# A Parallel Bundle Framework for Asynchronous Subspace Optimisation of Nonsmooth Convex Functions

F. Fischer, C. Helmberg

Preprint 2013-8

Preprintreihe der Fakultät für Mathematik
ISSN 1614-8835

**Impressum:**
**Herausgeber:**
Der Dekan der
Fakultät für Mathematik
an der Technischen Universität Chemnitz
**Sitz:**
Reichenhainer Straße 39
09126 Chemnitz
**Postanschrift:**
09107 Chemnitz
Telefon: (0371) 531-22000
Telefax: (0371) 531-22009
E-Mail: dekanat@mathematik.tu-chemnitz.de
**Internet:**
http://www.tu-chemnitz.de/mathematik/
ISSN 1614-8835 (Print)

# A Parallel Bundle Framework for Asynchronous Subspace Optimisation of Nonsmooth Convex Functions

Frank Fischer[*] Christoph Helmberg[*]

2nd July 2013

An algorithmic framework is presented for optimising general convex functions by non synchronised 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 analysing 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 generalises a practical method proposed by the authors for Lagrangian relaxation of large scale packing problems and simplifies the analysis.

**Keywords:** bundle methods, parallel programming, convex optimisation
**MSC 2010:** 90C06; 65Y05, 90C25, 65K05

## 1 Introduction

Let $f\colon \mathbb{R}^M \to \mathbb{R}$, $M = \{1, \dots, m\}$ be a non-smooth, 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 optimisation problem

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

Bundle methods are an established tool for solving such optimisation problems [1, 6]. In a nutshell, they collect subgradient information from the set

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

[*] Fakultät für Mathematik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany. {frank.fischer, helmberg}@mathematik.tu-chemnitz.de

and, based on this, they determine, in minimal 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

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

With respect to a current *centre of stability* $\hat{y}$ a next *candidate* is determined by

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

where the (here fixed) *weight* $u > 0$ penalises 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.*, Theorem 9.15 in [1]) Let $\hat{y}^k$ and $\Delta^k$ denote the centre 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 $\hat{y}^{k_0}$ minimises $f$ 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 [2]. It consists of a main routine (Algorithm 1) that initialises and manages access to global data and maintains a number $N_\Pi \in \{1, \ldots, |M|\}$ of parallel processes performing the actual optimisation independently and asynchronously. To provide a brief sketch of the main idea, the global data includes an access index $\sigma \in \mathbb{N}_0$, the current centre $\hat{y}^{\langle\sigma\rangle}$, 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 $\pi$ 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 $\pi$ tries to reserve an unblocked subset of coordinates $J^{(\pi)} \subseteq M$ defining an affine subspace

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

so that, according to current dependency information, optimising $f$ over $\mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle})$ promises a significant share of $\Delta^{\langle\underline{\pi}\rangle}$. This will always be possible for at least one process. If this fails due to blocking conditions, $\pi$ frees access again and stops immediately without having modified any data (a new process may be restarted once the global data is 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, optimising $f$ over $\mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle})$, increases the access index $\sigma$ and frees global access.

**2. Subspace optimisation.** Process $\pi$ forms the subproblem and optimises 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 $\pi$ at access counter $\underline{\pi}$, or it stops if the reduction in predicted decrease on the subproblem enables $\pi$ 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 $\pi$ 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 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 $\sigma$, and checks whether the new global data satisfies 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 case of global termination, all processes are stopped immediately and the algorithm ends. Otherwise $\pi$ 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 to ensure convergence of this non synchronised 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 optimising 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 timetabling problems can be found in [2].

**Literature review.** We give a short overview over existing parallelisation approaches for convex optimisation, also see [2]. The most straight forward 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.* [3, 7, 8], and evaluate each function $f_r$ in an isolated parallel process, see, *e. g.*, [9] for an example within a bundle algorithm.

Another approach are variable transformation algorithms [4, 11]. Similar 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 synchronised 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 [10] 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

**Algorithm 1:** ParallelBundle

```
Input   : Problem (P), parameters
            • $\tau_1 \in (0,1), \varepsilon > 0$
            • $N_\Pi \in \{1,\dots,|M|\}$
Output : Approximate solution $y \in \mathbb{R}^M$
// Initialisation
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 pre-specified dependencies)
if $\Delta^{\langle 0\rangle} \le \varepsilon(|f(\hat{y}^{\langle 0\rangle})| + 1)$ then
    then do not start any process
    set $y \leftarrow \hat{y}^{\langle 0\rangle}$
    return
// Main loop
while Less than $N_\Pi$ processes are running do
    Start a new process $\pi = (\underline{\pi}, \bar{\pi}) \leftarrow (-1,-1)$.
    Each process $\pi$ performs the following steps independently
        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
            $(\bar{w}^{\langle \underline{\pi}+1\rangle}, \hat{y}^{\langle \underline{\pi}+1\rangle}, S^{\langle \underline{\pi}+1\rangle}, \Delta^{\langle \underline{\pi}+1\rangle}) \leftarrow (\bar{w}^{\langle \underline{\pi}\rangle}, \hat{y}^{\langle \underline{\pi}\rangle}, S^{\langle \underline{\pi}\rangle}, \Delta^{\langle \underline{\pi}\rangle})$
            $B^{\langle \underline{\pi}+1\rangle} \leftarrow B^{\langle \underline{\pi}\rangle} \cup J^{(\pi)}$
            $\bar{\pi} \leftarrow \infty$, $\sigma \leftarrow \underline{\pi} + 1$
            Free exclusive access to global data
        Subspace optimisation
            Solve a subspace problem on $J^{(\pi)}$ yielding $\bar{y}^{(\pi)}$ and $\bar{w}^{(\pi)}$
        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.
            UpdateSubspace($\pi, \bar{y}^{(\pi)}, \bar{w}^{(\pi)}$)
            compute $\Delta^{\langle \bar{\pi}+1\rangle}$
            $B^{\langle \bar{\pi}+1\rangle} \leftarrow B^{\langle \bar{\pi}\rangle} \setminus J^{(\pi)}$
            $\sigma \leftarrow \bar{\pi} + 1$
            if $\Delta^{\langle \bar{\pi}+1\rangle} \le \varepsilon(|f(\hat{y}^{\langle \bar{\pi}+1\rangle})| + 1)$ then
                set $y \leftarrow \hat{y}^{\langle \sigma\rangle}$
                TERMINATE all processes and STOP.
            Free exclusive access to global data
            STOP this process $\pi$.
```

