SIAM J. OPTIM.
Vol. 28, No. 2, pp. 1640–1669

DECOMPOSITION METHODS FOR COMPUTING DIRECTIONAL STATIONARY SOLUTIONS OF A CLASS OF NONSMOOTH NONCONVEX OPTIMIZATION PROBLEMS*

JONG-SHI PANG† AND MIN TAO‡

Abstract. Motivated by block partitioned problems arising from group sparsity representation and generalized noncooperative potential games, this paper presents a basic decomposition method for a broad class of multiblock nonsmooth optimization problems subject to coupled linear constraints on the variables that may additionally be individually constrained. The objective of such an optimization problem is given by the sum of two nonseparable functions minus a sum of separable, pointwise maxima of finitely many convex differentiable functions. One of the former two nonseparable functions is of the class LC^1, i.e., differentiable with a Lipschitz gradient, while the other summand is *multiconvex*. The subtraction of the separable, pointwise maxima of convex functions induces a partial difference-of-convex (DC) structure in the overall objective; yet with all three terms together, the objective is nonsmooth and non-DC, but is *blockwise directionally differentiable*. By taking advantage of the (negative) pointwise maximum structure in the objective, the developed algorithm and its convergence result are aimed at the computation of a *blockwise directional stationary solution*, which arguably is the sharpest kind of stationary solutions for this class of nonsmooth problems. This aim is accomplished by combining the alternating direction method of multipliers (ADMM) with a semilinearized Gauss–Seidel scheme, resulting in a decomposition of the overall problem into subproblems each involving the individual blocks. To arrive at a stationary solution of the desired kind, our algorithm solves multiple convex subprograms at each iteration, one per convex function in each pointwise maximum. In order to lessen the potential computational burden in each iteration, a probabilistic version of the algorithm is presented and its almost sure convergence is established.

Key words. nonconvex optimization, directional stationary points, Bregman regularization, alternating direction method of multipliers, block coordinate descent method, convergence analysis

AMS subject classifications. 90C26, 68W20

DOI. 10.1137/17M1110249

1. Introduction.

Originated from the family of splitting methods for solving monotone operators in the mid-1970s [22, 33, 21], and leading to the methods of multipliers for nonlinear programming in the 1980–1990s [5, 24, 14, 15, 16], the family of alternating direction method of multipliers (ADMM) has in recent years become extremely popular for solving convex programs [28, 27, 13] and has applications to many engineering domains such as image science [9, 10, 49, 54], machine learning [7, 45, 35], matrix completion [51], factorization [55], and rank minimization [20], and polynomial optimization [36], to name a few areas. See [17, Chapter 12] for a comprehensive summary of splitting methods for monotone variational inequalities up to 2002. The survey [7] and the edited volume [23] contain extensive recent references.

*Received by the editors January 3, 2017; accepted for publication (in revised form) January 10, 2018; published electronically May 15, 2018.
http://www.siam.org/journals/siopt/28-2/M111024.html
Funding: This work was funded by the National Science Foundation under grant IIS-1632971 and the Natural Science Foundation of China: NSFC–11301280, and the sponsorship of the Jiangsu overseas research and training program for university prominent young and middle-aged teachers and presidents. The second author was supported by the Natural Science Foundation of China: NSFC–11301280.
†The Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089-0193 (jongship@usc.edu).
‡Department of Mathematics, Nanjing University, Nanjing, 210093, People's Republic of China (taom@nju.edu.cn).

In the past few years, extensions of the ADMM to nonconvex programs have been investigated in [3, 25, 37, 29, 46, 47, 48]. This paper adds to this growing literature of the ADMM applied to nonconvex programs by considering a distinctive class of multiblock, nonsmooth optimization problem involving difference-of-convex (DC) functions of a certain kind.

Specifically, for given positive integers I, $\{J_i\}_{i=1}^{I}$, and $\{n_i\}_{i=1}^{I+1}$, and closed convex sets $X^i \subseteq \mathbb{R}^{n_i}$ with $\mathbf{X} \triangleq \prod_{i=1}^{I+1} X^i \subset \Omega \triangleq \prod_{i=1}^{I+1} \Omega^i$, where each Ω^i is an open set, by letting $n \triangleq \sum_{i=1}^{I+1} n_i$, the problem, whose feasible set we denote $\widehat{\mathbf{X}}$, is

$$\text{(1)} \qquad \left\{ \begin{array}{l} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \theta(x) \triangleq \varphi(x) + H(x) - \displaystyle\sum_{i=1}^{I} \max_{1 \le j \le J_i} g_{ij}(x^i) \\ \qquad\qquad \text{(only } I \text{ separable terms in the last sum)} \\ \text{subject to} \quad \displaystyle\sum_{i=1}^{I+1} A^i x^i = b \quad \text{and} \quad x \triangleq \left(x^i \right)_{i=1}^{I+1} \in \mathbf{X}; \end{array} \right.$$

here the defining terms in the overall objective function θ satisfy the following properties:

- $\varphi : \Omega \to \mathbb{R}$ is a continuously differentiable function;
- $H : \mathbf{X} \to \mathbb{R}$ is *multiconvex* [53]; i.e., for all $i = 1, \ldots, I+1$, $H(x^i, x^{-i})$ is convex on X^i for every fixed $x^{-i} \in X^{-i} \triangleq \prod_{j \neq i} X^j$, for each $i = 1, \ldots, I$ and $j = 1, \ldots, J_i$;
- each $g_{ij} : \Omega^i \to \mathbb{R}$ is convex and continuously differentiable.

Moreover, each $A^i \in \mathbb{R}^{\ell \times n_i}$ for $i = 1, \ldots, I+1$, and $b \in \mathbb{R}^\ell$ for some nonnegative integer ℓ. The case in which $\ell = 0$ pertains to the absence of coupling constraints of the variable blocks. In this setting, each resulting function $H(x^i, x^{-i}) - \max_{1 \le j \le J_i} g_{ij}(x^i)$ for $i = 1, \ldots, I$ is a nondifferentiable DC function for fixed x^{-i}; yet the overall objective θ is neither differentiable (because of the possible lack of joint differentiability of H in its arguments and the pointwise maxima) nor DC (because of the lack of such a requirement on the first two summands φ and H). Nevertheless, θ has some partial differentiability and a multi-DC structure. For the full set assumptions on problem (1), see subsection 3.2. Among these, some Lipschitz conditions are imposed on the gradient of the function φ and the partial gradient of H with respect to the distinguished block x^{I+1}; see assumption (A0) there. Moreover, the Lipschitz constants play an important role in ensuring the convergence of the algorithms to be developed. The concluding remarks at the end of the paper mention an extended class of problems not covered by this framework.

Besides extending the existing literature, the special structure of problem (1) arises from two applied sources. One is in sparsity representation [26] of data where surrogate sparsity functions [1] are used to approximate the well-known discontinuous univariate ℓ_0 function $|t|_0$, which equals 1 if $t \neq 0$ and equals 0 otherwise. The other source is a generalized noncooperative game with a potential function [19] that leads to the multiconvexity property of H. Some details of these applied problems can be found in the appendix. Our goal is to investigate the possibility of decomposing problem (1) into individual convex minimization subproblems over the individual subvectors, with the aim of computing a "blockwise directional stationary solution" of this problem; see section 2 for the definition. This goal is accomplished by the combination of three techniques: a well-known block coordinate method (BCDM) decomposing the objective and utilizing the partially linearized Gauss–Seidel (GS) scheme to update the subvectors sequentially; an ADMM to decouple the coupled linear constraints; and the handling of the pointwise maxima structure using the ε-argmax idea introduced in [40].

The contributions of our work are severalfold. One that distinguishes it from those in the references [4, 3, 25, 37, 29, 46, 47, 48] for nonconvex programs is that our combined BCDM-ADMM is shown to compute a d(irectional)-derivative based stationary solution [40] of problem (1). (See the latter reference for discussion of various concepts of stationary solutions of nondifferentiable DC programs.) The main convergence results are given by Theorem 4.3 for subsequential convergence and Proposition 4.4 for boundedness and sequential convergence, with the needed assumptions summarized in subsection 3.2 and proofs detailed in section 4. An example is provided to illustrate that without the special technique to handle the pointwise max functions for this class of problems, the standard ADMM as described in the cited literature computes a point that is far from being d-stationary, and thus has no chance to be a minimizer. Another contribution of our work is that the distinguished variable x^{I+1} is allowed to be constrained by a private closed convex set X^{I+1} that is linked to the coupling constraint in a certain way. This is a significant extension of the existing literature where such a variable is typically not privately constrained.

A departure of our work from the above-cited references is that while some of them have employed the Kurdyka–Lojasiewicz (KL) property [4, section 3.2] to establish the sequential convergence and also error bounds for the sequence of iterates produced by the ADMM algorithm in less general settings, we leave the treatment using the KL property for a subsequent work. Part of such a treatment would involve verifying or extending this property for the class of problems (1). Instead, we describe a probabilistic version of our deterministic algorithm that aims to alleviate the additional per-iteration computational effort of this algorithm for solving the class of nonconvex programs in question. Details including a convergence proof of this randomized algorithm are presented in section 5.

To close this introduction, we note that the framework (1) includes the case in which the coupling constraint is not present. In this case, the algorithm reduces to that of a convex programming based (partially linearized) block coordinate descent method (BCDM) of Gauss–Seidel type for solving a nonsmooth, nonconvex optimization problem with separable constraints and a particular nonsmooth structure in the objective; the convergence of such a method to a directional stationary point of the problem with such a pointwise max objective is a new result in the vast literature of the family of BCDMs—see [50] for a recent survey, [42, 30, 53] for works related to ours, and [52] for a stochastic version of the BCDM.

2. Preliminaries. In this section, we summarize some preliminary materials needed for the rest of the paper. First is the Bregman distance [8] that is a well-studied "pseudo metric" and has played an important role in various areas and algorithmic design for optimization problems. Formally, given a convex differentiable function ψ defined on an open convex domain $\mathcal{D} \subseteq \mathbb{R}^n$, the *Bregman distance* $D_\psi(x, y)$ between two vectors x and y in $\mathcal{D}$ is defined as

$$D_\psi(x, y) \triangleq \psi(x) - \psi(y) - \nabla\psi(y)^T(x - y).$$

Clearly, $D_\psi(x, y)$ reduces to $\|x - y\|^2$ if $\psi(x) = \|x\|_2^2$. We refer the reader to [8] for properties of the Bregman distance, which we will use freely in the analysis. We recall that a function $f : \mathcal{D} \to \mathbb{R}$ is α-strongly convex if, for some scalar $\alpha > 0$,

$$f(\tau x + (1-\tau)y) \leq \tau f(x) + (1-\tau) f(y) - \frac{\alpha}{2} \tau (1-\tau) \| x - y \|^2 \quad \forall \tau \in [0,1] \text{ and } x, y \in \mathcal{D}.$$

It is easy to show that if x^* is a minimizer of such a function f on a closed convex subset X of $\mathcal{D}$, then $f(x^*) \leq f(x) - \frac{\alpha}{2} \|x^* - x\|^2$ for all $x \in X$.

Given a closed convex set $X \subseteq \mathbb{R}^n$, the *lineality space* of X [43], denoted L_X, is the linear subspace consisting of vectors v such that $x + \tau v \in X$ for all scalars $\tau \in \mathbb{R}$. We denote the orthogonal complement of L_X by $L_X^{\perp}$. The normal cone, denoted $\mathcal{N}(x; X)$, of X at a vector $x \in X$ consists of all vectors u such that $u^T(y - x) \leq 0$. It is clear that $\mathcal{N}(x; X) \subseteq L_X^{\perp}$ for any $x \in X$; thus $\mathcal{N}(x; X) - \mathcal{N}(x'; X) \subseteq L_X^{\perp}$ for any x and x' in X.

2.1. Directional derivative-based stationarity. In general, given a constrained optimization problem $\text{minimize}_{x \in \mathcal{X}}\, \theta(x)$ with $\mathcal{X} \subseteq \mathbb{R}^N$ being a closed and convex set contained in the open convex set Υ and θ being directionally differentiable with directional derivatives at a vector $x \in \Upsilon$ given by

$$\theta'(x; d) \triangleq \lim_{\tau \downarrow 0} \frac{\theta(x + \tau d) - \theta(x)}{\tau}, \quad d \in \mathbb{R}^N,$$

a feasible vector $\bar{x} \in \mathcal{X}$ is a *d(irectional)-stationary point* if $\theta'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in \mathcal{X}$. When θ is continuously differentiable, the latter condition becomes $0 \in \nabla\theta(\bar{x}) + \mathcal{N}(\bar{x}; \mathcal{X})$. In the case of the objective in (1), noticing that $(\max_{1 \leq j \leq J_i} g_{ij})'(x^i; d^i) = \max_{j \in \mathcal{M}(x^i)} \nabla g_{ij}(x^i)^T d^i$ for all $x^i \in \Omega^i$ and all $d^i \in \mathbb{R}^{n_i}$, where

$$\mathcal{M}_i(x^i) \triangleq \operatorname*{argmax}_{1 \leq j \leq I_i} g_{ij}(x^i) = \left\{ j \mid g_{ij}(x^i) = \max_{1 \leq k \leq J_i} g_{ik}(x^i) \right\}$$

is the index set of maximizing functions in the pointwise maximum function

$$\max_{1 \leq k \leq J_i} g_{ik}(x^i),$$

we say that $\bar{x} \in \widehat{\mathbf{X}}$ is a *directional derivative based stationary solution* or a *blockwise d-stationary solution* of (1) if

$$\text{(2)} \quad \nabla\varphi(\bar{x})^T(x - \bar{x}) + \sum_{i=1}^{I+1} H(\bullet, \bar{x}^{-i})'(\bar{x}^i; x^i - \bar{x}^i) - \sum_{i=1}^{I} \max_{j \in \mathcal{M}_i(\bar{x}^i)} \nabla g_{ij}(\bar{x}^i)^T(x^i - \bar{x}^i) \geq 0 \quad \forall x \in \widehat{\mathbf{X}}.$$

If H is jointly directionally differentiable in all its components, and $H'(x; d) \geq \sum_{i=1}^{I+1} H(\bullet, x^{-i})'(x^i; d^i)$, then a block directional stationary point is indeed a directional stationary solution of (1) in its standard sense. Clearly, the condition (2) is equivalent to the following: for every tuple $\mathbf{j} = (j_i)_{i=1}^I$ with $j_i \in \mathcal{M}_i(\bar{x}^i)$ for all $i = 1, \ldots, I$,

$$\nabla\varphi(\bar{x})^T(x - \bar{x}) + \sum_{i=1}^{I+1} H(\bullet, \bar{x}^{-i})'(\bar{x}^i; x^i - \bar{x}^i) \geq \sum_{i=1}^{I} \nabla g_{ij_i}(\bar{x}^i)^T(x^i - \bar{x}^i) \quad \forall x \in \widehat{\mathbf{X}},$$

because $\max_{j \in \mathcal{M}_i(\bar{x}^i)} \nabla g_{ij}(\bar{x}^i)^T(x^i - \bar{x}^i) \geq \nabla g_{ij_i}(\bar{x}^i)^T(x^i - \bar{x}^i)$ for all $i = 1, \ldots, I$ and $j_i \in \mathcal{M}_i(\bar{x}^i)$, with equality holding for at least one such j_i. Based on this observation and employing the Bregman distance, we can characterize the stationarity condition (2) in terms of an optimality property of $\bar{x}$ in multiple convex programs over the same feasible set $\widehat{\mathbf{X}}$, which has a separable objective if the Bregman function is chosen to be separable in the subvectors.

