TECHNISCHE UNIVERSITÄT
CHEMNITZ

A Parallel Bundle Method for Asynchronous Subspace Optimization in Lagrangian Relaxation

F. Fischer, C. Helmberg

Preprint 2012-2

Fakultät für Mathematik

Impressum:
Herausgeber:
Der Dekan der
Fakultät für Mathematik
an der Technischen Universität Chemnitz
Sitz:
Reichenhainer Strae 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 Method for Asynchronous Subspace Optimization in Lagrangian Relaxation

Frank Fischer* Christoph Helmberg*

February 14, 2012

Abstract. An algorithmic approach is proposed for exploiting parallelization possibilities in large scale optimization models of the following generic type. Objects change their state over time subject to a limited availability of common resources. These are modeled by linear coupling constraints and result in few objects competing for the same resource at each point in time.

In a kind of asynchronous parallel coordinate descent, each independent process iteratively picks a free subset of violated constraints together with their interacting objects, improves the corresponding Lagrange multipliers by a bundle method to a certain level, and stores observed presumable dependencies leading to increased violation of other constraints in a common dependency graph. These dependencies have to be respected in future subset selections. No synchronization is required between the processes, for each subproblem the number of evaluations may differ arbitrarily. Under the assumption of boundedness of the set of dual optimizers we prove convergence of appropriate subsequences of the iterates to primal and dual optimal solutions of the relaxation. Preliminary computational results indicate that this approach may develop into a viable alternative to classical bundle methods using parallel evaluations.
Keywords: bundle methods, parallel programming, Lagrangian relaxation
MSC 2010: 90C06; 65Y05, 90C25, 65K05

1. Introduction

We consider the problem of solving structured optimization problems that arise from Lagrangian relaxation of linear constraints that couple a number of well solvable basic problems. Formally, given a finite number $\omega \in \mathbb{N}$ of compact ground sets $\Omega_v \subset \mathbb{R}^{n_v}$ with $n_v \in \mathbb{N}$ and $n := \sum_{v \in V} n_v$ for $v \in V := \{1, \dots, \omega\}$, the coupled problem reads

$$(P) \qquad \begin{array}{ll} \text{maximize} & h(x) := \sum_{v \in V} h_v(x_v) \\ \text{subject to} & Ax = b, \\ & x = (x_1^T, \dots, x_\omega^T)^T \in \Omega := \bigotimes_{v \in V} \Omega_v, \end{array}$$

where $h_v \colon \Omega_v \to \mathbb{R}$, $v \in V$, are arbitrary functions, $A \in \mathbb{R}^{M \times n}$, $b \in \mathbb{R}^M$ with constraint set $M = \{1, \dots, m\}$.

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

Problems of this kind appear generically in connection with multicommodity flow models in scheduling and planning applications. In such problems a sequence of valid states has to be determined for different objects (indexed by V) like vehicles, robots, machines where each object requires some resources (*e. g.*, route capacity, time slot restrictions, ...) in each state. The schedules therefore have to be chosen so that the resource constraints are not violated. For each object the sequence of states over time is then represented by some time expanded network and the resource restrictions are modeled using coupling equalities and inequalities (indexed by M). In such applications Lagrangian relaxation and decomposition has proven to be a valuable tool, see, *e. g.*, [5, 15, 16] for an overview on Lagrangian relaxation and applications, [4, 11] for applications in train timetabling and inventory management, [1] for further models based on multicommodity flows, or [23] for stochastic programming. The increasing problem sizes in this area as well as the availability of cheap parallel computing hardware require the development of decomposition techniques not only on the modeling but also on the algorithmic side. We contribute to this by proposing an asynchronous parallel approach for Lagrangian relaxation under the rather natural assumption in this setting, that most of the coupling constraints either couple only few subproblems or, if they couple many, coupling actually affects just a few subproblems that try to use the corresponding resource at the same time. In this, we assume without loss of generality (w. l. o. g.) that each row of A has at least one non-zero entry, *i. e.*,

$$A_{j,\bullet} \neq 0 \text{ for all } j \in M. \tag{1.1}$$

Lagrangian relaxation is applicable if for each $v \in V$ and all $y \in \mathbb{R}^M$ the subproblem

$$(P_v(y)) \qquad \begin{array}{rl} \text{maximize} & L_v(x_v, y) := h_v(x_v) - y^T A_{\bullet,v} x_v \\ \text{subject to} & x_v \in \Omega_v \end{array} \tag{1.2}$$

can be easily solved via an oracle that returns the optimal value

$$f_v(y) := \max_{x_v \in \Omega_v} L_v(x_v, y) \tag{1.3}$$

and an optimal solution $\hat{x}_v(y)$. The standard algorithmic approach is then to compute an upper bound for (P) by solving the Lagrangian Dual problem of (P)

$$(D) \qquad \min_{y \in \mathbb{R}^M} f(y) := b^T y + \sum_{v \in V} f_v(y)$$

via a subgradient or bundle method. Note that the optimal value of (D) may be larger than the optimal value of (P) in general because the functions h_v, $v \in V$, may be arbitrary and the sets Ω_v may be non-convex, in many applications even discrete. In fact, the optimal value of (D) is equal to the optimal value of the following "convexified" relaxation of the primal problem (P)

$$(\operatorname{conv} P) \qquad \begin{array}{rl} \text{maximize} & \bar{h}(x) := \sum_{v \in V} \bar{h}_v(x_v) \\ \text{subject to} & Ax = b \\ & x = (x_1^T, \ldots, x_\omega^T)^T, x_v \in \operatorname{conv} \Omega_v, v \in V, \end{array}$$

where for each $v \in V$ the function $\bar{h}_v \colon \operatorname{conv} \Omega_v \to \mathbb{R}$ is the negative closed convex hull of $-h$ with respect to (w. r. t.) $\operatorname{conv} \Omega_v$, *i. e.*, $\operatorname{epi}(-\bar{h}_v) = \overline{\operatorname{conv}} \operatorname{epi}(-h_v)$, see, *e. g.*, [14], Theorem 2.12.

We propose to extend this Lagrangian approach to a fully parallel approach by iteratively identifying subsets of y-coordinates that can be optimized in parallel in a form of asynchronous parallel coordinate descent, so that overall convergence to primal and dual optimal solutions is still guaranteed whenever the set of optimizers of (D) is bounded.

Here is a brief sketch of the algorithmic idea. A master process sets up the initial global data and then starts new parallel processes whenever less than some predefined number of processes are running. Each single process executes the following same four steps for solving at most one subspace optimization problem.

(a) **Subspace selection:** identify a "free" subset $J \subseteq M$ together with subproblems $V_J \subseteq V$ interacting with J, that promises a significant portion of a global predicted decrease measure, and block both in the global data against being used in other processes. If no appropriate subspace J is available, wait for the global data to change and then execute this step again.

(b) **Set up the subspace problem:** create the data for the subspace problem. Given the current global point $\hat{y}$, the subproblem changes y_J and uses fixed values $\hat{y}_{M\setminus J}$,

$$\min_{y_J \in \mathbb{R}^J} \left[b_J^T y_J + \sum_{v \in V_J} \max_{x_v \in \Omega_v} \left(h_v(x_v) - \hat{y}_{M\setminus J}^T A_{M\setminus J,v} x_v - y_J^T A_{J,v} x_v \right) \right].$$

(c) **Solve the subspace problem:** perform the steps of a bundle method until a certain stopping criterion is met.

(d) **Update global data with subspace solution:** check for progress impeding dependencies between J and $M \setminus J$, store relevant solution and dependency information in the global data, and free the subspace afterwards. If a certain global termination criterion is satisfied, stop all process and the entire algorithm, otherwise stop this process only (or continue as a new process with step (a)).

The subspace selection in step (a) will start with a single promising seed coordinate $j \in M$ that is then possibly extended to a larger subset J according to simple rules. Improving Lagrange multipliers on this subset may lead to increased violation of constraints in $M \setminus J$ and this may destroy convergence. If in step (d) the process detects increases in violation of significant size, these presumable dependencies are stored in a global dependency graph as arcs pointing form the seed j to the adverse coordinates. Future subspace selections in step (a) with seed j will be forced to include all these presumably dependent coordinates, as well. Note that this dynamic construction of the dependency graph requires no a priori knowledge about the problem.

If dependencies are strong and the graph forces $J = M$ eventually, the algorithm reduces to the standard sequential bundle method. In large scale applications, however, strong dependencies should only exist between rather small groups of multipliers and objects so that several processes can be started on more or less independent subproblems. In particular, the algorithm will be able to make good use of parallelism if the sets J and corresponding coupled subsets of V remain small. We will therefore present two algorithmic variants. The first and simpler variant assumes rather loose coupling in the sense that each constraint should couple only a few objects of V. This setting is well suited for getting acquainted with the basic mechanisms of our asynchronous parallel framework. For the second variant we consider the more realistic scenario that most indices $j \in M$ represent a resource constraint limiting the use of a specific resource at a given point in time to a few of many potential objects. In an actual solution most of these objects will not make use of the corresponding resource at this specific point in time. In the algorithm the states of the objects are represented by a convex combination of primal solutions and with respect to a current convex combination of primal solutions the corresponding Lagrange multiplier then typically affects only a few objects. Therefore, we keep track of actual constraint-object interactions by dynamically collecting for each $v \in V$ all constraints j that were "visited" by primal solutions of v. In the subspace selection step (a) of the second variant the selection of $J \subseteq M$ then entails the selection of all $v \in V$ that interacted with one of the constraints in J before and the update step (d) will have to make sure, that this selection was in fact sufficient to guarantee correctness and convergence.

In the proofs of convergence of the two algorithmic variants the main work is to prove correctness of the asynchronous updating schemes, which requires a fair amount of book-keeping. Once correctness is ensured, the finiteness of the dependency graph and the interaction sets allow to employ the standard bundle convergence mechanisms more or less directly (see, *e. g.*, [2] for an introduction and [13] for a detailed exposition) if boundedness of the set of dual optimizers is assumed. Without this assumption convergence is open and more elaborate techniques might be required. Further relevant questions open to further investigation concern the possibility to allow deletions in the dependency graph or to delete entries in the interaction sets. The extension to primal inequalities (or dual box constraints), however, should pose no major difficulties, see [10].

In order to highlight the difference of our asynchronous scheme to existing approaches we shortly review parallel algorithms of the literature. The classic parallelization approach for solving problem (D) via subgradient optimization is to exploit the decomposing structure of the problem by organizing the algorithm in a queen-worker approach. Here a single queen process does the global iteration, updating the current point in the dual space y, the evaluation of the dual function $f(y) = b^T y + \sum_{v \in V} f_v(y)$ is then distributed over several worker processes where each subproblem is solved independently by some of the workers. A computational study of this approach for a bundle algorithm can be found in [17].

Another class of algorithms are variable transformation algorithms for unconstrained [6] and constrained [22] optimization problems. In each iteration they select a set of subspaces either along coordinate directions [22] or more general directions [6] such that the subspaces span the whole space and then compute a new candidate point on each subspace where each subspace problem can be solved in an independent process. Afterwards a new global iterate is derived from the candidate points. Usually these approaches require smoothness of the objective function but can also be applied to non-smooth optimization problems using smoothing techniques as, *e. g.*, the Moreau-Yosida regularization [18].

A successful approach to non-smooth optimization uses incremental subgradient methods. These methods change y in each major iteration incrementally through a sequence of ω steps. In each step y is modified according to a subgradient direction of a single subproblem. In [21] this approach is extended to a fully parallel and asynchronous approach where the sequences in which the subproblem computations are started and in which y is modified may differ. In particular, when the candidate computed by a subproblem is considered as a new point, the current point y may differ from the point when that subproblem was started (in contrast to the non-parallel case). [21] proves the convergence of this approach under the assumption that the number of updates of y between the start of the subproblem's processing and the consideration of the subproblem's result are bounded by a constant. This implies that the overall number of evaluations is asymptotically equal for each subproblem, which is also the case for the synchronized parallel approaches above.

Other parallel approaches exploit the explicit static structure of convex optimization problems. Those so called splitting methods solve easier subproblems generated by the corresponding augmented Lagrangian function, which may be solved alternating or in parallel, and combine the results to solutions of the original problem in some iterative scheme, see, *e. g.*, [7, 8, 9, 19, 20].

The subspace selection of our approach is somewhat similar to the variable distribution approaches [3, 6, 22], but in contrast we do not require to select a set of subspaces that span the whole space. Furthermore there is no global synchronization step as in [7, 8] and no requirement on how often a certain subproblem must be evaluated. Indeed, some subproblems may require considerably fewer evaluations than others and this is a main source of efficiency. In contrast to the algorithms proposed in the papers above, there is no regularization condition or synchronization to ensure global convergence. Instead, convergence is guaranteed only by dependency analysis between the subspace evaluations.

This paper is organized as follows. In Section 2 we introduce the notation and the basic bundle framework used throughout the paper. Afterwards in Section 3 we develop the basic parallel bundle algorithm, which is designed to work on loosely coupled problems where each constraint acts only on a small number of subproblems. This approach is then extended in Section 4 to stronger coupled systems with additional structure so that active influence of the multipliers on the solution of the decoupled subproblems can be determined dynamically by the algorithm. We conclude the paper with some numerical tests in Section 5 comparing the parallel bundle algorithm proposed in this paper with the classical bundle method. These tests are very preliminary as it is not the focus of this paper to provide an extensive numerical study of the proposed algorithms. Still, the tests seem to indicate that the parallel bundle algorithm is superior to the classical bundle method for some problem classes of practical relevance and can therefore provide a useful alternative in applications.

2. General Setting

The *Lagrangian* function $L : \Omega \times \mathbb{R}^m \to \mathbb{R}$,

$$L(x, y) := h(x) + (b - Ax)^T y = b^T y + \sum_{v \in V} L_v(x_v, y), \tag{2.1}$$

gives rise to the respective (convex) dual functions, as introduced before,

$$f_v(y) = \max_{x_v \in \Omega_v} L_v(x_v, y) \quad (v \in V) \quad \text{and} \quad f(y) = b^T y + \sum_{v \in V} f_v(y) = \max_{x \in \Omega} L(x, y). \tag{2.2}$$

For $x \in \Omega$ the functions $L_v(x_v, \cdot)$ $(v \in V)$ and $L(x, \cdot)$ are linear in y with gradients

$$g_v(x_v) := -A_{\bullet,v} x_v \quad (v \in V) \quad \text{and} \quad g(x) := b - Ax = b + \sum_{v \in V} g_v(x_v), \tag{2.3}$$

so each point

$$\begin{aligned} w_v \in W_v &:= \{(l_v, x_v) \in \mathbb{R} \times \Omega_v : l_v = h_v(x_v)\} \quad (v \in V), \\ w \in W &:= \{(l, x) \in \mathbb{R}^V \times \Omega \colon (l_v, x_v) = w_v \in W_v, v \in V\} \end{aligned}$$

generates linear minorants

$$\begin{aligned} \hat{f}_{w_v,v}(y) &:= l_v + g_v(x_v)^T y \leq f_v(y), \\ \hat{f}_w(y) &:= \sum_{v \in V} l_v + g(x)^T y = b^T y + \sum_{v \in V} \hat{f}_{w_v,v}(y) \leq f(y). \end{aligned} \tag{2.4}$$

Convex combinations $w_v \in \operatorname{conv} W_v$ $(v \in V)$ and $w \in \operatorname{conv} W$ also yield such minorants.

A bundle method optimizing $f(y)$ over the full space $\mathbb{R}^m$ (see, *e. g.*, [12, 13]) starts at a given *center of stability* $\hat{y} \in \mathbb{R}^m$ and forms, given a compact $\widehat{W} \subseteq \operatorname{conv} W$, a model

$$\hat{f}_{\widehat{W}}(y) := \sup_{(l,x)\in\widehat{W}} \hat{f}_{(l,x)}(y) \leq f(y).$$

The bundle method [13] determines, for a given *weight* $u > 0$ and an augmenting term $\frac{u}{2}\|y - \hat{y}\|^2$ that penalizes points far from $\hat{y}$, the next candidate

$$\bar{y} = \operatorname*{argmin}_{y\in\mathbb{R}^M} \left[\hat{f}_{\widehat{W}}(y) + \tfrac{u}{2}\|y - \hat{y}\|^2\right].$$

For this candidate it checks whether the actual progress $f(\hat{y}) - f(\bar{y})$ is good in comparison to the predicted decrease $f(\hat{y}) - \hat{f}_{\widehat{W}}(\bar{y})$. If so, a *descent step* is made by setting $\hat{y} \leftarrow \bar{y}$. Otherwise, in a *null step*, the center is left unchanged but the model $\hat{f}_{\widehat{W}}$ is modified to improve the model in $\bar{y}$.

In our setting the compactness requirement on $\widehat{W}$ implies the following equality for the bundle subproblem [12]

$$\inf_{y\in\mathbb{R}^M} \sup_{w\in\widehat{W}} [\hat{f}_w(y) + \tfrac{u}{2}\|y - \hat{y}\|^2] = \sup_{w\in\operatorname{conv}\widehat{W}} \inf_{y\in\mathbb{R}^M} [\hat{f}_w(y) + \tfrac{u}{2}\|y - \hat{y}\|^2].$$

The inner optimization problem of the right hand side can be solved explicitly for given $w = (l, x) \in \operatorname{conv}\widehat{W}$,

$$\bar{y}(w) = \hat{y} - \tfrac{1}{u}g(x) = \hat{y} - \tfrac{1}{u}(b - Ax).$$

Therefore $\bar{y}$ is found by determining a (not necessarily unique) *primal aggregate*

$$\bar{w} = (\bar{l}, \bar{x}) \in \operatorname*{Argmax}_{(l,x)\in\operatorname{conv}\widehat{W}} \left[\sum_{v\in V} l_v + g(x)^T\hat{y} - \tfrac{1}{2u}\|g(x)\|^2\right].$$

The new objective value in $\bar{y} = \bar{y}(\bar{w})$ promised by this aggregate minorant is

$$\hat{f}_{\widehat{W}}(\bar{y}) = \hat{f}_{\bar{w}}(\bar{y}) = \hat{f}_{\bar{w}}(\hat{y}) - \tfrac{1}{u}\|g(\bar{x})\|^2. \tag{2.5}$$

Its difference to the objective value in the center is the *predicted decrease*

$$\Delta(\hat{y}, (\bar{l}, \bar{x})) := f(\hat{y}) - \hat{f}_{(\bar{l},\bar{x})}(\hat{y}) + \tfrac{1}{u}\|g(\bar{x})\|^2. \tag{2.6}$$

3. An Asynchronous Parallel Framework for Loose Coupling

Rather than optimizing over the full space the idea is to optimize in parallel over several subspaces. For this we will often refer to certain subsets of dual or primal variables. Let $V' \subseteq V$ and $J \subseteq M$ be subsets of the subproblems and constraints, resp., and let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^M$ be a primal and a dual vector. Then $x_{V'} = (x_v)_{v\in V'}$ and $y_J := (y_j)_{j\in J}$ denote the subvectors with index blocks in V' and indices in J. Similarly for a matrix $A \in \mathbb{R}^{M\times n}$ the matrix $A_{J,V'} \in \mathbb{R}^{J\times n_{V'}}$, $n_{V'} = \sum_{v\in V'} n_v$, denotes the submatrix with elements contained in rows J and column blocks corresponding to V'. If one of the sets contains only one element or one block, *i. e.*, $J = \{j\}$ or $V' = \{v\}$, we write as usual x_v, y_j or $A_{j,v}$.

The selection of subspaces will depend on the structure of the coupling constraints. We collect the dependencies of subproblems and constraints in the following notation illustrated in Fig. 1.

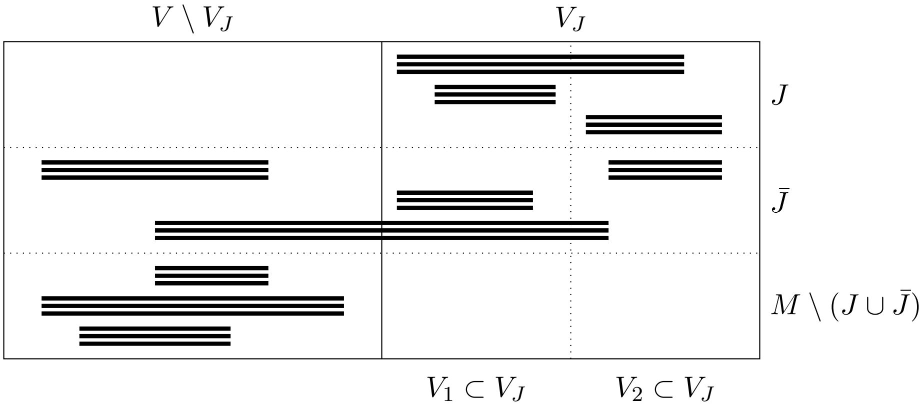


Figure 1: For $J \subseteq M$, V_J consists of all v interacting with some $j \in J$, $\bar{J}$ are the remaining constraints interacting with some $v \in V_J$.

Definition 1 Let $v \in V$, $j \in M$, and $J \subseteq M$. Then

$$J_v := \{j \in M \colon A_{j,v} \neq 0\}, \quad \text{the set of all constraints interacting with } v, \tag{3.1}$$
$$V_j := \{v \in V \colon j \in J_v\}, \quad \text{the set of all subproblems interacting with } j, \tag{3.2}$$
$$V_J := \bigcup_{j \in J} V_j, \quad \text{the set of all subproblems interacting with } J, \tag{3.3}$$
$$\bar{J} := \bigcup_{v \in V_J} (J_v \setminus J), \quad \text{constraints in } M \setminus J \text{ interacting with some } v \in V_J. \tag{3.4}$$

These definitions imply

$$\forall\, v \in V, \forall\, j \in M\colon \quad j \in J_v \Leftrightarrow v \in V_j, \tag{3.5}$$
$$\forall\, v \in V, \forall\, y \in \mathbb{R}^M\colon \quad (A_{J_v,v})^T y_{J_v} = (A_{\bullet,v})^T y, \tag{3.6}$$
$$\forall\, v \in V, \forall\, y, y' \in \mathbb{R}^M\colon \quad y_{J_v} = y'_{J_v} \;\Rightarrow\; f_v(y) = f_v(y'), \tag{3.7}$$
$$\forall\, v \in V, \forall\, w_v \in \operatorname{conv} W_v, \forall\, y, y' \in \mathbb{R}^M\colon \quad y_{J_v} = y'_{J_v} \;\Rightarrow\; \hat{f}_{w_v,v}(y) = \hat{f}_{w_v,v}(y'), \tag{3.8}$$
$$\forall\, J \subseteq M, \forall\, x, x' \in \mathbb{R}^n\colon \quad x_{V_J} = x'_{V_J} \Rightarrow g(x)_J = g(x')_J. \tag{3.9}$$

The algorithm starts with the initialization of some global data and then acts as a master process. This master process continues starting several parallel processes as long as less than some predefined number of processes are running. Each process executes the steps (a)–(d) outlined in the introduction. In this, steps (a) and (d) interact with global data and (b) and (c) do not interact with global data. For data consistency, at most one process is allowed to be in a step that interacts with global data. This is achieved using semaphores which block a process whenever it tries to start a globally interacting step if another process is currently dealing with global data. Once a process finished a step interacting with global data, one of the blocked processes (if there is one) is allowed to proceed. Because of this organization there is a unique sequence of global interactions. We denote this sequence by a global index marker $\sigma \in \mathbb{N}_0$ which is increased each time some elements of the global data are modified.

Throughout the algorithm the following objects are maintained globally, with an index marker $\sigma \in \mathbb{N}_0$ indicating the value after the σ-th writing access to the global data:

$\hat{y}^{\langle\sigma\rangle} \in \mathbb{R}^M$,	the current global center,
$f_v^{\langle\sigma\rangle} := f_v(\hat{y}^{\langle\sigma\rangle}) \in \mathbb{R}$,	the optimal primal value of $P_v(\hat{y}^{\langle\sigma\rangle})$ attained in some $\hat{x}_v^{\langle\sigma\rangle} \in \Omega_v$ for $v \in V$,
$(\bar{l}^{\langle\sigma\rangle}, \bar{x}^{\langle\sigma\rangle}) \in \operatorname{conv} W$,	the current global aggregate minorant,
$B^{\langle\sigma\rangle} \subseteq V$,	set of primal problems currently blocked by some processes,
$D^{\langle\sigma\rangle} = (M, E^{\langle\sigma\rangle})$,	digraph with arc set $E^{\langle\sigma\rangle}$ collecting presumed dependencies.

The global variables $\hat{y}^{\langle\sigma\rangle}$ and $\bar{x}^{\langle\sigma\rangle}$ correspond to the current dual and primal variables. Like for a standard bundle method, we will see that under reasonable assumptions each accumulation point of $(\hat{y}^{\langle\sigma\rangle})_\sigma$ is a dual optimal solution and there is a suitable subsequence of $(\bar{x}^{\langle\sigma\rangle})_\sigma$ converging to a primal optimal solution of (conv P).

Given the current center of stability $\hat{y}^{\langle\sigma\rangle} \in \mathbb{R}^M$ and the primal aggregate $\bar{x}_v^{\langle\sigma\rangle} \in \operatorname{conv} \Omega_v$ for all $v \in V$, the intuition underlying step (a) is to choose an initial coordinate $\hat{\jmath}$ and enlarge it to a subspace $J \subseteq M$ so that the subproblems V_J interacting with J strongly violate the coupling constraints $A_{J,V_J} \bar{x}_{V_J}^{\langle\sigma\rangle} = b$. The process will then try to improve the Lagrange multipliers $\hat{y}_J$ associated with the subspace J in order to reduce this violation. As long as this process runs, no other process is allowed to change any of the primal variables associated with V_J (for this, $B^{\langle\sigma\rangle}$ keeps track of all selected sets V_J) or the dual variables associated with $J \cup \bar{J}$ but other subspaces without such interaction may be selected and processed at the same time. If the sets V_j, $j \in J$, are relatively small subsets of V, *i. e.*, each constraint couples only a small number of subproblems (which is a reasonable assumption in many large scale applications), several disjoint subsets may be selected simultaneously.

For the selection of an appropriate subspace we investigate the contribution of a subset J to the predicted decrease Δ of (2.6),

$$\Delta^{\langle\sigma\rangle} := \sum_{v \in V} \left[f_v^{\langle\sigma\rangle} - \hat{f}_{(\bar{l}_v^{\langle\sigma\rangle}, \bar{x}_v^{\langle\sigma\rangle}),v}(\hat{y}^{\langle\sigma\rangle}) \right] + \tfrac{1}{u} \| g(\bar{x}^{\langle\sigma\rangle}) \|^2, \tag{3.10}$$

$$\Delta_J^{\langle\sigma\rangle} := \sum_{v \in V_J} \left[f_v^{\langle\sigma\rangle} - \hat{f}_{(\bar{l}_v^{\langle\sigma\rangle}, \bar{x}_v^{\langle\sigma\rangle}),v}(\hat{y}^{\langle\sigma\rangle}) \right] + \tfrac{1}{u} \| g(\bar{x}^{\langle\sigma\rangle})_J \|^2, \tag{3.11}$$

$$\bar{\Delta}_J^{\langle\sigma\rangle} := \sum_{v \in V \setminus V_J} \left[f_v^{\langle\sigma\rangle} - \hat{f}_{(\bar{l}_v^{\langle\sigma\rangle}, \bar{x}_v^{\langle\sigma\rangle}),v}(\hat{y}^{\langle\sigma\rangle}) \right] + \tfrac{1}{u} \| g(\bar{x}^{\langle\sigma\rangle})_{M \setminus (J \cup \bar{J})} \|^2, \tag{3.12}$$

$$\delta_{\bar{J}}^{\langle\sigma\rangle} := \tfrac{1}{u} \| g(\bar{x}^{\langle\sigma\rangle})_{\bar{J}} \|^2.$$

The value $\Delta_J^{\langle\sigma\rangle}$ may be interpreted as the predicted decrease if only the variables $\hat{y}_J$ along with the primal variables $\bar{x}_{V_J}$ are allowed to change and all other variables are fixed. The value $\bar{\Delta}_J^{\langle\sigma\rangle}$ is the predicted decrease that cannot be influenced by changing $\hat{y}_J$ or $\bar{x}_{V_J}$, and $\delta_{\bar{J}}^{\langle\sigma\rangle}$ is the residual part that can be influenced by both, the subproblems V_J and the subproblems not interacting with J, namely $V \setminus V_J$. The following observation justifies this.