of times (which is also the case in the "synchronised" approaches mentioned above).

In contrast to these approaches, our method has no explicit synchronisation or regularisation 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.

## 2 The Parallel Framework

In this section we recall from [2] 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 the third step. 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 $\sigma \in \mathbb{N}_0$. Each global object will be denoted by a superscript $\sigma$ in angle braces $\langle\sigma\rangle$.

Let $\pi$ be a process. Because of the uniqueness of the global index we can equip $\pi$ with the two special index markers $\underline{\pi}$ and $\bar{\pi}$ that correspond to the global index when $\pi$ 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 $\sigma \in \mathbb{N}_0 \cup \{\infty\}$:

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

The first group $\underline{\Pi}^{\langle\sigma\rangle}$ collects all processes that have successfully executed the subspace selection step before the algorithm reached $\sigma$, $\bar{\Pi}^{\langle\sigma\rangle}$ singles out all processes that have executed the subspace update step before $\sigma$ and the set $\Pi^{\langle\sigma\rangle}$ comprises all processes that are actively working on or just finishing a subproblem at $\sigma$, *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 $\sigma$ 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 [2]) 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* $\pi$ *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* $\pi$ *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.

As long as the algorithm runs, the main process will start new subprocesses working on different subspaces, in particular, each process $\pi$ will select and work on a subset of coordinates $J^{(\pi)} \subseteq M$. 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 centre $\hat{y}^{\langle\sigma\rangle} \in \mathbb{R}^M$,
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 $B^{\langle\sigma\rangle} = \bigcup_{\pi \in \Pi^{\langle\sigma\rangle}} J^{(\pi)}$.

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

The first main step of a process $\pi$ is to select appropriate subspace problems. Formally, any given current state $S^{\langle\sigma\rangle}$, current centre $\hat{y}^{\langle\sigma\rangle}$ and subspace $J^{(\pi)}$ 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 to jointly optimise over $J' \cup J$ whenever $J$ is selected as initial subspace. If it seems not necessary to include the coordinates together with $J$ but no other process should modify the values on $J'$ while $J$ is optimised 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

$$\forall\, S \in \mathcal{S} \colon F_X(S, \emptyset) = \emptyset, \tag{1}$$

$$\forall\, S \in \mathcal{S}, J \subseteq M \colon J \subseteq F_X(S, J) \tag{2}$$

$$\forall\, S \in \mathcal{S}, J, J' \subseteq M \colon F_X(S, J \cup J') = F_X(S, J) \cup F_X(S, J'), \tag{3}$$

$$\forall\, S \preceq S' \in \mathcal{S}, J \subseteq M \colon F_X(S, J) \subseteq F_X(S', J). \tag{4}$$

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 $\sigma$ with corresponding subspaces $J^{(\pi)}$ and $J^{(\eta)}$ there holds

$$\begin{aligned} J^{(\pi)} &= F_D^{\langle\underline{\pi}\rangle}(J) \text{ for some } J \subseteq M, \\ F_d^{\langle\underline{\pi}\rangle}(J^{(\pi)}) \cap J^{(\eta)} &= \emptyset = J^{(\pi)} \cap F_d^{\langle\underline{\eta}\rangle}(J^{(\eta)}). \end{aligned} \tag{5}$$

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

$$\forall\, S \in \mathcal{S}, \bar{w} \in W, \hat{y} \in \mathbb{R}^M \colon \Delta(S, M, \bar{w}, \hat{y}) = \Delta(\bar{w}, \hat{y}). \tag{6}$$

This function predicts the expected progress to be made by the subspace problem. Only subspace problems that are guaranteed to satisfy (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 $\pi$ 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 (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 (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 (Sel1) and (Sel2). 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 (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) 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 (6). □