Proposition 2.1. *Let φ be a differentiable function on Ω, H be multiconvex on $\mathbf{X}$, and g_{ij} be convex and differentiable on Ω^i for each $i = 1, \ldots, I$ and $j = 1, \ldots, J_i$. Let $\beta > 0$ be an arbitrary scalar and ψ be a convex differentiable function defined on Ω. Let $\bar{x} \triangleq (\bar{x}^i)_{i=1}^{I+1} \in \widehat{\mathbf{X}}$ be a feasible tuple of (1). Then (2) holds if and only if, for every tuple $\mathbf{j} \triangleq (j_i)_{i=1}^{I}$ with $j_i \in \mathcal{M}_i(\bar{x}^i)$ for all $i = 1, \ldots, I$,*

$$\bar{x} \in \underset{x \in \widehat{\mathbf{X}}}{\operatorname{argmin}} \underbrace{\bar{\theta}_{\mathbf{j}}(x)}_{\text{cvx in } x} \triangleq \nabla\varphi(\bar{x})^T(x - \bar{x}) + \sum_{i=1}^{I+1} H_i(x^i; \bar{x}^{-i})$$
$$- \sum_{i=1}^{I} \nabla g_{ij_i}(\bar{x}^i)^T(x^i - \bar{x}^i) + D_\psi(x; \bar{x}) + \frac{\beta}{2} \sum_{i=1}^{I+1} \| A^i x^i - A^i \bar{x}^i \|_2^2.$$

Proof. This follows from the following two facts: (a) $\nabla_x D_\psi(y, \bar{x})|_{y=\bar{x}} = 0$ and (b) $\nabla \| A^i x^i - A^i \bar{x}^i \|_2^2 |_{x^i = \bar{x}^i} = 0$. ∎

In the proposed method for computing a stationary point of (1) satisfying (2), we need to make use of the ε-argmax of the family of pointwise maximum functions for a given $\varepsilon > 0$; specifically, for each $x^i \in X^i$,

$$\mathcal{M}_{\varepsilon,i}(x^i) \triangleq \left\{ j \mid g_{ij}(x^i) \geq \max_{1 \leq k \leq J_i} g_{ik}(x^i) - \varepsilon \right\}.$$

This extended argmax set has the property that if $\{x^{\nu,i}\}_{\nu=1}^{\infty}$ is a sequence converging to x^i, then for any $\varepsilon > 0$ we have, for all ν sufficiently large,

$$\mathcal{M}_i(x^{\nu,i}) \subseteq \mathcal{M}_i(x^i) \subseteq \mathcal{M}_{\varepsilon,i}(x^{\nu,i}).$$

The second inclusion is the cornerstone for the convergence of the algorithm to a blockwise d-stationary solution of (1) to be presented in the next section. This inclusion suggests that in the generation of a sequence $\{x^{\nu,i}\}$ converging to a limit $\bar{x}^i$, in order to capture all the maximizing functions $\{g_{ij}(\bar{x}^i)\}_{j \in \mathcal{M}_i(\bar{x}^i)}$, it is essential to include functions g_{ij} that are ε away from the maximizing ones at each $x^{\nu,i}$. As is apparent from the stationarity condition (2), the inclusion of all maximizing functions $\{g_{ij}(\bar{x}^i)\}_{j \in \mathcal{M}_i(\bar{x}^i)}$ at $\bar{x}$ is part of the requirement for this point to be blockwise directionally stationary.

We define the augmented Lagrangian function as follows: for a given scalar $\beta > 0$, with z denoting the Lagrange multiplier of the coupling constraint $b = \sum_{i=1}^{I+1} A^i x^i$,

$$\mathcal{L}_\beta(x, z) \triangleq \theta(x) + z^T \left[b - \sum_{i=1}^{I+1} A^i x^i \right] + \frac{\beta}{2} \left\| b - \sum_{i=1}^{I+1} A^i x^i \right\|_2^2$$
$$= \theta(x) + \frac{\beta}{2} \left\| b - \sum_{i=1}^{I+1} A^i x^i - \frac{z}{\beta} \right\|_2^2 - \frac{\| z \|_2^2}{2\beta}.$$

2.2. Subgradient-based stationarity. Based on variational calculus for nonsmooth functions, stationarity can be defined in terms the concept of subgradients as follows. Consider the optimization problem

$$\underset{x \in \mathcal{X}}{\text{minimize}}\, f(x), \tag{3}$$

where, avoiding an extended-valued objective, we explicitly express the constraints by the closed convex set $\mathcal{X}$ in $\mathbb{R}^N$. A vector $v \in \mathbb{R}^N$ is a *subgradient* of f at a point x if there exist a sequence $\{x^k\}$ of vectors converging to x and a sequence $\{v^k\}$ of vectors converging to v such that for every k,

$$\liminf_{y \to x^k \text{ and } y \neq x^k} \frac{f(y) - f(x^k) - (\, y - x^k\,)^T v^k}{\|\, y - x^k\|} \;\geq\; 0.$$

We denote the set of subgradients of f at x by the standard notation $\partial f(x)$. We say that a vector $\bar{x} \in \mathcal{X}$ is a *subgradient-based stationary point* of (3) if $0 \in \partial f(\bar{x}) + \mathcal{N}(\bar{x}; \mathcal{X})$.

To clarify the difference between a subgradient-based stationary point and a directional stationary solution for the class of objectives in (1), we consider for simplicity the case where $f(x) = \varphi(x) - \max_{1 \leq i \leq I} g_i(x)$ with φ and g_i all continuously differentiable. By [38, Proposition 1.113], we have $\partial f(x) \subseteq \nabla \varphi(x) - \{\nabla g_i(x) \mid i \in \mathcal{M}(x)\}$. It follows from this inclusion that for any $v \in \partial f(x)$ and all $d \in \mathbb{R}^N$, $v^T d \geq \nabla \varphi(x)^T d - \max_{i \in \mathcal{M}(x)} \nabla g_i(x)^T d = f\,'(x; d)$. Thus if $\bar{x}$ is a directional stationary solution of (3), i.e., if $f\,'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in \mathcal{X}$, then $\bar{x}$ is subgradient-based stationary. Nevertheless, the example below shows that the converse is false.

EXAMPLE 1. Consider the univariate function $f(x) = \frac{3}{2}\,x^2 - \max(-x, 0) = \frac{3}{2}\,x^2 + \min(x, 0)$ and let $\mathcal{X}$ be the interval $[-1, 1]$. The function f has a unique directional stationary point on $\mathcal{X}$, namely $x = -1/3$, which is the (unique) global minimum of f on $\mathcal{X}$. Nevertheless, since $0 \in \partial \min(x, 0)\,|_{x=0}$, it follows that $x = 0$ is a subgradient based stationary point of f on the same interval. □

In summary, while the subgradient-based stationarity concept is supported by the rich theory of nonsmooth calculus [38, 44], such a stationary point can be quite unrelated to a minimizer of any kind. Therefore, while it may be possible to compute a subgradient-based stationary point as in the recent literature on the ADMM for nonsmooth nonconvex optimization problems, a question arises as to whether a decomposition algorithm can be designed to compute a point with a sharper stationarity property. The next section presents such an algorithm that utilities three ideas: the Gauss–Seidel sequential update scheme, an ADMM scheme to decouple the coupled linear constraint, and the ε-argmax idea to cover all potential binding functions in the pointwise maximum terms. Subsequently, a probabilistic version of the latter idea is also presented as a promising way to reduce the computational burden of the individual ε-argmax decomposition.

3. The combined BCDM and ADMM. Besides the use of a Bregman function for regularization, a distinguishing feature of our algorithm from existing ADMMs and BCDMs is the explicit treatment of the pointwise max term in the objective function. Previously appearing in [40], the use of a positive ε is essential to ensure the convergence to a desired directional stationary solution.

The algorithm below employs the sequential Gauss–Seidel idea. At iteration $\nu+1$ the most recently updated components $x^{\nu+1}_{<i} \triangleq (x^{\nu+1,k})_{k<i}$ along with those $x^{\nu}_{\geq i} \triangleq$

$(x^{\nu,k})_{k \geq i}$ from the immediate past iteration ν are employed to update the component $x^{\nu+1,i}$ in the current iteration. This is accomplished by solving $|\,\mathcal{M}_{\varepsilon,i}(x^{\nu,i})\,|$ convex subproblems, each defined, for $i \leq I$ and $j \in \mathcal{M}_{\varepsilon,i}(x^{\nu,i})$, by the function

$$\begin{aligned}
&\underbrace{\theta_{ij}^{\nu+1}(x^i; x^{\nu+1}_{<i}, x^{\nu}_{\geq i}; z^{\nu})}_{\text{convex in } x^i} \\
&\triangleq \underbrace{\nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)^T (\,x^i - x^{\nu,i}\,) - \underbrace{\nabla g_{ij}(x^{\nu,i})^T (\,x^i - x^{\nu,i}\,)}_{\text{varies with } j} - (\,z^{\nu}\,)^T A^{\,i} x^i}_{\text{linear in } x^i} \\
&+ \underbrace{H\left(x^{\nu+1}_{<i}, x^i, x^{\nu}_{>i}\right) + D_{\psi_i}(x^i, x^{\nu,i})}_{\text{convex in } x^i} + \frac{\beta}{2} \underbrace{\left\| b - \sum_{k<i} A^{\,k} x^{\nu+1,k} - A^{\,i} x^i - \sum_{k>i} A^{\,k} x^{\nu,k} \right\|_2^2}_{\text{augmented Lagrangian term}}
\end{aligned}$$

followed by a selection based on the test function

$$\begin{aligned}
\theta_{ij}^{\nu+1,\text{test}}(x^i; x^{\nu+1}_{<i}, x^{\nu}_{\geq i}; z^{\nu}) \;&\triangleq\; \nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)^T (\,x^i - x^{\nu,i}\,) + H\left(x^{\nu+1}_{<i}, x^i, x^{\nu}_{>i}\right) \\
-g_{ij}(x^i) + D_{\psi_i}(x^i, x^{\nu,i}) - (\,z^{\nu}\,)^T A^{\,i} x^i + \frac{\beta}{2} &\left\| b - \sum_{k<i} A^{\,k} x^{\nu+1,k} - A^{\,i} x^i - \sum_{k>i} A^{\,k} x^{\nu,k} \right\|_2^2.
\end{aligned}$$

Also define

$$\begin{aligned}
\underbrace{\theta_{I+1}^{\nu+1}(x^{I+1}; x^{\nu+1}_{\leq I}; z^{\nu})}_{\text{convex in } x^{I+1}} &\triangleq \nabla_{x^{I+1}}\varphi\left(x^{\nu+1}_{\leq I}, x^{\nu,I+1}\right)^T (\,x^{I+1} - x^{\nu,I+1}\,) \\
&\quad + H\left(x^{\nu+1}_{\leq I}, x^{I+1}\right) + D_{\psi_{I+1}}(x^{I+1}, x^{\nu,I+1}) - (\,z^{\nu}\,)^T A^{\,I+1} x^{I+1} \\
&\quad + \frac{\beta}{2} \left\| b - \sum_{k \leq I} A^{\,k} x^{\nu+1,k} - A^{\,I+1} x^{I+1} \right\|_2^2.
\end{aligned}$$

The combined BCDM-ADMM is presented in Algorithm 1. The special case of the algorithm where the coupling constraint is absent results in the BCDM for partitioned constrained problems. In this case, the multiplier z is absent and Step 2 is not needed.

Each of the subproblems in Step 1a corresponds to a function g_{ij} for $j \in \mathcal{M}_{\varepsilon,i}(x^{\nu,i})$ that is linearized at the current iterate $x^{\nu,i}$ along with a similar linearization of the function $\varphi(x^{\nu+1}_{<i}, \bullet, x^{\nu}_{>i})$ at the same iterate. There are two noteworthy features of these subproblems: the blockwise handling of the variables of the function H to exploit its multiconvexity and the use of separable Bregman functions $D_{\psi_i}(\bullet, x^{\nu,i})$ associated with given convex functions $\psi_i(x^i)$ for $i = 1, \ldots, I$ for regularization. Once the components $x^{\nu+1,i}$ are computed for all $i = 1, \ldots, I$, they are included in the vector $x^{\nu+1}_{\leq I} \triangleq (x^{\nu+1,k})_{k \leq I}$ to update the last component $x^{\nu+1,I+1}$ in a similar way. The algorithm also employs a positive threshold $\underline{\beta}$ that will be specified subsequently. Overall, the number of convex subprograms to compute the entire tuple $x^{\nu+1}$ is $\prod_{i=1}^{I} |\,\mathcal{M}_{\varepsilon,i}(x^{\nu,i})\,|$ for $x^{\nu+1}_{\leq I}$ and one more for the last component $x^{\nu+1,I+1}$. This amount of computation is significant if the latter Cartesian product contains

Algorithm 1 The deterministic combined BCDM-ADMM.

Let $x^0 \triangleq (x^{0,i})_{i=1}^{I+1} \in \mathbf{X}$ and z^0 be given. Also let scalars $\varepsilon > 0$ and $\beta > \underline{\beta}$ be given. Let $\{\psi_i\}_{i=1}^{I+1}$ be a family of convex differentiable functions.

while a prescribed termination criterion is not satisfied **do**

Step 1a. Solve $|\mathcal{M}_{\varepsilon,i}(x^{\nu,i})|$ convex subprograms

$$\left\{ \widehat{x}^{\,\nu+1,i,j} \in \operatorname*{argmin}_{x^i \in X^i} \theta_{ij}^{\nu+1}(x^i; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu) \right\}_{j \in \mathcal{M}_{\varepsilon,i}(x^{\nu,i})}. \tag{4}$$

Define $x^{\nu+1,i} \triangleq \widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}$, where $\widehat{j}_i^{\,\nu} \in \operatorname*{argmin}_{j \in \mathcal{M}_{\varepsilon,i}(x^{\nu,i})} \theta_{ij}^{\nu+1,\text{test}}(\widehat{x}^{\,\nu,i,j}; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu)$.

Step 1b. Let

$$x^{\nu+1,I+1} \in \operatorname*{argmin}_{x^{I+1} \in X^{I+1}} \theta_{I+1}^{\nu+1}(x^{I+1}; x_{\leq I}^{\nu+1}; z^\nu).$$

Step 2. Update the Lagrange multiplier:

$$z^{\nu+1} \triangleq z^\nu + \beta \left(b - \sum_{i=1}^{I+1} A^i x^{\nu+1,i} \right).$$

end while

return (x, z)

Algorithm 2 The deterministic BCDM.

The convex programs (4) in Step 1a (and similarly in Step 1b) of Algorithm 1 simplify to

$$\left\{ \begin{array}{l} \operatorname*{minimize}_{x^i \in X^i} \left[\nabla_{x^i} \varphi \left(x_{<i}^{\nu+1}, x_{\geq i}^{\nu} \right)^T (x^i - x^{\nu,i}) + H \left(x_{<i}^{\nu+1}, x^i, x_{>i}^{\nu} \right) \right. \\ \qquad \left. - \nabla g_{ij}(x^{\nu,i})^T (x^i - x^{\nu,i}) + D_{\psi_i}(x^i, x^{\nu,i}) \right] \end{array} \right\}_{j \in \mathcal{M}_{\varepsilon,i}(x^{\nu,i})}. \tag{5}$$

Step 2 is not needed.

a large number of elements. Subsequently, we will introduce a probabilistic version of the algorithm wherein we randomly select only one element from each $\mathcal{M}_{\varepsilon,i}(x^{\nu,i})$, thus simplifying the computational effort per iteration in the implementation of the algorithm.

The test function $\theta_{ij}^{\nu+1,\text{test}}(x^i; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu)$ differs from $\theta_{ij}^{\nu+1}(\bullet; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu)$ and $\mathcal{L}_\beta(x_{<i}^{\nu+1}, \bullet, x_{>i}^{\nu}, z^\nu)$ as follows: g_{ij} is not linearized in $\theta_{ij}^{\nu+1,\text{test}}(\bullet; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu)$, it is in $\theta_{ij}^{\nu+1}(\bullet; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu)$; a linearization of $\varphi(x_{<i}^{\nu+1}, \bullet, x_{>i}^{\nu})$ at $x^{\nu,i}$ and a single function g_{ij} are used in $\theta_{ij}^{\nu+1,\text{test}}(\bullet; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu)$; in $\mathcal{L}_\beta(x_{<i}^{\nu+1}, \bullet, x_{>i}^{\nu}, z^\nu)$, $\varphi(x_{<i}^{\nu+1}, \bullet, x_{>i}^{\nu})$ is not linearized at $x^{\nu,i}$ and all the functions g_{ij} for $j \in \mathcal{J}_i$ are used. Incidentally, if the function φ is convex (which is the case in [40] without the coupling constraint), then we could use $\varphi(x_{<i}^{\nu+1}, \cdot, x_{>i}^{\nu})$ without linearization in both $\theta_{ij}^{\nu+1}(\bullet; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu)$ and $\theta_{ij}^{\nu+1,\text{test}}(\bullet; x_{<i}^{\nu+1}, x_{\geq i}^{\nu}; z^\nu)$.

