Dykstra’s Splitting and an Approximate Proximal Point Algorithm for Minimizing the Sum of Convex Functions

Chin How Jeffrey Pang

Received: 20 May 2018 / Accepted: 20 March 2019 / Published online: 30 March 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
We show that Dykstra’s splitting for projecting onto the intersection of convex sets can be extended to minimize the sum of convex functions and a regularizing quadratic function. We give conditions for which convergence to the primal minimizer holds so that more than one convex function can be minimized at a time, the convex functions are not necessarily sampled in a cyclic manner, and the SHQP strategy for problems involving the intersection of more than one convex set can be applied. When the sum does not involve the regularizing quadratic function, we discuss an approximate proximal point method combined with Dykstra’s splitting to minimize this sum.

Keywords Dykstra’s splitting · Proximal point algorithm · Block coordinate minimization

Mathematics Subject Classification 90C25 · 65K05 · 68Q25 · 47J25

1 Introduction
Throughout this paper, the domains of the functions we consider are finite-dimensional Hilbert spaces. The aim of this paper is to combine Dykstra’s splitting and an approximate proximal point algorithm in order to minimize the sum of functions.
Dykstra’s Algorithm

Dykstra’s algorithm [1] solves the problem of projecting a point onto the intersection of several convex sets. This problem can be written as a minimization problem with a convex quadratic function plus the sum of indicator functions. The projection onto the intersection of these sets may be difficult, but each step of Dykstra’s algorithm requires only the projection onto one set at a time. Its convergence to a primal minimizer without constraint qualifications was established in [2]. Separately, Dykstra’s algorithm was rediscovered in [3], who noticed that it is block coordinate minimization on the dual problem and proved the convergence to a primal minimizer, but under a constraint qualification. This dual perspective was also noticed by [4], who built on [2] and used duality to prove the convergence to a primal minimizer without constraint qualifications. We shall assume that the sets have nonempty intersection. (An example of the case where we have empty intersection is discussed in [5] for example).

Dykstra’s algorithm can be made into a parallel algorithm by using the product space approach largely attributed to [6]. But this parallelization is slower than the original Dykstra’s algorithm because the dual variables are not updated in a Gauss-Seidel manner. (In other words, the dual variables are not updated with the most recent values of the other dual variables). It was also noticed in [7] (among other things) that the projections onto the sets need not be performed in a cyclic manner to achieve convergence. In [8], we studied a SHQP (supporting half-space and quadratic programming) heuristic for improving the convergence of Dykstra’s algorithm by noticing that the projection operations onto the sets generate half-spaces containing these sets, and the intersection of these half-spaces can be a better approximate of the intersection than each set alone.

Dykstra’s algorithm is related to the method of alternating projections for finding a point in the intersection more than one closed set. For more information on the various topics in Dykstra’s algorithm mentioned so far, we refer to [9–12].

Block Coordinate Minimization

For the problem of minimizing the sum of a smooth and a block separable function, one strategy is to minimize one block of the variables at a time, keeping the others fixed. This strategy is called block coordinate minimization, or alternating minimization. Nonasymptotic convergence rates of $O(1/k)$ to the optimal value were obtained for when the smooth function is not known to be strongly convex in [13,14]. We refer to these papers for more on the history of block coordinate minimization.

The smooth portion of the dual problem in Dykstra’s algorithm is a specific quadratic function, so block coordinate minimization for this problem coincides with a block coordinate proximal gradient approach in [15,16]. Convergence properties of minimizing over more than one block at a time were discussed. There is too much recent research on block coordinate minimization and block coordinate proximal gradient, so we refer the reader to the two recent references [17,18] and their references within.
Proximal Point Algorithm

The proximal point algorithm attributed to [19,20] is a method for finding minimizers of a function by iteratively minimizing the sum of the function with a quadratic regularizer. It was noticed in [21] that one can use the proximal point algorithm to minimize the sum of several convex functions by approximately solving a sequence of problems of form (2) using Dykstra’s algorithm. The rules there for moving to a new proximal center involve finding a primal feasible point that satisfies the optimality conditions approximately. But such a feasible point might not be found in a finite number of iterations when some of the functions are indicator functions, so a separate rule for moving the proximal center is needed.

Contributions of this Paper

Firstly, in Sect. 3, we extend Dykstra’s splitting for minimizing (2) so that

(A) the proof of convergence does not require constraint qualifications,
(B) the \( r \) in (2) is any number greater than or equal to 2, and
(C) \( h_i(\cdot) \) can be any closed convex function instead of the indicator function.

As mentioned earlier, [2] and [4] have features (A) and (B), [21] has (B) and (C), and [22] has (A) and (C). We are not aware of Dykstra’s splitting being proved to have features (A), (B), and (C). In addition, our analysis incorporates these features that are now rather standard in block coordinate minimization algorithms.

(D) The convex functions \( h_i(\cdot) \) are not necessarily sampled in a cyclic manner like in [7],
(E) more than one convex function \( h_i(\cdot) \) can be minimized at one time in the Dykstra’s splitting, and
(F) the SHQP strategy in [8] is applied.

The proof is largely adapted from [4]. This paper also updates the discussion of the SHQP strategy in [8] by pointing out that if the convex functions \( \delta_{C_i}^*(\cdot) \) are not necessarily sampled in a cyclic manner, then we just need one set of the form \( \tilde{C}^{m,w} \) in Algorithm 3.1 instead of multiple sets of this type as was done in [8].

Secondly, in Sect. 5, we show that one can minimize problems of form (1) where the feasible region is a compact set by combining Dykstra’s splitting on problems of the kind (2) and an approximate proximal point algorithm where the proximal center is moved once the KKT conditions are approximately satisfied. The compactness of the feasible region allows us to remove the constraint qualifications on the constraint sets for our results.

In Sect. 4, we show that if a dual minimizer exists and some processing is performed so that the dual multipliers related to the indicator functions are uniformly bounded throughout all iterations, an \( O(1/n) \) convergence of the dual problem (which leads to an \( O(1/\sqrt{n}) \) convergence to the primal minimizer) can be attained.
2 Preliminaries

Let $X$ be a finite-dimensional Hilbert space. In this paper, we aim to minimize the sum of convex functions

$$
\sum_{i=1}^{r} h_i(\cdot),
$$

where $h_i : X \to \mathbb{R} \cup \{\infty\}$ are closed proper convex functions. We refer to the natural extension of Dykstra’s algorithm for minimizing

$$
\min_{x} \frac{1}{2}\|x - x_0\|^2 + \sum_{i=1}^{r} h_i(x),
$$

where $h_i(\cdot)$ are generalized to be closed convex functions, as Dykstra’s splitting. (Dykstra’s algorithm corresponds to the case when $h_i(\cdot)$ are indicator functions of closed convex sets.) Instead of projections, one now uses proximal mappings. (See (12) for an example.) Dykstra’s splitting was studied in [21] and [23] for the case of $r \geq 2$, and they proved the convergence (to the primal minimizer) under constraint qualifications. It was also proved in [22] that Dykstra’s splitting converges for the case of $r = 2$ without constraint qualifications.

2.1 Other Methods for Minimizing the Sum of Functions

The dual ascent perspective in this paper (not just for the case of having only indicator functions in Dykstra’s algorithm) can also be traced to [24–26]. Nevertheless, these papers focus on different aspects compared to this paper.

When the constraint sets are either too big and have to be split up as the intersection of more than 1 set, or when these constraint sets are only revealed as the algorithm is run, it is beneficial to write these problems in form (1) where two or more of the $h_i(\cdot)$ are indicator functions. In such a case, as remarked in [27], the accelerated methods of [28] and further developed by [29,30] do not immediately apply (to the primal problem). We now recall other methods and observations on minimizing (1) when more than one of the functions $h_i(\cdot)$ are indicator functions and the algorithm can operate on a few of the functions $h_i(\cdot)$ at a time. As we have seen earlier, Dykstra’s algorithm is one such example.

In the case where all the functions $h_i(\cdot)$ in (1) are indicator functions, then this problem coincides with the problem of finding a point in the intersection of convex sets, which is a problem of much interest on its own (see for example [10,12,31]). We refer to this as the convex feasibility problem. The convex feasibility problem can be solved by the method of alternating projections and the Douglas–Rachford method. A discussion of the effectiveness of methods for the convex feasibility problem is [32].

Beyond the convex feasibility problem, various extensions of the subgradient method in [27,33,34] can solve problems of form (1). Another recent development is in superiorization (See, for example, [35]), where an algorithm for the convex feasibility problem is perturbed to try to reduce the value of the objective function. The
result is an algorithm that seeks feasibility at a rate comparable to algorithms for the feasibility problem, while achieving a superior objective value to what an algorithm for the feasibility problem alone would achieve. A comparison of projected subgradient methods and superiorization is given in [36].

A typical assumption on the constraint sets is that they have a Lipschitzian error bound, which is also equivalent to the stability of the intersection under perturbations. See, for example, [37–40].

Lastly, another method for minimizing (1) is the ADMM [41]. The ADMM is an effective method, but we feel that Dykstra’s splitting still has its own value. For example, as we shall see later, the different agents can minimize in any order, and convergence does not even require the existence of a dual minimizer.

We refer to survey [42] for other proximal techniques for minimizing (1).

2.2 Methodology

In Sect. 3, we prove the convergence of Dykstra’s splitting. We divide the functions into three different types, (A1)–(A3). We use the notation of [7], where they looked at Dykstra’s algorithm in an order that is not necessarily cyclic, in order to treat our case where we do the same. In [8], we had extended the proof of [2] as presented in [10]. But in this paper, where we extended to general convex functions $h_i(\cdot)$, we found it easier to extend the proof of [4] as the notation was more readily extendable. (An example is (17).) As a consequence of allowing $|S_{n,w}| > 1$ in Algorithm 3.1, we also try to make assumptions as weak as possible, which led us to Assumption 3.1(c)(d) and Proposition 3.4.

In Sect. 4, we extend the framework of block coordinate minimization in [13,14]. An additional consideration is that $|S_{n,w}| > 1$, and the data are not necessarily sampled in a cyclic manner.

In Sect. 5, we adapt the proximal point algorithm of [20] on how to shift the proximal point to solve (39) and (55). Our approach is to first look at the easier case (39) and then reduce the harder case (55) to the easier case (39).

2.3 Notation

We use “∂” to refer to either the subdifferential of a convex function, or the boundary of a set, which should be clear from context. The conjugate $\delta^*_C(\cdot)$ of the indicator function has the form $\delta^*_C(y) = \sup_{x \in C} \langle y, x \rangle$ and is also known as the support function.

3 Dykstra’s Splitting for the Sum of Convex Functions

Consider the primal problem

$$\begin{align*}
(P) \quad \alpha = \min_{x \in X} & \left[ \frac{1}{2} \|x - x_0\|^2 + \sum_{i=1}^{r_1} f_i(x) + \sum_{i=r_1+1}^{r_2} g_i(x) + \sum_{i=r_2+1}^{r} \delta_{C_i}(x) \right],
\end{align*}$$

$(3)$
where $X$ is a finite-dimensional Hilbert space, and

(A1) $f_i : X \to \mathbb{R}$ are convex functions such that $\text{dom} f_i(\cdot) = X$ for all $i \in \{1, \ldots, r_1\}$.

(A2) $g_i : X \to \mathbb{R}$ are lower semicontinuous and convex functions for all $i \in \{r_1 + 1, \ldots, r_2\}$.

(A3) $C_i$ are closed and convex subsets of $X$ for all $i \in \{r_2 + 1, \ldots, r\}$.

In this section, we generalize the proof in [4] to show that Dykstra’s splitting algorithm can be used to minimize problems of form (3).

