On Polyhedral Approaches to Scheduling Problems

Maurice Queyranne

Sauder School of Business at UBC

Based on joint work with many co-authors…

Scheduling Seminar

https://schedulingseminar.com/

June 22, 2022

Prelude

When formulating scheduling problems as mathematical optimization problems (MOPs)

- in particular, (mixed)-integer linear optimization problems

the primary design decision is which **variables** to use, e.g.,

- start/completion dates
- linear ordering variables
- traveling salesman variables
- assignment and positional date variables
- time-indexed variables

Constraints are then introduced to represent or approximate the feasible set (and the objective function)

- linear constraints give rise to **polyhedra**

The study of these scheduling polyhedra yields **structural** and **algorithmic** insights

Start (or Completion) **Date Variables**

- Decision variables are operation (or job) start times t_j
  - continuous variables only
- The combinatorial structure of (nonpreemptive) scheduling,
  - no two operations *i* and *j* on a same machine can overlap,

  may be modeled by **disjunctive constraints**
  - operation *i* is completed before operation *j* starts <u>or</u> *i* starts after *j* is completed:

$$t_j - t_i \geq d_{i,j} \quad \vee \quad t_i - t_j \geq d_{j,i}$$

Disjunctive Scheduling Constraints

$$t_j - t_i \geq d_{i,j} \quad \vee \quad t_i - t_j \geq d_{j,i}$$

- The **delay** $d_{i,j}$ may include operation i's processing time and a switchover cost: $d_{ij} = p_i + s_{ij}$

This allows to model

- release dates: $d_{0j} = r_j$ (with $t_0 = 0$)
- precedence constraints: $d_{j,i} = +\infty$, i.e.,

$$t_j - t_i \geq d_{i,j} \quad \text{if } i \rightarrow j$$

Mathematical Programming Study 24 (1985) 179–218
North-Holland

ON THE FACIAL STRUCTURE OF SCHEDULING POLYHEDRA

Egon BALAS

Carnegie-Mellon University, Schenley Park, Pittsburgh, PA 15213, USA

Received 28 November 1983
Revised manuscript received 28 January 1985

Dedicated to George B. Dantzig for his 70th birthday.

A well-known job shop scheduling problem can be formulated as follows. Given a graph G with node set N and with directed and undirected arcs, find an orientation of the undirected arcs that minimizes the length of a longest path in G. We treat the problem as a disjunctive program, without recourse to integer variables, and give a partial characterization of the scheduling polyhedron $P(N)$, i.e., the convex hull of feasible schedules. In particular, we derive all the facet inducing inequalities for the scheduling polyhedron $P(K)$ defined on some clique with node set K, and give a sufficient condition for such inequalities to also induce facets of $P(N)$. One of our results is that any inequality that induces a facet of $P(H)$ for some $H \subset K$, also induces a facet of $P(K)$. Another one is a characterization of adjacent facets in terms of the index sets of the nonzero coefficients of their defining inequalities. We also address the constraint identification problem, and give a procedure for finding an inequality that cuts off a given solution to a subset of the constraints.

Key words: Scheduling, Disjunctive Programming, Polyhedral Combinatorics, Facets of the Scheduling Polyhedron.

Job Shop Scheduling

Balas (1985) was motivated by job-shop scheduling:

- Each **job** consists of a sequence of nonpreemptive operations, each one on a specified machine
- **Makespan** objective: the time to complete all work

Formulated as finding a critical path in a “disjunctive” network with operations as nodes and directed **arcs** representing constraints on operation pairs:

- Precedence arcs (operations of the same job)
- Disjunctive arc pairs (no overlap when on the same machine)

A **selection** specifies a choice for each disjunctive pair

- Each selection leads to a critical path (project) network

Main Results in Balas (1995)

The set of all feasible start time vectors t is the union of (full-dimensional, unbounded) polyhedra

- one for each selection

The **scheduling polyhedron** is its closed convex hull

Main focus is on the **single machine** with release dates and no precedences (scheduling polyhedron on a **clique**)

- Extreme points, as “minimal” feasible schedules
- Extreme rays correspond to idle time insertion

and on its facets

- Facets with small support
- Lifting from subcliques
- Separation

Computational Experiments

Applegate & Cook (1991) implemented Balas's separation algorithm for small support cuts ($\leq$ 5 operations)

- Later generalized as Fenchel cuts (Boyd 1995), and "local cuts" (Applegate at al. 1998)

and compared, on 10x10 instances, with

- Combinatorial bounds (Carlier 1982)
- Large support "basic cuts" of Dyer and Wolsey (1990)

$\rightarrow$ Modest reductions in integrality gaps

(for makespan objective)