Observation 2 *For $J \subseteq M$ there holds $\Delta^{\langle\sigma\rangle} = \Delta_J^{\langle\sigma\rangle} + \bar{\Delta}_J^{\langle\sigma\rangle} + \delta_{\bar{J}}^{\langle\sigma\rangle}$ and if $f_v^{\langle\sigma\rangle} = f_v(\hat{y}^{\langle\sigma\rangle})$ for all $v \in V$, then $\Delta^{\langle\sigma\rangle} = \Delta(\hat{y}^{\langle\sigma\rangle}, (\bar{l}^{\langle\sigma\rangle}, \bar{x}^{\langle\sigma\rangle}))$.*

Proof. This follows from (2.2), (2.4) and (2.6) by direct computation. □

If $\Delta_J^{\langle\sigma\rangle}$ represents a significant portion of $\Delta^{\langle\sigma\rangle}$ and the process's attempt to reserve J and V_J as its subspace in step (a) is *successful*, it then optimizes over J and V_J by a bundle method thereby driving Δ_J to zero while not changing $\bar{\Delta}_J$, but this only helps to reduce the global predicted decrease $\Delta^{\langle\sigma\rangle}$ if $\delta_{\bar{J}}$ does not increase too much. If $\delta_{\bar{J}}$ develops unfavorably, say due to some coordinate $j^* \in \bar{J}$, a new edge $(\hat{\jmath}, j^*)$ will be inserted into $E^{\langle\sigma\rangle}$ in step (d) in order to memorize that future selections starting with index $\hat{\jmath}$ should always include index j^*, as well.

We point out two facts. First, there may be some processes which do never update the global data, either because the subspace selection in step (a) fails or because the algorithm stops while the process has not yet completed step (d). Because those processes have no relevant influence on the global data, we will mostly ignore them in the following (as long as we can guarantee that at least one process is running, see Observation 9 below). In contrast, a relevant process (one with a successful subspace selection step and a completed step (d)) interacts exactly twice with the global data, namely at the steps (a) and (d). Second, note that for a relevant process the global data at the beginning of step (d) may have changed since the process has set up its subproblem in step (b) because the processes run in parallel without synchronization and some other process may have changed the global data in the meantime.

Because at most one process is allowed to interact with the global data at the same time, there is a well-defined sequence of global interactions and we can identify these global states using the global counter $\sigma \in \mathbb{N}_0$. Furthermore each relevant process interacts exactly twice with the global data and it will be convenient to identify the status of each process via these two associated index markers. For this purpose each process π is equipped with two labels, $\underline{\pi}$ and $\overline{\pi}$. In initialization these labels are set to $\underline{\pi} = -1 = \overline{\pi}$. Whenever the process tries to acquire a subproblem in step (a), the label $\underline{\pi}$ is set to the current value of σ. If π succeeds in reserving a subspace at $\underline{\pi} = \sigma$ then this will be the final value of $\underline{\pi}$, the new status is indicated by setting $\overline{\pi} \leftarrow \infty$ and σ is increased. Once the process completed its subproblem and enters step (d), the label $\overline{\pi}$ is assigned the new current value of σ as its final value and σ will be increased again. In consequence, a process π with $\overline{\pi} \leq \underline{\pi}$ is waiting for a suitable subproblem, a process with $\overline{\pi} = \infty$ is currently working on its subproblem, and if $\underline{\pi} < \overline{\pi} < \infty$ the process has done its work.

3.1. The Subspace Problem Associated with Process π

Each process π solves a subproblem in a certain subspace indexed by a subset of M acquired in step (a). Each subspace problem itself is again a convex optimization problem of the same kind as the global dual problem (D). We mark all information that is associated with a specific process π with the superscript (π). In particular, we will use the following notation,

$J^{(\pi)} \subseteq M$,	the subset selected in step (a) that π may modify,	
$V^{(\pi)} := V_{J^{(\pi)}}$,	the subproblems that π optimizes,	(3.13)
$\bar{J}^{(\pi)} := \bigcup_{v \in V^{(\pi)}} J_v \setminus J^{(\pi)}$,	dual variables interacting with $V^{(\pi)}$ but not in $J^{(\pi)}$,	(3.14)
$\Omega^{(\pi)} = \bigotimes_{v \in V^{(\pi)}} \Omega_v$,	the primal ground set.	

The subspace $J^{(\pi)}$ determines everything else except for the constant cost coefficients $c^{(\pi)}$ defined next and to be determined in step (b) w. r. t. the global data:

$$c_v^{(\pi)} := -(A_{\bar{J}^{(\pi)},v})^T \hat{y}_{\bar{J}^{(\pi)}}^{\langle\pi\rangle} \qquad \text{for all } v \in V^{(\pi)}. \tag{3.15}$$

These give rise to modified cost functions

$$h_v^{(\pi)}(x_v) := h_v(x_v) + (c_v^{(\pi)})^T x_v \qquad \text{for } x_v \in \Omega_v, v \in V^{(\pi)}. \tag{3.16}$$

With the same notation now equipped with superscript (π),

$$L_v^{(\pi)}(x_v, y^{(\pi)}) := h_v^{(\pi)}(x_v) - (y^{(\pi)})^T A_{J^{(\pi)},v} x_v, \quad x_v \in \Omega_v, v \in V^{(\pi)}, y^{(\pi)} \in \mathbb{R}^{J^{(\pi)}}, \tag{3.17}$$

$$f_v^{(\pi)}(y^{(\pi)}) := \max_{x_v \in \Omega_v} L_v^{(\pi)}(x_v, y^{(\pi)}), \quad v \in V^{(\pi)}, y^{(\pi)} \in \mathbb{R}^{J^{(\pi)}}, \tag{3.18}$$

$$f^{(\pi)}(y^{(\pi)}) := b_{J^{(\pi)}}^T y^{(\pi)} + \sum_{v \in V^{(\pi)}} f_v^{(\pi)}(y^{(\pi)}), \quad y^{(\pi)} \in \mathbb{R}^{J^{(\pi)}}, \tag{3.19}$$

the subproblem that π solves is

$$(SP^{(\pi)}) \qquad \min_{y^{(\pi)} \in \mathbb{R}^{J^{(\pi)}}} f^{(\pi)}(y^{(\pi)}).$$

Introducing the rest of the notation in the same vein, let

$$W^{(\pi)} := \left\{ (l^{(\pi)}, x^{(\pi)}) \in \mathbb{R}^{V^{(\pi)}} \times \Omega^{(\pi)} : l_v^{(\pi)} = h_v(x_v^{(\pi)}), x_v^{(\pi)} \in \Omega_v, v \in V^{(\pi)} \right\}.$$

The process π manages a local version of the global data, namely,

$$\begin{aligned} \hat{y}^{(\pi)} &\in \mathbb{R}^{J^{(\pi)}}, && \text{the local center,} \\ (\bar{l}^{(\pi)}, \bar{x}^{(\pi)}) &\in \operatorname{conv} W^{(\pi)}, && \text{the local aggregate minorant.} \end{aligned}$$

Process π solves the problem approximately using a bundle method. Each point $(\bar{l}^{(\pi)}, \bar{x}^{(\pi)}) \in \operatorname{conv} W^{(\pi)}$ defines a minorant of $f^{(\pi)}$,

$$\hat{f}^{(\pi)}_{(\bar{l}^{(\pi)}, \bar{x}^{(\pi)})}(y^{(\pi)}) := b_{J^{(\pi)}}^T y^{(\pi)} + \sum_{v \in V^{(\pi)}} \hat{f}^{(\pi)}_{(\bar{l}_v^{(\pi)}, \bar{x}_v^{(\pi)}),v}(y^{(\pi)}) \le f^{(\pi)}(y^{(\pi)})$$

where

$$\hat{f}^{(\pi)}_{(\bar{l}_v^{(\pi)}, \bar{x}_v^{(\pi)}),v}(y^{(\pi)}) := \bar{l}_v^{(\pi)} + (c_v^{(\pi)})^T \bar{x}_v^{(\pi)} - (y^{(\pi)})^T A_{J^{(\pi)},v} \bar{x}_v^{(\pi)}. \tag{3.20}$$

The gradient of $\hat{f}^{(\pi)}_{(\bar{l}^{(\pi)}, \bar{x}^{(\pi)})}$ can be worked out to be

$$g^{(\pi)}(\bar{x}^{(\pi)}) := b_{J^{(\pi)}} - A_{J^{(\pi)},V^{(\pi)}} \bar{x}^{(\pi)}. \tag{3.21}$$

Indeed, the blocks of A appearing here but not in the gradients of (3.20) are zero blocks by definition of J_v and do not influence the result.

Given a center $\hat{y}^{(\pi)}$, an aggregate minorant $(\bar{l}^{(\pi)}, \bar{x}^{(\pi)}) \in \operatorname{conv} W^{(\pi)}$ and the minimizer

$$\bar{y}^{(\pi)} := \hat{y}^{(\pi)} - \tfrac{1}{u} g^{(\pi)}(\bar{x}^{(\pi)})$$

of the augmented model $\hat{f}^{(\pi)}_{(\bar{l}^{(\pi)}, \bar{x}^{(\pi)})}(y^{(\pi)}) + \frac{u}{2}\|y^{(\pi)} - \hat{y}^{(\pi)}\|^2$ as candidate, the *predicted decrease* of this subproblem is

$$\Delta^{(\pi)}(\hat{y}^{(\pi)}, (\bar{l}^{(\pi)}, \bar{x}^{(\pi)})) := f^{(\pi)}(\hat{y}^{(\pi)}) - \hat{f}^{(\pi)}_{(\bar{l}^{(\pi)}, \bar{x}^{(\pi)})}(\hat{y}^{(\pi)}) + \tfrac{1}{u}\|g^{(\pi)}(\bar{x}^{(\pi)})\|^2. \tag{3.22}$$

The definitions above immediately imply the following relations.

Observation 3

$$J_v \subseteq J^{(\pi)} \cup \bar{J}^{(\pi)} \qquad \textit{for all } v \in V^{(\pi)}, \tag{3.23}$$

$$J_v \cap J^{(\pi)} = \emptyset \qquad \textit{for all } v \in V \setminus V^{(\pi)}, \tag{3.24}$$

$$V_j \cap V^{(\pi)} = \emptyset \qquad \textit{for all } j \in M \setminus (J^{(\pi)} \cup \bar{J}^{(\pi)}). \tag{3.25}$$

Proof. Directly from the definitions. □

3.2. Linking Local and Global Information

There is a direct relation between the subproblems $L_v^{(\pi)}$, $v \in V^{(\pi)}$, of π and the global subproblems L_v, $v \in V^{(\pi)}$, as well as between the local and global function values, models, and gradients.

Observation 4 *Let $\pi = (\underline{\pi}, \overline{\pi})$ be a process, $v \in V^{(\pi)}$ and $y \in \mathbb{R}^M$ so that $y_{\bar{J}^{(\pi)}} = \hat{y}^{\langle \pi \rangle}_{\bar{J}^{(\pi)}}$.*

$$L_v^{(\pi)}(x_v, y_{J^{(\pi)}}) = L_v(x_v, y) \qquad \textit{for all } x_v \in \Omega_v, \tag{3.26}$$

$$f_v^{(\pi)}(y_{J^{(\pi)}}) = f_v(y), \tag{3.27}$$

$$\hat{f}^{(\pi)}_{(l_v,x_v),v}(y_{J^{(\pi)}}) = \hat{f}_{(l_v,x_v),v}(y) \qquad \textit{for all } (l,x) \in \operatorname{conv} W, \tag{3.28}$$

$$g^{(\pi)}(x_{V^{(\pi)}}) = g(x)_{J^{(\pi)}} \qquad \textit{for all } (l,x) \in \operatorname{conv} W. \tag{3.29}$$

Proof. There hold $J_v \subseteq J^{(\pi)} \cup \bar{J}^{(\pi)}$ by (3.23) and $A_{M \setminus J_v, v} = 0$ by (3.1), so

$$\begin{aligned} c_v^{(\pi)} - A^T_{J^{(\pi)},v} y_{J^{(\pi)}} &\overset{(3.15)}{=} -A^T_{\bar{J}^{(\pi)},v} \hat{y}^{\langle \pi \rangle}_{\bar{J}^{(\pi)}} - A^T_{J^{(\pi)},v} y_{J^{(\pi)}} \\ &= -A^T_{\bar{J}^{(\pi)},v} y_{\bar{J}^{(\pi)}} - A^T_{J^{(\pi)},v} y_{J^{(\pi)}} = -A^T_{\bullet,v} y. \end{aligned}$$

Thus, for $x_v \in \Omega_v$,

$$\begin{aligned} L_v^{(\pi)}(x_v, y_{J^{(\pi)}}) &\overset{(3.16)(3.17)}{=} h_v(x_v) + (c_v^{(\pi)})^T x_v - y^T_{J^{(\pi)}} A_{J^{(\pi)},v} x_v \\ &= \; h_v(x_v) - y^T A_{\bullet,v} x_v \overset{(1.2)}{=} L_v(x_v, y). \end{aligned}$$

This proves (3.26) and also (3.27) by (1.3) and (3.18). For (3.28),

$$\begin{aligned} \hat{f}^{(\pi)}_{(l_v,x_v),v}(y_{J^{(\pi)}}) &\overset{(3.20)}{=} l_v + (c_v^{(\pi)})^T x_v - (y_{J^{(\pi)}})^T A_{J^{(\pi)},v} x_v \\ &= l_v - y^T A_{\bullet,v} x_v \overset{(2.4)}{=} \hat{f}_{(l_v,x_v),v}(y). \end{aligned}$$

Finally, (3.29) relies on $A_{j, V \setminus V_j} = 0$ and $V_j \subseteq V^{(\pi)}$ for $j \in J^{(\pi)}$ by (3.13),

$$\begin{aligned} g^{(\pi)}(x_{V^{(\pi)}}) &\overset{(3.21)}{=} b_{J^{(\pi)}} - A_{J^{(\pi)},V^{(\pi)}} x_{V^{(\pi)}} - \underbrace{A_{J^{(\pi)},V \setminus V^{(\pi)}}}_{=0 \text{ by } (3.13)} x_{V \setminus V^{(\pi)}} \\ &= b_{J^{(\pi)}} - A_{J^{(\pi)},\bullet} x \overset{(2.3)}{=} g(x)_{J^{(\pi)}}. \end{aligned}$$

□

3.3. The Parallel Bundle Algorithm for Loose Coupling

Next we present the parallel bundle algorithm in detail. In this algorithm each process π is allowed to modify the global data on its associated subspace $J^{(\pi)}$ and for the subproblems $V^{(\pi)}$, but while π is running no other process may change this global data or the multipliers corresponding to $\bar{J}^{(\pi)}$. The importance of the latter requirement is clearly visible in Observation 4 and is achieved by condition (S3) in the algorithm below.

Algorithm *Parallel Bundle:*

1. **Initialization**
   - Choose parameters

- *relative predicted subspace decrease level* $\tau_1 \in (0, \frac{1}{2}\min\{\frac{1}{|M|}, \frac{1}{|V|}\})$,
- *subspace improvement level* $\tau_2 \in (0, 1)$,
- *subspace dependency level* $\tau_3 \in [0, 1 - \tau_2)$,
- *termination precision* $\varepsilon > 0$,
- *maximal number of parallel processes* $N_\Pi \le |M|$,
- *weight* $u > 0$ and *descent step parameter* $\rho \in (0, 1)$ for the bundle method.

- Set $\sigma \leftarrow 0$, $\hat{y}^{\langle 0 \rangle} \leftarrow 0$, $B^{\langle 0 \rangle} \leftarrow \emptyset$.
- Set $E^{\langle 0 \rangle} \leftarrow \emptyset$ (or use some prespecified dependencies).
- Determine all $f_v^{\langle 0 \rangle} = f_v(\hat{y}^{\langle 0 \rangle})$ and set $(\bar{l}^{\langle 0 \rangle}, \bar{x}^{\langle 0 \rangle})$ to the minorant defined by the optimal solutions.
- Compute $\Delta^{\langle 0 \rangle}$. If $\Delta^{\langle 0 \rangle} \le \varepsilon(|f(\hat{y}^{\langle 0 \rangle})| + 1)$ then do not start any process and **STOP**.

2. While less than N_Π processes are running, start a new process π setting $\underline{\pi} \leftarrow -1$, $\overline{\pi} \leftarrow -1$. Each process π does the following (with semaphores guarding the access).

(a) **Subspace selection**

- Secure exclusive access to global data, set $\underline{\pi} \leftarrow \sigma$.
- Find $\hat{\jmath} \in M$ and a subspace $J^{(\pi)} \subseteq M$, $\hat{\jmath} \in J^{(\pi)}$ with
  (S1) $\{j : (\hat{\jmath}, j) \in E^{\langle \underline{\pi} \rangle}\} \subseteq J^{(\pi)}$,
  (S2) $\Delta_{J^{(\pi)}}^{\langle \underline{\pi} \rangle} \ge \tau_1 \Delta^{\langle \underline{\pi} \rangle}$,
  (S3) $V^{(\pi)} \cap B^{\langle \underline{\pi} \rangle} = \emptyset$.
- If no such $\hat{\jmath}$ exists, the step is *unsuccessful*, do not change the global data, free access to it and restart at some later time at (a) when $\sigma > \underline{\pi}$.
- The step is *successful*, $B^{\langle \sigma+1 \rangle} \leftarrow B^{\langle \underline{\pi} \rangle} \cup V^{(\pi)}$, $\overline{\pi} \leftarrow \infty$ and keep

$$\left(\hat{y}^{\langle \sigma+1 \rangle}, (\bar{l}^{\langle \sigma+1 \rangle}, \bar{x}^{\langle \sigma+1 \rangle}), f_V^{\langle \sigma+1 \rangle}, E^{\langle \sigma+1 \rangle}\right) \leftarrow \left(\hat{y}^{\langle \underline{\pi} \rangle}, (\bar{l}^{\langle \underline{\pi} \rangle}, \bar{x}^{\langle \underline{\pi} \rangle}), f_V^{\langle \underline{\pi} \rangle}, E^{\langle \underline{\pi} \rangle}\right). \tag{3.30}$$

- Read required global data, set $\sigma \leftarrow \underline{\pi} + 1$, and free exclusive access.

(b) **Set up the subspace problem.**
There is no global interaction in this step!
Compute $c_v^{(\pi)}$ by (3.15) for all $v \in V^{(\pi)}$ and initialize the bundle method to solve problem $(SP^{(\pi)})$ with this $c^{(\pi)}$ starting from center $\hat{y}^{(\pi)} = \hat{y}_{J^{(\pi)}}^{\langle \underline{\pi} \rangle}$ and initial aggregate $(\bar{l}^{(\pi)}, \bar{x}^{(\pi)}) = (\bar{l}_{V^{(\pi)}}^{\langle \underline{\pi} \rangle}, \bar{x}_{V^{(\pi)}}^{\langle \underline{\pi} \rangle})$. In the first iteration we have then exactly a predicted decrease

$$\Delta^{(\pi)}(\hat{y}^{(\pi)}, (\bar{l}^{(\pi)}, \bar{x}^{(\pi)})) = \Delta_{J^{(\pi)}}^{\langle \underline{\pi} \rangle}.$$

(c) **Solve the subspace problem.**
There is no global interaction in this step!
Solve $(SP^{(\pi)})$ by the bundle method, iteratively generating candidates $\bar{y}^{(\pi)}$ together with primal aggregates $(\bar{l}^{(\pi)}, \bar{x}^{(\pi)}) \in \operatorname{conv} W^{(\pi)}$. In each iteration test the following conditions in this sequence.

1. If the predicted decrease on the subspace $J^{(\pi)}$ has been sufficiently reduced, *i. e.*,

$$\Delta^{(\pi)}(\hat{y}^{(\pi)}, (\bar{l}^{(\pi)}, \bar{x}^{(\pi)})) < \tau_2 \Delta_{J^{(\pi)}}^{\langle \underline{\pi} \rangle}, \tag{StopP}$$

go to (d).

2. If $\Delta^{(\pi)}(\hat{y}^{(\pi)}, (\bar{l}^{(\pi)}, \bar{x}^{(\pi)})) \geq \tau_2 \Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}$ and the descent step criterion

$$\Delta^{(\pi)}(\hat{y}^{(\pi)}, (\bar{l}^{(\pi)}, \bar{x}^{(\pi)})) \leq \tfrac{1}{\rho}(f^{(\pi)}(\hat{y}^{(\pi)}) - f^{(\pi)}(\bar{y}^{(\pi)})) \tag{StopD}$$

is satisfied, set $\hat{y}^{(\pi)} \leftarrow \bar{y}^{(\pi)}$ and go to (d).

(d) **Update global data with subspace solution.**

- Secure exclusive access to global data, set $\overline{\pi} \leftarrow \sigma$.
- Independent of (StopP) and (StopD), keep

$$\left(\hat{y}^{\langle\sigma+1\rangle}_{M\setminus J^{(\pi)}}, f^{\langle\sigma+1\rangle}_{V\setminus V^{(\pi)}}, (\bar{l}^{\langle\sigma+1\rangle}_{V\setminus V^{(\pi)}}, \bar{x}^{\langle\sigma+1\rangle}_{V\setminus V^{(\pi)}})\right) \leftarrow \left(\hat{y}^{\langle\overline{\pi}\rangle}_{M\setminus J^{(\pi)}}, f^{\langle\overline{\pi}\rangle}_{V\setminus V^{(\pi)}}, (\bar{l}^{\langle\overline{\pi}\rangle}_{V\setminus V^{(\pi)}}, \bar{x}^{\langle\overline{\pi}\rangle}_{V\setminus V^{(\pi)}})\right) \tag{3.31}$$

and update the subterms

$$(\bar{l}^{\langle\sigma+1\rangle}_{V^{(\pi)}}, \bar{x}^{\langle\sigma+1\rangle}_{V^{(\pi)}}) \leftarrow (\bar{l}^{(\pi)}_{V^{(\pi)}}, \bar{x}^{(\pi)}_{V^{(\pi)}}), \tag{3.32}$$

$$B^{\langle\sigma+1\rangle} \leftarrow B^{\langle\overline{\pi}\rangle} \setminus V^{(\pi)}, \tag{3.33}$$

- If π stopped with (StopD), update

$$\hat{y}^{\langle\sigma+1\rangle}_{J^{(\pi)}} \leftarrow \hat{y}^{(\pi)}, \tag{3.34}$$

$$f^{\langle\sigma+1\rangle}_{V^{(\pi)}} \leftarrow f^{(\pi)}_{V^{(\pi)}}(\hat{y}^{(\pi)}), \tag{3.35}$$

otherwise keep

$$\left(\hat{y}^{\langle\sigma+1\rangle}_{J^{(\pi)}}, f^{\langle\sigma+1\rangle}_{V^{(\pi)}}\right) \leftarrow \left(\hat{y}^{\langle\overline{\pi}\rangle}_{J^{(\pi)}}, f^{\langle\overline{\pi}\rangle}_{V^{(\pi)}}\right). \tag{3.36}$$

- If π stopped with (StopP) and

$$\delta^{\langle\sigma+1\rangle}_{\bar{J}^{(\pi)}} - \delta^{\langle\overline{\pi}\rangle}_{\bar{J}^{(\pi)}} > \tau_3 \Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}, \tag{Dep}$$

  - increase $E^{\langle\sigma+1\rangle} \leftarrow E^{\langle\overline{\pi}\rangle} \cup \{(\hat{\jmath}, j^*)\}$ by at least an arc $(\hat{\jmath}, j^*)$ with

$$j^* \in \operatorname{Argmax}\{g(\bar{x}^{\langle\sigma+1\rangle})^2_j - g(\bar{x}^{\langle\overline{\pi}\rangle})^2_j : j \in \bar{J}^{(\pi)}\},$$

  in all other cases keep $E^{\langle\sigma+1\rangle} \leftarrow E^{\langle\overline{\pi}\rangle}$.
- Set $\sigma \leftarrow \overline{\pi} + 1$ and compute $\Delta^{\langle\sigma\rangle}$. If $\Delta^{\langle\sigma\rangle} \leq \varepsilon(|f(\hat{y}^{\langle\sigma\rangle})| + 1)$ terminate *all* processes and **STOP**.
- Free access to global data and stop this process π.

Note, we do not require continuation of a previous model on the same subset J of variables in some later process. In fact, another subset might work on the same subproblem $v \in V$ and change $(\bar{l}^{\langle\sigma\rangle}_v, \bar{x}^{\langle\sigma\rangle}_v)$.

For the analysis, it will be convenient to arrange the processes in different groups for each value $\sigma \in \mathbb{N}_0 \cup \{\infty\}$ by forming these sets as soon as the algorithm's index marker exceeds this value σ or when the algorithm reached its final marker of at most this value,

$$\underline{\Pi}^{\langle\sigma\rangle} := \{\pi = (\underline{\pi}, \overline{\pi}) : \underline{\pi} < \overline{\pi} \text{ and } \underline{\pi} < \sigma\},$$

$$\overline{\Pi}^{\langle\sigma\rangle} := \{\pi = (\underline{\pi}, \overline{\pi}) : \underline{\pi} < \overline{\pi} < \sigma\},$$

$$\Pi^{\langle\sigma\rangle} := \{\pi = (\underline{\pi}, \overline{\pi}) : \underline{\pi} < \sigma \leq \overline{\pi}\} = \underline{\Pi}^{\langle\sigma\rangle} \setminus \overline{\Pi}^{\langle\sigma\rangle}.$$

The first group $\underline{\Pi}^{\langle\sigma\rangle}$ collects all processes that have successfully executed step (a) before the algorithm reached σ, $\overline{\Pi}^{\langle\sigma\rangle}$ singles out all processes that have executed (d) before σ and the set $\Pi^{\langle\sigma\rangle}$ comprises all processes that are actively working on or just finishing a subproblem at σ, *i. e.*, these are running in parallel. Each increment of σ by the algorithm is associated with the addition or deletion of exactly one process from this set of parallel processes as we show next.

Observation 5 *$\Pi^{\langle 0\rangle} = \emptyset$ and for $\sigma' \in \mathbb{N}_0$ there holds $\Pi^{\langle\sigma'\rangle} \neq \Pi^{\langle\sigma'+1\rangle}$ if and only if $|(\Pi^{\langle\sigma'\rangle} \setminus \Pi^{\langle\sigma'+1\rangle}) \cup (\Pi^{\langle\sigma'+1\rangle} \setminus \Pi^{\langle\sigma'\rangle})| = 1$. If $\Pi^{\langle\sigma'\rangle} = \Pi^{\langle\sigma'+1\rangle}$ then for all $\sigma \geq \sigma'$ we have $\Pi^{\langle\sigma\rangle} = \Pi^{\langle\sigma'\rangle}$ and there is no process π with $\underline{\pi} = \sigma$ or $\overline{\pi} = \sigma$.*

Proof. In the first statement $\Pi^{\langle 0\rangle} = \emptyset$ follows by definition and sufficiency is obvious. In order to show necessity observe that $\Pi^{\langle\sigma'\rangle} \neq \Pi^{\langle\sigma'+1\rangle}$ implies $|(\Pi^{\langle\sigma'\rangle} \setminus \Pi^{\langle\sigma'+1\rangle}) \cup (\Pi^{\langle\sigma'+1\rangle} \setminus \Pi^{\langle\sigma'\rangle})| \geq 1$, so there must be at least one process π that is in $\Pi^{\langle\sigma'\rangle}$ but not in $\Pi^{\langle\sigma'+1\rangle}$, *i. e.*, it satisfies $\overline{\pi} = \sigma'$, or that is in $\Pi^{\langle\sigma'+1\rangle}$ but not in $\Pi^{\langle\sigma'\rangle}$, *i. e.*, it satisfies $\underline{\pi} = \sigma'$. Due to the exclusive access steps (a) and (d) there is exactly one such process π with $\overline{\pi} \geq 0$. Indeed, any process $\pi \in \Pi^{\langle\sigma'+1\rangle}$ with $\underline{\pi} = \sigma'$ executed step (a) successfully at $\sigma = \sigma'$ (unsuccessful steps have $\overline{\pi} < 0$) and increased the marker to $\sigma = \sigma' + 1$ so that all further executions of steps (a) and (d) lead to larger numbers. An analogous argument holds for the case $\overline{\pi} = \sigma'$.

If $\Pi^{\langle\sigma'\rangle} = \Pi^{\langle\sigma'+1\rangle}$ then no process $\pi \in \Pi^{\langle\sigma'\rangle}$ executes (d) at $\sigma = \sigma'$ and there is no new successful execution of (a) at $\sigma = \sigma'$. Thus, σ is never increased above $\sigma' + 1$, all running processes $\pi \in \Pi^{\langle\sigma'\rangle}$ satisfy $\overline{\pi} = \infty$, all others have $\overline{\pi} < \sigma' + 1$ and none of these values ever change, so the claim follows. □

By this observation, the following set collects the markers $\sigma \in \mathbb{N}_0$ visited by the algorithm,

$$\Sigma := \{0\} \cup \{\sigma \in \mathbb{N} : \Pi^{\langle\sigma\rangle} \neq \Pi^{\langle\sigma-1\rangle}\}.$$

3.4. Consistency of the Updating Scheme

The most important steps of the algorithm, which guarantee that the parallel subspace process does not endanger global convergence, are the subspace selection step (a) and the update of the dependency graph $D^{\langle\sigma\rangle} = (M, E^{\langle\sigma\rangle})$ in step (d). The tests (S1)–(S3) enforce a proper selection of the subspaces.

