SIAM J. OPTIM.
Vol. 29, No. 3, pp. 1931–1957

CONVERGENCE RATES FOR PROJECTIVE SPLITTING*

PATRICK R. JOHNSTONE† AND JONATHAN ECKSTEIN†

Abstract. Projective splitting is a family of methods for solving inclusions involving sums of maximal monotone operators. First introduced by Eckstein and Svaiter in 2008, these methods have enjoyed significant innovation in recent years, becoming one of the most flexible operator-splitting frameworks available. While weak convergence of the iterates to a solution has been established, there have been few attempts to study convergence rates of projective splitting. The aim of this paper is to do so under various assumptions. To this end, it makes four main contributions. First, in the context of convex optimization, an $O(1/k)$ ergodic function convergence rate is established. Second, for strongly monotone inclusions, strong convergence is established as well as an ergodic $O(1/\sqrt{k})$ convergence rate for the distance from the iterates to the solution. Third, for inclusions featuring strong monotonicity and cocoercivity, linear convergence is established. We also consider the special case of one operator. In this case we show that projective splitting reduces to either the extragradient method or the proximal-point method, depending on whether forward or backward steps are used.

Key words. splitting algorithms, proximal algorithms, convergence rates

AMS subject classifications. 49M27, 47H05, 47J25

DOI. 10.1137/18M1203523

1. Introduction.

For a real Hilbert space $\mathcal{H}$, consider the problem of finding $z \in \mathcal{H}$ such that

$$0 \in \sum_{i=1}^{n} T_i z, \tag{1.1}$$

where $T_i : \mathcal{H} \to 2^{\mathcal{H}}$ are maximal monotone operators and additionally there exists a subset $\mathcal{I}_{\mathrm{F}} \subseteq \{1, \dots, n\}$ such that for all $i \in \mathcal{I}_{\mathrm{F}}$ the operator T_i is Lipschitz continuous. An important instance of this problem is

$$\min_{z \in \mathcal{H}} F(z), \qquad \text{where} \qquad F(z) \triangleq \sum_{i=1}^{n} f_i(z) \tag{1.2}$$

and every $f_i : \mathcal{H} \to (-\infty, +\infty]$ is closed, proper, and convex, with some subset of the functions also being Fréchet differentiable with Lipschitz-continuous gradients. Under appropriate constraint qualifications, (1.1) and (1.2) are equivalent. Problem (1.2) arises in a host of applications such as machine learning, signal and image processing, inverse problems, and computer vision; see [3, 5, 6] for some examples.

A relatively recently proposed class of operator-splitting algorithms that can solve (1.1), among other problems, is *projective splitting*. It originated with [11] and was then generalized to more than two operators in [12]. The related algorithm in [1] introduced a technique for handling compositions of linear and monotone operators, and

*Received by the editors July 27, 2018; accepted for publication (in revised form) May 8, 2019; published electronically July 23, 2019.
https://doi.org/10.1137/18M1203523
Funding: This material is based upon work supported by the National Science Foundation under grant CCF-1617617.
†Management Sciences and Information Systems, Rutgers Business School, Rutgers University, 100 Rockafeller Road, Piscataway, NJ 08854 (patrick.r.johnstone@gmail.com, jeckstei@business.rutgers.edu).

an extension to "block-iterative" and asynchronous operation was proposed in [4]—block-iterative operation meaning that only a subset of the operators making up the problem need to be considered at each iteration (this approach may be called "incremental" in the optimization literature). A restricted and simplified version of this framework appears in [9]. The recent work [15] incorporated forward steps into the projective splitting framework for any Lipschitz continuous operators and introduced backtracking and adaptive step-size rules (the set $\mathcal{I}_{\rm F}$ is named for "forward steps").

In general, projective splitting offers unprecedented levels of flexibility compared with previous operator-splitting algorithms (e.g., [20, 17, 24, 8]). The framework can be applied to arbitrary sums of maximal monotone operators, the step sizes can vary by operator and by iteration, compositions with linear operators can be handled, and block-iterative asynchronous implementations have been demonstrated.

In previous works on projective splitting, the main theoretical goal was to establish the weak convergence of the iterates to a solution of the monotone inclusion under study (either a special case or a generalization of (1.1)). This goal was achieved using Fejér-monotonicity arguments in coordination with the unique properties of projective splitting. The question of convergence rates has not been addressed, except in [18], which considers a different type of convergence rate than those investigated here; we discuss the differences between our analysis and that of [18] in more detail below.

Contributions. To this end, there are four main novel contributions in this paper.

1. For (1.2), we establish an ergodic $\mathrm{O}(1/k)$ function value convergence rate for iterates generated by projective splitting.
2. When one of the operators in (1.1) is strongly monotone, we establish strong, rather than weak, convergence in the general Hilbert space setting, without using the Haugazeau [13] modification employed to obtain general strong convergence in [4]. Furthermore, we derive an ergodic $\mathrm{O}(1/\sqrt{k})$ convergence rate for the distance from the iterates to the unique solution of (1.1).
3. If, additionally, $T_1, \ldots, T_{n-1}$ are cocoercive, we establish *linear* convergence to 0 of the distance from the iterates to the unique solution. If $n = 1$ (i.e. there is only one operator present), the linear convergence result only requires strong monotonicity of T_1.
4. We discuss the special cases of projective splitting when $n = 1$. Interestingly, projective splitting reduces to one of two well-known algorithms, depending on whether forward or backward steps are used. This observation has implications for the convergence rate analysis.

The primary engine of the analysis is a new summability lemma (see Lemma 6.2) in which important quantities of the algorithm are shown to be summable. This summability is directly exploited in the ergodic function value rate analysis in section 7. In section 9, the same lemma is used to show linear convergence when strong monotonicity and cocoercivity are present. With only strong monotonicity present, we also obtain strong convergence and rates using a novel analysis in section 8.

Our convergence rates apply directly to the variants of projective splitting discussed in [4, 11, 15]. The papers [9, 12] use a slightly different separating hyperplane formulation than ours but the difference is easy to resolve. However, our analysis does not allow for the asynchrony or block-iterative effects developed in [4, 9, 15]. In particular, at each iteration we assume that every operator is processed and that the computations use the most up-to-date information. Developing a convergence rate analysis that extends to asynchronous and block-iterative operation in a satisfactory manner is a matter for future work.

In [4, 9, 15], projective splitting was extended to handle the composition of each T_i with a linear operator. While it is possible to extend all of our convergence rate results to allow for this generalization under appropriate conditions, for readability we will not do so here.

In section 4 we consider the case when $n = 1$. In this case, we show that projective splitting reduces to the proximal point method [22] if one uses backward steps, or to a special case of the extragradient method (with no constraint) [16, 21] when one uses forward steps. Since projective splitting includes the proximal point method as a special case, the $\mathrm{O}(1/k)$ function value convergence rate derived in section 7 cannot be improved, since this is the best rate available for the proximal point method, as established in [7, Theorem 12]. The same is true for the linear convergence result in section 9, since the proximal point method cannot do better than this rate under strong monotonicity [2, Proposition 26.16].

The specific outline of the paper is as follows. Immediately below, we discuss in more detail the convergence rate analysis for projective splitting conducted in [18] and how it differs from our analysis. Section 2 presents notation, basic mathematical results, and assumptions. Section 3 introduces the projective splitting framework under study, along with some associated assumptions. Section 4 discusses special cases of projective splitting when $n = 1$. Section 5 recalls some important lemmas from [15]. Section 6 proves some new lemmas necessary for convergence rate results, including the key summability lemma (Lemma 6.2). Section 7 derives the ergodic $\mathrm{O}(1/k)$ function value convergence rate for (1.2). Section 8 derives strong convergence and convergence rates under strong monotonicity. Finally, section 9 establishes linear convergence under strong monotonicity and cocoercivity.

Comparison with [18, 19]. To the best of our knowledge, the only works attempting to quantify convergence rates of projective splitting are [18, 19], both by the same author. The analysis in [19] concerns a dual application of projective splitting and its convergence rate results are similar to those in [18]. In these works, convergence rates are not defined in the more customary way as they are in this paper, i.e., in terms of either the distance to the solution or the gap between the current function values and the optimal value of (1.2). Instead, in [18], they are defined in terms of an approximate solution criterion for the monotone inclusion under study, specifically (1.1) with $n = 2$. Without any enlargement being applied to the operators, the approximate solution condition is as follows: a point $(x, y) \in \mathcal{H}^2$ is said to be an ϵ-approximate solution of (1.1) with $n = 2$ if there exists $(a, b) \in \mathcal{H}^2$ s.t. $a \in T_1 x$, $b \in T_2 y$, and $\max\{\|a+b\|, \|x-y\|\} \leq \epsilon$. If this condition holds with $\epsilon = 0$, then $x = y$ is a solution to (1.1). The iteration complexity of a method is then defined in terms of the number of iterations required to produce a point (x^k, y^k) that is a ϵ-approximate solution in this sense.

This notion of iteration complexity/convergence rate is different than the one we use here: for the special case of (1.2), we determine the convergence rate of $F(z^k) - F^*$ to zero, where F^* is the optimal value and z^k is an appropriate sequence generated by the algorithm. Also, under an additional strong monotonicity assumption that was not made in [18], we determine the convergence rate of the distance from the points generated by the algorithm to a solution of (1.1). That is, we determine how fast $\|z^k - z^*\|$ converges to 0, where z^* solves (1.1).

In summary, we derive results of a more standard nature than [18, 19], but rely on stronger assumptions.

2. Mathematical preliminaries.

2.1. Notation, terminology, and basic lemmas. Throughout this paper we will use the standard convention that a sum over an empty set of indices, such as $\sum_{i=1}^{n-1} a_i$ with $n = 1$, is taken to be 0. A single-valued operator $A : \mathcal{H} \to \mathcal{H}$ is called *cocoercive* if, for some $\Gamma > 0$,

$$(\forall\, u, v \in \mathcal{H}) \qquad \langle u - v, A(u) - A(v)\rangle \geq \Gamma^{-1}\|A(u) - A(v)\|^2.$$

Every cocoercive operator is Γ-Lipschitz continuous, but the converse does not hold. For any maximal monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ and scalar $\rho > 0$ we will use the notation $J_{\rho A} \triangleq (I + \rho A)^{-1}$ to denote the *resolvent*, also known as the backward step, implicit step, or proximal operator with respect to A. In particular,

$$x = J_{\rho A}(a) \quad \Longrightarrow \quad \exists y \in Ax : x + \rho y = a, \tag{2.1}$$

and the x and y satisfying this relation are unique. Furthermore, $J_{\rho A}$ is defined everywhere and $\operatorname{range}(J_A) = \operatorname{dom}(A)$ [2, Proposition 23.2]. In the special case where $A = \partial f$ for some convex, closed, and proper function f, the proximal operator may be written as

$$J_{\rho\partial f}(a) = \arg\min_x \left\{\frac{1}{2}\|x - a\|^2 + \rho f(x)\right\}. \tag{2.2}$$

We also adopt the "Prox" notation introduced in [9]: for any maximal monotone operator T, vector a, and positive scalar ρ, we let $(x, y) = \mathrm{Prox}_A^\rho(a)$ denote the unique $(x, y) \in \operatorname{gra} A$ such that $x + \rho y = a$. In other words,

$$(x, y) = \mathrm{Prox}_A^\rho(a) \quad \Longleftrightarrow \quad x = J_{\rho A}(a) \text{ and } y = \rho^{-1}(a - x). \tag{2.3}$$

Finally, we will use the following two standard results.

LEMMA 2.1. *For any vectors* $v_1, \ldots, v_n \in \mathcal{H}$, $\|\sum_{i=1}^n v_i\|^2 \leq n \sum_{i=1}^n \|v_i\|^2$.

LEMMA 2.2. *For any* $x, y, z \in \mathcal{H}$,

$$2\langle x - y, x - z\rangle = \|x - y\|^2 + \|x - z\|^2 - \|y - z\|^2. \tag{2.4}$$

We will use a boldface $\mathbf{w} \triangleq (w_1, \ldots, w_{n-1})$ for elements of $\mathcal{H}^{n-1}$.

2.2. The main assumptions regarding problem (1.1). Define the *extended solution set* or *Kuhn–Tucker set* of (1.1) to be

$$\mathcal{S} \triangleq \left\{(z, w_1, \ldots, w_{n-1}) \in \mathcal{H}^n \;\middle|\; w_i \in T_i z,\ i = 1, \ldots, n-1,\ \ -\sum_{i=1}^{n-1} w_i \in T_n z\right\}. \tag{2.5}$$

Clearly $z \in \mathcal{H}$ solves (1.1) if and only if there exists $\mathbf{w} \in \mathcal{H}^{n-1}$ such that $(z, \mathbf{w}) \in \mathcal{S}$. Our main assumptions regarding (1.1) are as follows.

Assumption 2.3. $\mathcal{H}$ is a real Hilbert space and problem (1.1) conforms to the following:

1. for $i = 1, \dots, n$, the operators $T_i : \mathcal{H} \to 2^{\mathcal{H}}$ are monotone;
2. for all i in some possibly empty subset $\mathcal{I}_{\mathrm{F}} \subseteq \{1, \dots, n\}$, the operator T_i is L_i-Lipschitz continuous (and thus single valued) and $\mathrm{dom}(T_i) = \mathcal{H}$;
3. for all $i \in \mathcal{I}_{\mathrm{B}} \triangleq \{1, \dots, n\} \setminus \mathcal{I}_{\mathrm{F}}$, the operator T_i is maximal and the map $\mathrm{Prox}^{\rho}_{T_i} : \mathcal{H} \to \mathcal{H}^2$ can be computed to within the error tolerance specified in Assumption 3.2 (however, these operators are not precluded from also being Lipschitz continuous);
4. problem (1.1) has a solution, and therefore the solution set $\mathcal{S}$ as defined in (2.5) is nonempty.