Toward Simpler Polyhedral Descriptions

The **Polytime Equivalence of Optimization and Separation** (GLS 1981) suggests that a family $(P_f)_{f \in F}$ of polyhedra may be easier to describe if the associated linear optimization problem, $\min\{ w\, x : x \in P_f \}$ is solvable in polytime for any objective weights w

- For machine scheduling, using the **completion time variables C_j** instead of the start times,
  - a simple affine change of variables,

  this leads to considering **minsum** objectives of minimizing a weighted sum of completion times:

$$\min \Sigma_j\, w_j\, C_j$$

Single Machine Scheduling

A classical problem: $1||\Sigma w_j C_j$: Given

- one machine, which can only process one job at a time
- n jobs $J_1,\ldots, J_n$ released at date 0, with
  - known **processing times** $p_j \geq 0$ and
  - known **weights** $w_j > 0$ of each job J_j

find a schedule of these n jobs to minimize $\Sigma w_j C_j$

Solved by **Smith's Ratio Rule**: process the jobs as soon as possible and in "increasing" (nondecreasing) order of the **ratios p_j / w_j**

- a.k.a.: WSPT rule, $c\mu$ rule (a Gittins index)

Simple Scheduling Polyhedra

Theorem (Wolsey 1985, Q 1993)

- For problem $1 || \Sigma_{j \in N} w_j C_j$ the convex hull of the completion time vectors C of all feasible schedules is the polyhedron

$$P_g = \{ C \in \mathbf{R}^n : \Sigma_{j \in A} p_j C_j \geq g(A) \quad \forall A \subseteq N \}$$

where $g(A) = ½ (\Sigma_{j \in A} p_j)^2 + ½ \Sigma_{j \in A} p_j^2$

- The convex hull of the **permutation schedules** C (without idle time and without preemption) is its facet

$$Q_g = P_g \cap \{ C \in \mathbf{R}^n : \Sigma_{j \in N} p_j C_j = g(N) \}$$

More on Simple Scheduling Polyhedra

- g is a **supermodular** set function:

$$g(A \cup B) + g(A \cap B) \geq g(A) + g(B) \quad \forall A, B \in N$$

- The scheduling polyhedron $\boldsymbol{P_g}$ is therefore (an affine image of) a **supermodular polyhedron**,
  and $\boldsymbol{Q_g}$ is its **base polytope**
- Smith's rule is (up to an affine transformation) a version of the **greedy algorithm** for supermodular (or submodular) polyhedra:

$$\text{OPT} := \min\{\textstyle\sum_{j \in N} w_j C_j \;:\; \sum_{j \in A} p_j C_j \geq g(A) \; \forall A \subseteq N\}$$

$$= \min\{\textstyle\sum_{j \in N} (w_j / p_j)\, x_j : \sum_{j \in A} x_j \geq g(A) \; \forall A \subseteq N\}$$

Precedence Constrained Single Machine Scheduling

- Problem $1|\text{prec}|\sum w_j C_j$ is **NP-hard**

Series inequalities: the simple precedence constraints

$C_k - C_j \geq p_k$ generalize, when job subsets $J \to K$

i.e., when $j \to k$ for all $j \in J$ and $k \in K$

to: $$p(J)\sum_{k\in K} p_k C_k - p(K)\sum_{j\in J} p_j C_j \geq h(J,K)$$

where $h(J,K) = p(J)\, g(K) + p(K)\, g(J) - p(K)\sum_{j\in J} p_j^2$

Q &Wang (1991): $(N, \to)$ is **series-parallel** iff for any positive p_j's, the scheduling polyhedron is defined by

- the *parallel* inequalities $\sum_{j\in A} p_j C_j \geq g(A)\ \forall A \subseteq N$
- and the series inequalities for all $J \to K$

Series and Parallel Inequalities

The **separation problem** for series inequalities is:

- easy when prec. is series-parallel (Schulz 1996a)
- NP-hard for general prec. (Kobayashi et al. 2012)
- but polytime separation for a larger class of inequalities from projection of an extended formulation using also linear ordering variables (Wolsey 1989)

Computational experiments

- random instances, heuristic separation for series ineq.
- relaxation gap < 1% (Q and Wang 1991, $n \leq 160$)
- branch and cut (Margot et al. 2003, $n \leq 120$)

Extensions and subsequent related work

- characterizing facets, N-sparse prec, 2D Gantt charts…

A Change of Perspective: Approximation Algorithms

- Are scheduling polyhedra useful to derive approximation algorithms for scheduling problems?

