SIAM J. OPTIM.
Vol. 24, No. 2, pp. 795–822
© 2014 Society for Industrial and Applied Mathematics

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

A PARALLEL BUNDLE FRAMEWORK FOR ASYNCHRONOUS SUBSPACE OPTIMIZATION OF NONSMOOTH CONVEX FUNCTIONS*

FRANK FISCHER† AND CHRISTOPH HELMBERG‡

Abstract. An algorithmic framework is presented for optimizing general convex functions by nonsynchronized parallel processes. Each process greedily picks a suitable adaptive subset of coordinates and runs a bundle method on a corresponding restricted problem stopping whenever a descent step is encountered or predicted decrease is reduced sufficiently. No prior knowledge on the dependencies between variables is assumed. Instead, dependency information is collected automatically by analyzing aggregate subgradient properties required for ensuring convergence. Within this framework three strategies are discussed for supporting varying scenarios of structural independence: a single convex function, the sum of partially separable convex functions, and a scenario tuned to problem decomposition by Lagrangian relaxation of packing-type constraints. The theoretical framework presented here generalizes a practical method proposed by the authors for Lagrangian relaxation of large scale packing problems and simplifies the analysis.

Key words. bundle methods, parallel programming, convex optimization

AMS subject classifications. 90C06, 65Y05, 90C25, 65K05

DOI. 10.1137/120865987

1. Introduction.

Let $f\colon \mathbb{R}^M \to \mathbb{R}$, $M = \{1, \dots, m\}$, be a nonsmooth, convex function specified via a *first order oracle*, i.e., given a point $y \in \mathbb{R}^M$ the oracle returns the function value $f(y)$ and a subgradient $g \in \partial f(y)$ at y. We consider the optimization problem

$$\text{(P)} \qquad \begin{aligned} &\text{minimize} && f(y) \\ &\text{subject to} && y \in \mathbb{R}^M. \end{aligned}$$

Bundle methods are an established tool for solving such optimization problems [1, 7]. In a nutshell, they collect subgradient information from the set

$$W := \operatorname{conv}\left\{(l, g)\colon l = f(y) - \langle g, y\rangle, g \in \partial f(y), y \in \mathbb{R}^M\right\},$$

and, based on this, they determine, in their simplest form, an appropriate *aggregate* $\bar{w} = (\bar{l}, \bar{g}) \in W$ giving rise to a cutting plane model consisting of a single affine minorant

$$\hat{f}_{\bar{w}}(y) := \bar{l} + \langle \bar{g}, y\rangle \quad \le f(y) \quad \forall\, y \in \mathbb{R}^M.$$

Starting from a given *center of stability*, iteratively a next *candidate* is determined with respect to the current center of stability $\hat{y}$ and aggregate $\bar{w}$ by

$$\bar{y} = \arg\min\left\{\hat{f}_{\bar{w}}(y) + \tfrac{u}{2}\|y - \hat{y}\|^2\colon y \in \mathbb{R}^M\right\} = \hat{y} - \tfrac{1}{u}\bar{g},$$

*Received by the editors February 14, 2012; accepted for publication (in revised form) January 29, 2014; published electronically May 8, 2014. This work was supported by the *Bundesministerium für Bildung und Forschung* under grant 05M10OCD. Responsibility for the content rests with the authors.

http://www.siam.org/journals/siopt/24-2/86598.html

†Institut für Mathematik, Universität Kassel, D-34132 Kassel, Germany (frank.fischer@mathematik.uni-kassel.de).

‡Fakultät für Mathematik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany (helmberg@mathematik.tu-chemnitz.de).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where the (here fixed) *weight* $u > 0$ penalizes the distance to $\hat{y}$. The actual progress $f(\hat{y}) - f(\bar{y})$ is then compared to the *predicted decrease*

$$\Delta(\bar{w}, \hat{y}) := f(\hat{y}) - \hat{f}_{\bar{w}}(\bar{y}) = f(\hat{y}) - \hat{f}_{\bar{w}}(\hat{y}) + \frac{1}{u}\|\bar{g}\|^2 \geq 0.$$

If the actual progress achieves a fraction of at least $\rho \in (0,1)$ relative to this predicted decrease, the method performs a *descent step* by setting $\hat{y} \leftarrow \bar{y}$. Otherwise, in a *null step*, $\hat{y}$ remains unchanged, but the new subgradient information is used to update $\bar{w}$ so as to ensure a significant reduction in $\Delta(\bar{w}, \hat{y})$. Our method relies heavily on the following result.

THEOREM 1 (e.g., Theorems 10.14 and 10.15 in [1]). *Let $\hat{y}^k$ and Δ^k denote the center and predicted decrease in iteration k of the bundle method, then $f(\hat{y}^k) \to \inf f$. Moreover, if the method generates a finite number of descent steps with $\hat{y}^{k_0} = \hat{y}^k$ for $k \geq k_0$ then $\hat{y}^{k_0}$ minimizes f and the sequence $(\Delta^k)_{k>k_0}$ is monotonically decreasing to 0.*

The parallel bundle method proposed here extends and simplifies the analysis of the parallel bundle approach for Lagrangian relaxation presented in [3]. It consists of a main routine (Algorithm 1) that initializes and manages access to global data and maintains a number $N_\Pi \in \{1, \ldots, |M|\}$ of parallel processes performing the actual optimization independently and asynchronously. To provide a brief sketch of the main idea, the global data include an access index $\sigma \in \mathbb{N}_0$, the current center $\hat{y}^{\langle\sigma\rangle}$, the aggregate $\bar{w}^{\langle\sigma\rangle}$, the corresponding predicted decrease $\Delta^{\langle\sigma\rangle} := \Delta(\bar{w}^{\langle\sigma\rangle}, \hat{y}^{\langle\sigma\rangle})$, as well as some automatically updated dependency ($S^{\langle\sigma\rangle}$) and blocking information ($B^{\langle\sigma\rangle}$) to be explained later. Each parallel process π started by the main routine performs the following three steps.

1. Subspace selection. Securing exclusive access to the global data at some access index $\underline{\pi} = \sigma$, process π tries to reserve an unblocked subset of coordinates $J^{(\pi)} \subseteq M$ defining an affine subspace

$$\mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle}) := \hat{y}^{\langle\underline{\pi}\rangle} + \operatorname{span}\{e_j : j \in J^{(\pi)}\} = \left\{ y \in \mathbb{R}^M : y_{M\setminus J^{(\pi)}} = \hat{y}^{\langle\underline{\pi}\rangle}_{M\setminus J^{(\pi)}} \right\},$$

so that, according to current dependency information, optimizing f over $\mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle})$ promises a significant share of $\Delta^{\langle\underline{\pi}\rangle}$. If this fails due to blocking conditions, π frees access again and stops immediately without having modified any data (a new process may be restarted once the global data are changed in the third step by some other active process). If successful, it marks the coordinates $J^{(\pi)}$ as blocked in the global data, retrieves all information needed for setting up an independent convex subproblem aimed at, but not necessarily identical to, optimizing f over $\mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle})$, increases the access index σ, and frees global access. Note, whenever no other process is running, this subspace selection will always be successful.

2. Subspace optimization. Process π forms the subproblem and optimizes it by a separate bundle method. This is done without any global interaction. It stops (in finite time) if the progress on the subproblem guarantees a descent step for f with respect to the global situation encountered initially by π at access counter $\underline{\pi}$, or it stops if the reduction in predicted decrease on the subproblem enables π to update the global aggregate so as to significantly reduce the global predicted decrease encountered at $\underline{\pi}$. In hindsight these properties may fail to hold if the initial dependency information turns out to be incorrect, which must be checked in the next step.

3. Subspace update. This final step of process π starts by securing exclusive access to the global data, granted at some access index $\bar{\pi} = \sigma > \underline{\pi}$, and checks whether the solution computed for the subproblem matches expectations in view of

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ALGORITHM 1. PARALLELBUNDLE.

Input : Problem (P), parameters $\tau_1 \in (0,1)$, $\varepsilon > 0$, $N_\Pi \in \{1, \ldots, |M|\}$
Output : Approximate solution $y \in \mathbb{R}^M$
`// Initialization`
set $\sigma \leftarrow 0$, $\hat{y}^{\langle 0\rangle} \leftarrow 0$, $B^{\langle 0\rangle} \leftarrow \emptyset$
set $\bar{w}^{\langle 0\rangle} = (\bar{l}^{\langle 0\rangle}, \bar{g}^{\langle 0\rangle})$ an initial minorant `// requires a first evaluation`
set $\Delta^{\langle 0\rangle} \leftarrow \Delta(\bar{w}^{\langle 0\rangle}, \hat{y}^{\langle 0\rangle})$
set $S^{\langle 0\rangle} \leftarrow \emptyset$ (or use some prespecified dependencies)
1 **if** $\Delta^{\langle 0\rangle} \leq \varepsilon(|f(\hat{y}^{\langle 0\rangle})| + 1)$ **then**
  then do not start any process
  set $y \leftarrow \hat{y}^{\langle 0\rangle}$
  return
`// Main loop`
while *Less than* N_Π *processes are running* **do**
  Start a new process $\pi = (\underline{\pi}, \bar{\pi}) \leftarrow (-1, -1)$.
  Each *process* π *performs the following steps independently*
2 **Subspace selection**
    Secure exclusive access to global data $\underline{\pi} \leftarrow \sigma$
    if SELECTSUBSPACE$(\underline{\pi})$ = ***FALSE*** **then**
      `// Step is unsuccessful`
      Free exclusive access to global data
      STOP this process
3 $(\bar{w}^{\langle \underline{\pi}+1\rangle}, \hat{y}^{\langle \underline{\pi}+1\rangle}, S^{\langle \underline{\pi}+1\rangle}, \Delta^{\langle \underline{\pi}+1\rangle}) \leftarrow (\bar{w}^{\langle \underline{\pi}\rangle}, \hat{y}^{\langle \underline{\pi}\rangle}, S^{\langle \underline{\pi}\rangle}, \Delta^{\langle \underline{\pi}\rangle})$
4 $B^{\langle \underline{\pi}+1\rangle} \leftarrow B^{\langle \underline{\pi}\rangle} \cup J^{(\pi)}$
    $\bar{\pi} \leftarrow \infty$, $\sigma \leftarrow \underline{\pi} + 1$
    Free exclusive access to global data
  Subspace optimization
    Solve a subspace problem on $J^{(\pi)}$ yielding $\bar{y}^{(\pi)}$ and $\bar{w}^{(\pi)}$
5 **Subspace update**
    Secure exclusive access to global data
    $\bar{\pi} \leftarrow \sigma$
    `// Set` $(\bar{w}^{\langle \bar{\pi}+1\rangle}, \hat{y}^{\langle \bar{\pi}+1\rangle}, S^{\langle \bar{\pi}+1\rangle})$ `to the values computed.`
6 UPDATESUBSPACE$(\pi, \bar{y}^{(\pi)}, \bar{w}^{(\pi)})$
    compute $\Delta^{\langle \bar{\pi}+1\rangle}$
7 $B^{\langle \bar{\pi}+1\rangle} \leftarrow B^{\langle \bar{\pi}\rangle} \setminus J^{(\pi)}$
    $\sigma \leftarrow \bar{\pi} + 1$
8 **if** $\Delta^{\langle \bar{\pi}+1\rangle} \leq \varepsilon(|f(\hat{y}^{\langle \bar{\pi}+1\rangle})| + 1)$ **then**
      set $y \leftarrow \hat{y}^{\langle \sigma\rangle}$
      TERMINATE *all* processes and **STOP**.
    Free exclusive access to global data
    STOP this process π.

the new global situation. If this is not the case, additional tests ensure that this is either due to intermediate changes in the relevant global data, or that it results in adaptations of the dependency information. The process modifies the global data accordingly, frees its own blocked subspace, increases the access index σ, and checks

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

whether the new global data satisfy a global termination criterion (by the usual lack of progress criterion $\Delta^{\langle\bar{\pi}+1\rangle} \leq \varepsilon(|f(\hat{y}^{\langle\bar{\pi}+1\rangle})| + 1)$ for some fixed $\varepsilon > 0$). In the case of global termination, all processes are stopped immediately and the algorithm ends. Otherwise π frees exclusive access and only stops itself (alternatively, the process may be restarted with step one).

Two main challenges are associated with the proposed algorithmic framework. The first consists in establishing weak requirements on the properties of subproblems, the automatic generation of dependency information and blocking requirements that still allow us to ensure convergence of this nonsynchronized approach. After an introductory discussion of some notational issues associated with the parallel mechanism in section 2, this is fully addressed in section 3. The second challenge consists in describing appropriate subproblems and automatic dependency generation schemes for practical problem classes of f that promise computational advantages of this parallel approach by being able to locate and exploit small independent subspaces allowing for partial evaluations. For this we will consider three different scenarios of increasing complexity and increasing potential for practical efficiency. First, we discuss a simple model for general convex functions in section 4. Next, section 5 considers optimizing the sum of partially separable convex functions, where for each subfunction the set of variables influencing it is supposed to be small and known. Finally, section 6 proposes a more involved algorithmic approach intended for large scale models, where we now assume that the influence of variables on the subfunctions has yet to be detected. This last approach is motivated by Lagrangian relaxation of packing-type constraints. A description of a practical implementation of the approach of the last section as well as computational results on (artificial but realistic) train time-tabling problems can be found in [3].

Literature review. We give a short overview of existing parallelization approaches for convex optimization; also see [3]. The most straightforward approach is to exploit a separable objective function $f(y) = \sum_{r\in R} f_r(y)$, which appears generically in Lagrangian relaxation, e.g., [10, 9, 4], and evaluate each function f_r in an isolated parallel process; see, e.g., [11] for an example within a bundle algorithm.

Another approach is using variable transformation algorithms [5, 13]. Similarly to our approach, these algorithms select a set of subspaces and compute new candidates on each of the subspaces. Unlike our approach, however, they compute a new global iterate in a synchronized step as a combination of the subspace candidates.

Incremental subgradient methods may also be developed into an asynchronous parallel scheme. These methods change y according to the directions specified by the single subfunctions f_r in turn. In [12] this is extended to a fully asynchronous approach that only requires that the number of iterations between two evaluations of the same subfunction is bounded by a constant and each subfunction is evaluated asymptotically the same number of times (which is also the case in the "synchronized" approaches mentioned above). A similar approach is using incremental bundle methods [2], which compute the next iterate by evaluating only a few subfunctions and by approximating the other subfunctions by their current model.

In contrast to these approaches, our method has no explicit synchronization or regularization mechanism (e.g., our approach does no intermediate steepest-descent-like steps) and has, in the case of a separable objective function, no restriction on the number and frequency of evaluations of the single subfunctions. Indeed, the number of evaluations may differ significantly and this may be a main source of computational efficiency. Our framework ensures convergence via an automatic dependency detection mechanism between subspaces.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2. The parallel framework. In this section we recall from [3] how parallelism works in the proposed method and how we denote the different parallel subprocesses and their associated "local" data as well as the shared "global" data. For each process the only interaction with, and particularly the only change of, global data happens in the first and third steps. As in any parallel framework, simultaneous writing (or reading and writing) to the (shared) global data from more than one parallel subprocess must be forbidden using some locking mechanism. Because of this there is a *unique sequence* of the interactions of the processes with the global data. Thus we can assign each state of the global data a unique index $\sigma \in \mathbb{N}_0$, which will be increased each time the global data are modified. Each global object will be denoted by a superscript σ in angle braces $\langle\sigma\rangle$.

Let π be a process. Because of the uniqueness of the global index we can equip π with the two special index markers $\underline{\pi}$ and $\bar{\pi}$ that correspond to the global index when π executes its subspace selection, resp., update step.

For the analysis it will be convenient to arrange the processes in different groups for each global index $\sigma \in \mathbb{N}_0 \cup \{\infty\}$:

$$\underline{\Pi}^{\langle\sigma\rangle} := \{\pi = (\underline{\pi}, \bar{\pi})\colon \underline{\pi} < \bar{\pi} \text{ and } \underline{\pi} < \sigma\},$$
$$\bar{\Pi}^{\langle\sigma\rangle} := \{\pi = (\underline{\pi}, \bar{\pi})\colon \underline{\pi} < \bar{\pi} < \sigma\},$$
$$\Pi^{\langle\sigma\rangle} := \{\pi = (\underline{\pi}, \bar{\pi})\colon \underline{\pi} < \sigma \le \bar{\pi}\} = \underline{\Pi}^{\langle\sigma\rangle} \setminus \bar{\Pi}^{\langle\sigma\rangle}.$$

The first group $\underline{\Pi}^{\langle\sigma\rangle}$ collects all processes that have successfully executed the subspace selection step before the algorithm reached σ, $\bar{\Pi}^{\langle\sigma\rangle}$ singles out all processes that have executed the subspace update step before σ, and the set $\Pi^{\langle\sigma\rangle}$ comprises all processes that are actively working on or just finishing a subproblem at σ, i.e., these are running in parallel. Note that the set $\underline{\Pi}^{\langle\infty\rangle}$ is precisely the set of all processes that successfully selected a subspace. Each increment of σ by the algorithm is associated with the addition or deletion of exactly one process from this set of parallel processes.

