Math. Program., Ser. A (2017) 164:129–155
DOI 10.1007/s10107-016-1079-2

FULL LENGTH PAPER

Tight MIP formulations for bounded up/down times and interval-dependent start-ups

Maurice Queyranne[1,2] · **Laurence A. Wolsey**[1]

Received: 29 July 2015 / Accepted: 11 October 2016 / Published online: 31 October 2016

Abstract Switching machines on and off is an important aspect of unit commitment problems and production planning problems, among others. Here we study tight mixed integer programming formulations for two aspects of such problems: bounded length on- and off-intervals, and interval-dependent start-ups. The problem with both these aspects admits a general Dynamic Programming (shortest path) formulation which leads to a tight extended formulation with a number of binary variables that is quadratic in the number n of time periods. We are thus interested in tight formulations with a linear number of binary variables. For the bounded interval problem we present a tight network dual formulation based on new integer cumulative variables that allows us to simultaneously treat lower and upper bounds on the interval lengths and also to handle time-varying bounds. This in turn leads to more general results, including simpler proofs of known tight formulations for problems with just lower bounds. For the interval-dependent start-up problem we develop a path formulation that allows us to describe the convex hull of solutions in the space of machine state variables and interval-dependent start-up variables.

Keywords Production sequencing · Unit commitment · Bounded up/down times · Interval-dependent startups · Tight MIP formulations · Convex hulls

This research has been funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office.

✉ Laurence A. Wolsey
laurence.wolsey@uclouvain.be

Maurice Queyranne
maurice.queyranne@sauder.ubc.ca

1 CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium

2 Sauder School of Business, University of British Columbia, Vancouver, Canada

Mathematics Subject Classfication C44 · C61

1 Introduction

Switching machines on and off is an important aspect of unit commitment and production planning problems, among others. For unit commitment problems, four of the most commonly cited features are (1) nonlinear production costs, (2) start-up costs that are a function of the machine down time, (3) minimum up/down times and (4) ramp rates; whereas in production planning minimum run times may be necessary because it takes time for the product to stabilize or for other economic reasons, and maximum run times may be imposed because of machine deterioration, etc. Both these classes of problems are often formulated as mixed integer programs. As the successful solution of such problems often depends on the quality of the formulation, providing stronger formulations of different aspects of the problem may help significantly in obtaining good or optimal solutions. In particular a *best possible* (*tight*) formulation of a mixed integer set $X \subset \mathbb{Z}^n \times \mathbb{R}^p$ is provided by an explicit or implicit (with additional variables) description of the convex hull of X.

First we describe briefly some earlier work. There is a significant literature on MIP formulations of different aspects of unit commitment problems, see for instance Ostrowski et al. [14], Morales-España et al. [12] and numerous articles in the journal, *IEEE Transactions on Power Systems*, and on MIP formulations of production planning problems, see Pochet and Wolsey [16]. Among others Frangioni et al. [4] and Wu [21] discuss strong formulations of nonlinear production costs (1), and Damcı-Kurt et al. [2], Pan et al. [15], Morales-España et al. [11] and Knueven et al. [7] formulations of ramping constraints (4).

We consider the two other aspects: minimum/maximum up-down times (3), and time-varying (hot/warm/cold) start-up costs (2). In contrast with much of the work in the unit commitment literature and the production planning literature, our focus is on modeling, using binary variables, the combinatorial structure of solutions taking into account these aspects. For lower bounds on the length of on- and off-intervals, necessary inequalities can be found in Wolsey [20]. Malkin [9] showed that these inequalities describe the convex hull of solutions in the space of machine state and start-up variables, and Lee et al. [8] describe the convex hull in the space of the machine state variables only. Hedman et al. [6] discuss different formulations for minimum up/down times and show how their strength can be compared. See also [18]. For time independent start-up costs Morales-España et al. [11,12] and Viana and Pedroso [19] present basic MIP formulations using additional continuous variables representing power output (or production quantities).

In Sect. 2 we consider first the joint problem with both lower and upper bounds on the length of the on-intervals and interval-dependent start-up costs. We present a simple shortest path network formulation that provides a tight extended formulation with $O(n^2)$ constraints and variables and an $O(n^2)$ optimization algorithm for an n-period instance. Then in the following sections we examine cases in which it is possible to obtain a tight formulation with only $O(n)$ variables. In Sect. 3 we present a new tight network dual formulation for the problem with both lower and upper bounds on the length of the on- and off-intervals. This allows us to generalize and simplify earlier results (treating just lower bounds) of Malkin [9] and Lee et al. [8] and to treat time-varying bounds. Two of our findings may be surprising. First, in the joint space of the state and start-up variables a tight formulation for both lower and upper bounds on up and down times is obtained by simply juxtaposing formulations for each one of the four bound constraint types, see Eq. (56); this is not the case in the space of the state variables alone (Example 2). Second, in the latter space, formulations for maximum up-down times are much simpler (and smaller, with $O(n)$ inequalities) than those for minimum up-down times, which require an exponential number of inequalities (compare Proposition 5 with Theorem 3).

In Sect. 4 we turn to the problem of interval-dependent start-ups. Based on a different path formulation, we obtain via projection a description of the convex hull of solutions in the space of the machine state and start-up variables. In Sect. 5 we discuss possible extensions.

The subproblems treated here will normally be part of much larger unit commitment or production planning problems involving multiple machines and a variety of other complicating constraints. By finding tight and compact formulations for the single machine subproblems, one hopes to obtain tighter mixed integer formulations of the complete problems and improve the solution approaches by providing stronger dual bounds and/or good feasible solutions faster.

It is perhaps of interest that three different proof techniques are used in Sects. 3 and 4 to obtain the various polyhedral descriptions: integer "event count" variables in Sect. 3.1; dynamic programming functionals providing feasible solutions for an extended formulation in 3.2 as in Queyranne and Wolsey [17]; and node flows and Hoffman's circulation theorem, as in Martens et al. [10], for characterizing feasible flows in Sect. 4.

Notation

- For $p, q \in \mathbb{Z}$, $[p, q]$ denotes the set of integers lying in the interval from p to q, i.e. the set $\{p, p+1, \ldots, q\}$. In particular $[p, q] = \{p\}$ if $p = q$, and $[p, q] = \emptyset$ if $p > q$.
- For a set S, $x(S) \equiv \sum_{u \in S} x_u$. If $S = \emptyset$ then $x(S) \equiv 0$.

2 Problem definitions and initial formulations

2.1 The problem with on/off interval bounds

First we consider the problem with interval bounds. Given a discrete time horizon of n periods, there is first an interval (possibly empty) in which the machine is off, followed by interval in which it is on, then off, then on, etc. The first period of an on-interval is called a *start-up period* and the first period of an off-interval (apart from the initial off-interval) is called a *switch-off period*. Various constraints are considered. In particular one can have bounds on the lengths of the on- and off-intervals, as follows.

Bounds on-interval lengths We are given the following bounds:

$\alpha_t \geq 1$ is a lower bound on the length of an on-interval starting in period t;
$\beta_t \geq \alpha_t$ is an upper bound on the length of an on-interval starting in period t;

$\gamma_t \geq 1$ is a lower bound on the length of an off-interval starting in period t;
$\delta_t \geq \gamma_t$ is an upper bound on the length of an off-interval starting in period t.

We let $\boldsymbol{\alpha} = (\alpha_t)_{t\in[1,n]}$ denote the vector of lower bounds on the on-intervals, and $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ be defined similarly.

Decision variables As in Garver's "3-bin model" [5][1] we define the following binary decision variables:

$y_t = 1$ if period t is an on-period and $y_t = 0$ otherwise;
$z_t = 1$ if period t is a start-up period ($y_{t-1} = 0$ and $y_t = 1$);
$w_t = 1$ if period t is a switch-off period[2] ($y_{t-1} = 1$ and $y_t = 0$).

Let $\mathbf{y} = (y_t)_{t\in[1,n]}$ denote the vector of state variables, and $\mathbf{z}$ and $\mathbf{w}$ those of start-up and switch-off decisions. Note that the equation $y_t - y_{t-1} = z_t - w_t$ always links the state, start-up and switch-off variables. This allows us to eliminate either the $\mathbf{w}$ or the $\mathbf{z}$ variables. In Sect. 3.1 we shall do so and work in the "2-bin space" defined by the $\mathbf{y}$ and $\mathbf{z}$ variables, and then in Sect. 3.2 in the 1-bin space defined by projecting onto the subspace of the $\mathbf{y}$ variables only.

Initial conditions We assume that the initial state $y_0 \in \{0, 1\}$ of the machine is known and that the machine was last switched to its current state in the known period $-\tau \leq 0$. When the corresponding bounds (α_t and β_t if $y_0 = 1$, and γ_t and δ_t if $y_0 = 0$, for $t \leq 0$) are not constant over time, we also need to know their values, $\alpha_{-\tau}$ and $\beta_{-\tau}$ or $\gamma_{-\tau}$ and $\delta_{-\tau}$, that were in force when that switch occurred. This information may be summarized in our models by assuming that the switch just occurred in period 0, i.e., by letting

$$z_0 = y_0, \quad w_0 = 1 - y_0 \quad \text{and} \quad y_{-1} = 1 - y_0 \tag{1}$$

and modifying the bounds as follows:

- If $y_0 = 1$ then we adjust $\alpha_0 := \alpha_{-\tau} - \tau$ and $\beta_0 := \beta_{-\tau} - \tau$ to their residual values that still apply at $t = 0$. The bounds γ_0 and δ_0 on the off-intervals at $t = 0$ are irrelevant, we may simply let $\gamma_0 := 1$ and $\delta_0 := 1$.
- Else, $y_0 = 0$ and we similarly let $\gamma_0 := \gamma_{-\tau} - \tau$, $\delta_0 := \delta_{-\tau} - \tau$, $\alpha_0 := 1$ and $\beta_0 := 1$.

Terminal conditions We also need to specify how the lower bounds α_t and γ_t apply near the end n of the planning horizon, i.e., when $t + \alpha_t - 1 > n$ or $t + \gamma_t - 1 > n$. Alternative interpretations are possible:

- In a *hard horizon context*, the system is dismantled and ceases to exist after period n. Thus no startup should occur in any period t such that $t + \alpha_t - 1 > n$, and no switch-off should occur in any t such that $t + \gamma_t - 1 > n$, for this would lead to a violation of a lower bound constraint. This can be taken into account by adding to our models below the constraints $z_t = 0$ for all t such that $t + \alpha_t - 1 > n$ and $w_t = 0$ for all t such that $t + \gamma_t - 1 > n$.

[1] See also the overview of unit commitment models in [7, Section 2].

[2] Note that this definition of w_t differs by one period from that used in [16].

- In the more common *soft horizon context* the system will continue to operate after period n, and any residual lower bound constraint will then be taken care of later, and may be ignored in the present model. This arises in particular when using a *rolling horizon* approach, where the system is periodically reoptimized over successive planning horizons (such as $[1, n]$; then, after implementing period 1 decisions, $[2, n+1]$; etc.)

In either case we may redefine all $\alpha_t := \min\{\alpha_t, n+1-t\}$ and $\gamma_t := \min\{\gamma_t, n+1-t\}$, so after these modifications we have $t+\alpha_t-1 \le n$ and $t+\gamma_t-1 \le n$ for all $t \in [0, n]$.