There is a variation of the above two algorithms that is worth mentioning, namely, instead of linearizing the first summand φ, we may replace Step 1a by the problem

$$\widehat{x}^{\,\nu+1,i,j} \in \underset{x^i \in X^i}{\operatorname{argmin}} \left[\varphi\left(x^{\nu+1}_{<i}, x^i, x^{\nu}_{>i}\right) - \nabla g_{ij}(x^{\nu,i})^T (\, x^i - x^{\nu,i}\,) - (\, z^{\nu}\,)^T A^{\,i} x^i \right.$$

$$\left. + H\left(x^{\nu+1}_{<i}, x^i, x^{\nu}_{>i}\right) + D_{\psi_i}(x^i, x^{\nu,i}) + \frac{\beta}{2} \left\| b - \sum_{k<i} A^{\,k} x^{\nu+1,k} - A^{\,i} x^i - \sum_{k>i} A^{\,k} x^{\nu,k} \right\|_2^2 \right]$$

in which the function φ is not linearized. With the function ψ_i appropriately chosen so that the resulting function $\varphi(x^{\nu+1}_{<i}, \bullet, x^{\nu}_{>i}) + D_{\psi_i}(\bullet, x^{\nu,i})$ is strongly convex on X^i, the method can handle the case where the coupled function φ is not differentiable in the subvector x^i and for which the resulting subproblem can be efficiently solved. We omit further discussion of variations such as this one and instead focus on the algorithms as stated above; for a particular problem in which this strategy is employed fruitfully, see [34]. Needless to say, by using the function $\varphi(x^{\nu+1}_{<i}, \bullet, x^{\nu}_{>i})$ itself, it is expected that the solution of the subproblem is not difficult; moreover, if a global minimizing property of the solution is needed in the convergence proof, such a minimizing solution can be practically computed.

3.1. A numerical example. We use a modification of Example 1 to illustrate that variations of the algorithm could fail to converge to a directional stationary solution; one such variation does not employ a positive ε in the sets $\mathcal{M}_{\varepsilon,i}(x^{\nu,i})$ to set up the iterations.

EXAMPLE 2. Consider the following variation of Example 1 formulated to fit the framework of (1):

$$\underset{x_1, x_2}{\text{minimize}} \quad 2x_1^2 - \tfrac{1}{2}\, x_2^2 - \max(-x_1, 0) + \tfrac{1}{2}\, x_1 x_2 \quad \text{subject to}$$

$$x_1 - x_2 = 0 \text{ (coupling constraint)} \quad \text{and} \quad -1 \,\leq\, x_1 \,\leq\, 1 \text{ (private constraint).}$$

Similar to the previous example, it can be shown that this problem has two subgradient-based stationary solutions, $(0, 0)$ and $(-1/4, -1/4)$, among which only the latter is directional stationary. Under the identifications

$$\varphi(x_1, x_2) = \tfrac{1}{2} x_1 x_2, \quad H(x_1, x_2) = 2x_1^2 - \tfrac{1}{2} x_2^2, \quad I = 1,$$
$$J_1 = 2, \quad g_{11}(x_1) = 0, \quad \text{and} \quad g_{12}(x_1) = -x_1,$$

we apply the combined BCDM-ADMM with $\varepsilon = 0$ to this problem using the quadratic Bregman functions $\psi_1(x_1) = \frac{c}{2}\, x_1^2$ and $\psi_2(x_2) = \frac{c}{2}\, x_2^2$ for a constant $c > 1$. For any $\beta > 0$, we have,

$$x_1^{\nu+1} = \underset{-1 \leq x_1 \leq 1}{\operatorname{argmin}} \left[\frac{x_2^{\nu}}{2} (\, x_1 - x_1^{\nu}\,) + 2x_1^2 + \frac{c}{2} (\, x_1 - x_1^{\nu}\,)^2 - z^{\nu}\, x_1 + \frac{\beta}{2} (\, x_1 - x_2^{\nu}\,)^2 \right]$$

(when g_{11} is picked)

$$= \Pi_{[-1,1]} \left(\frac{-\frac{x_2^{\nu}}{2} + c\, x_1^{\nu} + z^{\nu} + \beta\, x_2^{\nu}}{4 + c + \beta} \right) \quad \text{(where } \Pi \text{ denotes the Euclidean projection)}$$

$$\begin{aligned} x_2^{\nu+1} &= \underset{x_2}{\operatorname{argmin}} \left[-\frac{1}{2}\, x_2^2 + \frac{x_1^{\nu+1}}{2}\,(\, x_2 - x_2^{\nu}\,) + \frac{c}{2}\,(\, x_2 - x_2^{\nu}\,)^2 + z^{\nu}\, x_2 + \frac{\beta}{2}\,(\, x_2 - x_1^{\nu+1}\,)^2 \right] \\ &= \frac{-\frac{x_1^{\nu+1}}{2} + c\, x_2^{\nu} - z^{\nu} + \beta\, x_1^{\nu+1}}{\beta + c - 1}, \quad \text{and} \\ z^{\nu+1} &= z^{\nu} - \beta\,(\, x_1^{\nu+1} - x_2^{\nu+1}\,). \end{aligned}$$

A similar formula for $x_1^{\nu+1}$ when g_{12} is picked can be similarly derived. We compared our algorithm with the standard ADMM [47] where (i) no linearization was applied to the nonlinear terms in the objective function, including the max term, (ii) there was no Bregman regularization, and (iii) the β-parameter was chosen sufficiently large so that the subproblems are convex. We ran the iterations with three different starting points of (x_1^0, x_2^0, z^0). The results of the iterations are plotted in Figure 1, which clearly shows (a) the convergence of our algorithm (with $\varepsilon = 0.01$) to the d-stationary point of $(-1/4, -1/4)$ when started at all three points, and (b) the convergence to the origin with the standard ADMM. In addition, we also ran our algorithm as described above with $\varepsilon = 0$ and linearization of φ. While convergence to the d-stationary point of $(-1/4, -1/4)$ was obtained with the second and third starting point, convergence to the origin was obtained when the algorithm was initiated at the first starting point.

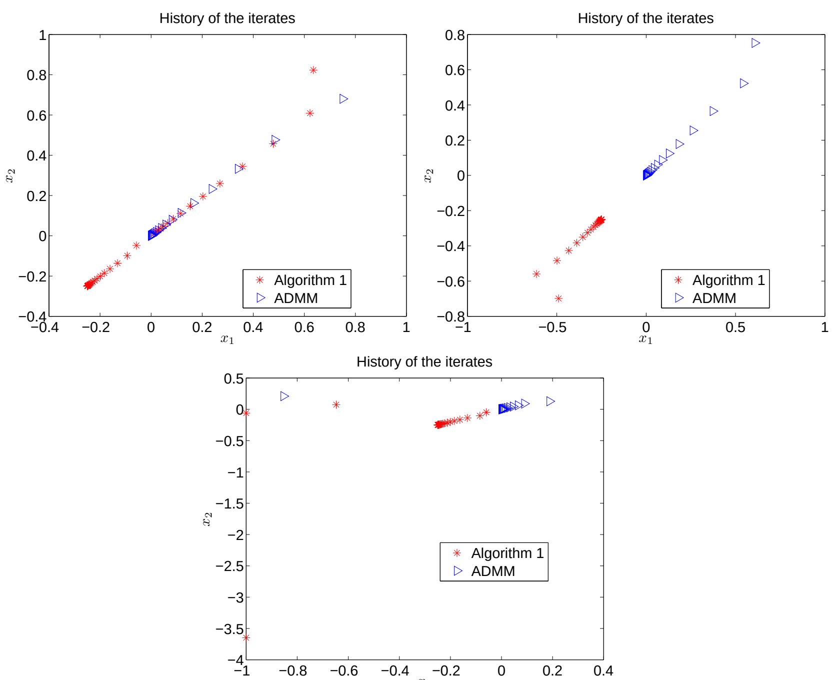

FIG. 1. *Algorithm* 1 *vs ADMM for Example* 2 *with different initial points:* $(1, 1, -1)$ *(top left);* $(-1, 1, 1)$ *(top right);* $(-10, -0.1, 10)$ *(bottom).*

In conclusion, this example shows the persistent convergence to a d-stationary solution of our algorithm with all three starting points; such convergence is not guaranteed with variations of the algorithm. ∎

3.2. Assumptions. We begin by summarizing the assumptions needed for the convergence proof of the combined BCDM-ADMM. These are fairly standard in the literature of this kind of problems. However, there are two novel features: (a) the presence of the pointwise max terms that complicates the convergence analysis, and (b) the private constraint set X^{I+1} that is not required to be polyhedral.

(A0) In addition to the basic setting in the definition of (1), the function φ is multi-LC^1 on $\mathbf{X}$ with modulus $\{ \text{Lip}^i_{\nabla\varphi} \}_{i=1}^{I+1}$, i.e., for all $i = 1, \dots, I+1$,

$$\| \nabla_{x^i}\varphi(x) - \nabla_{x^i}\varphi(y) \|_2 \,\leq\, \text{Lip}^i_{\nabla\varphi} \| x - y \|_2 \quad \forall\, x, y \,\in\, \mathbf{X}; \tag{6}$$

also, the partial gradient function $\nabla_{x^{I+1}} H$ is Lipschitz continuous on $\mathbf{X}$ with modulus $\text{Lip}^{I+1}_{\nabla H}$. □

Note that (6) implies, by the mean-value theorem for vector functions [39, **3.2.12**], that for every $i = 1, \dots, I+1$, and any $x^{-i} \in X^{-i}$ and any x^i and y^i in X^i,

$$\varphi(x^i, x^{-i}) - \varphi(y^i, x^{-i}) \,\leq\, \nabla_{x^i}\varphi(y^i, x^{-i})^T (x^i - y^i) + \frac{\text{Lip}^i_{\nabla\varphi}}{2} \| x^i - y^i \|_2^2. \tag{7}$$

The next assumption pertains to the choice of the Bregman functions and ensures in particular that each of the convex subprograms in Steps 1 and 2 has an optimal solution.

(A1) The functions ψ_i for $i = 1, \dots, I+1$ are σ_i-strongly convex on X^i; moreover ψ_{I+1} is LC^1 with $\text{Lip}_{\nabla\psi_{I+1}}$ being the Lipschitz modulus of the gradient $\nabla\psi_{I+1}$ on X^{I+1}. □

This assumption implies that each function $\theta^{\nu+1}_{ij}(\bullet, x^{\nu+1}_{<i}, x^{\nu}_{\geq i}; z^\nu)$ for $i = 1, \dots, I$ is strongly convex; thus each iterate $\widehat{x}^{\,\nu+1;i,j}$ is uniquely defined. So is $x^{\nu+1;I+1}$. The last assumption concerns the distinguished variable x^{I+1}. It is the reason why the formulation (1) requires separability in each of the functions g_{ij}. Specifically, for a nonseparable $g_{ij}(x)$, the duplication of variables via $\xi^i = x$ for all $i = 1, \dots, I$ will violate the assumption, thus jeopardizing the convergence proof. This assumption is not needed for the BCDM when the coupling constraint is absent. There are several equivalent ways to state the assumption. We first give a lemma asserting such equivalence.

LEMMA 3.1. *Let $X \subseteq \mathbb{R}^m$ be a closed convex set and $A \in \mathbb{R}^{\ell\times m}$. The following statements are equivalent:*

(a) *$[\, A^T\lambda + \mu = 0$ and $\mu \in L_X^\perp \,]$ implies $\lambda = 0$;*

(b) *there exists a positive constant, denoted $\gamma_{\min}$, such that*

$$\| A^T\lambda + \mu \|_2 \,\geq\, \sqrt{\gamma_{\min}}\, \| \lambda \|_2 \quad \forall\, \mu \,\in\, L_X^\perp \text{ and all } \lambda;$$

(c) *A has full row rank and $L_X^\perp \cap \text{Range}(A^T) = \{0\}$.*

Proof. (a) ⇒ (b). Consider the following optimization problem:

$$\underset{\lambda,\,\mu}{\text{minimize}}\ \| A^T\lambda + \mu \|_2^2 \quad \text{subject to} \quad \mu \,\in\, L_X^\perp \text{ and } \lambda \text{ satisfying } \| \lambda \|_1 \,= 1. \tag{8}$$

The feasible set is the union of finitely many polyhedra in the (λ, μ)-space. Since the objective is always nonnegative, it follows from the Frank–Wolfe theorem of quadratic programming [12, Theorem 2.8.1] that the program (8) attains a finite optimum objective value which must be nonnegative. If this value is zero, then there exists (λ, μ)

with $\lambda \neq 0$ and $\mu \in L_X^\perp$ such that $A^T\lambda + \mu = 0$. But this contradicts (a). Thus (8) attains a positive minimum objective value. This is sufficient to yield the existence of the desired scalar $\gamma_{\min}$.

(b) $\Rightarrow$ (c) $\Rightarrow$ (a). These implications are easy. ∎

Based on the above lemma, we introduce (recalling that X^{I+1} is closed and convex) the following assumption.

(A2) The pair (A^{I+1}, X^{I+1}) satisfies any one of the three conditions in Lemma 3.1. ∎

Two special cases of assumption (A2) are worth mentioning. One is the case in which $X^{I+1} = \mathbb{R}^{n_{I+1}}$ so that the distinguished variable x^{I+1} is not privately constrained. In this case, $L_{X^{I+1}}^\perp = \{0\}$ and (A2) reduces to the (standard) assumption that A^{I+1} has full row rank. The other special case is when X^{I+1} is a polyhedron given by, say,

$$X^{I+1} \triangleq \{x^{I+1} \in \mathbb{R}^{n_{I+1}} \mid C^{I+1}x^{I+1} \geq d^{I+1}\}$$

for some matrix C^{I+1} and vector d^{I+1} of appropriate order. In this case, (A2) holds if the matrix

$$\Xi^{I+1} \triangleq \begin{bmatrix} A^{I+1} \\ C^{I+1} \end{bmatrix} \begin{bmatrix} (A^{I+1})^T & (C^{I+1})^T \end{bmatrix}$$

is positive definite and the constant $\gamma_{\min}$ can be taken to be the smallest eigenvalue of Ξ^{I+1}. Since the positive definiteness of Ξ^{I+1} is equivalent to the implication that $(A^{I+1})^T\lambda + (C^{I+1})^T\xi = 0$ implies both $\lambda = 0$ and $\xi = 0$, it follows that (A2) is a significant weakening of this positive definiteness assumption, allowing in particular the multipliers of the constraints in X^{I+1} to be unbounded.

Associated with the Lipschitz constants and the constant $\gamma_{\min}$, we define, for a given vector $y^{I+1} \in X^{I+1}$, the modified augmented Lagrangian function:

$$(9)\quad \widehat{\mathcal{L}}_\beta(x, z; y^{I+1}) \triangleq \mathcal{L}_\beta(x, z) + \frac{4}{\beta\,\gamma_{\min}} \left[\left(\mathrm{Lip}_{\nabla\varphi}^{I+1} \right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}} \right)^2 \right] \| \, x^{I+1} - y^{I+1} \, \|_2^2.$$

The convergence of the algorithm relies on a nonincreasing and bounded-below property of the sequence of such modified augmented Lagrangian values $\{\widehat{\mathcal{L}}_\beta(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1})\}$ when the sequence of iterates $\{(x^{\nu+1}, z^{\nu+1})\}$ generated by the algorithm has an accumulation point. In turn, these crucial properties are derived under the choice of $\beta > \underline{\beta}$; see (11) for the definition of the lower bound $\underline{\beta}$. The boundedness of the sequence $\{(x^{\nu+1}, z^{\nu+1})\}$ is established under one of two growth assumptions on the objective function θ on the feasible set $\mathbf{X}$. See the discussion preceding Proposition 4.4 for details.

4. Convergence proof. The proof of convergence of the combined BCDM-ADMM is divided into several steps. First is a lemma that bounds the consecutive difference $\| \, z^{\nu+1} - z^\nu \, \|_2^2$ of the multipliers in terms of the differences of the primary variables. The lemma also bounds $\|z^{\nu+1}\|_2^2$ in terms of the latter variables. This lemma is where the special block x^{I+1} is needed.