OBSERVATION 2 (Obs. 5 in [3]). *The following relations hold:*

(i) $\Pi^{\langle 0\rangle} = \emptyset$*;*

(ii) *for* $\sigma' \in \mathbb{N}_0$ *there holds* $\Pi^{\langle\sigma'\rangle} \neq \Pi^{\langle\sigma'+1\rangle}$ *if and only if* $|(\Pi^{\langle\sigma'\rangle} \setminus \Pi^{\langle\sigma'+1\rangle}) \cup (\Pi^{\langle\sigma'+1\rangle} \setminus \Pi^{\langle\sigma'\rangle})| = 1$*;*

(iii) *if* $\Pi^{\langle\sigma'\rangle} = \Pi^{\langle\sigma'+1\rangle}$ *then for all* $\sigma \ge \sigma'$ *we have* $\Pi^{\langle\sigma\rangle} = \Pi^{\langle\sigma'\rangle}$ *and there is no process* π *with* $\underline{\pi} = \sigma$ *or* $\bar{\pi} = \sigma$*.*

In particular, if $\Pi^{\langle\sigma\rangle} \neq \Pi^{\langle\sigma+1\rangle}$ *then there is exactly* one *process* π *with* $\underline{\pi} = \sigma$ *or* $\bar{\pi} = \sigma$*.*

By this observation, the following set collects the markers $\sigma \in \mathbb{N}_0$ visited by the algorithm: $\Sigma := \{0\} \cup \{\sigma \in \mathbb{N}\colon \Pi^{\langle\sigma\rangle} \neq \Pi^{\langle\sigma-1\rangle}\}$.

3. The basic asynchronous parallel bundle method. The main difficulty in the analysis of the algorithm is due to dependencies between different subspaces. A subprocess improving the predicted decrease w.r.t. a certain subspace may worsen the predicted decrease along other directions not contained in the subspace. If this happens too often, the algorithm may not converge correctly. Therefore the parallel bundle method automatically collects *dependency information* between subspaces. For this, we use a *finite*, partially ordered set of *states* $(\mathcal{S}, \preceq)$ with a unique maximal element $\bar{S} \in \mathcal{S}$. The algorithm will change its current dependency state from S to $S' \succ S$ if some bad dependency is detected. The selection of further subspaces and creation of subspace problems working on these subspaces will be influenced by these states.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

As long as the algorithm runs, the main process will start new subprocesses working on different subspaces; in particular, each process π will select and work on a subset of coordinates $J^{\langle\pi\rangle} \subseteq M$. The purpose of the subspace problem is to either create a descent step along that subspace or reduce the global predicted decrease sufficiently. So the subspace must be carefully selected in order to provide sufficient decrease compared with the global expected progress. Furthermore, two different processes may influence each other if they work on the same or "strongly related" subspaces. In order to choose subspaces so that the mutual influence is small, the algorithm manages a set of *blocked constraints* B that may not be selected by further parallel subprocesses. Putting all these ingredients together, the algorithm handles the following global data:

1. the global center $\hat{y}^{\langle\sigma\rangle} \in \mathbb{R}^M$,
2. the global aggregated minorant $\bar{w}^{\langle\sigma\rangle} = (\bar{l}^{\langle\sigma\rangle}, \bar{g}^{\langle\sigma\rangle}) \in W$,
3. the global dependency state $S^{\langle\sigma\rangle} \in \mathcal{S}$,
4. the global blocked constraints $B^{\langle\sigma\rangle} \subseteq M$ satisfying $B^{\langle\sigma\rangle} = \bigcup_{\pi\in\Pi^{\langle\sigma\rangle}} J^{\langle\pi\rangle}$.

We refer to the global expected progress w.r.t. the data at σ by $\Delta^{\langle\sigma\rangle} = \Delta(\bar{w}^{\langle\sigma\rangle}, \hat{y}^{\langle\sigma\rangle})$.

The first main step of a process π is to select appropriate subspace problems. Formally, any given current state $S^{\langle\sigma\rangle}$, current center $\hat{y}^{\langle\sigma\rangle}$, and subspace $J^{\langle\pi\rangle}$ determine a subspace problem. Depending on the problem class of f and the corresponding choice of subspace problems it will be necessary to discern strong and weak forms of dependencies between subsets of coordinates. Strong forms of dependence of some coordinates J' on a given subset J will require us to jointly optimize over $J' \cup J$ whenever J is selected as the initial subspace. If it seems unnecessary to include the coordinates together with J, but no other process should modify the values on J' while J is optimized over, then we speak of weak dependence. In order to reflect detected or assumed dependencies between subspaces w.r.t. the current state S, we employ two functions, a *weakly dependent subspace function* F_d and a *strongly dependent subspace function* F_D:

$$F_X \colon \mathcal{S} \times 2^M \to 2^M$$

with

$$\forall\, S \in \mathcal{S} \colon F_X(S, \emptyset) = \emptyset, \tag{3.1}$$

$$\forall\, S \in \mathcal{S}, J \subseteq M \colon J \subseteq F_X(S, J), \tag{3.2}$$

$$\forall\, S \in \mathcal{S}, J, J' \subseteq M \colon F_X(S, J \cup J') = F_X(S, J) \cup F_X(S, J'), \tag{3.3}$$

$$\forall\, S \preceq S' \in \mathcal{S}, J \subseteq M \colon F_X(S, J) \subseteq F_X(S', J). \tag{3.4}$$

We will use the following short notation for the functions w.r.t. state $S^{\langle\sigma\rangle}$:

$$F_D^{\langle\sigma\rangle}(J) := F_D(S^{\langle\sigma\rangle}, J), \qquad F_d^{\langle\sigma\rangle}(J) := F_d(S^{\langle\sigma\rangle}, J).$$

The subspace selection step will guarantee that for each two processes $\pi \neq \eta \in \Pi^{\langle\sigma\rangle}$ running at σ with corresponding subspaces $J^{\langle\pi\rangle}$ and $J^{\langle\eta\rangle}$ there holds

$$J^{\langle\pi\rangle} = F_D^{\langle\pi\rangle}(J) \text{ for some } J \subseteq M,$$
$$F_d^{\langle\pi\rangle}(J^{\langle\pi\rangle}) \cap J^{\langle\eta\rangle} = \emptyset = J^{\langle\pi\rangle} \cap F_d^{\langle\eta\rangle}(J^{\langle\eta\rangle}). \tag{3.5}$$

Finally, in order to allow different progress measures for subproblems in dependence

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

on the state, we assume that there is a *subspace progress function*

$$\Delta\colon \mathcal{S} \times 2^M \times W \times \mathbb{R}^M \to \mathbb{R}_+$$

with

$$\forall\, S \in \mathcal{S}, \bar{w} \in W, \hat{y} \in \mathbb{R}^M\colon \Delta(S, M, \bar{w}, \hat{y}) = \Delta(\bar{w}, \hat{y}). \tag{3.6}$$

This function predicts the expected progress to be made by the subspace problem. Only subspace problems that are guaranteed to satisfy (3.5) and that have a large predicted decrease compared with the global predicted decrease will be chosen.

3.1. Subspace selection. The first step of a process π is the subspace selection step at access index $\underline{\pi}$. The goal is to select a subspace that predicts a large progress compared to the global predicted decrease. Formally, the algorithm checks whether for a subspace $J \subseteq M$ there holds

$$\Delta(S^{\langle\underline{\pi}\rangle}, J, \bar{w}^{\langle\underline{\pi}\rangle}, \hat{y}^{\langle\underline{\pi}\rangle}) \geq \tau_1 \Delta^{\langle\underline{\pi}\rangle} \tag{Sel1}$$

for some parameter $\tau_1 \in (0, 1)$. Furthermore, all processes running in parallel must satisfy (3.5). This is guaranteed by verifying

$$J \cap F_d^{\langle\underline{\pi}\rangle}(B^{\langle\underline{\pi}\rangle}) = \emptyset = F_d^{\langle\underline{\pi}\rangle}(J) \cap B^{\langle\underline{\pi}\rangle}. \tag{Sel2}$$

An important property is that it is always possible to select a subspace satisfying (Sel1) and (Sel2) if no other process is currently running. In particular, if $B^{\langle\underline{\pi}\rangle} = \emptyset$, then (3.6) implies that $J = M$ satisfies (Sel1), and (Sel2) is satisfied trivially. These observations motivate the following simple greedy algorithm to select an appropriate subspace. Starting with an empty subspace $J = \emptyset$, we add further variables so that (Sel2) remains valid until (Sel1) holds. If the set of blocked variables is empty, then this greedy strategy will succeed eventually by selecting $J = M$. The subspace selection procedure is shown in Algorithm 2.

OBSERVATION 3. *If Algorithm* 2 *returns a subspace* $\emptyset \neq J^{(\pi)} \subseteq M$*, then* $J^{(\pi)}$ *fulfils (Sel*1*) and (Sel*2*). If the algorithm is called without blocked variables, i.e.,* $B^{\langle\underline{\pi}\rangle} = \emptyset$*, then the returned subspace is not empty, i.e.,* $J^{(\pi)} \neq \emptyset$*.*

Proof. First observe that the returned subspace will satisfy (Sel1) because of the final test. Next we prove (Sel2) inductively for $J^{(\pi)}$ as constructed. The empty subspace $J^{(\pi)} = \emptyset$ satisfies (Sel2) by (3.1). Furthermore Y is only added to $J^{(\pi)}$ if $F_d^{\langle\underline{\pi}\rangle}(Y) \cap B^{\langle\underline{\pi}\rangle} = \emptyset = F_d^{\langle\underline{\pi}\rangle}(B^{\langle\underline{\pi}\rangle}) \cap Y$. Thus by (3.3) condition (Sel2) remains true when $J^{(\pi)}$ is increased. In particular, if $B^{\langle\underline{\pi}\rangle} = \emptyset$ then $F_d^{\langle\underline{\pi}\rangle}(B^{\langle\underline{\pi}\rangle}) = \emptyset$, too, so each considered variable will be added. Finally the set X is reduced in each iteration by at least one variable, so the loop will terminate after at most $|M|$ iterations. If $J^{(\pi)} = M$ then (Sel1) will be satisfied because of property (3.6). ∎

3.2. Subspace optimization and subspace update. The actual structure of the subspace problem is unknown to the main routine. However, the algorithm underlying the processes π has to guarantee certain conditions on the values returned in the update of the global data to $\sigma = \bar{\pi} + 1$. In particular, we make the following assumption.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ALGORITHM 2. SELECTSUBSPACE.

Input : global data at $\underline{\pi}$
Output : **TRUE** if a subspace can be selected
Changes : sets $J^{(\pi)}$
$X \leftarrow M \setminus F_d^{\langle\underline{\pi}\rangle}(B^{\langle\underline{\pi}\rangle})$, $J^{(\pi)} \leftarrow \emptyset$
while $J^{(\pi)}$ *does not satisfy (Sel*1*) and* $X \neq \emptyset$ **do**
1 select $j \in \operatorname{Arg\,max}\{\Delta(S^{\langle\underline{\pi}\rangle}, \{j\}, \bar{w}^{\langle\underline{\pi}\rangle}, \hat{y}^{\langle\underline{\pi}\rangle}) \colon j \in X\}$
$Y \leftarrow F_D^{\langle\underline{\pi}\rangle}(\{j\})$
2 **if** $F_d^{\langle\underline{\pi}\rangle}(Y) \cap B^{\langle\underline{\pi}\rangle} = \emptyset$ *and* $F_d^{\langle\underline{\pi}\rangle}(B^{\langle\underline{\pi}\rangle}) \cap Y = \emptyset$ **then**
$J^{(\pi)} \leftarrow J^{(\pi)} \cup Y$, $X \leftarrow X \setminus Y$
else
$X \leftarrow X \setminus \{j\}$
if $J^{(\pi)}$ *satisfies (Sel*1*)* **then**
return *TRUE*
else
$J^{(\pi)} \leftarrow \emptyset$
return *FALSE*

Assumption 4.

(i) Any process $\pi \in \bar{\Pi}^{\langle\infty\rangle}$ satisfies

$$\text{(Inv1)} \qquad S^{\langle\bar{\pi}\rangle} \preceq S^{\langle\bar{\pi}+1\rangle},$$

$$\text{(Inv2)} \qquad f(\hat{y}^{\langle\bar{\pi}\rangle}) \geq f(\hat{y}^{\langle\bar{\pi}+1\rangle}).$$

(ii) There exist constants $\tau_2, \tau_3 \in (0,1)$ so that either there is an infinite number of processes $\pi \in \bar{\Pi}^{\langle\infty\rangle}$ satisfying

$$\text{(Upd1)} \qquad f(\hat{y}^{\langle\underline{\pi}\rangle}) - f(\hat{y}^{\langle\bar{\pi}+1\rangle}) \geq \tau_2 \Delta^{\langle\underline{\pi}\rangle},$$

or there is a $\sigma_1 \in \Sigma$ so that each process $\pi \in \bar{\Pi}^{\langle\infty\rangle}$ with $\bar{\pi} \geq \sigma_1$ satisfies

$$\text{(Upd2)} \qquad \hat{y}^{\langle\sigma_1\rangle} = \hat{y}^{\langle\underline{\pi}\rangle} = \hat{y}^{\langle\bar{\pi}+1\rangle} \text{ and } \Delta^{\langle\bar{\pi}+1\rangle} \leq \tau_3 \Delta^{\langle\underline{\pi}\rangle}.$$

Conditions (Inv2), (Upd1), and (Upd2) require the processes to equip the global iterates with similar properties to those in classical bundle methods, which allows us to follow standard convergence arguments. Indeed, a process satisfying condition (Upd1) performs a global descent step of sufficient decrease in the objective value (compared with the global predicted decrease when π started). Together with condition (Inv2) an infinite number of such descent steps guarantees convergence if $\operatorname{Arg\,min} f$ is bounded; see section 3.3. Because in an asynchronous setting infinitely many processes may fail to satisfy (Upd1), condition (Inv2) cannot be dropped; see Example 5. Condition (Upd2) ensures that, in lack of further descent steps, eventually the global predicted decrease is reduced by a constant factor by each process, which will establish optimality of $\hat{y}^{\langle\sigma_1\rangle}$. Example 6 illustrates that processes restricted to small subspaces may not be able to satisfy Assumption 4(ii). Therefore some subspace dependency information needs to be collected over time and the states $S^{\langle\sigma\rangle}$ serve this purpose. The monotonicity condition (Inv1) helps to establish the validity of Assumption 4(ii) but is by itself not essential for convergence.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Example 5. To see the necessity of condition (Inv2), consider minimizing $f\colon \mathbb{R}^2 \to \mathbb{R}$,

$$f(y_1, y_2) = |y_1 + y_2|$$

with initial center $\hat{y}^{\langle 0\rangle} = (0,-1)^T \in \mathbb{R}^2$ and aggregate $\bar{w}^{\langle 0\rangle} = (\bar{l}^{\langle 0\rangle}, \bar{g}^{\langle 0\rangle}) = (1,(1,1)^T)$, and let processes π_1, π_2 with $\underline{\pi}_1 = 0$ and $\underline{\pi}_2 = 1$ pick the coordinate directions $J^{(\pi_1)} = \{1\}$ and $J^{(\pi_2)} = \{2\}$ as subspaces. Suppose both processes compute the optimizers of their respective subspaces, i.e., $\bar{y}^{(\pi_1)} = (1,-1)^T$ and $\bar{y}^{(\pi_2)} = (0,0)^T$; then process π_1 first updates coordinate 1 to 1 (resulting in $\hat{y}^{\langle 3\rangle} = (1,-1)^T$ in satisfaction of (Upd1)) and afterwards, in the absence of condition (Inv2), process π_2 updates coordinate 2 to 0 (giving $\hat{y}^{\langle 4\rangle} = (1,0)^T$). Note, the new center is symmetric to the initial one, admitting an analogous step back to the initial point. Thus, without (Inv2), Assumption 4(ii) is not sufficient for convergence.

Example 6. For illustrating subspace dependencies, let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(y_1, y_2) = \max\{-y_1, -y_2\}.$$

Starting in $\hat{y}^{\langle 0\rangle} = (0,0)^T \in \mathbb{R}^2$ with aggregate $(\bar{l}^{\langle 0\rangle}, \bar{g}^{\langle 0\rangle})$ satisfying $\bar{g}^{\langle 0\rangle} \in \partial f(\hat{y}^{\langle 0\rangle}) = \operatorname{conv}\{(-1,0)^T, (0,-1)^T\}$, at least one of the two coordinate directions has a sufficiently large predicted subspace decrease. Assume process π optimizes along subspace $J^{(\pi)} = \{1\}$. Because $\hat{y}^{\langle 0\rangle}$ is already an optimal solution on the subspace $\mathcal{L}(J^{(\pi)}, \hat{y}^{\langle 0\rangle})$, the process may reduce the predicted subspace decrease to 0 by updating the aggregate to $\bar{g} = (0,-1)^T$ but can neither satisfy (Upd1) nor (Upd2). This argument may be iterated for processes working alternatingly on coordinates 1 and 2. Violating (Upd2) in spite of a reduction in predicted subspace decrease implies an increase in other coordinates of $\bar{g}$ and these indicate dependencies that might be worth respecting in further subspace selections. The global state $S^{\langle\sigma\rangle}$ will be used to collect such information from earlier iterations and allows to change the subspace selection strategy.

In sections 4–6 three different algorithmic approaches for subspace optimization and subspace update will be proposed, each designed to work with increasingly detailed structural properties of f, that will collect different kinds of dependency information so that Assumption 4 can be guaranteed to hold for the selected subspaces.

3.3. Convergence analysis. Throughout we assume that the algorithm underlying the processes satisfies Assumption 4. The parallel bundle algorithm (Algorithm 1) is called with the problem to be solved and the parameters τ_1, ε, and u. In addition, the parameter N_Π is passed to the algorithm that specifies the maximal number of processes that should be started in parallel. Because each process works on a non-empty subspace and the subspaces of parallel processes are disjoint, the number of parallel processes is bounded by the number of variables, i.e., $|M|$. But in practice one would like to restrict the number of parallel processes to the number of processors available on the hardware platform to be used.