We emphasize that for the stronger results we prove in Sect. 4, we need a stronger assumption (A2′), which we will state in Sect. 4. The condition there ensures the validity of Step 1 of the claim in Theorem (i.e., $\|z_n, w_i\| \leq M_1$ for all $i \in \{r_1 + 1, \ldots, r_2\}$).

We note that the functions $\delta_{C_i}(\cdot)$ and $f_i(\cdot)$ can be written as $g_i(\cdot)$. But as we will see later, we will treat the functions of the three types differently in Algorithm 3.1. For convenience of future discussions, let $h : X \to \mathbb{R}$ and $h_i : X \to \mathbb{R}$ be the convex functions defined by

$$h(\cdot) := \sum_{i=1}^{r} h_i(\cdot), \quad \text{and} \quad h_i(\cdot) := \begin{cases} f_i(\cdot), & \text{if } i \in \{1, \ldots, r_1\}, \\ g_i(\cdot), & \text{if } i \in \{r_1 + 1, \ldots, r_2\}, \\ \delta_{C_i}(\cdot), & \text{if } i \in \{r_2 + 1, \ldots, r\}, \end{cases}$$

so that the objective function in (3) can be written simply as $\frac{1}{2}\|x - x_0\|^2 + h(x)$.

### 3.1 Algorithm Description and Commentary

The (Fenchel) dual of problem (3) is

$$(D) \quad \beta = \max_{z \in X^r} F(z),$$

where $F : X^r \to \mathbb{R}$ is defined by

$$F(z) := -\frac{1}{2} \left\| \left( \sum_{i=1}^{r} z_i \right) - x_0 \right\|^2 - \sum_{i=1}^{r} h_i^\ast(z_i) + \frac{1}{2}\|x_0\|^2.$$  

By weak duality, we have $\beta \leq \alpha$. (Actually $\beta = \alpha$ is true; we will see that later.)

If $\tilde{C}$ is any closed and convex set such that $\tilde{C} \subset \bar{C}$, where the set $\tilde{C}$ is defined by

$$\tilde{C} := \left[ \cap_{i=r_2+1}^{r} C_i \right] \cap \left[ \cap_{i=r_1+1}^{r} \text{cl dom } g_i(\cdot) \right],$$

then problem (3) has the same (primal) minimizer as

$$(P_{\tilde{C}}) \quad \alpha = \min_{x \in X} \left[ \frac{1}{2}\|x - x_0\|^2 + \sum_{i=1}^{r} h_i(x) + \delta_{\tilde{C}}(x) \right].$$
The dual of \((P_\tilde{C})\) is

\[
(D_\tilde{C}) \quad \beta = \max_{z \in \mathbb{R}^{r+1}} F_\tilde{C}(z),
\]

where \(F_\tilde{C} : \mathbb{R}^{r+1} \rightarrow \mathbb{R}\) is defined by

\[
F_\tilde{C}(z) := -\frac{1}{2} \left\| \left( \sum_{i=1}^{r+1} z_i \right) - x_0 \right\|^2 - \sum_{i=1}^{r} h_i^*(z_i) - \delta_{\tilde{C}}^*(z_{r+1}) + \frac{1}{2} \| x_0 \|^2. \tag{8}
\]

As detailed in [8], this observation leads us to construct a set \(\tilde{C}_{n,w}\) that changes in each iteration of our extended Dykstra’s algorithm in Algorithm 3.1 on the following page.

We list some observations of Algorithm 3.1. The choice of \(S_{n,w}\) in line 6 of Algorithm 3.1 allows for more than one block of \(z\) to be minimized in (9). If \(\bar{w} = r + 1\), the sets \(S_{n,w}\) are chosen to be \(\{w\}\), and \(r_1 = r_2 = 0\), then Algorithm 3.1 reduces to the extended Dykstra’s algorithm that was discussed in [8]. In general, \(\bar{w}\) is chosen to be large enough so that Assumption 3.1(b) can be satisfied.

**Remark 3.1** (Choice of \(H_{n,w}\)) An easy choice for \(H_{n,w}\) in line 11 of Algorithm 3.1 is to choose a half-space with outward normal \(z_{r+1}^{n,w}\) that supports the set \(\tilde{C}_{n,w}\). Another example of \(H_{n,w}\) is the intersection of the half-space mentioned earlier with a small number of half-spaces containing \(\tilde{C}\) defined in (6) that will allow \(H_{n,w}\) to approximate \(\tilde{C}\) well. At the start of Algorithm 3.1, we set \(H_{1,0} = \mathbb{R}^r\). Since we want \(\delta_{H_{1,0}}(z_{r+1}^{1,0})\) to be finite, we set \(z_{r+1}^{1,0} = 0\).

We have the following identities to simplify notation:

\[
v_{n,w} := \sum_{j=1}^{r+1} z_{n,w} j \tag{11a}
\]

and

\[
x_{n,w} := x_0 - v_{n,w}. \tag{11b}
\]

**Proposition 3.1** For all \(i \in S_{n,w}\), we have

(a) \(-x_{n,w} + \partial h_i^*(z_{i,n,w}) \ni 0, \)
(b) \(-z_i^{n,w} + \partial h_i(x_{n,w}) \ni 0, \)
(c) \(h_i(x_{n,w}) + h_i^*(z_{i,n,w}) = (x_{n,w}, z_{i,n,w}).\)

**Proof** By taking the optimality conditions in (9) with respect to \(z_i\) for \(i \in S_{n,w}\), we deduce (a). The equivalences of (a), (b), and (c) are standard. \(\square\)

Dykstra’s algorithm is traditionally written in terms of solving for the primal variable \(x\). For completeness, we show the equivalence between (9) and the primal minimization problem.
Algorithm 3.1 (Extended Dykstra’s algorithm) Consider the problem (3) along with the associated problems (4), (7), and (8).

Set some number $M \in \mathbb{R}_+ \cup \{\infty\}$, and let $\tilde{w}$ be a positive integer. Our extended Dykstra’s algorithm is as follows:

01 Define the set $H^{1.0}$ to be $H^{1.0} = X$.
02 Let $z^{1.0} \in X^{r+1}$ be the starting dual vector for (8), and let $z^{1.0}_{r+1} = 0$.
03 Let $x^{1.0} = x_0 - \sum_{i=1}^{r+1} z^{1.0}_i$.
04 For $n = 1, 2, \ldots$
05 For $w = 1, 2, \ldots, \tilde{w}$
06 Choose a subset $S_n, w \subset \{1, \ldots, r + 1\}$.
07 If $r + 1 \in S_n, w$, then

**Dual decrease with SHQP steps**

08 Choose $\tilde{C}^{n, w}$ to be any set such that $\tilde{C} \subset \tilde{C}^{n, w} \subset H^{n, w-1}$.
09 Let $z^{n, w}_i = z^{n, w-1}_i$ for all $i \notin S_n, w$.
10 Let $(z^{n, w}_i)_{i \in S_n, w}$ be defined through

$$z^{n, w}_i = \arg \max_{z \in X^{r+1}} \frac{1}{2} \left\| \left( \sum_{i \in S_n, w} z_i + \sum_{i \notin S_n, w} z^{n, w-1}_i \right) - x_0 \right\|^2$$

$$- \sum_{i=1}^{r} h_i^*(z_i) - \delta^*_{\tilde{C}^{n, w}}(z_{r+1}) + \frac{1}{2}\|x_0\|^2.$$

s.t. $z^{n, w}_i = z^{n, w-1}_i$ for all $i \notin S_n, w$.

11 Let $H^{n, w}$ be a set such that $\delta^*_{\tilde{C}^{n, w}}(z^{n, w}_i) = \delta^*_{H^{n, w}}(z^{n, w}_i) + 1$ and $\tilde{C} \subset H^{n, w}$.
12 Else

**Dual decrease**

13 Let $z^{n, w}_i = z^{n, w-1}_i$ for all $i \notin S_n, w$.
14 Define $(z^{n, w}_i)_{i \in S_n, w}$ through (9), except with $\delta^*_{\tilde{C}^{n, w}}(z_{r+1})$ omitted.
15 Let $H^{n, w}$ be $H^{n, w-1}$.
16 End If
17 End For

**Aggregating variables**

18 Find $z^{n+1, 0} \in X^{r+1}$ and $H^{n+1, 0} \supset \tilde{C}$ such that

$$z^{n+1, 0}_i = z^{n, \tilde{w}}_i \quad \text{for all} \quad i \in \{1, \ldots, r_2\}$$

$$\sum_{i=1}^{r+1} z^{n+1, 0}_i = \sum_{i=1}^{r+1} z^{n, \tilde{w}}_i$$

$$\|z^{n+1, 0}_i\| \leq M \quad \text{for all} \quad i \in \{r_2 + 1, \ldots, r\}$$

$$\sum_{i=r_2+1}^{r} \delta^*_{\tilde{C}_i}(z^{n+1, 0}_i) + \delta^*_{H^{n+1, 0}}(z^{n+1, 0}_{r+1}) \leq \sum_{i=r_2+1}^{r} \delta^*_{\tilde{C}_i}(z^{n, \tilde{w}}_i) + \delta^*_{H^{n, \tilde{w}}}(z^{n, \tilde{w}}_{r+1})$$

$$\sum_{i=1}^{r+1} \|z^{n+1, 0}_i\| \leq \sum_{i=1}^{r+1} \|z^{n, \tilde{w}}_i\|.$$

19 End For
Proposition 3.2 [On solving (9)] If a minimizer $z^{n,w}$ for (9) exists, then the $x^{n,w}$ in (11b) satisfies

$$x^{n,w} = \arg \min_{x \in X} \sum_{i \in S_{n,w}} h_i(x) + \frac{1}{2} \left\| x - \left( x_0 - \sum_{i \in S_{n,w}} z^{n,w}_i \right) \right\|^2. \quad (12)$$

Conversely, if $x^{n,w}$ solves (12) with the dual variables $\{\tilde{z}^{n,w}_i\}_{i \in S_{n,w}}$ satisfying

$$\tilde{z}^{n,w}_i \in \partial h_i(x^{n,w}) \text{ and } x^{n,w} - x_0 + \sum_{i \notin S_{n,w}} z^{n,w}_i + \sum_{i \in S_{n,w}} \tilde{z}^{n,w}_i = 0, \quad (13)$$

then $\{\tilde{z}^{n,w}_i\}_{i \in S_{n,w}}$ solves (9).

**Proof** For the first part, note that

$$\partial \left( h + \frac{1}{2} \left\| - (x_0 - \sum_{i \notin S_{n,w}} z^{n,w}_i) \right\|^2 \right)(x^{n,w})$$

$$\supseteq \sum_{i \in S_{n,w}} \partial h_i(x^{n,w}) + \left[ x^{n,w} - (x_0 - \sum_{i \in S_{n,w}} z^{n,w}_i) \right] \quad \text{Prop 3.1(b)}$$

$$\sum_{i \in S_{n,w}} z^{n,w}_i + x^{n,w} - x_0 + \sum_{i \notin S_{n,w}} z^{n,w}_i \equiv 0.$$

For the second part, note that we have $x^{n,w} \in \partial h^*_i(\tilde{z}^{n,w}_i)$ from the first part of (13), while the second part of (13) implies that 0 lies in the subdifferential of the objective function in (9). \qed

Remark 3.2 [Information needed to calculate (9)] We note that in (9), one only needs to have knowledge of the variables $v^{n,w-1}$ and $z^{n,w-1}_i$ for $i \in S_{n,w}$. Thus, Dykstra’s splitting may be suitable for problems where the communication costs are high compared to the costs of solving the proximal problems.