Schulz's Lemma (1996b): If vector C satisfies the parallel inequalities then **list scheduling** in order of the components of C gives a schedule C^L satisfying the **job-by-job bounds**: $C^L{}_j < 2\ C_j$ for each $j \in N$

$\Rightarrow \Sigma\ w_j\ C^{LP}{}_j > \frac{1}{2}$ OPT and $\Sigma\ w_j\ C^L{}_j < 2$ OPT for $1|\text{prec}|\ \Sigma\ w_j\ C_j$

- in fact, *every* feasible schedule consistent with the Sidney decomposition is 2-approximate (Chekuri and Motwani 1999, Margot et al 2003)

Many extensions and related work

- various scheduling environments, asymptotic optimality…

Release Dates & On-Line Scheduling

Each job J_j only becomes available for processing at its **release date** $r_j \geq 0$.

- In classical, "**off-line**" problems, the release dates r_j and all other problem data are known from the beginning.
- In **on-line** problems, one only learns the characteristics of a job J_j (including its release date r_j), or even its existence, at date r_j
- The off-line problem $1 | r_j | \Sigma\, w_j\, C_j$ is NP-hard.

Release Dates: WSPTA

WSPTA, a “natural” extension of the WSPT rule :
when the machine becomes available, process an ***Available*** *job with best WSPT ratio*

- WSPTA may produce schedules that are arbitrarily worse than the optimum (even with 2 jobs only !)

Asymptotic Optimality

An algorithm H is **asymptotically optimal** (relative to a given **instance class J**) if

$$\limsup_{|I|\to\infty} \frac{z^H(I)}{z^*(I)} = 1$$

where $|I|$ is the **size** of instance I

- here, the size $|I|$ is the number n of jobs

Why "Asymptotic Optimality"?

(1) Drawbacks of traditional performance measures (performance or approximation ratio, competitive ratio):

- conservative (worst case), hence "pessimistic"
- bad instances are often either:
  - very small, or
  - contrived, with extreme data values
- most results have limited practical value

(2) On large (uncontrived) instances, simple heuristics (e.g., based on Smith's ratio rule) may empirically outperform heuristics designed for better performance/competitive ratio

- e.g., Savelsbergh, Uma & Wein (2005)

(3) A well-accepted, desirable property in the study of stochastic processes and stochastic control

Asymptotic optimality of WSPTA

Theorem (Chou, Q. & Simchi-Levi, 2006):

The algorithm WSPTA is **asymptotically optimal** for every uniformly bounded instance class of problem $1|\ r_j$, *on-line*$|\ \Sigma\ w_j\, C_j$

The proof uses an LP relaxation with

- *a decomposition of the objective using "priority-work sets, and*
- *a "work delay lemma"*

Uniform Parallel Machines

This asymptotic optimality result extends to **uniform parallel machines** (machines with different speeds)

Theorem (Chou, Q. & Simchi-Levi, 2006):

The algorithm WSPTA is asymptotically optimal for every uniformly bounded instance class of $Qm \mid r_j$, *on-line*$\mid \Sigma\, w_j C_j$ with a fixed set of m uniform machines.

Stochastic On-Line Scheduling

Assume

- the processing times are **independent** random variables; and
- only the weight and **expected** processing time are known when a job appears (at its release date)

A scheduling policy is

- **adaptive** if it ***can*** use any information **available** at date t for making a decision at that date;
- **weakly non-anticipative** if it does not know the processing time of a job before the job is complete.

Stochastic On-Line Scheduling

Stochastic on-line scheduling on a single machine with release dates (Chou, Liu, Q. & Simchi-Levi, 2006)

Theorem: The WESPTA algorithm is **asymptotically optimal** relative to the class of all adaptive and weakly non-anticipative policies, for every uniformly bounded instance class of 1| r_j, p_j ~ stoch, *on-line*| $\Sigma\, w_j C_j$

The proof uses:

- *machine capacity constraints ("conservation laws")*
- *a Pmtn-WSPT algorithm applied to the deterministic instance with processing times* $\boldsymbol{E[p_j]}$
- *a Chernoff bound*
- *the notion of work delay for a given priority level*

Stochastic Shop Scheduling

Each job J_j consists of m operations $J_{j,1}, \ldots, J_{j,m}$

- operation $J_{j,i}$ must be processed on machine $\mathbf{M}_i$
- its processing time $p_{j,i}$ is a random variable

Let $p_j = (p_{j,1}, \ldots, p_{j,m})$ denote the resulting random vector

Assume these random vectors are independent

Stochastic Shop Scheduling (2)

Xia, Shantikumar & Glynn (2000): The processing times are **statistically exchangeable** across machines if, for every permutation α of $\{1, \ldots, m\}$, the random vector $(p_{j,\alpha(1)}, \ldots, p_{j,\alpha(m)})$ has the same distribution as random vector $p_j = (p_{j,1}, \ldots, p_{j,m})$

Theorem (Liu, Q. & Simchi-Levi, 2005)**:**

The WESPTA algorithm, extended to (non-permutation) **flowshop** and **open shop** problems, is asymptotically optimal, relative to the class of all adaptive and weakly non-anticipative policies, for every uniformly bounded instance class with independent job processing time vectors that are statistically exchangeable across machines.