We start with a simple but important observation, which is fundamental for the convergence analysis.

OBSERVATION 7. *The following relations hold:*

(i) *for all* $\sigma \in \Sigma$: $B^{\langle\sigma\rangle} = \bigcup_{\pi\in\Pi^{\langle\sigma\rangle}} J^{(\pi)}$;

(ii) *for all* $\sigma \in \Sigma$: $S^{\langle\sigma\rangle} \preceq S^{\langle\sigma+1\rangle}$;

(iii) *for all* $\sigma \in \Sigma$: $f(\hat{y}^{\langle\sigma\rangle}) \geq f(\hat{y}^{\langle\sigma+1\rangle})$;

(iv) *there is a* $\sigma_0 \in \Sigma$ *so that* $S^{\langle\sigma\rangle} = S^{\langle\sigma_0\rangle}$ *for all* $\sigma \in \Sigma$ *with* $\sigma \geq \sigma_0$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. The proof works by induction on σ. For $\sigma = 0$ we have $B^{\langle\sigma\rangle} = \emptyset$ and $\Pi^{\langle\sigma\rangle} = \emptyset$, so the relation holds trivially. Now suppose $\sigma + 1 \in \Sigma$ and the claim holds for $\sigma \in \Sigma$. By definition, $\sigma + 1 \in \Sigma$ implies $\Pi^{\langle\sigma\rangle} \neq \Pi^{\langle\sigma+1\rangle}$, so Observation 2 asserts the existence of a unique $\eta \in (\Pi^{\langle\sigma\rangle} \setminus \Pi^{\langle\sigma+1\rangle}) \cup (\Pi^{\langle\sigma+1\rangle} \setminus \Pi^{\langle\sigma\rangle})$ and this η either satisfies $\underline{\eta} = \sigma$ or $\bar{\eta} = \sigma$.

If $\underline{\eta} = \sigma$ we have $\Pi^{\langle\sigma\rangle} = \Pi^{\langle\sigma+1\rangle} \setminus \{\eta\}$ and process η executes a successful subspace selection step (Algorithm 2) at σ. Thus the process executes A1.4 (which is short for Algorithm 1, line 4) implying $B^{\langle\sigma+1\rangle} = B^{\langle\sigma\rangle} \cup J^{(\eta)}$ as well as $S^{\langle\sigma+1\rangle} = S^{\langle\sigma\rangle}$ and $\hat{y}^{\langle\sigma\rangle} = \hat{y}^{\langle\sigma+1\rangle}$ and the claim holds. If $\bar{\eta} = \sigma$ we have $\Pi^{\langle\sigma+1\rangle} = \Pi^{\langle\sigma\rangle} \setminus \{\eta\}$ and process η executes line A1.7 at σ. The latter implies $B^{\langle\sigma+1\rangle} = B^{\langle\sigma\rangle} \setminus J^{(\eta)}$. Furthermore condition (Inv1) implies $S^{\langle\sigma\rangle} \preceq S^{\langle\sigma+1\rangle}$ and (Inv2) implies $f(\hat{y}^{\langle\sigma\rangle}) \geq f(\hat{y}^{\langle\sigma+1\rangle})$.

The last statement follows from the second and the fact that $\mathcal{S}$ is a finite set. ∎

The (due to the finiteness of $\mathcal{S}$) trivial fact that after a finite number of global updates the dependency information encoded in $S^{\langle\sigma\rangle}$ does not change anymore, will be vital for the convergence analysis. Indeed, constant dependency information will imply that no new dependencies are encountered that endanger convergence.

OBSERVATION 8. *Let $\sigma \in \Sigma$ and $\pi, \eta \in \Pi^{\langle\sigma\rangle}$. Then* (3.5) *holds for π and η.*

Proof. This follows from test A2.2, Observation 7, and properties (3.3) and (3.4). ∎

The next result characterizes the situation when the algorithm terminates. In particular, the global algorithm is guaranteed to keep at least one running process as long as the termination criterion is not satisfied.

OBSERVATION 9. *For each $\sigma \in \Sigma$ the following statements are equivalent:*

(i) $\sigma = \max \Sigma$*;*

(ii) $\Pi^{\langle\sigma\rangle} = \Pi^{\langle\sigma+1\rangle}$*;*

(iii) *the algorithm* terminated *with σ being the last global index;*

(iv) $\Delta^{\langle\sigma\rangle} \leq \varepsilon(|f(\hat{y}^{\langle\sigma\rangle})| + 1)$ *and* $\Delta^{\langle\sigma'\rangle} > \varepsilon(|f(\hat{y}^{\langle\sigma'\rangle})| + 1)$ *for all $\sigma' < \sigma$.*

In particular, $\max \Sigma = \infty$ *if and only if the algorithm does not terminate and* $\Delta^{\langle\sigma\rangle} > \varepsilon(|f(\hat{y}^{\langle\sigma\rangle})| + 1)$ *for all $\sigma \in \mathbb{N}_0 = \Sigma$.*

Proof. We prove this by induction on σ assuming that the equivalence holds for smaller global indexes.

(i) $\iff$ (ii). This follows by the definition of Σ and Observation 2.

(iii) $\Rightarrow$ (iv). If $\sigma = 0$ then the algorithm must have been terminated in line A1.1, thus (iv) follows. If $\sigma > 0$ then by induction $\Delta^{\langle\sigma'\rangle} > \varepsilon(|f(\hat{y}^{\langle\sigma'\rangle})| + 1)$ for each $\sigma' < \sigma$ (otherwise σ' had been the last global index). The only possibility to terminate the algorithm is that the test in line A1.8 succeeds for some process π with $\bar{\pi} = \sigma - 1$.

(iv) $\Rightarrow$ (iii). By (iv) either $\sigma = 0$ or $\sigma > 0$ and $(\bar{w}^{\langle\sigma-1\rangle}, \hat{y}^{\langle\sigma-1\rangle}) \neq (\bar{w}^{\langle\sigma\rangle}, \hat{y}^{\langle\sigma\rangle})$. In the first case the termination test in A1.1 succeeds and the algorithm terminates right after the initialization step. In the second case there must be a process π with $\bar{\pi} = \sigma - 1$ setting the values for $\sigma = \bar{\pi} + 1$ in A1.6 (because if $\underline{\pi} = \sigma - 1$ the process would execute line A1.3 and $(\bar{w}^{\langle\sigma-1\rangle}, \hat{y}^{\langle\sigma-1\rangle}) = (\bar{w}^{\langle\sigma\rangle}, \hat{y}^{\langle\sigma\rangle})$). Consequently the termination test in line A1.8 will succeed and the algorithm will be terminated.

(iii) $\Rightarrow$ (ii). If the algorithm terminates with last global index σ then no process will ever execute a subspace selection step or an update step at some later index $\sigma' \geq \sigma$, hence $\Pi^{\langle\sigma\rangle} = \Pi^{\langle\sigma+1\rangle}$ by Observation 2.

(ii) $\Rightarrow$ (iii). From (ii) implies that no process ever executes a *successful* subspace selection step or an update step at some index $\sigma' \geq \sigma$. Assume that the algorithm has not been terminated with last global index σ. We consider two cases. First if $\exists \pi \in \Pi^{\langle\sigma\rangle} \neq \emptyset$ then at least one process is still running at σ. Because each subspace optimization is finite by assumption this process will eventually execute

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

its update step at $\bar{\pi} \geq \sigma$ and would increase σ, a contradiction. Now assume $\Pi^{\langle\sigma\rangle} = \emptyset$. In particular this is the case if $\sigma = 0$. Then Observation 7 implies $B^{\langle\sigma\rangle} = \emptyset$. The algorithm will therefore try to start a new process π at $\underline{\pi} = \sigma$ and because of Observation 3 this step will be successful. This again causes σ to be increased, a contradiction. ∎

Next we show that under Assumption 4 the algorithm always drives the predicted decrease to zero on an appropriate subsequence whenever f is bounded from below.

LEMMA 10. *Suppose that an infinite number of processes satisfy (Upd1) and f is bounded from below. Let $\Sigma^{dc} := \{\underline{\pi}\colon \pi$ satisfies (Upd1)$\}$. Then $\liminf_{\sigma\in\Sigma^{dc}} \Delta^{\langle\sigma\rangle} = 0$.*

Proof. By Observation 7 the sequence $(f(\hat{y}^{\langle\sigma\rangle}))_{\sigma\in\Sigma}$ is nonincreasing and because f is bounded from below the sequence $f(\hat{y}^{\langle\underline{\pi}\rangle}) - f(\hat{y}^{\langle\bar{\pi}+1\rangle})$ converges to zero. Hence condition (Upd1) implies that $\Delta^{\langle\underline{\pi}\rangle}$ converges to zero, too. ∎

LEMMA 11. *Assume $|\Sigma| = \infty$ and there is only a finite number of descent steps, i.e., of processes satisfying (Upd1), and $\varepsilon = 0$, then $\lim_{\sigma\in\Sigma} \Delta^{\langle\sigma\rangle} = 0$.*

Proof. Because there is only a finite number or processes satisfying (Upd1), Assumption 4 ensures that there exists a $\sigma_1 \in \Sigma$ so that each process π with $\bar{\pi} \geq \sigma_1$ satisfies (Upd2). Define for $\sigma \in \Sigma$

$$\hat{\Delta}^{\langle\sigma\rangle} := \max\left(\{\Delta^{\langle\sigma\rangle}\} \cup \{\Delta^{\langle\underline{\pi}\rangle}\colon \pi \in \Pi^{\langle\sigma\rangle}\}\right)$$

and

$$\nu(\sigma) := \max\left(\{\sigma + N_\Pi + 1\} \cup \{\bar{\pi} + 1\colon \pi \in \Pi^{\langle\sigma+N_\Pi+1\rangle}\}\right).$$

We will show that for all $\sigma \geq \sigma_1$ there holds

(a) $\hat{\Delta}^{\langle\sigma\rangle} \geq \hat{\Delta}^{\langle\sigma+1\rangle}$,
(b) $\tau_3 \cdot \hat{\Delta}^{\langle\sigma\rangle} > \Delta^{\langle\bar{\pi}+1\rangle}$ for all processes π with $\bar{\pi} \geq \sigma$,
(c) $\tau_3 \cdot \hat{\Delta}^{\langle\sigma\rangle} > \hat{\Delta}^{\langle\nu(\sigma)\rangle}$.

Note that (a) and (c) imply $0 \leq \Delta^{\langle\sigma\rangle} \leq \hat{\Delta}^{\langle\sigma\rangle} \to 0$.

First we show (a). For this let π be the process with $\sigma \in \{\underline{\pi}, \bar{\pi}\}$. If $\sigma = \underline{\pi}$ then $\Pi^{\langle\sigma+1\rangle} = \Pi^{\langle\sigma\rangle} \cup \{\pi\}$ and $\hat{\Delta}^{\langle\sigma\rangle} = \hat{\Delta}^{\langle\sigma+1\rangle} = \Delta^{\langle\underline{\pi}\rangle}$ because no relevant global data changed in A1.3, so $\hat{\Delta}^{\langle\sigma\rangle} = \hat{\Delta}^{\langle\sigma+1\rangle}$. Now assume $\sigma = \bar{\pi}$, then $\Pi^{\langle\sigma+1\rangle} = \Pi^{\langle\sigma\rangle} \setminus \{\pi\}$. Because $\bar{\pi} \geq \sigma_1$ we know that π satisfies (Upd2). Consequently the update A1.6 implies $\Delta^{\langle\sigma+1\rangle} = \Delta^{\langle\bar{\pi}+1\rangle} < \tau_3 \cdot \Delta^{\langle\underline{\pi}\rangle} \leq \tau_3 \cdot \hat{\Delta}^{\langle\sigma\rangle}$, so $\hat{\Delta}^{\langle\sigma+1\rangle} \leq \hat{\Delta}^{\langle\sigma\rangle}$.

In order to prove (b), observe that either $\pi \in \Pi^{\langle\sigma\rangle}$, so $\Delta^{\langle\bar{\pi}+1\rangle} < \tau_3 \cdot \Delta^{\langle\underline{\pi}\rangle} \leq \tau_3 \cdot \hat{\Delta}^{\langle\sigma\rangle}$, or $\underline{\pi} \geq \sigma$, implying by (a) $\Delta^{\langle\bar{\pi}+1\rangle} < \tau_3 \cdot \Delta^{\langle\underline{\pi}\rangle} \leq \tau_3 \cdot \hat{\Delta}^{\langle\underline{\pi}\rangle} \leq \tau_3 \cdot \hat{\Delta}^{\langle\sigma\rangle}$.

Finally we show (c). Note, for $\eta \in \Pi^{\langle\nu(\sigma)\rangle}$ there holds $\underline{\eta} \geq \sigma + N_\Pi + 1$, because otherwise we had $\eta \in \Pi^{\langle\sigma+N_\Pi+1\rangle}$, hence $\bar{\eta} + 1 \leq \nu(\sigma)$ contradicting $\eta \in \Pi^{\langle\nu(\sigma)\rangle}$. This implies $\hat{\Delta}^{\langle\nu(\sigma)\rangle} = \Delta^{\langle\gamma\rangle}$ for some $\gamma \geq \sigma + N_\Pi + 1$. Let

$$\pi = \operatorname*{arg\,max}_{\pi\in\Pi^{\langle\gamma\rangle}} \{\bar{\pi}\colon \Delta^{\langle\bar{\pi}+1\rangle} = \Delta^{\langle\gamma\rangle}\}.$$

Then we have $\bar{\pi} \geq \sigma$, because otherwise $\bar{\pi} < \sigma$ was the last update before γ and so there would exist processes $\pi_i \neq \pi$, $i \in \{0, \ldots, N_\Pi\}$, with $\underline{\pi}_i = \sigma + i$. Consequently $\hat{\Delta}^{\langle\nu(\sigma)\rangle} = \Delta^{\langle\bar{\pi}+1\rangle} \overset{\bar{\pi}\geq\sigma}{<} \tau_3 \cdot \hat{\Delta}^{\langle\sigma\rangle}$ by (b). ∎

Note that the sequence of global predicted decrease $(\Delta^{\langle\sigma\rangle})_{\sigma\geq\sigma'}$ does not necessarily converge *monotonically* to zero for any $\sigma' \in \mathbb{N}$. However, replacing (Upd2)

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

by

$$\text{(Upd2}') \qquad \hat{y}^{\langle\bar{\pi}\rangle} = \hat{y}^{\langle\bar{\pi}+1\rangle} \quad \text{and} \quad \Delta^{\langle\bar{\pi}+1\rangle} \le \tau_3 \cdot \Delta^{\langle\bar{\pi}\rangle}$$

for dependency detection, one gets a nonincreasing sequence.

OBSERVATION 12. *Assume $|\Sigma| = \infty$ and there is only a finite number of processes satisfying (Upd1) with $\varepsilon = 0$, and suppose, starting from some $\sigma_1 \in \Sigma$, each subprocess $\pi \in \bar{\Pi}^{\langle\infty\rangle}$ with $\bar{\pi} \ge \sigma_1$ guarantees (Upd2′) instead of (Upd2). Then the sequence $(\Delta^{\langle\sigma\rangle})_{\sigma\ge\sigma_1}$ is nonincreasing and $\lim_{\sigma\ge\sigma_1} \Delta^{\langle\sigma\rangle} = 0$.*

Proof. Given a $\sigma_1 \in \Sigma$ so that all processes π with $\bar{\pi} \ge \sigma_1$ satisfy (Upd2′), in particular, $\Delta^{\langle\bar{\pi}+1\rangle} \le \tau_3 \Delta^{\langle\bar{\pi}\rangle}$. Let $\sigma \ge \sigma_1$ and let π be the process with $\sigma \in \{\underline{\pi}, \bar{\pi}\}$. If $\sigma = \underline{\pi}$ then $\Delta^{\langle\sigma+1\rangle} = \Delta^{\langle\sigma\rangle}$ because no relevant global data are changed. Otherwise $\Delta^{\langle\sigma+1\rangle} = \Delta^{\langle\bar{\pi}+1\rangle} \le \tau_3 \Delta^{\langle\bar{\pi}\rangle} = \tau_3 \Delta^{\langle\sigma\rangle}$. This proves that $(\Delta^{\langle\sigma\rangle})_{\sigma\ge\sigma_1}$ is nonincreasing. In particular, $\Delta^{\langle\bar{\pi}+1\rangle} \le \tau_3 \Delta^{\langle\bar{\pi}\rangle} \le \tau_3 \Delta^{\langle\underline{\pi}\rangle}$, so (Upd2) also holds for all these processes, proving $\lim_{\sigma\ge\sigma_1} \Delta^{\langle\sigma\rangle} = 0$ by Lemma 11. ∎

COROLLARY 13. *If f is bounded from below and $\varepsilon = 0$, the predicted decrease $\Delta^{\langle\sigma\rangle} = f(\hat{y}^{\langle\sigma\rangle}) - \hat{f}_{\bar{w}^{\langle\sigma\rangle}}(\hat{y}^{\langle\sigma\rangle}) + \frac{1}{u}\|\bar{g}^{\langle\sigma\rangle}\|^2$ goes to zero on the subsequence $\Sigma^* \subseteq \Sigma$ with*

$$\Sigma^* := \begin{cases} \Sigma & \text{if } |\Sigma^{dc}| < \infty, \\ \Sigma^{dc} & \text{otherwise,} \end{cases}$$

where Σ^{dc} is the sequence of descent steps according to Lemma 10. In particular, both terms $f(\hat{y}^{\langle\sigma\rangle}) - \hat{f}_{\bar{w}^{\langle\sigma\rangle}}(\hat{y}^{\langle\sigma\rangle})$ and $\|\bar{g}^{\langle\sigma\rangle}\|$ go to zero, for the subsequence Σ^*.