We use $\mathcal{I}_{\mathrm{B}}$ for the set $\{1, \dots, n\} \setminus \mathcal{I}_{\mathrm{F}}$ because the operators in this set are processed by *backward steps* (resolvent operations) in Algorithm 3.1.

PROPOSITION 2.4 (Johnstone and Eckstein [15, Lemma 3]). *Under Assumption* 2.3*,* $\mathcal{S}$ *from* (2.5) *is closed and convex.*

3. The algorithm. Projective splitting is a special case of a general separator-projector method for finding a point in a closed and convex set. At each iteration the method constructs an affine function $\varphi_k : \mathcal{H}^n \to \mathbb{R}$ that separates the current point from the target set $\mathcal{S}$ defined in (2.5). In other words, if p^k is the current point in $\mathcal{H}^n$ generated by the algorithm, $\varphi_k(p^k) > 0$ and $\varphi_k(p) \le 0$ for all $p \in \mathcal{S}$. The next point is then the projection of p^k onto the hyperplane $\{p : \varphi_k(p) = 0\}$, subject to a relaxation factor β_k. What makes projective splitting an operator-splitting method is that the hyperplane is constructed through individual calculations on each operator T_i, either resolvent calculations or forward steps.

3.1. The hyperplane. Let $p \triangleq (z, \mathbf{w}) = (z, w_1, \dots, w_{n-1})$ be a generic point in $\mathcal{H}^n$. For $\mathcal{H}^n$, we adopt the following norm and inner product for some $\gamma > 0$:

$$\|(z, \mathbf{w})\|^2 \triangleq \gamma \|z\|^2 + \sum_{i=1}^{n-1} \|w_i\|^2 \; \left\langle (z^1, \mathbf{w}^1), (z^2, \mathbf{w}^2) \right\rangle \triangleq \gamma \langle z^1, z^2 \rangle + \sum_{i=1}^{n-1} \langle w_i^1, w_i^2 \rangle. \tag{3.1}$$

Define the following function for all $k \ge 1$ and $p \in \mathcal{H}^n$:

$$\varphi_k(p) \triangleq \sum_{i=1}^{n-1} \langle z - x_i^k, y_i^k - w_i \rangle + \left\langle z - x_n^k, y_n^k + \sum_{i=1}^{n-1} w_i \right\rangle, \tag{3.2}$$

where the (x_i^k, y_i^k) are some given points satisfying $y_i^k \in T_i x_i^k$ for $i = 1, \dots, n$. For Lemma 3.1 below, it is sufficient to take (x_i^k, y_i^k) to be an arbitrary pair in gra T_i for $i = 1, \dots, n$. In the next subsection, we will specify how to choose $\{(x_i^k, y_i^k)\}_{i=1}^n$ so as to construct a separating hyperplane between the current point and the solution set $\mathcal{S}$. The function φ_k is a special case of the separator function used in [4]. The following lemma proves some basic properties of φ_k; similar results are given in [1, 4, 9] in the case when $\gamma = 1$.

LEMMA 3.1 (Johnstone and Eckstein [15, Lemma 4]). *Let* φ_k *be defined as in* (3.2)*. Then we have the following.*

1. φ_k *is affine on* $\mathcal{H}^n$*. Thus, its gradient is constant:* $\nabla \varphi_k(p_1) = \nabla \varphi_k(p_2) \triangleq \nabla \varphi_k$ *for all* $p_1, p_2 \in \mathcal{H}^n$*.*

2. *With respect to inner product* (3.1) *defined on* $\mathcal{H}^n$, *the gradient of* φ_k *is*

$$\nabla\varphi_k = \left(\frac{1}{\gamma}\left(\sum_{i=1}^{n} y_i^k\right),\ x_1^k - x_n^k, x_2^k - x_n^k, \ldots, x_{n-1}^k - x_n^k\right). \tag{3.3}$$

3. *Suppose that the operators* $T_i : \mathcal{H} \to 2^{\mathcal{H}}$ *are monotone and* $y_i^k \in T_i x_i^k$ *for* $i = 1, \ldots, n$. *Then* $\varphi_k(p) \leq 0$ *for all* $p \in \mathcal{S}$ *defined in* (2.5).
4. *Suppose* $y_i^k \in T_i x_i^k$ *for* $i = 1, \ldots, n$ *and* $\nabla\varphi_k = 0$. *Then* $(x_n^k, y_1^k, \ldots, y_{n-1}^k) \in \mathcal{S}$ *and* x_n^k *is a solution of problem* (1.1).

3.2. Projective splitting. Algorithm 3.1 is the projective splitting framework for which we will derive convergence rates. It is a special case of the framework of [15] without asynchrony or block-iterative features. In particular, we assume at each iteration that the method processes every operator T_i using the most up-to-date information possible. The frameworks of [4, 9, 15] also allow for the monotone operators to be composed with linear operators. As mentioned in the introduction, our analysis may be extended to this situation, but for readability we will not do so here.

Algorithm 3.1 is a special case of the separator-projector algorithm applied to finding a point in $\mathcal{S}$ using the affine function φ_k defined in (3.2) (see [15, Lemma 6]). For the variables of Algorithm 3.1 defined on lines 19 and 20, we define $p^k \triangleq (z^k, \mathbf{w}^k) \triangleq (z^k, w_1^k, \ldots, w_{n-1}^k)$ for all $k \geq 1$. The points $(x_i^k, y_i^k) \in \operatorname{gra} T_i$ are chosen so that $\varphi_k(p^k)$ is sufficiently large to guarantee the weak convergence of p^k to a solution [15, Theorem 1]. For $i \in \mathcal{I}_{\mathrm{B}}$, a single (possibly approximate) resolvent calculation is required, whereas for $i \in \mathcal{I}_{\mathrm{F}}$ two forward steps are required per iteration. Following (2.3), the calculation in line 4 is equivalent to finding (x_i^k, y_i^k) that satisfies $y_i^k \in T_i(x_i^k)$ and $x_i^k + \rho_i^k y_i^k \approx z^k + \rho_i^k w_i^k$, with a residual of e_i^k in the second condition.

The algorithm has the following parameters:

- for each $k \geq 1$ and $i = 1, \ldots, n$, a positive scalar step size ρ_i^k;
- for each $k \geq 1$, an overrelaxation parameter $\beta_k \in [\underline{\beta}, \overline{\beta}]$, where $0 < \underline{\beta} \leq \overline{\beta} < 2$;
- the fixed scalar $\gamma > 0$ from (3.1), which controls the relative emphasis on the primal and dual variables in the projection update in lines 19 and 20;
- sequences of errors $\{e_i^k\}_{k \geq 1}$ for $i \in \mathcal{I}_{\mathrm{B}}$, modeling inexact computation of the resolvent steps.

To ease the mathematical presentation, we use the following notation in Algorithm 3.1 and throughout the rest of the paper:

$$(\forall\, k \in \mathbb{N}) \qquad w_n^k \triangleq -\sum_{i=1}^{n-1} w_i^k. \tag{3.4}$$

Note that when $n = 1$ we have $w_n^k = 0$ by the convention at the start of section 2.1.

3.3. Conditions on the errors and the step sizes. We now state our assumptions regarding the computational errors and step sizes in Algorithm 3.1. Inequalities (3.5) and (3.6) are taken verbatim from [9]. Assumption (3.7) is new and necessary to derive our convergence rate results. Throughout the rest of the paper, let $\bar{K}$ be the iteration where Algorithm 3.1 terminates via line 17, with $\bar{K} = \infty$ if the method runs indefinitely.

Algorithm 3.1 Algorithm for solving (1.1).

Require: $(z^1, \mathbf{w}^1) \in \mathcal{H}^n$, $\gamma > 0$.
for $k = 1, 2, \ldots$ **do**
  for $i = 1, 2, \ldots, n$ **do**
    if $i \in \mathcal{I}_{\mathrm{B}}$ **then**
      $(x_i^k, y_i^k) = \mathrm{Prox}_{T_i}^{\rho_i^k}(z^k + \rho_i^k w_i^k + e_i^k)$
    else
      $x_i^k = z^k - \rho_i^k(T_i z^k - w_i^k)$,
      $y_i^k = T_i x_i^k$;
    end if
  end for
  $u_i^k = x_i^k - x_n^k, \quad i = 1, \ldots, n-1$,
  $v^k = \sum_{i=1}^n y_i^k$,
  $\pi_k = \|u^k\|^2 + \gamma^{-1}\|v^k\|^2$
  if $\pi_k > 0$ **then**
    $\varphi_k(p_k) = \langle z^k, v^k\rangle + \sum_{i=1}^{n-1}\langle w_i^k, u_i^k\rangle - \sum_{i=1}^n \langle x_i^k, y_i^k\rangle$,
    $\alpha_k = \beta_k \varphi_k(p_k)/\pi_k$
  else
    return $z^{k+1} \leftarrow x_n^k, w_1^{k+1} \leftarrow y_1^k, \ldots, w_{n-1}^{k+1} \leftarrow y_{n-1}^k$
  end if
  $z^{k+1} = z^k - \gamma^{-1}\alpha_k v^k$,
  $w_i^{k+1} = w_i^k - \alpha_k u_i^k, \quad i = 1, \ldots, n-1$,
  $w_n^{k+1} = -\sum_{i=1}^{n-1} w_i^{k+1}$;
end for

Assumption 3.2. For some $\sigma \in [0,1[$ and $\delta \geq 0$, the following hold for all $1 \leq k \leq \bar{K}$ and $i \in \mathcal{I}_{\mathrm{B}}$:

$$\langle z^k - x_i^k, e_i^k\rangle \geq -\sigma\|z^k - x_i^k\|^2, \tag{3.5}$$

$$\langle e_i^k, y_i^k - w_i^k\rangle \leq \rho_i^k \sigma \|y_i^k - w_i^k\|^2, \tag{3.6}$$

$$\|e_i^k\|^2 \leq \delta\|z^k - x_i^k\|^2. \tag{3.7}$$

A sufficient condition to imply all three of these inequalities is

$$\|e_i^k\| \leq \min\left\{\sigma\rho_i^k\|y_i^k - w_i^k\|, \min\{\sqrt{\delta}, \sigma\}\|z^k - x_i^k\|\right\}. \tag{3.8}$$

We also remark that, when $\sqrt{\delta} \leq \sigma$, condition (3.7) implies (3.5), so (3.5) could be dropped in such cases. For each $i \in \mathcal{I}_{\mathrm{B}}$, we will show in (6.4) that the sequence $\{\|z^k - x_i^k\|\}$ is square-summable, so an eventual consequence of (3.7) will be that the corresponding error sequence must be square-summable, that is,

$$\sum_{k=1}^{\bar{K}} \|e_i^k\|^2 < \infty.$$

Assumption 3.3. The step sizes satisfy

$$\underline{\rho} \triangleq \min_{i=1,\ldots,n} \left\{ \inf_{1\leq k\leq \bar{K}} \rho_i^k \right\} > 0, \tag{3.9}$$

$$(\forall i \in \mathcal{I}_\mathrm{B}) \quad \overline{\rho}_i \triangleq \sup_{1\leq k\leq \bar{K}} \rho_i^k < \infty, \tag{3.10}$$

$$(\forall\, i \in \mathcal{I}_\mathrm{F}) \quad \overline{\rho}_i \triangleq \sup_{1\leq k\leq \bar{K}} \rho_i^k < \frac{1}{L_i}. \tag{3.11}$$

From here on, we let $\overline{\rho} \triangleq \max_{i\in\mathcal{I}_\mathrm{B}} \overline{\rho}_i$ and $\bar{L} \triangleq \max_{i\in\mathcal{I}_\mathrm{F}} L_i$ with the convention that $\overline{\rho} = 0$ if $\mathcal{I}_\mathrm{B} = \{\emptyset\}$ and $\bar{L} = 0$ if $\mathcal{I}_\mathrm{F} = \{\emptyset\}$.

The recent work [15] includes several extensions to the basic forward-step step-size constraint (3.11), under which one still obtains weak convergence of the iterates to a solution. Section 4.1 of [15] presents a backtracking line search that may be used when the constant L_i is unknown. Section 4.2 of [15] presents a backtrack-free adaptive step size for the special case in which T_i is affine but L_i is unknown. Our convergence rate analysis holds for these extensions, but for readability we omit the details.

4. Simplifications when $n = 1$.

Suppose $n = 1$, in which case we have either $1 \in \mathcal{I}_\mathrm{B}$ or $1 \in \mathcal{I}_\mathrm{F}$. In either case, using the convention discussed at the beginning of section 2.1, we have $w_1^k = 0$ for all k and the affine function defined in (3.2) becomes

$$\varphi_k(p) = \langle z - x^k, y^k\rangle,$$

where we have dropped the unnecessary subscript and written $x_1^k = x^k$ and $y_1^k = y^k$.