Back to Deterministic Shop Scheduling: From Makespan to Min-Sum Objectives

Q and Sviridenko (2002): For shop scheduling problems with

- multiple **stages**, each consisting of m_h parallel machines
- jobs J_j consisting of operations O_{hj} each on a prescribed stage, with processing time p_{hj} and release dates r_{hj} , and with precedence constraints between operations
- **sets *S* of operations** with $C_S := \max\{C_{hj} : (h,j) \in S\}$
  - operation and job completion times, stage loads,…

and set weights w_S

Assume a **ρ-approximation** for the **makespan** (without release dates!), then using an LP relaxation in operations C_{hj}

- constraints for release dates, simple precedence constraints, stage capacity (parallel inequalities) and **job capacity** constraints

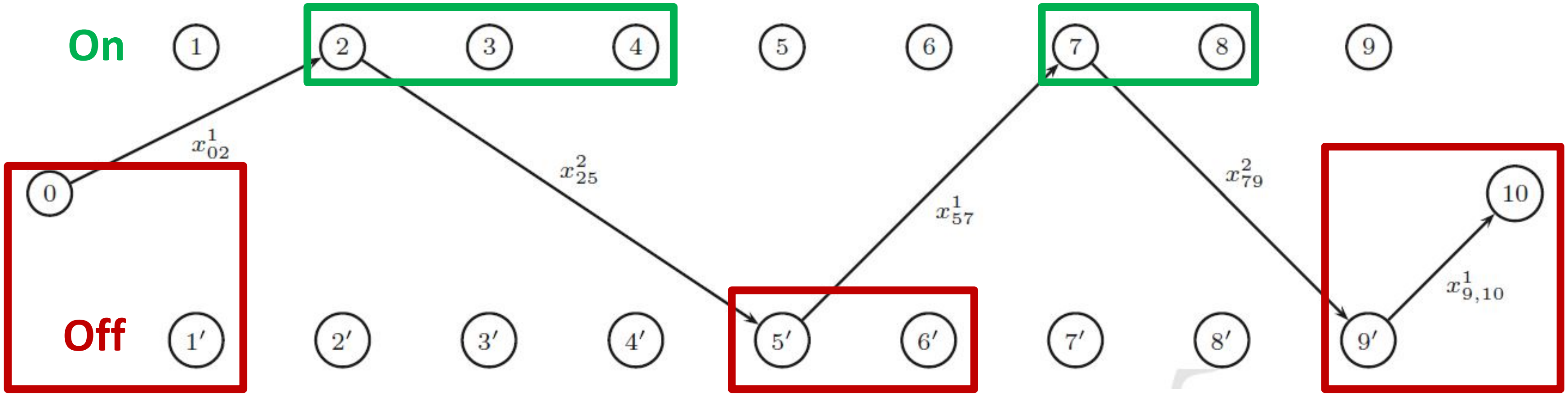
leads to a **2*e*ρ-approximation** for the **min-sum** obj. $\Sigma\, w_S\, C_S$

Time-Indexed Formulations

Single Unit Commitment Problem, Single Machine Production Planning

- Finite, discrete time horizon, $T = [1, \ldots, n]$
- Machine is in one of two states in each period $t \in T$:
  - $\boldsymbol{y_t} = 1$ if machine is On, 0 if it is Off
  - given initial state y_0 (and relevant earlier history)
- Actions, in each $t \in T$:
  - Switch on: $\boldsymbol{z_t} = 1$ if $y_{t-1} = 0$ and $y_t = 1$
  - Switch off: $\boldsymbol{w_t} = 1$ if $y_{t-1} = 1$ and $y_t = 0$
    - 3-bin (binary variables) model [Garver, 1963]
    - $z_t = (y_t - y_{t-1})^+, \quad w_t = (y_{t-1} - y_t)^+, \quad \boxed{z_t - w_t = y_t - y_{t-1}}$
    - Leads to 2-bin models, say, with **y** and **z** variables only
    - Also 1-bin model, with the **y** state variables only

General Dynamic Programming Approach



- General objective function (additive over intervals)
- General constraints (on individual intervals)

Shortest path formulation (tight extended formulation):

- $O(n^2)$ binary variables (flow $\mathbf{x}^1$ and $\mathbf{x}^2$, state $\mathbf{y}$, possibly switch on $\boldsymbol{z}$ and/or switch off $\boldsymbol{w}$), and $O(n)$ constraints

Question: Tight formulations with $O(n)$ variables (in either the 2-bin or 1-bin model)?