Furthermore, if $|\Sigma^{dc}| < \infty$, then for σ^* the first index of the final global center, i.e., $\hat{y}^{\langle\sigma\rangle} = \hat{y}^{\langle\sigma^*\rangle}$ for all $\sigma \ge \sigma^*$, the sequence of global candidates $\bar{y}^{\langle\sigma\rangle} := \hat{y}^{\langle\sigma\rangle} - \frac{1}{u}\bar{g}^{\langle\sigma\rangle}$, $\sigma \in \Sigma$, satisfies $\lim_{\sigma\in\Sigma} \bar{y}^{\langle\sigma\rangle} = \hat{y}^{\langle\sigma^*\rangle}$.

Proof. Depending on whether an infinite number of descent steps occurs or not, the claim follows either from Lemma 10 or 11 (or Observation 9 for $|\Sigma| < \infty$). The convergence to zero follows from the fact $f(\hat{y}^{\langle\sigma\rangle}) - \hat{f}_{\bar{w}^{\langle\sigma\rangle}}(\hat{y}^{\langle\sigma\rangle}) \ge 0$ and $\|\bar{g}^{\langle\sigma\rangle}\| \ge 0$. Note that $\bar{g}^{\langle\sigma\rangle} \to 0$ implies $\|\hat{y}^{\langle\sigma\rangle} - \bar{y}^{\langle\sigma\rangle}\| \to 0$ implying the last statement. ∎

THEOREM 14. *Suppose the set of optimizers* Arg min f *is nonempty and bounded. If Σ is finite, $\hat{y}^{\langle\max\Sigma\rangle}$ is an optimal solution of (P). Otherwise the sequence $(\hat{y}^{\langle\sigma\rangle})_{\sigma\in\Sigma}$ has at least one cluster point and each cluster point is an optimal solution of (P).*

Proof. The finite case follows from Observation 9, so assume $|\Sigma| = \infty$. Let $f^* := \min\{f(y)\colon y \in \mathbb{R}^M\}$. The boundedness of the level set $\{y\colon f(y) \le f^*\}$ implies the boundedness of all level sets, particularly of the set $L := \{y\colon f(y) \le f(\hat{y}^{\langle 0\rangle})\}$. Because $(f(\hat{y}^{\langle\sigma\rangle}))_\sigma$ is nonincreasing (see Observation 7), we have $\hat{y}^{\langle\sigma\rangle} \in L$ for all $\sigma \in \Sigma$ and therefore the sequence $(\hat{y}^{\langle\sigma\rangle})_\sigma$ is bounded, so it has at least one cluster point. Let y^* be a cluster point of $(\hat{y}^{\langle\sigma\rangle})_\sigma$. So there is a subsequence $(\hat{y}^{\langle\sigma\rangle})_{\sigma\in\Sigma'} \to y^*$ for some Σ' with

$$\Sigma' \subseteq \begin{cases} \Sigma & \text{if } |\Sigma^{\mathrm{dc}}| < \infty, \\ \Sigma^{\mathrm{dc}} & \text{otherwise.} \end{cases}$$

Let $y \in \mathbb{R}^M$ be an arbitrary point. Then by continuity of f we get

$$f(y) \ge \hat{f}_{\bar{w}^{\langle\sigma\rangle}}(y) = \underbrace{f(\hat{y}^{\langle\sigma\rangle})}_{\overset{\Sigma'}{\to} f(y^*)} - \underbrace{\langle \bar{g}^{\langle\sigma\rangle}, y - \hat{y}^{\langle\sigma\rangle}\rangle}_{\overset{\Sigma'}{\to} 0 \text{ by Corollary 13}} \overset{\Sigma'}{\to} f(y^*),$$

proving that $f(y^*)$ is indeed a minimum of f. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Remark 15. In practice $\varepsilon > 0$ and the algorithm is terminated by the traditional stopping criterion $\Delta^{\langle\sigma\rangle} \le \varepsilon(|f(\hat{y}^{\langle\sigma\rangle})| + 1)$ already used and discussed, e.g., in [8]. Via the aggregate minorant it provides the rather weak bound

$$\forall\, y \in \mathbb{R}^m\colon f(y) \ge f(\hat{y}^{\langle\sigma\rangle}) - \varepsilon(|f(\hat{y}^{\langle\sigma\rangle})| + 1) - \|\bar{g}\|(\|y - \hat{y}^{\langle\sigma\rangle}\| - \tfrac{1}{u}\|\bar{g}\|).$$

Informally, within a radius of $\frac{1}{u}\|\bar{g}\|$ around $\hat{y}^{\langle\sigma\rangle}$ no solution is better by more than an ε-fraction. For other possibilities and their discussion see, e.g., [1].

4. A simple subproblem for general convex functions. In this section we describe a simple algorithmic approach for setting up proper subspace problems for the processes so that Assumption 4 can be guaranteed. This approach, however, will in general not be useful in practice, because solving the subspace problem requires the same computational effort as the global problem. It merely serves as a first example for a fully workable asynchronous parallel bundle method. In the next sections we will present other variants that exploit structural properties of the objective function in order to construct more efficient subspace problems.

The most straightforward subspace problem for a process π working on an affine subspace indexed by $J^{(\pi)} \subseteq M$ is

$$(Sub_{\mathrm{s}}^{\pi}) \qquad \begin{aligned} &\text{minimize} && f(y) \\ &\text{subject to} && y \in \mathcal{L}^{(\pi)}, \end{aligned}$$

i.e., we just restrict the optimization to the affine subspace $\mathcal{L}^{(\pi)} = \mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle})$. Given a center $\hat{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$ and an aggregated minorant $\bar{w}^{(\pi)} = (\bar{l}^{(\pi)}, \bar{g}^{(\pi)}) \in W$, the candidate of this subspace problem is given by

$$\bar{y}^{(\pi)} = \arg\min\left\{\hat{f}_{\bar{w}^{(\pi)}}(y) + \frac{u}{2}\|y - \hat{y}^{(\pi)}\|^2\colon y \in \mathcal{L}^{(\pi)}\right\}$$

and so the expected progress can be worked out to be

$$\Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{(\pi)}) = f(\hat{y}^{(\pi)}) - \hat{f}_{\bar{w}^{(\pi)}}(\bar{y}^{(\pi)}) = f(\hat{y}^{(\pi)}) - \bar{l}^{(\pi)} - \langle\bar{g}^{(\pi)}, \hat{y}^{(\pi)}\rangle + \frac{1}{u}\|\bar{g}^{(\pi)}_{J^{(\pi)}}\|^2.$$

In particular, for $\hat{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$ and each $\bar{w} \in W$ there holds

$$\Delta(\bar{w}, \hat{y}^{(\pi)}) = \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}, \hat{y}^{(\pi)}) + \frac{1}{u}\|\bar{g}^{(\pi)}_{M\setminus J^{(\pi)}}\|^2. \tag{4.1}$$

A bundle method solving (Sub_{s}^{π}) will eventually drive $\Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{(\pi)})$ to zero, but this only helps to reduce the global progress $\Delta(\bar{w}^{(\pi)}, \hat{y}^{(\pi)})$ if the last term does not grow too much. Thus, we will use this last term to detect strong subspace dependencies, which we store in an adaptive *dependency digraph* $G = (M, E)$ on node set M with directed edges $E \subseteq \{(i, j)\colon i, j \in M, i \ne j\}$. The meaning of an edge $(j, j') \in E$ is as follows. If a process tries to select j for inclusion into $J^{(\pi)}$ in A2.1, the subspace is presumed to depend strongly on the variable j', so it must include j' as well. Thus, a dependency digraph defines a strongly dependent subspace function F_D^{gra}. In particular, let

$$\mathcal{S}^{\mathrm{gra}} := \{E \subseteq M^2\colon \forall\, j \in M, (j, j) \notin E\}$$

and with $S = E \in \mathcal{S}^{\mathrm{gra}}$

$$F_D^{\mathrm{gra}}(S, J) := J \cup \{j' \in M\colon \exists\, j \in J, (j, j') \in S\}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

It is easy to check that $(\mathcal{S}^{\mathrm{gra}}, \preceq)$, where $S \preceq S'$ if and only if $S \subseteq S'$, is a valid set of states with the complete digraph yielding the unique maximal element $\bar{S}^{\mathrm{gra}} \in \mathcal{S}^{\mathrm{gra}}$, and that F_D^{gra} satisfies the requirements for dependent subspace functions.

We will use the dependency graph based function F_D^{gra} together with the states $\mathcal{S}^{\mathrm{gra}}$ to track these dependencies. We do not use any weak dependencies, i.e., if a subspace J is selected, no other variables have to be kept fixed. So the following states and functions are used:

$$\mathcal{S}^{\mathrm{s}} := \mathcal{S}^{\mathrm{gra}}, \quad F_D^{\mathrm{s}}(S, J) := F_D^{\mathrm{gra}}(S, J), \quad F_d^{\mathrm{s}}(S, J) := J, \quad \Delta_{\mathrm{s}}(S, J, \bar{w}, \hat{y}) := \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}, \hat{y}).$$

Obviously F_d^{s} satisfies the conditions for weakly dependent subspace functions.

Algorithm 2 determines $J^{(\pi)}$ w.r.t. these dependency functions in the subspace selection step and so condition (Sel1) guarantees

$$\Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{\langle \underline{\pi} \rangle}, \hat{y}^{\langle \underline{\pi} \rangle}) \geq \tau_1 \Delta^{\langle \underline{\pi} \rangle}.$$

The subspace problem is solved by a bundle method with center $\hat{y}^{(\pi)} = \hat{y}^{\langle \underline{\pi} \rangle}$, descent parameter $\varrho_2 \in (0, 1)$, and a termination parameter $\varrho_1 \in (0, 1)$ until the aggregate minorant $\bar{w}^{(\pi)} \in W$ and the candidate $\bar{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$ either satisfy

$$\text{(Upd1}_{\mathrm{s}}\text{)} \qquad \begin{aligned} \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{(\pi)}) &\geq \varrho_1 \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{\langle \underline{\pi} \rangle}, \hat{y}^{\langle \underline{\pi} \rangle}) \\ f(\hat{y}^{\langle \underline{\pi} \rangle}) - f(\bar{y}^{(\pi)}) &\geq \varrho_2 \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{(\pi)}), \end{aligned}$$

or

$$\text{(Upd2}_{\mathrm{s}}\text{)} \qquad \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{(\pi)}) < \varrho_1 \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{\langle \underline{\pi} \rangle}, \hat{y}^{\langle \underline{\pi} \rangle}).$$

By Theorem 1 the bundle process ensures that one of the two criteria is met in finite time. Let $\bar{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$ be the last candidate point of the bundle method. With this the process runs Algorithm 3 to update the global data to index $\sigma = \bar{\pi} + 1$ as described next.

ALGORITHM 3. UPDATESUBSPACE$^{\mathrm{s}}$.

Input : π, final candidate $\bar{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$, aggregate minorant $\bar{w}^{(\pi)} \in W$
Changes : global data at $\bar{\pi} + 1$
if π *stopped because of (Upd1$_s$)* **then**
  $\bar{w}^{\langle \bar{\pi}+1 \rangle} \leftarrow \bar{w}^{(\pi)}$
  $\hat{y}^{\langle \bar{\pi}+1 \rangle} \in \operatorname{Arg\,min} \{ f(y) \colon y \in \{ \hat{y}^{\langle \bar{\pi} \rangle}, \bar{y}^{(\pi)}, \hat{y}^{\langle \bar{\pi} \rangle} + (\bar{y}^{(\pi)} - \hat{y}^{\langle \underline{\pi} \rangle}) \} \}$
  $S^{\langle \bar{\pi}+1 \rangle} \leftarrow S^{\langle \bar{\pi} \rangle}$
else // π `stopped because of (Upd2`$_{\mathrm{s}}$`)`
  $\bar{w}^{\langle \bar{\pi}+1 \rangle} \leftarrow \bar{w}^{(\pi)}$, $\hat{y}^{\langle \bar{\pi}+1 \rangle} \leftarrow \hat{y}^{\langle \bar{\pi} \rangle}$
  if *(Dep$_s$) holds* **then** $S^{\langle \bar{\pi}+1 \rangle} \leftarrow S^{\langle \bar{\pi} \rangle} \cup \{(j, j')\}$ with (j, j') be defined in (4.2)
  else $S^{\langle \bar{\pi}+1 \rangle} \leftarrow S^{\langle \bar{\pi} \rangle}$

We investigate the case that the subprocess stopped because of (Upd2$_{\mathrm{s}}$). In view of (4.1), condition (Upd2) might not hold if $\frac{1}{u} \|\bar{g}^{\langle \bar{\pi}+1 \rangle}_{M \setminus J^{(\pi)}}\|^2 > \frac{1}{u} \|\bar{g}^{\langle \underline{\pi} \rangle}_{M \setminus J^{(\pi)}}\|^2$. If this happens, the subprocess presumes to have detected a new strong dependency between $J^{(\pi)}$ and $M \setminus J^{(\pi)}$. In fact, the process checks if the following condition holds for some $\varrho_3 \in (0, 1 - \varrho_1)$:

$$\text{(Dep}_{\mathrm{s}}\text{)} \qquad \frac{1}{u} \|\bar{g}^{\langle \bar{\pi}+1 \rangle}_{M \setminus J^{(\pi)}}\|^2 - \frac{1}{u} \|\bar{g}^{\langle \underline{\pi} \rangle}_{M \setminus J^{(\pi)}}\|^2 > \varrho_3 \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{\langle \underline{\pi} \rangle}, \hat{y}^{\langle \underline{\pi} \rangle}).$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

If this condition holds then the improvement in global progress is small compared to the gain generated by the subspace optimization. In this case we increase the dependency graph by choosing an edge

$$(4.2)\quad (j,j') \in \operatorname{Arg\,max}\left\{(\bar{g}_{\hat{\jmath}}^{\langle\bar{\pi}+1\rangle})^2 - (\bar{g}_{\hat{\jmath}}^{\langle\underline{\pi}\rangle})^2 \colon (\tilde{\jmath},\hat{\jmath}) \in (J^{(\pi)} \times (M \setminus J^{(\pi)})) \setminus S^{\langle\bar{\pi}\rangle}\right\},$$

and set $S^{\langle\bar{\pi}+1\rangle} := S^{\langle\bar{\pi}\rangle} \cup \{(j,j')\}$. Note that this implies $S^{\langle\bar{\pi}+1\rangle} \succ S^{\langle\bar{\pi}\rangle}$ (see below).

It remains to verify the validity of Assumption 4. Conditions (Inv1) and (Inv2) are clearly fulfilled, so we check that the subproblem satisfies (Upd1) and (Upd2).

LEMMA 16. *For $\tau_2 \in (0, \varrho_2\varrho_1\tau_1]$ and $\tau_3 \in [1-\tau_1(1-\varrho_1-\varrho_3),1)$ there hold*

(i) *if a subprocess π terminates with (Upd1$_s$) then it satisfies (Upd1);*

(ii) *if there is only a finite number of processes satisfying (Upd1), then there is a $\sigma_1 \in \Sigma$ so that each process π with $\bar{\pi} \geq \sigma_1$ satisfies (Upd2).*

Proof. We start with (i). By (Upd1$_s$) there holds with the subspace selection condition (Sel1)

$$\begin{aligned} f(\hat{y}^{\langle\underline{\pi}\rangle}) - f(\bar{y}^{(\pi)}) &\geq \varrho_2\Delta_s^{(\pi)}(\bar{w}^{(\pi)},\hat{y}^{\langle\underline{\pi}\rangle}) \geq \varrho_2\varrho_1\Delta_s^{(\pi)}(\bar{w}^{\langle\underline{\pi}\rangle},\hat{y}^{\langle\underline{\pi}\rangle}) \\ &\geq \varrho_2\varrho_1\tau_1\Delta^{\langle\underline{\pi}\rangle} \geq \tau_2\Delta^{\langle\underline{\pi}\rangle}. \end{aligned}$$

The choice of $\hat{y}^{\langle\bar{\pi}+1\rangle}$ in this case implies $f(\hat{y}^{\langle\bar{\pi}+1\rangle}) \leq \min\{f(\hat{y}^{\langle\bar{\pi}\rangle}), f(\bar{y}^{(\pi)})\}$, so (Upd1) holds.