Formulations. A simple, "natural" integer programming formulation is as follows:

$$z_t \ge y_t - y_{t-1} \qquad t \in [1, n]; \tag{2}$$

$$z_t \le y_t \qquad t \in [1, n]; \tag{3}$$

$$z_t \le 1 - y_{t-1} \qquad t \in [1, n]; \tag{4}$$

$$z_t \le y_u \qquad u \in [t,\ t+\alpha_t-1],\quad t \in [0, n]; \tag{5}$$

$$z_t \le \sum_{u=t+1}^{t+\beta_t} (1 - y_u) \qquad t : t \ge 0 \text{ and } t+\beta_t \le n; \tag{6}$$

$$w_t \le 1 - y_u \qquad u \in [t,\ t+\gamma_t-1],\quad t \in [0, n]; \tag{7}$$

$$w_t \le \sum_{u=t+1}^{t+\delta_t} y_u \qquad t : t \ge 0 \text{ and } t+\delta_t \le n; \tag{8}$$

$$y_t - y_{t-1} = z_t - w_t \qquad t \in [1, n]; \tag{9}$$

$$\mathbf{y},\ \mathbf{z},\ \mathbf{w} \in \{0, 1\}^n. \tag{10}$$

Here (2)–(4) and (10) model the link between start-ups and on-periods given that a machine is always on or off for at least one period; (5) ensures that the machine is on for α_t periods after a start-up in t; (6) that the machine is off in at least one of the β_t periods that follow a start-up in t; etc. Note that w_t is just the slack variable in (2).

If the lower bounds $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are bounded above by a constant, formulation (2)–(10) has $O(n)$ constraints. In general, however, the number of constraints is $O(n^2)$. We now present a tighter formulation that only uses $O(n)$ constraints. Recall that given two formulations P and Q of a mixed integer set $X \subset \mathbb{R}^p \times \mathbb{R}^q$ (i.e., polyhedra such that $X = P \cap (\mathbb{Z}^p \times \mathbb{R}^q) = Q \cap (\mathbb{Z}^p \times \mathbb{R}^q)$), P is *tighter* than Q if $P \subseteq Q$.

Proposition 1 *The formulation defined by constraints* (2)*,*

$$\sum_{\substack{u \in [0,t]:\\ u+\alpha_u > t}} z_u \le y_t \qquad t \in [1, n] \tag{11}$$

$$\sum_{\substack{u \in [0,t]:\\ u+\gamma_u > t}} w_u \le 1 - y_t \qquad t \in [1, n] \tag{12}$$

(6)*,* (8) *and* (9) *with* $\mathbf{y}, \mathbf{z}, \mathbf{w} \in [0, 1]^n$ *is tighter than* (2)–(9) *with* $\mathbf{y}, \mathbf{z}, \mathbf{w} \in [0, 1]^n$.

Proof We show that $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfies (2)–(10) if and only if it satisfies the constraints in this alternative formulation. First, note that since $\alpha_t \geq 1$ and $\gamma_t \geq 1$ constraints (11) and (12) imply (3) and (4), respectively. Next we show that (11) is a strengthening of (5), i.e., that replacing in (5) with (11) in (2)–(10) gives a tighter formulation. Indeed, exchanging the u and t indices we can rewrite (5) in the equivalent form:

$$z_u \leq y_t \qquad \text{all } u \text{ such that } 0 \leq u \leq t \leq u + \alpha_u - 1 \text{ and all } t \in [1, n],$$

where we safely ignore the case $t = 0$ which is taken care of by the initial conditions. If for any $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfying (2)–(10) the left hand side of (11) is nonzero then let u be the least index in the summation range for which $z_u = 1$. By (5) the condition $u + \alpha_u > t$ implies $y_j = 1$ for all $j \in [u, t]$ and thus, by (4), $z_{j+1} = 0$ for all such j. Therefore the left hand side of (11) is at most one, and when it is equal to one then so is y_t. This shows that (11) is indeed a strengthening of (5).

A similar argument, exchanging the on and off states, the lower bounds $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ and also the start-ups and switch-offs, shows that (12) is a strengthening of (7). □

Inequalities (11) and (12) generalize the minimum run time and minimum down time[3] inequalities, respectively, of Wolsey [20, p. 235] (called "turn on" and "turn off inequalities" in Rajan and Takriti [18]) to the present case of nonconstant lower bounds $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

In the rest of this paper we shall eliminate the switch-off variables $\mathbf{w}$ [which can be recovered using Eq. (9)] and focus on formulations using only the $(\mathbf{y}, \mathbf{z})$ or the $\mathbf{y}$ variables.

Feasible sets and their convex hulls Let $Z(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$ denote the set of vectors $(\mathbf{y}, \mathbf{z})$ for which there exists $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfying (2)–(10). More generally, for any set $B \subseteq \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$ of interval bound types, let $Z(B)$ denote the set of vectors $(\mathbf{y}, \mathbf{z})$ for which there exists $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfying (2)–(4), (9)–10 and the subset of (5)–(8) corresponding to B; i.e., $Z(B)$ is the set of state and start-up vectors of all solutions satisfying the interval bound constraints specified by B. Let $Y(B)$ denote the projection of $Z(B)$ onto the coordinate subspace of the state variables $\mathbf{y}$. If $B = \emptyset$ then $Z(\emptyset)$ is the set of feasible $(\mathbf{y}, \mathbf{z})$ vectors for the problem without interval bounds, and $Y(\emptyset) = \{0, 1\}^n$. For B nonempty, by definition we have

$$Z(B) = \bigcap_{\epsilon \in B} Z(\boldsymbol{\epsilon}) \text{ and } Y(B) = \bigcap_{\epsilon \in B} Y(\boldsymbol{\epsilon}) \text{ for } \emptyset \neq B \subseteq \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\} \tag{13}$$

(where we simplify notation by omitting braces, i.e., where $Z(\boldsymbol{\epsilon})$ stands for $Z(\{\boldsymbol{\epsilon}\})$, and also $Z(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$ for $Z(\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\})$, etc.) Note also that $Z(B)$ is the set of all $(\mathbf{y}, \mathbf{z})$ which, with $\mathbf{w}$ defined through (9), satisfy the natural or alternate formulation above with only those constraints from (6) and (8), or (11) and (12), that correspond to whether bound types $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are in B. Our goal in Sect. 3 will be to describe the convex hulls $conv(Z(B))$ and $conv(Y(B))$ for all $B \subseteq \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$.

[3] The minimum down-times are denoted by β in Wolsey [20], while we use γ in the present paper.

2.2 The problem with interval-dependent start-ups

The costs of a start-up may depend on the time that the machine has been idle – in particular one talks of hot, warm and cold start-ups [11,12,19]. We consider a general model of start-up types defined by thresholds.

Start-up thresholds For each $t \in [1, n]$ let

$P_t \geq 1$ denote the number of start-up types available in period t, and
$\boldsymbol{\theta}_\mathbf{t} = \left(\theta_t^0, \theta_t^1, \ldots, \theta_t^{P_t}\right)$ the *threshold vector* for period t, with components satisfying

$$0 \leq \theta_t^0 < \theta_t^1 < \cdots < \theta_t^{P_t} \tag{14}$$

For $p = 1, \ldots, P_t$ a *type-p start-up* occurs in period t if the machine has been idle for between $\theta_t^{p-1} + 1$ and θ_t^p periods just before t.

Startup type decision variables Let

$z_t^p = 1$ if if there is a type-p start-up in period t and $z_t^p = 0$ otherwise.

Let $\mathbf{z_t} = (z_t^p)_{p \in [1, P_t]}$ be the start-up type vector in period t. Note that

$$z_t = \sum_{p=1}^{P_t} z_t^p \qquad t \in [1, n]. \tag{15}$$

As above, we assume we know the initial state $y_0 \in \{0, 1\}$ and the most recent period $-\tau \leq 0$ when the machine was switched to its current state. While this information suffices, we shall assume, to simplify the expression of constraints (18) and (19) below, that we actually know the state y_t of the machine for all $t \in \left[-\max_u \left\{\theta_u^{P_u}\right\},\ 0\right]$.

Formulation A basic integer programming formulation is:

$$\sum_{p=1}^{P_t} z_t^p \geq y_t - y_{t-1} \qquad t \in [1, n]; \tag{16}$$

$$\sum_{p=1}^{P_t} z_t^p \leq y_t \qquad t \in [1, n]; \tag{17}$$

$$z_t^p \leq 1 - y_{t-1-j} \qquad j \in [0, \theta_t^{p-1}], \quad p \in [1, P_t], \quad t \in [1, n]; \tag{18}$$

$$z_t^p \leq \sum_{j=\theta_t^{p-1}+1}^{\theta_t^p} y_{t-1-j} \qquad p \in [1, P_t], \quad t \in [1, n]; \tag{19}$$

$$\mathbf{z_t} \in \{0, 1\}^{P_t} \qquad t \in [1, n]; \tag{20}$$

$$\mathbf{y} \in \{0, 1\}^n. \tag{21}$$

Inequality (18) implies that if there is a type-p start-up in t, the machine was off in the interval $[t-\theta_t^{p-1}-1, t-1]$, whereas (19) implies that the machine was on sometime in the interval $\left[t-\theta_t^{p}-1,\ t-\theta_t^{p-1}-2\right]$.

Let $\mathbf{\Theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_n})_{t\in[1,n]}$ and $H(\mathbf{\Theta})$ denote the set of vectors $(\mathbf{y}, \mathbf{z_1}, \ldots, \mathbf{z_n})$ satisfying (16)–(21). In Sect. 4 we will describe $conv(H(\mathbf{\Theta}))$.

2.3 Optimization and a tight extended formulation for the general problem

We now consider the general model combining the interval length bounds $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$ of Sect. 2.1 and the interval dependent start-ups of Sect. 2.2 with thresholds $\mathbf{\Theta}$. Note that, for consistency, we must have for all $u \le t-1$

$$u+\gamma_u > t \ \text{ if and only if } \ u \ge t-\theta_t^0\ , \quad \text{and} \tag{22}$$

$$u+\delta_u < t \ \text{ if and only if } \ u < t-\theta_t^{P_t} \tag{23}$$

Indeed, condition (22) implies at least θ_t^0 off periods before t, and (23) at most $\theta_t^{P_t}$ such off periods. A formulation for this problem is thus (2)–(10), (16)–(20) and the linking constraints (15). Let $Z(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \mathbf{\Theta})$ denote its feasible set.

One can apply a general Dynamic Programming approach (see, e.g., [16]) to optimize a linear objective over $Z(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \mathbf{\Theta})$. The DP approach below may be viewed as a simplified version of that in Frangioni and Gentile [3]. For the purpose of the present paper it suffices to describe it as a shortest path problem in an acyclic digraph $D=(V,A)$. The nodes are $V=\{0\}\cup\{1,\ldots,n\}\cup\{1',\ldots,n'\}\cup\{n+1\}$ and the arcs are of two types, $A=A_1\cup A_2$: an arc $(i',j)\in A_1$ represents a switch-off in i followed by a start-up in j, and an arc $(i,j')\in A_2$ represents a start-up in i followed by a switch-off in j. More precisely, because of the bounds, $\{(i',j):\gamma\le j-i\le\delta\}\subseteq A_1$ and $\{(i,j'):\alpha\le j-i\le\beta\}\subseteq A_2$. The initial and terminal conditions define some additional arcs leaving node 0 and arriving at node $n+1$, respectively. Namely, if the machine is off in 0 and the last switch-off was in $-\tau\le 0$ then one includes the arcs $(0,t)\in A_1$ for $t\in\left[\max\{1,\ -\tau+\gamma_{-\tau}\},\ -\tau+\delta_{-\tau}\right]$, while if the machine is on in 0 and the last start-up was in $-\tau\le 0$ then one includes the arcs $(0,t')\in A_2$ for $t\in[\max\{1,\ -\tau+\alpha_{-\tau}\},\ -\tau+\beta_{-\tau}]$; if $i+\beta_i>n$ then there is an arc $(i,n+1)\in A_1$, and if $i+\delta_i>n$ then there is an arc $(i',n+1)\in A_2$.

