CASPT 2015

Comparing two dual relaxations of large scale train timetabling problems

Frank Fischer · Thomas Schlechte

Abstract Railway transportation and in particular train timetabling is one of the basic and source application areas of combinatorial optimization and integer programming. We will discuss two well established modeling techniques for the train timetabling problem. In this paper we focus on one major ingredient - the bounding by dual relaxations. We compare two classical dual relaxations of large scale time expanded train timetabling problems - the Lagrangean Dual and Lagrangean Decomposition. We discuss the convergence behavior and show limitations of the Lagrangean Decomposition approach for a configuration based model. We introduce a third dualization approach to overcome those limitations. Finally, we present promising preliminary computational experiments that show that our new approach indeed has superior convergence properties.

Keywords Train Timetabling · Lagrangean Relaxation · Duality · Bundle Methods

1 Introduction

The train timetabling problem (`TTP`) asks for schedules for a set of trains with fixed routes in an infrastructure network so that certain operational restrictions like station capacities and headway times are satisfied.

One major approach for large scale instances is based on time expanded networks for modeling train schedules. These models give rise to huge integer programming formulations and cannot be solved directly by standard solvers.

F. Fischer
University of Kassel, FB 10, Heinrich-Plett-Str. 40, 34132 Kassel,
Tel.: +49 (561) 804-4130
E-mail: frank.fischer@uni-kassel.de

T. Schlechte
Zuse Institute Berlin, Takustr. 7, 14195 Berlin,
E-mail: schlechte@zib.de
`http://www.zib.de/members/schlechte`

Recently, several mathematical techniques, e. g., dynamic graph generation (Fischer and Helmberg, 2012), bundle methods (Fischer and Helmberg, 2014), rapid branching (Borndörfer et al, 2013; Weider, 2007), were developed to overcome this situation.

In this paper we focus on one major ingredient - the bounding by dual relaxations. Approaches based on duality allow to exploit structural properties of the optimization problem so that the dual problem decomposes into several simpler subproblems. Such approaches are frequently used in large scale optimization, in particular duty and crew scheduling are success stories in which similar approaches utilizing the bundle method entered in a productive planning software, see Borndörfer et al (2005); Borndörfer et al (2008); Borndörfer et al (2013).

We discuss two well-established modeling techniques for the TTP, first a classical formulation based on packing constraints to allow for conflict-free routings of the trains and second a configuration based formulation which controls a conflict-free sequencing of trains on each track of the network via an extended formulation. We compare the dual relaxation approaches for both models and discuss their convergence behavior when solved by a bundle method (Bonnans et al, 2003). In particular, we show limitations of the Lagrangean Decomposition approach for a configuration based model.

Furthermore, we introduce a third relaxation approach, which is a combination of the two classical approaches, and show that it combines the good bounds of the configuration network and the better convergence of the packing model.

2 The Train Timetabling Problem

We briefly recall a formal description of the TTP next. The *infrastructure network* is a directed graph $G^I = (V^I, A^I)$, where the nodes V^I represent stations, junctions, and crossings and the arcs A^I represent connecting railway tracks. Furthermore, we are given a set of trains R and each train $r \in R$ is associated with a path $G^r = (V^r, A^r) \subseteq G^I$ in the infrastructure network. Between two trains $r, r' \in R$ using the same track $a \in A^r \cap A^{r'}$ there is a minimal headway time $h_a(r, r') \in \mathbb{N}$ (in minutes), which is the minimal time between the two trains entering this track. Furthermore, there are capacity constraints on the nodes that state that at most a certain number of trains $c_u \in \mathbb{N}$ may be at the same station $u \in V^I$ at the same time instance $t \in T$.

One of the most successful models in the literature for solving the TTP is based on time expanded networks, see, e. g., Caprara et al (2002); Borndörfer and Schlechte (2007). Given a set of discrete time steps $T = \{1, \dots, |T|\}$ (usually minutes), we have for each train $r \in R$ a *time expanded network* $G^r_T = (V^r_T, A^r_T)$ where $V^r_T = V^r \times T$ and

$$\begin{aligned} A^r_T = &\{((u, t_u), (v, t_v)) \colon (u, v) \in A^r, t_v - t_u = \bar{t}^r_{(u,v)}\} \\ &\cup \{((u, t_u), (u, t_u + 1)) \colon u \in V^r_{\text{wait}}\}, \end{aligned}$$

with $V^r_{\text{wait}} \subseteq V^r$ the nodes at which r might wait and $\bar{t}^r_{(u,v)}$ the *running time* of r over track $(u,v) \in A^r$. A feasible schedule of train r then corresponds to a path $P \subseteq G^r_T$ from the first to the last station. We denote the set of all feasible paths in G^r_T by $\mathcal{P}^r$. We associate a binary variable $x^r_a \in \{0,1\}$ with each $a \in A^r_T$ in each time expanded network, where $x^r_a = 1$ if and only if a is contained in the timetable of train r.