Remark 3.3 (On line 18 of Algorithm 3.1) If $M = \infty$ in Algorithm 3.1, then $z^{n+1,0}$ and $H^{n+1,0}$ can be set to be $\tilde{z}^{n,w}$ and $\tilde{H}^{n,w}$, respectively. We had to add this line to Algorithm 3.1 because the boundedness condition (10c) is necessary for our $O(1/n)$ convergence result in Sect. 4. This detail can be skipped for the discussions in this section and Sect. 5.

We need the following fact before we discuss how to find $z^{n+1,0}$ and $H^{n+1,0}$ satisfying (10).

Fact 3.1 (Aggregating half-spaces) Consider two half-spaces, say $H_1$ and $H_2$, which have (outward) normals $z_1$ and $z_2$. Assume that $\{z_1, z_2\}$ are linearly independent. Construct a third half-space $H_3$ with normal $z_1 + z_2$ such that $H_3 \supset H_1 \cap H_2$ and
\[ \partial H_3 \cap [H_1 \cap H_2] \neq \emptyset. \] Let \( x \) be any point on \( \partial H_1 \cap \partial H_2 \). We see that \( x \in \partial H_3 \). We have
\[
\delta_{H_1}^*(z_1) + \delta_{H_2}^*(z_2) = (z_1, x) + (z_2, x) = (z_1 + z_2, x) = \delta_{H_3}^*(z_1 + z_2).
\]

If \( \{z_1, z_2\} \) is linearly dependent instead, then \( H_3 := H_1 \cap H_2 \) is a half-space, and
\[
\delta_{H_3}^*(z_1 + z_2) \leq \delta_{H_1}^*(z_1) + \delta_{H_2}^*(z_2).
\]

Moreover, the inequality is strict if, for example, \( z_1 \neq 0, z_2 \neq 0 \) and \( H_1 \not\subseteq H_2 \). This fact can be generalized for more than two half-spaces.

We state some notation necessary for later. For any \( i \in \{1, \ldots, r + 1\} \) and \( n \in \{1, 2, \ldots\} \), let \( p(n, i) \) be
\[
p(n, i) = \max\{m : m \leq \bar{w}, i \in S_{n,m}\}.
\]

In other words, \( p(n, i) \) is the index \( m \) such that \( i \in S_{n,m} \) but \( i \notin S_{n,k} \) for all \( k \in \{m + 1, \ldots, \bar{w}\} \). It follows from lines 9 and 13 of Algorithm 3.1 that
\[
z_i^{n,p(n,i)} = z_i^{n,p(n,i)+1} = \ldots = z_i^{n,\bar{w}}.
\]

We now show one way to find \( z_i^{n+1,0} \) and \( H_i^{n+1,0} \) satisfying (10d).

**Proposition 3.3** [On satisfying (10d)] For all \( i \in \{r_2 + 1, \ldots, r\} \), set \( z_i^{n+1,0} \) to be \( \alpha_i z_i^{n,\bar{w}} \) where \( \alpha_i \) is a number in \([0, 1]\), so that (10c) is satisfied. Then,
\[
z_{r+1}^{n+1,0} = \sum_{i=r+1}^{r_2+1} z_i^{n,\bar{w}} - \sum_{i=r_2+1}^r z_i^{n+1,0} = z_{r+1}^{n,\bar{w}} + \sum_{i=r_2+1}^r (1 - \alpha_i) z_i^{n,\bar{w}}. \quad (15)
\]

For \( i \in \{r_2 + 1, \ldots, r\} \), recall that by the construction of \( z_i^{n,p(n,i)} \) in (9) and (12), the condition (13) implies that \( \tilde{H}_i^{n,i} \supseteq C_i \), where the half-space \( \tilde{H}_i^{n,i} \) is defined by
\[
\tilde{H}_i^{n,i} := \{x : (x - x_i^{n,p(n,i)}, z_i^{n,p(n,i)}) \leq 0\} \\
\overset{(14)}{=} \{x : (x - x_i^{n,p(n,i)}, z_i^{n,\bar{w}}) \leq 0\}. \quad (16)
\]

and \( x_i^{n,\bar{w}} \) is as defined in (12). We can check that
\[
\delta_{C_i}^*(\alpha z_i^{n,\bar{w}}) = \delta_{\tilde{H}_i^{n,i}}^*(\alpha z_i^{n,\bar{w}}) \text{ for any } \alpha \geq 0 \text{ and } i \in \{r_2 + 1, \ldots, r\}.
\]

Let \( I_n \subset \{r_2 + 1, \ldots, r\} \) be the set of indices \( i \) such that \( z_i^{n+1,0} \neq z_i^{n,\bar{w}} \). Let \( H_i^{n+1,0} \) be the half-space with outward normal \( z_{r+1}^{n+1,0} \) such that
\[ H_{n+1,0} \supset H_{n,\bar{w}} \cap \bigcap_{i \in I_n} \tilde{H}_{n,i} \]

and \[ \partial H_{n+1,0} \cap [H_{n,\bar{w}} \cap \bigcap_{i \in I_n} \tilde{H}_{n,i}] \neq \emptyset. \]

Then, (10d) is satisfied. Furthermore, (10d) is actually an equality if the normals \( \{z_i^{n,\bar{w}}\}_{i \in I_n \cup \{r+1\}} \) are linearly independent.

**Proof** The conclusion can be deduced from Fact 3.1. \( \square \)

One can check that the construction in Proposition 3.3 also leads to the conditions in (10). In particular, (10e) can be inferred from (15).

The other items in (10) are clear.

### 3.2 Convergence of Algorithm 3.1

We now prove the convergence of Algorithm 3.1. We first list assumptions that will ensure convergence to the primal minimizer.

**Assumption 3.1** We make a few assumptions on Algorithm 3.1:

(a) The objective value \( \alpha \) in (3) is a finite number.

(b) For all \( \{1, \ldots, r+1\} \), the sets \( S_{n,w} \) are chosen such that for all \( n, \bigcup_{w=1}^{\bar{w}} S_{n,w} = \{1, \ldots, r+1\} \).

(c) There are constants \( A \) and \( B \) such that \( \sum_{i=1}^{r+1} \|z_i^{n,\bar{w}}\| \leq \sqrt{n}A + B \) for all \( n \).

(d) Minimizers of (9) can be obtained in each step.

We give a brief commentary on Assumption 3.1. Assumption 3.1(a) together with the strong convexity of the primal problem says that (3) is feasible and a unique primal minimizer exists. As we will see later, the structure of the functions \( f_i(\cdot) \) for \( i \in \{1, \ldots, r_1\} \) implies that \( z_i^{n,w} \) is uniformly bounded for all \( i \in \{1, \ldots, r_1\} \). In Proposition 3.4, we shall introduce a condition on the choice of \( S_{n,w} \) that will ensure that Assumption 3.1(c) is satisfied.

We follow the proof in [4] to show that \( \lim_{n \to \infty} x^{n,\bar{w}} \) exists and is the minimizer of (P).
For any \( x \in X \) and \( z \in X^{r+1} \), the analogue of [4, (8)] is

\[
\frac{1}{2} \| x_0 - x \|^2 + \sum_{i=1}^{r} h_i(x) + \delta \tilde{C}(x) - F_{\tilde{C}}(z_1, \ldots, z_r, z_{r+1})
\]

\[= \frac{1}{2} \| x_0 - x \|^2 + \sum_{i=1}^{r} [h_i(x) + h_i^*(z_i)] - \left( x_0, \sum_{i=1}^{r+1} z_i \right) + \frac{1}{2} \left\| \sum_{i=1}^{r+1} z_i \right\|^2 + \delta \tilde{C}(x) + \delta \tilde{C}^*(z_{r+1})
\]

\[
\geq \frac{1}{2} \| x_0 - x \|^2 + \sum_{i=1}^{r+1} \langle x, z_i \rangle - \left( x_0, \sum_{i=1}^{r+1} z_i \right) + \frac{1}{2} \left\| \sum_{i=1}^{r+1} z_i \right\|^2
\]

\[
= \frac{1}{2} \left\| x_0 - x - \sum_{i=1}^{r+1} z_i \right\|^2 \geq 0.
\] \quad (17)

The theorem below generalizes [4, Theorem 1] for the setting (3).

**Theorem 3.1** Suppose Assumption 3.1 holds. For the sequence

\[
\{z^{n,w}\}_{1 \leq n < \infty}^{0 \leq w \leq \tilde{w}} \subset X^{r+1}
\]

generated by Algorithm 3.1 and the sequences

\[
\{v^{n,w}\}_{1 \leq n < \infty}^{0 \leq w \leq \tilde{w}} \subset X \text{ and } \{x^{n,w}\}_{1 \leq n < \infty}^{0 \leq w \leq \tilde{w}} \subset X
\]
deduced from (11), we have:

(i) The sum \( \sum_{n=1}^{\infty} \sum_{w=1}^{\tilde{w}} \| v^{n,w} - v^{n,w-1} \|^2 \) is finite and \( \{F_{H^{n,w}}(z^{n,w})\}_{n=1}^{\infty} \) is non-decreasing.

(ii) There is a constant \( \theta \) such that \( \| v^{n,w} \|^2 \leq \theta \) for all \( w \in \{1, \ldots, \tilde{w}\} \) and \( n \in \mathbb{N} \).

(iii) There exists a subsequence \( \{v^{n_k,w}\}_{k=1}^{\infty} \) of \( \{v^{n,w}\}_{n=1}^{\infty} \) which converges to some \( v^* \in X \) and that

\[
\lim_{k \to \infty} \langle v^{n_k,w} - v^{n_k,p(n_k,i)}, z^{n_k,w}_i \rangle = 0 \text{ for all } i \in \{1, \ldots, r+1\}.
\]

(iv) For the \( v^* \) in (iii), \( x_0 - v^* \) is the minimizer of the primal problem (P) and

\[
\lim_{k \to \infty} F_{H^{n_k,w}}(z^{n_k,w}) = \frac{1}{2} \| v^* \|^2 + \sum_{i=1}^{r} h_i(x_0 - v^*).
\]

The properties (i)–(iv) in turn imply that \( \lim_{n \to \infty} x^{n,w} \) exists, and \( x_0 - v^* \) is the primal minimizer of (3).

**Proof** We first show that (i)–(iv) imply the final assertion. For all \( n \in \mathbb{N} \) we have, from weak duality,

\[
F_{H^{n,w}}(z^{n,w}) \leq \beta \leq \alpha \leq \frac{1}{2} \| x_0 - (x_0 - v^*) \|^2 + \sum_{i=1}^{r} h_i(x_0 - v^*),
\] \quad (18)
hence $\beta = \alpha = \frac{1}{2} \|x_0 - (x_0 - v^*)\|^2 + h(x_0 - v^*)$, and that $x_0 - v^*$ equals $\arg\min_x h(x) + \frac{1}{2} \|x - x_0\|^2$. Since the values $\{F_{H^n, \tilde{w}}(z^n, \tilde{w})\}_{n=1}^\infty$ are nondecreasing in $n$, we have

$$\lim_{n \to \infty} F_{H^n, \tilde{w}}(z^n, \tilde{w}) = \frac{1}{2} \|x_0 - (x_0 - v^*)\|^2 + \sum_{i=1}^r h_i(x_0 - v^*),$$

and (substituting $x = x_0 - v^*$ in (17))

$$\frac{1}{2} \|x_0 - (x_0 - v^*)\|^2 + h(x_0 - v^*) - F_{H^n, \tilde{w}}(z^n, \tilde{w}) \geq \frac{1}{2} \|x_0 - (x_0 - v^*) - v^n, \tilde{w}\|^2$$

(17),(11a)

$$\leq \frac{1}{2} \|x_0 - (x_0 - v^*) - v^n, \tilde{w}\|^2$$

(11b)

$$\leq \frac{1}{2} \|x^n, \tilde{w} - (x_0 - v^*)\|^2.$$

Hence, $\lim_{n \to \infty} x^n, \tilde{w}$ is the minimizer in (P).