It is well-known that the corresponding flow polyhedron, described by the following flow variables and in which 1 unit enters at node 0 and leaves at node $n+1$, has integral extreme points corresponding to the paths P from 0 to $n+1$.

Flow variables Let

$x_{ij}^1=1$ iff $(i',j)\in A_1\cap P$, and $x_{ij}^1=0$ otherwise;
$x_{ij}^2=1$ iff $(i,j')\in A_2\cap P$, and $x_{ij}^2=0$ otherwise.

Thus $x_{ij}^1=1$ if there is a switch-off in period i followed by a start-up in period j, while $x_{ij}^2=1$ if there is a start-up in period i followed by a switch-off in j.

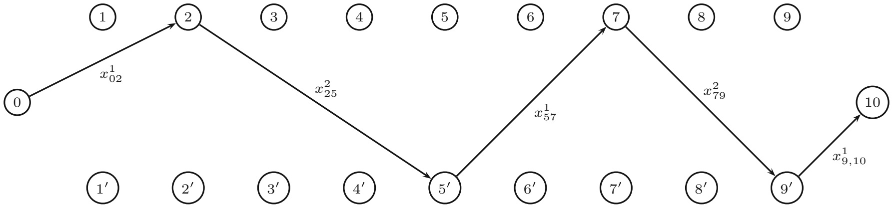

Fig. 1 A path for an instance with $n = 9$ and $y_0 = 0$. The on-intervals are [2, 4] and [7, 8]

Example 1 In Fig. 1 the path corresponding to a feasible solution of an instance with $y_0 = 0$ and $n = 9$ periods is shown. The on-intervals are the intervals [2, 4] and [7, 8]. The corresponding solution path is $x^1_{02} = x^2_{25} = x^1_{57} = x^2_{79} = x^1_{9,10} = 1$. The last switch-off occurred in period $-\tau = -2$. Assume a start-up after 1 or 2 off periods is "hot", and it is "cold" after 3, 4, 5 or 6 periods. Thus we have the time-independent thresholds $\boldsymbol{\theta_t} = (\theta^0, \theta^1, \theta^2) = (0, 2, 6)$ for all t, and the first start-up is cold ($p = 2$), as the off-interval $[-2, 1]$ is of length $3 \in [\theta^1 + 1, \theta^2]$, while the second start-up is hot ($p = 1$), as the off-interval [5, 6] is of length $2 \in [\theta^0 + 1, \theta^1]$.

Theorem 1 *(i) The polyhedron Q:*

$$\sum_{(0,j)\in A_1} x^1_{0j} + \sum_{(0,j')\in A_2} x^2_{0j} = 1; \tag{24}$$

$$\sum_{(i',t)\in A_1} x^1_{it} - \sum_{(t,j')\in A_2} x^2_{tj} = 0 \qquad t \in [1, n]; \tag{25}$$

$$\sum_{(i,t')\in A_2} x^2_{it} - \sum_{(t',j)\in A_1} x^1_{tj} = 0 \qquad t \in [1, n]; \tag{26}$$

$$\sum_{\substack{(i,j')\in A_2 \\ i \le t < j}} x^2_{ij} = y_t \qquad t \in [1, n]; \tag{27}$$

$$\sum_{\substack{(i',t)\in A_1 \\ t-\theta^p_t \le i < t-\theta^{p-1}_t}} x^1_{it} = z^p_t \qquad p \in [1, P_t], \quad t \in [1, n]; \tag{28}$$

$$\mathbf{x^1} \in \mathbb{R}^{A_1}_+; \quad \mathbf{x^2} \in \mathbb{R}^{A_2}_+; \tag{29}$$

$$\mathbf{z_t} \in [0, 1]^{P_t} \qquad t \in [1, n]; \tag{30}$$

$$\mathbf{y} \in [0, 1]^n. \tag{31}$$

is integral.

(ii) $proj_{y,z}(Q) = conv\,(Z(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\Theta}))$.

(iii) For any objective $(\mathbf{f}, \mathbf{c_1}, \ldots, \mathbf{c_n})$*, the linear program* $\max\left\{\mathbf{fy} - \sum_t \mathbf{c_t z_t} : (\mathbf{y}, \mathbf{z}, \mathbf{x^1}, \mathbf{x^2}) \in Q\right\}$ *solves the optimization problem* $\max\left\{\mathbf{fy} - \sum_t \mathbf{c_t z_t} : (\mathbf{y}, \mathbf{z}) \in Z(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\Theta})\right\}$.

Proof Constraints (24), (25), and (26) are flow conservation constraints at nodes 0, $\{1,\ldots,n\}$ and $\{1',\ldots,n'\}$ respectively. Equation (27) indicates that the machine is on in t if and only if there is a start-up in $i \le t$ followed by a switch-off in j with $j > t$. Equation (28) shows that there is a type-p start-up in t if there is a switch-off in $[t-\theta_t^p, t-\theta_t^{p-1}-1]$ followed by a start-up in t. Using these equations to eliminate the $\mathbf{y}$ and $\mathbf{z}$ variables, one can rewrite the objective function as a linear function $\sum_{ij\in A_1}\phi^1_{ij}x^1_{ij}+\sum_{ij\in A_2}\phi^2_{ij}x^2_{ij}$ in the $(\mathbf{x^1},\mathbf{x^2})$ variables (where $\phi^1_{ij}=-c^p_j$ if $j-i\in[\theta_j^{p-1}+1,\theta_j^p]$ and $\phi^2_{ij}=\sum_{u=i}^{j-1}f_u$), leaving a linear program over the path polytope (24)–(26), (29). □

The acyclic digraph contains about $\sum_t((\beta_t-\alpha_t+1)+(\delta_t-\gamma_t+1))=O(n^2)$ arcs. Thus:

Corollary 1 *The optimization problem over $Z(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\Theta})$ can be solved as a longest path problem in an acyclic digraph with $O(n^2)$ arcs in $O(n^2)$ time.*

When the upper bounds β_t and δ_t (or the differences $\beta_t-\alpha_t$ and $\delta_t-\gamma_t$) are bounded above by a constant, the acyclic digraph contains $O(n)$ arcs and a longest path can be found in linear time.

There are various ways in which Theorem 1 and its corollary can be used when solving a problem containing the constraints defining $Z(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\Theta})$ among others. In decomposition algorithms such as Dantzig-Wolfe or Lagrangian relaxation the optimization problem is solved repeatedly, while the polyhedron Q can be used to generate cuts in the space of the $(\mathbf{y},\mathbf{z_1},\ldots,\mathbf{z_n})$ variables. More specifically, given a solution $(\bar{\mathbf{y}},\bar{\mathbf{z}}_1,\ldots,\bar{\mathbf{z}}_n)$, one fixes the variables at these values in formulation Q and then uses the dual variables in the resulting linear programming feasibility problem to generate an inequality that is valid for *conv* $(Z(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\Theta}))$ and cuts off the point $(\bar{\mathbf{y}},\bar{\mathbf{z}}_1,\ldots,\bar{\mathbf{z}}_n)$, see for example [1,13].

3 Polyhedral results for on/off interval bounds

In this Section we consider, as in Sect. 2.1, bounds on on/off intervals (without interval-dependent start-up costs). Throughout Sect. 3 we assume that these bounds satisfy the *weak monotonicity* condition,

$$t+\epsilon_t \le u+\epsilon_u \qquad \text{for } 0\le t<u\le n \text{ and all } \boldsymbol{\epsilon}\in\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta},\}. \tag{32}$$

In particular this implies that by waiting one period, one cannot be forced to switch on or off earlier. As before, we assume the initial state $y_0\in\{0,1\}$ of the machine is known and that the bounds $(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta})$ and variables (z_0,w_0) have been modified to reflect the initial and terminal conditions, see Sect. 2.1 for details. Note that these modifications preserve weak monotonicity.

3.1 Convex hull in the (y, z)-space

We first present valid formulations $P_Z(B)$ for the sets $Z(B)$, namely polytopes such that $Z(B)=P_Z(B)\cap(\mathbb{Z}^n\times\mathbb{Z}^n)$, for every set $B\subseteq\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\delta}\}$ of interval bound types. For every $\boldsymbol{\epsilon} \in B$ let

$$s_B(\boldsymbol{\epsilon}, t) = \min\{u \in [0, t] : u + \epsilon_u > t\} \qquad \text{for all } t \in [1, n]. \tag{33}$$

Since all $\epsilon_t \geq 1$, $s_B(\boldsymbol{\epsilon}, t)$ is well defined. If $\boldsymbol{\alpha} \notin B$ (resp., $\boldsymbol{\gamma} \notin B$) then we could equivalently have all $\alpha_t = 1$ (resp., all $\gamma_t = 1$), and thus we let $s_B(\boldsymbol{\alpha}, t) = t$ (resp., $s_B(\boldsymbol{\gamma}, t) = t$) in these cases. If $\boldsymbol{\beta} \notin B$ (resp., $\boldsymbol{\delta} \notin B$) then we could have all $\beta_t = n$ (resp., all $\delta_t = n$), and thus we let $s_B(\boldsymbol{\beta}, t) = 0$ (resp., $s_B(\boldsymbol{\delta}, t) = 0$). By the weak monotonicity assumption (32) we have for any B and every $\boldsymbol{\epsilon} \in \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$ and $t \in [1, n]$,

$$u \in [s_B(\boldsymbol{\epsilon}, t), t] \quad \text{if and only if} \quad (u \in [0, t] \text{ and } u + \epsilon_u > t)\,. \tag{34}$$

Proposition 2 *For every set $B \subseteq \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$ of interval bound types the polytope $P_Z(B)$*

$$z_0 = y_0; \tag{35}$$

$$z_t \geq y_t - y_{t-1} \qquad t \in [1, n]; \tag{36}$$

$$\sum_{u=s_B(\boldsymbol{\alpha},t)}^{t} z_u \leq y_t \qquad t \in [1, n]; \tag{37}$$

$$y_t \leq \sum_{u=s_B(\boldsymbol{\beta},t)}^{t} z_u \qquad t \in [1, n]; \tag{38}$$

$$\sum_{u=s_B(\boldsymbol{\gamma},t)}^{t} z_u \leq 1 - y_{s_B(\boldsymbol{\gamma},t)-1} \qquad t \in [1, n] : s_B(\boldsymbol{\gamma}, t) \geq 1; \tag{39}$$

$$1 - y_{s_B(\boldsymbol{\delta},t)-1} \leq \sum_{u=s_B(\boldsymbol{\delta},t)}^{t} z_u \qquad t \in [1, n] : s_B(\boldsymbol{\delta}, t) \geq 1; \tag{40}$$

$$0 \leq y_t \leq 1 \qquad t \in [1, n]; \tag{41}$$

$$0 \leq z_t \leq 1 \qquad t \in [1, n] \tag{42}$$

is a formulation for $Z(B)$.

Proof First, note that the initial conditions (1) imply (35) and also the possibility that $s_B(\boldsymbol{\alpha}, t) = 0$ and $s_B(\boldsymbol{\beta}, t) = 0$. On the other hand, by (1) the constraints (39) and (40) are trivially satisfied when $s_B(\boldsymbol{\gamma}, t) = 0$ and when $s_B(\boldsymbol{\delta}, t) = 0$, respectively, so we may restrict these constraints to periods t such that $s_B(\boldsymbol{\gamma}, t) \geq 1$ and $s_B(\boldsymbol{\delta}, t) \geq 1$, respectively.