**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 (Sel1) 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 (Sel1) then
    return TRUE
else
    $J^{(\pi)} \leftarrow \emptyset$
    return FALSE
```

## 3.2 Subspace Optimisation and Subspace Update

The concrete structure of the subspace problem is unknown to the main routine. However, the algorithm underlying the processes $\pi$ 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.

**Assumption 4.**

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

$$S^{\langle\bar{\pi}\rangle} \preceq S^{\langle\bar{\pi}+1\rangle}, \tag{Inv1}$$

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

*(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*

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

*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*

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

If condition (Upd1) holds then the process performed a descent step with a sufficiently large decrease in the objective value (compared with the global expected progress when $\pi$ started). In contrast, (Upd2) guarantees that the global predicted decrease has been reduced significantly without causing a descent step.

## 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 $\tau_1$, $\varepsilon$ and $u$. In addition, the parameter $N_\Pi$ 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 5.** *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* $\sigma \geq \sigma_0$.

**Proof.** The proof works by induction on $\sigma$. 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 $\eta$ 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 $\eta$ executes a successful subspace selection step (Algorithm 2) at $\sigma$. Thus the process executes A1.4 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 $\eta$ executes line A1.7 at $\sigma$. 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 6.** *Let* $\sigma \in \Sigma$ *and* $\pi, \eta \in \Pi^{\langle\sigma\rangle}$. *Then (5) holds for* $\pi$ *and* $\eta$.

**Proof.** This follows from test A2.2, Observation 5 and properties (3) and (4). □

The next result characterises 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 7.** *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* $\sigma$ *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 $\sigma$ assuming that the equivalence holds for smaller global indexes.

**(i) $\iff$ (ii):** This follows by definition of $\Sigma$ 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 $\sigma'$ 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 $\pi$ with $\bar{\pi} = \sigma - 1$.

**(iv) $\Rightarrow$ (iii):** (iv) implies that 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 initialisation step. In the second case there must be a process $\pi$ 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 $\sigma$ 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):** (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 $\sigma$. We consider two cases. First if $\exists\, \pi \in \Pi^{\langle\sigma\rangle} \neq \emptyset$ then at least one process is still running at $\sigma$. Because each subspace optimisation is finite by assumption this process will eventually execute its update step at $\bar{\pi} \geq \sigma$ and would increase $\sigma$, a contradiction. Now assume $\Pi^{\langle\sigma\rangle} = \emptyset$. In particular this is the case if $\sigma = 0$. Then Observation 5 implies $B^{\langle\sigma\rangle} = \emptyset$. The algorithm will therefore try to start a new process $\pi$ at $\underline{\pi} = \sigma$ and because of Observation 3 this step will be successful. This causes again $\sigma$ 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 8.** *Suppose an infinite number of processes satisfies (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 5 the sequence $(f(\hat{y}^{\langle\sigma\rangle}))_{\sigma\in\Sigma}$ is non-increasing 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 9.** *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 $\pi$ 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 $\pi$ 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 $\pi$ 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 $\pi$ 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

$$\pi = \arg\max_{\pi\in\bar{\Pi}^{\langle\gamma\rangle}} \{\bar{\pi} \colon \Delta^{\langle\bar{\pi}+1\rangle} = \Delta^{\langle\gamma\rangle}\}.$$

Then we have $\bar{\pi} \geq \sigma$, because otherwise $\bar{\pi} < \sigma$ was the last update before $\gamma$ 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) by

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

for dependency detection, one gets a non-increasing sequence.

**Observation 10.** *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 $\bar{\pi} \geq \sigma_1$ guarantees (Upd2') instead of (Upd2). Then the sequence $(\Delta^{\langle\sigma\rangle})_{\sigma\geq\sigma_1}$ is non-increasing and* $\lim_{\sigma\geq\sigma_1} \Delta^{\langle\sigma\rangle} = 0$.

**Proof.** Given a $\sigma_1 \in \Sigma$ so that all processes $\pi$ with $\bar{\pi} \geq \sigma_1$ satisfy (Upd2'), in particular $\Delta^{\langle\bar{\pi}+1\rangle} \leq \tau_3 \Delta^{\langle\bar{\pi}\rangle}$. Let $\sigma \geq \sigma_1$ and let $\pi$ 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 is changed. Otherwise $\Delta^{\langle\sigma+1\rangle} = \Delta^{\langle\bar{\pi}+1\rangle} \leq \tau_3 \Delta^{\langle\bar{\pi}\rangle} = \tau_3 \Delta^{\langle\sigma\rangle}$. This proves that $(\Delta^{\langle\sigma\rangle})_{\sigma\geq\sigma_1}$ is non-increasing. In particular, $\Delta^{\langle\bar{\pi}+1\rangle} \leq \tau_3 \Delta^{\langle\bar{\pi}\rangle} \leq \tau_3 \Delta^{\langle\underline{\pi}\rangle}$, so (Upd2) also holds for all these processes, proving $\lim_{\sigma\geq\sigma_1} \Delta^{\langle\sigma\rangle} = 0$ by Lemma 9. □

**Corollary 11.** *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*

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

