1} = (I + Q^\top Q)^{-1}(u_B^k - Qv_B^k), \quad (21a)$$

$$v_A^{k+1} = Qu_A^{k+1}. \quad (21b)$$

Since $(I + Q^\top Q)$ 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 [5]. Projection onto a cone in (20b) 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.

V. CONCLUSION

In this paper we provide necessary and sufficient conditions for the linear convergence of an optimization method whose associated fixed-point operator is averaged. The proposed framework is based on properties of such a fixed-point operator. We show that some 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.

REFERENCES

- [1] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Sim.*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [2] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, ser. Springer Optimization and Its Applications. Springer New York, 2011, vol. 49, pp. 185–212.
- [3] S. Richter, C. Jones, and M. Morari, "Computational complexity certification for real-time MPC with input constraints based on the fast gradient method," *IEEE T. Automat. Contr.*, vol. 57, no. 6, pp. 1391–1403, 2012.

- [4] B. O’Donoghue, G. Stathopoulos, and S. Boyd, “A splitting method for optimal control,” *IEEE T. Contr. Syst. T.*, vol. 21, no. 6, pp. 2432–2442, Nov 2013.
- [5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [6] N. Parikh and S. Boyd, “Proximal algorithms,” *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.
- [7] E. Ryu and S. Boyd, “A primer on monotone operator methods,” *Appl. Comput. Math.*, vol. 15, no. 1, 2016.
- [8] S. Richter, “Computational complexity certification of gradient methods for real-time model predictive control,” Ph.D. dissertation, ETH-Zürich, 2012.
- [9] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, “Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems,” *IEEE T. Automat. Contr.*, vol. 60, no. 3, pp. 644–658, 2015.
- [10] P. Giselsson and S. Boyd, “Linear convergence and metric selection in Douglas-Rachford splitting and ADMM,” *IEEE T. Automat. Contr.*, 2017.
- [11] Z.-Q. Luo and P. Tseng, “Error bounds and convergence analysis of feasible descent methods: a general approach,” *Ann. Oper. Res.*, vol. 46–47, no. 1, pp. 157–178, 1993.
- [12] D. Boley, “Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs,” *SIAM J. Optimiz.*, vol. 23, no. 4, pp. 2183–2207, 2013.
- [13] H. M. Phan, “Linear convergence of the Douglas-Rachford method for two closed sets,” *Optimization*, vol. 65, no. 2, pp. 369–385, 2016.
- [14] J. Eckstein and D. Bertsekas, “An alternating direction method for linear programming,” 1990.
- [15] P.-W. Wang and C.-J. Lin, “Iteration complexity of feasible descent methods for convex optimization,” *J. Mach. Learn. Res.*, vol. 15, pp. 1523–1548, 2014.
- [16] 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.
- [17] I. Necoara, “Computational complexity certification for dual gradient method: Application to embedded MPC,” *Syst. Control Lett.*, vol. 81, pp. 49–56, 2015.
- [18] F. Deutsch and H. Hundal, “The rate of convergence of Dykstra’s cyclic projections algorithm: The polyhedral case,” *Numer. Func. Anal. Opt.*, vol. 15, no. 5–6, pp. 537–565, 1994.
- [19] H. Bauschke, J. Bello Cruz, T. Nghia, H. Phan, and X. Wang, “The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle,” *J. Approx. Theory*, vol. 185, pp. 63–79, 2014.
- [20] R. Nishihara, S. Jegelka, and M. I. Jordan, “On the convergence rate of decomposable submodular function minimization,” in *Adv. Neur. In. 27*, 2014, pp. 640–648.
- [21] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 1st ed. Springer Publishing Company, Incorporated, 2011.
- [22] H. H. Bauschke, D. Noll, and H. M. Phan, “Linear and strong convergence of algorithms involving averaged nonexpansive operators,” *J. Math. Anal. Appl.*, vol. 421, no. 1, pp. 1–20, 2015.
- [23] G. Banjac and P. J. Goulart, “Tight global linear convergence rate bounds for operator splitting methods,” available: http://www.optimization-online.org/DB_HTML/2016/03/5364.html, March 2016.
- [24] J. von Neumann, *Functional Operators, Vol. II*. Princeton University Press, 1950.
- [25] H. Bauschke and J. Borwein, “On the convergence of von Neumann’s alternating projection algorithm for two sets,” *Set-Valued Anal.*, vol. 1, no. 2, pp. 185–212, 1993.
- [26] B. O’Donoghue, E. Chu, N. Parikh, and S. Boyd, “Conic optimization via operator splitting and homogeneous self-dual embedding,” *J. Optimiz. Theory Appl.*, pp. 1–27, 2016.
- [27] P. L. Combettes and I. Yamada, “Compositions and convex combinations of averaged nonexpansive operators,” *J. Math. Anal. Appl.*, vol. 425, no. 1, pp. 55–70, 2015.