It remains to prove assertions (i)–(iv).

**Proof of (i):** We note that if $r + 1 \in S_{n, w}$, then

$$F_{H^n, w-1}(z^n, w-1) \leq F_{\tilde{C}, n, w}(z^n, w-1) \leq F_{\tilde{C}, n, w}(z^n, w) \leq F_{H^n, w}(z^n, w) - \frac{1}{2} \|v^n, w - v^n, w-1\|^2$$

(9),(11a)

$$= F_{H^n, w}(z^n, w) - \frac{1}{2} \|v^n, w - v^n, w-1\|^2.$$ (19)

The first inequality comes from the fact that since $\tilde{C}^{n, w} \subset H^{n, w-1}$ (from line 8 of Algorithm 3.1), then $\delta^{n, w} \leq \delta^{n, w-1}$. The second inequality comes from the fact that $\{z_i^{n, w}\}_{i \in S_{n, w}}$ is a minimizer of the mapping

$$\{z_i\}_{i \in S_{n, w}} \mapsto \sum_{i \in S_{n, w}} h_i(z_i) + \frac{1}{2} \left\| \left( \sum_{i \in S_{n, w}} z_i \right) - \left( x_0 - \sum_{i \notin S_{n, w}} z_i^{n, w} \right) \right\|^2,$$

with the $\frac{1}{2} \|v^n, w - v^n, w-1\|^2$ arising from the quadratic term.

When $r + 1 \notin S_{n, w}$, then we can make use of the fact that $z_{r+1}^{n, w} = z_{r+1}^{n, w-1}$ and $\tilde{C}^{n, w} = H^{n, w-1} = H^{n, w}$ to see that the inequality (19) carries through as well.

Recall that through (10d), $F_{H^{n+1, 0}}(z^{n+1, 0}) \geq F_{H^n, \tilde{w}}(z^n, \tilde{w})$. Combining (19) over all $m \in \{1, \ldots, n\}$ and $w \in \{1, \ldots, \tilde{w}\}$, we have

$$F_{H^{1, 0}}(z^{1, 0}) + \sum_{m=1}^n \sum_{w=1}^\tilde{w} \|v^{m, w} - v^{m, w-1}\|^2 \leq F_{H^n, \tilde{w}}(z^n, \tilde{w}).$$

Next, $F_{H^n, \tilde{w}}(z^n, \tilde{w}) \leq \alpha$ by weak duality. The proof of the claim is complete.
Proof of (ii): Substituting $x$ in (17) to be the primal minimizer $x^*$ and $z$ to be $z_{n,w}$, we have

$$
\frac{1}{2} \|x_0 - x^*\|^2 + \sum_{i=1}^{r} h_i(x^*) - F_{H^{1,0}}(z^{1,0})
\geq \frac{1}{2} \|x_0 - x^*\|^2 + \sum_{i=1}^{r} h_i(x^*) - F_{H^{n,w}}(z_{n,w})
\geq \frac{1}{2} \|x_0 - x^*\|^2 - \sum_{i=1}^{r} z_{n,w} \parallel z_i \parallel^2 \geq \frac{1}{2} \|x_0 - x^* - v_{n,w}\|^2.
$$

The conclusion is immediate.

Proof of (iii): We first make use of the technique in [11, Lemma 29.1] (which is in turn largely attributed to [2]) to show that

$$
\liminf_{n \to \infty} \left( \sum_{w=1}^{\bar{w}} \parallel v_{n,w} - v_{n,w-1}\parallel \right) \sqrt{n} = 0.
$$

Seeking a contradiction, suppose instead that there is an $\epsilon > 0$ and $\bar{n} > 0$ such that if $n > \bar{n}$, then \( \left( \sum_{w=1}^{\bar{w}} \parallel v_{n,w} - v_{n,w-1}\parallel \right) \sqrt{n} > \epsilon \). By the Cauchy–Schwarz inequality, we have $\epsilon^2 \frac{n}{\bar{w}} < \left( \sum_{w=1}^{\bar{w}} \parallel v_{n,w} - v_{n,w-1}\parallel \right)^2 \leq \bar{w} \sum_{w=1}^{\bar{w}} \parallel v_{n,w} - v_{n,w-1}\parallel^2$. This contradicts the earlier claim in (i) that $\sum_{n=1}^{\infty} \sum_{w=1}^{\bar{w}} \parallel v_{n,w} - v_{n,w-1}\parallel^2$ is finite.

Next, we recall Assumption 3.1(c) that there are constants $A$ and $B$ such that $\sum_{i=1}^{r+1} \parallel z_{i}^{n,w}\parallel \leq A \sqrt{n} + B$ for all $n$. Through (20), we find a sequence $\{n_k\}_{k=1}^{\infty}$ such that $\lim_{k \to \infty} \left( \sum_{w=1}^{\bar{w}} \parallel v_{n_k,w} - v_{n_k,w-1}\parallel \right) \sqrt{n_k} = 0$. Thus,

$$
\lim_{k \to \infty} \left( \sum_{w=1}^{\bar{w}} \parallel v_{n_k,w} - v_{n_k,w-1}\parallel \parallel z_{i}^{n_k,w}\parallel \right) = 0 \text{ for all } i \in \{1, \ldots, r + 1\}.
$$

Moreover,

$$
|\langle v_{n_k,w} - v_{n_k,p(n_k,i)}, z_{i}^{n_k,w} \rangle| \leq \parallel v_{n_k,w} - v_{n_k,p(n_k,i)}\parallel \parallel z_{i}^{n_k,w}\parallel \leq \left( \sum_{w=1}^{\bar{w}} \parallel v_{n_k,w} - v_{n_k,w-1}\parallel \right) \parallel z_{i}^{n_k,w}\parallel.
$$

By (ii), there exists a further subsequence of $\{v_{n_k,w}\}_{k=1}^{\infty}$ which converges to some $v^* \in X$. Combining (21) and (22) gives (iii).
Proof of (iv): From earlier results, we obtain

\[
- \sum_{i=1}^{r} h_i(x_0 - v^*) - \delta_{H^{n_k,\bar{\omega}}}(x_0 - v^*)
\]

(17) \[
= \frac{1}{2} \| x_0 - (x_0 - v^*) \|^2 - F_{H^{n_k,\bar{\omega}}}(z^{n_k,\bar{\omega}})
\]

Alg 3.1 line 15 \[
= \frac{1}{2} \| x_0 - (x_0 - v^*) \|^2 - F_{H^{n_k,p(n_k,i)}}(z^{n_k,\bar{\omega}})
\]

(8),(14) \[
= \frac{1}{2} \| v^* \|^2 + \sum_{i=1}^{r} h_i^*(z_i^{n_k,p(n_k,i)}) + \delta_{H^{n_k,p(n_k,i)}}(z_i^{n_k,\bar{\omega}})
\]

Prop 3.1(c),i \in S_{n,p(n,i)} \[
= \frac{1}{2} \| v^* \|^2 + \sum_{i=1}^{r+1} (x_0 - v^{n_k,p(n_k,i)}, z_i^{n_k,\bar{\omega}})
\]

(14) \[
= \frac{1}{2} \| v^* \|^2 - \sum_{i=1}^{r+1} (x_0 - v^{n_k,p(n_k,i)}, z_i^{n_k,\bar{\omega}})
\]

- \sum_{i=1}^{r} h_i(x_0 - v^{n_k,p(n_k,i)}) - \langle x_0, v^{n_k,\bar{\omega}} \rangle + \frac{1}{2} \| v^{n_k,\bar{\omega}} \|^2

(11a) \[
= \frac{1}{2} \| v^* \|^2 - \frac{1}{2} \| v^{n_k,\bar{\omega}} \|^2 - \sum_{i=1}^{r+1} (v^{n_k,p(n_k,i)} - v^{n_k,\bar{\omega}}, z_i^{n_k,\bar{\omega}})
\]

- \sum_{i=1}^{r} h_i(x_0 - v^{n_k,p(n_k,i)})

(23)

Since \( \lim_{k \to \infty} v^{n_k,\bar{\omega}} = v^* \), we have \( \lim_{k \to \infty} \frac{1}{2} \| v^* \|^2 - \frac{1}{2} \| v^{n_k,\bar{\omega}} \|^2 = 0 \). The term \( \sum_{i=1}^{r+1} (v^{n_k,p(n_k,i)} - v^{n_k,\bar{\omega}}, z_i^{n_k,\bar{\omega}}) \) converges to 0 by (iii). Next, recall from (12) that \( x_0 - v^{n_k,p(n_k,i)} \in C_i \). Recall from the end of the proof of (iii) that \( x_0 - v^* \) equals \( \lim_{k \to \infty} x_0 - v^{n_k,p(n_k,i)} \), so \( x_0 - v^* \in C_i \). Hence \( x_0 - v^* \in \cap_{i=r+1}^{r+1} C_i \). Since \( H^{n_k,\bar{\omega}} \) was designed so that we have \( \cap_{i=r+1}^{r+1} C_i \subseteq H^{n_k,\bar{\omega}} \), we then have \( x_0 - v^* \in H^{n_k,\bar{\omega}} \), so \( \delta_{H^{n_k,\bar{\omega}}}(x_0 - v^*) = 0 \). Lastly, by the lower semicontinuity of \( h_i(\cdot) \), we have

\[
- \lim_{k \to \infty} \sum_{i=1}^{r} h_i(x_0 - v^{n_k,p(n_k,i)}) \leq - \sum_{i=1}^{r} h_i(x_0 - v^*)
\]

Therefore, (23) becomes an equation in the limit, which leads to

\[
\lim_{k \to \infty} F_{H^{n_k,\bar{\omega}}}(z^{n_k,\bar{\omega}}) = \frac{1}{2} \| v^* \|^2 + \sum_{i=1}^{r} h_i(x_0 - v^*).
\]

\( \square \)

We now show some reasonable conditions that guarantee Assumption 3.1(c).
Proposition 3.4 [Satisfying Assumption 3.1(c)] Assumption 3.1(c) is satisfied when all of the following conditions on \( S_n, w \) hold:

1. There are only finitely many \( S_n, w \) for which \( S_n, w \cap \{ r_1 + 1, \ldots, r + 1 \} \) contains more than one element.
2. There are constants \( M_1 > 0 \) and \( M_2 > 0 \) such that the size of the set \( \left\{ (m, w) : m \leq n, w \in \{ 1, \ldots, \bar{w} \}, |S_{m, w}| > 1 \right\} \) is bounded by \( M_1 \sqrt{n} + M_2 \) for all \( n \).