Bounded UP/Down Times

We are given bounds on interval lengths:

- lower and upper bounds $1 \le \alpha_t \le \beta_t$ on the length of an on-interval starting in period t
- lower and upper bounds $1 \le \gamma_t \le \delta_t$ on the length of an off-interval starting in period t
  - These (nonstationary) bounds may be adjusted to take into account initial and terminal conditions

A “Natural” MIP Formulation

$$z_t \geq y_t - y_{t-1} \qquad t \in [1, n]; \qquad (2)$$

$$z_t \leq y_t \qquad t \in [1, n]; \qquad (3)$$

$$z_t \leq 1 - y_{t-1} \qquad t \in [1, n]; \qquad (4)$$

$$z_t = y_t\,(1 - y_{t-1})$$

$$z_t \leq y_u \qquad u \in [t,\ t + \alpha_t - 1],\ \ t \in [0, n]; \qquad (5)$$

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

$$w_t \leq 1 - y_u \qquad u \in [t,\ t + \gamma_t - 1],\ \ t \in [0, n]; \qquad (7)$$

$$w_t \leq \sum_{u=t+1}^{t+\delta_t} y_u \qquad t : t \geq 0 \text{ and } t + \delta_t \leq n; \qquad (8)$$

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

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

A Tighter Formulation

- Replace the "forward looking" lower bound constraints

$$z_t \leq y_t \qquad t \in [1, n]; \qquad (3)$$

$$z_t \leq y_u \qquad u \in [t,\ t + \alpha_t - 1],\ \ t \in [0, n]; \qquad (5)$$

with the "backward looking" constraints

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

- Similarly, replace

$$z_t \leq 1 - y_{t-1} \qquad t \in [1, n]; \qquad (4)$$

$$w_t \leq 1 - y_u \qquad u \in [t,\ t + \gamma_t - 1],\ \ t \in [0, n]; \qquad (7)$$

with

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

A Simple Observation

Proposition

The resulting formulation (2), (11), (6), (12), (8), (9) with $\mathbf{y}, \mathbf{z}, \mathbf{w} \in [0,1]^n$ is valid and tighter than the "natural" formulation (2)-(9) with $\mathbf{y}, \mathbf{z}, \mathbf{w} \in [0,1]^n$

- Generalizes results in [Wolsey, 1998] and [Rajan & Takriti, 2005] to nonstationary lower bounds

Questions: When is such formulation **ideal**

- in the 3-bin space, or the 2-bin space of the ($\mathbf{y}$, $\mathbf{z}$) variables?
- in the 1-bin space of the state variables $\mathbf{y}$?

Subsets of Bound Types and Polytopes

Questions: When is such formulation **ideal**

- in the 3-bin space, or the 2-bin space of the $(\mathbf{y}, \mathbf{z})$ variables?
- in the 1-bin space of the state variables $\mathbf{y}$?

We generalize these questions to any subset $B \subseteq \{\alpha, \beta, \gamma, \delta\}$ of bound types in force: let

- $Z(B) = \{(\mathbf{y}, \mathbf{z}) \in \{0,1\}^{n+n}$ satisfying the bound constraints in $B\}$
- $Y(B) = \text{proj}_y Z(B)$

Remark: $Z(\emptyset) = \{(\mathbf{y}, \mathbf{z}) \in \{0,1\}^{n+n} : (2)\text{-}(4)\}$ and $Y(\emptyset) = \{0,1\}^n$

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

Weak Monotonicity and Earliest Interval Starts

The given bounds in B satisfy weak monotonicity if

$$t + \varepsilon_t \leq u + \varepsilon_u \quad \text{for} \quad 0 \leq t < u \leq n \quad \text{and} \quad \boldsymbol{\varepsilon} \in B$$

- By waiting one period one cannot be forced to switch on or off earlier

For every $\boldsymbol{\varepsilon} \in \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}\}$ and $t \in [1, n]$ let

$$s_B(\boldsymbol{\varepsilon}, \boldsymbol{t}) = \min\{ u \in [0, n] : u + \varepsilon_u > t \}$$

denote the earliest start period of an on- (if $\boldsymbol{\varepsilon} = \boldsymbol{\alpha}$ or $\boldsymbol{\beta}$) or (otherwise) off- interval that may end after t.