If $1 \in \mathcal{I}_\mathrm{B}$, suppose $e_1^k = 0$ for all $k \geq 1$ so the resolvent is calculated exactly on line 4 of Algorithm 3.1. Using $w_1^k = 0$ yields, for all $k \geq 1$,

$$x^k + \rho^k y^k = z^k \qquad \text{and} \qquad y^k \in Tx^k,$$

where we have written $T_1 = T$ and $\rho_1^k = \rho^k$. This implies that $\varphi_k(p^k) = \rho^k\|y^k\|^2$. Furthermore, $\nabla_z\varphi_k = \gamma^{-1}y^k$ and so $\|\nabla\varphi_k\|^2 = \|\nabla_z\varphi_k\|^2 = \gamma\cdot\gamma^{-2}\|y^k\|^2 = \gamma^{-1}\|y^k\|^2$. Therefore, using (5.1), we have, for all $k \geq 1$,

$$\begin{aligned} z^{k+1} &= z^k - \frac{\beta_k\varphi_k(p^k)}{\|\nabla\varphi_k\|^2}\nabla_z\varphi_k \\ &= z^k - \beta_k\rho^k y^k \\ &= (1-\beta_k)z^k + \beta_k x^k \\ &= (1-\beta_k)z^k + \beta_k J_{\rho^k T}(z^k). \end{aligned}$$

Thus, when $n = 1$, $1 \in \mathcal{I}_\mathrm{B}$, and $e_1^k = 0$ for all $k \geq 1$, projective splitting reduces to the relaxed proximal point method of [10]; see also [2, Theorem 23.41]. In fact, when one allows for approximate evaluation of resolvents, projective splitting with $n = 1 \in \mathcal{I}_\mathrm{B}$ reduces to the hybrid projection proximal point method of Solodov and Svaiter [23, Algorithm 1.1]. However, the error criterion of [23, equation (1.1)] is more restrictive than the conditions (3.5), (3.6) that we use in Assumption 3.2.

On the other hand, if $1 \in \mathcal{I}_\mathrm{F}$, considering lines 6–7 of the algorithm with $w_1^k = 0$ yields for all $k \geq 1$

$$x^k = z^k - \rho^k Tz^k \qquad \text{and} \qquad y^k = Tx^k.$$

Furthermore, we have

$$\varphi_k(p^k) = \rho^k \langle Tz^k, Tx^k \rangle \qquad \text{and} \qquad \nabla_z \varphi_k = \gamma^{-1} Tx^k.$$

Therefore, for all $k \geq 1$,

$$z^{k+1} = z^k - \frac{\beta_k \varphi_k(p^k)}{\|\nabla \varphi_k\|^2} \nabla_z \varphi_k = z^k - \frac{\beta_k \rho^k \langle Tz^k, Tx^k \rangle}{\|Tx^k\|^2} Tx^k.$$

Thus, the method reduces to

$$x^k = z^k - \rho^k T z^k,$$
$$z^{k+1} = z^k - \tilde{\rho}^k T x^k,$$

where

$$\tilde{\rho}^k \triangleq \frac{\beta_k \rho^k \langle Tz^k, Tx^k \rangle}{\|Tx^k\|^2}. \tag{4.1}$$

This is the unconstrained version of the extragradient method [16]. When $\beta_k = 1$, the step size (4.1) corresponds to the extragradient step size proposed by Iusem and Svaiter in [14]. Furthermore, in the unconstrained case, the backtracking line search for the extragradient method also proposed in [14] is almost equivalent to the line search we proposed for processing individual operators within projective splitting in [15], except that Iusem and Svaiter use a more restrictive termination condition (necessary perhaps because they also account for the constrained case of the extragradient method).

While these observations may be of interest in their own right, they also have implications for the convergence rate analysis below; in particular, that projective splitting reduces to the proximal point method suggests the $\mathrm{O}(1/k)$ ergodic convergence rate for (1.2) derived in section 7 cannot be improved beyond a constant factor. This is because the same rate is unimprovable for the proximal point method [7, Theorem 12]. The same is true for the linear convergence rate, since it is known that the proximal point method converges linearly under strong monotonicity [2, Proposition 26.16].

5. Results similar to [15]. In the results to follow in the remainder of the paper, $p^k \triangleq (z^k, w_1^k, \ldots, w_{n-1}^k)$ and $\{(x_i^k, y_i^k)\}_{i=1}^n$ are the iterates generated by Algorithm 3.1.

LEMMA 5.1. *Suppose Assumption* 2.3 *holds. For all* $1 \leq k < \bar{K}$,

1. *the iterates can be written as*

$$p^{k+1} = p^k - \alpha_k \nabla \varphi_k(p^k), \tag{5.1}$$

where $\alpha_k = \beta_k \varphi_k(p^k)/\|\nabla \varphi_k\|^2$ *and* φ_k *is defined in* (3.2),

2. *for all* $p^* \triangleq (z^*, w_1^*, \ldots, w_{n-1}^*) \in \mathcal{S}$,

$$\|p^{k+1} - p^*\|^2 \leq \|p^k - p^*\|^2 - \beta_k(2 - \beta_k)\|p^{k+1} - p^k\|^2, \tag{5.2}$$

which implies that

$$\sum_{t=1}^{k} \|p^t - p^{t+1}\|^2 \leq \tau \|p^1 - p^*\|^2, \qquad \text{where } \tau \triangleq \underline{\beta}^{-1}(2 - \overline{\beta})^{-1}, \tag{5.3}$$

$$\|z^k - z^*\| \leq \gamma^{-\frac{1}{2}} \|p^k - p^*\| \leq \gamma^{-\frac{1}{2}} \|p^1 - p^*\|, \tag{5.4}$$

$$\|w_i^k - w_i^*\| \leq \|p^k - p^*\| \leq \|p^1 - p^*\|, \qquad i = 1, \ldots, n-1. \tag{5.5}$$

Proof. The update (5.1) follows from algebraic manipulation of lines 10–21 and consideration of (3.2) and (3.3). Inequalities (5.2) and (5.3) result from Algorithm 3.1 being a separator-projector algorithm [15, Lemma 6]. A specific reference proving these results is [11, Proposition 1]. ∎

The following lemma places an upper bound on the gradient of the affine function φ_k at each iteration. A similar result was proved in [15, Lemma 11], but since that result is slightly different, we include the full proof here.

LEMMA 5.2. *Suppose Assumptions* 2.3, 3.2, *and* 3.3 *hold and recall the affine function* φ_k *defined in* (3.2). *For all* $1 \le k \le \bar{K}$,

$$\|\nabla\varphi_k\|^2 \le \xi_1 \sum_{i=1}^{n} \|z^k - x_i^k\|^2,$$

where

$$\xi_1 \triangleq 2n\left[1 + 2\gamma^{-1}\left(\bar{L}^2|\mathcal{I}_{\mathrm{F}}| + \underline{\rho}^{-2}(1+\delta)\right)\right] < \infty. \tag{5.6}$$

Proof. Using Lemma 3.1,

$$\|\nabla\varphi_k\|^2 = \gamma^{-1}\left\|\sum_{i=1}^{n} y_i^k\right\|^2 + \sum_{i=1}^{n-1}\|x_i^k - x_n^k\|^2. \tag{5.7}$$

Using Lemma 2.1, we begin by writing the second term on the right of (5.7) as

$$\sum_{i=1}^{n-1}\|x_i^k - x_n^k\|^2 \le 2\sum_{i=1}^{n-1}(\|x_i^k - z^k\|^2 + \|z^k - x_n^k\|^2) \le 2n\sum_{i=1}^{n}\|z^k - x_i^k\|^2. \tag{5.8}$$

We next consider the first term in (5.7). Rearranging the update equations as given in lines 4 and 6 of Algorithm 3.1, we may write

$$y_i^k = (\rho_i^k)^{-1}(z^k - x_i^k + \rho_i^k w_i^k + e_i^k), \qquad i \in \mathcal{I}_{\mathrm{B}}, \tag{5.9}$$

$$T_i z^k = (\rho_i^k)^{-1}(z^k - x_i^k + \rho_i^k w_i^k), \qquad i \in \mathcal{I}_{\mathrm{F}}. \tag{5.10}$$

Note that (5.9) rewrites line 4 of Algorithm 3.1 using (2.3). The first term on the right-hand side of (5.7) may then be written as

$$\begin{aligned}
\left\|\sum_{i=1}^{n} y_i^k\right\|^2 &= \left\|\sum_{i\in\mathcal{I}_{\mathrm{B}}} y_i^k + \sum_{i\in\mathcal{I}_{\mathrm{F}}}(T_i z^k + y_i^k - T_i z^k)\right\|^2 \\
&\overset{\text{(a)}}{\le} 2\left\|\sum_{i\in\mathcal{I}_{\mathrm{B}}} y_i^k + \sum_{i\in\mathcal{I}_{\mathrm{F}}} T_i z^k\right\|^2 + 2\left\|\sum_{i\in\mathcal{I}_{\mathrm{F}}}\left(y_i^k - T_i z^k\right)\right\|^2 \\
&\overset{\text{(b)}}{=} 2\left\|\sum_{i=1}^{n}\left(\rho_i^k\right)^{-1}\left(z^k - x_i^k + \rho_i^k w_i^k\right) + \sum_{i\in\mathcal{I}_{\mathrm{B}}}\left(\rho_i^k\right)^{-1} e_i^k\right\|^2 \\
&\qquad + 2\left\|\sum_{i\in\mathcal{I}_{\mathrm{F}}}\left(T_i x_i^k - T_i z^k\right)\right\|^2
\end{aligned}$$

$$\overset{(c)}{\leq} 4\left\|\sum_{i=1}^{n}\left(\rho_i^k\right)^{-1}\left(z^k - x_i^k + \rho_i^k w_i^k\right)\right\|^2 + 4\left\|\sum_{i\in\mathcal{I}_\mathrm{B}}\left(\rho_i^k\right)^{-1} e_i^k\right\|^2$$
$$+ 2|\mathcal{I}_\mathrm{F}|\sum_{i\in\mathcal{I}_\mathrm{F}}\left\|T_i x_i^k - T_i z^k\right\|^2$$
$$\overset{(d)}{\leq} 4n\underline{\rho}^{-2}\left(\sum_{i=1}^{n}\left\|z^k - x_i^k\right\|^2 + \sum_{i\in\mathcal{I}_\mathrm{B}}\|e_i^k\|^2\right) + 2|\mathcal{I}_\mathrm{F}|\sum_{i\in\mathcal{I}_\mathrm{F}}(L_i^2\|x_i^k - z^k\|^2)$$
$$\overset{(e)}{\leq} \xi_1'\sum_{i=1}^{n}\|z^k - x_i^k\|^2, \tag{5.11}$$

where

$$\xi_1' \triangleq 4n(\bar{L}^2|\mathcal{I}_\mathrm{F}| + \underline{\rho}^{-2}(1+\delta)),$$

where we recall that $\bar{L} = \max_{i\in\mathcal{I}_\mathrm{F}} L_i$. In the above, (a) uses Lemma 2.1, while (b) is obtained by substituting expressions (5.9) and (5.10) into the first squared norm and using $y_i^k = T_i x_i^k$ for $i \in \mathcal{I}_\mathrm{F}$ in the second. Next, (c) uses Lemma 2.1 once more on both terms. Inequality (d) uses Lemma 2.1, the Lipschitz continuity of T_i, $\sum_{i=1}^n w_i^k = 0$, and Assumption 3.3. Finally, (e) follows by collecting terms and using Assumption 3.2. Combining (5.7), (5.8), and (5.11) establishes the lemma with ξ_1 as defined in (5.6). ∎

LEMMA 5.3. *Suppose that assumptions* 2.3, 3.2, *and* 3.3 *hold. Then, for all* $1 \leq k \leq \bar{K}$,

$$\varphi_k(p^k) \geq \xi_2\sum_{i=1}^{n}\|z^k - x_i^k\|^2,$$

where

$$\xi_2 \triangleq \min\left\{(1-\sigma)\overline{\rho}^{-1},\ \min_{j\in\mathcal{I}_\mathrm{F}}\{\overline{\rho}_j^{-1} - L_j\}\right\} > 0. \tag{5.12}$$

Furthermore, for all such k,

$$\varphi_k(p^k) + \sum_{i\in\mathcal{I}_\mathrm{F}} L_i\|z^k - x_i^k\|^2 \geq (1-\sigma)\underline{\rho}\sum_{i\in\mathcal{I}_\mathrm{B}}\|y_i^k - w_i^k\|^2 + \underline{\rho}\sum_{i\in\mathcal{I}_\mathrm{F}}\|T_i z^k - w_i^k\|^2. \tag{5.13}$$

Proof. The claimed results are special cases of those given in [15, Lemmas 12 and 13]. ∎

6. New lemmas needed to derive convergence rates.

LEMMA 6.1. *Suppose Assumptions* 2.3, 3.2, *and* 3.3 *hold, and recall* α_k *computed on line* 15 *of Algorithm* 3.1. *For all* $1 \leq k < \bar{K}$, *it holds that* $\alpha_k \geq \underline{\alpha} \triangleq \underline{\beta}\xi_2/\xi_1 > 0$, *where* ξ_1 *and* ξ_2 *are as defined in* (5.6) *and* (5.12).

Proof. By Lemma 5.1, α_k defined on line 15 of Algorithm 3.1 may be expressed as

$$\alpha_k = \frac{\beta_k\varphi_k(z^k, \mathbf{w}^k)}{\|\nabla\varphi_k\|^2}. \tag{6.1}$$

By Lemma 5.2, $\|\nabla\varphi_k\|^2 \le \xi_1 \sum_{i=1}^n \|z^k - x_i^k\|^2$, where ξ_1 is defined as in (5.6). Furthermore, Lemma 5.3 implies that $\varphi_k(z^k, \mathbf{w}^k) \ge \xi_2 \sum_{i=1}^n \|z^k - x_i^k\|^2$, where ξ_2 is defined as in (5.12). Combining these two inequalities with (6.1) and $\beta_k \ge \underline{\beta}$ yields $\alpha_k \ge \underline{\beta}\xi_1/\xi_2 = \underline{\alpha}$. Using (5.6) and (5.12), ξ_1 and ξ_2 are positive and finite by Assumptions 3.2 and 3.3. Since $\beta_k > \underline{\beta} > 0$, we conclude that $\underline{\alpha} > 0$. ◻

The next lemma is the key to proving the $\mathrm{O}(1/k)$ function value convergence rate and linear convergence rate under strong monotonicity and cocoercivity. Essentially, it shows that several key quantities of Algorithm 3.1 are square-summable, within known bounds.