Proof We only need to prove this result for when only condition (2) holds and \( S_n \cap \{ r_1 + 1, \ldots, r + 1 \} \) always contains at most one element. We have

\[
\sum_{i=1}^{r+1} \| z_i^{n, \bar{w}} \| \leq \sum_{i=1}^{r+1} \| z_i^{0} \| + \sum_{i=1}^{r+1} \| \bar{w} z_i^{n, w} - z_i^{n, w-1} \| \\
\leq \sum_{i=1}^{r+1} \| z_i^{n-1, \bar{w}} \| + \sum_{i=1}^{r+1} \| z_i^{n, w} - z_i^{n, w-1} \|. \tag{24}
\]

Hence,

\[
\sum_{i=1}^{r+1} \| z_i^{n, \bar{w}} \| \leq \sum_{i=1}^{r+1} \| z_i^{1, \bar{w}} \| + \sum_{m=1}^{n} \sum_{i=1}^{r+1} \| \bar{w} z_i^{n, w} - z_i^{n, w-1} \|. \tag{25}
\]

So it suffices to show that there are numbers \( A' \) and \( B' \) such that

\[
\sum_{m=1}^{n} \sum_{i=1}^{r+1} \| \bar{w} z_i^{m, w} - z_i^{m, w-1} \| \leq A' \sqrt{n} + B'. \tag{25}
\]

The sum of the left-hand side of (25) can be written as

\[
\sum_{(m, w) \in \bar{S}_{n,1}} \sum_{i=1}^{r+1} \| z_i^{m, w} - z_i^{m, w-1} \| + \sum_{(m, w) \in \bar{S}_{n,2}} \sum_{i=1}^{r+1} \| z_i^{m, w} - z_i^{m, w-1} \|, \tag{26}
\]

where

\[
\bar{S}_{n,1} = \left\{ (m, w) : |S_{m, w}| = 1, m \leq n, w \in \{ 1, \ldots, \bar{w} \} \right\}, \tag{27a}
\]

and \( \bar{S}_{n,2} = \left\{ (m, w) : |S_{m, w}| > 1, m \leq n, w \in \{ 1, \ldots, \bar{w} \} \right\}. \tag{27b} \]
First, there is a constant $M_3$ such that

$$
\sum_{(m, w) \in \bar{S}_{n, 1}} \sum_{i=1}^{r+1} \| z_i^{m, w} - z_i^{m, w-1} \|
$$

is $1$ in (27a), (11a)

$$
\sum_{(m, w) \in \bar{S}_{n, 1}} \| v_m, w - v_m, w-1 \| \leq \sqrt{\bar{w} n} \left( \sum_{w=1}^{\bar{w}} \sum_{m=1}^{n} \| v_m, w - v_m, w-1 \|^2 \right)
$$

Thm 3.1(i)

$$
\leq \sqrt{n} M_3.
$$

(28)

Next, we estimate the second sum in (26). For each $(m, w) \in \bar{S}_{n, 2}$, by condition (1), there is a unique $i_{m, w} \in S_{m, w} \cap \{ r_1 + 1, \ldots, r + 1 \}$. We have

$$
z_{i_{m, w}}^{m, w} - z_{i_{m, w}}^{m, w-1} = v_{i_{m, w}}^{m, w} - v_{i_{m, w}}^{m, w-1} - \sum_{j \in S_{m, w} \backslash \{i_{m, w}\}} (z_j^{m, w} - z_j^{m, w-1}).
$$

(29)

For each $j \in S_{m, w} \backslash \{i_{m, w}\}$, we have

$$
z_j^{m, w} \in \partial f_j(x_m, w) \Rightarrow \partial f_j(x_0 - v_m, w).
$$

Proposition 3.1(b)

Together with the fact that $v^{m, w}$ is bounded from Theorem 3.1(ii) and the fact that $f_j(\cdot)$ are Lipschitz on bounded domains, we deduce that $z_j^{m, w}$ and $z_j^{m, w-1}$ are bounded for all $j \in \{1, \ldots, r\}$ by standard convex analysis. Since $S_{m, w} \backslash \{i_{m, w}\} \subset \{1, \ldots, r\}$, every term on the right-hand side of (29) is bounded, so there is a constant $M_4 > 0$ such that $\| z_i^{m} - z_i^{m-1} \| \leq M_4$. Therefore, condition (2) implies

$$
\sum_{(m, w) \in \bar{S}_{n, 2}} \sum_{i=1}^{r+1} \| z_i^{m, w} - z_i^{m, w-1} \| \leq M_4(M_1 \sqrt{n} + M_2).
$$

(30)

Combining (28) and (30) into (26) gives the conclusion we need.

$\square$

### 4 $O(1/n)$ Convergence when a Dual Minimizer Exists

In this section, we show that for problem (3), if Algorithm 3.1 is applied with some finite $M$ and a minimizer for the dual problem exists, then the rate of convergence of the dual objective function is $O(1/n)$, which leads to the $O(1/\sqrt{n})$ rate of convergence to the primal minimizer.
We recall a lemma on the convergence rates of sequences.

**Lemma 4.1** (Sequence convergence rate) Let $\alpha > 0$. Suppose the sequence of non-negative numbers $\{a_k\}_{k=0}^{\infty}$ is such that

$$a_k \geq a_{k+1} + a_k^2 a_{k+1}^2$$

for all $k \in \{1, 2, \ldots\}$.

1. [13, Lemma 6.2] If furthermore, $a_1 \leq \frac{1.5}{\alpha}$ and $a_2 \leq \frac{1.5}{2\alpha}$, then

$$a_k \leq \frac{1.5}{\alpha} a_k$$

for all $k \in \{1, 2, \ldots\}$.

2. [14, Lemma 3.8] For any $k \geq 2$,

$$a_k \leq \max \left\{ \left(\frac{1}{2}\right)^{(k-1)/2} a_0, \frac{4}{\alpha k} \right\}.$$

In addition, for any $\epsilon > 0$, if

$$k \geq \max \left\{ \frac{2}{\ln(2)} [\ln(a_0) + \ln(1/\epsilon)], \frac{4}{\alpha \epsilon} \right\} + 1,$$

then $a_n \leq \epsilon$.

Instead of condition (A2) after (3), we assume a stronger condition on $g(\cdot)$:

(A2’) $g_i : X \to \mathbb{R}$ are convex functions such that $\text{dom} g_i(\cdot)$ are open sets for all $i \in \{r_1 + 1, \ldots, r_2\}$.

In other words, the functions $g_i(\cdot)$ are such that if $\lim_{j \to \infty} x_j$ lies in $\partial \text{dom} g_i(\cdot)$, then $\lim_{j \to \infty} g_i(x_j) = \infty$.

We have the following theorem.

**Theorem 4.1** [O(1/n) convergence of dual function] Suppose conditions (1) and (2) in Proposition 3.4 and Assumption 3.1 are satisfied and Algorithm 3.1 is run with finite $M$. If a dual minimizer to (4) exists, then the convergence rate of the dual objective value is $O(1/n)$. This in turn implies that the convergence rate of $\{\|x^{n,w} - x^*\|_n\}$ is $O(1/\sqrt{n})$.

**Proof** Let $V_n = -F_{H^{n,w}}(z^{n,w})$. Recall that $\{V_n\}$ is nonincreasing by Theorem 3.1(i). We want to show that $V_n - (-\beta) \leq O(1/n)$.

First, from line 8 of Algorithm 3.1, we have $H^{n,w} \supseteq \tilde{C}^{n,w+1}$, so

$$\frac{1}{2}\|v^{n,w} - x_0\|^2 + \sum_{i=1}^{r} h_i^*(z_{i}^{n,w}) + \delta_{H^{n,w}}^*(z_{r+1}^{n,w}) - \frac{1}{2}\|x_0\|^2$$

$$\geq \frac{1}{2}\|v^{n,w+1} - x_0\|^2 + \sum_{i=1}^{r} h_i^*(z_{i}^{n,w+1}) + \delta_{\tilde{C}^{n,w+1}}^*(z_{r+1}^{n,w+1}) - \frac{1}{2}\|x_0\|^2$$

$$\geq \frac{1}{2}\|v^{n,w+1} - x_0\|^2 + \sum_{i=1}^{r} h_i^*(z_{i}^{n,w+1}) + \delta_{\tilde{C}^{n,w+1}}^*(z_{r+1}^{n,w+1}) - \frac{1}{2}\|x_0\|^2$$

\[\square\] Springer
Let the vector $s$ exist. Making use of the elementary fact that $s \in X$ be constructed by appending $z$ for all $i$. Claim There is a constant $M_4$ such that $\|z_i^{n,w}\| \leq M_4$ for all $n \geq 0$, $w \in \{1, \ldots, \bar{w}\}$ and $i \in \{1, \ldots, r+1\}$. Step 1 The claim above is true for all $n \geq 0$, $i \in \{1, \ldots, r_2\}$ and $w \in \{0, \ldots, \bar{w}\}$. The limit $\lim_{n \to \infty} x_i^{n,\bar{w}}$ must lie in the interior of the domains of $f_i(\cdot)$ and $g_i(\cdot)$ for all $i \in \{1, \ldots, r_2\}$ (by Assumption (A2') stated just before the statement of this
theorem). It is well known that the subgradients of a convex function are bounded in the interior of its domain, so there is a constant $M_1$ such that $\|z_i^{n,w}\| \leq M_1$ for all $i \in \{1, \ldots, r_2\}$ and $w \in \{0, \ldots, \tilde{w}\}$.

**Step 2** The claim is true for all $n \geq 0$, $i \in \{1, \ldots, r + 1\}$ and $w = 0$.

Since we assumed that Algorithm 3.1 was run with a finite $M$, by (10c), $\|z_i^{n,0}\| \leq M$ for all $i \in \{r_2 + 1, \ldots, r\}$ and $n \geq 0$. Next, we show that $M_1$ can be made larger if necessary so that $\|z_{r+1}^{n,0}\| \leq M_1$ for all $n \geq 0$. Seeking a contradiction, suppose that there is a subsequence $\{n_k\}$ such that $\lim_{k \rightarrow \infty} \|z_{r+1}^{n_k,0}\| = \infty$. Then, this would mean that $\lim_{k \rightarrow \infty} \|\sum_{i=1}^{r+1} z_{i}^{n_k,0}\| = \infty$. Recalling (17) for the special case where $x = x^*$ (the primal minimizer), we have

$$\frac{1}{2} \left\| x_0 - x^* - \sum_{i=1}^{r+1} z_i^{n,0} \right\|^2 \leq \frac{1}{2} \left\| x_0 - x^* \right\|^2 + \sum_{i=1}^{r} h_i(x^*) - F_{H^0}(z^{n,0})$$

This implies that there is a $M_2 > 0$ such that $\|v^{n,w} - v^{n,w-1}\| \leq M_2$ for all $n \geq 0$ and $w \in \{1, \ldots, \tilde{w}\}$.

Next, we recall from Theorem 3.1(i) that $\sum_{n=1}^{\infty} \sum_{w=1}^{\tilde{w}} \|v^{n,w} - v^{n,w-1}\|^2$ is finite. This means that there is a $M_3 > 0$ such that $\|v^{n,w} - v^{n,w-1}\| \leq M_3$ for all $n \geq 0$ and $w \in \{1, \ldots, \tilde{w}\}$. Since the $z_i^{n,w}$ were chosen by condition (1) in Proposition 3.4, then if $n$ is large enough, if $S_{n,w} \cap \{r_2 + 1, \ldots, r + 1\} \neq \emptyset$, then there is an $i_n, w \in S_{n,w}$ such that $S_{n,w}\{i_n, w\} \subset \{1, \ldots, r_2\}$. We have

$$z_{i_n,w}^{n,w} \overset{(11a)}{=} z_{i_n,w}^{n,w-1} + v^{n,w} - v^{n,w-1} - \sum_{j \in S_{n,w}\{i_n,w\}} [z_j^{n,w} - z_j^{n,w-1}]. \quad (35)$$

Then, we have

$$\|z_{i_n,w}^{n,w}\| \overset{(35)}{\leq} \|z_{i_n,w}^{n,w-1}\| + \|v^{n,w} - v^{n,w-1}\| + \sum_{j \in S_{n,w}\{i_n,w\}} \|z_j^{n,w}\| + \|z_j^{n,w-1}\|$$

$$\leq M_3 + 2 \sum_{j \in S_{n,w}\{i_n,w\}} \|z_j^{n,w-1}\| + M_2 + 2r_2M_1.$$

This would easily imply that $\|z_i^{n,w}\| \leq M_4$ for some $M_4 > 0$ for all $n \geq 0$, $i \in \{1, \ldots, r + 1\}$, and $w \in \{1, \ldots, \tilde{w}\}$ as needed, ending the proof of the claim.
Now,
\[
\sum_{w=1}^{\bar{w}} \| v^{n,w} - v^{n,w-1} \| \leq \sqrt{2 \bar{w}} \sqrt{\frac{1}{2} \sum_{w=1}^{\bar{w}} \| v^{n,w} - v^{n,w-1} \|^2} \leq \sqrt{2 \bar{w}} \sqrt{V_n - V_{n+1}}.
\]  
(36)