- If the lower bound type $\boldsymbol{\alpha} \notin B$ (resp., $\boldsymbol{\gamma} \notin B$) then let all $\alpha_t = 1$ (resp., all $\gamma_t = 1$) and $s_B(\boldsymbol{\alpha}, t) = t$ (resp., $s_B(\boldsymbol{\gamma}, t) = t$)
- If the upper bound type $\boldsymbol{\beta} \notin B$ (resp., $\boldsymbol{\delta} \notin B$) then let all $\beta_t = n$ (resp., all $\delta_t = n$) and $s_B(\boldsymbol{\beta}, t) = 0$ (resp., $s_B(\boldsymbol{\delta}, t) = 0$)

By weak monotonicity, for all B, $\boldsymbol{\varepsilon} \in B$ and $t \in [1, n]$

$$u \in [s_B(\boldsymbol{\varepsilon}, t), t] \quad \text{iff} \quad (u \in [0, n] \text{ and } u + \varepsilon_u > t)$$

Formulation for $Z(B)$

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

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

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

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

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

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

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

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

Ideal Formulation

Theorem: Under weak monotonicity, for every B the polytope $P_Z(B)$ defined by (35)-(42) is an ideal formulation of $Z(B)$

- i.e., $P_Z(B) = conv\ Z(B)$

- How do we prove this?

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., $u_t = \sum_{j=1}^{t} w_j$.

with initial values $v_0 = y_0$ and $u_0 = 0$

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

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

$$y_t = v_t - u_t \qquad t \in [1, n] \qquad (44)$$

Extended Formulation

$$u_{-1} = y_0 - 1; \qquad (46)$$

$$v_{-1} = u_0 = 0; \qquad (47)$$

$$v_0 = y_0; \qquad (48)$$

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

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

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

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

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

$$0 \leq v_t - u_t \leq 1 \qquad t \in [1, n]; \qquad (54)$$

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

Proposition: The polytope ***Q_{UV}(B)*** defined by (46)-(55) and the linking constraints (43)-(44) is an extended formulation for conv $Z(B)$.

Ideal Extended Formulation

Theorem: The polytopes $Q_{UV}(B)$ and $P_Z(B)$ are integral, and $P_Z(B) = conv\ Z(B)$.

- Generalizes results in [Malkin, 2003] and [Rajan & Takriti, 2005] to upper bounds and nonstationary data

Remark: in the $(\mathbf{y}, \mathbf{z})$ space the different bound types do not interact:

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

In the 1-Bin Space of the State Variables y?

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

- 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, $Even(S, y) \equiv y_{j_1} - y_{j_2} + \cdots - y_{j_k}$.
- $\text{Length}(S) \equiv j_k - j_1$.

Alternating inequality: $Odd(S, \mathbf{y}) \le (\text{or} \ge)\ \mu$ for some integer μ

The case of lower bounds only:

Theorem: conv $Y(\alpha, \gamma)$ is given by the alternating inequalities:

(i) $Odd(S, \mathbf{y}) \ge 0$ $\quad S \subseteq [s_B(\alpha, t), t]$ odd, and $t \in [0, n]$

(ii) $Odd(S, \mathbf{y}) \le 1$ $\quad S \subseteq [s_B(\gamma, t), t]$ odd, and $t \in [0, n]$

How do we prove this?

Let polytope $P = \{\ \mathbf{y} \in [0, 1]^n : \text{(i) and (ii)}\ \}$:

- (i) and (ii) are valid for $Y(\alpha, \gamma)$, hence $\text{proj}_y P_Z(\alpha, \gamma) \subseteq P$

Separating the Alternating Inequalities

To prove the converse inclusion, $P \subseteq \text{proj}_y P_Z(\boldsymbol{\alpha}, \boldsymbol{\gamma})$, consider $\mathbf{y}^* \in P$ and let for all t

$$F(t) = \max\{\, Odd(S, \mathbf{y}^*)\ : S \subseteq [0, t], |S| \text{ odd}\}$$

$$G(t) = \max\{Even(S, \mathbf{y}^*) : S \subseteq [0, t], |S| \text{ even}\}$$

- The resulting vectors **F** and **G** can be computed in linear time by Dynamic Programming
- In particular, the DP recursions imply that $(\mathbf{u}, \mathbf{v}) = (\mathbf{G}, \mathbf{F})$ satisfy (46)-(55), i.e., that $(\mathbf{G}, \mathbf{F}) \in \text{proj}_{u,v} Q_{UV}(\boldsymbol{\alpha}, \boldsymbol{\gamma})$
- Therefore $\mathbf{y}^* \in \text{proj}_y Q_{UV}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = \text{proj}_y P_Z(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ QED