Condition (S1) ensures that all dependencies implicated by the dependency graph $D^{\langle\sigma\rangle}$ are respected. If two subspaces have a dependency in the sense that an improvement on one subspace may worsen the other, future selections should always include the second with the first. Those dependencies are detected in the update step (d).

Condition (S2) guarantees that the progress made on the subspace achieves sufficient decrease on the whole space. This means, if a descent step occurs on the subspace this step is also a good descent step on the whole space, and if no descent step occurs, the reduction in predicted decrease $\Delta^{\langle\sigma\rangle}_{J(\pi)}$ on the subspace leads to a significant reduction of the predicted decrease $\Delta^{\langle\sigma\rangle}$ on the whole space.

Condition (S3) ensures that two processes running in parallel can interact only in a limited way. The next statement quantifies this limitation.

Lemma 6 *For $\sigma \in \Sigma$,*

(i) $B^{\langle\sigma\rangle} = \bigcup_{\pi\in\Pi^{\langle\sigma\rangle}} V^{(\pi)}$,

(ii) $V^{(\pi)} \cap V^{(\pi')} = \emptyset$ *for* $\pi, \pi' \in \Pi^{\langle\sigma\rangle}$ *with* $\pi \neq \pi'$,

(iii) $J^{(\pi')} \cap (J^{(\pi)} \cup \bar{J}^{(\pi)}) = J^{(\pi)} \cap (J^{(\pi')} \cup \bar{J}^{(\pi')}) = \emptyset$ *for* $\pi, \pi' \in \Pi^{\langle\sigma\rangle}$ *with* $\pi \neq \pi'$,

(iv) $\hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}\cup\bar{J}^{(\pi)}} = \hat{y}^{\langle\sigma\rangle}_{J^{(\pi)}\cup\bar{J}^{(\pi)}}$, $(\bar{l}^{\langle\underline{\pi}\rangle}_{V^{(\pi)}}, \bar{x}^{\langle\underline{\pi}\rangle}_{V^{(\pi)}}) = (\bar{l}^{\langle\sigma\rangle}_{V^{(\pi)}}, \bar{x}^{\langle\sigma\rangle}_{V^{(\pi)}})$, *and* $f^{\langle\underline{\pi}\rangle}_{V^{(\pi)}} = f^{\langle\sigma\rangle}_{V^{(\pi)}}$ *for* $\pi \in \Pi^{\langle\sigma\rangle}$.

Proof. The proof works by induction on σ. For $\sigma = 0$ we have $B^{\langle\sigma\rangle} = \emptyset$ and $\Pi^{\langle\sigma\rangle} = \emptyset$, thus (i)–(iv) hold 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 5 asserts the existence of a unique $\eta \in (\Pi^{\langle\sigma\rangle} \setminus \Pi^{\langle\sigma+1\rangle}) \cup (\Pi^{\langle\sigma+1\rangle} \setminus \Pi^{\langle\sigma\rangle})$ and this η either satisfies $\underline{\eta} = \sigma$ or $\overline{\eta} = \sigma$.

If $\underline{\eta} = \sigma$ we have $\Pi^{\langle\sigma\rangle} = \Pi^{\langle\sigma+1\rangle} \setminus \{\eta\}$ and process η executed a successful step (a) at σ, so $B^{\langle\sigma+1\rangle} = B^{\langle\sigma\rangle} \cup V^{(\eta)}$ and (i) as well as (iv) hold. By induction, (ii) and (iii) only need to be verified for $\pi = \eta$ and $\pi' \in \Pi^{\langle\sigma\rangle}$. Because $V^{(\pi')} \subseteq B^{\langle\sigma\rangle}$ and (S3) was satisfied for η at σ, (ii) follows from $\emptyset = V^{(\eta)} \cap B^{\langle\sigma\rangle} \supseteq V^{(\eta)} \cap V^{(\pi')}$. For (iii) assume, w. l. o. g., there exists a $j \in J^{(\eta)} \cap (J^{(\pi')} \cup \bar{J}^{(\pi')})$. Assumption (1.1) implies $\emptyset \neq V_j$ and by (3.3) and (3.13) we have $V_j \subseteq V_{J^{(\eta)}} = V^{(\eta)}$. If $j \in J^{(\pi')}$ then by the same argument $V_j \subseteq V^{(\pi')}$ and therefore $\emptyset \neq V_j \subseteq V_{J^{(\pi')}} \cap V_{J^{(\eta)}}$, a contradiction to (ii). If $j \in \bar{J}^{(\pi')}$ then by (3.14) there is a $v \in V^{(\pi')}$ so that $j \in J_v$ which implies by (3.5) $v \in V_j$ and as above $v \in V_{J^{(\pi')}} \cap V_{J^{(\eta)}} \neq \emptyset$.

If $\overline{\eta} = \sigma$ we have $\Pi^{\langle\sigma+1\rangle} = \Pi^{\langle\sigma\rangle} \setminus \{\eta\}$ and process η executed a step (d) at σ. By the latter, $B^{\langle\sigma+1\rangle} = B^{\langle\sigma\rangle} \setminus V^{(\eta)}$ and all values for indices $M \setminus J^{(\eta)}$ and $V \setminus V_{J^{(\eta)}}$ are left unchanged. In view of the validity of (ii) and (iii) for $\pi = \eta$ at σ, (i)–(iv) hold by induction also for $\sigma + 1$ and its remaining processes. □

When a process π stops at $\overline{\pi}$, the relevant subspace information for π has not been modified and at $\overline{\pi}+1$ the data on its selected subspace and subproblems is consistent with the terminal status of the bundle method of π.

Lemma 7 *Given* $\pi \in \overline{\Pi}^{\langle\infty\rangle}$ *assume* $f^{\langle\underline{\pi}\rangle}_{V^{(\pi)}} = f_{V^{(\pi)}}(\hat{y}^{\langle\underline{\pi}\rangle})$. *Then*

$$\hat{y}^{\langle\underline{\pi}\rangle}_{\bar{J}^{(\pi)}} = \hat{y}^{\langle\sigma\rangle}_{\bar{J}^{(\pi)}} \qquad \text{for all } \sigma \in \{\underline{\pi}, \dots, \overline{\pi}+1\}, \tag{3.37}$$

$$\hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}} = \hat{y}^{\langle\sigma\rangle}_{J^{(\pi)}} \qquad \text{for all } \sigma \in \{\underline{\pi}, \dots, \overline{\pi}\}, \tag{3.38}$$

$$f^{\langle\underline{\pi}\rangle}_{V^{(\pi)}} = f^{\langle\sigma\rangle}_{V^{(\pi)}} = f_{V^{(\pi)}}(\hat{y}^{\langle\sigma\rangle}) = f^{(\pi)}_{V^{(\pi)}}(\hat{y}^{\langle\sigma\rangle}_{J^{(\pi)}}) \qquad \text{for all } \sigma \in \{\underline{\pi}, \dots, \overline{\pi}\}, \tag{3.39}$$

$$(\bar{l}^{\langle\underline{\pi}\rangle}_{V^{(\pi)}}, \bar{x}^{\langle\underline{\pi}\rangle}_{V^{(\pi)}}) = (\bar{l}^{\langle\sigma\rangle}_{V^{(\pi)}}, \bar{x}^{\langle\sigma\rangle}_{V^{(\pi)}}) \qquad \text{for all } \sigma \in \{\underline{\pi}, \dots, \overline{\pi}\}, \tag{3.40}$$

$$\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}} = \Delta^{\langle\sigma\rangle}_{J^{(\pi)}} \qquad \text{for all } \sigma \in \{\underline{\pi}, \dots, \overline{\pi}\}, \tag{3.41}$$

and with $\hat{y}^{(\pi)}$ *and* $(\bar{l}^{(\pi)}, \bar{x}^{(\pi)}) \in \operatorname{conv} W_{V^{(\pi)}}$ *be the final values of* π *in step (d) at* $\overline{\pi}$

$$\hat{y}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}} = \hat{y}^{(\pi)}, \qquad f^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}} = f^{(\pi)}_{V^{(\pi)}}(\hat{y}^{(\pi)}) = f_{V^{(\pi)}}(\hat{y}^{\langle\overline{\pi}+1\rangle}), \tag{3.42}$$

$$\hat{y}^{\langle\overline{\pi}+1\rangle}_{M\setminus J^{(\pi)}} = \hat{y}^{\langle\overline{\pi}\rangle}_{M\setminus J^{(\pi)}}, \qquad f^{\langle\overline{\pi}+1\rangle}_{V\setminus V^{(\pi)}} = f^{\langle\overline{\pi}\rangle}_{V\setminus V^{(\pi)}}, \tag{3.43}$$

$$(\bar{l}^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}}, \bar{x}^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}}) = (\bar{l}^{(\pi)}, \bar{x}^{(\pi)}), \qquad (\bar{l}^{\langle\overline{\pi}+1\rangle}_{V\setminus V^{(\pi)}}, \bar{x}^{\langle\overline{\pi}+1\rangle}_{V\setminus V^{(\pi)}}) = (\bar{l}^{\langle\overline{\pi}\rangle}_{V\setminus V^{(\pi)}}, \bar{x}^{\langle\overline{\pi}\rangle}_{V\setminus V^{(\pi)}}), \tag{3.44}$$

$$\Delta^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}} = \Delta^{(\pi)}(\hat{y}^{(\pi)}, (\bar{l}^{(\pi)}, \bar{x}^{(\pi)})), \qquad \bar{\Delta}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}} = \bar{\Delta}^{\langle\overline{\pi}\rangle}_{J^{(\pi)}}. \tag{3.45}$$

Proof. For $\sigma \in \{\underline{\pi}+1, \dots, \overline{\pi}\}$ we have $\pi \in \Pi^{\langle\sigma\rangle}$, hence, for these values of σ, Lemma 6 (iv) implies (3.37), (3.38), (3.40), and the first equation of (3.39) (the remaining two will be proved below). With these (3.41) follows from the definition (3.11) together with (3.7)–(3.9) using (3.23).

The values for $\sigma = \overline{\pi} + 1$ are set by π when executing step (d) at $\overline{\pi}$, so (3.44) follows from (3.31) and (3.32). Likewise, (3.31) establishes (3.43) and also completes the result for (3.37) because the definition of $\bar{J}^{(\pi)}$ (3.14) implies $\bar{J}^{(\pi)} \subseteq M \setminus J^{(\pi)}$. Because of (3.37) we may invoke Observation 4 for $\hat{y}^{\langle\underline{\pi}\rangle}$ up to $\hat{y}^{\langle\overline{\pi}+1\rangle}$ throughout this proof. In particular, to complete (3.39),

$$f_{V^{(\pi)}}(\hat{y}^{\langle\sigma\rangle}) \overset{(3.27)}{=} f^{(\pi)}_{V^{(\pi)}}(\hat{y}^{\langle\sigma\rangle}_{J^{(\pi)}}) \overset{(3.38)}{=} f^{(\pi)}_{V^{(\pi)}}(\hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}) \overset{(3.27)}{=} f_{V^{(\pi)}}(\hat{y}^{\langle\underline{\pi}\rangle}) = f^{\langle\underline{\pi}\rangle}_{V^{(\pi)}} \text{ for all } \sigma \in \{\underline{\pi}, \ldots, \overline{\pi}\}.$$

Now consider (3.42). If step (c) was ended by (StopP), then $\hat{y}^{(\pi)} = \hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}} \overset{(3.38)}{=} \hat{y}^{\langle\overline{\pi}\rangle}_{J^{(\pi)}} \overset{(3.36)}{=} \hat{y}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}$ and $f_{V^{(\pi)}}(\hat{y}^{\langle\overline{\pi}+1\rangle}) \overset{(3.27)}{=} f^{(\pi)}_{V^{(\pi)}}(\hat{y}^{(\pi)}) \overset{(3.39)}{=} f^{\langle\overline{\pi}\rangle}_{V^{(\pi)}} \overset{(3.36)}{=} f^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}}$. If step (c) was ended by (StopD) then $\hat{y}^{(\pi)} \overset{(3.34)}{=} \hat{y}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}$ and $f^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}} \overset{(3.35)}{=} f^{(\pi)}_{V^{(\pi)}}(\hat{y}^{(\pi)}) \overset{(3.27)}{=} f_{V^{(\pi)}}(\hat{y}^{\langle\overline{\pi}+1\rangle})$, so (3.42) holds.

The left hand side equation of (3.45) follows by

$$\begin{aligned}
\Delta^{(\pi)}(\hat{y}^{(\pi)}, (\bar{l}^{(\pi)}, \bar{x}^{(\pi)})) &\overset{(3.22)}{=} \sum_{v \in V^{(\pi)}} \left[f^{(\pi)}_v(\hat{y}^{(\pi)}) - \hat{f}^{(\pi)}_{(\bar{l}^{(\pi)}, \bar{x}^{(\pi)}),v}(\hat{y}^{(\pi)}) \right] + \tfrac{1}{u}\|g^{(\pi)}(\bar{x}^{(\pi)})\|^2 \\
&\overset{(3.42),(3.44)}{=} \sum_{v \in V^{(\pi)}} \left[f^{\langle\overline{\pi}+1\rangle}_v - \hat{f}^{(\pi)}_{(\bar{l}^{\langle\overline{\pi}+1\rangle}_v, \bar{x}^{\langle\overline{\pi}+1\rangle}_v),v}(\hat{y}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}) \right] + \tfrac{1}{u}\|g^{(\pi)}(\bar{x}^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}})\|^2 \\
&\overset{(3.28),(3.29)}{=} \sum_{v \in V^{(\pi)}} \left[f^{\langle\overline{\pi}+1\rangle}_v - \hat{f}_{(\bar{l}^{\langle\overline{\pi}+1\rangle}_v, \bar{x}^{\langle\overline{\pi}+1\rangle}_v),v}(\hat{y}^{\langle\overline{\pi}+1\rangle}) \right] + \tfrac{1}{u}\|g(\bar{x}^{\langle\overline{\pi}+1\rangle})_{J^{(\pi)}}\|^2 \\
&\overset{(3.11),(3.13)}{=} \Delta^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}.
\end{aligned}$$

In order to show $\bar{\Delta}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}} = \bar{\Delta}^{\langle\overline{\pi}\rangle}_{J^{(\pi)}}$ it suffices to check that none of the values involved in (3.12) change when π executes (d) at $\overline{\pi}$. With $V_{J^{(\pi)}} = V^{(\pi)}$ by (3.13) this follows from (3.43) and (3.44) because for $v \in V \setminus V_{J^{(\pi)}}$ we have $J_v \subseteq M \setminus J^{(\pi)}$ by (3.24) and so $\hat{y}^{\langle\overline{\pi}+1\rangle}_{J_v} = \hat{y}^{\langle\overline{\pi}\rangle}_{J_v}$, $f^{\langle\overline{\pi}+1\rangle}_v = f^{\langle\overline{\pi}\rangle}_v$, and $(\bar{l}^{\langle\overline{\pi}+1\rangle}_v, \bar{x}^{\langle\overline{\pi}+1\rangle}_v) = (\bar{l}^{\langle\overline{\pi}\rangle}_v, \bar{x}^{\langle\overline{\pi}\rangle}_v)$. Thus, (3.8) establishes $\hat{f}_{(\bar{l}^{\langle\overline{\pi}+1\rangle}_v, \bar{x}^{\langle\overline{\pi}+1\rangle}_v),v}(\hat{y}^{\langle\overline{\pi}+1\rangle}) = \hat{f}_{(\bar{l}^{\langle\overline{\pi}\rangle}_v, \bar{x}^{\langle\overline{\pi}\rangle}_v),v}(\hat{y}^{\langle\overline{\pi}\rangle})$ for $v \in V \setminus V_{J^{(\pi)}}$, while $g(\bar{x}^{\langle\overline{\pi}+1\rangle})_{M\setminus(J^{(\pi)}\cup\bar{J}^{(\pi)})} = g(\bar{x}^{\langle\overline{\pi}\rangle})_{M\setminus(J^{(\pi)}\cup\bar{J}^{(\pi)})}$ follows via (3.9) because $V_{M\setminus(J^{(\pi)}\cup\bar{J}^{(\pi)})} \cap V_{J^{(\pi)}} = \emptyset$ by (3.25). Thus, $\bar{\Delta}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}} = \bar{\Delta}^{\langle\overline{\pi}\rangle}_{J^{(\pi)}}$ and (3.45) holds. □

Throughout the algorithm, the global data is consistent and the arc set of the dependency graph may only increase.

Lemma 8 *For all $\sigma \in \Sigma$,*

$$\begin{aligned}
f^{\langle\sigma\rangle}_V &= f_V(\hat{y}^{\langle\sigma\rangle}), & (\bar{l}^{\langle\sigma\rangle}, \bar{x}^{\langle\sigma\rangle}) &\in \operatorname{conv} W, \\
\Delta^{\langle\sigma\rangle} &= \Delta(\hat{y}^{\langle\sigma\rangle}, (\bar{l}^{\langle\sigma\rangle}, \bar{x}^{\langle\sigma\rangle})), & E^{\langle\sigma\rangle} &\subseteq E^{\langle\sigma+1\rangle} \subseteq \{(i,j) : i,j \in M, i \neq j\}.
\end{aligned} \tag{3.46}$$

Furthermore, $\Delta^{\langle\sigma\rangle} > \varepsilon(|f(\hat{y}^{\langle\sigma\rangle})| + 1)$ for all $\sigma \in \Sigma$ with $\sigma + 1 \in \Sigma$.

Proof. The proof is by induction on σ. For $\sigma = 0$ the claim holds by the initialization step. Suppose now $\sigma + 1 \in \Sigma$ and the claim holds for $\sigma \in \Sigma$. If $\sigma + 1$ is reached by a step (a) then none of the involved variables are changed and the relations still hold. Otherwise $\sigma + 1$ is reached by a step (d) executed by some process π with $\overline{\pi} = \sigma$. For $v \in V \setminus V^{(\pi)}$ (3.24) yields $J_v \subseteq M \setminus J^{(\pi)}$ and by (3.31) $\hat{y}^{\langle\overline{\pi}+1\rangle}_{J_v} = \hat{y}^{\langle\overline{\pi}\rangle}_{J_v}$, $f^{\langle\overline{\pi}+1\rangle}_v = f^{\langle\overline{\pi}\rangle}_v$, $(\bar{l}^{\langle\overline{\pi}+1\rangle}_v, \bar{x}^{\langle\overline{\pi}+1\rangle}_v) = (\bar{l}^{\langle\overline{\pi}\rangle}_v, \bar{x}^{\langle\overline{\pi}\rangle}_v) \in \operatorname{conv} W_v$, so for $v \in V \setminus V^{(\pi)}$ (3.46) holds by induction because $f_v(y)$ only depends on y_{J_v} by (3.7). For $v \in V^{(\pi)}$ the claim follows for (3.46) directly from (3.42) and (3.44). The correctness of (3.46) for $\overline{\pi} + 1$ implies the correctness of $\Delta^{\langle\overline{\pi}+1\rangle}$ by Observation 2. The claim for $E^{\langle\overline{\pi}+1\rangle}$ follows directly from step (d). Finally, $\Delta^{\langle\overline{\pi}+1\rangle} \leq \varepsilon(|f(\hat{y}^{\langle\overline{\pi}+1\rangle})| + 1)$ leads to the termination of the algorithm in step (d) at $\overline{\pi}$ and then $\overline{\pi} + 1 = \max \Sigma$. □

Next we show that the algorithm always starts at least one working process as long as the stopping criterion is not met.

Observation 9 *For $\sigma \in \Sigma$ with $\Pi^{\langle\sigma\rangle} = \emptyset$ there holds $\Pi^{\langle\sigma+1\rangle} = \{\pi\}$ for some π with $\underline{\pi} = \sigma$ if and only if $\Delta^{\langle\sigma\rangle} > \varepsilon(|f(\hat{y}^{\langle\sigma\rangle})|+1)$. In words, if no process is running, at least one process is started with a successful step (a) if and only if the stopping criterion is not satisfied for the current global data.*

Proof. Lemma 6 (i) and (1.1) ($J^{(\pi)} \neq \emptyset \Rightarrow V^{(\pi)} \neq \emptyset$) assert that $\Pi^{\langle\sigma\rangle} = \emptyset$ is equivalent to $B^{\langle\sigma\rangle} = \emptyset$. For $B^{\langle\sigma\rangle} = \emptyset$ conditions (S1)–(S3) can always be satisfied. Indeed, choosing $J^{(\pi)} = M$ satisfies (S1) because M trivially observes all dependencies of $E^{\langle\sigma\rangle}$, (S2) because $\Delta_M^{\langle\pi\rangle} = \Delta^{\langle\underline{\pi}\rangle}$ by (3.10) and (3.11), and (S3) because $B^{\langle\sigma\rangle} = \emptyset$. In particular, the algorithm starts with setting $\sigma = 0$ and $B^{\langle 0\rangle} = \emptyset$ and reaches step (a) if and only if the stopping criterion is not satisfied in step 1. For $0 < \sigma \in \Sigma$, $\Pi^{\langle\sigma\rangle} = \emptyset$ requires $\Pi^{\langle\sigma-1\rangle} = \{\pi'\}$ for some process π' with $\overline{\pi}' = \sigma - 1$ by Observation 5. Thus, π' executes a step (d) at $\sigma - 1$ and the algorithm continues if and only if the stopping criterion is not satisfied for the global data of σ. □

Already very small subspaces may suffice to satisfy the selection criteria. Indeed, the following observation proves that there is always a one dimensional subspace, *i. e.*, an index set with one element that satisfies (S2) (without considering blocking or implications required by E).

Observation 10 *Let $\sigma \in \Sigma$. For $J \subseteq J' \subseteq M$ there holds $\Delta_J^{\langle\sigma\rangle} \leq \Delta_{J'}^{\langle\sigma\rangle}$. There always exists a $j \in M$ with $\Delta_{\{j\}}^{\langle\sigma\rangle} \geq \frac{1}{2}\Delta^{\langle\sigma\rangle}\min\{\frac{1}{m}, \frac{1}{|V|}\}$.*

Proof. For $v \in V$, (3.46) asserts $(\bar{l}_v^{\langle\sigma\rangle}, \bar{x}_v^{\langle\sigma\rangle}) \in \operatorname{conv} W_v$ and

$$f_v^{\langle\sigma\rangle} \overset{(3.46)}{=} f_v(\hat{y}^{\langle\sigma\rangle}) \overset{(2.4)}{\geq} \hat{f}_{(\bar{l}_v^{\langle\sigma\rangle}, \bar{x}_v^{\langle\sigma\rangle}),v}(\hat{y}^{\langle\sigma\rangle}),$$

so the first statement concerning $\Delta_J^{\langle\sigma\rangle}$ is an immediate consequence of (3.11).

For the second statement, recall that

$$\Delta^{\langle\sigma\rangle} \overset{(3.10)}{=} \left(\sum_{v\in V}\left[f_v^{\langle\sigma\rangle} - \hat{f}_{(\bar{l}_v^{\langle\sigma\rangle}, \bar{x}_v^{\langle\sigma\rangle}),v}(\hat{y}^{\langle\sigma\rangle})\right]\right) + \left(\frac{1}{u}\sum_{j\in M} g(\bar{x}^{\langle\sigma\rangle})_j^2\right).$$

At least one of both summands is greater than or equal to $\frac{1}{2}\Delta^{\langle\sigma\rangle}$. If this is true for the first one, for at least one $v \in V$ the term $f_v^{\langle\sigma\rangle} - \hat{f}_{(\bar{l}_v^{\langle\sigma\rangle}, \bar{x}_v^{\langle\sigma\rangle}),v}(\hat{y}^{\langle\sigma\rangle})$ is greater than or equal to $\frac{1}{2}\Delta^{\langle\sigma\rangle}\frac{1}{|V|}$ and any $j \in J_v$ satisfies the claim. If it is true for the second one, for at least one $j \in M$ the term $\frac{1}{u}g(\bar{x}^{\langle\sigma\rangle})_j^2$ is greater than or equal to $\frac{1}{2}\Delta^{\langle\sigma\rangle}\frac{1}{m}$. □

The next result establishes that no process runs forever, so existing dependencies between subspaces have to be discovered eventually.

Lemma 11 *For each $\sigma \in \Sigma$, each process $\pi \in \underline{\Pi}^{\langle\sigma\rangle}$ either stops at $\overline{\pi} < \infty$ or it is terminated by another process π' executing step (d) at $\overline{\pi}' < \infty$ with the global data of $\overline{\pi}' + 1$ satisfying the termination criterion.*

Proof. Each process $\pi \in \underline{\Pi}^{\langle\sigma\rangle}$ satisfies (S2) at $\underline{\pi}$, therefore $\Delta_{J(\pi)}^{\langle\underline{\pi}\rangle} \geq \tau_1 \Delta^{\langle\underline{\pi}\rangle} > 0$ by Lemma 7. Then π runs a standard bundle method. After a finite number of its iterations either a descent step occurs or the predicted decrease drops below $\tau_2 \Delta_{J(\pi)}^{\langle\underline{\pi}\rangle} > 0$, see, *e. g.*, [12, 13]. Thus, if π is not terminated externally before, one of the conditions (StopP) or (StopD) is satisfied in finite time. □

Corollary 12 *If* $|\Sigma| < \infty$ *then for* $\sigma = \max \Sigma$ *there holds* $\Delta^{\langle\sigma\rangle} \leq \varepsilon(|f(\hat{y}^{\langle\sigma\rangle})| + 1)$. *If* $|\Sigma| = \infty$ *then* $\Pi^{\langle\infty\rangle} = \emptyset$, $\underline{\Pi}^{\langle\infty\rangle} = \overline{\Pi}^{\langle\infty\rangle}$, *and* $\Sigma = \dot{\bigcup}_{\pi \in \overline{\Pi}^{\langle\infty\rangle}} \{\underline{\pi}, \overline{\pi}\}$.

Proof. This follows from Observation 5, Observation 9, and Lemma 11. □

In the next section we show that the predicted decrease satisfies $\liminf_{\sigma \in \mathbb{N}_0} \Delta^{\langle\sigma\rangle} \to 0$ if f is bounded from below, so in this case the algorithm is finite whenever $\varepsilon > 0$.

3.5. Convergence Analysis

First we clarify the relation between the global progress and that of a single process.

Lemma 13 *For* $\pi \in \overline{\Pi}^{\langle\infty\rangle}$

$$0 \leq f^{(\pi)}(\hat{y}_{J(\pi)}^{\langle\underline{\pi}\rangle}) - f^{(\pi)}(\hat{y}_{J(\pi)}^{\langle\overline{\pi}+1\rangle}) = f(\hat{y}^{\langle\overline{\pi}\rangle}) - f(\hat{y}^{\langle\overline{\pi}+1\rangle}),$$

i. e., *the global progress achieved when* π *stores its subspace solution in the global data is exactly the progress made by* π *on* $J^{(\pi)}$. *In particular, the sequence* $(f(\hat{y}^{\langle\sigma\rangle}))_\sigma$ *is non-increasing.*

Proof. First observe that for π the initial value of the center $\hat{y}^{(\pi)}$ is $\hat{y}_{J(\pi)}^{\langle\underline{\pi}\rangle}$ by step (b) and the final center is $\hat{y}_{J(\pi)}^{\langle\overline{\pi}+1\rangle}$ by (3.42), so the left hand inequality follows from the properties of the bundle method employed in step (c) of π. By Lemma 8 the requirement for Lemma 7 is met, so we may use its results for proving the second equation,

$$\begin{aligned}
0 \leq f^{(\pi)}(\hat{y}_{J(\pi)}^{\langle\underline{\pi}\rangle}) - f^{(\pi)}(\hat{y}_{J(\pi)}^{\langle\overline{\pi}+1\rangle}) &\overset{(3.19)}{=} b_{J(\pi)}^T(\hat{y}_{J(\pi)}^{\langle\underline{\pi}\rangle} - \hat{y}_{J(\pi)}^{\langle\overline{\pi}+1\rangle}) + \sum_{v \in V^{(\pi)}} \left[f_v^{(\pi)}(\hat{y}_{J(\pi)}^{\langle\underline{\pi}\rangle}) - f_v^{(\pi)}(\hat{y}_{J(\pi)}^{\langle\overline{\pi}+1\rangle}) \right] \\
&\overset{(3.38)}{=} b_{J(\pi)}^T(\hat{y}_{J(\pi)}^{\langle\overline{\pi}\rangle} - \hat{y}_{J(\pi)}^{\langle\overline{\pi}+1\rangle}) + \sum_{v \in V^{(\pi)}} \left[f_v^{(\pi)}(\hat{y}_{J(\pi)}^{\langle\overline{\pi}\rangle}) - f_v^{(\pi)}(\hat{y}_{J(\pi)}^{\langle\overline{\pi}+1\rangle}) \right] \\
&\overset{(3.39),(3.42),(3.43),(3.46)}{=} b^T(\hat{y}^{\langle\overline{\pi}\rangle} - \hat{y}^{\langle\overline{\pi}+1\rangle}) + \sum_{v \in V} \left[f_v(\hat{y}^{\langle\overline{\pi}\rangle}) - f_v(\hat{y}^{\langle\overline{\pi}+1\rangle}) \right] \\
&\overset{(2.2)}{=} f(\hat{y}^{\langle\overline{\pi}\rangle}) - f(\hat{y}^{\langle\overline{\pi}+1\rangle}).
\end{aligned}$$

Observation 5 implies that for any $\sigma \in \Sigma$ without a $\pi \in \overline{\Pi}^{\langle\infty\rangle}$ satisfying $\overline{\pi} = \sigma$ there is a process π with $\underline{\pi} = \sigma$ which executes step (a) at $\underline{\pi}$. Therefore (3.30) and (2.2) guarantee $f(\hat{y}^{\langle\underline{\pi}\rangle}) = f(\hat{y}^{\langle\underline{\pi}+1\rangle})$ in this case, which establishes that $(f(\hat{y}^{\langle\sigma\rangle}))_\sigma$ is non-increasing. □

Next we show that the algorithm always drives the predicted decrease to zero on an appropriate subsequence.