Then, combining the above, we have
\[
V_n - V^* \overset{(34),(36)}{\leq} \left[ \sum_{i=1}^{r+1} \| z_i^{n,\bar{w}} \| + \sum_{i=1}^{r+1} \| z_i^* \| \right] \sqrt{2 \bar{w}} \sqrt{V_n - V_{n+1}} \leq \left[ (r+1)M_4 + \sum_{i=1}^{r+1} \| z_i^* \| \right] \sqrt{2 \bar{w}} \sqrt{V_n - V_{n+1}}.
\]  
(37)

Letting \( M_5 = (r+1)M_4 + \sum_{i=1}^{r+1} \| z_i^* \| \) and rearranging (37), we have
\[
V_n - V^* \geq V_{n+1} - V^* + \frac{1}{2 \bar{w} M_5^2} (V_{n+1} - V^*)^2.
\]

Applying Lemma 4.1 gives the first statement of our conclusion. The second statement comes from substituting \( x = x^* \) in (17) and noticing that
\[
x_0 - x^* - \sum_{i=1}^{r+1} z_i = x^{n,\bar{w}} - x^*.
\]

\[ \square \]

Remark 4.1 (Nonexistence of dual minimizers) An example of a problem where dual minimizers do not exist is in [3, page 9]. Lemma 2 in [4] shows that if \( \| x^n - P_C(x_0) \| = O(n^{-\lambda}) \) for some \( \lambda > 1 \) (where \( x_i^n \) is in the notation of [4]), then there exist dual minimizers.

5 Approximate Proximal Point Algorithm

Consider the problem of minimizing
\[
\sum_{i=1}^{r} h_i(x).
\]  
(38)

If one of the functions \( h_i(\cdot) \) is strongly convex, then it can be written as \( h_i(\cdot) = \tilde{h}_i(\cdot) + \frac{c_i}{2} \| \cdot - x_0 \|^2 \) for some convex \( \tilde{h}_i(\cdot) \) and \( c_i > 0 \), and (38) can be minimized using Dykstra’s splitting algorithm of Sect. 3. In this section, we propose an approximate
proximal point method for minimizing (38) without splitting $h_i(\cdot)$. We first present Algorithm 5.1 and prove that all its cluster points are minimizers of the parent problem. Then, in Sect. 5.2, we show that the Dykstra splitting investigated in Sect. 3 can find an approximate primal minimizer required in Algorithm 5.1.

5.1 An Approximate Proximal Point Algorithm

Consider the problem of minimizing $h : X \rightarrow \mathbb{R}$, where

$$h(\cdot) = \delta_D(\cdot) + \sum_{i=1}^{r_2} h_i(\cdot),$$

and each $h_i : X \rightarrow \mathbb{R}$ is a closed and convex function whose domain is an open set, and $D$ is a compact convex set in $X$. This setting is less general than that of (3), since it does not allow for all lower semicontinuous convex functions, and we only allow for one compact set $D$ instead of $r - r_2$ sets.

Algorithm 5.1 shows an approximate proximal point algorithm, where one solves a regularized version of (39) and shifts the proximal center $x_k$ when an approximate KKT condition is satisfied.

Algorithm 5.1 (Approximate proximal point algorithm) Consider the problem of minimizing $h(\cdot)$ of form (39). Let $\{\gamma_j\}_{j=1}^\infty \subset \mathbb{R}$ be a sequence such that $\lim_{j \rightarrow \infty} \gamma_j = 0$. Let $x_0 \in X$. Our algorithm is as follows: For $j = 1, \ldots$

- Find an approximate minimizer $x_j$ of $\min \delta_D(\cdot) + \frac{1}{2} \|\cdot - x_{j-1}\|^2 + \sum_{i=1}^{r_2} h_i(\cdot)$.

Specifically, find $x_j \in X, z^{(j)}(j), e^{(j)}(j) \in X^{r_2+1}$ and a closed and convex set $D^j \supset D$ such that

$$z_i^{(j)} \in \partial h_i(x_j + e_i^{(j)}) \text{ for all } i \in \{1, \ldots, r_2\},$$

$$z_0^{(j)} \in N_{D^j}(x_j + e_0^{(j)}),$$

$$\left\|(x_j - x_{j-1}) + \sum_{i=0}^{r_2} z_i^{(j)}\right\| \leq \gamma_j,$$

$$\|e_i^{(j)}\| \leq \gamma_j \text{ for all } i \in \{0, \ldots, r_2\}.$$  

end For.

If $r_2 = 1$, $D = X$, and $h_1(\cdot)$ were allowed to be any lower semicontinuous convex function, then Algorithm 5.1 would resemble the classical proximal point algorithm. Define the operator $T : X \rightarrow X$ by

$$T(x) := \text{prox}_{h}(x) := \arg\min_{x'} h(x') + \frac{1}{2}\|x' - x\|^2.$$  

This operator has some favorable properties in monotone operator theory. We prove our first result.

\@ Springer
Lemma 5.1 (Approximate of $T(\cdot)$) Consider the problem (39). Let $T(\cdot)$ be as defined in (41). Suppose $D \subset X$ is compact and convex. For all $\epsilon > 0$, there is a $\gamma > 0$ such that for all $x, x^+ \in X$ and $z, e \in X^{r+1}$ such that $d(x, D) \leq \gamma$ and

$$z_i \in \partial h_i(x^+ + e_i) \text{ for all } i \in \{1, \ldots, r_2\},$$

$$z_0 \in N_{\tilde{D}}(x^+ + e_0),$$

$$\left\| (x^+ - x) + \sum_{i=0}^{r_2} z_i \right\| \leq \gamma,$$

$$\left\| e_i \right\| \leq \gamma \text{ for all } i \in \{0, \ldots, r_2\},$$

where $\tilde{D} \supset D$ is a closed and convex set, we have $\left\| x^+ - T(x) \right\| \leq \epsilon$.

Proof Seeking a contradiction, suppose otherwise. Then, there exists a $\epsilon > 0$ such that for all positive integers $k$, there are $x_k, x_k^+ \in X, z^{(k)}, e^{(k)} \in X^{r+1}$ and a closed and convex set $D^{(k)} \supset D$ such that

$$d(x_k, D) \leq 1/k,$$

$$z_i^{(k)} \in \partial h_i(x_k^+ + e_i^{(k)}),$$

$$z_0^{(k)} \in N_{D^{(k)}}(x_k^+ + e_0^{(k)}),$$

$$\left\| (x_k^+ - x_k) + \sum_{i=0}^{r_2} z_i^{(k)} \right\| \leq 1/k,$$

$$\left\| e_i^{(k)} \right\| \leq 1/k \text{ for all } i \in \{0, \ldots, r_2\},$$

but

$$\left\| x_k^+ - T(x_k) \right\| \geq \epsilon.$$  (43)

Letting $k \not\to \infty$, we can assume (by taking subsequences if necessary) that

$$\lim_{k \to \infty} x_k = \bar{x}, \quad \lim_{k \to \infty} x_k^+ = \bar{x}^+ \text{ and } \lim_{k \to \infty} e^{(k)} = 0.$$  (44)

There are two cases we need to consider.

**Case 1** $\bar{x}^+$ lies in the interior of $\text{dom} h_i$ for all $i \in \{1, \ldots, r_2\}$.

Making use of the fact that convex functions are locally Lipschitz in the interior of their domains, we obtain the boundedness of $\{z^{(k)}\}$. We can assume (by taking subsequences if necessary) that $\lim_{k \to \infty} z^{(k)} = \bar{z}$. Taking the limits of (42) as $k \to \infty$ gives $\bar{z}_i \in \partial h_i(\bar{x}^+), \bar{z}_0 \in N_D(\bar{x}^+)$ and $\bar{x}^+ - \bar{x} + \sum_{i=0}^{r_2} \bar{z}_i = 0$, which would in turn imply that $\bar{x}^+ = T(\bar{x})$. It is well known that $T(\cdot)$ is nonexpansive and hence continuous, so

$$0 < \epsilon \leq \lim_{k \to \infty} \left\| x_k^+ - T(x_k) \right\| = \left\| \bar{x}^+ - T(\bar{x}) \right\| = 0,$$

where $\tilde{D} \supset D$ is a closed and convex set, we have $\left\| x^+ - T(x) \right\| \leq \epsilon$.  

Proof Seeking a contradiction, suppose otherwise. Then, there exists a $\epsilon > 0$ such that for all positive integers $k$, there are $x_k, x_k^+ \in X, z^{(k)}, e^{(k)} \in X^{r+1}$ and a closed and convex set $D^{(k)} \supset D$ such that

$$d(x_k, D) \leq 1/k,$$

$$z_i^{(k)} \in \partial h_i(x_k^+ + e_i^{(k)}),$$

$$z_0^{(k)} \in N_{D^{(k)}}(x_k^+ + e_0^{(k)}),$$

$$\left\| (x_k^+ - x_k) + \sum_{i=0}^{r_2} z_i^{(k)} \right\| \leq 1/k,$$

$$\left\| e_i^{(k)} \right\| \leq 1/k \text{ for all } i \in \{0, \ldots, r_2\},$$

but

$$\left\| x_k^+ - T(x_k) \right\| \geq \epsilon.$$  (43)

Letting $k \not\to \infty$, we can assume (by taking subsequences if necessary) that

$$\lim_{k \to \infty} x_k = \bar{x}, \quad \lim_{k \to \infty} x_k^+ = \bar{x}^+ \text{ and } \lim_{k \to \infty} e^{(k)} = 0.$$  (44)

There are two cases we need to consider.

**Case 1** $\bar{x}^+$ lies in the interior of $\text{dom} h_i$ for all $i \in \{1, \ldots, r_2\}$.

Making use of the fact that convex functions are locally Lipschitz in the interior of their domains, we obtain the boundedness of $\{z^{(k)}\}$. We can assume (by taking subsequences if necessary) that $\lim_{k \to \infty} z^{(k)} = \bar{z}$. Taking the limits of (42) as $k \to \infty$ gives $\bar{z}_i \in \partial h_i(\bar{x}^+), \bar{z}_0 \in N_D(\bar{x}^+)$ and $\bar{x}^+ - \bar{x} + \sum_{i=0}^{r_2} \bar{z}_i = 0$, which would in turn imply that $\bar{x}^+ = T(\bar{x})$. It is well known that $T(\cdot)$ is nonexpansive and hence continuous, so

$$0 < \epsilon \leq \lim_{k \to \infty} \left\| x_k^+ - T(x_k) \right\| = \left\| \bar{x}^+ - T(\bar{x}) \right\| = 0,$$
a contradiction.

Case 2 $\bar{x}^+$ lies on the boundary of $\text{dom } h_i$ for some $i \in \{1, \ldots, r_2\}$.