*where $\Sigma^{dc}$ is the sequence of descent steps according to Lemma 8. 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 $\Sigma^*$.*

**Proof.** Depending on whether an infinite number of descent steps occurs or not, the claim follows either from Lemma 8 or Lemma 9 (or Observation 7 for $|\Sigma| < \infty$). The last statement follows from the fact $f(\hat{y}^{\langle\sigma\rangle}) - \hat{f}_{\bar{w}^{\langle\sigma\rangle}}(\hat{y}^{\langle\sigma\rangle}) \geq 0$ and $\|\bar{g}^{\langle\sigma\rangle}\| \geq 0$. □

**Theorem 12.** *Suppose the set of optimisers* $\operatorname{Arg\,min} f$ *is nonempty and bounded. If* $\Sigma$ *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 7, so assume $|\Sigma| = \infty$. Let $f^* := \min\{f(y)\colon y \in \mathbb{R}^M\}$. The boundedness of the level set $\{y\colon f(y) \leq f^*\}$ implies the boundedness of all level sets, particularly of the set $\mathcal{S} := \{y\colon f(y) \leq f(\hat{y}^{\langle 0\rangle})\}$. Because $(f(\hat{y}^{\langle\sigma\rangle}))_\sigma$ is non-increasing, see Observation 5, we have $\hat{y}^{\langle\sigma\rangle} \in \mathcal{S}$ 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 $\Sigma'$ 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

$$f(y) \geq \hat{f}_{\bar{w}^{\langle\sigma\rangle}}(y) = \underbrace{f(\hat{y}^{\langle\sigma\rangle})}_{\xrightarrow{\Sigma'} f(y^*)} - \underbrace{\langle \bar{g}^{\langle\sigma\rangle}, y - \hat{y}^{\langle\sigma\rangle}\rangle}_{\xrightarrow{\Sigma'} 0 \text{ by Corollary 11}} \overset{\Sigma'}{\to} f(y^*),$$

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

# 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 straight forward subspace problem for a process $\pi$ working on an affine subspace indexed by $J^{(\pi)} \subseteq M$ is

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

*i. e.*, we just restrict the optimisation to the affine subspace $\mathcal{L}^{(\pi)} = \mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle})$. Given a centre $\hat{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$ 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

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

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

$$\Delta(\bar{w}, \hat{y}^{(\pi)}) = \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}, \hat{y}^{(\pi)}) + \frac{1}{u}\|\bar{g}^{(\pi)}_{M\setminus J^{(\pi)}}\|^2. \tag{7}$$

A bundle method solving $(\mathit{Sub}^{\pi}_{\mathrm{s}})$ 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 $E \subseteq \{(i,j)\colon i,j \in M, i \neq j\}$. 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 $F_D^{\mathrm{gra}}$. In particular, let

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

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

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

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 $F_D^{\mathrm{gra}}$ satisfies the requirements for dependent subspace functions.

We will use the dependency graph based function $F_D^{\mathrm{gra}}$ 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 $F_d^{\mathrm{s}}$ 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

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

The subspace problem is solved by a bundle method with centre $\hat{y}^{(\pi)} = \hat{y}^{\langle\bar{\pi}\rangle}$, 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

$$\Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{(\pi)}) \geq \varrho_1 \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{\langle\bar{\pi}\rangle}, \hat{y}^{\langle\bar{\pi}\rangle}) \text{ and } f(\hat{y}^{\langle\bar{\pi}\rangle}) - f(\bar{y}^{(\pi)}) \geq \varrho_2 \Delta_{\mathrm{s}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{(\pi)}), \tag{Upd1$_{\mathrm{s}}$}$$

or

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

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.

We investigate the case that the subprocess stopped because of (Upd2$_{\mathrm{s}}$). In view of (7), condition (Upd2) might not hold if $\frac{1}{u}\|\bar{g}^{\langle\bar{\pi}+1\rangle}_{M\setminus J^{(\pi)}}\|^2 > \frac{1}{u}\|\bar{g}^{\langle\bar{\pi}\rangle}_{M\setminus J^{(\pi)}}\|^2$. 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 $\varrho_3 \in (0, 1-\varrho_1)$,

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

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

$$(j,j') \in \operatorname{Arg\,max}\left\{(\bar{g}_{\hat{j}}^{\langle\bar{\pi}+1\rangle})^2 - (\bar{g}_{\hat{j}}^{\langle\underline{\pi}\rangle})^2 \colon (\tilde{j},\hat{j}) \in (J^{(\pi)} \times (M \setminus J^{(\pi)})) \setminus S^{\langle\bar{\pi}\rangle}\right\}, \tag{8}$$

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).