Lemma 14 *Suppose an infinite number of descent steps occurs and* f *is bounded from below. Then*

$$\liminf_{\sigma \in \mathbb{N}_0} \Delta^{\langle\sigma\rangle} \to 0.$$

Proof. Let π be a process for which a descent step occurs, *i. e.*, π is stopped because of condition (StopD). By (S2) we have $\Delta^{\langle\underline{\pi}\rangle} \leq \frac{1}{\tau_1}\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}$. By Lemma 8 and Lemma 7 (3.44) the final predicted decrease of π that caused the descent step, is $\Delta^{(\pi)}(\hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}, (\bar{l}^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}}, \bar{x}^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}}))$. Because (StopD) and not the preceding test (StopP) has caused π to stop, we have

$$\tau_2\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}} \overset{(StopP)}{\leq} \Delta^{(\pi)}(\hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}, (\bar{l}^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}}, \bar{x}^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}})) \overset{(StopD)}{\leq} \frac{1}{\rho}\left(f^{(\pi)}(\hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}) - f^{(\pi)}(\hat{y}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}})\right).$$

Putting all together and using Lemma 13 we get

$$\Delta^{\langle\underline{\pi}\rangle} \leq \frac{1}{\tau_1}\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}} \leq \frac{1}{\tau_1\tau_2\rho}\left(f^{(\pi)}(\hat{y}^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}) - f^{(\pi)}(\hat{y}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}})\right) = \frac{1}{\tau_1\tau_2\rho}\left(f(\hat{y}^{\langle\overline{\pi}\rangle}) - f(\hat{y}^{\langle\overline{\pi}+1\rangle})\right).$$

Because f is bounded from below and the sequence $(f(\hat{y}^{\langle\sigma\rangle}))_\sigma$ is non-increasing by Lemma 13, the right hand side of the inequality above converges to zero. □

Lemma 15 *Assume there is only a finite number of descent steps and $\varepsilon = 0$, then*

$$\lim_{\sigma\in\Sigma} \Delta^{\langle\sigma\rangle} = 0.$$

Proof. If $|\Sigma| < \infty$, then the statement holds by Corollary 12. Therefore we may assume $|\Sigma| = \infty$.

Lemma 8 implies $\Delta^{\langle\sigma\rangle} > 0$ for all $\sigma \in \Sigma$ and the dependency graph $D^{\langle\sigma\rangle}$ can only be increased. Because M is a finite set there must be a $\sigma' \in \Sigma$ such that for each $\sigma \geq \sigma'$ we have $E^{\langle\sigma\rangle} = E^{\langle\sigma'\rangle}$ and all processes π with $\overline{\pi} > \underline{\sigma} := \min\left(\{\sigma'\} \cup \{\underline{\pi}' : \pi' \in \Pi^{\langle\sigma'\rangle}\}\right)$ do *not* perform a descent step.

Let $\sigma > \sigma'$, then by Corollary 12 there is a process π such that $\sigma \in \{\underline{\pi}, \overline{\pi}\}$. If $\sigma = \underline{\pi}$ we know by (3.30) and Lemma 8 that $\Delta^{\langle\underline{\pi}\rangle} = \Delta^{\langle\underline{\pi}+1\rangle}$. So assume $\sigma = \overline{\pi}$. Because $\sigma > \sigma'$ process π satisfied condition (StopP) and $E^{\langle\sigma+1\rangle} = E^{\langle\sigma\rangle}$, so (Dep) is not satisfied,

$$\delta^{\langle\overline{\pi}+1\rangle}_{\bar{J}^{(\pi)}} - \delta^{\langle\overline{\pi}\rangle}_{\bar{J}^{(\pi)}} \leq \tau_3\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}.$$

Invoking Observation 2 twice for the subspace $J^{(\pi)}$ of π but once for the data of $\overline{\pi}$ and once for $\overline{\pi}+1$ yields the relations

$$\begin{aligned}\Delta^{\langle\overline{\pi}\rangle} &= \Delta^{\langle\overline{\pi}\rangle}_{J^{(\pi)}} + \delta^{\langle\overline{\pi}\rangle}_{\bar{J}^{(\pi)}} + \bar{\Delta}^{\langle\overline{\pi}\rangle}_{J^{(\pi)}},\\ \Delta^{\langle\overline{\pi}+1\rangle} &= \Delta^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}} + \delta^{\langle\overline{\pi}+1\rangle}_{\bar{J}^{(\pi)}} + \bar{\Delta}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}.\end{aligned}$$

We claim that $\Delta^{\langle\overline{\pi}+1\rangle} \leq (1-\tau)\Delta^{\langle\overline{\pi}\rangle}$ for some constant $0 < \tau < 1$ independent of π. Indeed, by Lemma 8 we may invoke Lemma 7, so (3.41) implies $\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}} = \Delta^{\langle\overline{\pi}\rangle}_{J^{(\pi)}}$ and (3.45) gives $\bar{\Delta}^{\langle\overline{\pi}\rangle}_{J^{(\pi)}} = \bar{\Delta}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}$. The subspace selection condition (S2) asserts $\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}} \geq \tau_1\Delta^{\langle\underline{\pi}\rangle}$ and stopping condition (StopP) implies $\Delta^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}} \overset{(3.45)}{=} \Delta^{(\pi)}(\hat{y}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}, (\bar{l}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}, \bar{x}^{\langle\overline{\pi}+1\rangle}_{V^{(\pi)}})) < \tau_2\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}$. This yields

$$\begin{aligned}\Delta^{\langle\overline{\pi}\rangle} - \Delta^{\langle\overline{\pi}+1\rangle} &= (\Delta^{\langle\overline{\pi}\rangle}_{J^{(\pi)}} - \Delta^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}) + (\bar{\Delta}^{\langle\overline{\pi}\rangle}_{J^{(\pi)}} - \bar{\Delta}^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}) + (\delta^{\langle\overline{\pi}\rangle}_{\bar{J}^{(\pi)}} - \delta^{\langle\overline{\pi}+1\rangle}_{\bar{J}^{(\pi)}}) \qquad (3.47)\\ &= (\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}} - \Delta^{\langle\overline{\pi}+1\rangle}_{J^{(\pi)}}) + (\delta^{\langle\overline{\pi}\rangle}_{\bar{J}^{(\pi)}} - \delta^{\langle\overline{\pi}+1\rangle}_{\bar{J}^{(\pi)}})\\ &\geq (1-\tau_2-\tau_3)\Delta^{\langle\underline{\pi}\rangle}_{J^{(\pi)}}\\ &\geq \underbrace{\tau_1(1-\tau_2-\tau_3)}_{=:\tau\in(0,1)}\Delta^{\langle\underline{\pi}\rangle}.\end{aligned}$$

Note that this shows $\Delta^{\langle\overline{\pi}\rangle} - \Delta^{\langle\overline{\pi}+1\rangle} \geq 0$ for all $\overline{\pi} = \sigma \geq \underline{\sigma}$. Together with $\Delta^{\langle\underline{\pi}\rangle} = \Delta^{\langle\underline{\pi}+1\rangle}$ (see above) we get therefore that the sequence $(\Delta^{\langle\sigma\rangle})_{\sigma\geq\underline{\sigma}}$ is non-increasing. Because $\overline{\pi} > \sigma'$ we have $\underline{\pi} \geq \underline{\sigma}$ and thus $\Delta^{\langle\underline{\pi}\rangle} \geq \Delta^{\langle\overline{\pi}\rangle}$. From (3.47) we obtain

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

and so

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

Together with the case $\sigma = \underline{\pi}$ above we get $\lim_{\sigma\in\mathbb{N}_0} \Delta^{\langle\sigma\rangle} = 0$, which completes the proof.□

Corollary 16 *If f is bounded from below and $\varepsilon = 0$, the predicted decrease $\Delta^{\langle\sigma\rangle} = f(\hat{y}^{\langle\sigma\rangle}) - \hat{f}_{(\bar{l}^{\langle\sigma\rangle},\bar{x}^{\langle\sigma\rangle})}(\hat{y}^{\langle\sigma\rangle}) + \frac{1}{u}\|g(\bar{x}^{\langle\sigma\rangle})\|^2$ goes to zero for an appropriate subsequence $\Sigma^* \subseteq \Sigma$. In particular, $f(\hat{y}^{\langle\sigma\rangle}) - \hat{f}_{(\bar{l}^{\langle\sigma\rangle},\bar{x}^{\langle\sigma\rangle})}(\hat{y}^{\langle\sigma\rangle})$ and $\|g(\bar{x}^{\langle\sigma\rangle})\|$ go to zero, too, for the subsequence Σ^*.*

Proof. Depending on whether an infinite number of descent steps occurs or not the claim follows either from Lemma 14 or Lemma 15. The last statement follows from the fact $f(\hat{y}^{\langle\sigma\rangle}) - \hat{f}_{(\bar{l}^{\langle\sigma\rangle},\bar{x}^{\langle\sigma\rangle})}(\hat{y}^{\langle\sigma\rangle}) \geq 0$ and $\|g(\bar{x}^{\langle\sigma\rangle})\| \geq 0$. □

Theorem 17 *Suppose $\emptyset \neq \operatorname{Argmin} f$ is bounded. Then for an appropriate subsequence $\Sigma^* \subseteq \Sigma$ the sequences $(\hat{y}^{\langle\sigma\rangle})_{\sigma\in\Sigma^*}$ and $(\bar{x}^{\langle\sigma\rangle})_{\sigma\in\Sigma^*}$ that are generated by the parallel bundle algorithm have the following properties.*

(i) each accumulation point of $(\hat{y}^{\langle\sigma\rangle})_{\sigma\in\Sigma^*}$ is an optimal solution of (D),

(ii) each accumulation point of $(\bar{x}^{\langle\sigma\rangle})_{\sigma\in\Sigma^*}$ is an optimal solution of $(\operatorname{conv} P)$.

Proof. 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 Lemma 13, 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. Likewise, $(\bar{l}^{\langle\sigma\rangle}, \bar{x}^{\langle\sigma\rangle})_\sigma$ lies in the compact set $\operatorname{conv} W$ by Lemma 8.

Let $\Sigma' \subseteq \Sigma$ be a subsequence that drives $\Delta^{\langle\sigma\rangle}$ to zero, according to Corollary 16. Let $(l^*, x^*), y^*$ be accumulation points of $(\bar{l}^{\langle\sigma\rangle}, \bar{x}^{\langle\sigma\rangle})_{\sigma\in\Sigma^*}$, $(\hat{y}^{\langle\sigma\rangle})_{\sigma\in\Sigma^*}$ for an appropriate subsequence $\Sigma^* \subseteq \Sigma'$, then Corollary 16 asserts $g(x^*) = 0$ and therefore x^* is a feasible solution of $(\operatorname{conv} P)$. By definition of $\bar{h}_v$ and W_v we have $l_v \leq \bar{h}_v(x_v)$ for all $(l_v, x_v) \in \operatorname{conv} W_v$ $(v \in V)$, thus $f(\hat{y}^{\langle\sigma\rangle}) - \hat{f}_{(\bar{l}^{\langle\sigma\rangle},\bar{x}^{\langle\sigma\rangle})}(\hat{y}^{\langle\sigma\rangle}) \to 0$ implies

$$\begin{aligned}\bar{h}(x^*) \quad \overset{\sigma\in\Sigma^*}{\longleftarrow} \quad \sum_{v\in V}\Big[\bar{h}_v(\bar{x}_v^{\langle\sigma\rangle}) + \underbrace{(\hat{y}^{\langle\sigma\rangle})^T}_{\text{bounded}}\underbrace{g(\bar{x}_v^{\langle\sigma\rangle})}_{\to 0}\Big] &\geq \sum_{v\in V}\Big[\bar{l}_v^{\langle\sigma\rangle} + (\hat{y}^{\langle\sigma\rangle})^T g(\bar{x}_v^{\langle\sigma\rangle})\Big] \\ &= \hat{f}_{(\bar{l}^{\langle\sigma\rangle},\bar{x}^{\langle\sigma\rangle})}(\hat{y}^{\langle\sigma\rangle}) \quad \overset{\sigma\in\Sigma^*}{\longrightarrow} \quad f(y^*).\end{aligned}$$

Thus x^* is an optimal solution of $(\operatorname{conv} P)$ and y^* is an optimal solution of (D). □

4. Extension to Stronger Coupling

In many applications the assumption that V_j is small for most $j \in M$ is actually too strong. Consider, *e. g.*, a constraint for a common resource ensuring that only a limited number of all objects may make use of this resource at specific point in time. Even though such a constraint $j \in M$ couples many subproblems, it typically influences but a few of them,

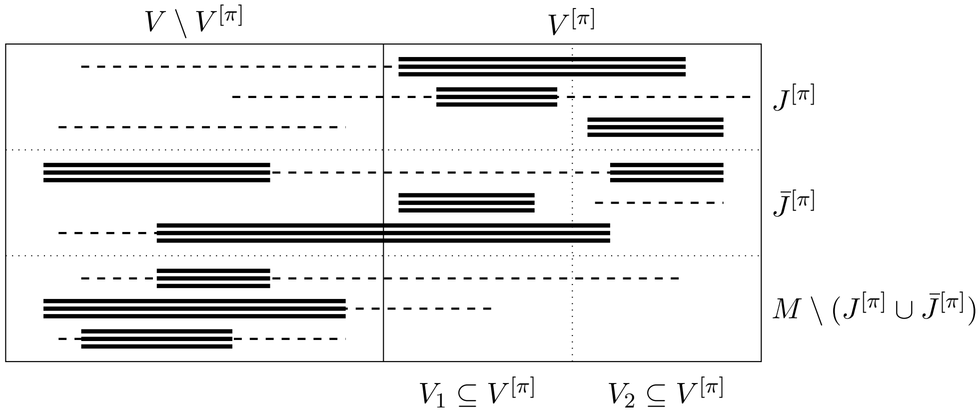


Figure 2: Bold lines show $J_v^{\{\pi\}}$, dashed lines $J_v \setminus J_v^{\{\pi\}}$ for all $v \in V$.

because most objects need this resource at some other time. A typical example for this situation, the train timetabling problem, is given in Section 5. In the extended approach this is exploited by keeping track of those constraints j and subproblems $v \in V$, that have proven to interact for at least one feasible solution x_v up to the current marker σ, and the optimization process is restricted to these.

In order to clearly discern the new objects of this extended version from the objects of the last section, we will use superscript $\{\sigma\}$ for index markers and $[\pi]$ for objects belonging to a process π. The algorithm maintains sets $J_v^{\{\sigma\}} \subseteq J_v$ for $v \in V$ which collect the indices of all those constraints acting on subproblem v, whose Lagrange multipliers presumably influence the optimal solution of $P_v(y)$ or its value. In the corresponding restricted subproblems the constraints $J_v \setminus J_v^{\{\sigma\}}$ will be ignored in the following sense. When a process π selects its subspace $J^{[\pi]} \subseteq M$ and corresponding subproblems $V^{[\pi]} \subset V$ at $\sigma = \underline{\pi}$ (see Fig. 2), it assumes that only the multipliers belonging to $M^{[\pi]} := \bigcup_{v \in V^{[\pi]}} J_v^{\{\underline{\pi}\}}$ have an influence on the solution of these subproblems. Once π finishes its work at $\sigma = \overline{\pi}$, it will not only need to include newly discovered influences due to new optimal solutions $\hat{x}_v$ for some $v \in V^{[\pi]}$ in updated sets $J_v^{\{\overline{\pi}+1\}}$, but possibly also encounter modified sets $J_v^{\{\overline{\pi}\}} \neq J_v^{\{\underline{\pi}\}}$ for some $v \in V^{[\pi]}$ due to changes of the Lagrange multipliers in $M \setminus M^{[\pi]}$ by other processes that might or might not invalidate the results of π, so that they have to be discarded. We start by studying conditions that ensure that results remain valid.

4.1. Conditions of Global Validity for Restricted Subproblems

Sufficient conditions that for every $v \in V$ the optimal solutions computed with respect to the restricted constraint set $J_v^{\{\sigma\}}$ are in fact optimal for the full subproblem with constraint set J_v, are established by the following three properties that will be required to hold for the current multiplier vector $\hat{y}^{\{\sigma\}}$ and the primal aggregate $\bar{x}^{\{\sigma\}}$,

$$\hat{y}_j^{\{\sigma\}} \cdot (A_{j,v} x_v) \geq 0 \qquad \text{for all } v \in V, x_v \in \Omega_v, j \in J_v \setminus J_v^{\{\sigma\}}, \tag{E1}$$

$$A_{J_v \setminus J_v^{\{\sigma\}},v} \hat{x}_v = 0 \qquad \text{for all } v \in V \text{ for some } \hat{x}_v \in \operatorname{Argmax}\{h_v(x_v) - (\hat{y}^{\{\sigma\}}_{J_v^{\{\sigma\}}})^T A_{J_v^{\{\sigma\}},v} x_v \colon x_v \in \Omega_v\}, \tag{E2}$$

$$A_{J_v \setminus J_v^{\{\sigma\}},v} \bar{x}_v^{\{\sigma\}} = 0 \qquad \text{for all } v \in V. \tag{E3}$$

Properties (E1) and (E2) ensure (see Observation 21 below) that for $v \in V$ an optimal solution $\hat{x}_v$ computed for (1.2) with respect to the restricted subspace of multiplier indices $J_v^{\{\sigma\}}$ is actually optimal when the remaining indices $J_v \setminus J_v^{\{\sigma\}}$ of this $\hat{y}^{\{\sigma\}}$ are included as well. Condition (E3) guarantees that the global affine minorant associated with $(\bar{l}^{\{\sigma\}}, \bar{x}^{\{\sigma\}})$ only depends on the restricted subspace $J_v^{\{\sigma\}}$.

Remark 18 Condition (E1) is usually not checked explicitly but guaranteed by exploiting the problem structure. For example, if $A_{j,v} \geq 0$ and $x_v \geq 0$ for all $x_v \in \Omega_v$ (which is often the case for combinatorial problems) the condition is equivalent to the condition $\hat{y}_j^{\{\sigma\}} \geq 0$.

The next definition and observation provide the basis for these requirements.

Definition 19 Given $v \in V$, a multiplier vector $y \in \mathbb{R}^M$ and a set $J'_v \subseteq J_v$, call a point $x_v \in \operatorname{conv} \Omega_v$ (y, J'_v)-*consistent* if

$$y_j \cdot (A_{j,v} x_v) = 0 \text{ for all } j \in M \setminus J'_v. \tag{C1}$$

The set J'_v is (y, v)-*consistent* if

$$y_j \cdot (A_{j,v} x_v) \geq 0 \text{ for all } x_v \in \Omega_v,\ j \in J_v \setminus J'_v \text{ and} \tag{C2}$$

$$\text{there exists a } (y, J'_v)\text{-consistent } \hat{x}_v \in \operatorname{Argmax}\{h_v(x_v) - (y_{J'_v})^T A_{J'_v,v} x_v \colon x_v \in \Omega_v\}. \tag{C3}$$

In this case $\hat{x}_v$ is called a *witness* (of (y, v)-consistency). A family $J'_V := (J'_v)_{v \in V}$ is y-*consistent*, if J'_v is (y, v)-consistent for each $v \in V$.

Observation 20 *Given $v \in V$, $y \in \mathbb{R}^M$ and sets $J'_v \subseteq J''_v \subseteq J_v$, suppose J'_v is (y, v)-consistent with a witness $\hat{x}_v \in \Omega_v$. Then J''_v is also (y, v)-consistent and $\hat{x}_v$ is a witness for this. In particular, $\hat{x}_v$ is an optimal solution for $P_v(y)$ with*

$$f_v(y) = L_v(\hat{x}_v, y) = h_v(\hat{x}_v) - (y_{J'_v})^T A_{J'_v,v} \hat{x}_v. \tag{4.1}$$

Proof. Let $x_v \in \Omega_v$, $v \in V$, be an arbitrary primal point. Then

$$\begin{aligned}
h_v(x_v) - (y_{J''_v})^T A_{J''_v,v} x_v &= h_v(x_v) - (y_{J'_v})^T A_{J'_v} x_v - \underbrace{(y_{J''_v \setminus J'_v})^T A_{J''_v \setminus J'_v,v} x_v}_{\geq 0 \text{ by (C2)}} \\
&\leq h_v(x_v) - (y_{J'_v})^T A_{J'_v,v} x_v \\
&\overset{(C3)}{\leq} h_v(\hat{x}_v) - (y_{J'_v})^T A_{J'_v,v} \hat{x}_v \\
&= h_v(\hat{x}_v) - (y_{J'_v})^T A_{J'_v,v} \hat{x}_v - \underbrace{(y_{J''_v \setminus J'})^T A_{J''_v \setminus J',v} \hat{x}_v}_{=0 \text{ as (C1) holds for } \hat{x}_v}.
\end{aligned}$$

Therefore $\hat{x}_v$ is an optimal solution for the subproblem induced by J''_v with the same objective value. Putting $J''_v = J_v$, (4.1) follows from (1.2), (1.3) and (3.1). □

The next observation connects y-consistency to (E1) and (E2).

Observation 21 *Let $\sigma \in \mathbb{N}_0$ with $\hat{y}^{\{\sigma\}} \in \mathbb{R}^M$ and $J_v^{\{\sigma\}} \subseteq J_v$, $v \in V$, so that (E1) and (E2) hold. Then $J_V^{\{\sigma\}} := (J_v^{\{\sigma\}})_{v \in V}$ is $\hat{y}^{\{\sigma\}}$-consistent.*

Proof. For $v \in V$, (E1) implies (C2) and (E2) establishes (C3) with (C1) for the witness. □

It will be exploited repeatedly that the conditions remain true whenever the sets $J_v^{\{\sigma\}}$ are increased but no other global data is changed.

Observation 22 *Let $\sigma \in \mathbb{N}_0$ so that (E1)–(E3) hold and assume $J_v^{\{\sigma\}} \subseteq J_v^{\{\sigma+1\}}$ for $v \in V$.*

(i) If $\hat{y}^{\{\sigma\}} = \hat{y}^{\{\sigma+1\}}$ then (E1) and (E2) hold for $\sigma + 1$.

(ii) If $\bar{x}^{\{\sigma\}} = \bar{x}^{\{\sigma+1\}}$ then (E3) holds for $\sigma + 1$.

Proof. (E1) and (E3) follow directly from the definition. For (E2) applying Observation 20 for $J' = J_v^{\{\sigma\}}$ and $J'' = J_v^{\{\sigma+1\}}$ implies that any $\hat{x}_v^{\{\sigma\}}$ satisfying (E2) for σ also satisfies (E2) for $\sigma + 1$. $\square$

4.2. Global Objects and Data

Next we specify the objects with their modifications that are required for exploiting the weaker dependency assumptions. The new algorithm maintains the following global data indexed by the global index marker $\sigma \in \mathbb{N}_0$,

$\hat{y}^{\{\sigma\}} \in \mathbb{R}^M$,	the current global center,
$f_v^{\{\sigma\}} := f_v(\hat{y}^{\{\sigma\}}) \in \mathbb{R}$,	the optimal primal value of $P_v(\hat{y}^{\langle\sigma\rangle})$ attained in some $\hat{x}_v^{\{\sigma\}} \in \Omega_v$ for $v \in V$,
$(\bar{l}^{\{\sigma\}}, \bar{x}^{\{\sigma\}}) \in \operatorname{conv} W$,	the current global aggregate minorant,
$B^{\{\sigma\}} \subseteq V$,	set of primal problems currently blocked by some processes,
$D^{\{\sigma\}} = (M, E^{\{\sigma\}})$,	digraph with arc set $E^{\{\sigma\}}$ collecting presumed dependencies,
and additionally	
$J_v^{\{\sigma\}} \subseteq J_v$,	the constraints that subproblem $v \in V$ has interacted with,
$B_M^{\{\sigma\}} \subseteq M$,	set of constraints currently blocked by some processes.

In our first algorithm the blocked constraints were implicitly given via the blocked subproblems $B^{\langle\sigma\rangle} \subset V$, but in the new algorithm the actual relation between subproblems and constraints changes during the algorithm, so the new blocking set $B_M^{\{\sigma\}}$ is needed to track the blocked constraints explicitly.

4.3. The Restricted Subspace Problem Associated with Process π

In contrast to the previous algorithm, in the extended algorithm a process π selects the subproblems $V^{[\pi]}$ w.r.t. the actual global dependency information $J_v^{\{\sigma\}}$, $v \in V$, at the time $\sigma = \underline{\pi}$ when the process starts. This gives rise to the following objects for process π.

$J^{[\pi]} \subseteq M$,	subset that π may modify, selected in step (a),	
$V_j^{[\pi]} := \{v \in V : j \in J_v^{\{\underline{\pi}\}}\}$,	the subproblems currently interacting with j,	(4.2)
$V^{[\pi]} := \bigcup_{j \in J^{[\pi]}} V_j^{[\pi]}$,	the subproblems that π optimizes over,	(4.3)
$J_v^{[\pi]} := J_v^{\{\underline{\pi}\}} \cap J^{[\pi]}$,	dual variables interacting with v and in $J^{[\pi]}$,	
$\bar{J}_v^{[\pi]} := J_v^{\{\underline{\pi}\}} \setminus J^{[\pi]}$,	dual variables interacting with v and not in $J^{[\pi]}$,	
$\bar{J}^{[\pi]} := \bigcup_{v \in V^{[\pi]}} \bar{J}_v^{[\pi]}$,	dual variables interacting with $V^{[\pi]}$ but not in $J^{[\pi]}$,	(4.4)
$\Omega^{[\pi]} = \bigotimes_{v \in V^{[\pi]}} \Omega_v$,	primal ground set.	

The important difference is the definition of $V^{[\pi]}$ which contains only those problems that *actually* interact with the selected constraints $J^{[\pi]}$. Note that setting $J_v^{\{\underline{\pi}\}} := J_v$ for all $v \in V$, *i. e.*, having all possible dependencies, implies $V^{[\pi]} = V^{(\pi)}$ and $\bar{J}^{[\pi]} = \bar{J}^{(\pi)}$. As before, there are some simple relations involving the sets defined above.

Observation 23 *(see Observation 3)*

$$
\begin{aligned}
&J_v^{\{\pi\}} = J_v^{[\pi]} \cup \bar{J}_v^{[\pi]} \subseteq J^{[\pi]} \cup \bar{J}^{[\pi]} && \textit{for all } v \in V^{[\pi]}, && (4.5)\\
&J_v^{\{\pi\}} \cap J^{[\pi]} = \emptyset && \textit{for all } v \in V \setminus V^{[\pi]}, && (4.6)\\
&J_v^{\{\sigma\}} \subseteq J_v && \textit{for all } v \in V, \sigma \in \mathbb{N}_0. && (4.7)
\end{aligned}
$$

Proof. Directly from the definitions. □

We split the global predicted decrease $\Delta^{\{\sigma\}}$ at some index σ according to the selected subspace and subproblems of a specific process π, but this time the analysis will be simpler if we split the coordinates $M \setminus J^{[\pi]}$ of the aggregate subgradient with respect to the dependencies known to or generated by π in the final step at $\overline{\pi}$,

$$
\begin{aligned}
\hat{J}^{[\pi]} &:= \textstyle\bigcup_{v \in V^{[\pi]}} J_v^{\{\overline{\pi}+1\}} \setminus J^{[\pi]} && (4.8)\\
\Delta^{\{\sigma\}} &:= \sum_{v \in V} \left[f_v^{\{\sigma\}} - \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}}) \right] + \tfrac{1}{u} \|g(\bar{x}^{\{\sigma\}})\|^2, && (4.9)\\
\Delta_\pi^{\{\sigma\}} &:= \sum_{v \in V^{[\pi]}} \left[f_v^{\{\sigma\}} - \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}}) \right] + \tfrac{1}{u} \|g(\bar{x}^{\{\sigma\}})_{J^{[\pi]}}\|^2, && (4.10)\\
\bar{\Delta}_\pi^{\{\sigma\}} &:= \sum_{v \in V \setminus V^{[\pi]}} \left[f_v^{\{\sigma\}} - \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}}) \right] + \tfrac{1}{u} \|g(\bar{x}^{\{\sigma\}})_{M \setminus (J^{[\pi]} \cup \hat{J}^{[\pi]})}\|^2, && (4.11)\\
\delta_\pi^{\{\sigma\}} &:= \tfrac{1}{u} \|g(\bar{x}^{\{\sigma\}})_{\hat{J}^{[\pi]}}\|^2.
\end{aligned}
$$

Observation 24 *For each process π and each global index σ there holds*

$$
\Delta^{\{\sigma\}} = \Delta_\pi^{\{\sigma\}} + \bar{\Delta}_\pi^{\{\sigma\}} + \delta_\pi^{\{\sigma\}}
$$

and if $f_v^{\{\sigma\}} = f_v(\hat{y}^{\{\sigma\}})$ for all $v \in V$, then $\Delta^{\{\sigma\}} = \Delta(\hat{y}^{\{\sigma\}}, (\bar{l}^{\{\sigma\}}, \bar{x}^{\{\sigma\}}))$.