We cannot use the method in Case 1 as some components of $\{z^{(k)}\}$ might be unbounded. We now consider the perturbed functions $h_{i,k}(\cdot)$ defined by

$$h_{i,k}(x) := h_i(x + e_i^{(k)}).$$

(45)

Let $\tilde{h}_k : X \to \mathbb{R}$ be defined by $\tilde{h}_k(\cdot) = \delta_{D(\cdot)}(\cdot) + \sum_{i=1}^{r_2} h_{i,k}(\cdot)$. Then, conditions (40) imply that

$$x_k^+ = \text{prox}_{\tilde{h}_k}(x_k + d_k) = \arg \min_x \tilde{h}_k(x) + \frac{1}{2} \|x - (x_k + d_k)\|^2,$$

(46)

where $d_k$ is marked in (42b). Suppose $\tilde{i}$ is such that $\bar{x}^+$ lies on the boundary of $\text{dom } h_{\tilde{i}}$. Then, we have that $\lim_{k \to \infty} h_{\tilde{i}}(x_k^+) = \infty$. Since $D$ is bounded, $\inf_{x \in D} h_i(x)$ is a finite number for all $i$, which implies that

$$\lim_{k \to \infty} \tilde{h}_k(x_k^+) + \frac{1}{2} \|x_k^+ - (x_k + d_k)\|^2 = \infty.$$

(47)

Next, let $x_k' = \text{prox}_h(x_k)$ and $x' = \text{prox}_h(\bar{x})$. By the continuity properties of $T(\cdot) = \text{prox}_h(\cdot)$ and $\lim_{k \to \infty} x_k = \bar{x}$ in (44), we must have $\lim_{k \to \infty} x_k' = x'$. It is clear that $x'$ lies in $\text{dom}(h_{\tilde{i}})$. Since we assumed that $\text{dom}(h_i)$ is open, $x' \in \text{int } \text{dom}(h_i)$ for all $i$. We then have

$$\lim_{k \to \infty} \tilde{h}_k(x') + \frac{1}{2} \|x' - x_k\|^2 = h(x') + \frac{1}{2} \|x' - \bar{x}\|^2 < \infty.$$

(48)

But on the other hand, since $\lim_{k \to \infty} d_k = 0$, we have

$$\lim_{k \to \infty} \tilde{h}_k(x') + \frac{1}{2} \|x' - (x_k + d_k)\|^2 \overset{(46)}{\geq} \lim_{k \to \infty} \tilde{h}_k(x_k^+) + \frac{1}{2} \|x_k^+ - (x_k + d_k)\|^2 \overset{(47)}{=} \infty.$$

(49)

Formulas (48) and (49) are contradictory, so $\bar{x}^+$ must lie in the interior of all $\text{dom } h_i$ for all $i$, which reduces to case 1.

Thus, we are done. \hfill \Box

To simplify notation in the next two results, we define the set $A$ to be

$$A := \arg \min_x h(x).$$

We have another lemma.

**Lemma 5.2** For all $\epsilon > 0$, there exists $\gamma > 0$ such that for all $w$ such that $d(w, D) \leq \gamma$, we have

$$\|T(w) - w\| \leq \gamma$$

implies $d(w, A) \leq \epsilon$.\hfill \copyright Springer
Proof Seeking a contradiction, suppose otherwise. In other words, there is a $\bar{\epsilon} > 0$ such that for all $k > 0$, there is a $w_k$ such that $d(w_k, A) > \bar{\epsilon}$ but $\|T(w_k) - w_k\| \leq \frac{1}{k}$. By taking subsequences if necessary, let $\bar{w} = \lim_{k \to \infty} w_k$, which exists by the compactness of $D$. Taking limits as $k \to \infty$ gives us $\bar{w} = T(\bar{w})$, which will in turn imply that $d(\bar{w}, A) = 0$, a contradiction. $\square$

Theorem 5.1 (Cluster points of Algorithm 5.1) All cluster points of $\{x_j\}$ in Algorithm 5.1 are minimizers of $h(\cdot)$.

Proof For any $\epsilon_1 > 0$, we make use of Lemma 5.2 and obtain $\gamma_1 > 0$ such that

$$\|T(w) - w\| \leq \gamma_1 \text{ implies } d(w, A) \leq \epsilon_1.$$  \hspace{1cm} (50)

By Lemma 5.1, there exists some $K$ large enough so that

$$\|x_{k+1} - T(x_k)\| \leq \epsilon \text{ for all } k \geq K.$$  \hspace{1cm} (51)

Let $ε > 0$ be small enough so that $\frac{4\epsilon^2 + \gamma_1^2}{4\epsilon} > \text{diam}(D)$. Then, since

$$\lim_{k \to \infty} d(x_k, D) = 0$$

and $A \subset D$, we can increase $K$ if necessary so that

$$d(x_k, A) \leq \text{diam}(D) < \frac{4\epsilon^2 + \gamma_1^2}{4\epsilon} \text{ for all } k \geq K.$$  \hspace{1cm} (52)

Let $\bar{x}_k = P_A(x_k)$ so that $d(x_k, A) = \|x_k - \bar{x}_k\|$. It is well known from the theory of monotone operators that $T(\cdot)$ is firmly nonexpansive (see, for example, [11, Definition 4.1(i), Proposition 12.27]), so we have

$$\|T(x_k) - \bar{x}_k\|^2 + \|x_k - T(x_k)\|^2 \leq \|x_k - \bar{x}_k\|^2.$$  \hspace{1cm} (53)

Suppose $k \geq K$. We split our analysis into two cases.

Case 1 $d(x_k, A) > \epsilon_1$.

Then, (50) implies $\|T(x_k) - x_k\| > \gamma_1$. We have

$$d(x_{k+1}, A) \leq \|x_{k+1} - \bar{x}_k\|$$

$$\leq \|x_{k+1} - T(x_k)\| + \|T(x_k) - \bar{x}_k\|$$

$$\leq \epsilon + \sqrt{\|x_k - \bar{x}_k\|^2 - \|x_k - T(x_k)\|^2}$$

$$< \epsilon + \sqrt{d(x_k, A)^2 - \gamma_1^2}$$

rearrange (52)

$$\leq d(x_k, A) - \epsilon.$$  \hspace{1cm} (54)

Case 2 $d(x_k, A) \leq \epsilon_1$. 

Springer
We have
\[ d(x_{k+1}, A) \leq \|x_{k+1} - \bar{x}_k\| \leq \|x_k - T(x_k)\| + \|T(x_k) - \bar{x}_k\| \leq \epsilon + \|x_k - \bar{x}_k\| = \epsilon + d(x_k, A) \leq \epsilon + \epsilon_1. \] (51), (53)

The analysis in these two cases implies \( d(x_{k+1}, A) \leq \epsilon_1 + \epsilon \) for all \( k \) large enough. Since \( \epsilon_1 \) and \( \epsilon \) can be made arbitrarily small, any cluster point \( x' \) of \( \{x_k\}_{k=1}^\infty \) must thus satisfy \( d(x', A) = 0 \), or \( x' \in A \). \( \square \)

5.2 Satisfying (40) Using Dykstra’s Splitting

Consider the problem of minimizing \( h : X \rightarrow \mathbb{R} \), where
\[ h(x) = \sum_{i=1}^{r_2} h_i(x) + \sum_{i=r_2+1}^{r} \delta_{C_i}(x). \] (55)

and each \( h_i : X \rightarrow \mathbb{R} \) is a closed and convex function whose domain is an open set, and each \( C_i \) is a closed and convex set such that \( \cap_{i=r_2+1}^{r} C_i \) is compact. This formulation is slightly more general than that of (39). Theorem 5.2 shows that Algorithm 3.1 can find approximate minimizers to (55) that will satisfy the conditions for moving to a new proximal center in Algorithm 5.1.

**Theorem 5.2** Consider problem (55). For any \( \gamma > 0 \), there is some \( n > 0 \) such that when Algorithm 3.1 is applied to solve
\[ h(x) = \sum_{i=1}^{r_2} h_i(x) + \sum_{i=r_2+1}^{r} \delta_{C_i}(x) + \frac{1}{2} \|x - x_0\|^2, \] (56)

we have a set \( D^{(n)} \), and points \( x^{(n)} \in X \) and \( \tilde{z} \in X \) such that
\[ \|x^{n,\bar{w}} - x^{n,\bar{w}}\| \leq \gamma \text{ for all } w \in \{1, \ldots, \bar{w}\}, \] (57a)
\[ z_i^{n,\bar{w}} \in \partial h_i(x^{n,p(n,i)}) \text{ for all } i \in \{1, \ldots, r\}, \] (57b)
\[ \sum_{i=1}^{r_2} z_i^{n,\bar{w}} + \tilde{z} + x^{n,\bar{w}} - x_0 \| \leq \gamma, \] (57c)
\[ \cap_{i=r_2+1}^{r} C_i \subset D^{(n)}, \] (57d)
\[ \tilde{z} \in N_{D^{(n)}}(x^{(n)}), \] (57e)
\[ \|x^{(n)} - x^{n,\bar{w}}\| \leq \gamma. \] (57f)
Proof Define $z_{0}^{n,\bar{w}}$ to be

$$z_{0}^{n,\bar{w}} := \sum_{i=r+1}^{r+1} z_{i}^{n,\bar{w}}. \quad (58)$$

Theorem 3.1 says that for any $\gamma > 0$, we can find $n > 0$ such that the first three conditions in (57) hold if $\tilde{z}$ were chosen to be $z_{0}^{n,\bar{w}}$. We separate into two cases and discuss how the set $D^{(n)}$ (which will actually be either the whole space $X$ or a half-space) and the point $x^{(n)}$ are constructed.

Case 1 $\lim \inf_{n \to \infty} \|z_{0}^{n,\bar{w}}\| = 0$.

By taking subsequences $\{n_{k}\}_{k}$, we can assume that $\lim_{k \to \infty} \|z_{0}^{n_{k},\bar{w}}\| = 0$. Then, the set $D^{(n)}$ can be chosen to be $X$, and $x^{(n)}$ can be chosen to be $x^{*}$, the minimizer of (56). The vector $\tilde{z}$ can be chosen to be zero, and the inequalities in (57) can be easily seen to be satisfied.

Case 2 $\lim \inf_{n \to \infty} \|z_{0}^{n,\bar{w}}\| > 0$.

Recall that for $i \in \{r+1, \ldots, r\}$, the dual vector $z_{i}^{n,\bar{w}}$ was constructed so that $z_{i}^{n,\bar{w}} \in NC_{i}(x^{n,p(n,i)})$ and that $z_{i}^{n,\bar{w}}_{r+1} \in N_{H^{n,i}}(x^{n,p(n,r+1)})$. For all $i \in \{r+1, \ldots, r+1\}$, define the half-space $H^{n,i}$ to be the half-space with $x^{n,p(n,i)}$ on its boundary and outward normal vector $z_{i}^{n,\bar{w}}$ like in (16). Note that $H^{n,i} \supset C_{i}$ for $i \in \{r+1, \ldots, r\}$, and $\tilde{H}^{n,r+1} \supset H^{n,i}$.

Let $D^{(n)}$ be the half-space with outward normal $z_{0}^{n,\bar{w}}$ such that we have $D^{(n)} \supset \bigcap_{i=r+1}^{r} H^{n,i}$, and $\partial D^{(n)} \cap \bigcap_{i=r+1}^{r} H^{n,i} \neq \emptyset$. Through Fact 3.1, this choice of $D^{(n)}$ would give us

$$\delta_{D^{(n)}}^{*}(z_{0}^{n,\bar{w}}) \leq \sum_{i=r+1}^{r} \delta_{C_{i}}^{*}(z_{i}^{n,\bar{w}}) + \delta_{H^{n,i}}^{*}(z_{r+1}^{n,\bar{w}}). \quad (59)$$

We now show how to satisfy (57e) and (57f) with $\tilde{z} = z_{0}^{n,\bar{w}}$. Let $x^{*}$ be the optimal primal solution of (56). We now want to show that we can choose a further subsequence if necessary so that $\lim_{k \to \infty} d(\partial D^{(n_{k})}, x^{*}) = 0$. From the definition of the support function and the fact that $x^{*} \in \bigcap_{i=1}^{r} C_{i} \subset D^{(n)}$, we have

$$\delta_{D^{(n)}}^{*}(z_{0}^{n,\bar{w}}) - \langle x^{*}, z_{0}^{n,\bar{w}} \rangle = \delta_{D^{(n)}}^{*}(z_{0}^{n,\bar{w}}) - \|z_{0}^{n,\bar{w}}\| d(\partial D^{(n)}, x^{*}). \quad (60)$$

We now mimic (17) to obtain

$$\frac{1}{2}\|x_{0} - x^{*}\|^{2} + \sum_{i=1}^{r} h_{i}(x^{*}) + \sum_{i=r+1}^{r} \delta_{C_{i}}(x^{*}) = \sum_{i=r+1}^{r} \delta_{C_{i}}(x^{*}) - F_{H^{n,i}}(z_{0}^{n,\bar{w}}, z_{r+1}^{n,\bar{w}})$$
To this paper. (In [43], we also extended the algorithm to time-varying graphs, which is not covered in this paper.) Consider a graph $G$.