```
Algorithm 3: UpdateSubspace^s
  Input    : $\pi$, final candidate $\bar{y}^{(\pi)} \in \mathcal{L}^{(\pi)}$, aggregate minorant $\bar{w}^{(\pi)} \in W$
  Changes  : global data at $\bar{\pi}+1$
  if $\pi$ 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 // $\pi$ stopped because of (Upd2$_s$)
      $\bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}^{(\pi)}$, $\hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}^{\langle\bar{\pi}\rangle}$
      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 (8)
      else $S^{\langle\bar{\pi}+1\rangle} \leftarrow S^{\langle\bar{\pi}\rangle}$
```

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

**Lemma 13.** *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 $\pi$ 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 $\pi$ with $\bar{\pi} \geq \sigma_1$ satisfies (Upd2).*

**Proof.** We start with (i). By (Upd1$_\mathrm{s}$) there holds with the subspace selection condition (Sel1)

$$f(\hat{y}^{\langle\underline{\pi}\rangle}) - f(\bar{y}^{(\pi)}) \geq \varrho_2\Delta_\mathrm{s}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle}) \geq \varrho_2\varrho_1\Delta_\mathrm{s}^{(\pi)}(\bar{w}^{\langle\underline{\pi}\rangle}, \hat{y}^{\langle\underline{\pi}\rangle}) \geq \varrho_2\varrho_1\tau_1\Delta^{\langle\underline{\pi}\rangle} \geq \tau_2\Delta^{\langle\underline{\pi}\rangle}.$$

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 $\Sigma$ 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 $\sigma_0$ be the index defined in Observation 5. 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 $\sigma_1$ a process $\pi$ 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 stopped because of (Upd2$_\mathrm{s}$) without satisfying (Dep$_\mathrm{s}$). Then with (7) it follows

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

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 optimising over the whole space (only the steps are restricted to the subspace, but the function evaluations remain the same). In particular, each process $\pi$ 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 $R$ 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

$$\begin{aligned} \hat{f}_w(y) &= \sum_{r \in R} \hat{f}_{w_r,r}(y) = l + \langle g, y\rangle, \\ \hat{f}_{w_r,r}(y) &= l_r + \langle g_r, y\rangle, \end{aligned}$$

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 $\pi$ selecting coordinates $J^{(\pi)}$

$$R^{(\pi)} := R_{J^{(\pi)}}, \qquad \bar{J}^{(\pi)} := \bigcup_{r \in R^{(\pi)}} J_r \setminus J^{(\pi)}, \tag{9}$$

then for a given centre $\hat{y} \in \mathbb{R}^M$ 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 optimise 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.

$$\begin{aligned} &\text{minimise} \quad f^{(\pi)}(y) := \sum_{r \in R^{(\pi)}} f_r(y), \\ &\text{subject to} \quad y \in \mathcal{L}^{(\pi)}, \end{aligned} \tag{$Sub_{\mathrm{p}}^{\pi}$}$$

with $\mathcal{L}^{(\pi)} = \mathcal{L}(J^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle})$. The set of minorants is given by $W^{(\pi)} = \bigtimes_{r \in R^{(\pi)}} W_r$. A centre 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

$$\Delta(\bar{w}, \hat{y}) = \Delta^{(\pi)}_{\mathrm{p}}(\bar{w}_{R^{(\pi)}}, \hat{y}) + \bar{\Delta}^{(\pi)}_{\mathrm{p}}(\bar{w}_{R \setminus R^{(\pi)}}, \hat{y}) + \frac{1}{u} \|\bar{g}_{\bar{J}^{(\pi)}}\|^2, \tag{10}$$

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 14.** *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 $f^{(\pi)}(y) = f^{(\pi)}(y')$ and $\Delta^{(\pi)}_{p}(w_{R^{(\pi)}}, y) = \Delta^{(\pi)}_{p}(w_{R^{(\pi)}}, y')$.*

*(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 (9) 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 ($Sub^{\pi}_{\mathrm{p}}$) is directly reflected in the term $\Delta_{\mathrm{p}}^{(\pi)}$ if the components $\bar{J}^{(\pi)}$ are kept fix. This motivates the following weakly dependent subspace function (it satisfies conditions (1)–(4))

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

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

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

and

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

Employed within a successful subspace selection step of Algorithm 2 for $\pi$ this yields

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

and ensures a “disjointness” of parallel subspace problems.

**Observation 15.** *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 $j \in J^{(\eta)} \cup \bar{J}^{(\eta)} = F_d^{\mathrm{p}}(S, J^{(\eta)})$. This is a contradiction to (5), which holds by Observation 6. □

For subspace optimisation 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

$$\Delta_{\mathrm{p}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle}) \geq \varrho_1 \Delta_{\mathrm{p}}^{(\pi)}(\bar{w}_{R^{(\pi)}}^{\langle\underline{\pi}\rangle}, \hat{y}^{\langle\underline{\pi}\rangle}) \text{ and } f^{(\pi)}(\hat{y}^{\langle\underline{\pi}\rangle}) - f^{(\pi)}(\bar{y}^{(\pi)}) \geq \varrho_2 \Delta_{\mathrm{p}}^{(\pi)}(\bar{w}^{(\pi)}, \hat{y}^{\langle\underline{\pi}\rangle}), \tag{Upd1$_\mathrm{p}$}$$

or

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

Observation 14 and (10) indicate that the critical term, which may cause an increase in the global progress even if $\Delta_{\mathrm{p}}^{(\pi)}$ is reduced, is the last of (10). Therefore, the subprocess will use the following dependency test for some $\varrho_3 \in (0, 1 - \varrho_1)$

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

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

$$(j, j') \in \operatorname{Arg\,max} \Big\{ (\bar{g}_{\hat{\jmath}}^{\langle\bar{\pi}+1\rangle})^2 - (\bar{g}_{\hat{\jmath}}^{\langle\bar{\pi}\rangle})^2 \colon (\tilde{\jmath}, \hat{\jmath}) \in (J^{(\pi)} \times \bar{J}^{(\pi)}) \setminus S^{\langle\bar{\pi}\rangle} \Big\} \tag{11}$$

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 centre does not change on $J^{(\pi)} \cup \bar{J}^{(\pi)}$ as long as $\pi$ runs and changes in $J^{(\pi)}$ at most when $\pi$ finishes.

**Algorithm 4:** UPDATESUBSPACE$^{\mathrm{P}}$