Proof. Direct computation. □

Process π computes cost coefficients

$$
c_v^{[\pi]} := -A_{\bar{J}_v^{[\pi]},v}^T \hat{y}_{\bar{J}_v^{[\pi]}}^{\{\pi\}} \qquad \text{for } v \in V^{[\pi]}, \qquad (4.12)
$$

and forms augmented cost functions

$$
h_v^{[\pi]}(x_v) := h_v(x_v) + (c_v^{[\pi]})^T x_v \qquad \text{for } x_v \in \Omega_v, v \in V^{[\pi]}. \qquad (4.13)
$$

Using the correspondingly adapted notation

$$
\begin{aligned}
L_v^{[\pi]}(x_v, y^{[\pi]}) &:= h_v^{[\pi]}(x_v) - (y_{J_v^{[\pi]}}^{[\pi]})^T A_{J_v^{[\pi]},v} x_v && \text{for } x_v \in \Omega_v, v \in V^{[\pi]}, y^{[\pi]} \in \mathbb{R}^{J^{[\pi]}}, && (4.14)\\
f_v^{[\pi]}(y^{[\pi]}) &:= \max_{x_v \in \Omega_v} L_v^{[\pi]}(x_v, y^{[\pi]}) && \text{for } v \in V^{[\pi]}, y^{[\pi]} \in \mathbb{R}^{J^{[\pi]}}, && (4.15)\\
f^{[\pi]}(y^{[\pi]}) &:= b_{J^{[\pi]}}^T y^{[\pi]} + \sum_{v \in V^{[\pi]}} f_v^{[\pi]}(y^{[\pi]}) && \text{for } y^{[\pi]} \in \mathbb{R}^{J^{[\pi]}}, && (4.16)
\end{aligned}
$$

the subproblem that π solves is

$$
(SP^{[\pi]}) \qquad \min_{y^{[\pi]} \in \mathbb{R}^{J^{[\pi]}}} f^{[\pi]}(y^{[\pi]}).
$$

Affine minorants will be generated by the set

$$W^{[\pi]} := \left\{(l^{[\pi]}, x^{[\pi]}) \in \mathbb{R}^{V^{[\pi]}} \times \Omega^{[\pi]} \colon l_v^{[\pi]} = h_v(x_v^{[\pi]}), x_v^{[\pi]} \in \Omega_v, v \in V^{[\pi]}\right\}.$$

The local version of the global data of process π is

$$\begin{aligned} \hat{y}^{[\pi]} &\in \mathbb{R}^{J^{[\pi]}}, && \text{local center,} \\ \hat{x}^{[\pi]} &\in \Omega^{[\pi]}, && \hat{x}_v^{[\pi]} \text{ maximizes (4.15) in } \hat{y}^{[\pi]},\ v \in V^{[\pi]}, \\ (\bar{l}^{[\pi]}, \bar{x}^{[\pi]}) &\in \operatorname{conv} W^{[\pi]}, && \text{local aggregate minorant.} \end{aligned}$$

Process π solves $(SP^{[\pi]})$ approximately using a bundle method. Each $(\bar{l}^{[\pi]}, \bar{x}^{[\pi]}) \in \operatorname{conv} W^{[\pi]}$ defines minorants of $f_v^{[\pi]}$ and $f^{[\pi]}$

$$\begin{aligned} \hat{f}^{[\pi]}_{(\bar{l}_v^{[\pi]}, \bar{x}_v^{[\pi]}),v}(y^{[\pi]}) &:= \bar{l}_v^{[\pi]} + (c_v^{[\pi]})^T \bar{x}_v^{[\pi]} - (y^{[\pi]}_{J_v^{[\pi]}})^T A_{J_v^{[\pi]},v} \bar{x}_v^{[\pi]} && \text{for } v \in V^{[\pi]}, y^{[\pi]} \in \mathbb{R}^{J^{[\pi]}}, \quad (4.17) \\ \hat{f}^{[\pi]}_{(\bar{l}^{[\pi]}, \bar{x}^{[\pi]})}(y^{[\pi]}) &:= b_{J^{[\pi]}}^T y^{[\pi]} + \sum_{v \in V^{[\pi]}} \hat{f}^{[\pi]}_{(\bar{l}_v^{[\pi]}, \bar{x}_v^{[\pi]}),v}(y^{[\pi]}) && \text{for } y^{[\pi]} \in \mathbb{R}^{J^{[\pi]}}. \end{aligned}$$

Due to the selective use of rows of A in (4.14), the gradient of the affine function $\hat{f}^{[\pi]}_{(\bar{l}^{[\pi]}, \bar{x}^{[\pi]})}$ is somewhat clumsy to state,

$$g^{[\pi]}(\bar{x}^{[\pi]}) := b_{J^{[\pi]}} - \left[\sum_{v \in V_j^{[\pi]}} A_{j,v} \bar{x}_v^{[\pi]}\right]_{j \in J^{[\pi]}}. \tag{4.18}$$

With this, the next candidate w. r. t. a given center $\hat{y}^{[\pi]}$ and an aggregate minorant $(\bar{l}^{[\pi]}, \bar{x}^{[\pi]}) \in \operatorname{conv} W^{[\pi]}$ is determined as the minimizer

$$\bar{y}^{[\pi]} := \hat{y}^{[\pi]} - \tfrac{1}{u} g^{[\pi]}(\bar{x}^{[\pi]})$$

of the augmented model $\hat{f}^{[\pi]}_{(\bar{l}^{[\pi]}, \bar{x}^{[\pi]})}(y^{[\pi]}) + \frac{u}{2}\|y^{[\pi]} - \hat{y}^{[\pi]}\|^2$, giving the predicted decrease

$$\Delta^{[\pi]}(\hat{y}^{[\pi]}, (\bar{l}^{[\pi]}, \bar{x}^{[\pi]})) := f^{[\pi]}(\hat{y}^{[\pi]}) - \hat{f}^{[\pi]}_{(\bar{l}^{[\pi]}, \bar{x}^{[\pi]})}(\hat{y}^{[\pi]}) + \tfrac{1}{u}\|g^{[\pi]}(\bar{x}^{[\pi]})\|^2. \tag{4.19}$$

4.4. Linking Local and Global Information

In analogy to the standard setting of Section 3 we establish the conditions that ensure a direct relation between the subproblems of process π and the global subproblems.

Observation 25 *Given a process π, $v \in V^{[\pi]}$ and $y \in \mathbb{R}^M$ with $y_{\bar{J}_v^{[\pi]}} = \hat{y}^{\{\pi\}}_{\bar{J}_v^{[\pi]}}$, there hold*

$$\begin{aligned} L_v^{[\pi]}(x_v, y_{J^{[\pi]}}) &= h_v(x_v) - (y_{J_v^{\{\pi\}}})^T A_{J_v^{\{\pi\}},v} x_v && \textit{for } x_v \in \Omega_v, && (4.20) \\ \hat{f}^{[\pi]}_{(l_v, x_v),v}(y_{J^{[\pi]}}) &= l_v - y^T_{J_v^{\{\pi\}}} A_{J_v^{\{\pi\}},v} x_v && \textit{for } (l_v, x_v) \in \operatorname{conv} W_v. && (4.21) \end{aligned}$$

Proof. Due to $y_{\bar{J}_v^{[\pi]}} = \hat{y}^{\{\pi\}}_{\bar{J}_v^{[\pi]}}$ we have $c_v^{[\pi]} \overset{(4.12)}{=} -A^T_{\bar{J}_v^{[\pi]},v} y_{\bar{J}_v^{[\pi]}}$, so by $J_v^{\{\pi\}} = J_v^{[\pi]} \cup \bar{J}_v^{[\pi]}$ definitions (4.13) and (4.14) prove (4.20) and definition (4.17) shows (4.21). ☐

Observation 26 *Given a process π, $v \in V^{[\pi]}$ and $y \in \mathbb{R}^M$ with $y_{\bar{J}_v^{[\pi]}} = \hat{y}^{\{\pi\}}_{\bar{J}_v^{[\pi]}}$, suppose $J_v^{\{\pi\}}$ is (y, v)-consistent with witness $\hat{x}_v$. Then*

$$f_v^{[\pi]}(y_{J^{[\pi]}}) = L_v^{[\pi]}(\hat{x}_v, y_{J^{[\pi]}}) = L_v(\hat{x}_v, y) = f_v(y). \tag{4.22}$$

Proof. Together with

$$f_v^{[\pi]}(y_{J^{[\pi]}}) \overset{(4.15)}{=} \max_{x_v\in\Omega_v} L_v^{[\pi]}(x_v, y_{J^{[\pi]}}) \overset{(4.20)(C3)}{=} h_v(\hat{x}_v) - (y_{J_v^{\{\pi\}}})^T A_{J_v^{\{\pi\}},v}\hat{x}_v \overset{(4.20)}{=} L_v^{[\pi]}(\hat{x}_v, y_{J^{[\pi]}})$$

this is a direct consequence of (4.1). □

Observation 27 *Given a process π, $v \in V^{[\pi]}$, $y \in \mathbb{R}^M$ with $y_{\bar{J}_v^{[\pi]}} = \hat{y}^{\{\pi\}}_{\bar{J}_v^{[\pi]}}$, and $(l_v, x_v) \in \operatorname{conv} W_v$, suppose x_v is $(y, J_v^{\{\pi\}})$-consistent. Then*

$$\hat{f}^{[\pi]}_{(l_v,x_v),v}(y_{J^{[\pi]}}) = \hat{f}_{(l_v,x_v),v}(y). \tag{4.23}$$

Proof. $\hat{f}^{[\pi]}_{(l_v,x_v),v}(y_{J^{[\pi]}}) \overset{(4.21)}{=} l_v - y^T_{J_v^{\{\pi\}}} A_{J_v^{\{\pi\}},v} x_v \overset{(C1)}{=} l_v - y^T A_{\bullet,v} x_v \overset{(2.4)}{=} \hat{f}_{(l_v,x_v),v}(y).$ □

Observation 28 *Given a process π and an $x \in \operatorname{conv}\Omega$ satisfying*

$$A_{j,v}x_v = 0 \quad \text{for } v \in V_j \setminus V_j^{[\pi]}, j \in J^{[\pi]}, \tag{4.24}$$

there holds

$$g^{[\pi]}(x_{V^{[\pi]}}) = g(x)_{J^{[\pi]}}. \tag{4.25}$$

Proof. We use $A_{j,V\setminus V_j} = 0$ for all $j \in M$ (see the definition (3.2) of V_j),

$$\begin{aligned} g^{[\pi]}(x_{V^{[\pi]}}) &\overset{(4.18)}{=} b_{J^{[\pi]}} - \left[\sum_{v\in V_j^{[\pi]}} A_{j,v}x_v + \sum_{v\in V_j\setminus V_j^{[\pi]}} \underbrace{A_{j,v}x_v}_{=0 \text{ by } (4.24)} \right]_{j\in J^{[\pi]}} \\ &= b_{J^{[\pi]}} - A_{J^{[\pi]},\bullet}x \overset{(2.3)}{=} g(x)_{J^{[\pi]}}. \end{aligned}$$

□

4.5. The Parallel Bundle Algorithm for Stronger Coupling

Now we present the extended parallel bundle algorithm in detail. While step 2.(d) needs several modifications, steps 1. and 2.(a)–(c) require only minor adaptations and we mark these by boxes.

Algorithm *Extended Parallel Bundle:*

1. **Initialization**
   - Choose parameters
     - *relative predicted subspace decrease level* $\tau_1 \in (0, \frac{1}{2}\min\{\frac{1}{|M|}, \frac{1}{|V|}\})$,
     - *subspace improvement level* $\tau_2 \in (0, 1)$,
     - *subspace dependency level* $\tau_3 \in [0, 1 - \tau_2)$ and
     - *termination precision* $\varepsilon > 0$,
     - *maximal number of parallel processes* $N_\Pi \leq |M|$.
     - *weight* $u > 0$ and *descent step parameter* $\rho \in (0, 1)$ for the bundle method.
   - Set $\sigma \leftarrow 0$, $\hat{y}^{\{0\}} \leftarrow 0$, $B^{\{0\}} \leftarrow \emptyset$, $\boxed{B_M^{\{0\}} \leftarrow \emptyset}$.
   - Set $E^{\{0\}} \leftarrow \emptyset$ (or use some prespecified dependencies).
   - Determine all $f^{\{0\}} = f_v(\hat{y}^{\{0\}})$ and set $(\bar{l}^{\{0\}}, \bar{x}^{\{0\}})$ to the minorant defined by the optimal solutions.

- $\boxed{\text{Set } J_v^{\{0\}} \leftarrow \{j \in M \colon A_{j,v}\bar{x}_v^{\{0\}} \neq 0\} \text{ for all } v \in V}$ and possibly enlarge them further until each is $(\hat{y}^{\{0\}}, v)$-consistent.
- If $\Delta^{\{0\}} \leq \varepsilon(|f(\hat{y}^{\{0\}})| + 1)$ then **STOP**.

2. While less than N_Π processes are running, start a new process π setting $\underline{\pi} \leftarrow -1$, $\overline{\pi} \leftarrow -1$. Each process π does the following (with semaphores guarding the access).

   (a) **Subspace selection**

   - Secure exclusive access to global data, set $\underline{\pi} \leftarrow \sigma$.
   - Find $\hat{\jmath} \in M$ and a subspace $J^{[\pi]} \subseteq M$ with

     [S1] $\{j \colon (\hat{\jmath}, j) \in E^{\{\underline{\pi}\}}\} \subseteq J^{[\pi]}$,

     [S2] $\Delta_\pi^{\{\underline{\pi}\}} \geq \tau_1 \Delta^{\{\underline{\pi}\}}$,

     [S3] $V^{[\pi]} \cap B^{\{\underline{\pi}\}} = \emptyset$,

     [S4] $\boxed{B_M^{\{\underline{\pi}\}} \cap (J^{[\pi]} \cup \bar{J}^{[\pi]}) = \emptyset}$.
   - If no such $\hat{\jmath}$ exists, the step is *unsuccessful*, do not change the global data, free access to it and restart at some later time at (a) when $\sigma > \underline{\pi}$.
   - The step is *successful*, set $B^{\{\sigma+1\}} \leftarrow B^{\{\underline{\pi}\}} \cup V^{[\pi]}$, $\boxed{B_M^{\{\sigma+1\}} \leftarrow B_M^{\{\underline{\pi}\}} \cup J^{[\pi]}}$, $\overline{\pi} \leftarrow \infty$ and keep

$$\left(\hat{y}^{\{\sigma+1\}}, (\bar{l}^{\{\sigma+1\}}, \bar{x}^{\{\sigma+1\}}), f_V^{\{\sigma+1\}}, E^{\{\sigma+1\}}, \boxed{J_V^{\{\sigma+1\}}}\right) \leftarrow \left(\hat{y}^{\{\underline{\pi}\}}, (\bar{l}^{\{\underline{\pi}\}}, \bar{x}^{\{\underline{\pi}\}}), f_V^{\{\underline{\pi}\}}, E^{\{\underline{\pi}\}}, \boxed{J_V^{\{\underline{\pi}\}}}\right). \quad (4.26)$$

   - Read required global data, set $\sigma \leftarrow \underline{\pi} + 1$ and free exclusive access.

   (b) **Set up the subspace problem.**
   There is no global interaction in this step!
   Compute $c_v^{[\pi]}$ by (4.12) for all $v \in V^{[\pi]}$ and initialize the bundle method to solve problem $(SP^{[\pi]})$ with this $c^{[\pi]}$ starting from center $\hat{y}^{[\pi]} = \hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}$ and initial aggregate $(\bar{l}^{[\pi]}, \bar{x}^{[\pi]}) = (\bar{l}_{V^{[\pi]}}^{\{\underline{\pi}\}}, \bar{x}_{V^{[\pi]}}^{\{\underline{\pi}\}})$. In the first iteration we have then exactly a predicted decrease

$$\Delta^{[\pi]}(\hat{y}^{[\pi]}, (\bar{l}^{[\pi]}, \bar{x}^{[\pi]})) = \Delta_\pi^{\{\underline{\pi}\}}.$$

   (c) **Solve the subspace problem.**
   There is no global interaction in this step!
   Solve $(SP^{[\pi]})$ by the bundle method, iteratively generating candidates $\bar{y}^{[\pi]}$ together with primal aggregates $(\bar{l}^{[\pi]}, \bar{x}^{[\pi]}) \in \operatorname{conv} W^{[\pi]}$. In each iteration test the following conditions in this sequence.

   1. If the predicted decrease on the subspace $J^{[\pi]}$ has been sufficiently reduced, *i. e.*,

$$\Delta^{[\pi]}(\hat{y}^{[\pi]}, (\bar{l}^{[\pi]}, \bar{x}^{[\pi]})) < \tau_2 \Delta_\pi^{\{\underline{\pi}\}}, \quad \text{(StopP')}$$

   $\boxed{\text{consider } \hat{x}^{[\pi]} \text{ to be the original optimizer at } \hat{y}^{[\pi]}}$ and go to (d).

2. If $\Delta^{[\pi]}(\hat{y}^{[\pi]}, (\bar{l}^{[\pi]}, \bar{x}^{[\pi]})) \geq \tau_2 \Delta_\pi^{\{\pi\}}$ and the descent step criterion

$$\Delta^{[\pi]}(\hat{y}^{[\pi]}, (\bar{l}^{[\pi]}, \bar{x}^{[\pi]})) \leq \tfrac{1}{\rho}(f^{[\pi]}(\hat{y}^{[\pi]}) - f^{[\pi]}(\bar{y}^{[\pi]})), \tag{StopD'}$$

is satisfied, set $\hat{y}^{[\pi]} \leftarrow \bar{y}^{[\pi]}$, $\boxed{\text{let } \hat{x}^{[\pi]} \text{ be the optimizer at } \bar{y}^{[\pi]}}$ and go to (d).

(d) **Update global data with subspace solution.**

- Secure exclusive access to global data and set $\overline{\pi} \leftarrow \sigma$.
- Independent of (StopP') and (StopD'), keep

$$\begin{aligned}&\left(\hat{y}^{\{\sigma+1\}}_{M\setminus J^{[\pi]}}, (\bar{l}^{\{\sigma+1\}}_{V\setminus V^{[\pi]}}, \bar{x}^{\{\sigma+1\}}_{V\setminus V^{[\pi]}}), f^{\{\sigma+1\}}_{V\setminus V^{[\pi]}}\right) \leftarrow \\ &\qquad \left(\hat{y}^{\{\overline{\pi}\}}_{M\setminus J^{[\pi]}}, (\bar{l}^{\{\overline{\pi}\}}_{V\setminus V^{[\pi]}}, \bar{x}^{\{\overline{\pi}\}}_{V\setminus V^{[\pi]}}), f^{\{\overline{\pi}\}}_{V\setminus V^{[\pi]}}\right)\end{aligned} \tag{4.27}$$

and update the subterms

$$B^{\{\sigma+1\}} \leftarrow B^{\{\overline{\pi}\}} \setminus V^{[\pi]}, B_M^{\{\sigma+1\}} \leftarrow B_M^{\{\overline{\pi}\}} \setminus J^{[\pi]} \tag{4.28}$$

- Keep dependencies $J^{\{\sigma+1\}}_{V\setminus V_{J^{[\pi]}}} \leftarrow J^{\{\overline{\pi}\}}_{V\setminus V_{J^{[\pi]}}}$, update for all $v \in V^{[\pi]}$

$$\begin{aligned}J_v^{\{\sigma+1\}} \leftarrow J_v^{\{\overline{\pi}\}} \cup \{j \in & J_v \setminus J_v^{\{\overline{\pi}\}} \colon A_{j,v}\hat{x}_v^{[\pi]} \neq 0 \text{ or } \ A_{j,v}\bar{x}_v^{[\pi]} \neq 0 \ \text{ or} \\ & (j \in J^{[\pi]} \ \wedge \ \hat{y}_j^{[\pi]} \cdot A_{j,v}x_v < 0 \text{ for some } x_v \in \Omega_v)\}\end{aligned} \tag{4.29}$$

and for all $v \in V_{J^{[\pi]}} \setminus V^{[\pi]}$

$$\begin{aligned}J_v^{\{\sigma+1\}} \leftarrow J_v^{\{\overline{\pi}\}} \cup \{j \in J^{[\pi]} \cap J_v \setminus J_v^{\{\overline{\pi}\}} \colon \hat{y}_j^{[\pi]} \cdot A_{j,v}x_v < 0 \text{ for some} \\ x_v \in \Omega_v\}\end{aligned} \tag{4.30}$$

- Ensure consistency of the new solution by testing the following conditions:

$$\hat{y}_j^{[\pi]} A_{j,v}x_v \geq 0 \quad \text{for } v \in V^{[\pi]}, x_v \in \Omega_v, j \in J^{[\pi]} \cap (J_v^{\{\sigma+1\}} \setminus J_v^{\{\underline{\pi}\}}), \tag{4.31}$$

$$\hat{y}_j^{[\pi]} A_{j,v}\hat{x}_v^{[\pi]} = 0 = \hat{y}_j^{[\pi]} A_{j,v}\bar{x}_v^{[\pi]} \quad \text{for } v \in V^{[\pi]}, j \in J^{[\pi]} \cap (J_v^{\{\sigma+1\}} \setminus J_v^{\{\underline{\pi}\}}), \tag{4.32}$$

$$\hat{y}_j^{\{\overline{\pi}\}} A_{j,v}\hat{x}_v^{[\pi]} = 0 = \hat{y}_j^{\{\overline{\pi}\}} A_{j,v}\bar{x}_v^{[\pi]} \quad \text{for } v \in V^{[\pi]}, j \in (J_v^{\{\sigma+1\}} \setminus J_v^{\{\underline{\pi}\}}) \setminus J^{[\pi]}, \tag{4.33}$$

$$J_v^{\{\sigma+1\}} \cap J^{[\pi]} = \emptyset \quad \text{for } v \in V_{J^{[\pi]}} \setminus V^{[\pi]}. \tag{4.34}$$

If all of them hold, call π *good* and update

$$(\bar{l}^{\{\sigma+1\}}_{V^{[\pi]}}, \bar{x}^{\{\sigma+1\}}_{V^{[\pi]}}) \leftarrow (\bar{l}^{[\pi]}, \bar{x}^{[\pi]}), \tag{4.35}$$

otherwise call π *bad* and keep $(\bar{l}^{\{\sigma+1\}}_{V^{[\pi]}}, \bar{x}^{\{\sigma+1\}}_{V^{[\pi]}}) \leftarrow (\bar{l}^{\{\overline{\pi}\}}_{V^{[\pi]}}, \bar{x}^{\{\overline{\pi}\}}_{V^{[\pi]}})$.

- If π is good and stopped with (StopD'), update

$$\hat{y}^{\{\sigma+1\}}_{J^{[\pi]}} \leftarrow \hat{y}^{[\pi]}, \tag{4.36}$$

$$f^{\{\sigma+1\}}_{V^{[\pi]}} \leftarrow f^{[\pi]}_{V^{[\pi]}}(\hat{y}^{[\pi]}), \tag{4.37}$$

otherwise keep

$$\left(\hat{y}^{\{\sigma+1\}}_{J^{[\pi]}}, f^{\{\sigma+1\}}_{V^{[\pi]}}\right) \leftarrow \left(\hat{y}^{\{\overline{\pi}\}}_{J^{[\pi]}}, f^{\{\overline{\pi}\}}_{V^{[\pi]}}\right). \tag{4.38}$$

- If π is good, stopped with (StopP'), and satisfies

$$\delta_{\pi}^{\{\sigma+1\}} - \delta_{\pi}^{\{\overline{\pi}\}} > \tau_3 \Delta_{\pi}^{\{\underline{\pi}\}} \tag{Dep}$$

  - enlarge $E^{\{\sigma+1\}} \leftarrow E^{\{\overline{\pi}\}} \cup \{(\hat{\jmath}, j^*)\}$ at least by an arc $(\hat{\jmath}, j^*)$ with

$$j^* \in \text{Argmax}\{g(\bar{x}^{\{\sigma+1\}})_j^2 - g(\bar{x}^{\{\overline{\pi}\}})_j^2 : j \in \bar{J}^{[\pi]}\},$$

  in all other cases keep $E^{\{\sigma+1\}} \leftarrow E^{\{\overline{\pi}\}}$.
- Set $\sigma \leftarrow \overline{\pi} + 1$ and compute $\Delta^{\{\sigma\}}$. If π is good and $\Delta^{\{\sigma\}} \le \varepsilon(|f(\hat{y}^{\{\sigma\}})| + 1)$ terminate *all* processes and **STOP**.
- Free access to global data and stop this process π.

The proofs below will show that in actual implementations the availability of $\hat{x}^{[\pi]}$ required for (4.29), (4.32) and (4.33) is only needed in the case of (StopD') but not for (StopP'), because in the latter case the property $A_{j,v}\hat{x}_v^{[\pi]} = 0$ holds for $j \in J_v \setminus J_v^{\{\underline{\pi}\}}$, $v \in V^{[\pi]}$ by induction. It is not needed at any other place and so we do not store it in the global data. Note also that the consistency tests (4.31)–(4.33) need to be performed w. r. t. to the initial dependencies $J_v^{\{\underline{\pi}\}}$ instead of the global $J_v^{\{\overline{\pi}\}}$ attained when the process stops. Indeed, the process has to verify that the assumptions under which it solved its subproblem are still true when it stops.

For the analysis we will collect, for each value $\sigma \in \mathbb{N}_0 \cup \{\infty\}$, the processes in the sets

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

in the same way and with the same interpretation as before.

Observation 29 $\Pi^{\{0\}} = \emptyset$ *and for* $\sigma' \in \mathbb{N}_0$ *there holds* $\Pi^{\{\sigma'\}} \neq \Pi^{\{\sigma'+1\}}$ *if and only if* $|(\Pi^{\{\sigma'\}} \setminus \Pi^{\{\sigma'+1\}}) \cup (\Pi^{\{\sigma'+1\}} \setminus \Pi^{\{\sigma'\}})| = 1$. *If* $\Pi^{\{\sigma'\}} = \Pi^{\{\sigma'+1\}}$ *then for all* $\sigma \ge \sigma'$ *we have* $\Pi^{\{\sigma\}} = \Pi^{\{\sigma'\}}$ *and there is no process* π *with* $\underline{\pi} = \sigma$ *or* $\overline{\pi} = \sigma$.

Proof. Identical to the proof of Observation 5. □

The set of markers $\sigma \in \mathbb{N}_0$ visited by the algorithm will again be denoted by

$$\Sigma := \{0\} \cup \{\sigma \in \mathbb{N} : \Pi^{\{\sigma\}} \neq \Pi^{\{\sigma-1\}}\}.$$

4.6. Consistency of the Updating Scheme

In the new setting the two conditions [S3] and [S4] ensure that two processes running in parallel choose subspaces in a coordinated way.