LEMMA 4.1. *It holds that*

(10)

$$\gamma_{\min} \left\| z^{\nu+1} - z^{\nu} \right\|_2^2 \leq 4 \left\{ \left[\left(\text{Lip}_{\nabla\varphi}^{I+1} \right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}} \right)^2 \right] \| x^{\nu,I+1} - x^{\nu-1,I+1} \|_2^2 \right.$$

$$\left. + \left[\max\left(\left(\text{Lip}_{\nabla\varphi}^{I+1} \right)^2, \left(\text{Lip}_{\nabla\psi_{I+1}} \right)^2 \right) + \left(\text{Lip}_{\nabla H}^{I+1} \right)^2 \right] \sum_{i=1}^{I+1} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2 \right\}$$

$$\textit{and} \quad \gamma_{\min} \left\| z^{\nu+1} \right\|_2^2 \leq 2 \left\{ \left(\text{Lip}_{\nabla\varphi}^{I+1} + \text{Lip}_{\nabla\psi_{I+1}} \right)^2 \left\| x^{\nu+1,I+1} - x^{\nu,I+1} \right\|^2 \right.$$

$$\left. + \left[\left(\text{Lip}_{\nabla\varphi}^{I+1} + \text{Lip}_{\nabla H}^{I+1} \right) \| x^{\nu+1} \| + \| \nabla_{x^{I+1}} \varphi(0) \|_2 + \| \nabla_{x^{I+1}} H(0) \|_2 \right]^2 \right\}.$$

Proof. By applying the optimality conditions of the problem,

$$\underset{x^{I+1} \in X^{I+1}}{\text{argmin}} \; \theta_{I+1}^{\nu+1}(x^{I+1}; x_{\leq I}^{\nu+1}; z^{\nu}),$$

we deduce

$$\mu^{\nu,I+1} \triangleq \nabla_{x^{I+1}} \varphi \left(x_{i \leq I}^{\nu+1,i}, x^{\nu,I+1} \right) - (A^{I+1})^T z^{\nu} + \nabla_{x^{I+1}} H(x^{\nu+1}) + \nabla\psi_{I+1}(x^{\nu+1})$$

$$- \nabla\psi_{I+1}(x^{\nu}) + \beta \, (A^{I+1})^T \left(\sum_{k=1}^{I+1} A^k x^{\nu+1,k} - b \right) \in -\mathcal{N}(x^{\nu,I+1}; X^{I+1}),$$

yielding

$$(A^{I+1})^T z^{\nu+1} + \mu^{\nu,I+1} = \nabla_{x^{I+1}} \varphi \left(x_{i \leq I}^{\nu+1,i}, x^{\nu,I+1} \right) + \nabla_{x^{I+1}} H(x^{\nu+1})$$

$$+ \nabla\psi_{I+1}(x^{\nu+1,I+1}) - \nabla\psi_{I+1}(x^{\nu,I+1}).$$

Subtracting this equation from the one from the previous iteration, we deduce

$$\begin{aligned} &(A^{I+1})^T \left(z^{\nu+1} - z^{\nu} \right) + \left(\mu^{\nu+1,I+1} - \mu^{\nu,I+1} \right) \\ &\quad = \left[\nabla_{x^{I+1}} \varphi \left(x_{i \leq I}^{\nu+1,i}, x^{\nu,I+1} \right) - \nabla_{x^{I+1}} \varphi \left(x_{i \leq I}^{\nu,i}, x^{\nu-1,I+1} \right) \right] \\ &\quad\quad + \left[\nabla_{x^{I+1}} H(x^{\nu+1}) - \nabla_{x^{I+1}} H(x^{\nu}) \right] + \left[\nabla\psi_{I+1}(x^{\nu+1,I+1}) - \nabla\psi_{I+1}(x^{\nu,I+1}) \right] \\ &\quad\quad - \left[\nabla\psi_{I+1}(x^{\nu,I+1}) - \nabla\psi_{I+1}(x^{\nu-1,I+1}) \right]. \end{aligned}$$

Taking squared norms on both sides, employing Lemma 3.1, and making use of the Cauchy–Schwartz inequality and the assumed LC^1 conditions, we obtain

$$\begin{aligned} &\gamma_{\min} \left\| z^{\nu+1} - z^{\nu} \right\|_2^2 \\ &\quad \leq 4 \left\{ \left(\text{Lip}_{\nabla\varphi}^{I+1} \right)^2 \left[\sum_{i \leq I} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2 + \| x^{\nu,I+1} - x^{\nu-1,I+1} \|_2^2 \right] \right. \\ &\quad\quad + \left(\text{Lip}_{\nabla H}^{I+1} \right)^2 \| x^{\nu+1} - x^{\nu} \|_2^2 \\ &\quad\quad \left. + \left(\text{Lip}_{\nabla\psi_{I+1}} \right)^2 \left[\| x^{\nu+1,I+1} - x^{\nu,I+1} \|_2^2 + \| x^{\nu,I+1} - x^{\nu-1,I+1} \|_2^2 \right] \right\} \end{aligned}$$

$$= 4\Bigg\{ \left[\left(\text{Lip}_{\nabla \varphi}^{I+1} \right)^2 + \left(\text{Lip}_{\nabla H}^{I+1} \right)^2 \right] \sum_{i \leq I} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2$$

$$+ \left[\left(\text{Lip}_{\nabla H}^{I+1} \right)^2 + \left(\text{Lip}_{\nabla \psi_{I+1}} \right)^2 \right] \| x^{\nu+1,I+1} - x^{\nu,I+1} \|_2^2$$

$$+ \left[\left(\text{Lip}_{\nabla \varphi}^{I+1} \right)^2 + \left(\text{Lip}_{\nabla \psi_{I+1}} \right)^2 \right] \| x^{\nu,I+1} - x^{\nu-1,I+1} \|_2^2 \Bigg\},$$

establishing the desired bound on $\| z^{\nu+1} - z^\nu \|_2^2$. To obtain the bound on $\| z^{\nu+1} \|_2^2$, we have

$$(A^{I+1})^T z^{\nu+1} + \mu^{\nu,I+1} = \left[\nabla_{x^{I+1}} \varphi \left(x_{i \leq I}^{\nu+1,i}, x^{\nu,I+1} \right) - \nabla_{x^{I+1}} \varphi \left(x_{i \leq I}^{\nu+1,i}, x^{\nu+1,I+1} \right) \right]$$
$$+ \left[\nabla_{x^{I+1}} \varphi \left(x^{\nu+1} \right) - \nabla_{x^{I+1}} \varphi(0) \right] + \nabla_{x^{I+1}} \varphi(0) + \left[\nabla_{x^{I+1}} H(x^{\nu+1}) - \nabla_{x^{I+1}} H(0) \right]$$
$$+ \nabla_{x^{I+1}} H(0) + \left[\nabla \psi_{I+1}(x^{\nu+1,I+1}) - \nabla \psi_{I+1}(x^{\nu,I+1}) \right].$$

Taking norms on both sides, we deduce the desired bound on $\| z^{\nu+1} \|_2^2$. ∎

We next upper bound the differences

$$\text{diff}_{\beta,i}^\nu \triangleq \mathcal{L}_\beta \left(x_{\leq i}^{\nu+1}, x_{>i}^\nu, z^\nu \right) - \mathcal{L}_\beta \left(x_{<i}^{\nu+1}, x_{\geq i}^\nu, z^\nu \right) \quad \text{for } i = 1, \ldots, I+1,$$
$$\text{diff}_{\beta,z}^\nu \triangleq \mathcal{L}_\beta \left(x_{\leq I+1}^{\nu+1}, z^{\nu+1} \right) - \mathcal{L}_\beta \left(x_{\leq I+1}^{\nu+1}, z^\nu \right)$$

by invoking the subprograms in Steps 1a and 1b and the update formula of $z^{\nu+1}$ in Step 2. Adding the bounds for the above differences, we can in turn bound the difference $\mathcal{L}_\beta(x^{\nu+1}, z^{\nu+1}) - \mathcal{L}_\beta(x^\nu, z^\nu)$ of the augmented Lagrangian function at two consecutive tuples $(x^{\nu+1}, z^{\nu+1})$ and (x^ν, z^ν).

LEMMA 4.2. *It holds that*

$$\mathcal{L}_\beta \left(x^{\nu+1}, z^{\nu+1} \right) - \mathcal{L}_\beta \left(x^\nu, z^\nu \right)$$
$$\leq \frac{4}{\beta \gamma_{\min}} \left[\left(\text{Lip}_{\nabla \varphi}^{I+1} \right)^2 + \left(\text{Lip}_{\nabla \psi_{I+1}} \right)^2 \right] \| x^{\nu,I+1} - x^{\nu-1,I+1} \|_2^2$$
$$+ \sum_{i=1}^{I+1} \left\{ \frac{4}{\beta \gamma_{\min}} \left[\max \left(\left(\text{Lip}_{\nabla \varphi}^{I+1} \right)^2, \left(\text{Lip}_{\nabla \psi_{I+1}} \right)^2 \right) + \left(\text{Lip}_{\nabla H}^{I+1} \right)^2 \right] + \frac{\text{Lip}_{\nabla \varphi}^i - \sigma_i}{2} \right\} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2.$$

Proof. By the definition of the augmented Lagrangian function, we have

$$\text{diff}_{\beta,i}^\nu = \theta \left(x_{\leq i}^{\nu+1}, x_{>i}^\nu \right) - \theta \left(x_{<i}^{\nu+1}, x_{\geq i}^\nu \right) - (z^\nu)^T A^i (x^{\nu+1,i} - x^{\nu,i})$$
$$+ \frac{\beta}{2} \left\{ \left\| b - \sum_{k=1}^{i} A^k x^{\nu+1,k} - \sum_{k=i+1}^{I} A^k x^{\nu,k} \right\|_2^2 - \left\| b - \sum_{k=1}^{i-1} A^k x^{\nu+1,k} - \sum_{k=i}^{I} A^k x^{\nu,k} \right\|_2^2 \right\}.$$

In turn, since $x^{\nu+1,i} = \widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}$, we have, for all $i = 1, \dots, I$,

$$\begin{aligned}
&\theta\left(x^{\nu+1}_{\leq i}, x^{\nu}_{>i}\right) - \theta\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right) \\
&\quad = \left[\varphi\left(x^{\nu+1}_{\leq i}, x^{\nu}_{>i}\right) - \varphi\left(x^{\nu+1}_{<i} x^{\nu}_{\geq i}\right)\right] \\
&\qquad + \left[H\left(x^{\nu+1}_{\leq i}, x^{\nu}_{>i}\right) - H\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)\right] - \left[\max_{1\leq j\leq J_i} g_{ij}(x^{\nu+1,i}) - \max_{1\leq j\leq J_i} g_{ij}(x^{\nu,i})\right] \\
&\quad \leq \left[\varphi\left(x^{\nu+1}_{<i}, \widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}, x^{\nu}_{>i}\right) - \varphi\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)\right] \\
&\qquad + \left[H\left(x^{\nu+1}_{<i}, \widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}, x^{\nu}_{>i}\right) - H\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)\right] \\
&\qquad - \left[g_{i\widehat{j}_i^{\,\nu}}(\widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}) - g_{ij}(x^{\nu,i})\right] \quad (\text{for any } j \in \mathcal{M}_i(x^{\nu,i})) \\
&\quad \leq \nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)^T \left(\widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}} - x^{\nu,i}\right) \\
&\qquad + \left[H\left(x^{\nu+1}_{<i}, \widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}, x^{\nu}_{>i}\right) - H\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)\right] \\
&\qquad + \frac{\mathrm{Lip}^i_{\nabla\varphi}}{2} \| \widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}} - x^{\nu,i} \|_2^2 - \left[g_{i\widehat{j}_i^{\,\nu}}(\widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}) - g_{ij}(x^{\nu,i})\right],
\end{aligned}$$

where the last inequality follows from (7). Thus, by the definition of $\widehat{j}_i^{\,\nu}$ and using the gradient inequality $g_{ij}(\widehat{x}^{\,\nu+1,i,j}) - g_{ij}(x^{\nu,i}) \geq \nabla g_{ij}(x^{\nu,i})^T(\widehat{x}^{\,\nu+1,i,j} - x^{\nu,i})$ of the convex function g_{ij}, we deduce that, for any $j \in \mathcal{M}_i(x^{\nu,i})$,

$$\begin{aligned}
\mathrm{diff}^{\nu}_{\beta,i} \leq\; & \nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)^T \left(\widehat{x}^{\,\nu+1,i,j} - x^{\nu,i}\right) \\
& + \left[H\left(x^{\nu+1}_{<i}, \widehat{x}^{\,\nu+1,i,j}, x^{\nu}_{>i}\right) - H\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)\right] \\
& - \left[D_{\psi_i}(\widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}, x^{\nu,i}) - D_{\psi_i}(\widehat{x}^{\,\nu+1,i,j}, x^{\nu,i})\right] + \frac{\mathrm{Lip}^i_{\nabla\varphi}}{2} \| x^{\,\nu+1,i} - x^{\nu,i} \|_2^2 \\
& - \nabla g_{ij}(x^{\nu,i})^T \left(\widehat{x}^{\,\nu+1,i,j} - x^{\nu,i}\right) - (z^{\nu})^T A^i \left(\widehat{x}^{\,\nu+1,i,j} - x^{\nu,i}\right) + \\
& + \frac{\beta}{2}\left\{ \left\| b - \sum_{k=1}^{i-1} A^k x^{\nu+1,k} - A^i \widehat{x}^{\,\nu+1,i,j} - \sum_{k=i+1}^{I} A^k x^{\nu,k} \right\|_2^2 \right. \\
& \qquad \left. - \left\| b - \sum_{k=1}^{i-1} A^k x^{\nu+1,k} - \sum_{k=i}^{I} A^k x^{\nu,k} \right\|_2^2 \right\},
\end{aligned}$$

which yields

$$\begin{aligned}
\mathrm{diff}^{\nu}_{\beta,i} \leq\; & \nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)^T (\widehat{x}^{\,\nu+1,i,j} - x^{\nu,i}) \\
& + \left[H\left(x^{\nu+1}_{<i}, \widehat{x}^{\,\nu+1,i,j}, x^{\nu}_{>i}\right) - H\left(x^{\nu+1}_{<i}, x^{\nu}_{\geq i}\right)\right] \\
& + \frac{\mathrm{Lip}^i_{\nabla\varphi} - \sigma_i}{2} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2 \quad (\text{by the } \sigma_i\text{-strong convexity of } \psi_i) \\
& + D_{\psi_i}(\widehat{x}^{\,\nu+1,i,j}, x^{\nu,i}) - \nabla g_{ij}(x^{\nu,i})^T(\widehat{x}^{\,\nu+1,i,j} - x^{\nu,i}) \\
& - (z^{\nu})^T A^i \left(\widehat{x}^{\,\nu+1,i,j} - x^{\nu,i}\right)
\end{aligned}$$

$$+ \frac{\beta}{2}\left\{ \left\| b - \sum_{k=1}^{i-1} A^k x^{\nu+1,k} - A^i \widehat{x}^{\,\nu+1,i,j} - \sum_{k=i+1}^{I} A^k x^{\nu,k} \right\|_2^2 - \left\| b - \sum_{k=1}^{i-1} A^k x^{\nu+1,k} - \sum_{k=i}^{I} A^k x^{\nu,k} \right\|_2^2 \right\}$$

$$\leq \frac{\mathrm{Lip}^i_{\nabla\varphi} - \sigma_i}{2} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2 \quad \text{(by the definition of } \widehat{x}^{\,\nu+1,i,j}).$$

Similarly, we can establish the desired bound for $\mathrm{diff}^\nu_{\beta,I+1}$ by dropping the g_{ij} functions. To bound $\mathrm{diff}^\nu_{\beta,z}$, we note that

$$\mathrm{diff}^\nu_{\beta,z} = (z^{\nu+1} - z^\nu)^T \left[b - \sum_{i=1}^{I+1} A^i x^{\nu+1,i} \right] = \frac{1}{\beta} \| z^{\nu+1} - z^\nu \|_2^2$$

$$\leq \frac{4}{\beta\,\gamma_{\min}} \left\{ \left[\max\left(\left(\mathrm{Lip}^{I+1}_{\nabla\varphi} \right)^2, \left(\mathrm{Lip}_{\nabla\psi_{I+1}} \right)^2 \right) + \left(\mathrm{Lip}^{I+1}_{\nabla H} \right)^2 \right] \sum_{i=1}^{I+1} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2 + \left[\left(\mathrm{Lip}^{I+1}_{\nabla\varphi} \right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}} \right)^2 \right] \| x^{\nu,I+1} - x^{\nu-1,I+1} \|_2^2 \right\} \quad \text{(by Lemma 4.1).}$$

Adding this to the bounds

$$\mathrm{diff}^\nu_{\beta,i} \leq \frac{\mathrm{Lip}^i_{\nabla\varphi} - \sigma_i}{2} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2,$$

for all $i = 1, \ldots, I+1$, we deduce the desired bound on $\mathcal{L}_\beta(x^{\nu+1}, z^{\nu+1}) - \mathcal{L}_\beta(x^\nu, z^\nu)$. □

Recalling the modified augmented Lagrangian function (9), from Lemma 4.2 we deduce that

$$\widehat{\mathcal{L}}_\beta(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}) - \widehat{\mathcal{L}}_\beta(x^\nu, z^\nu; x^{\nu-1,I+1})$$

$$\leq \sum_{i=1}^{I} \left\{ \frac{4}{\beta\,\gamma_{\min}} \left[\max\left(\left(\mathrm{Lip}^{I+1}_{\nabla\varphi} \right)^2, \left(\mathrm{Lip}_{\nabla\psi_{I+1}} \right)^2 \right) + \left(\mathrm{Lip}^{I+1}_{\nabla H} \right)^2 \right] + \frac{\mathrm{Lip}^i_{\nabla\varphi} - \sigma_i}{2} \right\} \| x^{\nu+1,i} - x^{\nu,i} \|_2^2$$

$$+ \left\{ \frac{4}{\beta\,\gamma_{\min}} \left[2\left(\left(\mathrm{Lip}^{I+1}_{\nabla\varphi} \right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}} \right)^2 \right) + \left(\mathrm{Lip}^{I+1}_{\nabla H} \right)^2 \right] + \frac{\mathrm{Lip}^{I+1}_{\nabla\varphi} - \sigma_{I+1}}{2} \right\} \| x^{\nu+1,I+1} - x^{\nu,I+1} \|_2^2.$$

To ensure that the sequence $\{\widehat{\mathcal{L}}_\beta(x^\nu, z^\nu; x^{\nu-1,I+1})\}$ is nonincreasing, we postulate the following.