```
Input    : $\pi$, 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 $\pi$ stopped because of (Upd$1_p$) 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 // $\pi$ stopped because of (Upd$2_{\mathrm{p}}$)
    $(\bar{w}^{\langle\bar{\pi}+1\rangle}_{R^{(\pi)}}, \bar{w}^{\langle\bar{\pi}+1\rangle}_{R\setminus R^{(\pi)}}) \leftarrow (\bar{w}^{(\pi)}, \bar{w}^{\langle\bar{\pi}\rangle}_{R\setminus R^{(\pi)}})$
    $\hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}^{\langle\bar{\pi}\rangle}$
    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 (11)
    else  $S^{\langle\bar{\pi}+1\rangle} \leftarrow S^{\langle\bar{\pi}\rangle}$
```

**Corollary 16.** *Let $\pi \in \bar{\Pi}^{\langle\infty\rangle}$, 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 first statement relies on the subspace selection in conjunction with the update of the centre in Algorithm 4. Assume, for contradiction, $\hat{y}_j^{\langle\underline{\pi}\rangle} \neq \hat{y}_j^{\langle\bar{\pi}\rangle}$ for some $j \in J^{(\pi)} \cup \bar{J}^{(\pi)} = F_d^{\langle\underline{\pi}\rangle}(J^{(\pi)})$. Then there must be a process $\eta$ with $\underline{\pi} < \bar{\eta} < \bar{\pi}$ and $\hat{y}_j^{\langle\bar{\eta}\rangle} \neq \hat{y}_j^{\langle\bar{\eta}+1\rangle}$. However, in all cases $\hat{y}_{j'}^{\langle\bar{\eta}\rangle} = \hat{y}_{j'}^{\langle\bar{\eta}+1\rangle}$ for $j' \notin J^{(\eta)}$, hence $j \in J^{(\eta)}$. This implies $j \in J^{(\eta)} \cap F_d^{\langle\underline{\pi}\rangle}(J^{(\pi)}) \neq \emptyset$, contradicting (5) by Observation 6 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 17.** *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 $\pi$ terminates with (Upd$1_p$) 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 $\pi$ with $\bar{\pi} \geq \sigma_1$ satisfies (Upd2) and (Upd2').*

*(iii) All processes satisfy (Inv2).*

**Proof.** By Corollary 16 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 $\pi$ stops with (Upd$1_{\mathrm{p}}$). Observation 14 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

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

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 5 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 $\pi$ with $\bar{\pi} \geq \sigma''$ we have $\underline{\pi} \geq \sigma'$.

We use the following short notation for any process $\pi$ 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 $\pi$ be an arbitrary process with $\bar{\pi} \geq \sigma''$. Then $\pi$ satisfies $\Delta_\pi^{\langle\bar{\pi}+1\rangle} \overset{(\mathrm{Upd2_p})}{<} \varrho_1 \Delta_\pi^{\langle\underline{\pi}\rangle}$ and $(\mathrm{Dep_p})$ does not hold, hence by (10) and Observation 14

$$\begin{aligned}
\Delta^{\langle\bar{\pi}\rangle} - \Delta^{\langle\bar{\pi}+1\rangle} &= (\Delta_\pi^{\langle\bar{\pi}\rangle} - \Delta_\pi^{\langle\bar{\pi}+1\rangle}) + (\bar{\Delta}_\pi^{\langle\bar{\pi}\rangle} - \bar{\Delta}_\pi^{\langle\bar{\pi}+1\rangle}) + \frac{1}{u}\left(\|\bar{g}_{\bar{J}(\pi)}^{\langle\bar{\pi}\rangle}\|^2 - \|\bar{g}_{\bar{J}(\pi)}^{\langle\bar{\pi}+1\rangle}\|^2\right) \\
&= (\Delta_\pi^{\langle\underline{\pi}\rangle} - \Delta_\pi^{\langle\bar{\pi}+1\rangle}) + \frac{1}{u}\left(\|\bar{g}_{\bar{J}(\pi)}^{\langle\bar{\pi}\rangle}\|^2 - \|\bar{g}_{\bar{J}(\pi)}^{\langle\bar{\pi}+1\rangle}\|^2\right) \\
&\geq (1 - \varrho_1 - \varrho_3)\Delta_\pi^{\langle\underline{\pi}\rangle} \\
&\geq \underbrace{\tau_1(1 - \varrho_1 - \varrho_3)}_{=:\tau\in(0,1)} \Delta^{\langle\underline{\pi}\rangle} \geq 0.
\end{aligned} \tag{12}$$

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 $\pi$ with $\bar{\pi} \geq \sigma_1$. Consequently, relation (12) implies for each process $\pi$ 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 18.** Note that the main algorithm requires the objective value in the current centre $f(\hat{y}^{\langle\sigma\rangle})$ 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 centre is changed by a process $\pi$ from $\hat{y}^{\langle\bar{\pi}\rangle}$ to $\hat{y}^{\langle\bar{\pi}+1\rangle}$. This follows from the proof of Observation 5, 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 $\pi$ 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 generalise 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 19.** Consider Lagrangian relaxation decoupling a problem of the form

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

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.*, [5] 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 optimiser $\hat{x}_r$ influenced by $J'_r := \{i \in J_r \colon [A_r \hat{x}_r]_i \neq 0 \lor i \in J\}$. 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 an $\hat{y}'$ 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 optimiser $\hat{x}'_r$ 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 20.** 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

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

We say the (global) data at $\sigma$ is 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 21.** For the packing type application described in Example 19 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 optimiser $\hat{x}_r$ exists with $[A_r \hat{x}_r]_{M \setminus J} = 0$. Therefore starting with $J \leftarrow J'_r$ and iterating the process of finding $\hat{x}_r \in \operatorname{Arg\,max}_{x_r \in \mathcal{X}_r}(c_r^T x_r - \hat{y}^T_{|J} A_r x_r)$, 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 22.** *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.** 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 $y_{J_r^{\langle\underline{\pi}\rangle}}$. In order to emphasise this difference in the objects related to a process $\pi$ we now use a superscript $[\pi]$ in brackets. In particular, let $J^{[\pi]}$ be the subspace associated with a process $\pi$, then we define

$$\begin{aligned} J_r^{[\pi]} &:= J_r^{\langle\underline{\pi}\rangle} \cap J^{[\pi]}, & \bar{J_r}^{[\pi]} &:= J_r^{\langle\underline{\pi}\rangle} \setminus J^{[\pi]}, \\ \bar{J}^{[\pi]} &:= \bigcup_{r \in R^{[\pi]}} \bar{J_r}^{[\pi]}, & R^{[\pi]} &:= \left\{ r \in R \colon J_r^{\langle\underline{\pi}\rangle} \cap J^{[\pi]} \neq \emptyset \right\} \\ f_r^{[\pi]}(y) &:= f_r(y_{|J_r^{\langle\underline{\pi}\rangle}}), & f^{[\pi]}(y) &:= \sum_{r \in R^{[\pi]}} f_r^{[\pi]}(y). \end{aligned} \tag{13}$$

Using these notations the subspace problem associated with process $\pi$ reads

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

A centre 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

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

where

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

This gives rise to the expected progress

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

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

**Observation 23.** *For $\pi \in \bar{\Pi}^{\langle\infty\rangle}$, $\hat{y} \in \mathbb{R}^M$, $\bar{w} \in W$ with $(\hat{y}, \bar{w}_r = (\bar{l}_r, \bar{g}_r), J_r^{\langle\pi\rangle})$ consistent for $r \in R$ there holds*

$$\Delta(\bar{w}, \hat{y}) = \Delta^{[\pi]}(\bar{w}_{R^{[\pi]}}, \hat{y}) + \bar{\Delta}^{[\pi]}(\bar{w}, \hat{y}) + \frac{1}{u}\|\bar{g}_{\bar{J}^{[\pi]}}\|^2, \tag{14}$$

*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

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

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 is 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 centre is not changed on $J^{[\pi]} \cup \bar{J}^{[\pi]}$ by other processes as long as $\pi$ 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

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

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

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

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

**Observation 24.** *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.** 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 (5), which holds by Observation 6.□

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

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

or

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

Observation 23 indicates that the critical term is again the third one of (14), so that the subprocess will use the following dependency test for some $\varrho_3 \in (0, 1 - \varrho_1)$

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

If condition (Dep$_\mathrm{x}$) is fulfilled, the dependency graph is enlarged by adding an edge according to (11).

There is one important difference to 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 centre cannot change on $J^{[\pi]} \cup \bar{J}^{[\pi]}$ as long as $\pi$ runs and may only change in $J^{[\pi]}$ when $\pi$ finishes.

**Corollary 25.** *Let $\pi$ 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 $j \notin J^{[\pi]}$.*

**Algorithm 5:** UpdateSubspace$^{\mathrm{x}}$