Lemma 30 *For* $\sigma \in \Sigma$,

(i) $B^{\{\sigma\}} = \bigcup_{\pi \in \Pi^{\{\sigma\}}} V^{[\pi]}$, $B_M^{\{\sigma\}} = \bigcup_{\pi \in \Pi^{\{\sigma\}}} J^{[\pi]}$,

(ii) $V^{[\pi]} \cap V^{[\pi']} = \emptyset$ *for* $\pi, \pi' \in \Pi^{\{\sigma\}}$ *with* $\pi \neq \pi'$,

(iii) $J^{[\pi']} \cap (J^{[\pi]} \cup \bar{J}^{[\pi]}) = \emptyset = J^{[\pi]} \cap (J^{[\pi']} \cup \bar{J}^{[\pi']})$ *for* $\pi, \pi' \in \Pi^{\{\sigma\}}$ *with* $\pi \neq \pi'$,

(iv) $\hat{y}_{J^{[\pi]} \cup \bar{J}^{[\pi]}}^{\{\underline{\pi}\}} = \hat{y}_{J^{[\pi]} \cup \bar{J}^{[\pi]}}^{\{\sigma\}}$, $(\bar{l}_{V^{[\pi]}}^{\{\underline{\pi}\}}, \bar{x}_{V^{[\pi]}}^{\{\underline{\pi}\}}) = (\bar{l}_{V^{[\pi]}}^{\{\sigma\}}, \bar{x}_{V^{[\pi]}}^{\{\sigma\}})$, *and* $f_{V^{[\pi]}}^{\{\underline{\pi}\}} = f_{V^{[\pi]}}^{\{\sigma\}}$ *for* $\pi \in \Pi^{\{\sigma\}}$,

(v) $J_v^{\{\sigma'\}} \subseteq J_v^{\{\sigma\}} \subseteq J_v$ *for* $\sigma' \in \Sigma$ *with* $\sigma' < \sigma$ *and* $v \in V$,

(vi) $J_v^{[\pi]} = J_v^{\{\underline{\pi}\}} \cap J^{[\pi]} = J_v^{\{\sigma\}} \cap J^{[\pi]}$ *for* $v \in V^{[\pi]}$, $\pi \in \Pi^{\{\sigma\}}$,

(vii) *(E1), (E2), (E3) are satisfied for* σ.

(viii) $J_v^{\{\underline{\pi}\}}$ *is* $(\hat{y}^{\{\sigma\}}, v)$*-consistent for a witness* $\hat{x}_v$ *with* $A_{M \setminus J_v^{\{\underline{\pi}\}}} \hat{x}_v = 0$ *for* $v \in V^{[\pi]}$, $\pi \in \Pi^{\{\sigma\}}$.

(ix) $\bar{x}_v^{\{\sigma\}}$ *is* $(\hat{y}^{\{\sigma\}}, J_v^{\{\underline{\pi}\}})$*-consistent with* $A_{M \setminus J_v^{\{\underline{\pi}\}}} \bar{x}_v^{\{\sigma\}} = 0$ *for* $v \in V^{[\pi]}$, $\pi \in \Pi^{\{\sigma\}}$.

Proof. The proof works by induction on σ. For $\sigma = 0$ the initialization step sets $B^{\{\sigma\}} = \emptyset = B_M^{\{\sigma\}}$ and $J_v^{\{\sigma\}} \subseteq J_v$ $(v \in V)$ satisfying (E1)–(E3), because $\hat{y}^{\{\sigma\}} = 0$, $\hat{x}^{\{\sigma\}} = \bar{x}^{\{\sigma\}}$ and $A_{j,v}\bar{x}_v^{\{\sigma\}} \neq 0 \Rightarrow j \in J_v^{\{\sigma\}}$ for $v \in V$. Together with $\Pi^{\{\sigma\}} = \emptyset$ this proves (i)–(ix).

Now suppose $\sigma + 1 \in \Sigma$ and the claim holds for $\sigma \in \Sigma$. By definition, $\sigma + 1 \in \Sigma$ implies $\Pi^{\{\sigma\}} \neq \Pi^{\{\sigma+1\}}$, so Observation 29 asserts the existence of a unique $\eta \in (\Pi^{\{\sigma\}} \setminus \Pi^{\{\sigma+1\}}) \cup (\Pi^{\{\sigma+1\}} \setminus \Pi^{\{\sigma\}})$ and this η either satisfies $\underline{\eta} = \sigma$ or $\overline{\eta} = \sigma$.

Case $\underline{\eta} = \sigma$:
We have $\Pi^{\{\sigma\}} = \Pi^{\{\sigma+1\}} \setminus \{\eta\}$ and process η executed a successful step (a) at σ. In this case (4.26) implies that most of the global data remains unchanged.

(i), (iv)–(vii): These hold because in a successful step (a), $B^{\{\sigma+1\}} = B^{\{\sigma\}} \cup V^{[\eta]}$, $B_M^{\{\sigma+1\}} = B_M^{\{\sigma\}} \cup J^{[\eta]}$ are the only global objects changing.

(ii), (iii): By induction, these only need to be verified for $\pi = \eta$ and $\pi' \in \Pi^{\{\sigma+1\}} \setminus \{\eta\} = \Pi^{\{\sigma\}}$. Because $V^{[\pi']} \subseteq B^{\{\sigma\}}$ by (i) and [S3] was satisfied for η at σ, (ii) follows from $\emptyset = V^{[\eta]} \cap B^{\{\sigma\}} \supseteq V^{[\eta]} \cap V^{[\pi']}$. Applying the corresponding argument with [S4] proves the left hand side equation of (iii). For the right hand side equation of (iii) assume there exists a $j \in J^{[\eta]} \cap (J^{[\pi']} \cup \bar{J}^{[\pi']})$, then by [S4] we have $j \in \bar{J}^{[\pi']}$. By definition (4.4) there is a $v \in V^{[\pi']}$ with $j \in J_v^{\{\underline{\pi}'\}}$. Now $\underline{\pi}' < \underline{\eta} = \sigma$ by the choice of η, so by (v) $J_v^{\{\underline{\pi}'\}} \subseteq J_v^{\{\underline{\eta}\}}$ and with $j \in J^{[\eta]}$ we conclude $v \in V^{[\eta]}$ by (4.3), contradicting (ii).

(viii),(ix): Because the data involved in these conditions is not modified, the conditions hold by induction for $\pi' \in \Pi^{\{\sigma+1\}} \setminus \{\eta\}$. So consider η, fix some $v \in V^{[\eta]}$ and observe $J_v^{[\eta]} \cup \bar{J}_v^{[\eta]} \overset{(4.5)}{=} J_v^{\{\underline{\eta}\}} \overset{(4.26)}{=} J_v^{\{\sigma+1\}}$ (and $\sigma + 1 = \underline{\eta} + 1$). So (viii) is guaranteed by Observation 21 because (E1),(E2) hold for $\sigma + 1$ by (vii). Likewise, (E3) establishes (ix).

Case $\overline{\eta} = \sigma$:
We have $\Pi^{\{\sigma+1\}} = \Pi^{\{\sigma\}} \setminus \{\eta\}$ and process η executed a step (d) at σ.

(i)–(iv): By step (d), $B^{\{\sigma+1\}} = B^{\{\sigma\}} \setminus V^{[\eta]}$, $B_M^{\{\sigma+1\}} = B^{\{\sigma\}} \setminus J^{[\eta]}$ and for indices $M \setminus J^{[\eta]}$ and $V \setminus V^{[\eta]}$ all values relevant for (iv) are left unchanged. In view of the validity of (ii) and (iii) for $\pi = \eta$ at σ, (i)–(iv) hold by induction also for $\sigma + 1$ and its remaining processes.

(v): Its correctness follows directly from the operations performed on $J_v^{\{\sigma+1\}}$ for $v \in V$ in step (d).

(vi): Because of (v) and the induction hypothesis it suffices to show $J_v^{\{\sigma+1\}} \cap J^{[\pi]} \subseteq J_v^{\{\underline{\pi}\}} \cap J^{[\pi]} = J_v^{\{\sigma\}} \cap J^{[\pi]}$ for $v \in V^{[\pi]}$, $\pi \in \Pi^{\{\sigma+1\}}$. Assume, for contradiction, for some $\pi' \in \Pi^{\{\sigma+1\}} = \Pi^{\{\sigma\}} \setminus \{\eta\}$ there is a $v \in V^{[\pi']}$ with some $j \in (J_v^{\{\sigma+1\}} \cap J^{[\pi']}) \setminus J_v^{\{\sigma\}}$. Now $v \notin V^{[\eta]}$ by (ii) for $\pi = \eta$ at σ and, as this j was added to $J_v^{\{\sigma\}}$ by step (d) of η at σ, this forces $j \in J^{[\eta]}$ which contradicts (iii) for $\pi = \eta$ at σ.

(vii): We first check (E3). For $v \in V \setminus V^{[\eta]}$ there holds $\bar{x}_v^{\{\sigma+1\}} = \bar{x}_v^{\{\sigma\}}$ by (4.27) and Observation 22 (ii) together with (v) guarantees that (E3) also holds at $\sigma + 1$. For $v \in V^{[\eta]}$ and a bad η, we have $\bar{x}_v^{\{\sigma+1\}} = \bar{x}_v^{\{\sigma\}}$ which allows the same argument as before. If η is good, the update (4.35) ensures $\bar{x}_v^{\{\sigma+1\}} = \bar{x}_v^{[\eta]}$ and all indices $j \in M \setminus J_v^{\{\underline{\eta}\}}$ that violate $A_{j,v}\bar{x}_v^{[\eta]} = 0$ are explicitly included in $J_v^{\{\sigma+1\}}$, so that (E3) is satisfied for these v, as well.

For checking (E1) and (E2) first consider the case that η is a bad process or stopped by (StopP'). Then $\hat{y}^{\{\sigma+1\}} = \hat{y}^{\{\sigma\}}$ and by (v) and Observation 22 (i) conditions (E1) and (E2) hold for $\sigma + 1$. It remains to consider the case that η is good and stopped by (StopD').

For (E1) only the indices $j \in J^{[\eta]}$ need to be checked because by (4.27) the other conditions do not change and hold by induction. For $j \in J^{[\eta]}$ (3.1) implies $A_{j,v} = 0$ unless $v \in V_{J^{[\eta]}}$, so the update (4.29) and (4.30) to $J_v^{\{\sigma+1\}}$ in (d) adds all indices that do not yet satisfy the condition and (E1) holds at $\sigma + 1$.

We split the verification of (E2) into the three cases $v \in V \setminus V_{J^{[\eta]}}$, $v \in V_{J^{[\eta]}} \setminus V^{[\eta]}$ and $v \in V^{[\eta]}$. If $v \in V \setminus V_{J^{[\eta]}}$ then the dependency update (4.29) and (4.30) do not change $J_v^{\{\sigma\}}$ and thus $J_v^{\{\sigma\}} = J_v^{\{\sigma+1\}}$. Furthermore definition (3.3) implies $J_v^{\{\sigma+1\}} \cap J^{[\eta]} = \emptyset$. If $v \in V_{J^{[\eta]}} \setminus V^{[\eta]}$ then the dependency update (4.29) and (4.30) together with the successful test (4.34) ensure $J_v^{\{\sigma\}} = J_v^{\{\sigma+1\}}$ and $J_v^{\{\sigma+1\}} \cap J^{[\eta]} = \emptyset$ as well. The induction hypothesis for (E2) asserts the existence of a $\hat{x}_v \in \operatorname{Argmax}\{h_v(x_v) - (\hat{y}^{\{\sigma\}}_{J_v^{\{\sigma\}}})^T A_{J_v^{\{\sigma\}},v} x_v \colon x_v \in \Omega_v\}$ with $A_{j,v}\hat{x}_v = 0$ for $j \in M \setminus J_v^{\{\sigma\}} = M \setminus J_v^{\{\sigma+1\}}$. Regarding $J_v^{\{\sigma\}} = J_v^{\{\sigma+1\}}$, $J_v^{\{\sigma+1\}} \cap J^{[\eta]} = \emptyset$ and the update (4.27) the relevant multipliers have not changed, *i. e.* $\hat{y}^{\{\sigma\}}_{J_v^{\{\sigma\}}} = \hat{y}^{\{\sigma+1\}}_{J_v^{\{\sigma+1\}}}$, and $\hat{x}_v$ remains a valid witness for the correctness of (E2) at $\sigma + 1$.

The case $v \in V^{[\eta]}$ will be useful in further proofs, as well:

Claim: *For a good process η stopped by (StopD') and $v \in V^{[\eta]}$ the point $\hat{x}_v^{[\eta]}$ is a valid witness for $J_v^{\{\underline{\eta}\}}$ to be $(\hat{y}^{\{\sigma+1\}}, v)$-consistent with $A_{M \setminus J_v^{\{\sigma+1\}},v}\hat{x}_v^{[\eta]} = 0$.*

Indeed, update (4.29) in (d) to $J_v^{\{\sigma+1\}}$ includes all indices $j \in J_v \setminus J_v^{\{\sigma\}}$ with $A_{j,v}\hat{x}_v^{[\eta]} \neq 0$, so $A_{j,v}\hat{x}_v^{[\eta]} = 0$ for $j \in J_v \setminus J_v^{\{\sigma+1\}}$. Next we proof that (C2) holds. We have to show that $\hat{y}_j^{\{\sigma+1\}} A_{j,v} x_v \geq 0$ for all $x_v \in \Omega_v$ and all $j \in J_v \setminus J_v^{\{\underline{\eta}\}}$.

- If $j \in (J_v \setminus J_v^{\{\underline{\eta}\}}) \setminus J^{[\eta]}$ then $\hat{y}_j^{\{\sigma\}} = \hat{y}_j^{\{\sigma+1\}}$ by (4.27) and the induction hypothesis for (viii) for $\sigma = \overline{\eta}$ and $\pi = \eta$ implies the assertion.
- If $j \in (J_v \setminus J_v^{\{\sigma+1\}}) \cap J^{[\eta]}$ then the dependency update (4.29) implies the assertion.
- If $j \in (J_v^{\{\sigma+1\}} \setminus J_v^{\{\underline{\eta}\}}) \cap J^{[\eta]}$ then the consistency test (4.31) implies the assertion.

Putting all together (C2) holds. Similarly, the dependency update (4.29) together with the successful tests (4.32) and (4.33) certify that $\hat{x}_v^{[\eta]}$ satisfies (C1) for $J_v^{\{\underline{\eta}\}}$.

Finally regarding the choice of $\hat{x}^{[\eta]} \in \operatorname{Argmax}\{L^{[\eta]}(x_v, \hat{y}^{[\eta]})\colon x_v \in \Omega_v\}$ and the facts $\hat{y}^{\{\sigma+1\}}_{J^{[\eta]}} = \hat{y}^{[\pi]}$ by (4.36) and $\hat{y}^{\{\sigma+1\}}_{\bar{J}^{[\eta]}_v} = \hat{y}^{\{\eta\}}_{\bar{J}^{[\eta]}_v}$ by (4.27) and (iv), (4.20) establishes that $\hat{x}^{[\eta]}_v$ satisfies (C3), *i. e.*, $\hat{x}^{[\eta]}_v$ is a witness for $J^{\{\eta\}}_v$ to be $(\hat{y}^{\{\sigma+1\}}, v)$-consistent. This completes the proof of the claim.

Now Observation 20 allows to conclude that $J^{\{\sigma+1\}}_v$ is $(\hat{y}^{\{\sigma+1\}}, v)$-consistent with witness $\hat{x}^{[\eta]}_v$ and (E2) holds.

(viii): Fix $\pi' \in \Pi^{\{\sigma+1\}}$ and $v \in V^{[\pi']}$. By induction there exists a witness $\hat{x}_v$ with $A_{M\setminus J^{\{\underline{\pi}'\}}_v, v}\hat{x}_v = 0$ for $J^{\{\underline{\pi}'\}}_v$ to be $(\hat{y}^{\{\sigma\}}, v)$-consistent and we show that step (d) of η does not invalidate this witness. In view of (iv) and $J^{\{\underline{\pi}'\}}_v \subseteq \bar{J}^{[\pi']} \cup J^{[\pi']}$ by (4.5), (C3) holds because $\hat{x}_v$ satisfies (C1) trivially. By (3.1) process η has no influence on the validity of (C2) unless there is some $j \in J^{[\eta]} \cap J_v$, in this case $v \in V_{J^{[\eta]}} \setminus V^{[\eta]}$ by (ii). If the value $\hat{y}^{[\eta]}_j$ leads to a violation of (C2) for any such j, this j is included in $J^{\{\sigma+1\}}_v$ in the corresponding update (4.30) in step (d), the subsequent test (4.34) fails and renders η a bad process resulting in $\hat{y}^{\{\sigma+1\}} = \hat{y}^{\{\sigma\}}$, which satisfies (C2) by induction.

(ix): Fix $\pi' \in \Pi^{\{\sigma+1\}}$ and $v \in V^{[\pi']}$. By (iv) we know $\bar{x}^{\{\underline{\pi}'\}}_v = \bar{x}^{\{\sigma+1\}}_v$ and applying (E3) for $\sigma = \underline{\pi}'$ implies the assertion by induction. □

The inductive properties allow to clarify the development of the global values relevant for a process during its running time.

Lemma 31 *Given $\pi \in \overline{\Pi}^{\{\infty\}}$ assume $f^{\{\underline{\pi}\}}_{V^{[\pi]}} = f_{V^{[\pi]}}(\hat{y}^{\{\underline{\pi}\}})$. Then for all $\sigma \in \{\underline{\pi}, \dots, \overline{\pi}\}$*

$$\hat{y}^{\{\underline{\pi}\}}_{J^{[\pi]}\cup\bar{J}^{[\pi]}} = \hat{y}^{\{\sigma\}}_{J^{[\pi]}\cup\bar{J}^{[\pi]}}, \qquad f^{[\pi]}_{V^{[\pi]}}(\hat{y}^{\{\sigma\}}_{J^{[\pi]}}) = f^{\{\underline{\pi}\}}_{V^{[\pi]}} = f^{\{\sigma\}}_{V^{[\pi]}} = f_{V^{[\pi]}}(\hat{y}^{\{\sigma\}}), \tag{4.39}$$

$$(\bar{l}^{\{\underline{\pi}\}}_{V^{[\pi]}}, \bar{x}^{\{\underline{\pi}\}}_{V^{[\pi]}}) = (\bar{l}^{\{\sigma\}}_{V^{[\pi]}}, \bar{x}^{\{\sigma\}}_{V^{[\pi]}}), \qquad \hat{f}^{[\pi]}_{(\bar{l}^{\{\sigma\}}_v, \bar{x}^{\{\sigma\}}_v), v}(\hat{y}^{\{\sigma\}}_{J^{[\pi]}}) = \hat{f}_{(\bar{l}^{\{\sigma\}}_v, \bar{x}^{\{\sigma\}}_v), v}(\hat{y}^{\{\sigma\}}) \quad (v \in V^{[\pi]}), \tag{4.40}$$

and

$$\hat{y}^{\{\overline{\pi}+1\}}_{M\setminus J^{[\pi]}} = \hat{y}^{\{\overline{\pi}\}}_{M\setminus J^{[\pi]}}, \quad (\bar{l}^{\{\overline{\pi}+1\}}_{V\setminus V^{[\pi]}}, \bar{x}^{\{\overline{\pi}+1\}}_{V\setminus V^{[\pi]}}) = (\bar{l}^{\{\overline{\pi}\}}_{V\setminus V^{[\pi]}}, \bar{x}^{\{\overline{\pi}\}}_{V\setminus V^{[\pi]}}), \quad f^{\{\overline{\pi}+1\}}_{V\setminus V^{[\pi]}} = f^{\{\overline{\pi}\}}_{V\setminus V^{[\pi]}}, \tag{4.41}$$

in particular,

$$\hat{y}^{\{\underline{\pi}\}}_{\bar{J}^{[\pi]}_v} = \hat{y}^{\{\sigma\}}_{\bar{J}^{[\pi]}_v} \qquad \text{for all} \quad v \in V^{[\pi]}, \ \sigma \in \{\underline{\pi}, \dots, \overline{\pi}+1\}. \tag{4.42}$$

Furthermore, if π is bad,

$$\hat{y}^{\{\overline{\pi}+1\}}_{J^{[\pi]}} = \hat{y}^{\{\overline{\pi}\}}_{J^{[\pi]}}, \qquad f^{\{\overline{\pi}+1\}}_{V^{[\pi]}} = f_{V^{[\pi]}}(\hat{y}^{\{\overline{\pi}+1\}}), \tag{4.43}$$

$$(\bar{l}^{\{\overline{\pi}+1\}}_{V^{[\pi]}}, \bar{x}^{\{\overline{\pi}+1\}}_{V^{[\pi]}}) = (\bar{l}^{\{\overline{\pi}\}}_{V^{[\pi]}}, \bar{x}^{\{\overline{\pi}\}}_{V^{[\pi]}}). \tag{4.44}$$

Otherwise π is good and with $\hat{y}^{[\pi]} \in \mathbb{R}^{J^{[\pi]}}$ and $(\bar{l}^{[\pi]}, \bar{x}^{[\pi]}) \in \operatorname{conv} W^{[\pi]}$ being the final values of π in step (d), it satisfies for $v \in V^{[\pi]}$

$$\hat{y}^{\{\overline{\pi}+1\}}_{J^{[\pi]}} = \hat{y}^{[\pi]}, \qquad f^{[\pi]}_v(\hat{y}^{[\pi]}) = f^{\{\overline{\pi}+1\}}_v = f_v(\hat{y}^{\{\overline{\pi}+1\}}), \tag{4.45}$$

$$(\bar{l}^{\{\overline{\pi}+1\}}_v, \bar{x}^{\{\overline{\pi}+1\}}_v) = (\bar{l}^{[\pi]}_v, \bar{x}^{[\pi]}_v), \qquad \hat{f}^{[\pi]}_{(\bar{l}^{[\pi]}_v, \bar{x}^{[\pi]}_v), v}(\hat{y}^{[\pi]}) = \hat{f}_{(\bar{l}^{\{\overline{\pi}+1\}}_v, \bar{x}^{\{\overline{\pi}+1\}}_v), v}(\hat{y}^{\{\overline{\pi}+1\}}), \tag{4.46}$$

$$g^{[\pi]}(\bar{x}^{\{\sigma\}}_{V^{[\pi]}}) = g(\bar{x}^{\{\sigma\}})_{J^{[\pi]}} \qquad \text{for } \sigma \in \{\underline{\pi}, \dots, \overline{\pi}+1\}. \tag{4.47}$$

Proof. First consider(4.39) and (4.40) for $\sigma = \underline{\pi}$. At $\underline{\pi}$ the process π executed the step (a) and by (4.26) we have $\hat{y}^{\{\underline{\pi}+1\}} = \hat{y}^{\{\underline{\pi}\}}$, $\bar{x}_v^{\{\underline{\pi}+1\}} = \bar{x}_v^{\{\underline{\pi}\}}$ and $J_v^{\{\underline{\pi}+1\}} = J_v^{\{\underline{\pi}\}}$ for all $v \in V$, so with $\pi \in \Pi^{\{\underline{\pi}+1\}}$ items (viii),(ix) of Lemma 30 also hold for π at $\sigma = \underline{\pi}$, thus Observation 26 together with the assumption $f_{V^{[\pi]}}^{\{\underline{\pi}\}} = f_{V^{[\pi]}}(\hat{y}^{\{\underline{\pi}\}})$ proves $f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}) = f_v(\hat{y}^{\{\underline{\pi}\}})$ and Observation 27 proves $\hat{f}^{[\pi]}_{(\bar{l}_v^{\{\underline{\pi}\}},\bar{x}_v^{\{\underline{\pi}\}}),v}(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}) = \hat{f}_{(\bar{l}_v^{\{\underline{\pi}\}},\bar{x}_v^{\{\underline{\pi}\}}),v}(\hat{y}^{\{\underline{\pi}\}})$ for $v \in V^{[\pi]}$.

For $\sigma \in \{\underline{\pi}+1,\dots,\overline{\pi}\}$ we have $\pi \in \Pi^{\{\sigma\}}$, hence, for these values of σ, Lemma 30 (iv) implies the first and third equation in (4.39), the first in (4.40) and (4.42). For the remaining equations we may again use Lemma 30 (iv),(viii) together with (4.42) to invoke Observation 26 and Observation 27 for $\hat{y}^{\{\sigma\}}$. This completes (4.40) and for (4.39) we obtain

$$f_v^{\{\underline{\pi}\}} = f_v(\hat{y}^{\{\underline{\pi}\}}) \overset{(4.22)}{=} f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}) \overset{(4.39)}{=} f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\sigma\}}) \overset{(4.22)}{=} f_v(\hat{y}^{\{\sigma\}}) \qquad \text{for } v \in V^{[\pi]}, \sigma \in \{\underline{\pi},\dots,\overline{\pi}\}.$$

The values for $\sigma = \overline{\pi}+1$ are set by π when executing step (d) at $\overline{\pi}$, so (4.27) establishes (4.41) and also completes the result for (4.42) because the definition (4.4) of $\bar{J}^{[\pi]}$ implies $\bar{J}^{[\pi]} \subseteq M \setminus J^{[\pi]}$.

If π was stopped by (StopP') then $\hat{y}^{[\pi]} = \hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}} \overset{(4.39)}{=} \hat{y}_{J^{[\pi]}}^{\{\overline{\pi}\}}$ and in this case or if π is bad, (4.38) sets $\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}} = \hat{y}_{J^{[\pi]}}^{\{\overline{\pi}\}}$ and $f_{V^{[\pi]}}^{\{\overline{\pi}+1\}} = f_{V^{[\pi]}}^{\{\overline{\pi}\}}$, so (4.39) implies (4.43) if π is bad and (4.45) if π is good but stopped by (StopP'). For (4.45) there remains to consider the case of π being good and stopped by (StopD'). Then $\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}} = \hat{y}^{[\pi]}$ and $f_{V^{[\pi]}}^{\{\overline{\pi}+1\}} = f_{V^{[\pi]}}^{[\pi]}(\hat{y}^{[\pi]})$ by (4.36) and (4.37). The claim of the proof of Lemma 30 (vii) shows that for each $v \in V^{[\pi]}$ the set $J_v^{\{\underline{\pi}\}}$ is $(\hat{y}^{\{\overline{\pi}+1\}}, v)$-consistent with witness $\hat{x}_v^{[\pi]}$. Thus, for $v \in V^{[\pi]}$, (4.42) allows to use Observation 26 for proving $f_v^{[\pi]}(\hat{y}^{[\pi]}) = f_v(\hat{y}^{\{\overline{\pi}+1\}})$.

If π is identified as bad, (4.44) holds because the old primal aggregate is kept explicitly in step (d), while for a good process the update (4.35) gives the first equation of (4.46). The second equation follows from Observation 27 because by (4.42) and the successful tests (4.32) and (4.33) the point $\bar{x}^{[\pi]}$ satisfies (C1) for $J_v^{\{\underline{\pi}\}}$ and is therefore $(\hat{y}^{\{\overline{\pi}+1\}}, J_v^{\{\underline{\pi}\}})$-consistent.

In order to see (4.47) for a good process π, we prove that (4.24) holds for $\bar{x}^{\{\sigma\}}$, $\sigma \in \{\underline{\pi},\dots,\overline{\pi}+1\}$, then the result follows from Observation 28. Consider a $v \in V_j \setminus V_j^{[\pi]}$ for some $j \in J^{[\pi]}$. The successful test (4.34) certifies $j \notin J_v^{\{\overline{\pi}+1\}}$ and $J_v^{\{\overline{\pi}+1\}} \supseteq J_v^{\{\sigma\}}$ by Lemma 30 (v). By Lemma 30 (vii), (E3) holds at σ and implies $A_{j,v}\bar{x}_v^{\{\sigma\}} = 0$. □

As for the basic version, the global data is consistent throughout and the arc set of the dependency graph may only increase.

Lemma 32 *For all $\sigma \in \Sigma$,*

$$\begin{aligned} f_V^{\{\sigma\}} &= f_V(\hat{y}^{\{\sigma\}}), & (\bar{l}^{\{\sigma\}}, \bar{x}^{\{\sigma\}}) &\in \operatorname{conv} W, \\ \Delta^{\{\sigma\}} &= \Delta(\hat{y}^{\{\sigma\}}, (\bar{l}^{\{\sigma\}}, \bar{x}^{\{\sigma\}})), & E^{\{\sigma\}} &\subseteq E^{\{\sigma+1\}} \subseteq \{(i,j) : i,j \in M, i \neq j\}. \end{aligned} \tag{4.48}$$

Furthermore, $\Delta^{\{\sigma\}} > \varepsilon(|f(\hat{y}^{\{\sigma\}})|+1)$ for all $\sigma \in \Sigma$ with $\sigma+1 \in \Sigma$.