LEMMA 6.2. *Suppose Assumptions* 2.3, 3.2, *and* 3.3 *hold. If* $\bar{K} = \infty$, *then* $\varphi_k(p^k) \to 0$ *and* $\nabla\varphi_k \to 0$. *Furthermore, for all* $1 \le k < \bar{K}$,

$$\sum_{t=1}^{k} \|z^{t+1} - z^t\|^2 \le \gamma^{-1}\tau\|p^1 - p^*\|^2, \tag{6.2}$$

$$\sum_{t=1}^{k}\sum_{i=1}^{n-1} \|w_i^{t+1} - w_i^t\|^2 \le \tau\|p^1 - p^*\|^2, \tag{6.3}$$

$$\sum_{i=1}^{n} \|z^k - x_i^k\|^2 \le \frac{\xi_1}{\underline{\beta}^2\xi_2^2}\|p^{k+1} - p^k\|^2, \quad \sum_{t=1}^{k}\sum_{i=1}^{n} \|z^t - x_i^t\|^2 \le \frac{\tau\xi_1}{\underline{\beta}^2\xi_2^2}\|p^1 - p^*\|^2, \tag{6.4}$$

$$\sum_{i=1}^{n} \|w_i^k - y_i^k\|^2 \le E_1\|p^{k+1} - p^k\|^2, \qquad \sum_{t=1}^{k}\sum_{i=1}^{n} \|w_i^t - y_i^t\|^2 \le \tau E_1\|p^1 - p^*\|^2, \tag{6.5}$$

$$\sum_{t=1}^{k}\sum_{i\in\mathcal{I}_{\mathrm{F}}} \|w_i^t - T_i z^t\|^2 \le \tau E_1\|p^1 - p^*\|^2, \tag{6.6}$$

where

$$E_1 \triangleq 2(1-\sigma)^{-1}\underline{\rho}^{-1}(1 + \xi_2^{-1}\bar{L}(1 + \underline{\rho}\bar{L}))\frac{\xi_1}{\underline{\beta}^2\xi_2}, \tag{6.7}$$

τ *is as defined in* (5.3), *and* ξ_1 *and* ξ_2 *are as defined in* (5.6) *and* (5.12).

Proof. Fix any $1 \le k < \bar{K}$. First, recall (5.3) in Lemma 5.1. Since we have $\|p^{k+1} - p^k\|^2 = \gamma\|z^{k+1} - z^k\|^2 + \sum_{i=1}^{n-1} \|w_i^{k+1} - w_i^k\|^2$, inequalities (6.2) and (6.3) follow immediately from (5.3). Next, Lemma 5.1 also implies that $p^{k+1} - p^k = -(\beta_k\varphi_k(p^k)/\|\nabla\varphi_k\|^2)\nabla\varphi_k$, and therefore that

$$\|p^{k+1} - p^k\| = \frac{\beta_k\varphi_k(p^k)}{\|\nabla\varphi_k\|}. \tag{6.8}$$

As argued in Lemma 6.1, Lemmas 5.2 and 5.3 imply that

$$\frac{\varphi_k(p^k)}{\|\nabla\varphi_k\|^2} \ge \frac{\xi_2}{\xi_1} \qquad \Longrightarrow \qquad \|\nabla\varphi_k\|^2 \le \frac{\xi_1}{\xi_2}\varphi_k(p^k). \tag{6.9}$$

Since $\|\nabla\varphi_k\|^2 \ge 0$, it follows that $\varphi_k(p^k) \ge 0$ and we may take the square root of both sides of the right-hand inequality in (6.9). We then substitute the resulting inequality into (6.8) and use the lower bound on β_k to derive

$$\varphi_k(p^k) = \beta_k^{-1}\|\nabla\varphi_k\|\|p^{k+1} - p^k\| \le \underline{\beta}^{-1}\left(\frac{\xi_1}{\xi_2}\right)^{\frac{1}{2}}\sqrt{\varphi_k(p^k)}\,\|p^{k+1} - p^k\|,$$

which in turn leads to

$$\sqrt{\varphi_k(p^k)} \le \underline{\beta}^{-1}\left(\frac{\xi_1}{\xi_2}\right)^{\frac{1}{2}}\|p^{k+1}-p^k\| \quad \Longrightarrow \quad \varphi_k(p^k) \le \frac{\xi_1}{\xi_2\underline{\beta}^2}\|p^{k+1}-p^k\|^2. \tag{6.10}$$

Therefore,

$$\sum_{t=1}^{k}\varphi_t(p^t) \le \frac{\xi_1}{\xi_2\underline{\beta}^2}\sum_{t=1}^{k}\|p^{t+1}-p^t\|^2 \le \frac{\tau\xi_1}{\underline{\beta}^2\xi_2}\|p^1-p^*\|^2. \tag{6.11}$$

If $\bar{K}=\infty$, (6.11) implies that $\varphi_k(p^k)\to 0$, which in conjunction with the right-hand inequality in (6.9) implies that $\nabla\varphi_k \to 0$. Combining (6.10) with Lemma 5.3 yields the first part of (6.4). Applying (5.3) from Lemma 5.1 then yields the second part of (6.4).

Finally, we establish (6.5) and (6.6). Inequality (5.13) from Lemma 5.3 implies that

$$\begin{aligned}(1-\sigma)\underline{\rho}\sum_{i\in\mathcal{I}_{\mathrm{B}}}\|w_i^k-y_i^k\|^2+\underline{\rho}\sum_{i\in\mathcal{I}_{\mathrm{F}}}\|w_i^k-T_iz^k\|^2 &\le \varphi_k(p^k)+\bar{L}\sum_{i\in\mathcal{I}_{\mathrm{F}}}\|x_i^k-z^k\|^2\\ &\le (1+\xi_2^{-1}\bar{L})\frac{\xi_1}{\underline{\beta}^2\xi_2}\|p^{k+1}-p^k\|^2,\end{aligned} \tag{6.12}$$

where in the second inequality we have used (6.10) and (6.4). Together with (5.3) from Lemma 5.1, (6.12) implies (6.6). Furthermore, for any $i\in\mathcal{I}_{\mathrm{F}}$, we may write $w_i^k-y_i^k=w_i^k-T_iz^k+T_iz^k-y_i^k$, from which Lemma 2.1 can be used to obtain

$$\|w_i^k-T_iz^k\|^2 \ge \frac{1}{2}\|w_i^k-y_i^k\|^2-\|T_iz^k-y_i^k\|^2. \tag{6.13}$$

Substituting (6.13) into the left-hand side of (6.12) yields

$$\begin{aligned}(1-\sigma)\underline{\rho}&\sum_{i\in\mathcal{I}_{\mathrm{B}}}\|w_i^k-y_i^k\|^2+\frac{1}{2}\underline{\rho}\sum_{i\in\mathcal{I}_{\mathrm{F}}}\|w_i^k-y_i^k\|^2\\ &\le (1+\xi_2^{-1}\bar{L})\frac{\xi_1}{\underline{\beta}^2\xi_2}\|p^{k+1}-p^k\|^2+\underline{\rho}\sum_{i\in\mathcal{I}_{\mathrm{F}}}\|T_iz^k-T_ix_i^k\|^2\\ &\le (1+\xi_2^{-1}\bar{L})\frac{\xi_1}{\underline{\beta}^2\xi_2}\|p^{k+1}-p^k\|^2+\underline{\rho}\bar{L}^2\sum_{i\in\mathcal{I}_{\mathrm{F}}}\|z^k-x_i^k\|^2\\ &\le (1+\xi_2^{-1}\bar{L}(1+\underline{\rho}\bar{L}))\frac{\xi_1}{\underline{\beta}^2\xi_2}\|p^{k+1}-p^k\|^2,\end{aligned}$$

where first inequality follows by substituting $y_k^k=T_ix_i^k$ for $i\in\mathcal{I}_{\mathrm{F}}$, the second inequality uses the Lipschitz continuity of T_i for $i\in\mathcal{I}_{\mathrm{F}}$, and the final inequality uses (6.4). Because $0\le\sigma<1$, we may replace the coefficients in front of the two left-hand sums by $(1-\sigma)/2$, which can only be smaller, and obtain

$$\sum_{i=1}^{n}\|w_i^k-y_i^k\|^2 \le 2(1-\sigma)^{-1}\underline{\rho}^{-1}(1+\xi_2^{-1}\bar{L}(1+\underline{\rho}\bar{L}))\frac{\xi_1}{\underline{\beta}^2\xi_2}\|p^{k+1}-p^k\|^2, \tag{6.14}$$

which yields the first part of (6.5). The second part follows by applying (5.3) to (6.14). $\square$

The final technical lemma shows that the sequences $\{x_i^k\}$ and $\{y_i^k\}$ are bounded for $i = 1, \ldots, n$, and computes specific bounds on their norms.

LEMMA 6.3. *Suppose Assumptions 2.3, 3.2, and 3.3 hold. The sequences $\{x_i^k\}$ and $\{y_i^k\}$ are bounded for $i = 1, \ldots, n$. In particular, for all $i = 1, \ldots, n$ and $1 \le k \le \bar{K}$,*

$$\|x_i^k\|^2 \le \min_{p^* \in \mathcal{S}} \left\{ 2\left(\frac{\tau\xi_1}{\underline{\beta}^2\xi_2^2} + 2\gamma^{-1} \right) \|p^1 - p^*\|^2 + 4\gamma^{-1}\|p^*\|^2 \right\} \triangleq B_x^2, \tag{6.15}$$

$$\|y_i^k\|^2 \le \min_{p^* \in \mathcal{S}} \{2(2n + \tau E_1)\|p^1 - p^*\|^2 + 4n\|p^*\|^2\} \triangleq B_y^2. \tag{6.16}$$

Proof. Assumption 2.3 asserts $\mathcal{S}$ is nonempty, so let $p^* \in \mathcal{S}$ and fix any $1 \le k < \bar{K}$. From Lemma 2.1, we obtain

$$\|p^k\|^2 = \|p^k - p^* + p^*\|^2 \le 2\|p^k - p^*\|^2 + 2\|p^*\|^2 \le 2\|p^1 - p^*\|^2 + 2\|p^*\|^2, \tag{6.17}$$

where the final inequality uses (5.2). It immediately follows that

$$\|z^k\|^2 \le \gamma^{-1}\|p^k\|^2 \le 2\gamma^{-1}\|p^1 - p^*\|^2 + 2\gamma^{-1}\|p^*\|^2, \tag{6.18}$$

$$\sum_{i=1}^{n-1} \|w_i^k\|^2 \le \|p^k\|^2 \le 2\|p^1 - p^*\|^2 + 2\|p^*\|^2. \tag{6.19}$$

Furthermore, using the definition of w_n^k and Lemma 2.1, we also have

$$\|w_n^k\|^2 = \left\| \sum_{i=1}^{n-1} w_i^k \right\|^2 \le (n-1)\sum_{i=1}^{n-1} \left\|w_i^k\right\|^2 \le 2n\|p^1 - p^*\|^2 + 2n\|p^*\|^2. \tag{6.20}$$

Therefore, for all $i = 1, \ldots, n$, we have

$$\|x_i^k\|^2 \le 2\|z^k - x_i^k\|^2 + 2\|z^k\|^2 \le 2\left(\frac{\tau\xi_1}{\underline{\beta}^2\xi_2^2} + 2\gamma^{-1} \right)\|p^1 - p^*\|^2 + 4\gamma^{-1}\|p^*\|^2, \tag{6.21}$$

where the second inequality uses (6.18) and (6.4). Then, for $i = 1, \ldots, n$,

$$\|y_i^k\|^2 \le 2\|y_i^k - w_i^k\|^2 + 2\|w_i^k\|^2 \le 2(2n + \tau E_1)\|p^1 - p^*\|^2 + 4n\|p^*\|^2, \tag{6.22}$$

where the second inequality uses (6.5) and (6.18)–(6.20). Note that the factor of n is only necessary when $i = n$, but for simplicity we will just use a single bound for all i.

Finally, the set $\mathcal{S}$ is closed and convex from Proposition 2.4, and the expressions (6.21) and (6.22) are continuous functions of p^*. Thus, we may minimize these bounds over $\mathcal{S}$, yielding (6.15) and (6.16). □

7. Function value convergence rate.

7.1. Assumptions. We now consider the optimization problem given in (1.2), which we repeat here:

$$F^* \triangleq \min_{z \in \mathcal{H}} F(z) = \min_{z \in \mathcal{H}} \sum_{i=1}^{n} f_i(z). \tag{7.1}$$

Assumption 7.1. Problem (7.1) conforms to the following:

1. $\mathcal{H}$ is a real Hilbert space;
2. each $f_i : \mathcal{H} \to (-\infty, +\infty]$ is convex, closed, and proper;
3. there exists some subset $\mathcal{I}_{\mathrm{F}} \subseteq \{1, \dots, n\}$ s.t., for all $i \in \mathcal{I}_{\mathrm{F}}$, f is Fréchet differentiable everywhere and ∇f_i is L_i-Lipschitz continuous;
4. for $i \in \mathcal{I}_{\mathrm{B}} \triangleq \{1, \dots, n\} \setminus \mathcal{I}_{\mathrm{F}}$, $\mathrm{Prox}^{\rho}_{\partial f_i} : \mathcal{H} \to \mathcal{H}^2$ can be computed to within the error tolerance specified in Assumption 3.2 (however, these functions are not precluded from also having Lipschitz continuous gradients);
5. letting $T_i = \partial f_i$ for $i = 1, \dots, n$, the solution set $\mathcal{S}$ in (2.5) is nonempty.