APPENDIX A PROOF OF THEOREM 1

To show that linear regularity is a necessary property of an operator T for (3) to hold, we combine the generalized triangle

inequality

$$d_{\text{fix } T}(x) \leq d_{\text{fix } T}(Tx) + \|x - Tx\| \quad (22)$$

with (3) to produce

$$(1 - \beta)d_{\text{fix } T}(x) \leq d_{\text{fix } T}(x) - d_{\text{fix } T}(Tx) \leq \|x - Tx\|,$$

which proves that T is linearly regular.

The authors in [22, Lemma 3.8] showed that linear regularity is a sufficient condition for (3) when T is α -averaged. However, we derive this result with a better convergence factor, and thus we repeat some arguments from [22] for the sake of completeness.

Suppose that T is linearly regular with constant κ , that is

$$\frac{1}{\kappa^2} d_{\text{fix } T}^2(x) \leq \|x - Tx\|^2. \quad (23)$$

By taking $y = \mathbb{P}_{\text{fix } T}(x)$, (5) implies

$$\begin{aligned} \|Tx - \mathbb{P}_{\text{fix } T}(Tx)\|^2 + \frac{1-\alpha}{\alpha} \|x - Tx\|^2 \\ \leq \|Tx - \mathbb{P}_{\text{fix } T}(x)\|^2 + \frac{1-\alpha}{\alpha} \|x - Tx\|^2 \\ \leq \|x - \mathbb{P}_{\text{fix } T}(x)\|^2, \end{aligned}$$

where the first inequality exploits properties of the projection operator, i.e. $d_C(z) = \|z - \mathbb{P}_C(z)\| = \inf\{\|z - y\| \mid y \in C\}$. The above inequality can be rewritten as

$$d_{\text{fix } T}^2(Tx) + \frac{1-\alpha}{\alpha} \|x - Tx\|^2 \leq d_{\text{fix } T}^2(x). \quad (24)$$

Combining (23) and (24) implies the following

$$\frac{1}{\kappa^2} d_{\text{fix } T}^2(x) \leq \frac{\alpha}{1-\alpha} (d_{\text{fix } T}^2(x) - d_{\text{fix } T}^2(Tx)), \quad (25)$$

hence

$$d_{\text{fix } T}(Tx) \leq \sqrt{1 - \frac{1-\alpha}{\alpha\kappa^2}} d_{\text{fix } T}(x). \quad (26)$$

Note that from (25) we have

$$\frac{1}{\kappa^2} d_{\text{fix } T}^2(x) \leq \frac{\alpha}{1-\alpha} d_{\text{fix } T}^2(x),$$

which means that $\frac{1}{\kappa^2} \leq \frac{\alpha}{1-\alpha}$. Since $\alpha \in (0, 1)$ and $\kappa > 0$, we have

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

so (26) satisfies condition (3).

APPENDIX B PROOF OF PROPOSITION 1

Proof of (i).

Since T_1 is α_1 -averaged, then

$$T_0 := (1 - \frac{1}{\alpha_1}) \text{Id} + \frac{1}{\alpha_1} T_1 \quad (27)$$

is nonexpansive [27], which itself implies that

$$T_2 := (1 - \frac{\alpha_2}{\alpha_1}) \text{Id} + \frac{\alpha_2}{\alpha_1} T_1 = (1 - \alpha_2) \text{Id} + \alpha_2 T_0 \quad (28)$$

is α_2 -averaged (see Definition 1(ii)). 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 (28) we have $T_2 x = x$, and vice versa. Since T_1 is κ_1 -linearly regular, i.e.

$$d_{\text{fix } T_1}(x) \leq \kappa_1 \|x - T_1 x\|,$$

we have

$$\|x - T_2 x\| = \frac{\alpha_2}{\alpha_1} \|x - T_1 x\| \geq \frac{\alpha_2}{\alpha_1} \frac{1}{\kappa_1} d_{\text{fix } T_2}(x),$$

which proves that T_2 is linearly regular with constant $\frac{\alpha_1}{\alpha_2} \kappa_1$.

Proof of (ii).

The convergence factor of T_2 as a function of α_2 is

$$\beta_2(\alpha_2) = \sqrt{1 - \frac{1-\alpha_2}{\alpha_2\kappa_2^2}} = \sqrt{1 - \frac{(1-\alpha_2)\alpha_2}{\alpha_1^2\kappa_1^2}}.$$

This function is minimized when $(1 - \alpha_2)\alpha_2$ is maximized, which implies that the smallest convergence factor is attained for $\alpha_2 = \frac{1}{2}$.