Next, we show that in the natural formulation (2)–(10) constraints (6) may be replaced with constraints (38). Indeed, consider any $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfying (2)–(10). For any $t \in [1, n]$ such that $y_t = 1$ let $u \leq t$ denote the latest period with $z_u = 1$, therefore $u + \beta_u - 1 \geq t$ and the right hand side of (38) is at least one. Thus $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfies (38). Conversely, if $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfies (2)–(5) and (7)–(10) but there exists some period u at which (6) is violated, i.e., such that $z_u = 1$ and $y_j = 1$ for all $j \in [u, u + \beta_u]$, then let $t = u + \beta_u$. By weak monotonicity, $s_B(\boldsymbol{\beta}, t) \geq u + 1$. Constraints (2) imply that $z_j = 0$ for all $j \in [u + 1, t]$. Thus the right hand side of (38) is zero whereas $y_t = 1$, violating constraint (38). Therefore (6) and (38) are indeed equivalent for all $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfying (2)–(5) and (7)–(10).

A similar argument, exchanging the on and off states, the upper bounds $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ and also the start-ups and switch-offs, shows that (8) may be replaced with (40).

Finally, it now suffices to verify that, after substituting the switch-off variables w_u in (8) and, by (34), replacing the summation ranges in (11)–(12) with intervals $[s_B(\boldsymbol{\epsilon}, t), t]$, formulation (36)–(42) is equivalent to the formulation in Proposition 1. □

Remark 1 If $s_B(\boldsymbol{\alpha}, t) = 0$ then $\alpha_0 > t \geq 1$ and, by our treatment of the initial conditions, we must have $y_0 = z_0 = 1$. This then implies $z_u = 0$ for all $u \in [1, t]$ and constraint (37) correctly forces $y_t = 1$. If $s_B(\boldsymbol{\beta}, t) = 0$ then $\beta_0 > t \geq 1$ and again $z_0 = 1$, so constraint (38) has no effect. Similarly, if $s_B(\boldsymbol{\gamma}, t) = 0$ then $\gamma_0 > t \geq 1$ and thus $y_0 = 0$, $w_0 = 1$ and $y_{-1} = 1$, and constraint (39) correctly forces $y_u = 0$ for all $u \in [1, t]$, whereas with $s_B(\boldsymbol{\delta}, t) = 0$ constraint (38) has no effect.

Remark 2 In the natural formulation (2)–(10) the interval bounds are enforced by the $O(n^2)$ inequalities (5)–(8) in which individual state change decisions (z_t and w_t) constrain future states (y_u with $u \geq t$). In the formulation (35)–(42) of Proposition 2, these bounds are enforced by the $O(n)$ inequalities (37)–(40), in which a single state variable constrains past decisions. Thus we may think of the natural formulation as being "forward looking" (current decisions constrain future states), and of the alternate formulation as "backward looking" (current states constrain past decisions.)

We will show that, under the weak monotonicity assumption, formulation (35)–(42) is not only smaller than the natural formulation, but also that it is a tightest possible formulation of $Z(B)$, i.e., that $conv(P_Z(B)) = Z(B)$. For this, we now introduce integer "count" (or cumulative) variables:

Integer count variables For all $t \in [1, n]$ let

$v_t \in \mathbb{Z}_+$ denote the number of start-ups in the interval $[0, t]$, i.e., $v_t = \sum_{j=0}^{t} z_j = y_0 + \sum_{j=1}^{t} z_j$, and

$u_t \in \mathbb{Z}_+$ denote the number of switch-offs in the interval $[1, t]$, i.e., $u_t = \sum_{j=1}^{t} w_j$.

These variables form the integer vectors $\mathbf{v}$ and $\mathbf{u}$ and we may define their initial values $v_0 = y_0$ and $u_0 = 0$. Since on- and off-intervals alternate, v_t and u_t differ by at most one unit. More precisely:

Observation 1 *There is a one-to-one unimodular transformation between the variables* $(\mathbf{u}, \mathbf{v})$ *and* $(\mathbf{y}, \mathbf{z})$ *given by:*

$$z_t = v_t - v_{t-1} \qquad t \in [1, n]; \tag{43}$$

$$y_t = v_t - u_t \qquad t \in [1, n] \tag{44}$$

where $v_0 = y_0$. *With* $u_0 = 0$ *and, by the initial condition* $z_0 = y_0$ *in* (1), *equations* (43)–(44) *are also satisfied for* $t = 0$ *if we let* $v_{-1} = 0$. *In addition, one has the link to the switch-off variables given by:*

$$w_t = z_t + y_{t-1} - y_t = u_t - u_{t-1} \qquad t \in [0, n] \tag{45}$$

where the case $t = 0$ *follows from the initial conditions* $w_0 = y_{-1} = 1 - y_0$ *if we let* $u_{-1} = y_0 - 1$.

Proposition 3 *Let* $Q_{UV}(B) \subset \mathbb{R}^n \times \mathbb{R}^n$ *be the polytope:*

$$u_{-1} = y_0 - 1; \tag{46}$$

$$v_{-1} = u_0 = 0; \tag{47}$$

$$v_0 = y_0; \tag{48}$$

$$u_t - u_{t-1} \geq 0 \qquad t \in [1, n]; \tag{49}$$

$$u_t - v_{s_B(\boldsymbol{\alpha},t)-1} \leq 0 \qquad t \in [1, n]; \tag{50}$$

$$v_{s_B(\boldsymbol{\beta},t)-1} - u_t \leq 0 \qquad t \in [1, n]; \tag{51}$$

$$v_t - u_{s_B(\boldsymbol{\gamma},t)-1} \leq 1 \qquad t \in [1, n] : s_B(\boldsymbol{\gamma}, t) \geq 1; \tag{52}$$

$$v_t - u_{s_B(\boldsymbol{\delta},t)-1} \geq 1 \qquad t \in [1, n] : s_B(\boldsymbol{\delta}, t) \geq 1; \tag{53}$$

$$0 \leq v_t - u_t \leq 1 \qquad t \in [1, n]; \tag{54}$$

$$0 \leq v_t - v_{t-1} \leq 1 \qquad t \in [1, n]. \tag{55}$$

$Q_{UV}(B)$ *with the linking Eqs.* (43)–(44) *defines an extended formulation for* conv $(Z(B))$ *under the unimodular transformation of Observation 1.*

Proof Constraints (46)–(48) set initial values for the u and v variables according to Observation 1. Constraints (49)–(55) are obtained from (36)–(42) by substitution. □

Remark 3 Note that the inequalities in (52) and (53) actually hold for all $t \in [1, n]$. Indeed, if $s_B(\boldsymbol{\gamma}, t) < 1$ then $\gamma_0 > t \geq 1$ and thus $y_0 = 0$, which implies $z_j = 0$ for all $j \in [0, \gamma_0 - 1] \supseteq [0, t]$, and thus $v_t = 0$ and (52) holds with equality. The same argument also implies that (53) holds with equality when $s_B(\boldsymbol{\delta}, t) < 1$.

Theorem 2 *For every set* $B \subseteq \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$ *of interval bound types the polytopes* $P_Z(B)$ *and* $Q_{UV}(B)$ *are integral and* $P_Z(B) = $ conv $(Z(B))$.

Proof To see that the polyhedron $Q_{UV}(B)$ is integral, we observe that each constraint in (46)–(55) has at most one +1 and one -1 coefficient. Thus the corresponding matrix is the dual of a network matrix and is totally unimodular. As the right hand-side is integer, the extreme point solutions are integer. Adding the defining Eqs. (43) and (44) preserves integrality. Combined with Proposition 3, the claim follows. □

Thus, under the weak monotonicity assumption (32), for $B = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$ formulation (35)–(42) is not only smaller than the natural formulation (2)–(10), it is also tighter. For any nonempty B it also follows from Theorem 2 that we can obtain the convex hull $conv(Z(B) = P_Z(B)$ of the corresponding feasible set by simply adding the corresponding constraint(s) from (37)–(40) to the formulation $P_Z(\emptyset)$. Therefore, perhaps surprisingly,

$$conv(Z(B)) = \bigcap_{\epsilon \in B} conv(Z(\epsilon)) \qquad \text{for all } B \subseteq \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\},\ B \neq \emptyset. \tag{56}$$

In other words, in the $(\mathbf{y}, \mathbf{z})$ space the various bound types on the interval lengths do not interact, i.e., combining different bound types does not give rise to new facets. For $Z(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ with lower bounds $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ constant over time, inequalities (37) and (39) were, as mentioned above, presented in Wolsey [20, p. 235]. Malkin [9], and later Rajan and Takriti [18], proved the integrality of the polytope (36)–(37), (39) and (41)–(42). For this reason, Hedman et al. conclude that (37) and (39) are "the preferred valid inequalities to model the minimum up and down time constraints" [6, p. 5]. This formulation was also tested computationally by Rajan and Takriti [18].

3.2 Convex hull in the y-space

We now consider the question of describing the convex hulls of the projections $Y(B)$ of the sets $Z(B)$ onto the $\mathbf{y}$-subspace. As the projection of a polytope with integer vertices is a polytope with integer vertices, the convex hull can be obtained in principle by using Fourier-Motzkin to eliminate the $\mathbf{z}$ variables from the constraints defining the polytopes $P_Z(B)$.

To describe an important family of (known) valid inequalities, we introduce some notation.

Definition Let $S = \{j_1, \ldots, j_k\}$ with $1 \leq j_1 < \cdots < j_k \leq n$.

- If $k = |S|$ is odd, $Odd(S, y) \equiv y_{j_1} - y_{j_2} + \cdots - y_{j_{k-1}} + y_{j_k}$ and
- if $k = |S|$ is even, $Even(S, y) \equiv y_{j_1} - y_{j_2} + \cdots - y_{j_k}$.
- Length$(S) \equiv j_k - j_1$.

An inequality of the form

$$Odd(S, \mathbf{y}) \leq \mu \ \text{ or } \ Odd(S, \mathbf{y}) \geq \mu$$

for some integer μ is called an alternating inequality.

Observation 2 *Let $S = \{j_1, \ldots, j_k\} \subseteq [t, \tau]$. With k even, there exists $T \subseteq [t, \tau - 1]$ such that*

$$Even(S, \mathbf{y}) = \sum_{j \in T} (y_j - y_{j+1}),$$

and vice versa; and with k odd there exists $U \subseteq [t, \tau - 2]$ such that

$$Odd(S, \mathbf{y}) = \sum_{j \in U} (y_j - y_{j+1}) + y_{j_k},$$

and vice versa.

Proof With k even, it suffices to take $T = [j_1, j_2-1] \cup [j_3, j_4-1] \cup \cdots \cup [j_{k-1}, j_k-1]$. For the converse, if T consist of intervals with $T = [p_1, q_1] \cup [p_2, q_2] \cup \cdots \cup [p_r, q_r]$, then it suffices to take $S = \{p_1, q_1+1, p_2, q_2+1, \ldots, p_r, q_r+1\}$. The case with k odd follows. □

Proposition 4 *(i) All alternating inequalities* $Odd(S, \mathbf{y}) \geq 0$ *with* $S \subseteq [s_B(\boldsymbol{\alpha}, t), t]$ *and* $t \in [0, n]$ *are valid for* $Y(\boldsymbol{\alpha})$.

(ii) All alternating inequalities $Odd(S, \mathbf{y}) \leq 1$ *with* $S \subseteq [s_B(\boldsymbol{\gamma}, t), t]$ *and* $t \in [0, n]$ *are valid for* $Y(\boldsymbol{\gamma})$.

(iii) The inequalities $\sum_{j=t}^{t+\beta_t} y_j \leq \beta_t$ *are valid for* $Y(\boldsymbol{\beta})$ *for all* $t \in [0, n]$ *such that* $t + \beta_t \leq n$.