(A3) For every $i = 1, \ldots, I+1$, $\sigma_i > \mathrm{Lip}^i_{\nabla\varphi}$. □

It then follows that for $\beta > \underline{\beta}$, where

$$\underline{\beta} \triangleq \frac{8\left[2\left(\left(\text{Lip}_{\nabla\varphi}^{I+1}\right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}}\right)^2\right) + \left(\text{Lip}_{\nabla H}^{I+1}\right)^2\right]}{\gamma_{\min}} \max_{1\le i\le I+1}\left\{\frac{1}{\sigma_i - \text{Lip}_{\nabla\varphi}^i}\right\}, \tag{11}$$

a constant $c_\beta > 0$ exists such that, for all ν,

$$\widehat{\mathcal{L}}_\beta(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}) - \widehat{\mathcal{L}}_\beta(x^\nu, z^\nu; x^{\nu-1,I+1}) \le -c_\beta \| x^{\nu+1} - x^\nu \|_2^2. \tag{12}$$

THEOREM 4.3. *Under assumptions* (A0)–(A3) *and the choice* (11) *of* $\underline{\beta}$, *every accumulation point (if it exists) of the sequence* $\{x^\nu\}$ *produced by the deterministic BCDM-ADMM with a* $\beta > \underline{\beta}$ *satisfies the stationarity condition* (2).

Proof. Let x^∞ be the limit of a convergent subsequence $\{x^{\nu+1}\}_{\nu\in\kappa}$. We first show that $\{z^{\nu+1}\}_{\nu\in\kappa}$ is bounded. In turn, from the bound on $\|z^{\nu+1}\|_2$ it suffices to show that $\{x^{\nu+1,I+1} - x^{\nu,I+1}\}_{\nu\in\kappa}$ is bounded. By the definition of $x^{\nu+1,I+1}$, we have

$$\begin{aligned}
&\widehat{\mathcal{L}}_\beta\left(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}\right) \\
&= \mathcal{L}_\beta\left(x^{\nu+1}, z^{\nu+1}\right) + \frac{4}{\beta\,\gamma_{\min}}\left[\left(\text{Lip}_{\nabla\varphi}^{I+1}\right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}}\right)^2\right]\|x^{\nu+1,I+1} - x^{\nu,I+1}\|_2^2 \\
&= \theta(x^{\nu+1}) + \left(z^{\nu+1}\right)^T\left[b - \sum_{i=1}^{I+1} A^i x^{\nu+1,i}\right] + \frac{\beta}{2}\left\| b - \sum_{i=1}^{I+1} A^i x^{\nu+1,i}\right\|_2^2 \\
&\quad + \frac{4}{\beta\gamma_{\min}}\left[\left(\text{Lip}_{\nabla\varphi}^{I+1}\right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}}\right)^2\right]\|x^{\nu+1,I+1} - x^{\nu,I+1}\|_2^2 \\
&= \theta(x^{\nu+1}) + \frac{\beta}{2}\left\| b - \sum_{i=1}^{I+1} A^i x^{\nu+1,i} + \frac{z^{\nu+1}}{\beta}\right\|_2^2 - \frac{\|z^{\nu+1}\|_2^2}{2\,\beta} \\
&\quad + \frac{4}{\beta\,\gamma_{\min}}\left[\left(\text{Lip}_{\nabla\varphi}^{I+1}\right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}}\right)^2\right]\|x^{\nu+1,I+1} - x^{\nu,I+1}\|_2^2 \\
&\ge \theta(x^{\nu+1}) + \frac{\beta}{2}\left\| b - \sum_{i=1}^{I+1} A^i x^{\nu+1,i} + \frac{z^{\nu+1}}{\beta}\right\|_2^2 \\
&\quad + \frac{4}{\beta\gamma_{\min}}\left[\left(\text{Lip}_{\nabla\varphi}^{I+1}\right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}}\right)^2\right]\|x^{\nu+1,I+1} - x^{\nu,I+1}\|_2^2 \\
&\quad - \frac{1}{\beta\,\gamma_{\min}}\left\{\left(\text{Lip}_{\nabla\varphi}^{I+1} + \text{Lip}_{\nabla\psi_{I+1}}\right)^2\left\| x^{\nu+1,I+1} - x^{\nu,I+1}\right\|_2^2\right. \\
&\quad \left. + \left[\left(\text{Lip}_{\nabla\varphi}^{I+1} + \text{Lip}_{\nabla H}^{I+1}\right)\| x^{\nu+1}\|_2 + \|\nabla_{x^{I+1}}\varphi(0)\|_2 + \|\nabla_{x^{I+1}} H(0)\|_2\right]^2\right\} \\
&\ge \theta(x^{\nu+1}) - \frac{1}{\beta\,\gamma_{\min}}\left[\left(\text{Lip}_{\nabla\varphi}^{I+1} + \text{Lip}_{\nabla H}^{I+1}\right)\| x^{\nu+1}\|_2 + \|\nabla_{x^{I+1}}\varphi(0)\|_2 + \|\nabla_{x^{I+1}} H(0)\|_2\right]^2.
\end{aligned}$$

Since the subsequence $\{x^{\nu+1}\}_{\nu\in\kappa}$ is assumed to have an accumulation point, thus is bounded, it follows that the subsequence $\{\widehat{\mathcal{L}}_\beta(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1})\}_{\nu\in\kappa}$ is bounded below. Since the entire sequence $\{\widehat{\mathcal{L}}_\beta(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1})\}$ is nonincreasing, it must be bounded below, and hence converges. Consequently, the sequence of consecutive differences $\{\widehat{\mathcal{L}}_\beta(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}) - \widehat{\mathcal{L}}_\beta(x^\nu, z^\nu; x^{\nu-1,I+1})\}$ converges to zero; thus the entire sequence of consecutive differences of the iterates $\{x^{\nu+1} - x^\nu\}$ also converges to zero, by (12). Thus the subsequence $\{z^{\nu+1}\}_{\nu\in\kappa}$ is bounded by (10). By the first inequality of Lemma 4.1, the entire sequence of consecutive differences of the multipliers $\{z^{\nu+1} - z^\nu\}$ converges to zero. Hence the limit x^∞ is feasible to (1). It remains to show that, for every tuple $\mathbf{j} = (j_i)_{i=1}^I$ with $j_i \in \mathcal{M}_i(x^\infty)$ for all $i = 1, \dots, I$,

$$\nabla\varphi(x^\infty)^T(\,x - x^\infty\,) + \sum_{i=1}^{I+1} H(\bullet, x^{\infty,-i})'(x^{\infty,i}; x^i - x^{\infty,i}) \geq \sum_{i=1}^{I} \nabla g_{ij_i}(x^{\infty,i})^T(\,x^i - x^{\infty,i}\,) \quad \forall\, x \in \widehat{\mathbf{X}}.$$

Let z^∞ be an accumulation point of the subsequence $\{z^{\nu+1}\}_{\nu\in\kappa}$. Without loss of generality, by working with an infinite subset of κ if necessary, we may assume that $\{z^{\nu+1}\}_{\nu\in\kappa}$, and thus $\{z^\nu\}_{\nu\in\kappa}$ also, converges to z^∞; moreover, there exists $\widehat{j}_i^{\,\infty} = \widehat{j}_i^{\,\nu}$ for all $\nu \in \kappa$. Let an arbitrary $x \in \widehat{\mathbf{X}}$ and a tuple $\mathbf{j}$ as specified above be given. It then follows that $j_i \in \mathcal{M}_{\varepsilon,i}(x^{\nu,i})$ for all $\nu \in \kappa$ sufficiently large and all $i = 1, \dots, I$. By the definition of the index $\widehat{j}_i^{\,\nu}$ that gives $x^{\nu+1,i} = \widehat{x}^{\,\nu+1,i,\widehat{j}_i^{\,\nu}}$ in particular, and by the minimizing property of $\widehat{x}^{\,\nu+1,i,j_i}$ we have

$$\begin{aligned}
&\nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^\nu_{\geq i}\right)^T(\,x^{\nu+1,i} - x^{\nu,i}\,) + H\left(x^{\nu+1}_{\leq i}, x^\nu_{>i}\right) - \underbrace{g_{i\,\widehat{j}_i^{\,\nu}}(x^{\nu+1,i})}_{= g_{i\,\widehat{j}_i^{\,\infty}}(x^{\nu+1,i})} \\
&\quad - (\,z^\nu\,)^T A^i x^{\nu+1,i} + D_{\psi_i}(x^{\nu+1,i}, x^{\nu,i}) + \frac{\beta}{2}\left\| b - \sum_{k\leq i} A^k x^{\nu+1,k} - \sum_{k>i} A^k x^{\nu,k}\right\|_2^2 \\
&\leq \nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^\nu_{\geq i}\right)^T(\,\widehat{x}^{\,\nu+1,i,j_i} - x^{\nu,i}\,) + H\left(x^{\nu+1}_{<i}, \widehat{x}^{\,\nu+1,i,j_i}, x^\nu_{>i}\right) - g_{ij_i}(\widehat{x}^{\,\nu+1,i,j_i}) \\
&\quad - (\,z^\nu\,)^T A^i \widehat{x}^{\,\nu+1,i,j_i} + D_{\psi_i}(\widehat{x}^{\,\nu+1,i,j_i}, x^{\nu,i}) \\
&\quad + \frac{\beta}{2}\left\| b - \sum_{k<i} A^k x^{\nu+1,k} - A^i\widehat{x}^{\,\nu+1,i,j_i} - \sum_{k>i} A^k x^{\nu,k}\right\|_2^2 \\
&\leq \nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^\nu_{\geq i}\right)^T(\,\widehat{x}^{\,\nu+1,i,j_i} - x^{\nu,i}\,) + H\left(x^{\nu+1}_{<i}, \widehat{x}^{\,\nu,i,j_i}, x^\nu_{>i}\right) \\
&\quad - g_{ij_i}(x^{\nu,i}) - \nabla g_{ij_i}(x^{\nu,i})^T(\,\widehat{x}^{\,\nu+1,i,j_i} - x^{\nu,i}\,) \quad \text{(by the convexity of } g_{ij}) \\
&\quad - (\,z^\nu\,)^T A^i \widehat{x}^{\,\nu+1,i,j_i} + D_{\psi_i}(\widehat{x}^{\,\nu+1,i,j_i}, x^{\nu,i}) \\
&\quad + \frac{\beta}{2}\left\| b - \sum_{k<i} A^k x^{\nu+1,k} - A^i\widehat{x}^{\,\nu+1,i,j_i} - \sum_{k>i} A^k x^{\nu,k}\right\|_2^2 \\
&\leq \nabla_{x^i}\varphi\left(x^{\nu+1}_{<i}, x^\nu_{\geq i}\right)^T(\,x^i - x^{\nu,i}\,) + H\left(x^{\nu+1}_{<i}, x^i, x^\nu_{>i}\right) \\
&\quad - g_{ij_i}(x^{\nu,i}) - \nabla g_{ij_i}(x^{\nu,i})^T(\,x^i - x^{\nu,i}\,) \\
&\quad - (\,z^\nu\,)^T A^i x^i + D_{\psi_i}(x^i, x^{\nu,i}) + \frac{\beta}{2}\left\| b - \sum_{k<i} A^k x^{\nu+1,k} - A^i x^i - \sum_{k>i} A^k x^{\nu,k}\right\|_2^2.
\end{aligned}$$

Taking the limit $\nu(\in\kappa) \to \infty$, and using the proved convergence of $\{x^{\nu+1} - x^\nu\} \to 0$

and the feasibility of x^∞ (thus the satisfaction of the coupling constraint), we deduce

$$\begin{aligned}
&H(x^\infty) - \max_{1\leq j\leq J_i} g_{ij}(x^{\infty,i}) - (z^\infty)^T A^i x^{\infty,i} \\
&\quad\leq H(x^\infty) - g_{i\widehat{j}_i^\infty}(x^{\infty,i}) - (z^\infty)^T A^i x^{\infty,i} \\
&\quad\leq \nabla_{x^i}\varphi(x^\infty)^T(x^i - x^{\infty,i}) + H\left(x^\infty_{<i}, x^i, x^\infty_{>i}\right) - \max_{1\leq j\leq J_i} g_{ij}(x^{\infty,i}) \\
&\qquad - \nabla g_{ij_i}(x^{\infty,i})^T(x^i - x^{\infty,i}) \\
&\qquad - (z^\infty)^T A^i x^i + D_{\psi_i}(x^i, x^{\infty,i}) + \frac{\beta}{2}\left\| A^i x^{\infty,i} - A^i x^i \right\|_2^2,
\end{aligned}$$

where the first inequality holds because $\max_{1\leq j\leq J_i} g_{ij}(x^{\infty,i}) \geq g_{i\widehat{j}_i^\nu}(x^{\infty,i})$ and the second inequality holds because $j_i \in \mathcal{M}_i(x^\infty)$. Hence, for all $i = 1, \dots, I$,

$$\begin{aligned}
H(x^\infty) - (z^\infty)^T A^i x^{\infty,i} \leq\ & \nabla_{x^i}\varphi(x^\infty)^T(x^i - x^{\infty,i}) + H\left(x^\infty_{<i}, x^i, x^\infty_{>i}\right) \\
& - \nabla g_{ij_i}(x^{\infty,i})^T(x^i - x^{\infty,i}) - (z^\infty)^T A^i x^i \\
& + D_{\psi_i}(x^i, x^{\infty,i}) + \frac{\beta}{2}\left\| A^i x^{\infty,i} - A^i x^i \right\|_2^2.
\end{aligned} \tag{13}$$

By the definition of $x^{\nu+1,I+1}$, we have

$$\begin{aligned}
&\nabla_{x^{I+1}}\varphi\left(x^{\nu+1}_{\leq I}, x^{\nu,I+1}\right)^T (x^{\nu+1,I+1} - x^{\nu,I+1}) + H\left(x^{\nu+1}_{\leq I}, x^{\nu+1,I+1}\right) \\
&\quad + D_{\psi_{I+1}}(x^{\nu+1,I+1}, x^{\nu,I+1}) - (z^\nu)^T A^{I+1} x^{\nu+1,I+1} \\
&\qquad + \frac{\beta}{2}\left\| b - \sum_{k\leq I} A^k x^{\nu+1,k} - A^{I+1} x^{\nu+1,I+1} \right\|_2^2 \\
&\leq \nabla_{x^{I+1}}\varphi\left(x^{\nu+1}_{\leq I}, x^{\nu,I+1}\right)^T (x^{I+1} - x^{\nu,I+1}) + H\left(x^{\nu+1}_{\leq I}, x^{I+1}\right) \\
&\quad + D_{\psi_{I+1}}(x^{I+1}, x^{\nu,I+1}) - (z^\nu)^T A^{I+1} x^{I+1} + \frac{\beta}{2}\left\| b - \sum_{k\leq I} A^k x^{\nu+1,k} - A^{I+1} x^{I+1} \right\|_2^2.
\end{aligned}$$

Passing to the limit $\nu(\in \kappa) \to \infty$, we deduce that

$$\begin{aligned}
H(x^\infty) - (z^\infty)^T A^{I+1} x^{\infty,I+1} \leq\ & \nabla_{x^{I+1}}\varphi(x^\infty)^T(x^{I+1} - x^{\infty,I+1}) + H(x^\infty_{\leq I}, x^{I+1}) - \\
& (z^\infty)^T A^{I+1} x^{I+1} + D_{\psi_{I+1}}(x^{I+1}, x^{\infty,I+1}) + \frac{\beta}{2}\left\| A^{I+1} x^{\infty,I+1} - A^{I+1} x^{I+1} \right\|_2^2.
\end{aligned}$$

Summing up the last inequality and the previous I inequalities (13), using the fact that $\sum_{i=1}^{I+1} A^i x^{\infty,i} = b = \sum_{i=1}^{I+1} A^i x^i$, and by Proposition 2.1, we obtain the desired stationarity of x^∞. ∎

The above theorem asserts neither the boundedness of the (primal) sequence $\{x^\nu\}$ nor its sequential convergence. The result below addresses these two issues under either one of two growth assumptions on the objective function that are easily satisfied if $\mathbf{X}$ is bounded. The first condition is a growth requirement on the combined objective θ of the order $\|x\|^{1+\delta}$ for some positive δ; the second condition is a weakening of the first, requiring only a coercivity property on θ, namely, $\theta(x) \to \infty$ for $x \in \mathbf{X}$ with $\|x\| \to \infty$, but adding the requirement that $H + \varphi$ have bounded partial gradients with respect to the distinguished variable x^{I+1}.