```
    Input    : $\pi$, 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 $\pi$ stopped because of (Upd1$_x$) then
1       $\hat{y}' \leftarrow \hat{y}^{\langle\bar{\pi}\rangle} + (\bar{y}^{[\pi]} - \hat{y}^{\langle\underline{\pi}\rangle})$
        $(\bar{w}'_{R^{[\pi]}}, \bar{w}'_{R\setminus R^{[\pi]}}) \leftarrow (\bar{w}^{[\pi]}, \bar{w}^{\langle\bar{\pi}\rangle}_{R\setminus R^{[\pi]}})$
        // update consistency information
2       for $r \in R$ do  $J_r^{\langle\bar{\pi}+1\rangle} \leftarrow \mathcal{C}_r(\hat{y}', \bar{w}', J_r^{\langle\bar{\pi}\rangle})$
        // no changes in $\mathcal{S}^{\mathtt{gra}}$-part, $S^{\langle\bar{\pi}\rangle} \rightsquigarrow S^{\langle\bar{\pi}+1\rangle}$
3       if $J_r^{\langle\bar{\pi}+1\rangle} \neq J_r^{\langle\underline{\pi}\rangle}$ for some $r \in R$ then // update $S^{\langle\bar{\pi}+1\rangle} \succ S^{\langle\underline{\pi}\rangle}$
            $\hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}^{\langle\bar{\pi}\rangle}$, $\bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}^{\langle\bar{\pi}\rangle}$
        else
            $\hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}'$, $\bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}'$
    else // $\pi$ stopped because of (Upd2$_\mathrm{x}$)
        $\hat{y}^{\langle\bar{\pi}+1\rangle} \leftarrow \hat{y}^{\langle\bar{\pi}\rangle}$
4       $(\bar{w}'_{R^{[\pi]}}, \bar{w}'_{R\setminus R^{[\pi]}}) \leftarrow (\bar{w}^{[\pi]}, \bar{w}^{\langle\bar{\pi}\rangle}_{R\setminus R^{[\pi]}})$
5       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 $S^{\langle\bar{\pi}+1\rangle} \succ S^{\langle\underline{\pi}\rangle}$
            for $r \in R^{[\pi]}$ do  $J_r^{\langle\bar{\pi}+1\rangle} \leftarrow J_r^{\langle\bar{\pi}\rangle} \cup \{j \in J_r \setminus J_r^{\langle\bar{\pi}\rangle} : (\bar{g}'_r)_j \neq 0\}$
            // no changes in $\mathcal{S}^{\mathtt{gra}}$-part, $S^{\langle\bar{\pi}\rangle} \rightsquigarrow S^{\langle\bar{\pi}+1\rangle}$
            $\bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}^{\langle\bar{\pi}\rangle}$
        else
            if (Dep$_x$) holds then // update $S^{\langle\bar{\pi}+1\rangle} \succ S^{\langle\underline{\pi}\rangle}$
                add $(j, j')$ as defined in (11) to the $\mathcal{S}^{\mathrm{gra}}$-part of $S^{\langle\bar{\pi}\rangle} \rightsquigarrow S^{\langle\bar{\pi}+1\rangle}$
            else $S^{\langle\bar{\pi}+1\rangle} \leftarrow S^{\langle\bar{\pi}\rangle}$
            $\bar{w}^{\langle\bar{\pi}+1\rangle} \leftarrow \bar{w}'$
```

**Proof.** The proof is analogous to the proof of Corollary 16. □

**Observation 26.** *The global data is consistent for all $\sigma \in \Sigma$.*

**Proof.** Initialisation ensures consistency for $\sigma = 0$. Inductively, relevant information for consistency is only changed in Algorithm 5. Observation 22 ensures that active subspaces may be increased safely if neither centre nor aggregate is changed. In the (Upd2$_\mathrm{x}$) branch, A5.5 ensures that aggregates are only stored if they satisfy consistency, the centre remains unchanged. In the (Upd1$_\mathrm{x}$) 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 centre and aggregate only in case of proved consistency. □

**Lemma 27.** *If $\pi$ stops because of (Upd1$_x$) 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 $\pi$ be a process that stops because of ($\mathrm{Upd1_x}$). 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 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 (13) 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 25 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 (13) of $f_r^{[\pi]}$, the fact that consistency holds by (i) and Observation 26 prove (ii) and (iii). □

**Lemma 28.** *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 terminates with (Upd1$_x$), then an infinite number of processes satisfies (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 $\pi$ 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 5 there is an $\sigma_0 \in \Sigma$ so that $S^{\langle\sigma\rangle} = S^{\langle\sigma_0\rangle}$ for all $\sigma \geq \sigma_0$. Hence, for any process $\pi$ 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 $\pi$ be a process terminating with ($\mathrm{Upd1_x}$). If $S^{\langle\underline{\pi}\rangle} \neq S^{\langle\bar{\pi}+1\rangle}$ then $\pi$ does not modify the centre, 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 27 implies

$$\begin{aligned}