Now consider (ii). If Σ is finite, $\sigma_1 = \max\Sigma$ is a valid choice. If $|\Sigma| = \infty$ and there are only finitely many descent steps, there is a $\sigma_d \in \Sigma$ with $\hat{y}^{\langle\sigma\rangle} = \hat{y}^{\langle\sigma_d\rangle}$ for all $\sigma \geq \sigma_d$. Let σ_0 be the index defined in Observation 7. We show that $\sigma_1 = 1 + \max\{\sigma_0\} \cup \{\bar{\pi} : \pi \in \Pi^{\langle\sigma_d\rangle}\}$ satisfies the requirements. By (i) and the choice of σ_1 a process π with $\bar{\pi} \geq \sigma_1$ satisfies $\hat{y}^{\langle\sigma_1\rangle} = \hat{y}^{\langle\underline{\pi}\rangle} = \hat{y}^{\langle\bar{\pi}+1\rangle}$ and stops because of (Upd2$_s$) without satisfying (Dep$_s$). Then with (4.1) it follows

$$\begin{aligned} \Delta^{\langle\underline{\pi}\rangle} - \Delta^{\langle\bar{\pi}+1\rangle} &= \left(\Delta_s^{(\pi)}(\bar{w}^{\langle\underline{\pi}\rangle},\hat{y}^{\langle\underline{\pi}\rangle}) - \Delta_s^{(\pi)}(\bar{w}^{\langle\bar{\pi}+1\rangle},\hat{y}^{\langle\bar{\pi}+1\rangle})\right) \\ &\quad + \frac{1}{u}\left(\|\bar{g}_{M\setminus J^{(\pi)}}^{\langle\underline{\pi}\rangle}\|^2 - \|\bar{g}_{M\setminus J^{(\pi)}}^{\langle\bar{\pi}+1\rangle}\|^2\right) \\ &\geq (1-\varrho_1-\varrho_3)\Delta_s^{(\pi)}(\bar{w}^{\langle\underline{\pi}\rangle},\hat{y}^{\langle\underline{\pi}\rangle}) \geq \tau_1(1-\varrho_1-\varrho_3)\Delta^{\langle\underline{\pi}\rangle}, \end{aligned}$$

and consequently

$$\Delta^{\langle\bar{\pi}+1\rangle} \leq (1-\tau_1(1-\varrho_1-\varrho_3))\Delta^{\langle\underline{\pi}\rangle} \leq \tau_3\Delta^{\langle\underline{\pi}\rangle}. \qquad \square$$

5. Partially separable functions. The main drawback of the simple subspace problem presented in the previous section is that solving a subspace problem requires the same computational effort as optimizing over the whole space (only the steps are restricted to the subspace, but the function evaluations remain the same). In particular, each process π working on a subspace $J^{(\pi)}$ still has to call the same oracle of f at its candidates. From a practical point of view one would only expect a gain from the parallel algorithm if the computations and evaluations associated with the subspace problems are significantly easier than the computations for the whole problem. This seems only to be possible, if the objective function f has a special structure.

In this section we develop an alternative subspace problem that exploits the structure of a "partially separable" objective function. In particular, given a finite set R

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

and subspaces $J_r \subset M$, $r \in R$, we assume that f has the form

$$f(y) := \sum_{r \in R} f_r(y),$$

where $f_r \colon \mathbb{R}^M \to \mathbb{R}$ are convex functions with the property

$$f_r(y) = f_r(y') \text{ for all } y, y' \text{ with } y_{J_r} = y'_{J_r},\ r \in R.$$

In other words, a function f_r only depends on the variables y_j with $j \in J_r$, and there holds

$$\forall\, y \in \mathbb{R}^M, \forall\, g \in \partial f_r(y) \colon g_{M \setminus J_r} = 0.$$

The subdifferential of f at $y \in \mathbb{R}^M$ has the form $\partial f(y) = \sum_{r \in R} \partial f_r(y)$. With

$$W_r = \operatorname{conv}\left\{(l_r, g_r) \colon l_r = f_r(y) - \langle g_r, y\rangle, g_r \in \partial f_r(y), y \in \mathbb{R}^M\right\},$$

each minorant is defined by $w = (w_r)_{r \in R} = (l_r, g_r)_{r \in R} \in W = \bigtimes_{r \in R} W_r$ with

$$\hat f_w(y) = \sum_{r \in R} \hat f_{w_r, r}(y) = l + \langle g, y\rangle,$$
$$\hat f_{w_r, r}(y) = l_r + \langle g_r, y\rangle,$$

being the affine minorant of f (in slight abuse of notation we will regard $w = (l, g) \in \mathbb{R} \times \mathbb{R}^M$ also as an element of W). An immediate consequence is that the subspace problem associated with a subspace $J^{(\pi)} \subseteq M$ depends only on the functions f_r, $r \in R$, with $J_r \cap J^{(\pi)} \neq \emptyset$. We define for a subspace $J \subseteq M$

$$R_J := \{r \in R \colon J_r \cap J \neq \emptyset\}, \qquad \bar J := \bigcup_{r \in R_J} J_r \setminus J,$$

and for a process π selecting coordinates $J^{(\pi)}$

$$R^{(\pi)} := R_{J^{(\pi)}}, \qquad \bar J^{(\pi)} := \bigcup_{r \in R^{(\pi)}} J_r \setminus J^{(\pi)}, \tag{5.1}$$

then for a given center $\hat y \in \mathbb{R}^M$ there holds

$$\min\left\{f(y) \colon y \in \mathcal{L}(J^{(\pi)}, \hat y)\right\} = \min\left\{f_{R^{(\pi)}}(y) \colon y \in \mathcal{L}(J^{(\pi)}, \hat y)\right\} + f_{R \setminus R^{(\pi)}}(\hat y).$$

Hence, in order to optimize on a subspace $J^{(\pi)}$, a process only has to evaluate the functions f_r with $r \in R^{(\pi)}$, which is significantly easier than the evaluation of f if $|R^{(\pi)}| \ll |R|$. These considerations motivate the following choice of a subspace problem:

$$(Sub^{\pi}_{\mathrm{p}}) \qquad \begin{aligned} &\text{minimize} && f^{(\pi)}(y) := \sum_{r \in R^{(\pi)}} f_r(y) \\ &\text{subject to} && y \in \mathcal{L}^{(\pi)} \end{aligned}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

with $\mathcal{L}^{(\pi)} = \mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\pi\rangle})$. The set of minorants is given by $W^{(\pi)} = \bigtimes_{r\in R^{(\pi)}} W_r$. A center point $\hat{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$ and an aggregated minorant $\bar{w}^{(\pi)} = (\bar{l}^{(\pi)}, \bar{g}^{(\pi)}) \in W^{(\pi)}$ define the candidate

$$\bar{y}^{(\pi)} = \arg\min \left\{ \hat{f}^{(\pi)}_{\bar{w}^{(\pi)}}(y) + \frac{u}{2}\|y - \hat{y}^{(\pi)}\|^2 \colon y \in \mathcal{L}^{(\pi)} \right\},$$

where

$$\hat{f}^{(\pi)}_{\bar{w}^{(\pi)}}(y) = \sum_{r\in R^{(\pi)}} \hat{f}_{\bar{w}^{(\pi)}_r, r}(y).$$

The expected subspace progress in this case for $\bar{w}^{(\pi)} \in W^{(\pi)}$ and $\hat{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$ can be worked out to be

$$\begin{aligned}\Delta^{(\pi)}_{\mathrm{p}}(\bar{w}^{(\pi)}, \hat{y}^{(\pi)}) &= f^{(\pi)}(\hat{y}^{(\pi)}) - \hat{f}^{(\pi)}_{\bar{w}^{(\pi)}}(\bar{y}^{(\pi)}) \\ &= f^{(\pi)}(\hat{y}^{(\pi)}) - \sum_{r\in R^{(\pi)}} \left(\bar{l}^{(\pi)}_r + \langle \bar{g}^{(\pi)}_r, \hat{y}^{(\pi)} \rangle \right) + \frac{1}{u}\|\bar{g}^{(\pi)}_{J^{(\pi)}}\|^2.\end{aligned}$$

In particular, for $\bar{w} = (\bar{l}, \bar{g}) = (\bar{w}_r)_{r\in R} = (\bar{l}_r, \bar{g}_r)_{r\in R} \in W$ there holds

$$\Delta(\bar{w}, \hat{y}) = \Delta^{(\pi)}_{\mathrm{p}}(\bar{w}_{R^{(\pi)}}, \hat{y}) + \bar{\Delta}^{(\pi)}_{\mathrm{p}}(\bar{w}_{R\setminus R^{(\pi)}}, \hat{y}) + \frac{1}{u}\|\bar{g}_{\bar{J}^{(\pi)}}\|^2, \tag{5.2}$$

where

$$\bar{\Delta}^{(\pi)}_{\mathrm{p}}(\bar{w}_{R\setminus R^{(\pi)}}, \hat{y}) = \sum_{r\in R\setminus R^{(\pi)}} \left(f_r(\hat{y}) - \hat{f}_{\bar{w}_r, r}(\hat{y}) \right) + \frac{1}{u}\|\bar{g}_{M\setminus(J^{(\pi)}\cup\bar{J}^{(\pi)})}\|^2.$$

In fact, this decomposes the global expected progress $\Delta(\bar{w}, \hat{y})$ into three parts, the first only depending on $\hat{y}_{J^{(\pi)}\cup\bar{J}^{(\pi)}}$ and $\bar{w}_{R^{(\pi)}}$, the second *not* depending on $\hat{y}_{J^{(\pi)}}$ and $\bar{w}_{R^{(\pi)}}$, and only the last term depending on complete $\bar{w} = (\bar{l}, \bar{g}) \in W$.

Observation 17. *Let $y, y' \in \mathbb{R}^M$ and $w \in W$.*

(i) *If $y_{J^{(\pi)}\cup\bar{J}^{(\pi)}} = y'_{J^{(\pi)}\cup\bar{J}^{(\pi)}}$ then $f^{(\pi)}(y) = f^{(\pi)}(y')$ as well as $\Delta^{(\pi)}_p(w_{R^{(\pi)}}, y) = \Delta^{(\pi)}_p(w_{R^{(\pi)}}, y')$.*

(ii) *If $y_{M\setminus J^{(\pi)}} = y'_{M\setminus J^{(\pi)}}$ then $f_r(y) = f_r(y')$, $r \in R\setminus R^{(\pi)}$, and $\bar{\Delta}^{(\pi)}_p(w_{M\setminus R^{(\pi)}}, y) = \bar{\Delta}^{(\pi)}_p(w_{M\setminus R^{(\pi)}}, y')$.*

Proof. Direct computation while exploiting that for each $r \in R\setminus R^{(\pi)}$ the function f_r does not depend on $y_{J^{(\pi)}}$ by definition (5.1) of $R^{(\pi)}$. ∎

If, on the one hand, $y, y' \in \mathcal{L}^{(\pi)}$ then $y_{M\setminus J^{(\pi)}} = y'_{M\setminus J^{(\pi)}}$, i.e. the subspace $\mathcal{L}^{(\pi)}$ does not influence the term $\bar{\Delta}^{(\pi)}_{\mathrm{p}}$. On the other hand, the first term $\Delta^{(\pi)}_{\mathrm{p}}$ only depends on the variables $J^{(\pi)} \cup \bar{J}^{(\pi)}$, hence, progress made by the subspace problem (Sub^{π}_{p}) is directly reflected in the term $\Delta^{(\pi)}_{\mathrm{p}}$ if the components $\bar{J}^{(\pi)}$ are kept fix. This motivates the following weakly dependent subspace function (it satisfies conditions (3.1)–(3.4))

$$F^{\mathrm{p}}_d(S, J) = J \cup \bar{J}.$$

As in section 4, we use the dependency digraph based strongly dependent subspace function and $\Delta^{(\pi)}_{\mathrm{p}}$ for the subspace progress function

$$\mathcal{S}^{\mathrm{p}} := \mathcal{S}^{\mathrm{gra}}, \qquad F^{\mathrm{p}}_D(S, J) := F^{\mathrm{gra}}_D(S, J),$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

and

$$\Delta^{\mathrm{p}}(S, J, \bar{w}, \hat{y}) := \sum_{r \in R_J} \left(f_r(\hat{y}) - \hat{f}_{\bar{w}_r, r}(\hat{y}) \right) + \frac{1}{u} \|\bar{g}_J\|^2.$$

Employed within a successful subspace selection step of Algorithm 2 for π this yields

$$\Delta^{\mathrm{p}}(S^{\langle \underline{\pi} \rangle}, J^{(\pi)}, \bar{w}^{\langle \underline{\pi} \rangle}, \hat{y}^{\langle \underline{\pi} \rangle}) = \Delta^{(\pi)}_{\mathrm{p}}(\bar{w}^{\langle \underline{\pi} \rangle}_{R^{(\pi)}}, \hat{y}^{\langle \underline{\pi} \rangle}) \geq \tau_1 \Delta^{\langle \underline{\pi} \rangle}$$

and ensures a "disjointness" of parallel subspace problems.

OBSERVATION 18. *Let $\sigma \in \Sigma$ and $\pi, \eta \in \Pi^{\langle \sigma \rangle}$, $\pi \neq \eta$, be two parallel processes. Then $R^{(\pi)} \cap R^{(\eta)} = \emptyset$.*

Proof. Assume there exists an $r \in R^{(\pi)} \cap R^{(\eta)}$. By definition there must be a $j \in J^{(\pi)} \cap J_r$. Furthermore $r \in R^{(\eta)}$ implies $J_r \subseteq J^{(\eta)} \cup \bar{J}^{(\eta)}$, hence $j \in J^{(\eta)} \cup \bar{J}^{(\eta)} = F^{\mathrm{p}}_d(S, J^{(\eta)})$. This is a contradiction to (3.5), which holds by Observation 8. ∎

For subspace optimization and subspace update we follow the same steps as in section 4. The subspace problem is solved by a bundle method with descent parameter $\varrho_2 \in (0,1)$ and termination parameter $\varrho_1 \in (0,1)$ until the aggregate minorant $\bar{w}^{(\pi)} \in W^{(\pi)}$ and the candidate $\bar{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$ satisfy

$$\begin{aligned} \Delta^{(\pi)}_{\mathrm{p}}(\bar{w}^{(\pi)}, \hat{y}^{\langle \underline{\pi} \rangle}) &\geq \varrho_1 \Delta^{(\pi)}_{\mathrm{p}}(\bar{w}^{\langle \underline{\pi} \rangle}_{R^{(\pi)}}, \hat{y}^{\langle \underline{\pi} \rangle}) \\ f^{(\pi)}(\hat{y}^{\langle \underline{\pi} \rangle}) - f^{(\pi)}(\bar{y}^{(\pi)}) &\geq \varrho_2 \Delta^{(\pi)}_{\mathrm{p}}(\bar{w}^{(\pi)}, \hat{y}^{\langle \underline{\pi} \rangle}), \end{aligned} \tag{Upd1$_{\mathrm{p}}$}$$

or

$$\Delta^{(\pi)}_{\mathrm{p}}(\bar{w}^{(\pi)}, \hat{y}^{\langle \underline{\pi} \rangle}) < \varrho_1 \Delta^{(\pi)}_{\mathrm{p}}(\bar{w}^{\langle \underline{\pi} \rangle}_{R^{(\pi)}}, \hat{y}^{\langle \underline{\pi} \rangle}). \tag{Upd2$_{\mathrm{p}}$}$$

Observation 17 and (5.2) indicate that the critical term, which may cause an increase in the global progress even if $\Delta^{(\pi)}_{\mathrm{p}}$ is reduced, is the last of (5.2). Therefore, the subprocess will use the following dependency test for some $\varrho_3 \in (0, 1 - \varrho_1)$:

$$\frac{1}{u} \left\| \bar{g}^{\langle \bar{\pi}+1 \rangle}_{\bar{J}^{(\pi)}} \right\|^2 - \frac{1}{u} \left\| \bar{g}^{\langle \bar{\pi} \rangle}_{\bar{J}^{(\pi)}} \right\|^2 > \varrho_3 \Delta^{(\pi)}_{\mathrm{p}} \left(\bar{w}^{\langle \underline{\pi} \rangle}_{R^{(\pi)}}, \hat{y}^{\langle \underline{\pi} \rangle} \right). \tag{Dep$_{\mathrm{p}}$}$$

Note that in contrast to (Dep$_{\mathrm{s}}$) we compare the new values at $\bar{\pi}+1$ with the values at $\bar{\pi}$ and not at $\underline{\pi}$. This will allow for stronger convergence results. If condition (Dep$_{\mathrm{p}}$) is fulfilled, then the dependency graph is enlarged by adding an edge

$$(j, j') \in \operatorname{Arg\,max} \left\{ (\bar{g}^{\langle \bar{\pi}+1 \rangle}_{\hat{\jmath}})^2 - (\bar{g}^{\langle \bar{\pi} \rangle}_{\hat{\jmath}})^2 \colon (\tilde{\jmath}, \hat{\jmath}) \in (J^{(\pi)} \times \bar{J}^{(\pi)}) \setminus S^{\langle \bar{\pi} \rangle} \right\} \tag{5.3}$$

and $S^{\langle \bar{\pi}+1 \rangle} := S^{\langle \bar{\pi} \rangle} \cup \{(j, j')\}$. Putting all together, the update algorithm for the partially separable subspace problems is displayed in Algorithm 4.

In order to prove Assumption 4 we first observe that the center does not change on $J^{(\pi)} \cup \bar{J}^{(\pi)}$ as long as π runs and changes in $J^{(\pi)}$ at most when π finishes.

COROLLARY 19. *Let $\pi \in \bar{\Pi}^{\langle \infty \rangle}$, then $\hat{y}^{\langle \underline{\pi} \rangle}_j = \hat{y}^{\langle \bar{\pi} \rangle}_j$ for all $j \in J^{(\pi)} \cup \bar{J}^{(\pi)}$ and $\hat{y}^{\langle \bar{\pi} \rangle}_j = \hat{y}^{\langle \bar{\pi}+1 \rangle}_j$ for all $j \notin J^{(\pi)}$.*

Proof. The first statement relies on the subspace selection in conjunction with the update of the center in Algorithm 4. Assume, for contradiction, $\hat{y}^{\langle \underline{\pi} \rangle}_j \neq \hat{y}^{\langle \bar{\pi} \rangle}_j$ for some $j \in J^{(\pi)} \cup \bar{J}^{(\pi)} = F^{\langle \underline{\pi} \rangle}_d(J^{(\pi)})$. Then there must be a process η with $\underline{\pi} < \bar{\eta} < \bar{\pi}$ and $\hat{y}^{\langle \bar{\eta} \rangle}_j \neq \hat{y}^{\langle \bar{\eta}+1 \rangle}_j$. However, in all cases $\hat{y}^{\langle \bar{\eta} \rangle}_{j'} = \hat{y}^{\langle \bar{\eta}+1 \rangle}_{j'}$ for $j' \notin J^{(\eta)}$, hence $j \in J^{(\eta)}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ALGORITHM 4. UPDATESUBSPACE$^{\mathrm{P}}$.