When $|\mathcal{I}_{\mathrm{B}}| \geq 1$, it is possible for (7.1) to have a finite solution, and yet for $\mathcal{S}$ to be empty. For constraint-qualification-like conditions precluding such pathological situations, see, for example, [2, Corollary 16.50].

LEMMA 7.2. *If Assumption* 7.1 *holds, then Assumption* 2.3 *holds for* (1.1) *with* $T_i = \partial f_i$ *for* $i = 1, \dots, n$. *If* $(z, \mathbf{w}) \in \mathcal{S}$, *then* z *is a solution to* (7.1). *The solution value* F^* *is finite.*

Proof. Closed, convex, and proper functions have maximal monotone subdifferentials [2, Theorem 21.2]. That z is a solution to (7.1) for any $(z, \mathbf{w}) \in \mathcal{S}$ is a consequence of Fermat's rule [2, Proposition 27.1] and the elementary fact that $\sum_i \partial f_i(x) \subseteq \partial(\sum_i f_i)(x)$. Since $\mathcal{S}$ is nonempty, a solution to (7.1) exists. For any $(z, \mathbf{w}) \in \mathcal{S}$, the function f_i must be subdifferentiable and hence finite at z for all $i = 1, \dots, n$, so $F^* = \sum_{i=1}^n f_i(z)$ must be finite. ∎

In the following analysis, we consider two types of convergence rates. First, we will establish a rate of the form

$$\sum_{i=1}^n f_i(\bar{x}_i^k) - F^* = \mathrm{O}(1/k) \quad \text{and} \quad \|\bar{x}_i^k - \bar{x}_j^k\| = \mathrm{O}(1/k) \quad \forall i, j = 1, \dots, n,$$

where $\bar{x}_i^k$ is the averaging of the sequence $\{x_i^k\}$ defined in (7.3). With an additional Lipschitz continuity assumption, we can derive a more direct rate of the form

$$F(\bar{x}_j^k) - F^* = \sum_{i=1}^n f_i(\bar{x}_j^k) - F^* = \mathrm{O}(1/k) \tag{7.2}$$

for an appropriate index j. This additional assumption is as follows.

Assumption 7.3. If $|\mathcal{I}_{\mathrm{B}}| > 1$, there exists $\mathcal{I}_{\mathrm{L}} \subseteq \mathcal{I}_{\mathrm{B}}$ such that $|\mathcal{I}_{\mathrm{L}}| \geq |\mathcal{I}_{\mathrm{B}}| - 1$, and for all $i \in \mathcal{I}_{\mathrm{L}}$ the function f_i is M_i-Lipschitz continuous on the ball $\mathcal{B}_x \triangleq \{x : \|x\| \leq B_x\}$, where B_x is defined in (6.15).

By virtue of Assumption 7.3, note that, for all $i \in \mathcal{I}_{\mathrm{L}}$, $\mathcal{B}_x \subseteq \mathrm{dom}(f_i)$.

It is not surprising that Assumption 7.3 is required to derive convergence rates of the form (7.2): suppose the first two functions f_1 and f_2 are the respective indicator functions of closed nonempty convex sets $\mathcal{C}_1$ and $\mathcal{C}_2$, that is, $f_i(x) = 0$ if $x \in \mathcal{C}_i$ and otherwise $f_i(x) = +\infty$. Since neither function is differentiable, $\{1, 2\} \subseteq \mathcal{I}_{\mathrm{B}}$. Since neither function is Lipschitz continuous, $\{1, 2\} \notin \mathcal{I}_{\mathrm{L}}$, unless $\mathcal{B}_x \subseteq \mathcal{C}_1 \cap \mathcal{C}_2$. This last situation is of little interest, since the iterates remain inside $\mathcal{C}_1 \cap \mathcal{C}_2$ for all iterations and thus the constraints encoded by f_1 and f_2 may be disregarded.

Instead, suppose $\mathcal{B}_x \not\subseteq \mathcal{C}_1 \cap \mathcal{C}_2$ and thus $\{1, 2\} \notin \mathcal{I}_{\mathrm{L}}$. Then $|\mathcal{I}_{\mathrm{L}}| \leq |\mathcal{I}_{\mathrm{B}}| - 2$, which violates Assumption 7.3. Suppose anyway that we could establish a convergence rate of $\mathrm{O}(1/k)$ (or similar) for some sequence of points $\tilde{x}^k$ generated by the algorithm. But this would imply that $\tilde{x}^k \in \mathcal{C}_1 \cap \mathcal{C}_2$ for all k, meaning that we would have to be able to solve an arbitrary two-set convex feasibility problem in a single iteration. Of course, it is easy to construct a counterexample in which Algorithm 3.1 does not find a point in the intersection of two closed convex sets within a finite number of iterations.

So, to derive a convergence rate of the form (7.2) we allow for only one function to be non-Lipschitz and thus able to encode a "hard" constraint. Any other nonsmooth functions present in the problem must be Lipschitz continuous on the bounded set $\mathcal{B}_x$, which contains all points potentially encountered by the algorithm.

7.2. Main result.

THEOREM 7.4. *Suppose Assumptions* 3.2, 3.3, *and* 7.1 *hold. For* $1 \le k \le \bar{K}$, *let*

$$(\forall i = 1, \ldots, n) \quad \bar{x}_i^k \triangleq \frac{\sum_{t=1}^k \alpha_t x_i^t}{\sum_{t=1}^k \alpha_t}, \tag{7.3}$$

where α_t *is calculated on line* 15 *of Algorithm* 3.1. *Fixing any* $p^* \in \mathcal{S}$, *one has the following.*

1. *If* $\bar{K} < \infty$,

$$(\forall j = 1, \ldots, n) \quad \sum_{i=1}^n f_i(x_j^{\bar{K}}) = \sum_{i=1}^n f_i(z^{\bar{K}+1}) = F^*. \tag{7.4}$$

2. *For all* $1 \le k < \bar{K}$,

$$\sum_{i=1}^n f_i(\bar{x}_i^k) - F^* \le \frac{E_2\|p^1 - p^*\|^2 + E_3\|p^1 - p^*\|}{k}, \tag{7.5}$$

where

$$E_2 \triangleq \frac{\tau\xi_1}{2\underline{\beta}\xi_2}\left(\frac{2}{\tau} + 4 + 2E_1 + \overline{\rho}_n(1 + \gamma E_1) + \frac{\gamma\delta\xi_1}{\underline{\beta}^2\xi_2^2}\right), \tag{7.6}$$

$$E_3 \triangleq 2\sqrt{n}\|p^*\|\frac{\xi_1}{\underline{\beta}\xi_2}. \tag{7.7}$$

Furthermore,

$$(\forall i, l = 1, \ldots, n) \quad \|\bar{x}_i^k - \bar{x}_l^k\| \le \frac{4\|p^1 - p^*\|}{\underline{\alpha}k} = \frac{4\xi_1\|p^1 - p^*\|}{\underline{\beta}\xi_2 k}. \tag{7.8}$$

3. *Additionally, suppose Assumption* 7.3 *holds. If* $|\mathcal{I}_{\mathrm{B}}| \ge 1$ *and* $\mathcal{I}_{\mathrm{L}} \ne \mathcal{I}_{\mathrm{B}}$, *then* $|\mathcal{I}_{\mathrm{B}} \setminus \mathcal{I}_{\mathrm{L}}| = 1$ *by Assumption* 7.3. *In this case let* j *be the unique element in* $\mathcal{I}_{\mathrm{B}} \setminus \mathcal{I}_{\mathrm{L}}$. *Otherwise if* $|\mathcal{I}_{\mathrm{B}}| \ge 1$ *and* $\mathcal{I}_{\mathrm{L}} = \mathcal{I}_{\mathrm{B}}$, *then choose* j *to be any element of* $\mathcal{I}_{\mathrm{B}}$. *On the other hand, if* $|\mathcal{I}_{\mathrm{B}}| = 0$, *then choose* j *to be any index in* $\mathcal{I}_{\mathrm{F}}$. *Then, for all* $1 \le k < \bar{K}$,

$$F(\bar{x}_j^k) - F^* = \sum_{i=1}^n f_i(\bar{x}_j^k) - F^* \le \frac{E_2\|p^1 - p^*\|^2 + E_4\|p^1 - p^*\|}{k}, \tag{7.9}$$

where

$$E_4 \triangleq E_3 + \frac{4\xi_1}{\underline{\beta}\xi_2}\left(\sum_{i \in \mathcal{I}_{\mathrm{L}}} M_i + n(B_y + 2\bar{L}B_x)\right). \tag{7.10}$$

Before embarking on the proof of this result, we state and prove three lemmas that we will use to establish (7.5).

LEMMA 7.5. *Under the assumptions and notation of Theorem* 7.4, *fix some* $p^* = (x^*, \mathbf{w}^*) \in \mathcal{S}$. *Then Algorithm* 3.1 *produces sequences such that*

$$(\forall\, k \in \{1,\dots,\bar{K}\}) \qquad \sum_{i=1}^n f_i(\bar{x}_i^k) - F^* \le \frac{\xi_1}{\underline{\beta}\xi_2} \cdot \frac{\sum_{t=1}^k \alpha_t (A_1^t + A_2^t)}{k},$$

where $\bar{x}_i^k$ *is as defined in* (7.3) *and*

$$A_1^k \triangleq \frac{\gamma}{\alpha_k}\langle z^k - z^{k+1}, x_n^k - x^*\rangle, \qquad A_2^k \triangleq \sum_{i=1}^{n-1}\langle y_i^k, x_i^k - x_n^k\rangle. \tag{7.11}$$

Proof. Since f_i is convex,

$$\sum_{i=1}^n f_i(\bar{x}_i^k) - F^* = \sum_{i=1}^n f_i\left(\frac{\sum_{t=1}^k \alpha_t x_i^t}{\sum_{t=1}^k \alpha_t}\right) - F^* \le \frac{\sum_{t=1}^k \alpha_t \left(\sum_{i=1}^n f_i(x_i^t) - F^*\right)}{\sum_{t=1}^k \alpha_t}. \tag{7.12}$$

Since Lemma 6.1 implies $\alpha_k \ge \underline{\alpha}$, we will aim to show that $\sum_{t=1}^k \alpha_t(\sum_{i=1}^n f_i(x_i^t) - F^*)$ is bounded for all $1 \le k < \bar{K}$ (recall that $\bar{K}$ may be $+\infty$).

The projection updates on lines 19, 20 of Algorithm 3.1 mean that, for all $1 \le k < \bar{K}$,

$$z^{k+1} = z^k - \gamma^{-1}\alpha_k \sum_{i=1}^n y_i^k, \tag{7.13}$$

$$w_i^{k+1} = w_i^k - \alpha_k(x_i^k - x_n^k), \quad i = 1,\dots,n-1. \tag{7.14}$$

Since $p^* = (x^*, \mathbf{w}^*) \in \mathcal{S}$, by Lemma 7.2 we get $F^* = \sum_{i=1}^n f_i(x^*)$. Therefore,

$$\begin{aligned}
\sum_{i=1}^n f_i(x_i^k) - F^* &= \sum_{i=1}^n f_i(x_i^k) - \sum_{i=1}^n f_i(x^*) \\
&\overset{\text{(a)}}{\le} \sum_{i=1}^n \langle y_i^k, x_i^k - x^*\rangle \\
&= \left\langle \sum_{i=1}^n y_i^k, x_n^k - x^* \right\rangle + \sum_{i=1}^{n-1}\langle y_i^k, x_i^k - x_n^k\rangle \\
&\overset{\text{(b)}}{=} \underbrace{\frac{\gamma}{\alpha_k}\langle z^k - z^{k+1}, x_n^k - x^*\rangle}_{\triangleq A_1^k} + \underbrace{\sum_{i=1}^{n-1}\langle y_i^k, x_i^k - x_n^k\rangle}_{\triangleq A_2^k}.
\end{aligned} \tag{7.15}$$

In the above, (a) uses that $y_i^k \in \partial f_i(x_i^k)$ and (b) uses (7.13). We now show that $\alpha_k A_1^k$ and $\alpha_k A_2^k$ both have a finite sum over k.