PROPOSITION 4.4. *Under the assumptions of Theorem* 4.3*, if either*

(a) *for some* $\delta > 0$, $\liminf_{\|x\|\to\infty} \left\{ \frac{\theta(x)}{\| x \|^{1+\delta}} \mid x \in \mathbf{X} \right\} > 0$*; or*

(b) $\nabla_{x^{I+1}}(H + \varphi)$ *is bounded and* θ *is coercive, both on* $\mathbf{X}$*,*

hold, then the sequence $\{(x^{\nu+1}, z^{\nu+1})\}$ *is bounded, and thus has an accumulation point. If in addition* $\{x^{\nu+1}\}$ *has an isolated accumulation point, then the sequence converges.*

Proof. In the proof of Theorem 4.3, we have shown that, for all ν, $\theta(x^{\nu+1})$ is bounded above by

$$\widehat{\mathcal{L}}_\beta\left(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}\right) + \frac{1}{\beta\,\gamma_{\min}} \Big[\left(\text{Lip}^{I+1}_{\nabla\varphi} + \text{Lip}^{I+1}_{\nabla H} \right) \| x^{\nu+1} \|_2 + \| \nabla_{x^{I+1}}\varphi(0) \|_2 + \| \nabla_{x^{I+1}} H(0) \|_2 \Big]^2,$$

which yields, since the sequence $\{ \widehat{\mathcal{L}}_\beta(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}) \}$ is nonincreasing,

$$\theta(x^{\nu+1}) \le \widehat{\mathcal{L}}_\beta\left(x^1, z^1; x^{0,I+1}\right) + \frac{1}{\beta\,\gamma_{\min}} \Big[\left(\text{Lip}^{I+1}_{\nabla\varphi} + \text{Lip}^{I+1}_{\nabla H} \right) \| x^{\nu+1} \|_2 + \| \nabla_{x^{I+1}}\varphi(0) \|_2 + \| \nabla_{x^{I+1}} H(0) \|_2 \Big]^2.$$

The liminf assumption in (a) therefore yields the boundedness of $\{x^{\nu+1}\}$.

Suppose condition (b) holds. From the proof of Lemma 4.1, we may deduce

$$\gamma_{\min} \| z^{\nu+1} \|_2^2 \le 4 \left\{ \left[\left(\text{Lip}^{I+1}_{\nabla\varphi} \right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}} \right)^2 \right] \| x^{\nu+1,I+1} - x^{\nu,I+1} \|_2^2 + \| \nabla_{x^{I+1}}(\varphi + H)(x^{\nu+1}) \|_2^2 \right\},$$

which shows in particular that the sequence of multipliers $\{z^{\nu+1}\}$ is bounded if the sequence of primal variables $\{x^{\nu+1}\}$ is. We have

$$\begin{aligned}
&\widehat{\mathcal{L}}_\beta\left(x^1, z^1; x^{0,I+1}\right) \ge \widehat{\mathcal{L}}_\beta\left(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}\right) \\
&= \mathcal{L}_\beta(x^{\nu+1}, z^{\nu+1}) + \frac{4}{\beta\,\gamma_{\min}} \left[\left(\text{Lip}^{I+1}_{\nabla\varphi} \right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}} \right)^2 \right] \| x^{\nu+1,I+1} - x^{\nu,I+1} \|_2^2 \\
&= \theta(x^{\nu+1}) + \frac{\beta}{2} \left\| b - \sum_{i=1}^{I+1} A^i x^{\nu+1,i} + \frac{z^{\nu+1}}{\beta} \right\|_2^2 - \frac{\| z^{\nu+1} \|_2^2}{2\,\beta} + \\
&\qquad \frac{4}{\beta\,\gamma_{\min}} \left[\left(\text{Lip}^{I+1}_{\nabla\varphi} \right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}} \right)^2 \right] \| x^{\nu+1,I+1} - x^{\nu,I+1} \|_2^2 \\
&\ge \theta(x^{\nu+1}) + \frac{\beta}{2} \left\| b - \sum_{i=1}^{I+1} A^i x^{\nu+1,i} - \frac{z^{\nu+1}}{\beta} \right\|_2^2 - \frac{2}{\beta\,\gamma_{\min}} \| \nabla_{x^{I+1}}(\varphi + H)(x^{\nu+1}) \|_2^2 + \\
&\qquad \frac{2}{\beta\,\gamma_{\min}} \left[\left(\text{Lip}^{I+1}_{\nabla\varphi} \right)^2 + \left(\text{Lip}_{\nabla\psi_{I+1}} \right)^2 \right] \| x^{\nu+1,I+1} - x^{\nu,I+1} \|_2^2.
\end{aligned}$$

Hence,

$$\theta(x^{\nu+1}) \le \widehat{\mathcal{L}}_\beta\left(x^1, z^1; x^{0,I+1}\right) + \frac{2}{\beta\,\gamma_{\min}} \sup \left\{ \| \nabla_{x^{I+1}}(\varphi + H)(x) \|_2^2 \mid x \in \mathbf{X} \right\},$$

establishing the boundedness of $\{x^{\nu+1}\}$ under the assumed coercivity condition of θ.

The last statement of the proposition follows from [17, Proposition 8.3.10] since $\{x^{\nu+1} - x^{\nu}\} \to 0$. ∎

Specializing Theorem 4.3 and Proposition 4.4 to the case without the coupling constraint, we obtain the following convergence result of the BCDM. In applying the proposition, it suffices to realize that the boundedness condition of $\nabla_{x^{I+1}}(H+\varphi)$ on $\mathbf{X}$ is needed to deal with the coupling constraint and its multiplier z; without them, there is no need for this condition. No further proof of the result is needed.

THEOREM 4.5. *Under assumptions* (A0), (A1), *and* (A3), *if θ is coercive on $\mathbf{X}$, then the sequence $\{x^{\nu+1}\}$ produced by the (deterministic) BCDM is bounded. Every one of its accumulation points is blockwise directional stationary solution of* (1) *without the coupling constraint. Moreover, if one such accumulation point is isolated, then the entire sequence converges to it.* ∎

5. Randomized choice of subproblems.

One way to reduce the number of convex subprograms solved at each iteration is by means of a randomized choice; this idea was first proposed in [40] for a nonsmooth DC program without regard to decomposition. The following randomized version of the combined BCDM-ADMM is the same as the deterministic version except that Step 1a is modified such that only one convex subprogram is solved (versus as many as $|\mathcal{M}_{\varepsilon,i}(x^{\nu,i})|$) for each $i = 1, \ldots, I$. The choice of the minimizing index $\widehat{j}_i^{\,\nu}$ employs an augmented Lagrangian based acceptance/rejection rule of the computed iterate.

To describe the randomized algorithm, it is useful to introduce some notation. For a given tuple $\mathbf{w} \triangleq ((x^i)_{i=1}^{I+1}, z)$ of primal blocks $x^i \in X^i$ and constraint multiplier z and for a tuple $\mathbf{s} \triangleq (s_i)_{i=1}^{I}$ of indices with $s_i \in \{1, \ldots, J_i\}$ for each $i = 1, \ldots, I$, define the tuple $\widehat{x}^{\mathbf{s}}(\mathbf{w}) \triangleq (\{\widehat{x}^{\,i,s_i}(\mathbf{w})\}_{i=1}^{I}, \widehat{x}^{\,I+1,\mathbf{s}}(\mathbf{w}))$ as the (unique) global minimizers of the respective semilinearized Gauss–Seidel subproblems

$$\left\{\begin{array}{l} \widehat{x}^{\,i,s_i}(\mathbf{w}) \in \displaystyle\operatorname*{argmin}_{\widehat{x}^{\,i}\in X^i} \nabla_{x^i}\varphi\left(\widehat{x}^{\,1,s_1}(\mathbf{w}),\ldots,\widehat{x}^{\,i-1,s_{i-1}}(\mathbf{w}),\widehat{x}^{\,i},x^{i+1},\ldots,x^{I+1}\right)^T \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \times\,(\,\widehat{x}^{\,i}-x^i\,) \\ \qquad\qquad + H\left(\widehat{x}^{\,1,s_1}(\mathbf{w}),\ldots,\widehat{x}^{\,i-1,s_{i-1}}(\mathbf{w}),\widehat{x}^{\,i},x^{i+1},\ldots,x^{I+1}\right) + D_{\psi_i}(\widehat{x}^{\,i},x^i) \\ \qquad\qquad - z^T A^i \widehat{x}^{\,i} + \dfrac{\beta}{2}\left\| b - \displaystyle\sum_{k<i} A^k \widehat{x}^{\,k,s_k}(\mathbf{w}) - A^i\widehat{x}^{\,i} - \sum_{k>i} A^k x^k \right\|_2^2 \\ \qquad\qquad - \nabla g_{is_i}(x^i)^T(\,\widehat{x}^{\,i}-x^i\,) \end{array}\right\}_{i=1}^{I}$$

plus the minimization over the last block

$$\begin{aligned} \widehat{x}^{\,I+1,\mathbf{s}}(\mathbf{w}) \in \operatorname*{argmin}_{\widehat{x}^{\,I+1}\in X^{I+1}} \Bigg\{ & \nabla_{x^{I+1}}\varphi\left(\widehat{x}^{\,1,s_1}(\mathbf{w}),\ldots,\widehat{x}^{\,I,s_I}(\mathbf{w}),\widehat{x}^{\,I+1}\right)^T(\,\widehat{x}^{\,I+1}-x^{I+1}\,) \\ & + H\left(\widehat{x}^{\,1,s_1}(\mathbf{w}),\ldots,\widehat{x}^{\,I,s_I}(\mathbf{w}),\widehat{x}^{\,I+1}\right) \\ & + D_{\psi_{I+1}}(\widehat{x}^{\,I+1},x^{I+1}) - z^T A^{I+1}\widehat{x}^{\,I+1} \\ & + \frac{\beta}{2}\left\| b - \sum_{k\le I} A^k\widehat{x}^{\,k,s_k}(\mathbf{w}) - A^{I+1}\widehat{x}^{\,I+1}\right\|_2^2 \Bigg\}. \end{aligned}$$

Let $\widehat{z}^{\mathbf{s}}(\mathbf{w}) \triangleq z+\beta\,(\,b-\sum_{i=1}^{I} A^i\widehat{x}^{\,i,s_i}(\mathbf{w})-A^{I+1}\widehat{x}^{\,I+1,\mathbf{s}}\,)$. We can establish the following result for the above defined tuple $\mathbf{x}^{\mathbf{s}}(\mathbf{w})$.

LEMMA 5.1. *For any given tuples* $\mathbf{w} \triangleq ((x^i)_{i=1}^{I+1}, z)$ *and* $\mathbf{s} \triangleq (s_i)_{i=1}^{I} \in \mathcal{M}(\mathbf{w})$ *it holds that*

$$\mathcal{L}_\beta\left(\left\{\widehat{x}^{\,i,s_i}(\mathbf{w})\right\}_{i\leq I}, x^{I+1}, z\right) - \sum_{i=1}^{I} \frac{\mathrm{Lip}^i_{\nabla\varphi} - \sigma_i}{2} \,\|\, \widehat{x}^{\,i,s_i}(\mathbf{w}) - x^i \,\|_2^2 \,\leq\, \mathcal{L}_\beta(\mathbf{w}).$$

Proof. We omit the proof since it is similar to the bound for $\mathrm{diff}^{\nu}_{\beta,i}$, $i = 1, \ldots, I$, employed in the proof of Lemma 4.2. ∎

Algorithm 3 The randomized combined BCDM-ADMM: fixed ε.

Let $x^0 \triangleq (x^{0,i})_{i=1}^{I+1} \in \prod_{i=1}^{I+1} X^i$ and z^0 be given. Also let scalars $\varepsilon > 0$, $\beta > \underline{\beta}$, and $p_{\min} \in (0,1)$ be given. Let $\nu = 0$. Compute the next iterate $x^{\nu+1} \triangleq (x^{\nu+1,i})_{i=1}^{I+1} \in \prod_{i=1}^{I+1} X^i$ by performing the following steps.

while a prescribed termination criterion is not satisfied **do**

Step 1a$_{\mathrm{ran}_1}$ Choose a random tuple $\mathbf{s}^{\nu+1} \triangleq (s_i^{\nu+1})_{i\leq I} \in \prod_{i=1}^{I} \mathcal{M}_{\varepsilon,i}(x^{\nu,i})$ and, for all $i = 1, \ldots, I$,

$$\mathrm{Prob}\left\{\text{ index } s_i^{\nu+1} \text{ is chosen } \mid \text{ given } (x^\nu, z^\nu)\right\} \triangleq p^{\nu+1,i,s_i^{\nu+1}} \,\geq\, p_{\min} \,>\, 0. \tag{14}$$

Let $\widehat{x}_{\leq I}^{\nu+1,\mathbf{s}^{\nu+1}} \triangleq \{\widehat{x}^{\,i,s_i^{\nu+1}}(\mathbf{w}^\nu)\}_{i\leq I}$, where $\mathbf{w}^\nu \triangleq (x^\nu, z^\nu)$. Let

$$x_{\leq I}^{\nu+1} \triangleq \begin{cases} x^\nu_{\leq I} & \text{if } \mathcal{L}_\beta\left(\widehat{x}_{\leq I}^{\nu+1,\mathbf{s}^{\nu+1}}, x^{\nu,I+1}; z^\nu\right) \\ & \quad + \displaystyle\sum_{i=1}^{I} \frac{\sigma_i - \mathrm{Lip}^i_{\nabla\varphi}}{2}\,\|x^{\nu,i} - \widehat{x}^{\,i,s_i^{\nu+1}}(\mathbf{w}^\nu)\|_2^2 > \mathcal{L}_\beta\left(x^\nu, z^\nu\right), \\ \widehat{x}_{\leq I}^{\nu+1,\mathbf{s}^{\nu+1}} & \text{otherwise.} \end{cases}$$

Steps 1b and 2. Same as Algorithm 1.

end while

return (x, z)

Several remarks are worthy of note. One, $\widehat{x}^{\,\nu+1,\mathbf{s}^{\nu+1}}$ is a random vector as its components depend on the randomly chosen tuple $\mathbf{s}^{\nu+1}$. Two, even though Steps 1b and 2 are the same as before, the last block $x^{\nu+1,I+1}$ and the multiplier $z^{\nu+1}$ depend on $x^{\nu+1}_{\leq I}$, thus these two variables are not deterministic either. Three, through the test that defines the new iterate $x^{\nu+1}_{\leq I}$, we accept or reject the I blocks $\{\widehat{x}^{\,i,s_i^{\nu+1}}(\mathbf{w}^\nu)\}_{i\leq I}$ all at once after I convex subprograms are solved. Different frequencies and variations of the latter test can be introduced but are omitted. Lastly, $x^{\nu+1,I+1}$ is equal to either $\widehat{x}^{\,I+1;\mathbf{s}}(\mathbf{w}^\nu)$ if $\widehat{x}_{\leq I}^{\nu+1,\mathbf{s}^{\nu+1}}$ is accepted as $x^{\nu+1}_{\leq I}$ or the minimizer of $\theta^{\nu+1}_{I+1}(\cdot; x^\nu_{\leq I}, z^\nu)$ on X^{I+1} if the test rejects $\widehat{x}_{\leq I}^{\nu+1,\mathbf{s}^{\nu+1}}$.

EXAMPLE 2 (continued). We applied the randomized choice of the subproblems to the example with fixed $\varepsilon = 0.1$, $c = 1.1$, and the initial iterate $(x_1^0, x_2^0, z^0) = (1, 1, -1)$ for which the deterministic version of the algorithm converges (see Figure 1) while the existing ADMM does not. In iterations 10 through 22, two subproblems were candidates that triggered the randomization. After 60 iterations, the optimal triple of $(-1/4, -1/4, -1/8)$ was obtained. In summary, for this example, we may draw the following two conclusions with the initial iterate $(x_1^0, x_2^0, z^0) = (1, 1, -1)$.

(a) Both the deterministic and randomized versions of the combined BCDM-ADMM with a fixed $\varepsilon = 0.1$ converge to the unique d-stationary point $(-1/4, -1/4)$.