Input : π, final candidate $\bar{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$, aggregate minorant $\bar{w}^{(\pi)} \in W^{(\pi)}$
Changes: global data at $\bar{\pi}+1$
if π *stopped because of (Upd1$_p$)* **then**
  $(\bar{w}^{\langle\bar{\pi}+1\rangle}_{R^{(\pi)}}, \bar{w}^{\langle\bar{\pi}+1\rangle}_{R\setminus R^{(\pi)}}) \leftarrow (\bar{w}^{(\pi)}, \bar{w}^{\langle\bar{\pi}\rangle}_{R\setminus R^{(\pi)}})$
  $\hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}^{\langle\bar{\pi}\rangle} + (\bar{y}^{(\pi)} - \hat{y}^{\langle\underline{\pi}\rangle})$
  $S^{\langle\bar{\pi}+1\rangle} \leftarrow S^{\langle\bar{\pi}\rangle}$
else // π `stopped because of (Upd2`$_{\mathrm{p}}$`)`
  $(\bar{w}^{\langle\bar{\pi}+1\rangle}_{R^{(\pi)}}, \bar{w}^{\langle\bar{\pi}+1\rangle}_{R\setminus R^{(\pi)}}) \leftarrow (\bar{w}^{(\pi)}, \bar{w}^{\langle\bar{\pi}\rangle}_{R\setminus R^{(\pi)}})$
  $\hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}^{\langle\bar{\pi}\rangle}$
  if *(Dep$_p$) holds* **then** $S^{\langle\bar{\pi}+1\rangle} \leftarrow S^{\langle\bar{\pi}\rangle} \cup \{(j,j')\}$ with (j,j') defined in (5.3)
  else $S^{\langle\bar{\pi}+1\rangle} \leftarrow S^{\langle\bar{\pi}\rangle}$

This implies $j \in J^{(\eta)} \cap F_d^{\langle\underline{\pi}\rangle}(J^{(\pi)}) \neq \emptyset$, contradicting (3.5) by Observation 8 because $\pi, \eta \in \Pi^{\langle\bar{\eta}\rangle}$.

For the second statement it suffices to observe that $\hat{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$, so for $j \notin J^{(\pi)}$ it is $\hat{y}_j^{(\pi)} = \hat{y}_j^{\langle\underline{\pi}\rangle}$ and $\hat{y}_j^{\langle\bar{\pi}+1\rangle} = \hat{y}_j^{\langle\bar{\pi}\rangle} + (\hat{y}_j^{(\pi)} - \hat{y}_j^{\langle\underline{\pi}\rangle}) = \hat{y}_j^{\langle\bar{\pi}\rangle}$. ∎

Condition (Inv1) is clearly fulfilled, so it remains to check (Inv2) and (Upd1)–(Upd2) in order to ensure validity of Algorithm 4.

LEMMA 20. *For $\tau_2 \in (0, \varrho_2\varrho_1\tau_1]$ and $\tau_3 \in [1-\tau_1(1-\varrho_1-\varrho_3), 1)$ there hold*
(i) *if a subprocess π terminates with (Upd1$_p$) then it satisfies (Upd1);*
(ii) *if there is only a finite number of processes satisfying (Upd1), then there is a $\sigma_1 \in \Sigma$ so that each process π with $\bar{\pi} \geq \sigma_1$ satisfies (Upd2) and (Upd2′);*
(iii) *all processes satisfy (Inv2).*

Proof. By Corollary 19 we know $\hat{y}^{\langle\bar{\pi}\rangle}_{J^{(\pi)}\cup\bar{J}^{(\pi)}} = \hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}\cup\bar{J}^{(\pi)}}$ and $\hat{y}^{\langle\bar{\pi}+1\rangle}_{M\setminus J^{(\pi)}} = \hat{y}^{\langle\bar{\pi}\rangle}_{M\setminus J^{(\pi)}}$. First assume that π stops with (Upd1$_{\mathrm{p}}$). Observation 17 implies $f^{(\pi)}(\hat{y}^{\langle\bar{\pi}\rangle}) = f^{(\pi)}(\hat{y}^{\langle\underline{\pi}\rangle})$ and $f_r(\hat{y}^{\langle\bar{\pi}\rangle}) = f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})$ for all $r \in R \setminus R^{(\pi)}$. Consequently,

$$\begin{aligned}
f(\hat{y}^{\langle\bar{\pi}\rangle}) - f(\hat{y}^{\langle\bar{\pi}+1\rangle}) &= \sum_{r\in R^{(\pi)}} \Big(f_r(\hat{y}^{\langle\bar{\pi}\rangle}) - f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})\Big) \\
&\quad + \sum_{r\in R\setminus R^{(\pi)}} \Big(f_r(\hat{y}^{\langle\bar{\pi}\rangle}) - f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})\Big) \\
&= \sum_{r\in R^{(\pi)}} \Big(f_r(\hat{y}^{\langle\underline{\pi}\rangle}) - f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})\Big) \\
&= f^{(\pi)}(\hat{y}^{\langle\underline{\pi}\rangle}) - f^{(\pi)}(\hat{y}^{\langle\bar{\pi}+1\rangle}) = f^{(\pi)}(\hat{y}^{\langle\underline{\pi}\rangle}) - f^{(\pi)}(\bar{y}^{(\pi)}) \\
&\overset{(\mathrm{Upd1_p})}{\geq} \varrho_2\Delta_{\mathrm{p}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle}) \overset{(\mathrm{Upd1_p})}{\geq} \varrho_2\varrho_1\Delta_{\mathrm{p}}^{(\pi)}(\bar{w}^{\langle\underline{\pi}\rangle}_{R^{(\pi)}}, \hat{y}^{\langle\underline{\pi}\rangle}) \\
&\overset{(\mathrm{Sel1})}{\geq} \varrho_2\varrho_1\tau_1\Delta^{\langle\underline{\pi}\rangle} \geq \tau_2\Delta^{\langle\underline{\pi}\rangle} \geq 0.
\end{aligned}$$

This proves (Inv2) as well as (Upd1) in this case.

Now assume that $|\Sigma| = \infty$ (otherwise $\sigma_1 = \max\Sigma$ is a valid choice) and there is only a finite number of processes satisfying (Upd1). Then by Observation 7 there is a $\sigma_0 \in \Sigma$ so that $S^{\langle\sigma\rangle} = S^{\langle\sigma_0\rangle}$ for all $\sigma \geq \sigma_0$. Putting both together, there is a $\sigma' \geq \sigma_0$ so that $\hat{y}^{\langle\sigma\rangle} = \hat{y}^{\langle\sigma'\rangle}$ for all $\sigma \geq \sigma'$. Set $\sigma'' := \max\{\bar{\pi}+1 \colon \pi \in \Pi^{\langle\sigma'\rangle}\}$. Then for each π with $\bar{\pi} \geq \sigma''$ we have $\underline{\pi} \geq \sigma'$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We use the following short notation for any process π and $\sigma \in \Sigma$:

$$\Delta_\pi^{\langle\sigma\rangle} := \Delta_{\mathrm{p}}^{(\pi)}(\bar{w}_{R^{(\pi)}}^{\langle\sigma\rangle}, \hat{y}^{\langle\sigma\rangle}), \qquad \bar{\Delta}_\pi^{\langle\sigma\rangle} := \bar{\Delta}_{\mathrm{p}}^{(\pi)}(\bar{w}_{R\setminus R^{(\pi)}}^{\langle\sigma\rangle}, \hat{y}^{\langle\sigma\rangle}).$$

Let π be an arbitrary process with $\bar{\pi} \geq \sigma''$. Then by ($\mathrm{Upd2_p}$) process π satisfies $\Delta_\pi^{\langle\bar{\pi}+1\rangle} < \varrho_1 \Delta_\pi^{\langle\underline{\pi}\rangle}$ and ($\mathrm{Dep_p}$) does not hold; hence, by (5.2) and Observation 17,

$$\begin{aligned}
\Delta^{\langle\bar{\pi}\rangle} - \Delta^{\langle\bar{\pi}+1\rangle} &= (\Delta_\pi^{\langle\bar{\pi}\rangle} - \Delta_\pi^{\langle\bar{\pi}+1\rangle}) + (\bar{\Delta}_\pi^{\langle\bar{\pi}\rangle} - \bar{\Delta}_\pi^{\langle\bar{\pi}+1\rangle}) + \frac{1}{u}\left(\|\bar{g}_{\bar{J}^{(\pi)}}^{\langle\bar{\pi}\rangle}\|^2 - \|\bar{g}_{\bar{J}^{(\pi)}}^{\langle\bar{\pi}+1\rangle}\|^2\right) \\
&= (\Delta_\pi^{\langle\underline{\pi}\rangle} - \Delta_\pi^{\langle\bar{\pi}+1\rangle}) + \frac{1}{u}\left(\|\bar{g}_{\bar{J}^{(\pi)}}^{\langle\bar{\pi}\rangle}\|^2 - \|\bar{g}_{\bar{J}^{(\pi)}}^{\langle\bar{\pi}+1\rangle}\|^2\right) \\
&\geq (1 - \varrho_1 - \varrho_3)\Delta_\pi^{\langle\underline{\pi}\rangle} \\
&\geq \underbrace{\tau_1(1 - \varrho_1 - \varrho_3)}_{=:\tau\in(0,1)} \Delta^{\langle\underline{\pi}\rangle} \geq 0. \qquad (5.4)
\end{aligned}$$

This proves $\Delta^{\langle\bar{\pi}\rangle} \geq \Delta^{\langle\bar{\pi}+1\rangle}$, hence $(\Delta^{\langle\sigma\rangle})_{\sigma\geq\sigma''}$ is a decreasing sequence. Let $\sigma_1 := \max\{\bar{\pi}+1 \colon \pi \in \Pi^{\langle\sigma''\rangle}\}$, then $\underline{\pi} \geq \sigma''$ for all π with $\bar{\pi} \geq \sigma_1$. Consequently, relation (5.4) implies for each process π with $\bar{\pi} \geq \sigma_1$,

$$\Delta^{\langle\bar{\pi}\rangle} - \Delta^{\langle\bar{\pi}+1\rangle} \geq \tau\Delta^{\langle\underline{\pi}\rangle} \geq \tau\Delta^{\langle\bar{\pi}\rangle}$$

and so

$$\Delta^{\langle\bar{\pi}+1\rangle} \leq (1-\tau)\Delta^{\langle\bar{\pi}\rangle} \leq \tau_3\Delta^{\langle\bar{\pi}\rangle}.$$

This shows that (Upd2) and (Upd2′) hold. ∎

Remark 21. Note that the main algorithm requires the objective value in the current center $f(\hat{y}^{\langle\sigma\rangle})$ in order to compute $\Delta^{\langle\sigma\rangle} = f(\hat{y}^{\langle\sigma\rangle}) - \bar{l}^{\langle\sigma\rangle} - \langle\bar{g}^{\langle\sigma\rangle}, \hat{y}^{\langle\sigma\rangle}\rangle + \frac{1}{u}\|\bar{g}^{\langle\sigma\rangle}\|^2$. However, there is no need to recompute the value if the center is changed by a process π from $\hat{y}^{\langle\bar{\pi}\rangle}$ to $\hat{y}^{\langle\bar{\pi}+1\rangle}$. This follows from the proof of Observation 7, which shows that $f_r(\hat{y}^{\langle\bar{\pi}\rangle}) = f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})$ for all $r \in R \setminus R^{(\pi)}$ and $f_r(\bar{y}^{(\pi)}) = f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})$ for all $r \in R^{(\pi)}$, i.e., the latter values equal the values computed by π on its subspace. Therefore no additional evaluation of f is required apart from the evaluations within a subprocess.

6. Lazy separable functions. In this section we generalize the subspace problem for partially separable functions presented in section 5. As described there, we can only hope for a significant reduction in the overall number of oracle calls if the subproblems couple only few functions f_r, $r \in R$, i.e., if for a given subspace $J \subseteq M$ we have $|R_J| \ll |R|$. In practice, e.g., in Lagrangian relaxation, we frequently encounter functions f_r with large sets $J_r \subseteq M$ while the actual solution just depends on a few of the coordinates in J_r. The following example serves as the central motivation for the approach presented in this section and requires the following notation. For $J \subseteq M$ and $y \in \mathbb{R}^M$ we denote by $y_{|J}$ the vector with values $[y_{|J}]_i = y_i$ for $i \in J$ and $[y_{|J}]_i = 0$ for $i \in M \setminus J$.

Example 22. Consider Lagrangian relaxation decoupling a problem of the form

$$\max\left\{\sum_{r\in R} c_r^T x_r \colon \sum_{r\in R} A_r x_r \leq b, x_r \in \mathcal{X}_r, r \in R\right\}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

with compact $\mathcal{X}_r \subseteq \mathbb{R}^{n_r}_+$, $A_r \in \mathbb{R}^{M\times n_r}_+$, for $r \in R$ and $b \in \mathbb{R}^M$. Introducing Lagrange multipliers $y \in \mathbb{R}^M_+$ (see, e.g., [6] for an extension of bundle methods to sign constraints) for the linear coupling constraints, the dual problem reads

$$\min_{y\in\mathbb{R}^M_+} f(y) = b^T y + \sum_{r\in R} f_r(y) \quad \text{with} \quad f_r(y) = \max_{x_r\in\mathcal{X}_r} (c_r^T x_r - y^T A_r x_r),\ r \in R,$$

and for $r \in R$ the sets J_r are determined by $J_r := \{i \in M\colon [A_r]_{i,\bullet} = 0\}$. Suppose now f_r is evaluated at some $\hat{y} \geq 0$ with support $J = \{i \in M\colon \hat{y}_i \neq 0\}$ yielding an optimizer $\hat{x}_r$ influenced by $J'_r := \{i \in J_r\colon [A_r\hat{x}_r]_i \neq 0 \vee i \in J\}$. Then it can be checked that $\hat{x}_r$ is optimal for any other $\hat{y}' \geq 0$ with $\hat{y}'_{J'_r} = \hat{y}_{J'_r}$, so arbitrary changes in $J_r \setminus J'_r$ do not require reevaluations of f_r. In fact, there is no need to consider values of a $\hat{y}'$ outside of J'_r even in the case of $\hat{y}'_{J'_r} \neq \hat{y}_{J'_r}$ as long as $f(\hat{y}'_{|J'_r})$ is attained for an optimizer $\hat{x}'_r$ satisfying $\{i \in J_r\colon [A_r\hat{x}'_r]_i \neq 0\} \subseteq J'_r$. If this fails to hold then a correct evaluation may next be attempted by simply enlarging J'_r by the support of the new gradient $A_r\hat{x}'_r$. However, any strict enlargement requires another evaluation with respect to the enlarged support which may entail yet another enlargement till a new stable support $J''_r \subseteq J_r$ is found for $\hat{y}'$.

For each $r \in R$ the algorithm will maintain an *active subspace* $J_r^{\langle\sigma\rangle} \subseteq J_r$, $\sigma \in \Sigma$, which collects the indexes of J_r that have shown some influence on the evaluation of f_r. The next definition may be interpreted as formulating structural requirements on f_r that allow the algorithm to automatically adapt the active subspaces.

DEFINITION 23. *Given $r \in R$, a point $\hat{y} \in \mathbb{R}^M$, a minorant $\bar{w}_r = (\bar{l}_r, \bar{g}_r) \in W_r$, and a set $J'_r \subseteq J_r$, the triple $(\hat{y}, \bar{w}_r, J'_r)$ is called* consistent *for r if*

(C) $f_r(y) = f_r(\hat{y}_{|J})$ *for all* $y_{|J} = \hat{y}_{|J}, y \in \mathbb{R}^M, J'_r \subseteq J \subseteq J_r$, *and* $(\bar{g}_r)_{M\setminus J'_r} = 0$.

We say the (global) data at σ are consistent if $(\hat{y}^{\langle\sigma\rangle}, \bar{w}_r^{\langle\sigma\rangle}, J_r^{\langle\sigma\rangle})$ is consistent for each $r \in R$.

By definition of J_r, any triple $(\hat{y}, w_r, J_r)$ is consistent for r. Algorithmically it is not possible to check consistency of subsets of J_r without additional structural knowledge. Therefore the algorithm assumes for each $r \in R$ the availability of a *consistency oracle* $\mathcal{C}_r\colon \mathbb{R}^M \times W_r \times 2^M \to 2^{J_r}$ returning an index set $J = \mathcal{C}_r(\hat{y}, w_r, J'_r)$ with $J'_r \subseteq J$ so that $(\hat{y}, w_r, J)$ is consistent for r ($J = J_r$ is always a feasible choice).