Corollary: There is a **linear time** separation algorithm for conv $Y(\boldsymbol{\alpha}, \boldsymbol{\gamma})$: given $\mathbf{y}^* \in [0,1]^n$ compute **F** and **G** and check whether $(\mathbf{G}, \mathbf{F})$ satisfies (46)-(55)

Upper Bounds Only

The case of upper bounds only is much easier:

Proposition: conv $Y(\beta, \delta)$ is given by

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

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

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

Corollary: conv $Y(\alpha, \gamma) = \text{conv } Y(\alpha) \cap \text{conv } Y(\gamma)$

conv $Y(\beta, \delta) = \text{conv } Y(\beta) \cap \text{conv } Y(\delta)$

- Combining the two lower bound types, or the two upper bound types, does not give rise to new facets when projecting onto the **y**-subspace

Related Polytopes?

On the other hand, the structure of conv $Y(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is not well understood

- even in the case of stationary bounds $\boldsymbol{\alpha} = \overline{\alpha}$ (i.e., $\alpha_t = \alpha$ for all t) and $\boldsymbol{\beta} = \overline{\beta}$
- For example: *If* $\overline{\alpha} < \overline{\beta} < 2\overline{\alpha}$, *the inequalities*

$$\sum_{u=t}^{t+\overline{\alpha}-1} y_u + \sum_{u=t+\overline{\alpha}+1}^{t+\overline{\beta}} y_u + \sum_{u=t+\overline{\beta}+2}^{t+\overline{\alpha}+\overline{\beta}+1} y_u \leq \overline{\alpha} + \overline{\beta} - 1$$

are valid, and facet defining when t and $n - t$ are sufficiently large

Remark: Exchanging the roles of the on and off machine states and switch on off actions, all results about $Y(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and its convex hull translate (with an affine change of variables) into equivalent results about $Y(\boldsymbol{\gamma}, \boldsymbol{\delta})$ and its convex hull

Some References (1)

1. Two "classic" papers:

W.E. Smith, 1956: Various optimizers for single-stage production. *Naval Research Logistics Quarterly* **3** (1&2) 59-66

W.L. Eastman, S. Even, I.M. Isaacs, 1964: Bounds for the optimal scheduling of *n* jobs on *m* processors. *Management Sc.* **11** (2) 268-279

2. Unpublished "survey":

M. Queyranne and A.S. Schulz, Polyhedral Approaches to Machine Scheduling, Report 408/1994, Fachbereich Mathematik, T.U. Berlin, 1994

3. Other references (with apologies for the many omissions…)

D. Applegate, R. Bixby, W. Cook, and V. Chvátal, 1998: On the solution of traveling salesman problems. *Documenta Math.* **III.**

D. Applegate and W. Cook, 1991: A computational study of the job-shop scheduling problem. *ORSA J. Computing* **3** (2) 149-156.

E. Balas, 1985: On the facial structure of scheduling polyhedra. In *Math. Prog. Study* **24** (Essays in Honor of George B. Dantzig Part I) 179-218

E.A. Boyd, 1995: On the convergence of Fenchel cutting planes in mixed-integer programming. *SIAM J. Optimization* **5** (2) 421-435

J. Carlier, 1982: The one-machine sequencing problem. Europ J. O.R.. **11** (1) 42-47

C. Chekuri and R. Motwani, 1999: Precedence constrained scheduling to minimize sum of weighted completion times on a single machine. *Discrete Appl. Math*. **98** 29–38

Some References (2)

M. Chou, M. Queyranne and D. Simchi-Levi, 2006: The asymptotic performance ratio of an on-line algorithm for uniform parallel machine scheduling with release dates. *Math. Programming* **106** (1) 137-157

M. Chou, H. Liu, M. Queyranne and D. Simchi-Levi, 2006: On the asymptotic optimality of a simple on-line algorithm for the stochastic single machine weighted completion time problem and its extensions. *Operations Res*. **54** (3), 464-474

M.E. Dyer and L.A. Wolsey, 1990. Formulating the single machine sequencing problem with release dates as a mixed integer program. *Discrete Applied Math*, **26** (2-3) 255-270

L.L. Garver, 1963: Power generation scheduling by integer programming–development of theory, power apparatus and systems. III *Trans. Am. Inst. Electr. Eng.* **81** (3) 730–734

M.X. Goemans, 1997: Improved approximation algorithms for scheduling with release dates. *Eighth Annual ACM-SIAM Symp. on Discrete Algorithms*, New Orleans, 591--598