(iv) The inequalities $\sum_{j=t}^{t+\delta_t} y_j \geq 1$ *are valid for* $Y(\boldsymbol{\delta})$ *for all* $t \in [0, n]$ *such that* $t + \delta_t \leq n$.

Proof The validity of the alternating inequalities is simple. Namely, replacing each z_t variable by either lower bound 0 or $y_t - y_{t-1}$ in (37) and (39), respectively, gives the first two sets of inequalities. The validity of the last two inequalities is immediate from the definitions of β_t and δ_t. □

We now examine the polytope described by the alternating inequalities as well as considering the separation problem for these inequalities. Given a prescribed initial state $y_0^* \in \{0, 1\}$ and a vector $\mathbf{y^*} \in [0, 1]^n$, define

$$F(t) = \max_{S \subseteq [0,t]\,:\,|S| \text{ odd}} Odd(S, \mathbf{y^*}) \quad \text{and} \quad G(t) = \max_{S \subseteq [0,t]\,:\,|S| \text{ even}} Even(S, \mathbf{y^*}).$$

The resulting vectors $\mathbf{F}$ and $\mathbf{G}$ are easy to compute and have interesting properties. To compute them in linear time, one has the recursions:

$$F(t) = \max\{F(t-1),\ G(t-1) + y_t^*\}$$
$$\text{and} \quad G(t) = \max\{F(t-1) - y_t^*,\ G(t-1)\}$$

with the initial values $F(0) = y_0^*$, $F(-1) = G(0) = 0$ and $G(-1) = y_0^* - 1$.

Lemma 1 *(i)* $0 \leq F(t) - F(t-1) \leq 1$ *for all* $t \in [0, n]$;

(ii) $0 \leq G(t) - G(t-1) \leq 1$ *for all* $t \in [0, n]$;

(iii) $0 \leq F(t) - G(t) = y_t^* \leq 1$ *for all* $t \in [0, n]$;

(iv) $F(t) - G(\tau) = \max_{S \subseteq [\tau,t]} Odd(S, \mathbf{y^*})$ *for* $0 \leq \tau \leq t \leq n$;

(v) $F(\tau) - G(t) = \min_{S \subseteq [\tau,t]} Odd(S, \mathbf{y^*})$ *for* $0 \leq \tau \leq t \leq n$.

Proof The first two range inequalities follow directly from the recursion and $\mathbf{y^*} \in [0, 1]^n$. $F(t) - G(t) = y_t^*$ is immediate from the definitions of $F(t)$ and $G(t)$. To establish *iv)*, note that $G(k+1) - G(k) = \max\{G(k),\ F(k) - y_{k+1}^*\} - G(k) = \max\{0,\ y_k^* - y_{k+1}^*\}$. Therefore for $\tau \leq t$, $F(t) - G(\tau) = \sum_{j=\tau}^{t-1}(G(j+1) - G(j)) + F(t) - G(t) = \sum_{j=\tau}^{t-1} \max\{0,\ y_j^* - y_{j+1}^*\} + y_t^* = \max_{S \subseteq [\tau,t]} Odd(S, \mathbf{y^*})$, where the last equation follows from Observation 2. The proof of *v)* is similar. □

This leads to a simple proof of the structure of $conv(Y(\boldsymbol{\alpha}, \boldsymbol{\gamma}))$.

Theorem 3 *$conv(Y(\boldsymbol{\alpha},\boldsymbol{\gamma}))$ is given by the two families of alternating inequalities in Proposition 4.*

Proof Let P be the polytope defined by the two families of alternating inequalities and $\mathbf{y} \in [0,1]^n$. From Proposition 4 and Theorem 2, $proj_y(P_Z(\boldsymbol{\alpha},\boldsymbol{\gamma})) \subseteq P$. To show the converse, suppose that $\mathbf{y^*} \in P \subseteq [0,1]^n$. We will show that $\mathbf{y^*} \in proj_y(P_Z(\boldsymbol{\alpha},\boldsymbol{\gamma}))$. Calculate F and G with respect to $\mathbf{y}^*_0$. Set $v_t = F(t)$ and $u_t = G(t)$ for all $t \in [-1,n]$. The initial values and Lemma 1 imply that $(\mathbf{u},\mathbf{v})$ satisfy (46)–(49), (51) and (53)–(55). As $\mathbf{y^*} \in P$, it follows that all the alternating inequalities are satisfied and thus $v_{s_B(\boldsymbol{\alpha},t)-1} - u_t = F(s_B(\boldsymbol{\alpha},t)-1) - G(t) \geq \min_{S \subseteq [s_B(\boldsymbol{\alpha},t),t]} Odd(S,\mathbf{y^*}) \geq 0$ satisfying (50). Similarly the fact that $\max_{S \subseteq [s_B(\boldsymbol{\gamma},t),t]} Odd(S,\mathbf{y^*}) \leq 1$ implies that (52) is satisfied. So $(\mathbf{u},\mathbf{v}) \in Q_{UV}(\boldsymbol{\alpha},\boldsymbol{\gamma})$ and the claim follows. □

Corollary 2 *There is a linear time separation algorithm for $conv(Y(\boldsymbol{\alpha},\boldsymbol{\gamma}))$. Given a point $\mathbf{y^*}$, it suffices to calculate $\mathbf{F}$ and $\mathbf{G}$, and verify if the point $(\mathbf{u},\mathbf{v}) = (\mathbf{G},\mathbf{F})$ lies in $Q_{UV}(\boldsymbol{\alpha},\boldsymbol{\gamma})$.*

When $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are constant over time, the validity of the alternating inequalities for $Y(\boldsymbol{\alpha})$ and $Y(\boldsymbol{\gamma})$, Theorem 3 and a linear time separation algorithm for the alternating inequalities are due to Lee et al. [8]. The approach presented above is more general and arguably simpler.

When there are only upper bounds on the lengths of the on- and off-intervals, the convex hull of $Y(\boldsymbol{\beta},\boldsymbol{\delta})$ is easily described.

Proposition 5 *$conv(Y(\boldsymbol{\beta},\boldsymbol{\delta}))$ is given by:*

$$\sum_{u=t}^{t+\beta_t} y_u \leq \beta_t \qquad t \in [0,n] : t+\beta_t \leq n; \tag{57}$$

$$\sum_{j=t}^{t+\delta_t} y_u \geq 1 \qquad t \in [0,n] : t+\delta_t \leq n; \tag{58}$$

$$y \in [0,1]^n. \tag{59}$$

Proof Every $\mathbf{y} \in conv(Y(\boldsymbol{\beta},\boldsymbol{\delta}))$ satisfies (57)–(59). Conversely every state vector $\mathbf{y} \in \{0,1\}^n$ which violates (57) or (58) contains an on-interval including $[t, t+\beta_t]$ or an off-interval including $[t, t+\delta_t]$ for some $t \in [0,n]$. So (57)–(59) is a valid formulation for $Y(\boldsymbol{\beta},\boldsymbol{\delta})$. Its constraint matrix has the consecutive 1's property. Thus it is totally unimodular. As the right-hand side vector is integer, the claim follows. □

Theorem 3 and Proposition 5 imply

$$conv(Y(\boldsymbol{\alpha},\boldsymbol{\gamma})) = conv(Y(\boldsymbol{\alpha})) \cap conv(Y(\boldsymbol{\gamma})) \quad \text{and} \quad conv(Y(\boldsymbol{\beta},\boldsymbol{\delta})) = conv(Y(\boldsymbol{\beta})) \cap conv(Y(\boldsymbol{\delta})),$$

i.e., combining the two lower bound types, or the two upper bound types, does not give rise to new facets when projecting onto the $\mathbf{y}$-subspace. This is no longer true, however, when mixing (nontrivial) lower and upper bounds. This is demonstrated briefly below for $Y(\boldsymbol{\alpha}, \boldsymbol{\beta})$ by an example and a description of two of the families of new facets that are encountered in the simple case where the bounds are constant over time.

In the rest of Sect. 3.2 we now assume that $\alpha_t = \overline{\alpha}$ and $\beta_t = \overline{\beta}$ for all $t \in [0, n]$ are constant over time (except for the terminal conditions, whereby $\alpha_t = \min\{\overline{\alpha}, n+1-t\}$), and we let $Y(\overline{\alpha}, \overline{\beta})$ denote $Y(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in this case.

Example 2 For an instance of $Y(\overline{\alpha}, \overline{\beta})$ with $\overline{\alpha} = 2$ and $\overline{\beta} = 3$ and $n \geq 8$, one obtains facet-defining inequalities of the form:

$$\begin{aligned} -y_t + y_{t+1} + y_{t+4} - y_{t+5} &\leq 1 \\ y_t + y_{t+1} + y_{t+3} - y_{t+4} &\leq 2 \\ -y_t + y_{t+1} + y_{t+3} + y_{t+4} &\leq 2 \\ y_t + y_{t+1} + y_{t+3} + y_{t+5} + y_{t+6} &\leq 4 \end{aligned}$$

as well as alternating inequalities and on-interval upper bound inequalities (57).

As a function of $\overline{\alpha}$ and $\overline{\beta}$, there are several families of new facet-defining inequalities. To give an idea of two general classes, we show classes that include the first and fourth inequalities in the Example.

Proposition 6 *If $\overline{\alpha} < \overline{\beta} < 2\overline{\alpha}$, the inequalities*

$$-y_{t-1} + y_t + y_{t+\overline{\beta}} - y_{t+\overline{\beta}+1} \leq 1 \qquad t \in [0, n - \overline{\beta} - 1] \tag{60}$$

and

$$\sum_{u=t}^{t+\overline{\alpha}-1} y_u + \sum_{u=t+\overline{\alpha}+1}^{t+\overline{\beta}} y_u + \sum_{u=t+\overline{\beta}+2}^{t+\overline{\alpha}+\overline{\beta}+1} y_u \leq \overline{\alpha} + \overline{\beta} - 1 \quad t \in [0, n - \overline{\alpha} - \overline{\beta} - 1] \tag{61}$$

are valid (and facet-defining) for $Y(\overline{\alpha}, \overline{\beta})$.

Proof Consider the first inequality (60). We show that any point $\mathbf{y}$ that violates the inequality is an infeasible state vector. By contradiction suppose that $\mathbf{y} \in Y(\overline{\alpha}, \overline{\beta})$ with $y_t = y_{t+\overline{\beta}} = 1$ and $y_{t-1} = y_{t+\overline{\beta}+1} = 0$. As the interval $[t, t+\overline{\beta}]$ contains more than $\overline{\beta}$ periods, $y_{t+j} = 0$ for some $j \in [1, \overline{\beta}-1]$. Suppose that $j \leq (\overline{\beta}-1)/2 \leq \overline{\alpha}-1$. Now, from the definition of $\overline{\alpha}$, $y_{t-1} = 0$ and $y_t = 1$ imply that $y_u = 1$ for all $u \in [t, t+\overline{\alpha}-1]$ which contradicts $y_{t+j} = 0$. The argument when $j > (\overline{\beta}-1)/2$ is similar.

Consider now the second inequality (61). Suppose that $\mathbf{y} \in Y(\overline{\alpha}, \overline{\beta})$ violates the inequality. So $y_u = 1$ for $u \in [t, t+\overline{\alpha}-1] \cup [t+\overline{\alpha}+1, t+\overline{\beta}] \cup [t+\overline{\beta}+2, t+\overline{\alpha}+\overline{\beta}+1]$. As the interval $[t, t+\overline{\beta}]$ contains more than $\overline{\beta}$ periods, one must have $y_{t+\overline{\alpha}} = 0$. Similarly $y_{t+\overline{\beta}+1} = 0$. Now, from the definition of $\overline{\alpha}$, $y_{t+\overline{\alpha}} = 0$ and $y_{t+\overline{\alpha}+1} = 1$ imply that $y_u = 1$ for all $u \in [t+\overline{\alpha}+1, t+2\overline{\alpha}]$. As $2\overline{\alpha} \geq \overline{\beta}+1$, this contradicts $y_{t+\overline{\beta}+1} = 0$.