Proof. The proof is by induction on σ. For $\sigma = 0$ the claim holds by the initialization step. Suppose now $\sigma+1 \in \Sigma$ and the claim holds for $\sigma \in \Sigma$. If $\sigma+1$ is reached by a step (a) then none of the involved variables are changed and the relations still hold. Otherwise $\sigma+1$ is reached by a step (d) executed by some process π with $\overline{\pi} = \sigma$. If π is bad, none of the data relevant for the claim is changed, so assume π is good, in which case $J_v^{\{\overline{\pi}\}} = J_v^{\{\overline{\pi}+1\}}$

for $v \in V \setminus V^{[\pi]}$. For $v \in V \setminus V^{[\pi]}$ the successful test (4.34) asserts $J_v^{\{\overline{\pi}+1\}} \subseteq M \setminus J^{[\pi]}$ and by (4.27) $\hat{y}^{\{\overline{\pi}+1\}}_{J_v^{\{\overline{\pi}\}}} = \hat{y}^{\{\overline{\pi}\}}_{J_v^{\{\overline{\pi}\}}}$, $f_v^{\{\overline{\pi}+1\}} = f_v^{\{\overline{\pi}\}}$ and $(\bar{l}_v^{\{\overline{\pi}+1\}}, \bar{x}_v^{\{\overline{\pi}+1\}}) = (\bar{l}_v^{\{\overline{\pi}\}}, \bar{x}_v^{\{\overline{\pi}\}})$. Thus $(\bar{l}_v^{\{\overline{\pi}\}}, \bar{x}_v^{\{\overline{\pi}\}}) \in \operatorname{conv} W_v$ holds by induction, while Lemma 30 (vii) implies the validity of (E1) and (E2) for $\overline{\pi}$ and $\overline{\pi}+1$. Furthermore the witness $\hat{x}_v$ of (E2) for $\overline{\pi}$ is also a witness for $\overline{\pi}+1$ because $J_v^{\{\overline{\pi}\}} = J_v^{\{\overline{\pi}+1\}}$. Therefore it holds $f_v^{\{\overline{\pi}\}} = f_v(\hat{y}^{\{\overline{\pi}\}}) = f_v(\hat{y}^{\{\overline{\pi}+1\}})$ via Observation 21 and (4.1). For $v \in V^{[\pi]}$ the claim follows for (4.48) directly from (4.45) and (4.46). The correctness of (4.48) for $\overline{\pi}+1$ implies the correctness of $\Delta^{\{\overline{\pi}+1\}}$ by Observation 24. The claim for $E^{\{\overline{\pi}+1\}}$ follows directly from step (d). Finally, for a bad process π, $\Delta^{\{\overline{\pi}+1\}} = \Delta^{\{\overline{\pi}\}}$ because none of the data involved is changed, and if π is good $\Delta^{\{\overline{\pi}+1\}} \le \varepsilon(|f(\hat{y}^{\{\sigma\}})| + 1)$ leads to a termination for the algorithm in step (d) at $\overline{\pi}$ and then $\overline{\pi}+1 = \max \Sigma$. □

In the extended version, the algorithm may well happen to start a processes π on a subspace $J^{[\pi]}$ with $V^{[\pi]} = \emptyset$ and this will not cause any problems, because eventually the constraint indices will be added to $J_v^{\{\overline{\pi}+1\}}$ for the corresponding subproblems. We have to make sure, however, that a gap between function value and linear minorant, $f_v^{\{\sigma\}} > \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}})$, for some $v \in V$ enables some process π to choose an index in $j \in J_v^{\{\sigma\}}$ so that v ends up in $V^{[\pi]}$.

Observation 33 *For given $\sigma \in \Sigma$ and $v \in V$, there holds $f_v^{\{\sigma\}} \ge \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}})$ and if $f_v^{\{\sigma\}} > \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}})$ then $J_v^{\{\sigma\}} \neq \emptyset$. A process π with $J^{[\pi]} = M$ satisfies $\Delta_\pi^{\{\underline{\pi}\}} = \Delta^{\{\underline{\pi}\}}$.*

Proof. For $v \in V$, (4.48) asserts $(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}) \in \operatorname{conv} W_v$ and

$$f_v^{\{\sigma\}} \overset{(4.48)}{=} f_v(\hat{y}^{\{\sigma\}}) \overset{(2.4)}{\ge} \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}}).$$

By Lemma 30 (v) the sets $J_v^{\{\sigma\}}$ can only grow for increasing σ. We may have $J_v^{\{0\}} = \emptyset$ after initialization and we show that $f_v(\hat{y}^{\{\sigma\}}) = \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}})$ as long as $J_v^{\{\sigma\}} = \emptyset$. Indeed, in this case $v \notin V^{[\pi]}$ for any process π with $\underline{\pi} \le \sigma$ by (4.3). Then step (d) implies $(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}) = (\bar{l}_v^{\{0\}}, \bar{x}_v^{\{0\}})$ and $f_v^{\{\sigma\}} = f_v^{\{0\}}$. Furthermore, by Lemma 30 (vii) condition (E3) ensures $A_{J_v,v}\bar{x}^{\{\sigma\}} = 0$. Hence, together with the initialization step,

$$\hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}}) \overset{(2.4)}{=} \bar{l}_v^{\{\sigma\}} = \bar{l}_v^{\{0\}} \overset{(2.4)}{=} \hat{f}_{(\bar{l}_v^{\{0\}}, \bar{x}_v^{\{0\}}),v}(\hat{y}^{\{0\}}) = f_v(\hat{y}^{\{0\}}) = f_v^{\{\sigma\}}.$$

For π with $J^{[\pi]} = M$ we obtain by (4.3) that $v \in V^{[\pi]}$ for all v with $f_v^{\{\sigma\}} > \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}})$, and the result follows from the definitions (4.9) and (4.10). □

With this, we can establish again that at least one working process is started as long as the stopping criterion is not met .

Observation 34 *For $\sigma \in \Sigma$ with $\Pi^{\{\sigma\}} = \emptyset$ there holds $\Pi^{\{\sigma+1\}} = \{\pi\}$ for some π with $\underline{\pi} = \sigma$ if and only if $\Delta^{\{\sigma\}} > \varepsilon(|f(\hat{y}^{\{\sigma\}})| + 1)$.*

Proof. Lemma 30 (i) asserts that $\Pi^{\{\sigma\}} = \emptyset$ is equivalent to $B^{\{\sigma\}} = \emptyset$ and $B_M^{\{\sigma\}} = \emptyset$. For $B^{\{\sigma\}} = \emptyset$ and $B_M^{\{\sigma\}} = \emptyset$ conditions [S1]–[S4] can always be satisfied. Indeed, choosing $J^{[\pi]} = M$ satisfies [S1] because M trivially observes all dependencies of $E^{\{\sigma\}}$, [S2] by Observation 33, and [S3] as well as [S4] because the two blocking sets are empty. In particular, the algorithm starts with setting $\sigma = 0$ and $B^{\{0\}} = \emptyset$, $B_M^{\{0\}} = \emptyset$ and reaches step (a) if and only if the stopping criterion is not satisfied in step 1. For $0 < \sigma \in \Sigma$, $\Pi^{\{\sigma\}} = \emptyset$ requires $\Pi^{\{\sigma-1\}} = \{\pi\}$ for some process π with $\overline{\pi} = \sigma - 1$ by Observation 29. Thus, π executes a step (d) at $\sigma-1$. If π is bad we have $\Delta^{\{\overline{\pi}+1\}} = \Delta^{\{\overline{\pi}\}} > \varepsilon(|f(\hat{y}^{\{\overline{\pi}+1\}})|+1)$ by induction, and if π is good, the algorithm continues if and only if the stopping criterion is not satisfied for the global data of σ. □

The following observation ensures that a subspace selection satisfying [S2] is even possible with a selected subspace containing only one constraint.

Observation 35 *For any π executing (a) at $\sigma \in \Sigma$ there always exists a $j \in M$ so that $J^{[\pi]} = \{j\}$ yields $\Delta_\pi^{\{\sigma\}} \geq \frac{1}{2}\Delta^{\{\sigma\}} \min\{\frac{1}{m}, \frac{1}{|V|}\}$.*

Proof. Considering, for $\sigma \in \Sigma$, Observation 34 and the definition

$$\Delta^{\{\sigma\}} \overset{(4.9)}{=} \left(\sum_{v \in V} \left[f_v^{\{\sigma\}} - \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}}) \right] \right) + \left(\frac{1}{u} \sum_{j \in M} g(\bar{x}^{\{\sigma\}})_j^2 \right),$$

at least one of both summands is greater than or equal to $\frac{1}{2}\Delta^{\{\sigma\}} > 0$. If it is true for the second summand, at least one of the terms $\frac{1}{u}g(\bar{x}^{\{\sigma\}})_j^2$ is greater than or equal to $\frac{1}{2}\Delta^{\{\sigma\}}\frac{1}{m}$ and then choosing $J^{[\pi]} = \{j\}$ is sufficient. If this is true for the first one, at least one of the terms $f_v^{\{\sigma\}} - \hat{f}_{(\bar{l}_v^{\{\sigma\}}, \bar{x}_v^{\{\sigma\}}),v}(\hat{y}^{\{\sigma\}})$ is greater than or equal to $\frac{1}{2}\Delta^{\{\sigma\}}\frac{1}{|V|}$. In this case we may choose a $j \in J_v^{\{\sigma\}} \neq \emptyset$ for such a $v \in V$ by Observation 33 and $J^{[\pi]} = \{j\}$ satisfies the requirement. □

Again, no process runs forever and, eventually, all required dependencies between subspaces will be identified.

Lemma 36 *For each $\sigma \in \Sigma$, each process $\pi \in \underline{\Pi}^{\{\sigma\}}$ either stops at $\overline{\pi} < \infty$ or it is terminated by another process π' executing step (d) at $\overline{\pi}' < \infty$ with the global data of $\overline{\pi}' + 1$ satisfying the termination criterion.*

Proof. Analogous to the proof of Lemma 11. □

Corollary 37 *If $|\Sigma| < \infty$ then for $\sigma = \max \Sigma$ there holds $\Delta^{\{\sigma\}} \leq \varepsilon(|f(\hat{y}^{\{\sigma\}})| + 1)$. If $|\Sigma| = \infty$ then $\Pi^{\{\infty\}} = \emptyset$, $\underline{\Pi}^{\{\infty\}} = \overline{\Pi}^{\{\infty\}}$, and $\Sigma = \dot{\bigcup}_{\pi \in \overline{\Pi}^{\{\infty\}}} \{\underline{\pi}, \overline{\pi}\}$.*

Proof. This follows from Observation 29, Observation 34, and Lemma 36. □

Observation 38 *The number of bad processes $\pi \in \overline{\Pi}^{\{\infty\}}$ is finite.*

Proof. First note that by Lemma 30 (v) the sets $J_v^{\{\sigma\}} \subseteq J_v$ can only grow for increasing σ until $J_v^{\{\sigma\}} = J_v$, so only a finite number of processes enlarges the sets $J_v^{\{\sigma\}}$. For a bad process $\pi \in \overline{\Pi}^{\{\infty\}}$ one of the tests (4.31)–(4.34) failed. A failure in (4.31)–(4.33) requires $J_v^{\{\overline{\pi}+1\}} \setminus J_v^{\{\underline{\pi}\}} \neq \emptyset$ for some $v \in V^{[\pi]}$ and a failed (4.34) implies $J_v^{\{\overline{\pi}+1\}} \setminus J_v^{\{\overline{\pi}\}} \neq \emptyset$ for some $v \in V_{J^{[\pi]}} \setminus V^{[\pi]}$, because $J_v^{\{\underline{\pi}\}} \cap J^{[\pi]} = \emptyset$ by (4.6). In consequence, there must have been a process π' (maybe π itself) with $\underline{\pi} < \overline{\pi}' \leq \overline{\pi}$ that modified $J_v^{\{\overline{\pi}'\}} \neq J_v^{\{\overline{\pi}'+1\}}$. Because $|\Pi^{\{\sigma\}}| \leq N_\Pi$ for all $\sigma \in \Sigma$, there can only be finitely many bad processes. □

It will be convenient to collect all good processes in the set

$$\Gamma = \{\pi \in \overline{\Pi}^{\{\infty\}} : \pi \text{ is good}\}.$$

4.7. Convergence Analysis

For the convergence analysis the same steps work out as for the first variant. In particular, the global progress matches that of a single process.

Lemma 39 *For* $\pi \in \Gamma$

$$0 \le f^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}) - f^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}) = f(\hat{y}^{\{\overline{\pi}\}}) - f(\hat{y}^{\{\overline{\pi}+1\}}),$$

i. e., *the global progress achieved when* π *stores its subspace solution in the global data is exactly the progress made by* π *on* $J^{[\pi]}$*. In particular, the sequence* $(f(\hat{y}^{\{\sigma\}}))_\sigma$*,* $\sigma \in \Sigma$*, is non-increasing.*

Proof. First observe that for π the initial value of the center $\hat{y}^{[\pi]}$ is $\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}$ by step (b) and the final center is $\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}$ by (4.45), so the left hand inequality follows from the properties of the bundle method employed in step (c) of π. By Lemma 32 the requirement for Lemma 31 is fulfilled, so we may use its results for proving the second equation,

$$\begin{aligned}
0 \le f^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}) - f^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}) &\overset{(4.16)}{=} b_{J^{[\pi]}}^T(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}} - \hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}) + \sum_{v\in V^{[\pi]}} \left[f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}) - f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}) \right] \\
&\overset{(4.39)}{=} b_{J^{[\pi]}}^T(\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}\}} - \hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}) + \sum_{v\in V^{[\pi]}} \left[f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}\}}) - f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}) \right] \\
&\overset{(4.39),(4.41),(4.45)}{=} b^T(\hat{y}^{\{\overline{\pi}\}} - \hat{y}^{\{\overline{\pi}+1\}}) + \sum_{v\in V} \left[f_v(\hat{y}^{\{\overline{\pi}\}}) - f_v(\hat{y}^{\{\overline{\pi}+1\}}) \right] \\
&\overset{(2.2)}{=} f(\hat{y}^{\{\overline{\pi}\}}) - f(\hat{y}^{\{\overline{\pi}+1\}}).
\end{aligned}$$

For any $\sigma \in \Sigma$ without a $\pi \in \Gamma$ satisfying $\overline{\pi} = \sigma$ there is either a process π with $\underline{\pi} = \sigma$ which executes step (a) at $\underline{\pi}$ or a bad process π which executes step (d) at $\overline{\pi} = \sigma$ and in both cases $f_V^{\{\sigma\}} = f_V^{\{\sigma+1\}}$ and $\hat{y}^{\{\sigma\}} = \hat{y}^{\{\sigma+1\}}$. Therefore (2.2) and Lemma 32 guarantee $f(\hat{y}^{\{\sigma\}}) = f(\hat{y}^{\{\sigma+1\}})$ for these σ, which establishes that $(f(\hat{y}^{\{\sigma\}}))_\sigma$ is non-increasing. □

For the next step we first have to establish the required relations between the predicted decrease of the global data and that of a process.

Observation 40 *For* $\pi \in \Gamma$

$$\Delta_\pi^{\{\overline{\pi}\}} = \Delta_\pi^{\{\underline{\pi}\}} = \Delta^{[\pi]}(\hat{y}^{\{\underline{\pi}\}}, (\bar{l}_{V^{[\pi]}}^{\{\underline{\pi}\}}, \bar{x}_{V^{[\pi]}}^{\{\underline{\pi}\}})), \tag{4.49}$$

$$\bar{\Delta}_\pi^{\{\overline{\pi}+1\}} = \bar{\Delta}_\pi^{\{\overline{\pi}\}}, \tag{4.50}$$

$$\Delta_\pi^{\{\overline{\pi}+1\}} = \Delta^{[\pi]}(\hat{y}^{\{\overline{\pi}+1\}}, (\bar{l}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}, \bar{x}_{V^{[\pi]}}^{\{\overline{\pi}+1\}})). \tag{4.51}$$

Proof. By Lemma 32 we may use all results of Lemma 31 in the following.

$$\begin{aligned}
\Delta_\pi^{\{\overline{\pi}\}} &\overset{(4.10)}{=} \sum_{v\in V^{[\pi]}} \left[f_v^{\{\overline{\pi}\}} - \hat{f}_{(\bar{l}_v^{\{\overline{\pi}\}}, \bar{x}_v^{\{\overline{\pi}\}}),v}(\hat{y}^{\{\overline{\pi}\}}) \right] + \tfrac{1}{u}\|g(\bar{x}^{\{\overline{\pi}\}})_{J^{[\pi]}}\|^2 \\
&\overset{(4.39),(4.40),(4.47)}{=} \sum_{v\in V^{[\pi]}} \left[f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}\}}) - \hat{f}_{(\bar{l}_v^{\{\overline{\pi}\}}, \bar{x}_v^{\{\overline{\pi}\}}),v}^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\overline{\pi}\}}) \right] + \tfrac{1}{u}\|g^{[\pi]}(\bar{x}_{V^{[\pi]}}^{\{\overline{\pi}\}})\|^2 \\
&\overset{(4.39),(4.40),(4.47)}{=} \sum_{v\in V^{[\pi]}} \left[f_v^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}) - \hat{f}_{(\bar{l}_v^{\{\underline{\pi}\}}, \bar{x}_v^{\{\underline{\pi}\}}),v}^{[\pi]}(\hat{y}_{J^{[\pi]}}^{\{\underline{\pi}\}}) \right] + \tfrac{1}{u}\|g^{[\pi]}(\bar{x}_{V^{[\pi]}}^{\{\underline{\pi}\}})\|^2 \\
&\overset{(4.39),(4.40),(4.47)}{=} \sum_{v\in V^{[\pi]}} \left[f_v^{\{\underline{\pi}\}} - \hat{f}_{(\bar{l}_v^{\{\underline{\pi}\}}, \bar{x}_v^{\{\underline{\pi}\}}),v}(\hat{y}^{\{\underline{\pi}\}}) \right] + \tfrac{1}{u}\|g(\bar{x}^{\{\underline{\pi}\}})_{J^{[\pi]}}\|^2 \overset{(4.10)}{=} \Delta_\pi^{\{\underline{\pi}\}}
\end{aligned}$$

To complete (4.49) note that by (4.19) the third line is in fact $\Delta^{[\pi]}(\hat{y}^{\{\pi\}}, (\bar{l}^{\{\pi\}}_{V^{[\pi]}}, \bar{x}^{\{\pi\}}_{V^{[\pi]}}))$. In order to prove (4.50) we first show (with $\hat{J}^{[\pi]}$ as defined in (4.8))

$$A_{j,v}\bar{x}_v^{\{\overline{\pi}\}} = A_{j,v}\bar{x}_v^{\{\overline{\pi}+1\}} \quad \text{for } j \in M \setminus (J^{[\pi]} \cup \hat{J}^{[\pi]}), v \in V. \tag{4.52}$$

For $v \in V \setminus V^{[\pi]}$ this is an immediate consequence of (4.41). For $v \in V^{[\pi]}$ consider some fixed $j \in M \setminus (J^{[\pi]} \cup \hat{J}^{[\pi]})$ and recall that, by (4.8) and Lemma 30 (v), $J_v^{\{\overline{\pi}\}} \subseteq J_v^{\{\overline{\pi}+1\}} \subseteq J^{[\pi]} \cup \hat{J}^{[\pi]}$ for $v \in V^{[\pi]}$. Using Lemma 30 (vii) we get $A_{j,v}\bar{x}_v^{\{\overline{\pi}\}} \stackrel{(E3)}{=} 0 \stackrel{(E3)}{=} A_{j,v}\bar{x}_v^{\{\overline{\pi}+1\}}$. With (4.52) and by recalling the definitions (4.11), (2.3) we obtain as a first ingredient for (4.50)

$$\tfrac{1}{u}\|g(\bar{x}^{\{\overline{\pi}\}})_{M\setminus(J^{[\pi]}\cup\hat{J}^{[\pi]})}\|^2 = \tfrac{1}{u}\|g(\bar{x}^{\{\overline{\pi}+1\}})_{M\setminus(J^{[\pi]}\cup\hat{J}^{[\pi]})}\|^2.$$

Next, by definition (2.4), establishing

$$\hat{f}_{(\bar{l}_v^{\{\overline{\pi}\}},\bar{x}_v^{\{\overline{\pi}\}}),v}(\hat{y}^{\{\overline{\pi}\}}) = \hat{f}_{(\bar{l}_v^{\{\overline{\pi}+1\}},\bar{x}_v^{\{\overline{\pi}+1\}}),v}(\hat{y}^{\{\overline{\pi}+1\}}) \quad \text{for } v \in V \setminus V^{[\pi]}$$

requires to show $\bar{l}_v^{\{\overline{\pi}\}} = \bar{l}_v^{\{\overline{\pi}+1\}}$ and $(\hat{y}_{J_v}^{\{\overline{\pi}\}})^T A_{J_v,v}\bar{x}_v^{\{\overline{\pi}\}} = (\hat{y}_{J_v}^{\{\overline{\pi}+1\}})^T A_{J_v,v}\bar{x}_v^{\{\overline{\pi}+1\}}$ for $v \in V \setminus V^{[\pi]}$. By (4.41) the only possible difference might be caused by $\hat{y}_j^{\{\overline{\pi}\}} \neq \hat{y}_j^{\{\overline{\pi}+1\}}$ for some $j \in J^{[\pi]}$. For a fixed $j \in J^{[\pi]}$ and $v \in V_j \setminus V^{[\pi]}$, however, the successful test (4.34) at $\overline{\pi}$, asserts $j \notin J_v^{\{\overline{\pi}+1\}} \supseteq J_v^{\{\overline{\pi}\}}$. Therefore (E3) ensures $A_{j,v}\bar{x}_v^{\{\overline{\pi}\}} = 0 = A_{j,v}\bar{x}_v^{\{\overline{\pi}+1\}}$. The final ingredient $f_v^{\{\overline{\pi}\}} = f_v^{\{\overline{\pi}+1\}}$ for $v \in V \setminus V^{[\pi]}$ is contained explicitly in (4.41) and this completes (4.50).

The last equation (4.51) is a direct consequence of the definitions (4.10) and (4.19) together with (4.45)–(4.47). □

From now on, the analysis follows exactly the same path as that of the first variant, and up to small adaptations due to the presence of bad processes the proofs match the first ones verbatim.

Lemma 41 *Suppose an infinite number of descent steps occurs and f is bounded from below. Then*

$$\liminf_{\sigma\in\mathbb{N}_0} \Delta^{\{\sigma\}} \to 0.$$

Proof. Let $\pi \in \Gamma$ be a good process for which a descent step occurs, *i. e.*, π is stopped because of condition (StopD'). By [S2] we have $\Delta^{\{\pi\}} \leq \frac{1}{\tau_1}\Delta_\pi^{\{\pi\}}$. By Lemma 32 and Lemma 31 (4.45),(4.46) the final predicted decrease of π that caused the descent step is $\Delta^{[\pi]}(\hat{y}^{\{\pi\}}_{J^{[\pi]}}, (\bar{l}^{\{\overline{\pi}+1\}}_{V^{[\pi]}}, \bar{x}^{\{\overline{\pi}+1\}}_{V^{[\pi]}}))$. Because (StopD') and not the preceding test (StopP') has caused π to stop, we have

$$\tau_2\Delta_\pi^{\{\pi\}} \stackrel{(StopP')}{\leq} \Delta^{[\pi]}(\hat{y}^{\{\pi\}}_{J^{[\pi]}}, (\bar{l}^{\{\overline{\pi}+1\}}_{V^{[\pi]}}, \bar{x}^{\{\overline{\pi}+1\}}_{V^{[\pi]}})) \stackrel{(StopD')}{\leq} \frac{1}{\rho}\left(f^{[\pi]}(\hat{y}^{\{\pi\}}_{J^{[\pi]}}) - f^{[\pi]}(\hat{y}^{\{\overline{\pi}+1\}}_{J^{[\pi]}})\right).$$

Putting all together and using Lemma 39 we get

$$\Delta^{\{\pi\}} \leq \frac{1}{\tau_1}\Delta_\pi^{\{\pi\}} \leq \frac{1}{\tau_1\tau_2\rho}\left(f^{[\pi]}(\hat{y}^{\{\pi\}}_{J^{[\pi]}}) - f^{[\pi]}(\hat{y}^{\{\overline{\pi}+1\}}_{J^{[\pi]}})\right) = \frac{1}{\tau_1\tau_2\rho}\left(f(\hat{y}^{\{\overline{\pi}\}}) - f(\hat{y}^{\{\overline{\pi}+1\}})\right).$$

Because f is bounded from below and the sequence $(f(\hat{y}^{\{\sigma\}}))_\sigma$ is non-increasing by Lemma 39, the right hand side of the inequality above converges to zero. □

Lemma 42 *Assume there is only a finite number of descent steps and $\varepsilon = 0$, then*

$$\lim_{\sigma \in \Sigma} \Delta^{\{\sigma\}} = 0.$$

Proof. If $|\Sigma| < \infty$, then the statement holds by Corollary 37. Therefore we may assume $|\Sigma| = \infty$.

Lemma 32 implies $\Delta^{\{\sigma\}} > 0$ for all $\sigma \in \Sigma$ and that the dependency graph $D^{\{\sigma\}}$ can only be increased. By Observation 38 and because M is a finite set there must be a $\sigma' \in \Sigma$ such that for each $\sigma \geq \sigma'$ we have $E^{\{\sigma\}} = E^{\{\sigma'\}}$ and all processes $\pi = (\underline{\pi}, \overline{\pi})$ with $\overline{\pi} > \underline{\sigma} := \min\{\sigma', \underline{\pi}' : \pi' \in \Pi^{\{\sigma'\}}\}$ are good and do *not* perform a descent step.

Let $\sigma > \sigma'$, then by Corollary 37 there is a process $\pi = (\underline{\pi}, \overline{\pi})$ such that $\sigma \in \{\underline{\pi}, \overline{\pi}\}$. If $\sigma = \underline{\pi}$ we know by (4.26) and Lemma 32 that $\Delta^{\{\underline{\pi}\}} = \Delta^{\{\underline{\pi}+1\}}$. So assume $\sigma = \overline{\pi}$. Because $\sigma > \sigma'$ we know that π satisfied condition (StopP') and that

$$\delta_\pi^{\{\overline{\pi}+1\}} - \delta_\pi^{\{\overline{\pi}\}} \leq \tau_3 \Delta_\pi^{\{\underline{\pi}\}}.$$

Invoking Observation 24 twice for the subspace $J^{[\pi]}$ of π but once for the data of $\overline{\pi}$ and once for $\overline{\pi}+1$ yields the relations

$$\begin{aligned} \Delta^{\{\overline{\pi}\}} &= \Delta_\pi^{\{\overline{\pi}\}} + \delta_\pi^{\{\overline{\pi}\}} + \bar{\Delta}_\pi^{\{\overline{\pi}\}}, \\ \Delta^{\{\overline{\pi}+1\}} &= \Delta_\pi^{\{\overline{\pi}+1\}} + \delta_\pi^{\{\overline{\pi}+1\}} + \bar{\Delta}_\pi^{\{\overline{\pi}+1\}}. \end{aligned}$$

We claim that $\Delta^{\{\overline{\pi}+1\}} \leq (1-\tau)\Delta^{\{\overline{\pi}\}}$ for some constant $0 < \tau < 1$ independent of π. Indeed, (4.49) implies $\Delta_\pi^{\{\underline{\pi}\}} = \Delta_\pi^{\{\overline{\pi}\}}$ and (4.50) gives $\bar{\Delta}_\pi^{\{\overline{\pi}\}} = \bar{\Delta}_\pi^{\{\overline{\pi}+1\}}$. The subspace selection condition [S2] asserts $\Delta_\pi^{\{\underline{\pi}\}} \geq \tau_1 \Delta^{\{\underline{\pi}\}}$ and stopping condition (StopP') (because π is good and does not perform a descent step) implies $\Delta_\pi^{\{\overline{\pi}+1\}} \overset{(4.51)}{=} \Delta^{[\pi]}(\hat{y}^{\{\overline{\pi}+1\}}, (\bar{l}_{J^{[\pi]}}^{\{\overline{\pi}+1\}}, \bar{x}_{V^{[\pi]}}^{\{\overline{\pi}+1\}})) < \tau_2 \Delta_\pi^{\{\underline{\pi}\}}$. This yields

$$\begin{aligned} \Delta^{\{\overline{\pi}\}} - \Delta^{\{\overline{\pi}+1\}} &= (\Delta_\pi^{\{\overline{\pi}\}} - \Delta_\pi^{\{\overline{\pi}+1\}}) + (\bar{\Delta}_\pi^{\{\overline{\pi}\}} - \bar{\Delta}_\pi^{\{\overline{\pi}+1\}}) + (\delta_\pi^{\{\overline{\pi}\}} - \delta_\pi^{\{\overline{\pi}+1\}}) \qquad (4.53) \\ &= (\Delta_\pi^{\{\underline{\pi}\}} - \Delta_\pi^{\{\overline{\pi}+1\}}) + (\delta_\pi^{\{\overline{\pi}\}} - \delta_\pi^{\{\overline{\pi}+1\}}) \\ &\geq (1 - \tau_2 - \tau_3)\Delta_\pi^{\{\underline{\pi}\}} \\ &\geq \underbrace{\tau_1(1 - \tau_2 - \tau_3)}_{=:\tau \in (0,1)} \Delta^{\{\underline{\pi}\}}. \end{aligned}$$

Note that this shows $\Delta^{\{\overline{\pi}\}} - \Delta^{\{\overline{\pi}+1\}} \geq 0$ for all $\overline{\pi} = \sigma \geq \underline{\sigma}$. Together with $\Delta^{\{\underline{\pi}\}} = \Delta^{\{\underline{\pi}+1\}}$ (see above) we get therefore that the sequence $(\Delta^{\{\sigma\}})_{\sigma \geq \underline{\sigma}}$ is non-increasing. Because $\overline{\pi} > \sigma'$ we have $\underline{\pi} \geq \underline{\sigma}$ and thus $\Delta^{\{\underline{\pi}\}} \geq \Delta^{\{\overline{\pi}\}}$. From (4.53) we obtain

$$\Delta^{\{\overline{\pi}\}} - \Delta^{\{\overline{\pi}+1\}} \geq \tau \Delta^{\{\underline{\pi}\}} \geq \tau \Delta^{\{\overline{\pi}\}}$$

and so

$$\Delta^{\{\overline{\pi}+1\}} \leq (1-\tau)\Delta^{\{\overline{\pi}\}}.$$

Together with the case $\sigma = \underline{\pi}$ above we get $\lim_{\sigma \in \mathbb{N}_0} \Delta^{\{\sigma\}} = 0$, which completes the proof.□

Corollary 43 *If f is bounded from below and $\varepsilon = 0$, the predicted decrease $\Delta^{\{\sigma\}} = f(\hat{y}^{\{\sigma\}}) - \hat{f}_{(\bar{l}^{\{\sigma\}}, \bar{x}^{\{\sigma\}})}(\hat{y}^{\{\sigma\}}) + \frac{1}{u}\|g(\bar{x}^{\{\sigma\}})\|^2$ goes to zero for an appropriate subsequence $\Sigma^* \subseteq \Sigma$. In particular, $f(\hat{y}^{\{\sigma\}}) - \hat{f}_{(\bar{l}^{\{\sigma\}}, \bar{x}^{\{\sigma\}})}(\hat{y}^{\{\sigma\}})$ and $\|g(\bar{x}^{\{\sigma\}})\|$ go to zero, too, for the subsequence Σ^*.*

Proof. Analogous to the proof of Corollary 16. □

Theorem 44 *Suppose* $\emptyset \neq \operatorname{Argmin} f$ *is bounded. Then for an appropriate subsequence* $\Sigma^* \subseteq \Sigma$ *the sequences* $(\hat{y}^{\{\sigma\}})_{\sigma\in\Sigma^*}$ *and* $(\bar{x}^{\{\sigma\}})_{\sigma\in\Sigma^*}$ *that are generated by the extended parallel bundle algorithm have the following properties.*

(i) each accumulation point of $(\hat{y}^{\{\sigma\}})_{\sigma\in\Sigma^*}$ *is an optimal solution of* (D),

(ii) each accumulation point of $(\bar{x}^{\{\sigma\}})_{\sigma\in\Sigma^*}$ *is an optimal solution of* $(\operatorname{conv} P)$.