f(\hat{y}^{\langle\bar{\pi}\rangle}) - f(\hat{y}^{\langle\bar{\pi}+1\rangle}) &= \sum_{r \in R^{[\pi]}} \Big( f_r(\hat{y}^{\langle\bar{\pi}\rangle}) - f_r(\hat{y}^{\langle\bar{\pi}+1\rangle}) \Big) + \sum_{r \in R \setminus R^{[\pi]}} \underbrace{\Big( f_r(\hat{y}^{\langle\bar{\pi}\rangle}) - f_r(\hat{y}^{\langle\bar{\pi}+1\rangle}) \Big)}_{=0} \\
&= \sum_{r \in R^{[\pi]}} \Big( f_r^{[\pi]}(\hat{y}^{\langle\underline{\pi}\rangle}) - f_r^{[\pi]}(\bar{y}^{[\pi]}) \Big) = f^{[\pi]}(\hat{y}^{\langle\underline{\pi}\rangle}) - f^{[\pi]}(\bar{y}^{[\pi]}) \\
&\overset{(\mathrm{Upd1_x})}{\geq} \varrho_2 \Delta^{[\pi]}(\bar{w}^{[\pi]}, \hat{y}^{\langle\underline{\pi}\rangle}) \overset{(\mathrm{Upd1_x})}{\geq} \varrho_2\varrho_1 \Delta^{[\pi]}(\bar{w}^{\langle\underline{\pi}\rangle}_{R^{[\pi]}}, \hat{y}^{\langle\underline{\pi}\rangle}) \\
&\overset{(Sel1)}{\geq} \varrho_2\varrho_1\tau_1 \Delta^{\langle\underline{\pi}\rangle} \geq \tau_2 \Delta^{\langle\underline{\pi}\rangle} \geq 0.
\end{aligned}$$

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

Now let $\pi$ be a process terminating with ($\mathrm{Upd2_x}$). 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 ($\mathrm{Upd1_x}$). Exactly like in the proof of Lemma 17 we get a $\sigma_1 \in \Sigma$ so that each process $\pi$ with $\bar{\pi} \geq \sigma_1$ stops with ($\mathrm{Upd2_x}$), there holds $S^{\langle\underline{\pi}\rangle} = S^{\langle\bar{\pi}+1\rangle}$ and the dependency test ($\mathrm{Dep_x}$) 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 to employ Observation 23 and to follow exactly the same steps as in the proof of Lemma 17. □

**Remark 29.** Analogously to the partial separable case, there is no need to evaluate $f(\hat{y}^{\langle\sigma\rangle})$ when the centre 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 centre.

**Acknowledgement.** We thank an anonymous referee for suggesting to attempt an extension of [2] to general convex functions. This work was supported by the *Bundesministerium für Bildung und Forschung* under grant 05M10OCD. Responsibility for the content rests with the authors.

## References

[1] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical Optimization.* Springer, 2003.

[2] 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, D-09107 Chemnitz, Germany, 2012.

[3] M. L. Fisher. The Lagrangian relaxation method for solving integer programming problems. *Manage. Sci.*, 50(12):1861–1871, 2004.

[4] M. Fukushima. Parallel variable transformation in unconstrained optimization. *SIAM J. on Optim.*, 8(3):658–672, 1998.

[5] C. Helmberg and K. Kiwiel. A spectral bundle method with bounds. *Mathematical Programming*, 93:173–194, 2002.

[6] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I & II*, volume 305, 306 of *Grundlehren der mathematischen Wissenschaften.* Springer, Berlin, Heidelberg, 1993.

[7] C. Lemaréchal. Lagrangian relaxation. In M. Jünger and D. Naddef, editors, *Computational Combinatorial Optimization*, volume 2241 of *Lecture Notes in Computer Science*, pages 112–156. Springer Berlin / Heidelberg, 2001.

[8] C. Lemaréchal. The omnipresence of Lagrange. *Annals of Operations Research*, 153(1):9–27, 2007.

[9] D. Medhi. Parallel bundle-based decomposition for large-scale structured mathematical programming problems. *Ann. Oper. Res.*, 22(1-4):101–127, 1990.

[10] A. Nedić, D. Bertsekas, and V. Borkar. Distributed asynchronous incremental subgradient methods. In Y. C. Dan Butnariu and S. Reich, editors, *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, volume 8 of *Studies in Computational Mathematics*, pages 381–407. Elsevier, 2001.

[11] C. A. Sagastizábal and M. V. Solodov. Parallel variable distribution for constrained optimization. *Comput. Optim. Appl.*, 22(1):111–131, 2002.