It is a straightforward exercise to show that both inequality classes are facet-defining when t and $n-t$ are sufficiently large. □

Remark 4 Exchanging the roles of the on and off machine states, the set $Y(\boldsymbol{\gamma}, \boldsymbol{\delta})$ is isomorphic to $Y(\boldsymbol{\alpha}, \boldsymbol{\beta})$ under the change of variables $\mathbf{y} \to \mathbf{1} - \mathbf{y}$ and the change of parameters $\boldsymbol{\alpha} \to \boldsymbol{\gamma}$, $\boldsymbol{\beta} \to \boldsymbol{\delta}$ and $y_0 \to 1 - y_0$. Thus all results about $Y(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and its convex hull translate directly into equivalent results about $Y(\boldsymbol{\gamma}, \boldsymbol{\delta})$ and $conv(Y(\boldsymbol{\gamma}, \boldsymbol{\delta}))$.

Finding a complete description of $conv(Y(\overline{\alpha}, \overline{\beta}))$, and more generally of $conv(Y(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ and of other models combining bound types in the **y**-subspace, is an open problem.

4 Interval-dependent start-ups

Now we consider the problem in which the start-up costs depend on the number of periods during which the machine has been off as defined in Sect. 2.2, except that we now assume that the thresholds are time-invariant, i.e., $P_t = P$ and $\theta_t^p = \theta^p$ for all t and all p. Thus we let $\boldsymbol{\theta} = (\theta^p)_{p \in [0,P]}$ denote the threshold vector and $H(\boldsymbol{\theta})$ the corresponding set of feasible binary $(\mathbf{y}, \mathbf{z_1}, \ldots, \mathbf{z_n})$ vectors. Note that the thresholds imply time-invariant bounds $\overline{\gamma} = \theta^0 + 1$ and $\overline{\delta} = \theta^P$ on the off-intervals. On the other hand we assume no bounds on the on-intervals, in other words, we may assume time-invariant $\overline{\alpha} = 1$ and $\overline{\beta} = n + 1$.

We suppose without loss of generality that $y_0 = 1$ and that we are in a soft horizon context. The case in which $y_0 = y_{-1} = \cdots = y_{-\tau+1} = 0$, $y_{-\tau} = 1$ for some $\tau > 0$ can be treated by adding τ periods at the beginning of the n-period horizon and then setting $y_0 = 1$ and $y_1 = \cdots = y_\tau = 0$, as well as $z_t^p = 0$ for all $p \in [1, P]$ and all $t \in [1, \tau]$, in the augmented problem. In either case we may assume that the thresholds θ^p, and thus also the bounds $\overline{\gamma} = \theta_0 + 1$ and $\overline{\delta} = \theta^P$, are constant over the whole planning horizon.

This Section is devoted to proving the following:

Theorem 4 *$conv(H(\boldsymbol{\theta}))$ is described by the polytope*

$$y_0 = 1; \tag{62}$$

$$y_t + \sum_{\substack{p : \theta^p = \theta^{p-1}+1 \\ and\ t+\theta^p \le n}} z_{t+\theta^p}^p \le y_{t-1} + \sum_{p \in [1,P]} z_t^p \qquad t \in [1, n]; \tag{63}$$

$$\sum_{p \in [1,P]} z_t^p \le y_t \qquad t \in [1, n]; \tag{64}$$

$$y_t + \sum_{p \in [1,P]} \sum_{u=t+1}^{\min\{n,\, t+\theta^{p-1}+1\}} z_u^p \le 1 \qquad t \in [1, n]; \tag{65}$$

$$y_k + \sum_{p \in [1,P]} \sum_{u=k+1}^{k+\theta^p} z_u^p \ge 1 \qquad k \in [1,\, n - \theta^P]; \tag{66}$$

$$y_t + \sum_{p \in [1,P]} \sum_{u=\max\{t+1,\, k+\theta^p+1\}}^{\min\{n,\, t+\theta^{p-1}+1\}} z_u^p \le y_k + \sum_{p \in [1,P]} \sum_{u=k+1}^{\min\{k+\theta^p,\, t\}} z_u^p \qquad k \in [n - \theta^P + 1,\, t-2] \text{ and } t \in [n - \theta^P + 3,\, n]; \tag{67}$$

$$\mathbf{y} \in \mathbb{R}_+^n, \quad \mathbf{z_t} \in \mathbb{R}_+^P \qquad t \in [1, n]. \tag{68}$$

Notice that the first summation in (63) only arises for those instances in which $\theta^p = \theta^{p-1} + 1$, i.e., where some thresholds are immediately consecutive.

Roughly speaking constraints (63) generalize (36); (64) generalize constraints (37) (since $s_B(\boldsymbol{\alpha}, t) = t$); (65) generalize (39) and (66) generalize (40). More precisely, when $P = 1$, and thus $\overline{\gamma} = \theta^0 + 1$ and $\overline{\delta} = \theta^1$, constraints (63) reduce to the constraints (36) except for the very special case where $\theta^1 = \theta^0 + 1$ in which the length of every off-interval must be exactly θ^1; for the latter case, the resulting inequality $y_t + z_{t+\theta^1} \le y_{t-1} + z_t$ is the sum of the inequalities (39) for period $t + \theta^1$ and (40) for period $t + \theta^1 - 1$ (since $s_B(\boldsymbol{\gamma}, j) = s_B(\boldsymbol{\delta}, j) = j - \theta^1 + 1$ for all $j \ge \theta^1$). Constraints (64) reduce to (37) (as observed above); (65) for $t \in [1, n - \bar{\gamma}]$ reduce to (39) for $t \in [1 + \bar{\gamma}, n]$ (the inequalities (65) for $t \in [n - \bar{\gamma} + 1, n]$ are redundant); (66) for $t \in [1, n - \bar{\delta}]$ reduce to (40) for $t \in [1 + \bar{\delta}, n]$ and finally (67) takes the form $y_t \le y_k + \sum_{u=k+1}^{t} z_u^1$ (since $k > n - \theta^1$) and is the sum of inequalities (36).

Constraint (67) captures relationships between pairs (y_k, y_t) of states with $k < t$ and certain switch-on decisions z_u^p with $u > k$. We will prove its validity, and that of equivalent forms, in Proposition 7 and Corollary 3 below. Meanwhile, the next Lemma lists some properties of the inequality in (67) which justify the range restrictions on k and n in (67).

Lemma 2 *(i) When $k = t - 1$ the inequality in (67) reduces to that in (63); when $k = t$ it is trivially satisfied; when $k > t$ it simplifies to $y_t \le y_k$ which is not valid in general.*

(ii) The inequality in (67) is equivalent to

$$y_t + \sum_{p \in [1,P]} \sum_{u=k+\theta^p+1}^{\min\{n,\, t+\theta^{p-1}+1\}} z_u^p \;\le\; y_k + \sum_{p \in [1,P]} \sum_{u=k+1}^{t} z_u^p \tag{69}$$

and also to

$$y_t + \sum_{p \in [1,P(k)]} \sum_{u=t+1}^{\min\{n,\, t+\theta^{p-1}+1\}} z_u^p \le y_k + \sum_{p \in [1,P(k)]} \sum_{u=k+1}^{k+\theta^p} z_u^p + \sum_{p \in [P(k)+1,P]} \sum_{u=k+1}^{t} z_u^p. \tag{70}$$

where $P(k) = \max\{p : k + \theta^p \le n\}$.

(iii) If $k \le n - \theta^P$ then the inequality in (67) is implied by constraints (65) and (66).

Proof (i) Since $\theta^p \geq 1$ for all $p \in [1, P]$, the double summation on the left hand side of (67) simplifies to that of (63) and that on the right hand side to $\sum_{p\in[1,P]} z_t^p$ when $k = t-1$. These double summations vanish (because u has empty ranges) when $k \geq t$.
(ii) For a given p, if $k + \theta^p \geq t$, the corresponding terms in (67) are identical to those of (69). Else if $k + \theta^p < t$, it suffices to add $\sum_{u=k+\theta^p+1}^{t} z_u^p$ to the terms to obtain those of (69).

For given $p \leq P(k)$, if $k + \theta^p \leq t$, the corresponding terms in (67) are identical to those of (70). Else if $k + \theta^p > t$, it suffices to add $\sum_{u=t+1}^{k+\theta^p} z_u^p$ to the terms to obtain those of (70). For $p > P(k)$, the terms are identical.
(iii) If $k \leq n - \theta^P$, then $P(k) = P$ and inequality (70) is the difference of inequalities in (65) and (66). □

To prove Theorem 4 we show how the set $conv(H(\boldsymbol{\theta}))$ can be viewed as the solution set of a network flow (shortest path) model and then use Hoffman's circulation theorem to obtain $conv(H(\boldsymbol{\theta}))$ as the projection of this network flow polytope onto the subspace of the $(\mathbf{y}, \mathbf{z_1}, \ldots, \mathbf{z_n})$ variables. This network flow model is the path formulation of a variant of the Dynamic Programming algorithm of Sect. 2.3 which explicitly uses the state variables $\mathbf{y}$. As will become clear, this is made possible by the absence of (nontrivial) bounds on the length of the on-intervals.

The digraph $D = (V, A)$ has nodes t for all $t = 1, \ldots, n+1$, t' for all $t = 0, \ldots, n$, and $t \cdot p$ for all $p = 1, \ldots, P$ and $t = 1 \ldots, n$. The arcs are of the following types:

1. Arcs (t, t') for $t = 1, \ldots, n$. Flow on this arc indicates whether the machine is on in period t, i.e., is equal to y_t.
2. Arcs $(t', t+1)$ for $t = 1, \ldots, n$. This arc is used if the machine stays on from period t to $t+1$.
3. Start-up Arcs $(t \cdot p, t)$, for $t = 1, \ldots, n$ and $p = 1, \ldots, P$, used if a type-p start-up occurs in period t. Its flow is thus z_t^p.
4. Arcs $(t', (t+k) \ldots p)$ with $k \in [\theta^{p-1} + 2, \theta^p + 1]$, for $t = 1, \ldots, n$ and $p = 1, \ldots, P$. Such an arc is used if a switch-off in $t+1$ is followed by a type-p start-up in period $t+k$.
5. Arcs $(t', n+1)$ for $t \geq n - \overline{\delta}$, used if there is a switch-off in $t+1$ and the machine then remains off.
6. Initial Arcs $(0', t \cdot p)$ for $t \in [\theta^{p-1} + 2, \theta^p + 1]$. Terminal Arc $(0', n+1)$ if $\overline{\delta} \geq n$.
7. Return Arc $(n+1, 0')$, introduced to formulate the shortest $(0'$-$(n+1))$-path problem as a circulation problem. This arc will carry exactly one unit of flow.

Example 3 An instance of the digraph with $n = 6$, $P = 2$, $(\theta^0, \theta^1, \theta^2) = (0, 2, 4)$ (hence $\overline{\gamma} = 1$ and $\overline{\delta} = 4$) (and $y_0 = 1$) is shown in Fig. 2.