Now using (7.15) in (7.12) and the fact that $\alpha_t \geq \underline{\alpha}$ by Lemma 6.1, we obtain, for any $1 \leq k < \bar{K}$,

$$\begin{aligned}
\sum_{i=1}^{n} f_i(\bar{x}_i^k) - F^* &\leq \frac{\sum_{t=1}^k \alpha_t \left(\sum_{i=1}^n f_i(x_i^t) - F^*\right)}{\sum_{t=1}^k \alpha_t} \\
&\leq \frac{\sum_{t=1}^k \alpha_t (A_1^t + A_2^t)}{\sum_{t=1}^k \alpha_t} \\
&\leq \frac{\sum_{t=1}^k \alpha_t (A_1^t + A_2^t)}{\underline{\alpha} k} = \frac{\xi_1}{\underline{\beta}\xi_2} \cdot \frac{\sum_{t=1}^k \alpha_t (A_1^t + A_2^t)}{k}. \qquad (7.16)
\end{aligned}$$

∎

LEMMA 7.6. *Again, suppose that the assumptions and notation of Theorem* 7.4 *and Algorithm* 3.1 *are in force, and fix any* $k \in \{1, \ldots, \bar{K}\}$. *For* A_1^k *defined in* (7.11),

$$\sum_{t=1}^{k} \alpha_t A_1^t \leq \frac{\tau}{2} \left(\frac{1}{\tau} + 2 + \overline{\rho}_n (1 + \gamma E_1) + \frac{\gamma \delta \xi_1}{\underline{\beta}^2 \xi_2^2} \right) \|p^1 - p^*\|^2.$$

Proof. If $n \in \mathcal{I}_{\mathrm{B}}$, A_1^k can be simplified as follows:

$$\begin{aligned}
A_1^k &= \frac{\gamma}{\alpha_k} \langle z^k - z^{k+1}, x_n^k - x^* \rangle \\
&\overset{\text{(a)}}{=} \frac{\gamma}{\alpha_k} \langle z^k - z^{k+1}, z^k - x^* + \rho_n^k (w_n^k - y_n^k) + e_n^k \rangle \\
&= \frac{\gamma}{\alpha_k} \langle z^k - z^{k+1}, z^k - x^* \rangle + \frac{\gamma \rho_n^k}{\alpha_k} \langle z^k - z^{k+1}, w_n^k - y_n^k \rangle \\
&\quad + \frac{\gamma}{\alpha_k} \langle z^k - z^{k+1}, e_n^k \rangle \\
&\overset{\text{(b)}}{\leq} \frac{\gamma}{\alpha_k} \langle z^k - z^{k+1}, z^k - x^* \rangle + \frac{\gamma \rho_n^k}{\alpha_k} \langle z^k - z^{k+1}, w_n^k - y_n^k \rangle \\
&\quad + \frac{\gamma}{2\alpha_k} \left(\|z^k - z^{k+1}\|^2 + \|e_n^k\|^2 \right) \\
&\overset{\text{(c)}}{\leq} \frac{\gamma}{\alpha_k} \langle z^k - z^{k+1}, z^k - x^* \rangle + \frac{\gamma \rho_n^k}{\alpha_k} \langle z^k - z^{k+1}, w_n^k - y_n^k \rangle \\
&\quad + \frac{\gamma}{2\alpha_k} (\|z^k - z^{k+1}\|^2 + \delta \|z^k - x_n^k\|^2), \qquad (7.17)
\end{aligned}$$

where (a) uses the resolvent update on line 4 of the algorithm for the case when $i = n$, (b) uses Young's inequality, and (c) uses Assumption 3.2. On the other hand, if $n \in \mathcal{I}_{\mathrm{F}}$, A_1^k may be written as

$$A_1^k = \frac{\gamma}{\alpha_k} \langle z^k - z^{k+1}, z^k - x^* \rangle + \frac{\gamma \rho_n^k}{\alpha_k} \langle z^k - z^{k+1}, w_n^k - T_n z^k \rangle, \qquad (7.18)$$

where we have instead used the forward step on line 6. Let $\chi^k \triangleq w_n^k - y_n^k$ when $n \in \mathcal{I}_{\mathrm{B}}$ and $\chi^k \triangleq w_n^k - T_n z^k$ when $n \in \mathcal{I}_{\mathrm{F}}$. Furthermore, let

$$\theta_k \triangleq \frac{\gamma}{2\alpha_k} (\|z^k - z^{k+1}\|^2 + \delta \|z^k - x_n^k\|^2) \qquad (7.19)$$

when $n \in \mathcal{I}_{\mathrm{B}}$ and $\theta_k \triangleq 0$ if $n \in \mathcal{I}_{\mathrm{F}}$. Combining (7.17) and (7.18), we may write

$$A_1^k = \frac{\gamma}{\alpha_k} \langle z^k - z^{k+1}, z^k - x^* \rangle + \frac{\gamma \rho_n^k}{\alpha_k} \langle z^k - z^{k+1}, \chi^k \rangle + \theta_k. \qquad (7.20)$$

Rewriting the first term in (7.20) using (2.4), we obtain

$$\frac{\gamma}{\alpha_k}\langle z^k - z^{k+1}, z^k - x^*\rangle = \frac{\gamma}{2\alpha_k}\|z^k - z^{k+1}\|^2 + \frac{\gamma}{2\alpha_k}(\|z^k - x^*\|^2 - \|z^{k+1} - x^*\|^2). \tag{7.21}$$

The second term in (7.20) may be upper bounded using Young's inequality, as follows:

$$\frac{\gamma\rho_n^k}{\alpha_k}\langle z^k - z^{k+1}, \chi^k\rangle \le \frac{\gamma\rho_n^k}{2\alpha_k}\|z^k - z^{k+1}\|^2 + \frac{\gamma\rho_n^k}{2\alpha_k}\|\chi^k\|^2.$$

Thus, we obtain

$$\begin{aligned}\alpha_k A_1^k \le \frac{\gamma}{2}(1+\overline{\rho}_n)\|z^k - z^{k+1}\|^2 + \frac{\gamma}{2}\|z^k - x^*\|^2 \\ - \frac{\gamma}{2}\|z^{k+1} - x^*\|^2 + \frac{\gamma\overline{\rho}_n}{2}\|\chi^k\|^2 + \alpha_k\theta_k.\end{aligned} \tag{7.22}$$

Summing (7.22) over k yields, for $1 \le k < \bar{K}$,

$$\sum_{t=1}^k \alpha_t A_1^t \le \frac{\gamma}{2}\|z^1 - x^*\|^2 + \frac{\gamma}{2}(1+\overline{\rho}_n)\sum_{t=1}^k \|z^t - z^{t+1}\|^2 + \frac{\gamma\overline{\rho}_n}{2}\sum_{t=1}^k\|\chi^t\|^2 + \sum_{t=1}^k \alpha_t\theta_t. \tag{7.23}$$

We now consider the first three terms on the right-hand side of this relation, employing Lemma 6.2:

$$\begin{aligned}\frac{\gamma}{2}\|z^1 - x^*\|^2 &\le \frac{1}{2}\|p^1 - p^*\|^2 && \text{by (3.1)},\\ \frac{\gamma}{2}(1+\overline{\rho}_n)\sum_{t=1}^k\|z^t - z^{t+1}\|^2 &\le \frac{1}{2}(1+\overline{\rho}_n)\tau\|p^1 - p^*\|^2 && \text{by (6.2)},\\ \frac{\gamma\overline{\rho}_n}{2}\sum_{t=1}^k\|\chi^t\|^2 &\le \frac{\gamma\overline{\rho}_n}{2}\tau E_1\|p^1 - p^*\|^2 && \text{by (6.5) or (6.6)}.\end{aligned}$$

In the last inequality, we use (6.5) when $n \in \mathcal{I}_{\mathrm{B}}$ and (6.6) when $n \in \mathcal{I}_{\mathrm{F}}$, and E_1 is defined in (6.7). We now consider the last term in (7.23). When $n \in \mathcal{I}_{\mathrm{B}}$, we have

$$\begin{aligned}\sum_{t=1}^k \alpha_t\theta_t &= \frac{\gamma}{2}\left(\sum_{t=1}^k\|z^t - z^{t+1}\|^2 + \sum_{t=1}^k \delta\|z^t - x_n^t\|^2\right) && \text{by (7.19)}\\ &\le \frac{\gamma}{2}\left(\gamma^{-1}\tau\|p^1 - p^*\|^2 + \delta\frac{\tau\xi_1}{\underline{\beta}^2\xi_2^2}\|p^1 - p^*\|^2\right) && \text{by (6.2) and (6.4)}\\ &= \frac{\tau}{2}\left(1 + \frac{\gamma\delta\xi_1}{\underline{\beta}^2\xi_2^2}\right)\|p^1 - p^*\|^2.\end{aligned}$$

The resulting inequality also holds trivially when $n \in \mathcal{I}_{\mathrm{F}}$, since its left-hand side must be zero. Combining all these inequalities, we obtain

$$\begin{aligned}\sum_{t=1}^k \alpha_t A_1^t &\le \frac{1}{2}\left(1 + (1+\overline{\rho}_n)\tau + \gamma\overline{\rho}_n\tau E_1 + \tau\left(1 + \frac{\gamma\delta\xi_1}{\underline{\beta}^2\xi_2^2}\right)\right)\|p^1 - p^*\|^2\\ &= \frac{\tau}{2}\left(\frac{1}{\tau} + 2 + \overline{\rho}_n(1+\gamma E_1) + \frac{\gamma\delta\xi_1}{\underline{\beta}^2\xi_2^2}\right)\|p^1 - p^*\|^2.\end{aligned}$$

□

We next perform a similar summability analysis on the second term in (7.15), A_2^k.

LEMMA 7.7. *Consider the same setting as in the previous lemma and fix any* $k \in \{1, \ldots, \bar{K}\}$. *For* A_2^k *as defined in* (7.11),

$$\sum_{t=1}^{k} \alpha_t A_2^t \leq \frac{1}{2}\|p^1 - p^*\|^2 + \tau(1 + E_1)\|p^1 - p^*\|^2 + 2\sqrt{n}\|p^*\|\|p^1 - p^*\|.$$

Proof. We begin by fixing any $1 \leq k < \bar{K}$ and writing

$$\alpha_k A_2^k = \alpha_k \sum_{i=1}^{n-1}\langle y_i^k, x_i^k - x_n^k\rangle = \alpha_k \sum_{i=1}^{n-1}\langle w_i^k, x_i^k - x_n^k\rangle + \alpha_k \sum_{i=1}^{n-1}\langle y_i^k - w_i^k, x_i^k - x_n^k\rangle. \tag{7.24}$$

Now fix any $i = 1, \ldots, n-1$. Using (7.14),

$$x_i^k - x_n^k = -\frac{1}{\alpha_k}(w_i^{k+1} - w_i^k). \tag{7.25}$$

Substituting this into the first summand in (7.24) and then using (2.4), we have

$$\begin{aligned}
\alpha_k\langle w_i^k, x_i^k - x_n^k\rangle &= \langle w_i^k, w_i^k - w_i^{k+1}\rangle \\
&= \langle w_i^k - w_i^*, w_i^k - w_i^{k+1}\rangle + \langle w_i^*, w_i^k - w_i^{k+1}\rangle \\
&= \frac{1}{2}(\|w_i^k - w_i^*\|^2 + \|w_i^k - w_i^{k+1}\|^2 - \|w_i^{k+1} - w_i^*\|^2) \\
&\quad + \langle w_i^*, w_i^k - w_i^{k+1}\rangle.
\end{aligned} \tag{7.26}$$

Using (7.25) in the second summand in (7.24) and applying Young's inequality yields

$$\alpha_k\langle y_i^k - w_i^k, x_i^k - x_n^k\rangle = \langle y_i^k - w_i^k, w_i^k - w_i^{k+1}\rangle \leq \frac{1}{2}\|y_i^k - w_i^k\|^2 + \frac{1}{2}\|w_i^k - w_i^{k+1}\|^2. \tag{7.27}$$

Substituting (7.26) and (7.27) into (7.24) yields

$$\begin{aligned}
\alpha_k A_2^k \leq \frac{1}{2}\sum_{i=1}^{n-1}&(\|w_i^k - w_i^*\|^2 + 2\|w_i^k - w_i^{k+1}\|^2 - \|w_i^{k+1} - w_i^*\|^2 + \|y_i^k - w_i^k\|^2) \\
&+ \sum_{i=1}^{n-1}\langle w_i^*, w_i^k - w_i^{k+1}\rangle.
\end{aligned}$$

Summing this relation, we obtain that, for any $1 \leq k < \bar{K}$,

$$\begin{aligned}
\sum_{t=1}^{k}\alpha_t A_2^t &\leq \frac{1}{2}\sum_{i=1}^{n-1}\left(\|w_i^1 - w_i^*\|^2 + 2\sum_{t=1}^{k}\|w_i^t - w_i^{t+1}\|^2 + \sum_{t=1}^{k}\|y_i^t - w_i^t\|^2\right) \\
&\quad + \sum_{i=1}^{n-1}\langle w_i^*, w_i^1 - w_i^{k+1}\rangle \qquad (7.28) \\
&\leq \frac{1}{2}\sum_{i=1}^{n-1}\left(\|w_i^1 - w_i^*\|^2 + 2\sum_{t=1}^{k}\|w_i^t - w_i^{t+1}\|^2 + \sum_{t=1}^{k}\|y_i^t - w_i^t\|^2\right) \\
&\quad + \sum_{i=1}^{n-1}\|w_i^*\|(\|w_i^1 - w_i^*\| + \|w_i^{k+1} - w_i^*\|). \qquad (7.29)
\end{aligned}$$

By (6.3) and (6.5) from Lemma 6.2, we then obtain

$$\sum_{t=1}^{k}\sum_{i=1}^{n-1}(2\|w_i^t - w_i^{t+1}\|^2 + \|y_i^t - w_i^t\|^2) \leq 2\tau(1+E_1)\|p^1 - p^*\|^2.$$

Finally, we may use (5.5) to derive

$$\begin{aligned}\sum_{i=1}^{n-1}\|w_i^*\|(\|w_i^1 - w_i^*\| + \|w_i^{k+1} - w_i^*\|) &\leq 2\|p^1 - p^*\|\sum_{i=1}^{n-1}\|w_i^*\| \\ &\leq 2\sqrt{n}\|p^1 - p^*\|\|\mathbf{w}^*\| \leq 2\sqrt{n}\|p^1 - p^*\|\|p^*\|,\end{aligned}$$

where in the second inequality we have used Lemma 2.1. Therefore,

$$\sum_{t=1}^{k}\alpha_t A_2^t \leq \frac{1}{2}\|p^1 - p^*\|^2 + \tau(1+E_1)\|p^1 - p^*\|^2 + 2\sqrt{n}\|p^*\|\|p^1 - p^*\|. \qquad \square$$

Proof of Theorem 7.4. In the $\bar{K} < \infty$ case, it was established in [15, Lemma 5] that $x_1^{\bar{K}} = \cdots = x_n^{\bar{K}} = z^{\bar{K}+1}$ is a solution to (7.1), which establishes (7.4). We now address points 2 and 3. The proof consists of three separate parts, proving (7.5), (7.8), and (7.9), the first of which follows from the preceding three lemmas.