Proof. Analogous to the proof of Theorem 17. □

5. Numerical Experiments

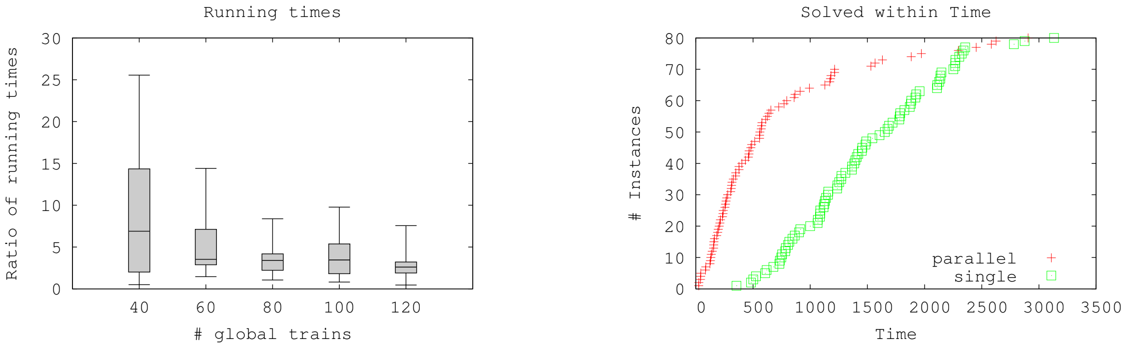
In this section we present first numerical results comparing the (Extended) Parallel Bundle Method proposed in this paper with a standard proximal bundle algorithm. Note that it is not the intent of this paper to provide an extensive numerical study. Indeed, the experiments are very preliminary. The purpose of the tests is to show that there are problems where the application of the parallel bundle method may give advantages over the classical bundle method and may therefore be a valid alternative in future applications.

We used simple randomly generated test instances for the *Train Timetabling Problem* (TTP), see, *e. g.*, [4] for a full description of the problem. The task of the TTP is to find a conflict free timetable for a set of trains in a given railway network. By conflict free we mean that certain capacity restrictions in the stations as well as headway distances between successive trains on the same track have to be fulfilled. In a real world TTP problem there are often certain local areas with many local short distance trains (*e. g.*, around big cities) and a number of long distance and freight trains running through the whole network and coupling those local areas. In order to mimic this structure, we generated in our tests a random network as a subset of the two dimensional square grid graph which is further divided into several (almost disjoint) subsquares. Those subsquares represent the local areas and are filled each afterwards with a certain number of local trains with random routes within that subsquare and random start time within a certain time window. Analogously a certain number of global trains is generated which have random routes within the whole network coupling the local areas. Each train has a running time of either 1 or 2 per arc chosen randomly, all nodes in the network have a random capacity of either 1 or 2 and all headway times are assumed to be one minute. Appendix A describes the generation process of the test instances in more detail.

A typical model for the TTP, see [4], is based on time-expanded networks for each train. In our case the network for each train is expanded up to twice its minimal running time from its start to its end node. It is easy to see, that this model is indeed an instance of (P) where each subproblem $(P_v(y))$ is a shortest-path-problem in one of the time expanded train networks. Therefore the classical bundle method as well as the parallel bundle method can be applied to solve the corresponding Lagrangian dual problem.

We ran 80 test instances with 4×4 subsquares, each of size 20×20, forming a large square of size 80×80 and 60 local trains per subsquare starting randomly distributed within three hours. The number of global trains that couple the local areas is increased from 40 to 120 trains in steps of 20 and they are started within three hours, too.

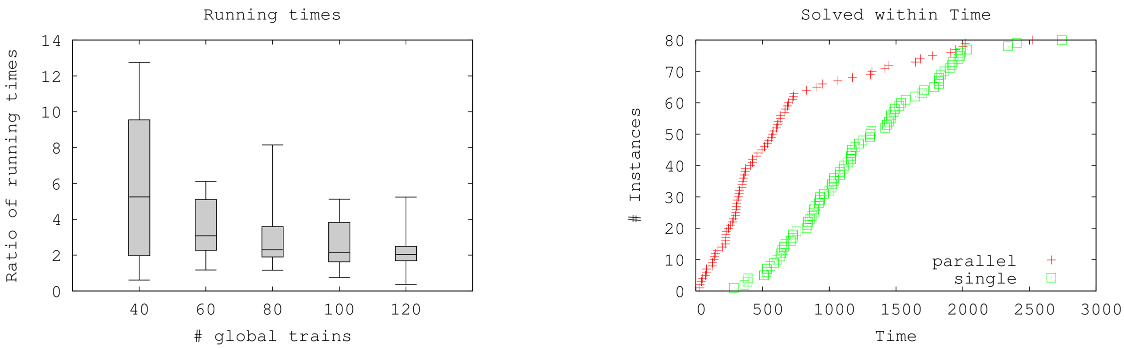
Both algorithms, the classical bundle method and the parallel bundle method, have been implemented in C++ and all tests were run on an Intel(R) Core(TM) i7 CPU with 8 processors and 12GB memory. We solved all instances up to a final termination precision of 10^{-4}. In a first test we solved all instances with the classical bundle method on one core and with the parallel bundle method on one core with exactly one subprocess. In the second test we ran the tests with the classical bundle method where the subproblem evaluations were executed in parallel on 4 cores and the parallel bundle method with up to 4 parallel processes. Figure 3 shows the results for the tests with one core. Denoting the running times by t_{parallel} for the parallel bundle method and t_{single} for the classical bundle method, diagram (a) shows minimum, the maximum, the 25%, 75% quantiles and the median of the relative running times $t_{\text{single}}/t_{\text{parallel}}$ over all instances with a certain number of global trains. Diagram (b) shows the number of instances solved up to a certain time for both algorithms. Figure 4 shows the analog results for the tests on four cores.



(a) minimum, maximum, 25%, 75% quantiles and median of $t_{\text{single}}/t_{\text{parallel}}$

(b) Performance profile

Figure 3: Running times for 80 test instances on 1 core.



(a) minimum, maximum, 25%, 75% quantiles and median of $t_{\text{single}}/t_{\text{parallel}}$

(b) Performance profile

Figure 4: Running times for 80 test instances on 4 cores.

The diagrams show that the parallel bundle method solves most instances within a shorter time than the classical bundle method on one as well as on four cores. Furthermore, as the number of global trains is increased the advantage of the parallel bundle method decreases. This matches expectations, because the classical bundle method has to evaluate all subproblems equally often whereas the parallel bundle method may take advantage of the decoupled structure of the problem with some easy and some more difficult areas. Therefore the parallel bundle method may evaluate certain subproblems less often while focusing on the more difficult ones. When the number of global trains is increased, coupling between the subproblems is increased as well, and hence most subproblems become similarly difficult. In this case the parallel bundle method looses its advantage.

It may come as a surprise, however, that the improvement of using four cores instead of one is not evident, not only for the parallel but also for the classical bundle method. A possible explanation might be that the parallel evaluation of a large number of combinatorial subproblems on multicore machines with common main memory seems to cause significant problems in accessing memory (cache) and the speedup in pure computing power is lost due to competing memory accesses. Therefore much better implementations of the algorithms are required in order to properly exploit the properties of modern hardware.

Still, the numerical results indicate that the proposed parallel bundle approach has the potential for developing into a useful alternative for large scale applications with appropriate structure.

Acknowledgment. This work was supported by the *Bundesministerium für Bildung und Forschung* under grant 03MS640D. Responsibility for the content rests with the authors.

A. Problem and model description of the TTP

A.1. Problem and model

One typical application for bundle methods is the solution of the Lagrangian relaxation of very large scale instances of combinatorial optimization problems. In out tests we focused on simple random instances of the *Train Timetabling Problem* (TTP), see, *e. g.*, [4]. The TTP can be roughly described as follows. Given an infrastructure network $G^I = (U^I, A^I)$ and a set of trains V with predefined routes $U^v = \{u_1^v, \dots, u_{k_v}^v\} \subset U^I$ with $(u_i^v, u_{i+1}^v) \in A^I$, $i = 1, \dots, k_v - 1$, a starting time $t_0^v \in \mathbb{R}_+$ and running time per arc $t_i^v \in \mathbb{R}_+$, the task is to determine a feasible timetable, *i. e.*, arrival and departure times for each train at each of its stations. This timetable has to fulfill several constraints like capacity constraints in the nodes (stations) and headway time constraints on the arcs. In detail, each node $u \in U^I$ has a maximal capacity $c_u \in \mathbb{N}$, which means at most c_u trains may be at node u at the same time, and for each arc $a \in A^I$ there is a minimal headway time $h_a > 0$ which means that two trains may enter arc a at times t_1 and t_2, resp., if and only if $|t_1 - t_2| \geq h_a$. Note that in practical applications the headway times may also depend on train-length, speed and other aspects.

A typical model for the TTP is as follows. First the time is discretized in a set $T = \{1, 2, \dots, t_{\max}\}$ (*e. g.*, in minutes). Then for each train $v \in V$ there is a time-expanded network $G_T^v = (U_T^v, A_T^v)$ with node set

$$U_T^v = U^v \times (\{t_0^v, t_0^v + 1, \dots, t_{\max}^v\} \cup \{u_{end}^v\})$$

and arc set

$$\begin{aligned} A_T^v = & \{((u_i^v, t), (u_i^v, t+1)) \colon u_i^v \in U^v, t_0^v \leq t < t_{\max}^v\} \\ & \cup \{((u_i^v, t), (u_{i+1}^v, t + t_i^v)) \colon u_i^v, u_{i+1}^v \in U^v, t_0^v \leq t \leq t_{\max}^v - t_i^v\} \\ & \cup \{((u_i^v, t), u_{end}^v) \colon (u_i^v, t) \in U_T^v\}. \end{aligned}$$

The arc set of such a network consists of three classes of arcs: wait arcs of the form $((u_i^v, t), (u_i^v, t+1))$, run-arcs of the form $((u_i^v, t), (u_{i+1}^v, t + t_i^v))$ and stop arcs $((u_i^v, t), u_{end}^v)$. The latter will always be assigned high costs for $u_i^v \neq u_{k_v}^v$ and zero for $u_i^v = u_{k_v}^v$, so the train-route will have small costs if and only if the train reaches its destination in time. Next we introduce binary variables $x^v \colon A_T^v \to \{0, 1\}$ for the arcs in G_T^v and get for each train $v \in V$ the optimization problem

$$(TTP_v) \qquad \begin{array}{ll} \text{maximize} & (-w^v)^T x^v \\ \text{subject to} & x^v \in \Omega_v, \end{array}$$

where

$$\Omega_v = \{x^v \in \{0,1\}^{A_T^v} \colon x^v \text{ represents a } (u_1^v, t_0^v) - u_{end}^v\text{-path in } G_T^v\}.$$

Now we can write the capacity constraints as

$$\sum_{v \in V} \sum_{e = ((u', t'), (u, t)) \in A_T^v} x_e^v \leq c_u, \qquad \text{for all } u \in U^I,\ t \in T \tag{A.1}$$

and the headway constraints as

$$\sum_{v \in V} \sum_{t = t_0}^{t + h_a - 1} \sum_{e = ((u, t), (u', t')) \in A_T^v} x_e^v \leq 1, \qquad \text{for all } (u, u') = a \in A^I \text{ and } t_0 \in T. \tag{A.2}$$

Putting all together the optimization problem reads

$$(TTP) \qquad \begin{array}{lll} \text{maximize} & \sum_{v\in V}(-w^v)^T x^v & \\ \text{subject to} & Ax \leq b & \\ & x^v \in \Omega_v, & v \in V, \end{array}$$

where the linear constraints (A.1) and (A.2) are collected in the linear inequalities $Ax \leq b$. It is easy to see that this model is indeed an instance of (P) except that the constraints are inequalities, but this has been implemented as well. The subproblems $(P_v(y))$ turn out to be

$$(TTP_v(y)) \qquad \begin{array}{lll} \text{maximize} & (-w^v)^T x^v - y^T A_{\bullet,v} x^v & \\ \text{subject to} & x^v \in \Omega_v, & v \in V, \end{array}$$

which are simple shortest path problems in an acyclic network with a linear cost function. Therefore those subproblems are easy to solve and the problem (TTP) fits in our framework.

A.2. Construction of test instances

Large scale real world instances of the TTP usually require much more fine grained models than those described in the previous section and often more advanced techniques like cutting plane approaches or dynamic graph generation [4] in order to be efficiently treatable. The current prototype implementation of the parallel bundle algorithm cannot handle these cases, yet, and therefore we restricted to randomly generated test instances. The construction of those instances is described in this section.

We assume the following properties of typical TTP instances. In a railway network there are some areas where many short distance local trains interact (*e. g.*, around big cities) and those areas overlap only slightly with other areas. Furthermore there are long-distance and freight trains that run through the whole network coupling those local areas. These assumptions are reasonable for large real world networks.

The node set is constructed as follows. Let $n_l, d_l \in \mathbb{N}$ two numbers. The node set consists of $n_l \times n_l$ subsquares $U^I_{i,j} = \{(u_1, u_2)\colon u_1 \in \{(i-1)d_l, \dots, id_l\}, u_2 \in \{(j-1)d_l, \dots, jd_l\}$ and the full node set $U^I = \bigcup_{i,j=1}^{n_l} U^I_{i,j}$. The arc set A^I is generated by inserting randomly a number of routes from a node of one border of the square node set to a node of the opposite site. The detailed approach can be seen in Algorithm 45.

Algorithm 45 *(Generation of infrastructure arcs.)*

1: ***Input:*** n
2: ***Output:*** $A^I \subseteq (U^I)^2$
3: ***for*** $k := 1$ *to* n ***do***
4: *Choose a direction* $d \in \{-e_1, e_1, -e_2, e_2\}$ *randomly.*
5: *Choose* $u_0 \in U^I$ *with* $d^T u_0$ *minimal* ▷ i. e., u_0 *is a border node*
6: $i \leftarrow 0$.
7: ***while*** $d^T u_i$ *not maximal* ***do*** ▷ i. e., u_i *has not reached the opposite border yet.*
8: *Choose* $u_{i+1} \in U$ *with*

$$\|u_{i+1} - u_i\| = 1,$$
$$(u_{i+1} - u_i)^T d \geq 0,$$
$$i = 0 \vee u_{i-1} \neq u_{i+1}$$

▷ i. e., *not in backward direction.*

$A^I \leftarrow A^I \cup \{(u_i, u_{i+1})\}$
$i \leftarrow i + 1.$
end while
end for

After the arc set is constructed a set of trains is constructed. For this a random border node is chosen and from there a random path with the arc set until another border node is reached. If a loop is created or the generated train route is too short the route is thrown away and the procedure is restarted. Algorithm 46 is called n_{local} times for each local area und n_{global} times for the whole network to generate local and global trains, resp..

Algorithm 46 *(Generation of trains.)*

Input: *Subsquare* $G' = (U', A')$ *with size* n.
Output: *Train route* $(u_1, \ldots, u_k)$
Choose random border node $u_0 \in U'$.
$i \leftarrow 0$
repeat
$i \leftarrow i + 1$
Choose u_i *as random neighbor of* u_{i-1}
if $u_i \in \{u_0, \ldots, u_{i-1}\}$ ***then***
go to 3.
end if
until u_i *border node*

In our test instances we chose $n_l = 20$, $d_l = 4$, $n_{\text{local}} = 60$, and $n_{\text{global}} \in \{40, 60, 80, 100, 120\}$ to generate 16 instances per choice of n_{global}.

B. Symbols

General symbols:

$M = \{1, \dots, m\}$	set of constraints
$V = \{1, \dots, \omega\}$	set of subproblems
$\Omega_v \subset \mathbb{R}^{n_v}, v \in V$	ground set of v
$\Omega = \bigotimes_{v \in V} \Omega_v \subset \mathbb{R}^n$	all ground sets
$h_v \colon \Omega_v \to \mathbb{R}$	objective function of subproblem v
$A \in \mathbb{R}^{M \times n}$	constraint matrix
$b \in \mathbb{R}^M$	right-hand-side of the constraints
$L_v(x_v, y), v \in V$	Lagrangian associated with subproblem $v \in V$
$f_v(y), v \in V$	optimal value of subproblem $v \in V$ with augmenting costs induced by y
$L(x, y)$	Lagrangian of full problem
$f(y)$	value of dual function at y
$g(x)$	subgradient generated $x \in \operatorname{conv} \Omega$
W_v	set of linear minorants of f_v generated by Ω_v
W	set of linear minorants of f generated by Ω
$\hat{f}_{w_v,v}(y),\ v \in V,\ w_v \in W_v$	value of linear minorant $w_v \in \operatorname{conv} W_v$ of f_v at y
$\hat{f}_w(y)$	value of linear minorant $w \in \operatorname{conv} W$ of f at y
$J \subset M$	some subspace of Lagrange multipliers
$J_v \subset M,\ v \in V$	constraint interacting with $v \in V$
$V_j \subset V,\ j \in J$	problems interacting with $j \in M$
$\bar{J} \subset M$	constraints interacting with V_J but not contained in J
$\Delta(y, w)$	predicted global decrease at y with aggregate minorant w
$\Delta_J(y, w)$	predicted decrease on subspace J
$\bar{\Delta}_J(y, w)$	predicted decrease not influences by subspace J
$\delta_{\bar{J}}(w)$	predicted decrease influenced by J and by $M \setminus J$
$\hat{y} \in \mathbb{R}^M$	center of stability
$\bar{y} \in \mathbb{R}^M$	candidate point
$(\bar{l}, \bar{x}) \in \operatorname{conv} W$	primal aggregate minorant
$\hat{x} \in \Omega$	optimal primal solution in center

Identifiers used in the standard parallel bundle algorithm in Section 3 denoting some global algorithmic information have a superscript $\langle\sigma\rangle$.

Σ	the set of all significant global steps
$\underline{\Pi}^{\langle\sigma\rangle}$	the set of all processes started before $\sigma \in \mathbb{N}$
$\overline{\Pi}^{\langle\sigma\rangle}$	the set of all processes finished before $\sigma \in \mathbb{N}$
$\Pi^{\langle\sigma\rangle}$	the set of all processes running at $\sigma \in \mathbb{N}$
$\hat{y}^{\langle\sigma\rangle} \in \mathbb{R}^M$	current global center
$f_v^{\langle\sigma\rangle}$	optimal value of f_v in current center $\hat{y}^{\langle\sigma\rangle}$ of subproblem $v \in V$
$(\bar{l}^{\langle\sigma\rangle}, \bar{x}^{\langle\sigma\rangle}) \in \operatorname{conv} W$	current global primal minorant
$B^{\langle\sigma\rangle} \subseteq V$	set of currently blocked subproblems
$D^{\langle\sigma\rangle}$	current dependency graph
$\Delta^{\langle\sigma\rangle}$	current predicted global decrease

Identifiers associated with the subspace problem optimized by a process π have a superscript (π).

π, η	processes
$\underline{\pi}$	global index at which π starts
$\overline{\pi}$	global index when π finishes
$V^{(\pi)} \subseteq V$	problems processed by π
$J^{(\pi)} \subseteq M$	subspace processed by π
$\bar{J}^{(\pi)} \subseteq M$	constraints not contained in $J^{(\pi)}$ but interacting with some $v \in V^{(\pi)}$
$\Omega^{(\pi)} = \bigotimes_{v \in V^{(\pi)}} \Omega_v$	ground set of process π
$\hat{y}^{(\pi)} \in \mathbb{R}^M$	current local center of process π
$c^{(\pi)}$	augmented cost term of subspace problem processed by π
$h^{(\pi)} \colon \Omega^{(\pi)} \to \mathbb{R}$	local objective function of π
$L_v^{(\pi)}(x_v, y^{(\pi)})$	local Lagrangian of process π associated with $v \in V^{(\pi)}$
$f^{(\pi)}(y^{(\pi)})$	local dual function of process π
$(\bar{l}^{(\pi)}, \bar{x}^{(\pi)}) \in \operatorname{conv} W^{(\pi)}$	local aggregate minorant of process π
$\hat{x}_v^{(\pi)} \in \Omega_v, v \in V^{(\pi)}$	optimal primal solution in current center $y^{(\pi)}$
$W^{(\pi)}$	linear minorants at $f^{(\pi)}$
$\hat{f}_{w_v,v}^{(\pi)}(y^{(\pi)})$	value of linear minorant $w_v \in \operatorname{conv} W_v$ of $f^{(\pi)}$ at $y^{(\pi)}$
$g^{(\pi)}(x^{(\pi)})$	subgradient of $f^{(\pi)}$ generated by $x^{(\pi)} \in \operatorname{conv} \Omega^{(\pi)}$
$\Delta^{(\pi)}$	predicted decrease in subspace problem of process π

In Section 4 for identifiers associated with the global states of the extended algorithm have a superscript $\{\sigma\}$, those associated with a process have a superscript $[\pi]$. The corresponding identifiers have the same meaning as in Section 3, the following additional identifiers appear:

$J_v^{\{\sigma\}}$	current active constraints of problem $v \in V$
$B_M^{\{\sigma\}} \subseteq M$	currently blocked constraints

and

$J_v^{[\pi]}$	constraints of $v \in V^{[\pi]}$ that are assumed to be active by π
$\bar{J}_v^{[\pi]}$	active constraints of $v \in V^{[\pi]}$ but not contained in $J^{[\pi]}$
$V_j^{[\pi]}$	problems for which $j \in J^{[\pi]}$ is active when π starts

References

[1] F. Babonneau, O. Du Merle, and J.-P. Vial. "Solving large-scale linear multicommodity flow problems with an active set strategy and proximal-ACCPM." In: *Oper. Res.* 54.1 (2006), pp. 184–197.

[2] J. F. Bonnans et al. *Numerical Optimization*. Springer, 2003.

[3] M. C. Ferris and O. L. Mangasarian. "Parallel Variable Distribution". In: *SIAM Journal on Optimization* 4.4 (1994), pp. 815–832. URL: http://link.aip.org/link/?SJE/4/815/1.

[4] F. Fischer and C. Helmberg. *Dynamic Graph Generation for Large Scale Operational Train Timetabling*. Preprint 2011-10. D-09107 Chemnitz, Germany: Technische Universität Chemnitz, Fakultät für Mathematik, 2011. URL: http://www.tu-chemnitz.de/mathematik/preprint/2011/PREPRINT_10.pdf.

[5] M. L. Fisher. "The Lagrangian Relaxation Method for Solving Integer Programming Problems". In: *Manage. Sci.* 50.12 Supplement (2004), pp. 1861–1871. ISSN: 0025-1909. URL: http://dl.acm.org/citation.cfm?id=1245920.1245938.

[6] M. Fukushima. "Parallel Variable Transformation in Unconstrained Optimization". In: *SIAM J. on Optimization* 8.3 (1998), pp. 658–672. ISSN: 1052-6234. URL: http://dl.acm.org/citation.cfm?id=588907.589327.

[7] D. Goldfarb and S. Ma. *Fast Alternating Linearization Methods for Minimizing the Sum of Two Convex Functions.* Tech. rep. arXiv:0912.4571. 2009. URL: http://www.optimization-online.org/DB_HTML/2009/12/2502.html.

[8] D. Goldfarb and S. Ma. *Fast Multiple Splitting Algorithms for Convex Optimization.* Tech. rep. arXiv:0912.4570. 2009. URL: http://www.optimization-online.org/DB_HTML/2009/12/2501.html.

[9] B. He, T. Min, and X. Yuan. *A splitting method for separate convex programming with linking linear constraints.* Tech. rep. 2008. URL: http://www.optimization-online.org/DB_HTML/2010/06/2665.html.

[10] C. Helmberg and K. C. Kiwiel. "A Spectral Bundle Method with Bounds". In: *Math. Programming* 93.2 (2002), pp. 173–194.

[11] C. Helmberg and S. Röhl. "A Case Study of Joint Online Truck Scheduling and Inventory Management for Multiple Warehouses". In: *Operations Research* 55.4 (July 2007), pp. 733–752.

[12] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I.* Vol. 305. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 1993.

[13] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II.* Vol. 306. Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 1993.

[14] C. Lemaréchal and A. Renaud. "A geometric study of duality gaps, with applications". In: *Mathematical Programming* 90.3 (2001), pp. 399–427. ISSN: 0025-5610. URL: http://dx.doi.org/10.1007/PL00011429.

[15] C. Lemaréchal. "Lagrangian Relaxation". In: *Computational Combinatorial Optimization.* Ed. by M. Jünger and D. Naddef. Vol. 2241. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2001, pp. 112–156. ISBN: 978-3-540-42877-0. URL: http://dx.doi.org/10.1007/3-540-45586-8_4.

[16] C. Lemaréchal. "The omnipresence of Lagrange". In: *Annals of Operations Research* 153.1 (2007), pp. 9–27. ISSN: 0254-5330. URL: http://dx.doi.org/10.1007/s10479-007-0169-1.

[17] D. Medhi. "Parallel bundle-based decomposition for large-scale structured mathematical programming problems". In: *Ann. Oper. Res.* 22.1-4 (1990), pp. 101–127. ISSN: 0254-5330. URL: http://portal.acm.org/citation.cfm?id=85598.85620.

[18] F. Meng et al. "Lagrangian-Dual Functions and Moreau–Yosida Regularization". In: *SIAM Journal on Optimization* 19.1 (2008), pp. 39–61. URL: http://link.aip.org/link/?SJE/19/39/1.

[19] M. Nayakkankuppam. "Solving large-scale semidefinite programs in parallel". In: *Mathematical Programming* 109.2 (2007), pp. 477–504. ISSN: 0025-5610. URL: http://dx.doi.org/10.1007/s10107-006-0032-1.

[20] I. Necoara and J. Suykens. "Interior-Point Lagrangian Decomposition Method for Separable Convex Optimization". In: *Journal of Optimization Theory and Applications* 143.3 (2009), pp. 567–588. ISSN: 0022-3239. URL: http://dx.doi.org/10.1007/s10957-009-9566-8.

[21] A. Nedić, D. Bertsekas, and V. Borkar. "Distributed asynchronous incremental subgradient methods". In: *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*. Ed. by Y. C. Dan Butnariu and S. Reich. Vol. 8. Studies in Computational Mathematics. Elsevier, 2001, pp. 381–407. URL: http://www.sciencedirect.com/science/article/pii/S1570579X01800239.

[22] C. A. Sagastizábal and M. V. Solodov. "Parallel Variable Distribution for Constrained Optimization". In: *Comput. Optim. Appl.* 22.1 (2002), pp. 111–131. ISSN: 0926-6003. URL: http://dl.acm.org/citation.cfm?id=584529.584535.

[23] R. Schultz. "Stochastic programming with integer variables". In: *Math. Programming* 97.1-2 (2003), pp. 285–309.