As in Sect. 2.3 the corresponding flow polyhedron (or path polytope) $P(D)$, described by the flow variables $\phi \in \mathbb{R}_+^A$ and in which the Return Arc $(n+1, 0')$ carries exactly one unit of flow, has integral extreme points corresponding to the paths from node $0'$ to node $n+1$. Since the flow $\phi_a = y_t$ on the type-1 arcs $a = (t, t')$ and z_t^p on the type-3 (startup) arcs $(t \cdot p, t)$, the polytope $H(\boldsymbol{\theta})$ is the projection of this flow polyhedron onto the coordinate subspace of the $(\mathbf{y}, \mathbf{z_1} \ldots, \mathbf{z_n})$ variables. That is,

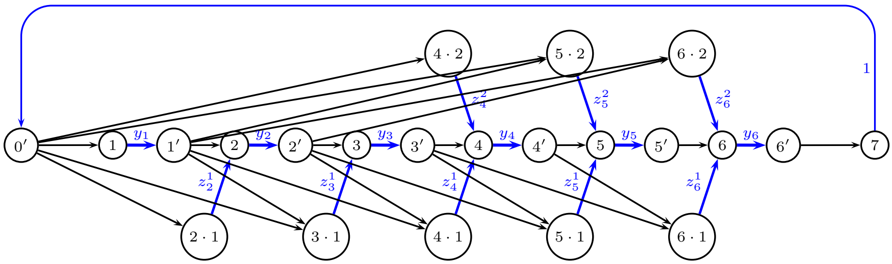

Fig. 2 Digraph for an instance with $n = 6$, $P = 2$, $\theta = (0, 2, 4)$ and $y_0 = 1$. Nodes $t \cdot p$ with $t < \theta^{p-1}+1$, their adjacent arcs, as well as type-5 arcs $(t', 7)$ with $t = 2, \dots, 5$ not drawn. Arcs a with flow value $l_a = u_a$ shown in *blue* (color figure online)

$(\mathbf{y}, \mathbf{z_1} \dots, \mathbf{z_n}) \in H(\boldsymbol{\theta})$ if and only if there exists a flow $\phi \in P(D)$ with components equal to $\mathbf{y}$ on the type-1 arcs and to $(\mathbf{z_1} \dots, \mathbf{z_n})$ on the type-3 arcs. These requirements can be enforced by adding lower and upper bounds $l_{t,t'} = u_{t,t'} = y_t$ on the flow on the type-1 arcs (t, t'), and $l_{t\cdot p,t} = u_{t\cdot p,t} = z_t^p$ on the flow on the type-3 arcs $(t \cdot p, t)$, as well as $l_{n+1,0'} = u_{n+1,0'} = 1$ on the Return Arc; the flow bounds on all other arcs a (types 2 and 4–6) are $l_a = 0$ and $u_a = +\infty$. This allows us to convert the decision question, whether $(\mathbf{y}, \mathbf{z_1} \dots, \mathbf{z_n}) \in H(\boldsymbol{\theta})$, into that of the existence of a feasible circulation satisfying all these flow bounds.

Given $X \subset V$, let $\overline{X} = V \setminus X$,

$$l(X, \overline{X}) = \sum_{(i,j)\in A \,:\, i\in X,\ j\in\overline{X}} l_{ij} \quad \text{and} \quad u(\overline{X}, X) = \sum_{(i,j)\in A \,:\, i\in\overline{X},\ j\in X} u_{ij}.$$

Observation 3 *To describe conv$(H(\boldsymbol{\theta}))$, it is necessary and sufficient to consider valid inequalities of the form*

$$l(X, \overline{X}) - u(\overline{X}, X) \le 0 \tag{71}$$

where $u(\overline{X}, X) < +\infty$.

This follows from Hoffman's circulation theorem, see in particular Martens et al. [10]. (Indeed the type-1 arcs (t, t') may be viewed as the result of splitting single nodes in a slightly simpler and more compact network so as to reveal, as in [10], the *node flows* y_t through these nodes.) We shall call the ordered bipartition $(X, \overline{X})$ of V a *Hoffman cut* and inequalities (71) *cut-inequalities*. Note that such inequalities in the $(\mathbf{y}, \mathbf{z_1}, \dots, \mathbf{z_n})$ variables have coefficients 0, +1 or −1.

Using this Observation, we first show the validity of the inequalities (63)–(68).

Proposition 7 *The inequalities* (63)–(68) *are valid for* $H(\boldsymbol{\theta})$.

Proof In each case we specify the set X that is used in Observation 3 to obtain the inequality as a cut inequality. Note that if an inequality, such as (63)–(68), is written in the form

$$\mathbf{y}(L_y) + \mathbf{z}(L_z) - \mathbf{y}(U_y) - \mathbf{z}(U_z) + b \le 0$$

with $L_z \cap U_z = \emptyset$ and is a cut inequality, then the corresponding Hoffman cut $(X, \overline{X})$ must have $u \cdot p \in X$ and $u \in \overline{X}$ for all $u \cdot p \in L_z$, and $u \in X$ and $u \cdot p \in \overline{X}$ for all $u \cdot p \in U_z$.

Taking $X = \{(t-1)', t\} \cup \{(t+\theta^p) \cdot p : \theta^p = \theta^{p-1} + 1 \text{ and } t + \theta^p \le n\}$ gives the valid inequality (63).
Taking $X = V \setminus \{t\}$ gives the valid inequality (64).
Taking $X = [1, t] \cup [0', (t-1)'] \cup \bigcup_{p=1}^{P} [1 \cdot p,\ \min\{n,\ t + \theta^{p-1} + 1\} \cdot p]$ gives the valid inequality (65).
Taking $X = [k+1, n+1] \cup [k', n'] \cup \bigcup_p [(k + \theta^p + 1) \cdot p, n \cdot p]$ gives the valid inequality (66).
Taking $X = [k+1, t] \cup [k', (t-1)'] \cup \bigcup_p [(k+\theta^p+1) \cdot p,\ \min\{n,\ t+\theta^{p-1}+1\} \cdot p]$ gives the valid inequality (67).
Finally, for (68) taking $X = \{t'\}$ gives $-y_t \le 0$, and taking $\overline{X} = \{t \cdot p\}$ gives $-z_t^p \le 0$. □

Corollary 3 *The inequality in* (67)*, as well as the equivalent inequalities* (69) *and* (70) *are valid for $H(\boldsymbol{\theta})$ for all k and t such that $1 \le k < t \le n$.*

Proof If $n - \theta^P < k < t - 1$ then (67) is valid by Proposition 7. Else, by Lemma 2 (i) it is valid if $k = t - 1$, and by Lemma 2 (iii) inequality (70) is implied by (65) and (66) if $k \le n - \theta^P$. □

To prove Theorem 4 we shall consider all possible Hoffman cuts $(X, \overline{X})$. A series of Observations and Propositions will allow us to restrict attention to a subfamily of these Hoffman cuts that suffices to determine $conv(H(\boldsymbol{\theta}))$. The first step (Observation 4) is to use the finiteness of the minimum cut to deduce that certain nodes must be assigned to $\overline{X}$ if one of their predecessors is in $\overline{X}$. Then in Proposition 8 we show that, unless all remaining unassigned nodes are assigned to X, the corresponding inequality is the sum of inequalities from (63)–(68) and therefore redundant. In Proposition 9 we show that certain other assignments also lead to inequalities that are the sum of other valid inequalities. Finally we consider the four sets of assignments remaining characterized by the different possible assignment of nodes $0'$ and $n+1$ to X or $\overline{X}$. In each case we show that the resulting inequalities are a nonnegative combination of the inequalities (63)–(68).

First, we only need to consider cut inequalities (71) with $u(\overline{X}, X)$ finite.

Observation 4 *Finiteness of $u(\overline{X}, X)$ implies:*

(a) if $t' \in \overline{X}$, then $t + 1 \in \overline{X}$ (type-2 arcs);
(b) if $t' \in \overline{X}$, then $(t+k) \cdot p \in \overline{X}$ for all $k \in [\theta^{p-1} + 2, \theta^p + 1]$ (type-4 arcs);
(c) if $t' \in \overline{X}$, then $n + 1 \in \overline{X}$ if $t \ge n - \overline{\delta}$ (type-5 arcs);

(d) *if* $0' \in \overline{X}$ *then* $t \cdot p \in \overline{X}$ *for all* $t \in [\theta^{p-1}+2, \theta^p+1]$, *and if* $0' \in \overline{X}$ *and* $\overline{\delta} > n$ *then* $n+1 \in \overline{X}$ *(type-6 arcs).*

Our approach is then to consider all possible X-*assignments* of the "main path" $PA = \{0', 1, 1', \ldots, n, n', n+1\}$, i.e. assignments of its nodes to X or $\overline{X}$. The following Proposition shows that facet-defining inequalities (71) are uniquely defined by the X-assignment of the main path.

Proposition 8 *Consider a cut inequality* (71) *other than a nonnegativity inequality* $-z_t^p \leq 0$. *If any node* $t \cdot p$, *other than those assigned to* $\overline{X}$ *in cases (b) and (d) of Observation 4, is assigned to* $\overline{X}$, *then the inequality is implied by a cut inequality with a larger set* X *and* $z_t^p \geq 0$.

Proof The contribution $l(X, \overline{X}) - u(\overline{X}, X)$ to the violation of (71) due to the arcs incident to node $t \cdot p \in \overline{X}$ is $-z_t^p$ if $t \in X$, and 0 if $t \in \overline{X}$. Since the inequality is not of the form $-z_t^p \leq 0$ we have $\overline{X} \neq \{t \cdot p\}$. Furthermore, since none of cases (b) and (d) of Observation 4 applies, all the predecessors u' of $t \cdot p$ are in X. We may therefore move $t \cdot p$ to X and obtain another Hoffman cut. There, the contribution to the violation of the arcs incident to node $t \cdot p$ becomes 0 if $t \in X$, and z_t^p if $t \in \overline{X}$. The given inequality is therefore the sum of the inequality (71) with $t \cdot p$ moved to X and the inequality $-z_t^p \leq 0$. □

The next Proposition restricts the X-assignments of the main path PA that need to be considered:

Proposition 9 *A cut inequality* (71), *other than* (64), *in which* $(t-1)'$ *and* t *are on opposite sides of the cut, i.e., assigned differently to* X *and* $\overline{X}$, *is implied by a cut inequality with a larger set* X *and inequality* (64) $\sum_p z_t^p \leq y_t$.

Proof Consider a Hoffman cut $(X, \overline{X})$ in which $(t-1)'$ and t are on opposite sides. The case in which $(t-1)' \in \overline{X}$ and $t \in X$ is excluded by Observation 4(a). Thus assume that $(t-1)' \in X$ and $t \in \overline{X}$. Since the inequality is not of the form (64) we must have the strict inclusion $\overline{X} \subset V \setminus \{t\}$ (as follows from the proof of Proposition 7), therefore we obtain another Hoffman cut if we move t to X. Here there are two possibilities.

(1) If $t' \in X$ then the contribution to the violation $l(X, \overline{X}) - u(\overline{X}, X)$ of the arcs incident to node t is $-y_t + \sum_{p\,:\,t\cdot p \in X} z_t^p$, while if t is moved to X then the contribution is $-\sum_{p\,:\,t\cdot p \in \overline{X}} z_t^p$.
(2) Else $t' \in \overline{X}$ and the contribution is $\sum_{p\,:\,t\cdot p \in X} z_t^p$, while if t is moved to X then the contribution is $y_t - \sum_{p\,:\,t\cdot p \in \overline{X}} z_t^p$.

The given inequality is therefore the sum of the inequality (71) with $t \cdot p$ moved to X and the inequality $\sum_p z_t^p - y_t \leq 0$. □

It follows that the remaining candidates to provide facet-defining inequalities are determined by the *flip* periods $\sigma(1) < \cdots < \sigma(K)$ (where $1 \leq \sigma(1)$ and $\sigma(K) \leq n$) in which nodes $\sigma(i)$ and $\sigma(i)'$ on the main path PA lie on opposite sides of the cut

$(X, \overline{X})$. Note that these flips determine the variables $y_{\sigma(1)}, \ldots, y_{\sigma(K)}$ which appear in the inequality and their coefficients, whose values alternate between $+1$ and -1 depending on whether $0' \in X$.