Part 1 (proof of (7.5)). Combining Lemmas 7.5, 7.6, and 7.7 establishes (7.5), with E_2 and E_3 as defined in (7.6) and (7.7).

Part 2 (proof of (7.8)). Repeating (7.14) yields, for all $i = 1,\ldots,n-1$,

$$w_i^{t+1} - w_i^t = -\alpha_t(x_i^t - x_n^t).$$

Summing this equation over $t = 1,\ldots,k$ and dividing by $\sum_{t=1}^k \alpha_t$ yields

$$(\forall i = 1,\ldots,n-1) \quad \bar{x}_n^k - \bar{x}_i^k = \frac{w_i^{k+1} - w_i^1}{\sum_{t=1}^k \alpha_t}. \tag{7.30}$$

Therefore, we obtain, for all $i, l \in \{1,\ldots,n-1\}$, that

$$\begin{aligned}\bar{x}_l^k - \bar{x}_i^k &= \frac{w_l^1 - w_i^1 - (w_l^{k+1} - w_i^{k+1})}{\sum_{t=1}^k \alpha_t} \\ &= \frac{w_l^1 - w_l^* - w_i^1 + w_i^* - (w_l^{k+1} - w_l^* - w_i^{k+1} + w_i^*)}{\sum_{t=1}^k \alpha_t}.\end{aligned}$$

Using (5.5) and the triangle inequality, we therefore have, for all $i, l \in \{1,\ldots,n-1\}$, that

$$\|\bar{x}_l^k - \bar{x}_i^k\| \leq \frac{\|w_l^1 - w_l^*\| + \|w_i^1 - w_i^*\| + \|w_l^{k+1} - w_l^*\| + \|w_i^{k+1} - w_i^*\|}{\sum_{t=1}^k \alpha_t} \leq \frac{4\|p^1 - p^*\|}{\underline{\alpha}k}, \tag{7.31}$$

where the second inequality uses Lemmas 5.1 and 6.1. Finally, we can also use (7.30) to obtain, for any $i = 1,\ldots,n-1$, that

$$\left\|\bar{x}_n^k - \bar{x}_i^k\right\| = \frac{\|w_i^1 - w_i^* - (w_i^k - w_i^*)\|}{\sum_{t=1}^k \alpha_t} \leq \frac{\|w_i^1 - w_i^*\| + \|w_i^k - w_i^*\|}{\sum_{t=1}^k \alpha_t} \leq \frac{2\|p^1 - p^*\|}{\underline{\alpha}k}, \tag{7.32}$$

where the second inequality again uses Lemmas 5.1 and 6.1. Together, (7.31), (7.32), and the definition of $\underline{\alpha}$ imply (7.8).

Part 3 (proof of (7.9)). Finally, we prove (7.9) under Assumption 7.3. We write

$$\sum_{i=1}^{n} f_i(\bar{x}_j^k) - F^* = \sum_{i=1}^{n} f_i(\bar{x}_i^k) - F^* + \sum_{i\neq j}(f_i(\bar{x}_j^k) - f_i(\bar{x}_i^k)). \tag{7.33}$$

The first summation on the right-hand side of this equation is $\mathrm{O}(1/k)$ by (7.5), which has already been established. So we now focus on the terms in the second summation.

First suppose $i \in \mathcal{I}_{\mathrm{B}}$ and $i \neq j$. Recall the constraints on $j \in \{1,\ldots,n\}$ in the third point of the theorem. These constraints, along with Assumption 7.3, imply that $\{1,\ldots,n\}\setminus\{j\} = \mathcal{I}_{\mathrm{L}} \cup \mathcal{I}_{\mathrm{F}}$. Thus, if $i \in \mathcal{I}_{\mathrm{B}}$ and $i \neq j$, since $\{1,\ldots,n\}\setminus\{j\} = \mathcal{I}_{\mathrm{L}} \cup \mathcal{I}_{\mathrm{F}}$, $\mathcal{I}_{\mathrm{L}} \subseteq \mathcal{I}_{\mathrm{B}}$, and $\mathcal{I}_{\mathrm{B}} \cap \mathcal{I}_{\mathrm{F}} = \{\emptyset\}$, it follows that i must be in $\mathcal{I}_{\mathrm{L}}$. Furthermore, since $\mathcal{B}_x$ is a convex set, $\bar{x}_j^k$ and $\bar{x}_i^k$ are both in $\mathcal{B}_x$. Therefore, we may use Assumption 7.3 to write

$$f_i(\bar{x}_j^k) - f_i(\bar{x}_i^k) \leq M_i\|\bar{x}_i^k - \bar{x}_j^k\| \leq \frac{4M_i\|p^1 - p^*\|}{\underline{\alpha}k}, \tag{7.34}$$

where the second inequality arises from (7.8).

Otherwise, we have $i \in \mathcal{I}_{\mathrm{F}}$ and $i \neq j$, and since $\mathrm{dom}(\nabla f_i) = \mathcal{H}$ we may use the subgradient inequality to write

$$\begin{aligned} f_i(\bar{x}_j^k) - f_i(\bar{x}_i^k) &\leq \langle \nabla f_i(\bar{x}_j^k), \bar{x}_j^k - \bar{x}_i^k\rangle \\ &= \langle \nabla f_i(x_i^k), \bar{x}_j^k - \bar{x}_i^k\rangle + \langle \nabla f_i(\bar{x}_j^k) - \nabla f_i(x_i^k), \bar{x}_j^k - \bar{x}_i^k\rangle \\ &\leq \langle y_i^k, \bar{x}_j^k - \bar{x}_i^k\rangle + L_i\|\bar{x}_j^k - x_i^k\|\|\bar{x}_j^k - \bar{x}_i^k\| \\ &\leq \|y_i^k\|\|\bar{x}_j^k - \bar{x}_i^k\| + L_i(\|\bar{x}_j^k\| + \|x_i^k\|)\|\bar{x}_j^k - \bar{x}_i^k\| \\ &= (\|y_i^k\| + L_i(\|\bar{x}_j^k\| + \|\bar{x}_i^k\|))\|\bar{x}_j^k - \bar{x}_i^k\|, \end{aligned} \tag{7.35}$$

where the second inequality uses that $y_i^k = \nabla f_i(x_i^k)$, $i \in \mathcal{I}_{\mathrm{F}}$, together with the Cauchy–Schwarz inequality and the Lipschitz continuity of ∇f_i. Lemma 6.3 assures us that $\|x_i^k\|, \|\bar{x}_j^k\| \leq B_x$ and $\|y_i^k\| \leq B_y$, which in conjunction with (7.8) and $L_i \leq \bar{L}$ leads to

$$f_i(\bar{x}_j^k) - f_i(\bar{x}_i^k) \leq \frac{4(B_y + 2\bar{L}B_x)\|p^1 - p^*\|}{\underline{\alpha}k}. \tag{7.36}$$

Substituting (7.5), (7.34), and (7.36) into (7.33), we obtain

$$\begin{aligned} \sum_{i=1}^{n} f_i(\bar{x}_j^k) - F^* &= \sum_{i=1}^{n} f_i(\bar{x}_i^k) - F^* + \sum_{i\neq j}(f_i(\bar{x}_j^k) - f_i(\bar{x}_i^k)) \\ &\leq \frac{E_2\|p^1 - p^*\|^2 + E_3\|p^1 - p^*\|}{k} \\ &\qquad + \sum_{i\in\mathcal{I}_{\mathrm{L}}} \frac{4M_i\|p^1 - p^*\|}{\underline{\alpha}k} + \frac{4|\mathcal{I}_{\mathrm{F}}|(B_y + 2\bar{L}B_x)\|p^1 - p^*\|}{\underline{\alpha}k} \\ &\leq \frac{E_2\|p^1 - p^*\|^2 + E_3\|p^1 - p^*\|}{k} \\ &\qquad + \frac{4\|p^1 - p^*\|}{\underline{\alpha}k}\left(\sum_{i\in\mathcal{I}_{\mathrm{L}}} M_i + n(B_y + 2\bar{L}B_x)\right). \end{aligned}$$

This establishes (7.9) with E_4 as defined in (7.10). ∎

8. Consequences of strong monotonicity. We now investigate the consequences of strong monotonicity in (1.1) for the convergence rate of projective splitting.

Assumption 8.1. For (1.1), there is some $l \in \{1, \dots, n\}$ for which T_l is μ-strongly monotone for some $\mu > 0$.

Note that under Assumption 8.1 it is immediate that the solution z^* of (1.1) is unique.

THEOREM 8.2. *Suppose Assumptions* 2.3, 3.2, 3.3, *and* 8.1 *hold. Let z^* be the unique solution to* (1.1). *Fix $p^* = (z^*, \mathbf{w}^*) \in \mathcal{S}$. If $\bar{K} = +\infty$, then $x_i^k \to z^*$ for $i = 1, \dots, n$ and $z^k \to z^*$. If $\bar{K} < \infty$, $x_i^{\bar{K}} = z^{\bar{K}+1} = z^*$ for $i = 1, \dots, n$. Furthermore, for all $1 \le k < \bar{K}$,*

$$\|x_{\text{avg}}^k - z^*\|^2 \le \frac{\xi_1(1+\tau)\|p^1 - p^*\|^2}{2\underline{\beta}\xi_2\mu k}, \tag{8.1}$$

where $x_{\text{avg}}^k \triangleq \frac{1}{k}\sum_{t=1}^k x_l^t$ and ξ_1 and ξ_2 are defined in (5.6) *and* (5.12), *respectively.*

Proof. Regarding the case when $\bar{K} < \infty$, the optimality of $x_i^{\bar{K}}$ and $z^{\bar{K}+1}$ was shown in [15, Lemma 5].

Let $w_n^* \triangleq -\sum_{i=1}^{n-1} w_i^*$. Using this notation and recalling the affine function defined in (3.2), we may write

$$\begin{aligned} -\varphi_k(z^*, \mathbf{w}^*) &= \sum_{i=1}^n \langle z^* - x_i^k, w_i^* - y_i^k \rangle \\ &= \sum_{i \ne l} \langle z^* - x_i^k, w_i^* - y_i^k \rangle + \langle z^* - x_l^k, w_l^* - y_l^k \rangle \\ &\ge \mu\|z^* - x_l^k\|^2, \end{aligned} \tag{8.2}$$

where the inequality uses that (x_i^k, y_i^k) and (z^*, w_i^*) are in gra T_i, along with the monotonicity of T_i for $i \ne l$, and also the strong monotonicity of T_l. Now, Lemma 5.3 implies that $\varphi_k(z^k, \mathbf{w}^k) \ge 0$. Therefore, for any $1 \le k < \bar{K}$,

$$\begin{aligned} \mu\|z^* - x_l^k\|^2 &\le \varphi_k(z^k, \mathbf{w}^k) - \varphi_k(z^*, \mathbf{w}^*) \\ &\overset{\text{(a)}}{=} \langle \nabla\varphi_k, p^k - p^* \rangle \\ &\overset{\text{(b)}}{=} \frac{1}{\alpha_k}\langle p^k - p^{k+1}, p^k - p^* \rangle \\ &\overset{\text{(c)}}{=} \frac{1}{2\alpha_k}(\|p^k - p^{k+1}\|^2 + \|p^k - p^*\|^2 - \|p^{k+1} - p^*\|^2) \\ &\overset{\text{(d)}}{\le} \frac{1}{2\underline{\alpha}}(\|p^k - p^{k+1}\|^2 + \|p^k - p^*\|^2 - \|p^{k+1} - p^*\|^2). \end{aligned}$$

Above, (a) uses that φ_k is affine, (b) employs (5.1) from Lemma 5.1, (c) uses (2.4), and (d) uses Lemma 6.1. Summing the resulting inequality yields, for all $1 \le k < \bar{K}$,

$$\mu\sum_{t=1}^k \|z^* - x_l^t\|^2 \le \frac{1}{2\underline{\alpha}}\left(\sum_{t=1}^k \|p^t - p^{t+1}\|^2 + \|p^1 - p^*\|^2\right) \le \frac{\xi_1(1+\tau)}{2\underline{\beta}\xi_2}\|p^1 - p^*\|^2, \tag{8.3}$$

where the second inequality uses (5.3) and the definition of $\underline{\alpha}$ in Lemma 6.1. So, when $\bar{K} = \infty$, we have $x_l^k \to z^*$. Lemma 6.2 asserts that $\nabla\varphi_k \to 0$ in the $\bar{K} = \infty$ case, so from (3.3) we conclude that $x_i^k \to z^*$ for $i = 1, \ldots, n$. Lemma 6.2 also establishes that $\varphi_k(p^k) \to 0$, so by Lemma 5.3, we have $\|z^k - x_i^k\| \to 0$ for all $i = 1, \ldots, n$ and hence $z^k \to z^*$. Finally, the convexity of the function $\|\cdot\|^2$ in conjunction with (8.3) yields (8.1). ∎

9. Linear convergence under strong monotonicity and cocoercivity. If we introduce a cocoercivity assumption along with strong monotonicity, we can derive linear convergence. We require that all but one of the operators be cocoercive. Without loss of generality, we designate T_n to be the operator that need not be cocoercive.

Assumption 9.1. In (1.1), suppose that, for $i = 1, \ldots, n-1$, the operator T_i is cocoercive with parameter $\Gamma_i > 0$. Let $\bar{\Gamma} \triangleq \max_{i=1,\ldots,n-1} \Gamma_i$.