(b) The standard ADMM with both linearization of φ and no such linearization converge to (0,0); so does the deterministic BCDM-ADMM with $\varepsilon = 0$. ◻

5.1. Convergence proof of randomization. We state and prove the almost sure subsequential convergence result for the randomized combined BCDM-ADMM. We omit the BCDM as it is a simplification of the combined BCDM-ADMM. The modified augmented Lagrangian function (9) continues to play an important role in the proof.

THEOREM 5.2. *Under assumptions* (A0)–(A3) *and the choice* (11) *of* $\underline{\beta}$, *suppose that for each* ν *the tuple* $\{s_1^{\nu+1}, \ldots, s_I^{\nu+1}\}$ *consists of independent random indices satisfying the condition* (14). *It holds that every accumulation point of the sequence* $\{x^\nu\}$ *produced by the randomized BCDM-ADMM with a fixed* $\varepsilon > 0$ *and with a* $\beta > \underline{\beta}$ *satisfies the stationarity condition* (2) *with probability one.*

Proof. The proof below combines that of the deterministic case and that of Proposition 7 in [40] for the case with no decomposition. Nevertheless, the details are more complicated than either one of the previous proofs. Throughout the proof, we write $\mathbf{w}^\nu \triangleq (x^\nu, z^\nu)$ and let $\mathcal{F}^\nu$ be the filtration generated by the iterates $\{\mathbf{w}^0, \ldots, \mathbf{w}^\nu\}$ produced by the randomization up to iteration ν. Let $\boldsymbol{\mathcal{M}}_\varepsilon(\mathbf{w}^\nu)$ consist of the tuples $\mathbf{s} \triangleq (s_i)_{i=1}^I$ with $s_i \in \mathcal{M}_{\varepsilon,i}(x^\nu)$ for all $i = 1, \ldots, I$ such that

$$\mathcal{L}_\beta\left(\left\{\widehat{x}^{\,i,s_i}(\mathbf{w}^\nu)\right\}_{i\le I}, x^{\nu,I+1}; z^\nu\right) + \sum_{i=1}^I \frac{\sigma_i - \operatorname{Lip}_{\nabla\varphi}^i}{2} \|\, x^{\nu,i} - \widehat{x}^{\,i,s_i}(\mathbf{w}^\nu)\,\|_2^2 \le \mathcal{L}_\beta\left(x^\nu, z^\nu\right).$$

Thus, if a tuple $\mathbf{s}^{\nu+1}$ belongs to $\boldsymbol{\mathcal{M}}_\varepsilon(\mathbf{w}^\nu)$, then the next iterate $x^{\nu+1} = \widehat{x}^{\,\mathbf{s}^{\nu+1}}(\mathbf{w}^\nu)$. Moreover, in this case we may follow the analysis of the deterministic case and deduce (see the derivation of the expression (12)),

$$\begin{aligned}
&\widehat{\mathcal{L}}_\beta\left(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}\right) - \widehat{\mathcal{L}}_\beta\left(x^\nu, z^\nu; x^{\nu-1,I+1}\right) \\
&= \widehat{\mathcal{L}}_\beta\left(\widehat{x}^{\,\mathbf{s}^{\nu+1}}(\mathbf{w}^\nu), z^\nu; x^{\nu,I+1}\right) - \widehat{\mathcal{L}}_\beta\left(x^\nu, z^\nu; x^{\nu-1,I+1}\right) \\
&\le \sum_{i=1}^I \left\{ \frac{4}{\beta\,\gamma_{\min}} \left[\max\left(\left(\operatorname{Lip}_{\nabla\varphi}^{I+1}\right)^2, \left(\operatorname{Lip}_{\nabla\psi_{I+1}}\right)^2 \right) + \left(\operatorname{Lip}_{\nabla H}^{I+1}\right)^2 \right] \right. \\
&\qquad\qquad \left. + \frac{\operatorname{Lip}_{\nabla\varphi}^i - \sigma_i}{2} \right\} \|\, \widehat{x}^{\,i,s_i^{\nu+1}}(\mathbf{w}^\nu) - x^{\nu,i}\,\|_2^2 \\
&\quad + \left\{ \frac{4}{\beta\,\gamma_{\min}} \left[2\left(\left(\operatorname{Lip}_{\nabla\varphi}^{I+1}\right)^2 + \left(\operatorname{Lip}_{\nabla\psi_{I+1}}\right)^2 \right) + \left(\operatorname{Lip}_{\nabla H}^{I+1}\right)^2 \right] \right. \\
&\qquad\qquad \left. + \frac{\operatorname{Lip}_{\nabla\varphi}^{I+1} - \sigma_{I+1}}{2} \right\} \|\, \widehat{x}^{\,I+1,\mathbf{s}^{\nu+1}}(\mathbf{w}^\nu) - x^{\nu,I+1}\,\|_2^2.
\end{aligned}$$

If the tuple $\mathbf{s}^{\nu+1} \notin \boldsymbol{\mathcal{M}}_\varepsilon(\mathbf{w}^\nu)$, then $x^{\nu+1,i} = x^{\nu,i}$ for all $i = 1, \ldots, I$. Thus,

$$\widehat{\mathcal{L}}_\beta\left(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}\right) - \widehat{\mathcal{L}}_\beta\left(x^\nu, z^\nu; x^{\nu-1,I+1}\right) = \widehat{\mathcal{L}}_\beta\left(x^\nu_{\leq I}, x^{\nu+1,I+1}, z^{\nu+1}; x^{\nu,I+1}\right) - \widehat{\mathcal{L}}_\beta\left(x^\nu, z^\nu; x^{\nu-1,I+1}\right).$$

We claim that the above difference is nonpositive, or, equivalently,

$$\begin{aligned}(15)\quad &\mathcal{L}_\beta\left(x^\nu_{\leq I}, x^{\nu+1,I+1}, z^{\nu+1}\right) \\ &\quad + \frac{4}{\beta\,\gamma_{\min}}\left[\left(\mathrm{Lip}^{I+1}_{\nabla\varphi}\right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}}\right)^2\right] \|\, x^{\nu+1,I+1} - x^{\nu,I+1}\,\|_2^2 \\ &\leq \mathcal{L}_\beta(x^\nu, z^\nu) + \frac{4}{\beta\,\gamma_{\min}}\left[\left(\mathrm{Lip}^{I+1}_{\nabla\varphi}\right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}}\right)^2\right] \|\, x^{\nu,I+1} - x^{\nu-1,I+1}\,\|_2^2.\end{aligned}$$

Similar to the proof of the bound for $\mathrm{diff}^{\nu+1}_{\beta,I+1}$ in Lemma 4.2, we may deduce

$$\mathcal{L}_\beta\left(x^\nu_{\leq I}, x^{\nu+1,I+1}, z^\nu\right) \leq \mathcal{L}_\beta\left(x^\nu, z^\nu\right) + \frac{\mathrm{Lip}^{I+1}_{\nabla\varphi} - \sigma_{I+1}}{2} \|\, x^{\nu+1,I+1} - x^{\nu,I+1}\,\|_2^2.$$

Moreover,

$$\begin{aligned}&\mathcal{L}_\beta\left(x^\nu_{\leq I}, x^{\nu+1,I+1}, z^{\nu+1}\right) - \mathcal{L}_\beta\left(x^\nu_{\leq I}, x^{\nu+1,I+1}, z^\nu\right) \\ &= (\, z^{\nu+1} - z^\nu\,)^T\left[b - \sum_{i=1}^I A^i x^{\nu,i} - A^{I+1}x^{\nu+1,I+1}\right] \\ &\leq \frac{4}{\beta\,\gamma_{\min}}\left\{\left[\max\left(\left(\mathrm{Lip}^{I+1}_{\nabla\varphi}\right)^2, \left(\mathrm{Lip}_{\nabla\psi_{I+1}}\right)^2\right) + \left(\mathrm{Lip}^{I+1}_{\nabla H}\right)^2\right]\sum_{i=1}^{I+1} \|\, x^{\nu+1,i} - x^{\nu,i}\,\|_2^2\right. \\ &\qquad \left. + \left[\left(\mathrm{Lip}^{I+1}_{\nabla\varphi}\right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}}\right)^2\right] \|\, x^{\nu,I+1} - x^{\nu-1,I+1}\,\|_2^2\right\} \\ &\qquad\qquad \text{(by a proof similar to that of Lemma 4.1)} \\ &= \frac{4}{\beta\,\gamma_{\min}}\left\{\left[\max\left(\left(\mathrm{Lip}^{I+1}_{\nabla\varphi}\right)^2, \left(\mathrm{Lip}_{\nabla\psi_{I+1}}\right)^2\right) + \left(\mathrm{Lip}^{I+1}_{\nabla H}\right)^2\right] \|\, x^{\nu+1,I+1} - x^{\nu,I+1}\,\|_2^2\right. \\ &\qquad \left. + \left[\left(\mathrm{Lip}^{I+1}_{\nabla\varphi}\right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}}\right)^2\right] \|\, x^{\nu,I+1} - x^{\nu-1,I+1}\,\|_2^2\right\}.\end{aligned}$$

Consequently,

$$\begin{aligned}&\mathcal{L}_\beta\left(x^\nu_{\leq I}, x^{\nu+1,I+1}, z^{\nu+1}\right) - \mathcal{L}_\beta(x^\nu, z^\nu) \\ &\leq \left\{\frac{4}{\beta\,\gamma_{\min}}\left[2\left(\left(\mathrm{Lip}^{I+1}_{\nabla\varphi}\right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}}\right)^2\right) + \left(\mathrm{Lip}^{I+1}_{\nabla H}\right)^2\right]\right. \\ &\qquad \left. + \frac{\mathrm{Lip}^{I+1}_{\nabla\varphi} - \sigma_{I+1}}{2}\right\} \|\, x^{\nu+1,I+1} - x^{\nu,I+1}\,\|_2^2 \\ &\quad + \frac{4}{\beta\,\gamma_{\min}}\left[\left(\mathrm{Lip}^{I+1}_{\nabla\varphi}\right)^2 + \left(\mathrm{Lip}_{\nabla\psi_{I+1}}\right)^2\right] \\ &\qquad \times\left[-\|\, x^{\nu+1,I+1} - x^{\nu,I+1}\,\|_2^2 + \|\, x^{\nu,I+1} - x^{\nu-1,I+1}\,\|_2^2\right],\end{aligned}$$

which yields the desired inequality (15) by the choice of β.

We note that at iteration $\nu+1$, every tuple $\mathbf{s}$ has probability equal to $\bar{p}^{\,\nu+1,\mathbf{s}} \triangleq \prod_{i=1}^{I} p^{\nu+1,i,s_i} \geq (p_{\min})^I$ of being picked. With this observation, taking conditional expectation $\mathbb{E}$ with respect to the filtration $\mathcal{F}^{\,\nu}$, we deduce, for some constant $c_\beta > 0$,

$$
\begin{aligned}
&\mathbb{E}\left[\widehat{\mathcal{L}}_\beta\left(x^{\nu+1}, z^{\nu+1}; x^{\nu,I+1}\right) \mid \mathcal{F}^{\,\nu}\right] \\
&\quad = \sum_{\mathbf{s}^{\nu+1} \in \boldsymbol{\mathcal{M}}_\varepsilon(\mathbf{w}^\nu)} \bar{p}^{\,\nu+1,\mathbf{s}^{\nu+1}}\, \widehat{\mathcal{L}}_\beta\left(\widehat{x}^{\,\mathbf{s}^{\nu+1}}(\mathbf{w}^\nu), z^{\nu+1}; x^{\nu,I+1}\right) \\
&\qquad + \sum_{\mathbf{s}^{\nu+1} \notin \boldsymbol{\mathcal{M}}_\varepsilon(\mathbf{w}^\nu)} \bar{p}^{\,\nu+1,\mathbf{s}^{\nu+1}}\, \widehat{\mathcal{L}}_\beta\left(x^\nu_{\leq I}, x^{\nu+1,I+1}, z^{\nu+1}; x^{\nu,I+1}\right) \\
&\quad \leq \widehat{\mathcal{L}}_\beta\left(x^\nu, z^\nu; x^{\nu,I}\right) - c_\beta\,(p_{\min})^I \sum_{\mathbf{s}^{\nu+1} \in \boldsymbol{\mathcal{M}}_\varepsilon(\mathbf{w}^\nu)} \left[\sum_{i=1}^{I} \left\| \widehat{x}^{\,i,s_i^{\nu+1}}(\mathbf{w}^\nu) - x^{\nu,i} \right\|_2^2 \right. \\
&\qquad\qquad \left. + \left\| \widehat{x}^{\,I+1,\mathbf{s}^{\nu+1}}(\mathbf{w}^\nu) - x^{\nu,I+1} \right\|_2^2 \right].
\end{aligned}
$$

Consequently, if the sequence $\{(x^\nu, z^\nu)\}$ possesses a convergent subsequence $\{\mathbf{w}^\nu \triangleq (x^\nu, z^\nu)\}_{\nu\in\kappa}$ with limit $\mathbf{w}^\infty \triangleq (x^\infty, z^\infty)$, then the random sequence $\{\widehat{\mathcal{L}}_\beta(x^\nu, z^\nu; x^{\nu,I+1})\}$ is bounded from below. By the supermartingale convergence theorem [6, Proposition 4.2], we may conclude that the latter sequence converges almost surely. Without loss of generality, we may assume, by working with a further subsequence of $\{\mathbf{w}^\nu\}_{\nu\in\kappa}$, that $\boldsymbol{\mathcal{M}}_\varepsilon(\mathbf{w}^\nu)$ is equal to the same $\widehat{\boldsymbol{\mathcal{M}}}_\varepsilon$ for all $\nu \in \kappa$. It follows that, for all $i = 1, \ldots, I$,

$$
\lim_{\nu(\in\kappa)\to\infty} \left\| \widehat{x}^{\,i,s_i}(\mathbf{w}^\nu) - x^{\nu,i} \right\| = \lim_{\nu(\in\kappa)\to\infty} \left\| \widehat{x}^{\,I+1,\mathbf{s}}(\mathbf{w}^\nu) - x^{\nu,I+1} \right\| = 0 \quad \forall\, \mathbf{s} \in \widehat{\boldsymbol{\mathcal{M}}}_\varepsilon.
$$

In what follows, we show that $\prod_{i=1}^{I} \mathcal{M}_i(x^{\infty,i}) \subseteq \widehat{\boldsymbol{\mathcal{M}}}_\varepsilon$. Let $\mathbf{s}$ be an arbitrary tuple in $\prod_{i=1}^{I} \mathcal{M}_i(x^{\infty,i})$. By Lemma 5.1, we have

$$
\begin{aligned}
&\mathcal{L}_\beta\left(\left\{\widehat{x}^{\,i,s_i}(\mathbf{w}^\infty)\right\}_{i\leq I}, x^{\infty,I+1}, z^\infty\right) - \sum_{i=1}^{I} \frac{\mathrm{Lip}^i_{\nabla\varphi} - \sigma_i}{2} \left\| \widehat{x}^{\,i,s_i}(\mathbf{w}^\infty) - x^{\infty,i} \right\|_2^2 \\
&\qquad\qquad \leq \mathcal{L}_\beta(\mathbf{w}^\infty).
\end{aligned}
$$

We claim that equality must hold. Assume not. It follows that, for all $\nu \in \kappa$ sufficiently large,

$$
\mathcal{L}_\beta\left(\left\{\widehat{x}^{\,i,s_i}(\mathbf{w}^\nu)\right\}_{i\leq I}, x^{\nu,I+1}, z^\nu\right) - \sum_{i=1}^{I} \frac{\mathrm{Lip}^i_{\nabla\varphi} - \sigma_i}{2} \left\| \widehat{x}^{\,i,s_i}(\mathbf{w}^\nu) - x^{\nu,i} \right\|_2^2 \leq \mathcal{L}_\beta(\mathbf{w}^\nu).
$$

This proves that $\mathbf{s} \in \boldsymbol{\mathcal{M}}_\varepsilon(\mathbf{w}^\nu) = \widehat{\boldsymbol{\mathcal{M}}}_\varepsilon$. Hence the inclusion $\prod_{i=1}^{I} \mathcal{M}_i(x^{\infty,i}) \subseteq \widehat{\boldsymbol{\mathcal{M}}}_\varepsilon$. Therefore, with probability one, it holds that $\lim_{\nu(\in\kappa)\to\infty} \|x^{\nu+1} - x^\nu\|_2 = 0$ and the subsequence $\{x^{\nu+1}\}_{\nu\in\kappa}$ also converges to x^∞. At this point, the remaining part of the proof of Theorem 4.3 can be applied. ∎

6. Conclusions. This paper has presented and analyzed the convergence of a combined BCDM-ADMM method for computing a directional stationary solution of a class of multiblock, nonsmooth, nonconvex optimization problems with private and coupled constraints. A randomized version of the algorithm is also introduced to lessen the computational effort per iteration of the basic algorithm if needed. Motivated by interesting applications, a future research direction is to develop decomposition methods for an extended class of problems where the functions g_{ij} in (1) are not separable in their respective blocks of variables and a coupled constraint is still present. The bottleneck to this extension is assumption (A2), which is key to the bound of the multiplier of the coupled constraint in a single-loop algorithm such as those presented in this paper. A resolution to this technical challenge may call for a two-loop algorithm wherein an inner loop would solve a Lagrangian relaxation of the problem with the coupled constraint lifted to the objective and an outer loop would update the multiplier accordingly. More research is needed to investigate this extension. Another topic omitted in this paper is the investigation of the KL inequality theory [4] to study the sequential convergence of the iterates produced by the combined BCDM-ADMM for computing directional stationary solutions of problem (1).