We now complete the proof of Theorem 4 by considering the four possible X-assignments of nodes $0'$ and $n+1$.

Case 1: $0', n+1 \in X$. By Proposition 9, $1 \in X$ and $n \in X$, so K is even, say $K = 2I$. The flips $\sigma(i)$ on the path PA lead to a partition $S_0, S_1, \ldots, S_I$ of the nodes $X \cap \{1, \ldots, n\}$ with $S_i = [\sigma(2i)+1, \sigma(2i+1)]$ where $\sigma(0) = 0$ and $\sigma(2I+1) = n$. The complement forms a partition $\widetilde{S}_0, \ldots, \widetilde{S}_{I-1}$ of $\widetilde{X} \cap \{1, \ldots, n\}$ with $\widetilde{S}_i = [\sigma(2i+1)+1, \sigma(2i+2)]$.

For each p, Proposition 8 then gives a collection of possibly overlapping sets $(\widetilde{T}_0^p, \ldots, \widetilde{T}_{I-1}^p)$ such that $\left(\cup_{j=0}^{I-1} \widetilde{T}_j^p\right) \cdot p = [1 \cdot p, n \cdot p] \cap \overline{X}$ where $\widetilde{T}_j^p = [\sigma(2j+1) + \theta^{p-1} + 2, \sigma(2j+2) + \theta^p] \cap [1, n]$.
Now the complement is a collection of disjoint sets $(T_0^p, \ldots, T_I^p)$ such that $\left(\cup_{j=0}^{I} T_j^p\right) \cdot p = [1 \cdot p, n \cdot p] \cap X$ with $T_0^p = [1, \sigma(1) + \theta^{p-1} + 1]$ and $T_j^p = [\sigma(2j) + \theta^p + 1, \sigma(2j+1) + \theta^{p-1} + 1] \cap [1, n]$ for $j \geq 1$. Some of the T_j^p may be empty. The resulting cut inequality is

$$\sum_{i=0}^{I-1} \left(y_{\sigma(2i+1)} + \sum_{p=1}^{P} \sum_{j=0}^{I} z^p \left(T_j^p \cap \widetilde{S}_i\right) \right) \leq \sum_{i=1}^{I} \left(y_{\sigma(2i)} + \sum_{p=1}^{P} \sum_{j=0}^{I-1} z^p \left(S_i \cap \widetilde{T}_j^p\right) \right).$$

Adding

$$\sum_{p=1}^{P} z^p \left(T_j^p \cap S_i\right)$$

to each side for all pairs (i, j) except $(0, 0)$ and (I, I), and using the following identities,

$$\begin{aligned} \widetilde{S}_0 \cup S_1 \cdots \cup \widetilde{S}_{I-1} \cup S_I &= [\sigma(1)+1, n]; \\ S_0 \cup \widetilde{S}_0 \cup S_1 \cdots \cap \widetilde{S}_{I-1} \cup S_I &= [1, n]; \\ T_I^p \cap \left(S_0 \cup \widetilde{S}_0 \cup S_1 \cdots \cap \widetilde{S}_{I-1}\right) &= \emptyset; \\ S_0 \cap \left(\widetilde{T}_0^p \cup T_1^p \cup \cdots \widetilde{T}_{I-1}^p \cup T_I^p\right) &= \emptyset; \\ T_0^p \cup \widetilde{T}_0^p \cup T_1^p \cup \cdots \widetilde{T}_{I-1}^p \cup T_I^p &= [1, n]; \qquad \text{and} \\ T_0^p \cup \widetilde{T}_0^p \cup T_1^p \cup \cdots \widetilde{T}_{I-1}^p &= [1, \sigma(2I) + \theta^p], \end{aligned}$$

this simplifies, after splitting the summation into three terms with $j = 0$, $j \in [1, I-1]$ and $j = I$ to:

$$\sum_{i=0}^{I-1} y_{\sigma(2i+1)} + \sum_{p=0}^{P} z^p\left(T_0^p \cap [\sigma(1)+1, n]\right) + \sum_{p=1}^{P}\sum_{j=1}^{I-1} z^p\left(T_j^p\right) + 0$$
$$\leq \sum_{i=1}^{I} y_{\sigma(2i)} + 0 + \sum_{p=1}^{P}\sum_{i=1}^{I-1} z^p(S_i) + \sum_{p=1}^{P} z^p(S_I \cap [1, \sigma(2I)+\theta^p]).$$

This is the sum of the inequalities:

$$\begin{aligned}
y_{\sigma(1)} + \sum_p z^p([\sigma(1)+1, \sigma(1)+\theta^{p-1}+1]) &\leq 1;\\
y_{\sigma(2i+1)} + \sum_p z^p\left(T_i^p\right) &\leq y_{\sigma(2i)} + \sum_p z^p(S_i) \text{for } i = 1, \ldots I-1; \text{ and}\\
-y_{\sigma(2I)} - \sum_p z^p([\sigma(2I)+1, \sigma(2I)+\theta^p]) &\leq -1,
\end{aligned}$$

namely inequality (65); $I-1$ inequalities (69) and inequality (66) (since $n+1 \in X$ implies $\sigma(2I)+\theta^P \leq n$).

Thus we have shown that every cut inequality with $0'$ and $n+1 \in X$, other than (64), is a nonnegative combination of inequalities in (63)–(67).

Case 2: $0' \in X, n+1 \in \overline{X}$. Here the flips occur at $\sigma(1), \ldots, \sigma(2I+1)$. The sets S_i, $\widetilde{S}_i$, $\widetilde{T}_i^p$ and T_i^p are unchanged, and sets $\widetilde{S}_I$ and $\widetilde{T}_I^p$ are added for all p. The resulting cut inequality is

$$\sum_{i=0}^{I}\left(y_{\sigma(2i+1)} + \sum_p\sum_{j=0}^{I} z^p\left(T_j^p \cap \widetilde{S}_i\right)\right) \leq \sum_{i=1}^{I}\left(y_{\sigma(2i)} + \sum_p\sum_{j=0}^{I} z^p\left(S_i \cap \widetilde{T}_j^p\right)\right) + 1$$

where the 1 appears because of the arc $(n+1, 0')$ with $n+1 \in \overline{X}$ and $0' \in X$. After adding the terms $\sum_p z^p\left(S_i \cap T_j^p\right)$ as before, the resulting inequality is the sum of

$$y_{\sigma(1)} + \sum_p z^p([\sigma(1)+1, \sigma(1)+\theta^{p-1}+1]) \leq 1$$

and the I inequalities

$$y_{\sigma(2i+1)} + \sum_p z^p(T_i^p) \leq y_{\sigma(2i)} + \sum_p z^p(S_i)$$

of the form (69).

Case 3: $0', n+1 \in \overline{X}$ can be treated as Case 2 by omitting sets S_0 and T_0^p.

Case 4: $0 \in \overline{X}, n+1 \in X$ is then similar.

5 Extensions

The results of Sect. 3 can be extended in different ways. One possibility is to include lower and/or upper bounds on the number of set-ups in a given interval. Bounds on $\sum_{t=\tau_1}^{\tau_2} z_t = v_{\tau_2} - v_{\tau_1 - 1}$ can be added to the Q_{UV} formulation without losing integrality. For $Y(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ with an upper bound on the number of setups, the constraint $\sum_{t=\tau_1}^{\tau_2} z_t \leq \Omega_{\tau_1,\tau_2}$ projects into another set of alternating inequalities $\max_{S \subseteq [\tau_1,\tau_2]} Odd(S, y) \leq \Omega_{\tau_1,\tau_2}$. These can be separated in linear time using the functions F and G and now the three sets of alternating inequalities give the convex hull. Another possibility is that the behaviour of the machine depends on the number of start-ups that have occurred. Thus $z_t^q = 1$ if the qth start-up is in period t. Here the Q_{UV} formulation can be generalized based on the binary variables v_t^q and u_t^q taking the value 1 if the qth start-up, respectively switch-off, occurs in or before t. Finally we note that one can generate inequalities for interval-dependent switch-offs, using a similar approach to that of Sect. 4.

References

1. Conforti, M., Cornuéjols, G., Zambelli, G.: Integer Programming. Springer, New York (2014)
2. Damcı-Kurt, P., Küçükyavuz, S., Rajan, D., Atamtürk, A.: A polyhedral study of ramping in unit commitment. Math. Program. Ser. A **158**, 175–206 (2016)
3. Frangioni, A., Gentile, C.: Solving nonlinear single-unit commitment problems with ramping constraints. Oper. Res. **54**(4), 767–775 (2006)
4. Frangioni, A., Gentile, C., Lacalandra, F.: Tighter approximated MILP formulations for unit commitment problems. IEEE Trans. Power Syst. **24**, 105–113 (2009)
5. Garver, L.L.: Power generation scheduling by integer programming–development of theory, power apparatus and systems. III Trans. Am. Inst. Electr. Eng. **81**(3), 730–734 (1963)
6. Hedman, K.W., O'Neill, R.P., Oren, S.S.: Analyzing valid inequalities of the generation unit commitment problem. In: Power Systems Conference and Exposition, 2009 PSCE '09. IEEE/PES, pp. 1–6 (2009)
7. Knueven, B., Ostrowski, J., Wang, J.: Generating cuts from the ramping polytope for the unit commitment problem Department of Industrial and Systems Engineering University of Tennessee(2016)
8. Lee, J., Leung, J., Margot, F.: Min-up/min-down polytopes. Discrete Optim. **1**, 77–85 (2004)
9. Malkin, P.: Minimum runtime and stoptime polyhedra. In: Working Paper, Core, Université catholique de Louvain (2003)
10. Martens, M., McCormick, S.T., Queyranne, M.: Separation, dimension, and facet algorithms for node flow polyhedra. Math. Program. Ser. B **24**, 317–348 (2010)
11. Morales-España, G., Gentile, C., Ramos, A.: Tight MIP formulations of the power-based unit commitment problem. OR Spectr. **37**(4), 1–22 (2015)
12. Morales-España, G., Latorre, J.M., Ramos, A.: Tight and compact MILP formulation for the thermal unit commitment problem. IEEE Trans. Power Syst. **28**, 4897–4908 (2013)
13. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. Wiley, New York (1988)
14. Ostrowski, J., Anjos, M.F., Vannelli, A.: Tight mixed integer linear programming formulations for the unit commitment problem. IEEE Trans. Power Syst. **1**(27), 39–46 (2012)
15. Pan, K., Guan, Y.: A Polyhedral Study of the Integrated Minimum-Up/Down Time and Ramping Polytope. http://www.optimization-online.org/DB_HTML/2015/08/5070.html (2015)
16. Pochet, Y., Wolsey, L.A.: Production Planning by Mixed Integer Programming. Springer, New York (2006)
17. Queyranne, M., Wolsey, L.A.: Modeling poset convex subsets. In: Cologne-Twente Workshop, Istanbul (2015)

18. Rajan, D., and Takriti, S.: Minimum up/down polytopes of the unit commitment problem with start-up costs. IBM Res. Rep. RC23628 (W0506-050), Thomas J. Watson Research Center, Yorktown Heights, NY (2005)
19. Viana, A., Pedroso, J.P.: A new MILP-based approach for unit commitment in power production planning. Electr. Power Energy Syst. **44**, 997–1005 (2013)
20. Wolsey, L.A.: Integer Programming. Wiley, New York (1998)
21. Wu, L.: A tighter piecewise linear approximation of quadratic cost curves for unit commitment problems. IEEE Trans. Power Syst. **26**, 2581–2583 (2011)