The headway restrictions impose that certain arcs must not be contained in the final timetable simultaneously if they correspond to some train runs violating a headway constraint. In particular, let $(r,a) \in A^r_T$ and $(r',a') \in A^{r'}_T$ be two arcs with $a = ((u,t_u),(v,t_v))$ and $a' = ((u,t'_u),(v,t'_v))$ with $t'_u - t_u < h_{(u,v)}(r,r')$, then those arcs must not be used both. Therefore, we add the following *headway constraints* for each pair of incompatible arcs

$$x^r_a + x^{r'}_{a'} \leq 1, \qquad \{(r,a),(r',a')\} \in H, \tag{1}$$

where H is the set of pairs of incompatible train arcs. Let

$$\mathcal{H} := \Big\{ x = (x^r)_{r\in R} \in \{0,1\}^{\sum_{r\in R}|A^r_T|} \colon x \text{ satisfies } (1) \Big\}$$

denote the set of integer vectors satisfying the headway constraints.

The capacity constraints in the nodes mean that at each time instance $t \in T$ at most $c_u \in \mathbb{N}$ trains may be in $u \in V^I$ at the same time. Hence, the sum over all arcs representing a train being in u at time t

$$K(u,t) := \{(r,a) \colon a = ((u',t'),(u,t)) \in A^r_T, r \in R\}$$

must be at most c_u

$$\sum_{(r,a)\in K(u,t)} x^r_a \leq c_u, \qquad u \in V^I, t \in T. \tag{2}$$

Putting all together, the `TTP` can be formulated as integer program as follows. Given arc weights $w^r \in \mathbb{R}^{|A^r_T|}$, $r \in R$,

$$\begin{aligned} \text{maximize} \quad & \sum_{r\in R} \langle w^r, x^r \rangle \\ \text{subject to} \quad & x^r \in \mathcal{P}^r, \qquad r \in R, \\ & (1),(2), \end{aligned} \tag{IP'}$$

i.e., we select for each train r a feasible schedule $x^r \in \mathcal{P}^r$ (with a slight abuse of notation x^r denotes the incidence vector of a path $P^r \in \mathcal{P}^r$), so that all paths satisfy the headway and capacity constraints.

In the rest of the paper we focus on the formulation of the headway restrictions. Therefore, for the sake of simplicity of presentation, we drop the capacity conditions (2) (but note Remark (1) below) and focus on the model

$$\begin{aligned} \text{maximize} \quad & \sum_{r\in R} \langle w^r, x^r \rangle \\ \text{subject to} \quad & x^r \in \mathcal{P}^r, \qquad r \in R, \\ & (1). \end{aligned} \tag{IP}$$

3 Dual Bounds

Because solving (IP) using standard solvers is impossible for instances of practical size, many solution approaches are based on relaxation methods to obtain bounds on the optimal solution. One of them is Lagrangean relaxation of the coupling constraints. Different approaches to apply Lagrangean relaxation have been proposed in the literature. We will discuss two of them next.

First, we write (IP) as

$$\begin{aligned} \text{maximize} \quad & \sum_{r\in R} \langle w^r, x^r \rangle \\ \text{subject to} \quad & x^r \in \mathcal{P}^r, \qquad r \in R, \\ & x \in \mathcal{H}, \end{aligned}$$

with $\mathcal{H}$ consisting of all integer vectors that satisfy the headway constraints. Its convex hull $\operatorname{conv}\mathcal{H}$ is a polytope. There are two general approaches for describing $\mathcal{H}$. The first uses inequalities, i. e.,

$$\mathcal{H} = \left\{ x^r \colon \sum_{r\in R} M^r x^r \leq b \right\}. \tag{3}$$

The second approach assumes that one is able to optimize over $\mathcal{H}$, i. e., one is able to solve

$$\min \left\{ \langle p, \tilde{x} \rangle \colon \tilde{x} \in \mathcal{H} \right\}. \tag{4}$$

We will briefly discuss both approaches in the context of the TTP.

3.1 Relaxation of Clique Constraints

The first approach uses the outer description of $\mathcal{H}$, i. e., one solves

$$\begin{aligned} \text{maximize} \quad & \sum_{r\in R} \langle w^r, x^r \rangle \\ \text{subject to} \quad & x^r \in \mathcal{P}^r, \qquad r \in R, \\ & \sum_{r\in R} M^r x^r \leq b. \end{aligned}$$

Because this problem is typically hard to solve one employs Lagrangean relaxation of the coupling constraints:

$$\min_{y\geq 0} \max \left\{ \sum_{r\in R} \langle w^r, x^r \rangle + y^T (b - \sum_{r\in R} M^r x^r) \colon x^r \in \mathcal{P}^r \right\}. \tag{LR-Clq}$$

The function

$$\varphi^{\text{clq}}(y) := \max\left\{y^T b + \sum_{r\in R}\langle w^r - {M^r}^T y, x^r\rangle \colon x^r \in \mathcal{P}^r, r\in R\right\}$$

is a non-smooth convex function and can be optimized using, e. g., bundle methods, see Hiriart-Urruty and Lemaréchal (1993). Note that the evaluation of the function φ^{clq} at some point y requires the solution of the subproblems $\max\left\{\langle w^r - {M^r}^T y, x^r\rangle \colon x^r \in \mathcal{P}^r\right\}$ for each $r \in R$, which are shortest path problems in the time expanded networks.