Remark 24. For the packing-type application described in Example 22 a consistency oracle $\mathcal{C}_r(\hat{y}, w_r, J'_r)$ can be implemented quite efficiently. Due to the nonnegativity of all quantities involved, i.e., $\hat{y} \geq 0$, $\mathcal{X}_r \subseteq \mathbb{R}^{n_r}_+$, $A_r \in \mathbb{R}^{M\times n_r}_+$, it suffices to find an index set J with $J'_r \subseteq J \subset J_r$ for which an optimizer $\hat{x}_r$ exists with $[A_r\hat{x}_r]_{M\setminus J} = 0$. Therefore starting with $J \leftarrow J'_r$ and iterating the process of finding $\hat{x}_r \in \operatorname{Arg\,max}_{x_r\in\mathcal{X}_r}(c_r^T x_r - \hat{y}_{|J}^T A_r x_r)$, stopping if $J' \leftarrow \{i \in M\colon [A_r\hat{x}_r]_i \neq 0\}$ satisfies $J' \subseteq J$, else repeating this with $J \leftarrow J \cup J'$, provides a valid J.

Consistency is maintained if only the subspace J'_r is enlarged.

OBSERVATION 25. *Given $r \in R$, $\hat{y} \in \mathbb{R}^M$, $w_r \in W_r$, and sets $J'_r \subseteq J''_r \subseteq J_r$, suppose $(\hat{y}, w_r, J'_r)$ is consistent. Then $(\hat{y}, w_r, J''_r)$ is consistent, too.*

Proof. The proof is obvious. ∎

The algorithmic approach follows the same steps as for partially separable functions, but this time we set up the subspace problem w.r.t. the *assumption* that the functions f_r only depend on the variables $y_{J_r^{\langle\pi\rangle}}$. In order to emphasize this difference in the objects related to a process π we now use a superscript $[\pi]$ in brackets. In

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

particular, let $J^{[\pi]}$ be the subspace associated with a process π, then we define

$$
\begin{aligned}
J_r^{[\pi]} &:= J_r^{\langle\underline{\pi}\rangle} \cap J^{[\pi]}, & \bar{J}_r{}^{[\pi]} &:= J_r^{\langle\underline{\pi}\rangle} \setminus J^{[\pi]},\\
\bar{J}^{[\pi]} &:= \bigcup_{r\in R^{[\pi]}} \bar{J}_r{}^{[\pi]}, & R^{[\pi]} &:= \Big\{ r \in R\colon J_r^{\langle\underline{\pi}\rangle} \cap J^{[\pi]} \neq \emptyset \Big\},\\
f_r^{[\pi]}(y) &:= f_r(y_{|J_r^{\langle\underline{\pi}\rangle}}), & f^{[\pi]}(y) &:= \sum_{r\in R^{[\pi]}} f_r^{[\pi]}(y).
\end{aligned} \tag{6.1}
$$

Using these notations the subspace problem associated with process π reads

$$
\begin{aligned}
&\text{minimize} && f^{[\pi]}(y)\\
&\text{subject to} && y \in \mathcal{L}^{[\pi]} := \mathcal{L}(J^{[\pi]}, \hat{y}^{\langle\underline{\pi}\rangle}).
\end{aligned} \tag{$\mathrm{Sub}_{\mathrm{x}}^{\pi}$}
$$

A center point $\hat{y}^{[\pi]} \in \mathcal{L}^{[\pi]}$ and an aggregated minorant $\bar{w}^{[\pi]} = (\bar{l}^{[\pi]}, \bar{g}^{[\pi]}) \in W^{[\pi]} := \bigtimes_{r\in R^{[\pi]}} W_r$ define the candidate

$$
\bar{y}^{[\pi]} = \arg\min \Big\{ \hat{f}^{[\pi]}_{\bar{w}^{[\pi]}}(y) + \frac{u}{2}\|y - \hat{y}^{[\pi]}\|^2 \colon y \in \mathcal{L}^{[\pi]} \Big\},
$$

where

$$
\hat{f}^{[\pi]}_{\bar{w}^{[\pi]}}(y) = \sum_{r\in R^{[\pi]}} \hat{f}_{\bar{w}_r^{[\pi]},r}(y_{|J_r^{\langle\underline{\pi}\rangle}}).
$$

This gives rise to the expected progress

$$
\begin{aligned}
\Delta^{[\pi]}(\bar{w}^{[\pi]}, \hat{y}^{[\pi]}) &= f^{[\pi]}(\hat{y}^{[\pi]}) - \hat{f}^{[\pi]}_{\bar{w}^{[\pi]}}(\bar{y}^{[\pi]})\\
&= \sum_{r\in R^{[\pi]}} \Big(f_r^{[\pi]}(\hat{y}^{[\pi]}) - \bar{l}_r^{[\pi]} + \langle (\bar{g}_r^{[\pi]})_{J_r^{\langle\underline{\pi}\rangle}}, \hat{y}^{[\pi]}_{J_r^{\langle\underline{\pi}\rangle}} \rangle \Big)\\
&\quad + \frac{1}{u} \Big\| \sum_{r\in R^{[\pi]}} (\bar{g}_r^{[\pi]})_{|J_r^{[\pi]}} \Big\|^2.
\end{aligned}
$$

Here, consistency is required in order to decompose the global expected progress as before.

Observation 26. *For $\pi \in \bar{\Pi}^{\langle\infty\rangle}$, $\hat{y} \in \mathbb{R}^M$, $\bar{w} \in W$ with $(\hat{y}, \bar{w}_r = (\bar{l}_r, \bar{g}_r), J_r^{\langle\underline{\pi}\rangle})$ consistent for $r \in R$ there holds*

$$
\Delta(\bar{w}, \hat{y}) = \Delta^{[\pi]}(\bar{w}_{R^{[\pi]}}, \hat{y}) + \bar{\Delta}^{[\pi]}(\bar{w}, \hat{y}) + \frac{1}{u}\|\bar{g}_{\bar{J}^{[\pi]}}\|^2, \tag{6.2}
$$

where

$$
\bar{\Delta}^{[\pi]}(\bar{w}, \hat{y}) = \sum_{R\setminus R^{[\pi]}} \big(f_r(\hat{y}) - \bar{l}_r - \langle \bar{g}_r, \hat{y} \rangle \big) + \frac{1}{u}\|\bar{g}_{M\setminus(J^{[\pi]}\cup\bar{J}^{[\pi]})}\|^2.
$$

Proof. By definition we have

$$
\Delta(\bar{w}, \hat{y}) = \tilde{\Delta} + \bar{\Delta}^{[\pi]}(\bar{w}, \hat{y}) + \frac{1}{u}\|\bar{g}_{\bar{J}^{[\pi]}}\|^2,
$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where

$$\tilde{\Delta} = \sum_{r \in R^{[\pi]}} \big(f_r(\hat{y}) - \bar{l}_r - \langle \bar{g}_r, \hat{y} \rangle\big) + \frac{1}{u} \|\bar{g}_{J^{[\pi]}}\|^2,$$

so it remains to show that $\tilde{\Delta} = \Delta^{[\pi]}(\bar{w}_{R^{[\pi]}}, \hat{y})$. The consistency assumption implies $f_r^{[\pi]}(\hat{y}) = f_r(\hat{y})$ for all $r \in R^{[\pi]}$, as well as $\langle \bar{g}_r, \hat{y} \rangle = \langle (\bar{g}_r)_{J_r^{\langle \pi \rangle}}, \hat{y}_{J_r^{\langle \pi \rangle}} \rangle$ by $\bar{g}_r = (\bar{g}_r)_{|J_r^{\langle \pi \rangle}}$ for all $r \in R$. Because $J_r^{[\pi]} = J_r^{\langle \pi \rangle} \cap J^{[\pi]} \neq \emptyset \iff r \in R^{[\pi]}$ by definition, there holds

$$\bar{g}_{|J^{[\pi]}} = \sum_{r \in R} (\bar{g}_r)_{|J^{[\pi]}} = \sum_{r \in R} (\bar{g}_r)_{|J_r^{[\pi]}} = \sum_{r \in R^{[\pi]}} (\bar{g}_r)_{|J_r^{[\pi]}},$$

completing the proof. ∎

The set of states and the strongly dependent subspace function are now a combination of the dependency graph function and the activity sets. In particular, let $\mathcal{S}^{\mathrm{act}} := \{(J'_r)_{r \in R} \colon J'_r \subseteq J_r, r \in R\}$ with $S \preceq S'$ for $S, S' \in \mathcal{S}^{\mathrm{act}}$ if and only if all sets J'_r of S are subsets of the corresponding sets in S'. Combining this with $\mathcal{S}^{\mathrm{gra}}$ we get $\mathcal{S}^{\mathrm{x}} = \mathcal{S}^{\mathrm{gra}} \times \mathcal{S}^{\mathrm{act}}$ with $(S_1, S_2) \preceq (S'_1, S'_2) \iff S_1 \preceq S'_1 \wedge S_2 \preceq S'_2$. Using the consistency oracle, the initial state is set to $S^{\langle 0 \rangle} := (\emptyset, (\mathcal{C}_r(\hat{y}^{\langle 0 \rangle}, \bar{w}_r^{\langle 0 \rangle}, \emptyset))_{r \in R})$. Thus, the global data are consistent for $\sigma = 0$.

We denote the activity set of $r \in R$ associated with a state $S \in \mathcal{S}^{\mathrm{x}}$ by $J_r(S)$. The corresponding strongly dependent subspace function is then

$$F_D^{\mathrm{x}}((S_1, S_2), J) := F_D^{\mathrm{gra}}(S_1, J).$$

The weakly dependent subspace function is defined analogously to section 5, guaranteeing that the center is not changed on $J^{[\pi]} \cup \bar{J}^{[\pi]}$ by other processes as long as π runs, i.e.,

$$F_d^{\mathrm{x}}(S, J) := J \cup \bar{J}(S),$$

where

$$\bar{J}(S) := \bigcup_{r \in R(S,J)} J_r(S) \setminus J, \quad R(S, J) := \{r \in R \colon J_r(S) \cap J \neq \emptyset\}.$$

Note that, in contrast to $F_d^{\mathrm{p}}(S, J)$, the function $F_d^{\mathrm{x}}(S, J)$ really depends on the state S. The expected subspace progress is given by

$$\Delta^{\mathrm{x}}(S, J, \bar{w}, \hat{y}) := \sum_{r \in R(S,J)} \Big(f_r(\hat{y}_{|J_r(S)}) - \hat{f}_{\bar{w}_r, r}(\hat{y}_{|J_r(S)})\Big) + \frac{1}{u} \Big\| \sum_{r \in R(S,J)} (\bar{g}_r)_{|J \cap J_r(S)} \Big\|^2.$$

As above one proves that in the case of consistent global data a successful call to Algorithm 2 of process π returns a $J_r^{[\pi]}$ with

$$\Delta^{\mathrm{x}}(S^{\langle \pi \rangle}, J^{[\pi]}, \bar{w}^{\langle \pi \rangle}, \hat{y}^{\langle \pi \rangle}) = \Delta^{[\pi]}(\bar{w}^{\langle \pi \rangle}_{R^{\langle \pi \rangle}}, \hat{y}^{\langle \pi \rangle}) \geq \tau_1 \Delta^{\langle \pi \rangle}.$$

Like in the partially separable case we get disjointness of the subfunctions associated with parallel processes.

Observation 27. *Let $\sigma \in \Sigma$ and $\pi, \eta \in \Pi^{\langle \sigma \rangle}$, $\pi \neq \eta$, be two parallel processes. Then $R^{[\pi]} \cap R^{[\eta]} = \emptyset$.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. Without loss of generality we may assume $\underline{\pi} < \underline{\eta}$. Suppose, for contradiction, there exists an $r \in R^{[\pi]} \cap R^{[\eta]}$. By definition there must be a $j \in J^{[\pi]} \cap J_r^{\langle \underline{\pi} \rangle}$, and $S^{\langle \underline{\pi} \rangle} \preceq S^{\langle \underline{\eta} \rangle}$ implies $J_r^{\langle \underline{\pi} \rangle} \subseteq J_r^{\langle \underline{\eta} \rangle}$. Because $r \in R^{[\eta]}$ we have $J_r^{\langle \underline{\eta} \rangle} \subseteq J^{[\eta]} \cup \bar{J}^{[\eta]}$, hence

$$j \in J_r^{\langle \underline{\pi} \rangle} \subseteq J_r^{\langle \underline{\eta} \rangle} \subseteq J^{[\eta]} \cup \bar{J}^{[\eta]} = F_d^{\mathrm{x}}(S^{\langle \underline{\eta} \rangle}, J^{[\eta]}),$$

so $J^{[\pi]} \cap F_d^{\mathrm{x}}(S^{\langle \underline{\eta} \rangle}, J^{[\eta]}) \neq \emptyset$. This is a contradiction to (3.5), which holds by Observation 8. ∎

The subspace process proceeds as in the separable case. The subspace problem is solved by a bundle method with descent parameter $\varrho_2 \in (0,1)$ and termination parameter $\varrho_1 \in (0,1)$ until the aggregate minorant $\bar{w}^{[\pi]} \in W^{[\pi]}$ and the candidate $\bar{y}^{[\pi]} \in \mathcal{L}^{[\pi]}$ either satisfy

$$\text{(Upd1}_{\mathrm{x}}\text{)} \qquad \begin{aligned} \Delta^{[\pi]}(\bar{w}^{[\pi]}, \hat{y}^{\langle \underline{\pi} \rangle}) &\geq \varrho_1 \Delta^{[\pi]}(\bar{w}^{\langle \underline{\pi} \rangle}_{R^{[\pi]}}, \hat{y}^{\langle \underline{\pi} \rangle}) \\ f^{[\pi]}(\hat{y}^{\langle \underline{\pi} \rangle}) - f^{[\pi]}(\bar{y}^{[\pi]}) &\geq \varrho_2 \Delta^{[\pi]}(\bar{w}^{[\pi]}, \hat{y}^{\langle \underline{\pi} \rangle}), \end{aligned}$$

or

$$\text{(Upd2}_{\mathrm{x}}\text{)} \qquad \Delta^{[\pi]}(\bar{w}^{[\pi]}, \hat{y}^{\langle \underline{\pi} \rangle}) < \varrho_1 \Delta^{[\pi]}(\bar{w}^{\langle \underline{\pi} \rangle}_{R^{[\pi]}}, \hat{y}^{\langle \underline{\pi} \rangle}).$$

Observation 26 indicates that the critical term is again the third one of (6.2), so that the subprocess will use the following dependency test for some $\varrho_3 \in (0, 1 - \varrho_1)$:

$$\text{(Dep}_{\mathrm{x}}\text{)} \qquad \frac{1}{u} \left\| \bar{g}^{\langle \bar{\pi}+1 \rangle}_{\bar{J}^{(\pi)}} \right\|^2 - \frac{1}{u} \left\| \bar{g}^{\langle \bar{\pi} \rangle}_{\bar{J}^{(\pi)}} \right\|^2 > \varrho_3 \Delta^{[\pi]} \left(\bar{w}^{\langle \underline{\pi} \rangle}_{R^{[\pi]}}, \hat{y}^{\langle \underline{\pi} \rangle} \right).$$

If condition (Dep$_{\mathrm{x}}$) is fulfilled, the dependency graph is enlarged by adding an edge according to (5.3).

There is one important difference from the partially separable case. In setting up the subproblem the contributing functions and their restricted evaluation spaces were determined under the assumption that for each $r \in R$ the relevant parameters of f_r were given by $J_r^{\langle \underline{\pi} \rangle}$. Before setting the new values we have to make sure that these assumptions are still valid. If at $\bar{\pi}$ the old sets $J_r^{\langle \underline{\pi} \rangle}$ prove to be too small, we dump the results and only make sure that the state increases strictly in comparison to $\underline{\pi}$. In the case of a descent step all functions potentially affected by $J^{[\pi]}$, inside as well as outside of $R^{[\pi]}$, have to be checked for their consistency, because (C) concerns the potential influence of all these coordinates (so those with $J^{[\pi]} \cap J_r \neq \emptyset$ would suffice). In the case of a null step only the aggregate may change, so checking for new nonzero coordinates in these suffices to ensure condition (C). Algorithm 5 implements these aspects in the subspace update.

As before, the center cannot change on $J^{[\pi]} \cup \bar{J}^{[\pi]}$ as long as π runs and may only change in $J^{[\pi]}$ when π finishes.

COROLLARY 28. *Let π be a process, then $\hat{y}_j^{\langle \underline{\pi} \rangle} = \hat{y}_j^{\langle \bar{\pi} \rangle}$ for all $j \in J^{[\pi]} \cup \bar{J}^{[\pi]}$ and $\hat{y}_j^{\langle \bar{\pi} \rangle} = \hat{y}_j^{\langle \bar{\pi}+1 \rangle}$ for all $j \notin J^{[\pi]}$.*

Proof. The proof is analogous to the proof of Corollary 19. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

OBSERVATION 29. *The global data are consistent for all* $\sigma \in \Sigma$.

Proof. Initialization ensures consistency for $\sigma = 0$. Inductively, relevant information for consistency is only changed in Algorithm 5. Observation 25 ensures that active subspaces may be increased safely if neither center nor aggregate are changed. In the (Upd2_x) branch, A5.5 ensures that aggregates are only stored if they satisfy consistency, the center remains unchanged. In the (Upd1_x) branch, consistency is taken care of explicitly by A5.2 (note, $\mathcal{C}_r$ is required to only increase active subspaces) and A5.3, which updates to the new center and aggregate only in the case of proved consistency. ∎

ALGORITHM 5. UPDATESUBSPACE$^\text{x}$.