M.X. Goemans, M. Queyranne, A.S. Schulz, M. Skutella, Y. Wang, 2002: Single machine scheduling with release dates. *SIAM J. Discrete Math.* **15** (2) 165-192

M.X. Goemans, D.P. Williamson, 2000: Two-dimensional Gantt charts and a scheduling algorithm of Lawler. *SIAM J. Discrete Math.* **13** (3) 281--294

M. Grötschel, L. Lovász, and A. Schrijver, 1981: The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica* **1** (2) 169-197

Y. Kobayashi, A.R. Mahjoub, and S.T. McCormick, 2012: Separation of Series Constraints for One-Machine Scheduling with Precedence. Talk given at T.U. Berlin, June 2012.

H. Liu, M. Queyranne and D. Simchi-Levi**,** 2005: On the asymptotic optimality of algorithms for the flow shop problem with release dates. *Naval Res. Logistics* **52** (3) 232-242

Some References (3)

A.R. Mahjoub and S.T. McCormick, 2010: Separation Algorithms for Single-Machine Scheduling with Precedence Constraints. U. Paris Dauphine.

P. Malkin, 2003: Minimum runtime and stoptime polyhedra. In: Working Paper, Core, Université catholique de Louvain

F. Margot, M. Queyranne, and Y. Wang, 2003: Decompositions, network flows and a precendece constrained single machine scheduling problem. *Operations Res.* **51** (6) 981-992

M. Mastrolilli, M. Queyranne, A.S. Schulz, O. Svensson & N.A. Uhan, 2008. Minimizing the sum of weighted completion times in a concurrent open shop. Working paper.

R.H. Möhring, A.S. Schulz, M. Uetz, 1999: Approximation in stochastic scheduling: the power of LP-based priority policies. *J. ACM* **46** 924-942

M. Queyranne, 1993: Structure of a simple scheduling polyhedron. *Math. Prog* **58** 263-285

M. Queyranne and M. Sviridenko, 2002: Approximation algorithms for shop scheduling problems with minsum objective*. J. Scheduling* **5** (4) 287-305

M. Queyranne and Y. Wang, 1991: Single-machine scheduling polyhedra with precedence constraints. *Math. O.R.* **16** (1) 1-20. Erratum: *Math O.R.* **20** (3) 768 (1995)

M. Queyranne and L.A. Wolsey, 2017: Tight MIP formulations for bounded up/down times and interval-dependent start-ups. *Math. Prog.* **164** (1) 129-155

D. Rajan and S. Takriti, 2005: Minimum up/down polytopes of the unit commitment problem with start-up costs. IBM Res. Rep. RC23628 (W0506-050), T.J. Watson Res. Center, Yorktown Heights, NY

Some References (4)

M.W.P. Savelsbergh, R. N. Uma, J. Wein, 1998: An Experimental Study of LP-Based Approximation Algorithms for Scheduling Problems. *INFORMS J. Computing* **17** (1) 123-136

A.S. Schulz, 1996a: Polytopes and scheduling, PhD thesis, Technische Universität Berlin

A.S. Schulz, 1996b: Scheduling to minimize total weighted completion time: Performance guarantees of LP-based heuristics and lower bounds. In: W.H.Cunningham, S.T. McCormick, and M.Queyranne, eds. *Integer Programming and Combinatorial Optimization*. LNCS 1084, Springer, 301–315

J.B. Sidney, 1975: Decomposition algorithms for single-machine sequencing with precedence relations and deferral costs. *Operations Res.* **23** 283-298

L. Stougie and A.P.A. Vestjens, 2002: Randomized algorithms for on-line scheduling problems: how low can't you go? *Oper. Res. Letters* **30** (2) 89-96

G. Weiss, 1990: Approximation results in parallel machines stochastic scheduling. *Annals Oper. Res.* **26** 195-242

G. Weiss, 1992. Turnpike optimality of Smith's rule in parallel machines stochastic scheduling. *Math. Oper. Res.* **17** 255-270

L. A. Wolsey, 1985: Mixed integer programming formulations for production planning and scheduling problems. Invited talk, 12th Internat'l Symp. on Math. Programming, MIT.

L. A. Wolsey, 1989: Formulating single machine scheduling problems with precedence constraints. CORE Discussion Papers 1989/024, Université catholique de Louvain

L.A. Wolsey, 1998: *Integer Programming*. Wiley.

C.H. Xia, J.G. Shanthikumar, P.W. Glynn, 2000: On the asymptotic optimality of the SPT rule for the flow shop average completion time problem. *Oper. Res.* **48** (4) 615-622