We look at an application of Algorithm 3.1, which we study in [43].

### 6 Application: A Distributed Optimization Algorithm

We look at an application of Algorithm 3.1, which we study in [43] as a follow-up to this paper. (In [43], we also extended the algorithm to time-varying graphs, which is not covered in this paper.) Consider a graph $G = (V, E)$. For each $i \in V$, let $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$ be closed and convex functions. We look at the case of a distributed optimization problem.
\[
\min_{x \in \mathbb{R}^m} \sum_{i \in V} \left[ f_i(x) + \frac{1}{2} \|x - \bar{x}_i\|^2 \right].
\] (62)

The difficulty in distributed optimization is that nodes \(i, j \in V\) may only communicate data if the edge \((i, j)\) exists. Consider the hyperplanes \(H_{(i, j)}\) defined by

\[
H_{(i, j)} := \{ x \in [\mathbb{R}^m]^{|V|} : x_i = x_j \}
\]

Note that if the graph \(G\) is connected, then \(\cap_{e \in E} H_e = D\), where \(D\) is the diagonal set defined by

\[
D := \{ x \in [\mathbb{R}^m]^{|V|} : x_1 = x_2 = \cdots = x_{|V|} \}.
\]

Let \(f_i: [\mathbb{R}^m]^{|V|} \to \mathbb{R}\) be defined as \(f_i(x) = f_i(x_i)\), and let \(\bar{x} \in [\mathbb{R}^m]^{|V|}\) be defined as \(\bar{x}_i = \bar{x}_i\). We write the problem (62) in the product space formulation as

\[
\min_{x \in [\mathbb{R}^m]^{|V|}} \sum_{i \in V} f_i(x) + \sum_{e \in E} \delta_{H_e}(x) + \frac{1}{2} \|x - \bar{x}\|^2.
\]

(The difference between the distributed formulation above and the product space formulation is that the product space formulation uses \(\delta_D(x)\) instead of \(\sum_{e \in E} \delta_{H_e}(x)\).)

This problem has a (modified) dual

\[
\min_{z_\alpha \in [\mathbb{R}^m]^{|V|}: \alpha \in V \cup E} \frac{1}{2} \left\| \bar{x} - \sum_{\alpha \in V \cup E} z_\alpha \right\|^2 + \sum_{i \in V} f_i^*(z_i) + \sum_{e \in E} \delta_{H_e}^*(z_e).
\]

We explained in [43] that even though there are many dual variables \(z_\alpha\), one only needs to keep track of \(x\) marked above and \([z_i]_{i}\) (i.e., the \(i\)-th coordinate of \(z_i\)). Also minimizing with respect to one of the variables \(z_i\) and \(z_e\) at a time changes the only one or two coordinates in \(x\), respectively, which leads to a distributed algorithm. The work in [43] is a follow-up to this paper, with the asynchronous operation there being a result of being able to minimize over the dual variables \(z_\alpha\) not necessarily in a cyclic manner. We mention that one application of the theory in this paper not covered in [43]: If some of the \(f_i(\cdot)\) in (62) were indicator functions of closed and convex sets, then the SHQP heuristic can be applied. The proximal point algorithm in Sect. 5 can be performed if one were able to check the conditions in the stated results, though it may not be easy in a distributed setting.

### 7 Infinite Dimensions

Finally, we mention that the results in [4] were proved in the infinite-dimensional case. We do not generalize to the infinite-dimensional case in this paper, but we remark that the ideas in [24–26] can be applied for the infinite-dimensional case.
8 Conclusions

In this paper, we looked at how the sum of convex functions and a regularizing quadratic function can be minimized through a dual block coordinate maximization strategy and consolidated some results on such problems. The functions can be sampled in a manner that is not necessarily cyclic. As mentioned, this has consequences in asynchronous computation in distributed optimization in [43]. In order to remove the regularizing quadratic function, a proximal point algorithm was proposed, but the conditions for moving to a new proximal center may not be easy to check in a distributed optimization problem. Possible future directions include devising simpler subproblems that would give dual ascent and maintain favorable convergence rates, alternative strategies for minimizing distributed problems without the regularizing quadratic function, and hardness results for minimizing the sum of convex functions in a distributed manner without the regularizing quadratic function.

Acknowledgements We acknowledge Grant R-146-000-214-112 from the Faculty of Science, National University of Singapore. We gratefully acknowledge discussions with Ting-Kei Pong on Dykstra’s splitting which led to this paper. We also thank the two anonymous referees and the editorial staff.

References

1. Dykstra, R.: An algorithm for restricted least-squares regression. J. Am. Stat. Assoc. 78, 837–842 (1983)
2. Boyle, J., Dykstra, R.: A method for finding projections onto the intersection of convex sets in Hilbert spaces. In: Dykstra, R., Robertson, T., Wright, F.T. (eds.) Advances in Order Restricted Statistical Inference. Lecture notes in Statistics, pp. 28–47. Springer, New York (1985)
3. Han, S.: A successive projection method. Math. Program. 40, 1–14 (1988)
4. Gaffke, N., Mathar, R.: A cyclic projection algorithm via duality. Metrika 36, 29–54 (1989)
5. Iusem, A., Pierro, A.D.: On the convergence of Han’s method of convex programming with quadratic objective. Math. Program. 52, 265–284 (1991)
6. Pierra, G.: Decomposition through formalization in a product space. Math. Program. 28, 96–115 (1984)
7. Hundal, H., Deutsch, F.: Two generalizations of Dykstra’s cyclic projections algorithm. Math. Program. 77, 335–355 (1997)
8. Pang, C.: The supporting halfspace—quadratic programming strategy for the dual of the best approximation problem. SIAM J. Optim. 26(4), 2591–2619 (2016)
9. Deutsch, F.: Accelerating the convergence of the method of alternating projections via a line search: a brief survey. In: Butnariu, D., Censor, Y., Reich, S. (eds.) Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications, pp. 203–217. Elsevier, Amsterdam (2001)
10. Deutsch, F.: Best Approximation in Inner Product Spaces. CMS Books in Mathematics. Springer, Berlin (2001)
11. Bauschke, H., Combettes, P.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)
12. Escalante, R., Raydan, M.: Alternating Projection Methods. SIAM, Philadelphia, PA (2011)
13. Beck, A., Tretuashvili, L.: On the convergence of block coordinate descent type methods. SIAM J. Optim. 23(4), 2037–2060 (2013)
14. Beck, A.: On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes. SIAM J. Optim. 25(1), 185–209 (2015)
15. Tseng, P., Yun, S.: A coordinate gradient descent method for nonsmooth separable minimization. Math. Program. Ser. B 117(117), 387–423 (2009)
16. Tseng, P., Yun, S.: Block-coordinate gradient descent method for linearly constrained nonsmooth separable optimization. J. Optim. Theory Appl. 140, 513–535 (2009)
17. Wright, S.: Coordinate descent algorithms. Math. Program. 151, 3–34 (2015)
18. Hong, M., Wang, X., Razaviyayn, M., Luo, Z.: Iteration complexity analysis of block coordinate descent methods. Math. Program. 163, 85–114 (2017)
19. Martinet, B.: Régularisation d’inéquations variationnelles par approximations successives. Rev. Française Informat. Rech. Opér. 4, 154–158 (1970)
20. Rockafellar, R.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. 14, 877–898 (1976)
21. Han, S.: A decomposition method and its application to convex programming. Math. Oper. Res. 14, 237–248 (1989)
22. Bauschke, H., Combettes, P.: A Dykstra-like algorithm for two monotone operators. Pac. J. Optim. 4, 383–391 (2008)
23. Tseng, P.: Dual coordinate ascent methods for non-strictly convex minimization. Math. Program. 59, 231–248 (1993)
24. Combettes, P., Dung, D., Vu, B.: Proximity for sums of composite functions. J. Math. Anal. Appl. 380(2), 680–688 (2011)
25. Combettes, P., Dung, D., Vu, B.: Dualization of signal recovery problems. Set-Valued Var. Anal. 18, 373–404 (2010)
26. Abboud, F., Chouzenoux, E., Pesquet, J.C., Chenot, J.H., Laborelli, L.: Dual block-coordinate forward–backward algorithm with application to deconvolution and deinterlacing of video sequences. J. Math. Imaging Vis. 59(3), 415–431 (2017)
27. Nedic, A.: Random algorithms for convex minimization problems. Math. Program. Ser. B 225, 225–253 (2011)
28. Nesterov, Y.: Introductory Lectures on Convex Optimization. Kluwer, London (2004)
29. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2(1), 183–202 (2009)
30. Tseng, P.: On accelerated proximal gradient methods for convex-concave optimization (2008)
31. Bauschke, H., Borwein, J.: On projection algorithms for solving convex feasibility problems. SIAM Rev. 38, 367–426 (1996)
32. Censor, Y., Chen, W., Combettes, P.L., Davidi, R., Herman, G.: On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints. Comput. Optim. Appl. 51, 1065–1088 (2012)
33. Neto, E.H., Pierro, A.D.: Incremental subgradients for constrained convex optimization: a unified framework and new methods. SIAM J. Optim. 20, 1547–1572 (2009)
34. Ram, S., Nedić, A., Veeravalli, V.: Incremental stochastic subgradient algorithms for convex optimization. SIAM J. Optim. 20, 691–717 (2009)
35. Censor, Y., Davidi, R., Herman, G.: Perturbation resilience and superiorization of iterative algorithms. Inverse Probl. 26(6), 065008 (2010)
36. Censor, Y., Davidi, R., Herman, G., Schulte, R., Tetrashvili, L.: Projected subgradient minimization versus superiorization. J. Optim. Theory Appl. 160, 730–747 (2014)
37. Bauschke, H., Borwein, J., Li, W.: Strong conical hull intersection property, bounded linear regularity, Jameson’s property (G), and error bounds in convex optimization. Math. Program., Ser. A 86(1), 135–160 (1999)
38. Burke, J., Deng, S.: Weak sharp minima revisited. II. Application to linear regularity and error bounds. Math. Program., Ser. B 104(2–3), 235–261 (2005)
39. Ng, K., Yang, W.: Regularities and their relations to error bounds. Math. Program., Ser. A 99, 521–538 (2004)
40. Kruger, A.: About regularity of collections of sets. Set-Valued Anal. 14, 187–206 (2006)
41. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1), 1–122 (2010)
42. Combettes, P., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Bauschke, H., Burachik, R., Combettes, P., Elser, V., Luke, D., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer, New York, NY (2011)
43. Pang, C.H.J.: Distributed deterministic asynchronous algorithms in time-varying graphs through Dykstra splitting. SIAM J. Optim. 29(1), 484–510 (2018)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.