Appendix. Two applied problems. We briefly discuss two classes of applied problems that provide the source for the framework (1). The first class of problems is topical, involving problems arising from (group) sparsity representation [26]; the other class involves a noncooperative generalized game with a multiconvex potential function and coupling constraints [18, 19].

Group sparsity representation. Let $\mathcal{G}_i \subseteq \{1, \ldots, m\}$ for $i = 1, \ldots, I$ be I nonoverlapping groups of parameters $\beta_{\mathcal{G}_i} \triangleq \{\, \beta_j \,\}_{j \in \mathcal{G}_i}$ partitioning the set of unknown coefficients $\{\beta_1, \ldots, \beta_m\}$ in the linear regression model: $y \approx \beta^T x$ with the vector input $x \in \mathbb{R}^m$ and scalar output $y \in \mathbb{R}$. Given are data $\{(x^i, y_i)\}_{i=1}^N$, a loss function $\ell(\beta)$ (e.g., the mean least-squares loss $\ell(\beta) = \frac{1}{N}\sum_{i=1}^N (\, y_i - \beta^T x^i \,)^2)$, constant $\gamma_1 > 0$ and sparsity function $P^{\,\text{ele}}(\beta)$ for the parameters, and $\gamma_2 > 0$ and sparsity functions $P_i^{\text{grp}}(\beta_{\mathcal{G}_i})$ for each group. The regression problem is

$$\underset{\beta \in \mathbb{R}^m}{\text{minimize}}\ \ell(\beta) + \gamma_1 \, P^{\,\text{ele}}(\beta) + \gamma_2 \sum_{i=1}^{I} P_i^{\text{grp}}(\beta_{\mathcal{G}_i}). \tag{16}$$

The sparsity functions $P^{\,\text{ele}}$ and each P_i^{grp} can be either *exact* or *surrogate*; all are of the DC type (see [1]). A simple example illustrating how (16) is of the form (1) is when the truncated $\|\bullet\|_1$ function is used, resulting in $P^{\,\text{ele}}(\beta) \triangleq \sum_{j=1}^m \min(\frac{1}{\tau_j}\,|\beta_j|, 1)$ for some positive scalars $\tau_j > 0$. Since

$$\min\left(\frac{1}{\tau_j}\,|\beta_j|, 1\right) = \frac{1}{\tau_j}\,|\beta_j| - \max\left(0, \frac{1}{\tau_j}\,|\beta_j| - 1\right),$$

the negative max term leads to the pointwise maximum terms in (1). Other univariate sparsity functions such as SCAD and MCP are (differentiable) DC functions; they also lead to an objective of the form (1). As an example of a coupled sparsity function, consider an exact K-sparsity function for some positive integer $K \geq 2$ that is used for a group $\mathcal{G}$ with s elements. Such a function has the property that its zeros are vectors with no more than K nonzero elements and is given by the following: for an s-dimensional vector w with $s \geq K$,

$$P_K(w) \triangleq \sum_{i=1}^{s} |\, w_i \,| - \sum_{k=1}^{K} |\, w_{[k]} \,| \, = \sum_{k=K+1}^{s} |\, w_{[k]} \,|,$$

where $\max_{1 \leq i \leq s} |w_i| \triangleq |w_{[1]}| \geq |w_{[2]}| \geq \cdots \geq |w_{[s]}| \triangleq \min_{1 \leq i \leq s} |w_i|$ is a nonincreasing ordering of $\{\, | w_i \,| \,\}_{i=1}^{s}$. The expression

$$\sum_{k=1}^{K} |\, w_{[k]} \,| \, = \, \underset{v \in \mathcal{E}(\Delta_{K,s})}{\text{maximum}} \, \sum_{i=1}^{s} v_i \,|\, w_i \,|,$$

where $\mathcal{E}(\Delta_{K,s})$ is the finite set of extreme points of the polytope $\Delta_{K,s} \triangleq \{\, v \in [\,0,1\,]^s \mid \sum_{i=1}^{s} v_i = K \,\}$, shows that such a sum leads to the last pointwise max sum in the objective function of (1).

A multiagent noncooperative game. Consider a noncooperative game with $I+1$ selfish players each with private strategy set $X^i \subseteq \mathbf{R}^{n_i}$ for $i = 1, \ldots, I+1$. Anticipating rivals' strategies $x^{-i} \in X^{-i}$, each player $i = 1, \ldots, I$ solves a (nonsmooth) problem

$$(17) \qquad \begin{array}{ll} \underset{x^i \in X^i}{\text{minimize}} \; \theta_i(x^i, x^{-i}) \triangleq \varphi(x) + \zeta_i(x^i, x^{-i}) - \displaystyle\max_{1 \leq j \leq J_i} g_{ij}(x^i) \\ \text{subject to} \;\; \displaystyle\sum_{k=1}^{I+1} A^k x^k \, = \, b, \end{array}$$

while player $I+1$, anticipating $(x^i)_{i=1}^{I}$, solves a problem without the max term in the objective:

$$(18) \qquad \begin{array}{ll} \underset{x^{I+1} \in X^{I+1}}{\text{minimize}} \; \theta_{I+1}\left(x^{I+1}, (x^i)_{i=1}^{I}\right) \triangleq \varphi(x) + \zeta_{I+1}\left(x^{I+1}, (x^i)_{i=1}^{I}\right) \\ \text{subject to} \;\; \displaystyle\sum_{k=1}^{I+1} A^k x^k \, = \, b. \end{array}$$

Assuming that the family of functions $\{\zeta_i(x)\}_{i=1}^{I+1}$ admits an inexact potential function $H(x)$ with the property that, for all $x \in \mathbf{X}$ and $y^i \in X^i$,

$$(19) \qquad \zeta_i(x^i, x^{-i}) - \zeta_i(y^i, x^{-i}) \, \geq \, H(x^i, x^{-i}) - H(y^i, x^{-i}) \quad \forall \; i \, = \, 1, \ldots, I+1,$$

one can show that the potential function H must be multiconvex, but not necessarily convex nor differentiable jointly in its arguments. The aggregate optimization problem

$$(20) \qquad \begin{array}{ll} \underset{x \in \mathbf{X}}{\text{minimize}} \; \theta(x) \triangleq \left[\varphi(x) + H(x) - \displaystyle\sum_{i=1}^{I} \max_{1 \leq j \leq J_i} g_{ij}(x^i) \right] \\ \text{subject to} \;\; \displaystyle\sum_{i=1}^{I+1} A^i x^i \, = \, b \end{array}$$

bears a close connection to the game in terms of their respective directional derivative based stationary solutions. Specifically, assuming that the common function $\varphi(x)$ is differentiable, if a tuple $(\bar{x}^{\,i})_{i=1}^{I+1}$ is a blockwise directional stationary solution of (20) satisfying the condition (2), then each $\bar{x}^{\,i}$ is a directional stationary solution of the optimization problem (17) for $i \leq I$ and (18) for $i = I+1$. Consequently, via the potential optimization formulation, the combined ADMM-BCD algorithm leads to a provably convergent one-loop best-response iterative algorithm for computing a tuple of strategies satisfying the first-order directional stationarity conditions for each player of the game. This is a significant advance towards solving a generalized Nash game in the presence of nonconvexity and nondifferentiability.

Acknowledgments. This work was completed during the second author's visit to the first author. The authors are grateful to Professor Defeng Sun of the National University of Singapore and two referees for providing some constructive comments on this paper. In particular, one referee's comment about the previous assumption (A2) has motivated us to obtain the present improved assumption by not requiring X^{I+1} to be polyhedral in particular.

REFERENCES

[1] M. Ahn, J.S. Pang, and J. Xin, *Difference-of-convex learning* I*: Directional stationarity, optimality, and sparsity*, SIAM J. Optim., 27 (2017), pp. 1637–1665.

[2] A. Alvarado, G. Scutari, and J.S. Pang, *A new decomposition method for multiuser DC-programming and its applications*, IEEE Trans. Signal Process., 62 (2014), pp. 2984–2998.

[3] B. Ames and M.Y. Hong, *Alternating direction method of multipliers for penalized zero-variance discriminant analysis*, Comput. Optim. Appl., 64 (2016), pp. 725–754.

[4] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Lojasiewicz inequality*, Math. Oper. Res, 35 (2010), pp. 438–457.

[5] D. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[6] D.P. Bertsekas and J.N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA, 1996.

[7] 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 (2010), http://dx.doi.org/10.1561/2200000016.

[8] L.M. Bregman, *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. Math. Phys., 7 (1967), pp. 200–217.

[9] R.H. Chan, M. Tao, and X.M. Yuan, *Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers*, SIAM J. Imag. Sci., 6 (2013), pp. 680–697.

[10] R.H. Chan, M. Tao and X.M. Yuan, *Linearized alternating direction method of multipliers for constrained linear least-squares problem*, East Asian J. Appl. Math., 2 (2012), pp. 326–341.

[11] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, PA, 1987.

[12] R.W. Cottle, J.S. Pang, and R.S. Stone, *The Linear Complementarity Problem*, Classics Appl. Math. 60, SIAM, Philadelphia, PA, 2009.

[13] Y. Cui, X.D. Li, D.F. Sun, and K.C. Toh, *On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions*, J. Optim. Theory Appl., 169 (2016), pp. 1013–1041.

[14] J. Eckstein, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1989; available at http://hdl.handle.net/1721.1/14356.

[15] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res, 18 (1993), pp. 202–226.

[16] 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.

[17] F. Facchinei and J.S. Pang *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[18] F. Facchinei and J.S. Pang, *Nash equilibria: The variational approach*, in Convex Optimization in Signal Processing and Communications, Y. Eldar and D. Palomar, eds., Cambridge University Press, Cambridge, 2009, pp. 443–493.

[19] F. Facchinei, V. Piccialli, and M. Sciandrone, *Decomposition algorithms for generalized potential games*, Comput. Optim. Appl., 50 (2011), pp. 237–262.

[20] M. Fazel, P.T. Kei, D. Sun, and P. Tseng, *Hankel matrix rank minimization with applications to system identification and realization*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 946–977.

[21] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, M. Fortin and R. Glowinski, eds., North-Holland, Amsterdam, 1983, pp. 299–340.

[22] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problem via finite-element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.

[23] R. Glowinski, S. Osher, and W.T. Yin, *Splitting Methods in Communication and Imaging, Science, and Engineering*, Springer International Publishing, Switzerland, 2016.

[24] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, Stud. Appl. Numer. Math., SIAM, Philadelphia, PA, 1989.

[25] K. Guo, D.R. Han and T.T. Wu, *Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints*, Int. J. Comput. Math., 94 (2016), pp. 1653–1669, www.tandfonline.com/doi/abs/10.1080/00207160.2016.1227432?journalCode=gcom20.

[26] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity: The Lasso and Generalizations*, Monogr. Statist. Appl. Probab. 143, Chapman and Hall/CRC Press, Boca Raton, FL, 2015.

[27] B.S. He, M. Tao, and X.M. Yuan, *A splitting method for separable convex programming*, IMA J. Numer. Anal., 35 (2015), pp. 394–426.

[28] B.S. He, M. Tao, and X.M. Yuan, *Alternating direction method with Gaussian back substitution for separable convex programming*, SIAM J. Optim., 22 (2012), pp. 313–340.

[29] M. Hong, Z.Q. Luo, M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM J. Optim., 26 (2016), pp. 337–364.

[30] M. Hong, M. Razaviyayn, Z.Q. Luo, and J.S. Pang, *A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing*, IEEE Signal Process. Mag., 33 (2015), pp. 57–77, doi:10.1109/MSP.2015.2481563.

[31] H.A. Le Thi and D.T. Pham, *Recent advances in DC programming and DCA*, in Transactions on Computational Collective Intelligence XIII, N.T. Nguyen and H.A. Le-Thi, eds., Lecture Notes in Comput. Sci. 8342, Springer, Berlin, Heidelberg, 2014, pp. 1–37.

[32] H.A. Le Thi and D.T. Pham, *The DC programming and DCA revised with DC models of real world nonconvex optimization problems*, Ann. Oper. Res., 133 (2005), pp. 25–46.

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

[34] Y.F. Lou and M. Yan, *Fast L_1-L_2 Minimization via a Proximal Operator*, preprint, arXiv:1609.09530.v2, 2016.

[35] Y.F. Lou, P.H. Yin, Q. He, and J. Xin, *Computing sparse representation in a highly coherent dictionary based on difference of $L1$ and $L2$*, J. Sci. Comput., 64 (2015), pp. 178–196.

[36] B. Jiang, S. Ma, and S. Zhang, *Alternating direction method of multipliers for real and complex polynomial optimization models*, Technical report, 2013; available at http://www.menet.umn.edu/~zhangs/Reports/2013_JMZ.pdf.

[37] G.Y. Li and T.K. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM J. Optim., 25 (2015), pp. 2434–2460.

[38] B. Mordukhovich, *Variational Analysis and Generalized Differentiation*, Springer, Berlin, 2006.

[39] J.M. Orgeta and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Classics Appl. Math. 30, SIAM, Philadelphia, PA, 2000.

[40] J.S. Pang, M. Razaviyayn, and A. Alvarado, *Computing B-stationary points of nonsmooth DC programs*, Math. Oper. Res., 42 (2017), pp. 95–118.

[41] D.T. Pham and H.A. Le Thi, *Convex analysis approach to DC programming: Theory, algorithm and applications*, Acta Math. Vietnam., 22 (1997), pp. 289–355.

[42] M. Razaviyayn, M. Hong, and Z.Q. Luo, *A unified convergence analysis of block successive minimization methods for nonsmooth optimization*, SIAM J. Optim., 23 (2013), pp. 1126–1153.

[43] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[44] R.T. Rockafellar and R.J.B. Wets, *Variational Analysis*, 3rd ed., Grundlehren Math. Wiss. 317, Springer, Berlin, Heidelberg, 2009.

[45] M. Tao and X.M. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim., 21 (2011), pp. 57–81.

[46] F.H. Wang, Z.B. Xu, and H.K. Xu, *Convergence of Bregman Alternating Direction Method with Multipliers for Nonconvex Composite Problems*, preprint, https://arxiv.org/pdf/1410.8625v3.pdf, 2014.

[47] Y. Wang, W.T. Yin and J.S. Zeng, *Global Convergence of ADMM in Nonconvex Nonsmooth Optimization*, preprint, http://arxiv.org/pdf/1511.06324v3.pdf, 2016.

[48] F.H. Wang, W.F. Cao, and Z.B. Xu, *Convergence of Multi-Block Bregman ADMM for Nonconvex Composite Problems*, preprint, https://arxiv.org/pdf/1505.03063v1.pdf, 2015.

[49] Z. Wen, C. Yang, X. Liu, and S. Marchesini, *Alternating direction methods for classical and ptychographic phase retrieval*, Inverse Problems, 28 (2012), pp. 1–18.

[50] S.J. Wright, *Coordinate descent algorithms*, Math. Program., 151 (2015), pp. 3–34.

[51] Y.Y. Xu, W.T. Yin, Z.W. Wen, and Y. Zhang, *An alternating direction algorithm for matrix completion with nonnegative factors*, Front. Math. China, 7 (2012), pp. 365–384.

[52] Y. Xu and W.T. Yin, *Block stochastic gradient iteration for convex and nonconvex optimization*, SIAM J. Optim., 25 (2015), pp. 1686–1716.

[53] Y. Xu and W.T. Yin, *A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion*, SIAM J. Imag. Sci., 6 (2013), pp. 1758–1789.

[54] P.H. Yin, Y.F. Lou, Q. He, and J. Xin, *Minimization of ℓ_{1-2} for compressed sensing*, SIAM J. Sci. Comput., 37 (2015), A536–A563.

[55] Y. Zhang, *An Alternating Direction Algorithm for Nonnegative Matrix Factorization*, Technical report, Rice University, Houston, TX, 2010; available at http://www.caam.rice.edu/~zhang/reports/tr1003.pdf.