If this assumption holds, then $T_1, \ldots, T_{n-1}$ are all single valued. If Assumption 8.1 also holds, this single-valuedness implies that not only is the solution z^* to (1.1) unique, but the extended solution set $\mathcal{S}$ must be a singleton of the form $\{(z^*, w_1^*, \ldots, w_{n-1}^*)\} = \{(z^*, T_1 z^*, \ldots, T_{n-1} z^*)\}$.

Note that if $n = 1$, Assumption 9.1 only requires that the single operator T_1 be strongly monotone (and not necessarily cocoercive). If $1 \in \mathcal{I}_{\mathrm{B}}$ in the $n = 1$ case, as discussed in section 4, projective splitting reduces to the proximal point method. Thus, our linear convergence for projective splitting in this case corresponds to the well-known result for the proximal point method [2, Proposition 26.16(i)]. On the other hand, if $n = 1 \in \mathcal{I}_{\mathrm{F}}$, projective splitting reduces to the extragradient method. If Assumption 9.1 holds in this case, then T_1 must be both strongly monotone and Lipschitz continuous because $1 \in \mathcal{I}_{\mathrm{F}}$, but it is not necessarily cocoercive. In this situation, our linear convergence result resembles a prior result for the extragradient method [23, Theorem 2.2].[1]

THEOREM 9.2. *Suppose Assumptions* 2.3, 3.2, 3.3, 8.1, *and* 9.1 *hold. Let* z^* *be the unique solution to* (1.1) *and take* $p^* \triangleq (z^*, \mathbf{w}^*) = (z^*, T_1 z^*, \ldots, T_{n-1} z^*) \in \mathcal{S}$. *If* $\bar{K} < \infty$, *then* $p^{\bar{K}+1} = p^*$. *For all* $1 \le k < \bar{K}$,

$$\|p^{k+1} - p^*\|^2 \le (1 - E_5)\|p^k - p^*\|^2, \tag{9.1}$$

where

$$E_5 \triangleq \frac{1}{2\tau}\left(\frac{8\xi_1^2(1+\gamma)^2 \max\{\mu^{-1}, \bar{\Gamma}\}^2 + 2\gamma\xi_1}{\underline{\beta}^2 \xi_2^2} + 2E_1\right)^{-1} \in \left]0,\, 1/4\right]. \tag{9.2}$$

Proof. For the finite-termination case, optimality of $p^{\bar{K}+1}$ was established in [15, Lemma 5]. Henceforth, we thus assume that $\bar{K} = \infty$. The key idea of the proof is to show that there exists $E_5' > 0$ s.t.

$$E_5'\|p^k - p^*\|^2 \le \|p^{k+1} - p^k\|^2, \tag{9.3}$$

which, when used with (5.2) of Lemma 5.1, implies

$$\|p^{k+1} - p^*\|^2 \le \|p^k - p^*\|^2 - \tau^{-1}\|p^{k+1} - p^k\|^2 \le (1 - \tau^{-1}E_5')\|p^k - p^*\|^2, \tag{9.4}$$

[1]However, Theorem 2.2 of [23] is slightly stronger than our result for this highly specialized case. While we assume T_1 to be strongly monotone, which is equivalent to T_1^{-1} being cocoercive, Theorem 2.2 of [23] only assumes T_1^{-1} to be Lipschitz continuous.

where τ is as defined in (5.3). We now establish (9.3). For $1 \leq k < \bar{K}$, we have $\varphi_k(p^k) \geq 0$, and hence

$$\begin{aligned}\varphi_k(p^k) - \varphi_k(p^*) &\geq \sum_{i=1}^{n} \langle z^* - x_i^k, w_i^* - y_i^k \rangle \\ &= \frac{1}{2}\sum_{i=1}^{n} \langle z^* - x_i^k, w_i^* - y_i^k \rangle + \frac{1}{2}\sum_{i=1}^{n} \langle z^* - x_i^k, w_i^* - y_i^k \rangle \\ &\geq \frac{\mu}{2}\|z^* - x_l^k\|^2 + \sum_{i=1}^{n-1} \frac{1}{2\Gamma_i}\|y_i^k - w_i^*\|^2.\end{aligned}$$

The last inequality here follows because

$$\sum_{i=1}^{n} \langle z^* - x_i^k, w_i^* - y_i^k \rangle \geq \mu\|z^* - x_l^k\|^2$$

by the same logic as in (8.2), and by the cocoerciveness of $T_1, \ldots, T_{n-1}$. We therefore obtain

$$\begin{aligned}\frac{\mu}{2}\|z^* - x_l^k\|^2 + \sum_{i=1}^{n-1} \frac{1}{2\Gamma_i}\|y_i^k - w_i^*\|^2 &\leq \varphi_k(p^k) - \varphi_k(p^*) \\ &= \langle \nabla\varphi_k, p^k - p^* \rangle \\ &= \frac{1}{\alpha_k}\langle p^k - p^{k+1}, p^k - p^* \rangle, \end{aligned} \tag{9.5}$$

where we have again used that φ_k is affine, and also (5.1). Continuing, we use Young's inequality applied to (9.5) to write, for any $\nu > 0$,

$$\frac{\mu}{2}\|z^* - x_l^k\|^2 + \sum_{i=1}^{n-1} \frac{1}{2\Gamma_i}\|y_i^k - w_i^*\|^2 \leq \frac{1}{2\nu\underline{\alpha}^2}\|p^{k+1} - p^k\|^2 + \frac{\nu}{2}\|p^k - p^*\|^2,$$

which implies

$$\|z^* - x_l^k\|^2 + \sum_{i=1}^{n-1} \|y_i^k - w_i^*\|^2 \leq \frac{\xi_5}{\nu\underline{\alpha}^2}\|p^{k+1} - p^k\|^2 + \xi_5\nu\|p^k - p^*\|^2, \tag{9.6}$$

where $\xi_5 \triangleq \max\{\mu^{-1}, \bar{\Gamma}\}$. From Lemma 2.1, we then deduce that

$$\begin{aligned}\|p^k - p^*\|^2 &= \gamma\|z^k - z^*\|^2 + \sum_{i=1}^{n-1} \|w_i^k - w_i^*\|^2 \\ &\leq 2\gamma\|z^* - x_l^k\|^2 + 2\gamma\|x_l^k - z^k\|^2 \\ &\quad + 2\sum_{i=1}^{n-1} \|y_i^k - w_i^*\|^2 + 2\sum_{i=1}^{n-1} \|y_i^k - w_i^k\|^2 \\ &\leq 2(1+\gamma)\left(\|z^* - x_l^k\|^2 + \sum_{i=1}^{n-1} \|y_i^k - w_i^*\|^2\right) \\ &\quad + 2\gamma\|x_l^k - z^k\|^2 + 2\sum_{i=1}^{n-1} \|y_i^k - w_i^k\|^2. \end{aligned} \tag{9.7}$$

Substituting the upper bound (9.6) for the term in parentheses in (9.7), and then using (6.4) and (6.5) from Lemma 6.2 on the other two terms, we conclude that

$$\|p^k - p^*\|^2 \le 2(1+\gamma)\left(\frac{\xi_5}{\nu\underline{\alpha}^2}\|p^{k+1}-p^k\|^2 + \xi_5\nu\|p^k-p^*\|^2\right) + \frac{2\gamma\xi_1}{\underline{\beta}^2\xi_2^2}\|p^{k+1}-p^k\|^2 + 2E_1\|p^{k+1}-p^k\|^2.$$

Rearranging this inequality yields

$$(1-2\nu(1+\gamma)\xi_5)\|p^k-p^*\|^2 \le \left(\frac{2(1+\gamma)\xi_5}{\nu\underline{\alpha}^2} + \frac{2\gamma\xi_1}{\underline{\beta}^2\xi_2^2} + 2E_1\right)\|p^{k+1}-p^k\|^2. \tag{9.8}$$

We now set

$$\nu = \frac{1}{4(1+\gamma)\xi_5}.$$

Using this value of ν in (9.8) implies (9.3) with

$$E_5' = \frac{1}{2}\left(\frac{8(1+\gamma)^2(\xi_5)^2}{\underline{\alpha}^2} + \frac{2\gamma\xi_1}{\underline{\beta}^2\xi_2^2} + 2E_1\right)^{-1}, \tag{9.9}$$

which when used with (9.4) yields the expression for E_5 in (9.2). Considering (9.2), we note that, since $\tau > 0$, $\xi_2 > 0$, $\underline{\beta} > 0$, $\mu > 0$, and $E_1 < \infty$, we must have $E_5 > 0$.

Finally, we show that $E_5 \le 1/4$ as claimed in (9.2). This means that the theorem cannot, under any circumstances, guarantee convergence in a single iteration, which would be implied by $E_5 = 1$. We write

$$\begin{aligned} E_5 &= \frac{1}{2\tau}\left(\frac{8\xi_1^2(1+\gamma)^2\max\{\mu^{-1},\bar{\Gamma}\}^2 + 2\gamma\xi_1}{\underline{\beta}^2\xi_2^2} + 2E_1\right)^{-1} \\ &\overset{(a)}{\le} \frac{\xi_2^2}{\gamma\xi_1} \\ &= \frac{\min\left\{(1-\sigma)\overline{\rho}^{-1},\ \min_{j\in\mathcal{I}_F}\left\{\overline{\rho}_j^{-1} - L_j\right\}\right\}^2}{2n\gamma\left[1+2\gamma^{-1}\left(\bar{L}^2|\mathcal{I}_F| + \underline{\rho}^{-2}(1+\delta)\right)\right]} \\ &\le \frac{\min\left\{(1-\sigma)\overline{\rho}^{-1},\ \min_{j\in\mathcal{I}_F}\left\{\overline{\rho}_j^{-1} - L_j\right\}\right\}^2}{4n\underline{\rho}^{-2}(1+\delta)} \\ &\overset{(b)}{\le} \frac{\left((1-\sigma)\overline{\rho}^{-1} + \sum_{j\in\mathcal{I}_F}\left(\overline{\rho}_j^{-1} - L_j\right)\right)^2}{4n\underline{\rho}^{-2}(1+\delta)(|\mathcal{I}_F|+1)^2} \\ &\overset{(c)}{\le} \frac{\overline{\rho}^{-2} + \sum_{j\in\mathcal{I}_F}\overline{\rho}_j^{-2}}{4n\underline{\rho}^{-2}(1+\delta)(|\mathcal{I}_F|+1)} \overset{(d)}{\le} \frac{1}{4n(1+\delta)} \le \frac{1}{4}, \end{aligned} \tag{9.10}$$

where we employ the following reasoning: first, (a) uses that all terms within the parentheses in (9.10) are positive, $\underline{\beta} < 2$, and $\tau^{-1} \le 1$. In (b), we use that the minimum of a set of numbers cannot exceed its average. Inequality (c) follows by observing that $(1-\sigma) \le 1$ and $-L_j \le 0$, and then using Lemma 2.1. Finally, (d) uses that $\underline{\rho} \le \overline{\rho}_j$ and $\underline{\rho} \le \overline{\rho}$. ∎

REFERENCES

[1] A. Alotaibi, P. L. Combettes, and N. Shahzad, *Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn–Tucker set*, SIAM J. Optim., 24 (2014), pp. 2076–2095.

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

[3] 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 Mach. Learn., 3 (2011), pp. 1–122.

[4] P. L. Combettes and J. Eckstein, *Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions*, Math. Program., 168 (2018), pp. 645–672.

[5] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, New York, 2011, pp. 185–212.

[6] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[7] D. Davis and W. Yin, *Convergence rate analysis of several splitting schemes*, in Splitting Methods in Communication, Imaging, Science, and Engineering, R. Glowinski, S. J. Osher, and W. Yin, eds., Springer, New York, 2016, pp. 115–163.

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

[9] J. Eckstein, *A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers*, J. Optim. Theory Appl., 173 (2017), pp. 155–182.

[10] J. Eckstein and D. P. Bertsekas, *On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[11] J. Eckstein and B. F. Svaiter, *A family of projective splitting methods for the sum of two maximal monotone operators*, Math. Program., 111 (2008), pp. 173–199.

[12] J. Eckstein and B. F. Svaiter, *General projective splitting methods for sums of maximal monotone operators*, SIAM J. Control Optim., 48 (2009), pp. 787–811.

[13] Y. Haugazeau, *Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes*, Ph.D. thesis, Université de Paris, Paris, 1968.

[14] A. Iusem and B. Svaiter, *A variant of Korpelevich's method for variational inequalities with a new search strategy*, Optimization, 42 (1997), pp. 309–321.

[15] P. R. Johnstone and J. Eckstein, *Projective Splitting with Forward Steps: Asynchronous and Block-Iterative Operator Splitting*, preprint, https://arxiv.org/abs/1803.07043, 2018.

[16] G. Korpelevich, *Extragradient method for finding saddle points and other problems*, Matekon, 13 (1977), pp. 35–49.

[17] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.

[18] M. P. Machado, *On the complexity of the projective splitting and Spingarn's methods for the sum of two maximal monotone operators*, J. Optim. Theory Appl., 178 (2018), pp. 153–190.

[19] M. P. Machado, *Projective method of multipliers for linearly constrained convex minimization*, Comput. Optim. Appl., 73 (2019), pp. 237–273.

[20] B. Mercier, *Lectures on Topics in Finite Element Solution of Elliptic Problems*, G. Vijayasundaram, ed., Tata Institute Lectures on Mathematics and Physics, Springer, New York, 1979.

[21] T. P. Nguyen, E. Pauwels, E. Richard, and B. W. Suter, *Extragradient method in optimization: Convergence and complexity*, J. Optim. Theory Appl., 176 (2018), pp. 137–162.

[22] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

[23] M. V. Solodov and B. F. Svaiter, *A hybrid projection-proximal point algorithm*, J. Convex Anal., 6 (1999), pp. 59–70.

[24] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446.