Input : π, final candidate $\bar{y}^{[\pi]} \in \mathcal{L}^{[\pi]}$, aggregate minorant $\bar{w}^{[\pi]} \in W^{[\pi]}$
Changes : global data at $\bar{\pi} + 1$
if π *stopped because of (Upd*1_x*)* **then**
1 $\quad \hat{y}' \leftarrow \hat{y}^{\langle\bar{\pi}\rangle} + (\bar{y}^{[\pi]} - \hat{y}^{\langle\underline{\pi}\rangle})$
$\quad (\bar{w}'_{R^{[\pi]}}, \bar{w}'_{R\setminus R^{[\pi]}}) \leftarrow (\bar{w}^{[\pi]}, \bar{w}^{\langle\bar{\pi}\rangle}_{R\setminus R^{[\pi]}})$
$\quad$ `// update consistency information`
2 $\quad$ **for** $r \in R$ **do** $J_r^{\langle\bar{\pi}+1\rangle} \leftarrow \mathcal{C}_r(\hat{y}', \bar{w}', J_r^{\langle\bar{\pi}\rangle})$
$\quad$ `// no changes in` $\mathcal{S}^{\text{gra}}$`-part,` $S^{\langle\bar{\pi}\rangle} \rightsquigarrow S^{\langle\bar{\pi}+1\rangle}$
3 $\quad$ **if** $J_r^{\langle\bar{\pi}+1\rangle} \neq J_r^{\langle\underline{\pi}\rangle}$ *for some* $r \in R$ **then** `// update` $S^{\langle\bar{\pi}+1\rangle} \succ S^{\langle\underline{\pi}\rangle}$
$\quad\quad \hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}^{\langle\bar{\pi}\rangle}$, $\bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}^{\langle\bar{\pi}\rangle}$
$\quad$ **else**
$\quad\quad \hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}'$, $\bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}'$
else `//` π `stopped because of (`Upd2_x`)`
$\quad \hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}^{\langle\bar{\pi}\rangle}$
4 $\quad (\bar{w}'_{R^{[\pi]}}, \bar{w}'_{R\setminus R^{[\pi]}}) \leftarrow (\bar{w}^{[\pi]}, \bar{w}^{\langle\bar{\pi}\rangle}_{R\setminus R^{[\pi]}})$
5 $\quad$ **if** $(\bar{g}'_r)_j \neq 0$ *for some* $r \in R^{[\pi]}$, $j \in M \setminus J_r^{\langle\underline{\pi}\rangle}$ **then** `// update`
$\quad S^{\langle\bar{\pi}+1\rangle} \succ S^{\langle\underline{\pi}\rangle}$
$\quad\quad$ **for** $r \in R^{[\pi]}$ **do** $J_r^{\langle\bar{\pi}+1\rangle} \leftarrow J_r^{\langle\bar{\pi}\rangle} \cup \{j \in J_r \setminus J_r^{\langle\bar{\pi}\rangle} : (\bar{g}'_r)_j \neq 0\}$
$\quad\quad$ `// no changes in` $\mathcal{S}^{\text{gra}}$`-part,` $S^{\langle\bar{\pi}\rangle} \rightsquigarrow S^{\langle\bar{\pi}+1\rangle}$
$\quad\quad \bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}^{\langle\bar{\pi}\rangle}$
$\quad$ **else**
$\quad\quad$ **if** *(Dep*$_x$*) holds* **then** `// update` $S^{\langle\bar{\pi}+1\rangle} \succ S^{\langle\underline{\pi}\rangle}$
$\quad\quad\quad$ add (j, j') as defined in (5.3) to the $\mathcal{S}^{\text{gra}}$-part of $S^{\langle\bar{\pi}\rangle} \rightsquigarrow S^{\langle\bar{\pi}+1\rangle}$
$\quad\quad$ **else** $S^{\langle\bar{\pi}+1\rangle} \leftarrow S^{\langle\bar{\pi}\rangle}$
$\quad\quad \bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}'$

LEMMA 30. *If* π *stops because of (Upd*1_x*) and* $S^{\langle\underline{\pi}\rangle} = S^{\langle\bar{\pi}+1\rangle}$, *then due to Algorithm* 5 *there hold*

(i) $J_r^{\langle\underline{\pi}\rangle} = J_r^{\langle\bar{\pi}\rangle} = J_r^{\langle\bar{\pi}+1\rangle}$ *and* $(\hat{y}', \bar{w}', J_r^{\langle\bar{\pi}\rangle})$ *is consistent for all* $r \in R$,
(ii) $f_r^{[\pi]}(\hat{y}^{\langle\underline{\pi}\rangle}) = f_r(\hat{y}^{\langle\underline{\pi}\rangle}) = f_r(\hat{y}^{\langle\bar{\pi}\rangle})$, $f_r^{[\pi]}(\bar{y}^{[\pi]}) = f_r(\bar{y}^{[\pi]}) = f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})$ *for all* $r \in R^{[\pi]}$,
(iii) $f_r(\hat{y}^{\langle\bar{\pi}\rangle}) = f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})$ *for all* $r \in R \setminus R^{[\pi]}$.

Proof. Let π be a process that stops because of (Upd1_x). If $J_r^{\langle\underline{\pi}\rangle} \neq J_r^{\langle\bar{\pi}+1\rangle}$ for some $r \in R$ then $S^{\langle\underline{\pi}\rangle} \neq S^{\langle\bar{\pi}+1\rangle}$. Thus $J_r^{\langle\underline{\pi}\rangle} = J_r^{\langle\bar{\pi}\rangle} = J_r^{\langle\bar{\pi}+1\rangle}$ for $r \in R$ because the

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

sets $J_r^{\langle\sigma\rangle}$ only increase. This implies consistency of $(\hat{y}', \bar{w}', J_r^{\langle\bar{\pi}+1\rangle})$ for $r \in R$ by A5.2 and also shows $\hat{y}^{\langle\bar{\pi}+1\rangle} = \hat{y}'$.

By (6.1) we have $J_r^{\langle\bar{\pi}\rangle} \cap J^{[\pi]} = \emptyset$ for $r \in R \setminus R^{[\pi]}$ and $J_r^{\langle\underline{\pi}\rangle} \subset J^{[\pi]} \cup \bar{J}^{[\pi]}$ for $r \in R^{[\pi]}$. Corollary 28 ensures $\hat{y}^{\langle\underline{\pi}\rangle}_{J^{[\pi]}\cup\bar{J}^{[\pi]}} = \hat{y}^{\langle\bar{\pi}\rangle}_{J^{[\pi]}\cup\bar{J}^{[\pi]}}$ and $\bar{y}^{[\pi]} \in \mathcal{L}(J^{[\pi]}, \hat{y}^{\langle\underline{\pi}\rangle})$, so for each $r \in R \setminus R^{[\pi]}$ we obtain $\hat{y}'_{J_r^{\langle\underline{\pi}\rangle}} = \hat{y}^{\langle\bar{\pi}\rangle}_{J_r^{\langle\underline{\pi}\rangle}}$ while for $r \in R^{[\pi]}$ there holds $\hat{y}'_{J_r^{\langle\underline{\pi}\rangle}} = \bar{y}^{[\pi]}_{J_r^{\langle\underline{\pi}\rangle}}$. Now the definition (6.1) of $f_r^{[\pi]}$, the fact that consistency holds by (i) and Observation 29 prove (ii) and (iii). ∎

LEMMA 31. *For $\tau_2 \in (0, \varrho_2\varrho_1\tau_1]$ and $\tau_3 \in [1 - \tau_1(1 - \varrho_1 - \varrho_3), 1)$ there hold*

(i) *(Inv1) and (Inv2) hold;*
(ii) *if an infinite number of processes terminate with (Upd1$_x$), then an infinite number of processes satisfy (Upd1);*
(iii) *if there is only a finite number of processes satisfying (Upd1$_x$) then there is a $\sigma_1 \in \Sigma$ so that each process π with $\bar{\pi} \geq \sigma_1$ satisfies (Upd2).*

Proof. First note that (Inv1) holds because Algorithm 5 ensures $S^{\langle\bar{\pi}+1\rangle} \succeq S^{\langle\bar{\pi}\rangle}$ in all cases. Furthermore by Observation 7 there is a $\sigma_0 \in \Sigma$ so that $S^{\langle\sigma\rangle} = S^{\langle\sigma_0\rangle}$ for all $\sigma \geq \sigma_0$. Hence, for any process π with $\bar{\pi} \geq \sigma' := \max\{\bar{\eta} + 1 \colon \eta \in \Pi^{\langle\sigma_0\rangle}\}$ we have $\underline{\pi} \geq \sigma_0$ and so $S^{\langle\underline{\pi}\rangle} = S^{\langle\bar{\pi}\rangle} = S^{\langle\bar{\pi}+1\rangle}$.

Let π be a process terminating with (Upd1$_x$). If $S^{\langle\underline{\pi}\rangle} \neq S^{\langle\bar{\pi}+1\rangle}$ then π does not modify the center, so $\hat{y}^{\langle\bar{\pi}+1\rangle} = \hat{y}^{\langle\bar{\pi}\rangle}$ proving (Inv2). Now assume $S^{\langle\underline{\pi}\rangle} = S^{\langle\bar{\pi}+1\rangle}$ (particularly this is the case if $\bar{\pi} \geq \sigma'$). Then Lemma 30 implies

$$\begin{aligned}
f(\hat{y}^{\langle\bar{\pi}\rangle}) - f(\hat{y}^{\langle\bar{\pi}+1\rangle}) &= \sum_{r\in R^{[\pi]}} \Big(f_r(\hat{y}^{\langle\bar{\pi}\rangle}) - f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})\Big) \\
&\quad + \sum_{r\in R\setminus R^{[\pi]}} \underbrace{\Big(f_r(\hat{y}^{\langle\bar{\pi}\rangle}) - f_r(\hat{y}^{\langle\bar{\pi}+1\rangle})\Big)}_{=0} \\
&= \sum_{r\in R^{[\pi]}} \Big(f_r^{[\pi]}(\hat{y}^{\langle\underline{\pi}\rangle}) - f_r^{[\pi]}(\bar{y}^{[\pi]})\Big) = f^{[\pi]}(\hat{y}^{\langle\underline{\pi}\rangle}) - f^{[\pi]}(\bar{y}^{[\pi]}) \\
&\overset{(\text{Upd1}_x)}{\geq} \varrho_2\Delta^{[\pi]}(\bar{w}^{[\pi]}, \hat{y}^{\langle\underline{\pi}\rangle}) \overset{(\text{Upd1}_x)}{\geq} \varrho_2\varrho_1\Delta^{[\pi]}(\bar{w}^{\langle\underline{\pi}\rangle}_{R^{[\pi]}}, \hat{y}^{\langle\underline{\pi}\rangle}) \\
&\overset{(\text{Sel1})}{\geq} \varrho_2\varrho_1\tau_1\Delta^{\langle\underline{\pi}\rangle} \geq \tau_2\Delta^{\langle\underline{\pi}\rangle} \geq 0.
\end{aligned}$$

This proves (Inv2) as well as (Upd1) in this case.

Now let π be a process terminating with (Upd2$_x$). In this case $\hat{y}^{\langle\bar{\pi}+1\rangle} = \hat{y}^{\langle\bar{\pi}\rangle}$ proving (Inv2). Assume that there is only a finite number of processes satisfying (Upd1$_x$). Exactly as in the proof of Lemma 20 we get a $\sigma_1 \in \Sigma$ so that each process π with $\bar{\pi} \geq \sigma_1$ stops with (Upd2$_x$), there holds $S^{\langle\underline{\pi}\rangle} = S^{\langle\bar{\pi}+1\rangle}$, and the dependency test (Dep$_x$) is not fulfilled. In particular, $J_r^{\langle\underline{\pi}\rangle} = J_r^{\langle\bar{\pi}\rangle} = J_r^{\langle\bar{\pi}+1\rangle}$ allows us to employ Observation 26 and to follow exactly the same steps as in the proof of Lemma 20. ∎

Remark 32. Analogously to the partial separable case, there is no need to evaluate $f(\hat{y}^{\langle\sigma\rangle})$ when the center is changed because the function values $f_r(\hat{y}^{\langle\sigma\rangle})$ that change have already been computed by the subprocess that sets the new center.

7. Concluding remarks. It is worth highlighting the differences between the three approaches for generating subspace problems proposed in sections 4–6 by considering their application to finding the best Lagrangian multipliers y for the Lagrangian relaxation of the packing problem in Example 22. Recall that in this setting the pack-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ing problem is thereby decomposed into many independent primal subproblems, for each of which the actual influence of the multipliers on the optimal solution may vary as the algorithm progresses.

The general version of section 4 will create subspace problems that do not take advantage of the separable structure. Hence, in solving a subspace problem each oracle call requires us to evaluate the entire objective, i.e., to solve all independent primal subproblems irrespective of the actual influence of the selected multipliers. The update of the global data only accepts a new point if it has a better objective value than the previous global point, which requires an additional evaluation of the objective function.

In contrast, the subspace problem of section 5 exploits the separable structure as guaranteed a priori by the formal presence of the multipliers in the objective of the relaxed primal subproblems. In particular, for a selected subspace $J^{(\pi)}$, the subspace problem only has to evaluate the primal subproblems $R_{J^{(\pi)}}$, i.e., the primal subproblems having nonzero coefficients in some constraint associated with $J^{(\pi)}$. Furthermore the subspace selection step of this process guarantees that during its life time no other process may change any other multiplier influencing the primal subproblems $r \in R_{J^{(\pi)}}$ or require their evaluation. This ensures that the update of the global data needs no additional evaluation of the global objective function. Note that this subspace problem only requires separability of the objective function but not its origin from Lagrangian relaxation.

Finally, the third subspace problem of section 6 uses explicit knowledge of the Lagrangian relaxation structure. Instead of considering all primal subproblems that are coupled by a certain constraint $j \in J^{[\pi]}$ it only considers those for which the constraint appeared to be important (e.g., if $(\bar{g}_r^{\langle\sigma\rangle})_j \neq 0$). For a selection $J^{[\pi]}$ this approach therefore works on fewer primal subproblems. In consequence fewer variables are blocked, which allows for more parallel processes, and in each process each oracle call needs to evaluate fewer primal subproblems. There is, however, also a downside to this approach. Before writing back the results to the global data, an additional test is required of whether the presumed independencies of the primal subproblems and their multipliers still hold in view of new global information. If the test fails, the computed function values of the current subproblem might be incorrect and have to be discarded.

Some preliminary numerical experiments, mainly regarding the approach of section 6, have been presented in [3] for Lagrangian relaxation of a *train timetabling problem*. There, the primal subproblems are shortest path problems in time expanded networks, which are coupled by inequality constraints restricting the simultaneous usage of common resources like stations and tracks. For the instances considered, the number of subproblem evaluations required to solve the problem to a certain precision was greatly reduced by the asynchronous parallel bundle method. The amount of reduction depended, as expected, on the strength of the interdependencies of the constraints. When increasing the number of long-distance trains, which interact with many other trains, the gain in reduction of evaluations decreases. These proof-of-concept experiments, however, also proved that parallelism requires much more sophisticated implementations. Indeed, on a computer with four cores the gain in wall clock time was comparable when using a single process and when using up to four parallel processes. Still the experiments confirm that the parallel bundle method can be developed into a competitive alternative for large scale problems with separable structure.

Acknowledgment. We thank an anonymous referee for suggesting to attempt an extension of [3] to general convex functions.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 08/07/14 to 132.204.100.226. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

REFERENCES

[1] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical Optimization*, 2nd ed., Springer, Berlin, 2006.

[2] G. Emiel and C. Sagastizábal, *Incremental-like bundle methods with application to energy planning*, Comput. Optim. Appl., 46 (2010), pp. 305–332.

[3] F. Fischer and C. Helmberg, *A Parallel Bundle Method for Asynchronous Subspace Optimization in Lagrangian Relaxation*, preprint 2012-02, Technische Universität Chemnitz, Fakultät für Mathematik, Chemnitz, Germany, 2012.

[4] M. L. Fisher, *The Lagrangian relaxation method for solving integer programming problems*, Manag. Sci., 50 (2004), pp. 1861–1871.

[5] M. Fukushima, *Parallel variable transformation in unconstrained optimization*, SIAM J. Optim., 8 (1998), pp. 658–672.

[6] C. Helmberg and K. C. Kiwiel, *A spectral bundle method with bounds*, Math. Program., 93 (2002), pp. 173–194.

[7] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms* II, Grundlehren Math. Wiss. 306, Springer, Berlin, 1993.

[8] K. C. Kiwiel, *Proximity control in bundle methods for convex nondifferentiable minimization*, Math. Program., 46 (1990), pp. 105–122.

[9] C. Lemaréchal, *Lagrangian relaxation*, in Computational Combinatorial Optimization, Michael Jünger and Denis Naddef, eds., Lecture Notes in Comput. Sci. 2241, Springer, Berlin, 2001, pp. 112–156.

[10] C. Lemaréchal, *The omnipresence of Lagrange*, Ann. Oper. Res., 153 (2007), pp. 9–27.

[11] D. Medhi, *Parallel bundle-based decomposition for large-scale structured mathematical programming problems*, Ann. Oper. Res., 22 (1990), pp. 101–127.

[12] A. Nedić, D. P. Bertsekas, and V. S. Borkar, *Distributed asynchronous incremental subgradient methods*, in Inherently Parallel Algorithms in Feasibility and Optimization and their Applications, Y. C. Dan Butnariu and S. Reich, eds., Studies Comput. Math. 8, Elsevier, Amsterdam, 2001, pp. 381–407.

[13] C. A. Sagastizábal and M. V. Solodov, *Parallel variable distribution for constrained optimization*, Comput. Optim. Appl., 22 (2002), pp. 111–131.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.