The coupling constraints (1) alone are very weak in general so that the relaxation leads to weak bounds. Better bounds can be achieved by replacing (1) with *clique constraints*

$$\sum_{(r,a)\in C} x_a^r \leq 1, \qquad\qquad C\in\mathcal{C}, \tag{1'}$$

where $\mathcal{C}$ is a family of cliques of pairwise conflicting train arcs, i. e., for each $C \in \mathcal{C}$ we have

$$\forall\,(r,a),(r',a')\in C, (r,a)\neq(r',a)\colon \{(r,a),(r',a')\}\in H.$$

However, in general the family $\mathcal{C}$ is very large. Therefore the coupling constraints must be separated (which is possible when applying bundle methods, see, e. g., Helmberg (2004); Belloni and Sagastizábal (2009)) and typically only a small subset of $\mathcal{C}$ is considered at all (often only some cliques of two trains). This leads to stronger relaxations at the cost of more difficult algorithmic steps.

Remark 1 Note that capacity constraints (2) have a similar form as the clique constraints (1'). In practice one deals with them in very much the same way employing Lagrangean relaxation together with separation. This can be done for model (LR-Clq) as well as for all following models.

3.2 Extended Formulation

Another approach proposed in Borndörfer and Schlechte (2007) is based on an extended formulation. It introduces additional *configuration networks* in order to model (4). This leads to the formulation:

$$\begin{aligned}
\text{maximize}\quad & \sum_{r\in R}\langle w^r, x^r\rangle + \imath_{\mathcal{H}}(\tilde{x}) & & \\
\text{subject to}\quad & x^r \in \mathcal{P}^r, & r\in R, & \\
& x = \tilde{x}, & &
\end{aligned} \tag{IP-Cfg}$$

with $\imath_{\mathcal{H}}$ the indicator function of $\mathcal{H}$

$$\imath_{\mathcal{H}}(\tilde{x}) := \begin{cases} 0, & \tilde{x} \in \mathcal{H}, \\ -\infty, & \text{otherwise.} \end{cases}$$

Then one employs Lagrangean relaxation of the coupling constraint $x = \tilde{x}$ leading to

$$\min_{p \in \mathbb{R}^n} \varphi^{\text{cfg}}(p) := \left\{ \sum_{r \in R} \max_{x^r \in \mathcal{P}^r} \langle w^r - p^r, x^r \rangle + \max_{\tilde{x} \in \mathcal{H}} \langle p^r, \tilde{x} \rangle \right\} \tag{LR-Cfg}$$

This is also known as Fenchel duality approach in the literature, see, e. g., Boţ (2010). The idea is to define local feasible flows, which ensure headway feasibility on each infrastructure arc $a \in A^I$ and couple them appropriately with the train flows. Based on the already described time expanded networks G^r_T, we will define configuration networks G_a as illustrated in Figure 1.

The construction is as follows: Let s_a be an artificial source and t_a an artificial sink node to define a flow on track $a = (u, v) \in A^I$. The set

$$X_a := \{((u, t_u), (v, t_v)) \colon (u, v) = a\}$$

denotes all running arcs on track a. Let $L_a := \{(u, t_u) \colon ((u, t_u), (v, t_v)) \in X_a\}$ and $R_a := \{(v, t_v) \colon ((u, t_u), (v, t_v)) \in X_a\}$ be the associated sets of event nodes, i. e., representing departure and arrival, respectively, at the start and end station of track a. Note that all arcs in X_a go from L_a to R_a. We denote by $n(r, r', \tau) \in \mathbb{Z}$ for $(u, t_u) \in R_a$ the next possible departure time of train r' after train r has arrived at time τ:

$$n(r, r', \tau) = \tau - \bar{t}^r_{(u,v)} + h_a(r, r').$$

Now let $\overline{A}_a := \{((v, t_v), (u, t_u)) \colon (v, t_v) \in R_a, (u, t_u) \in L_a\}$ be a set of "return" arcs that go back in the opposite direction and represent the next potential departure on that track; they connect the end of a running arc on a (or node s_a) with all possible follow-on arcs (or node t_a) on a:

$$((v, t_v), (u, t_u)) \in \overline{A}_a \Leftrightarrow t_u \geq n(r, r', t_v).$$

In Figure 1, the construction is shown on a small set X_a. On the left-hand side, the set of running arcs of track a and the node sets L_a and R_a are shown. In the middle, the configuration network is constructed with dashed and dotted auxiliary arcs for the easy case of full block occupation, i. e., $h_a(r, r') = \bar{t}^r_{(u,v)} \Rightarrow n(r, r', \tau) = \tau$. Instead of constructing all possible return arcs each arrival node in R_a is only connected once with the time-line, i. e., with the next potential departure node L_a (or t_a). On the right-hand side of Figure 1 the reduced graph based on a time-line concept can be seen.

It is easy to see that the *configuration networks* are bipartite and acyclic, if all minimal headway times are strictly positive. Furthermore, Schlechte (2012) showed that there is a bijection from all s_a-t_a-paths in G_a to the set of valid

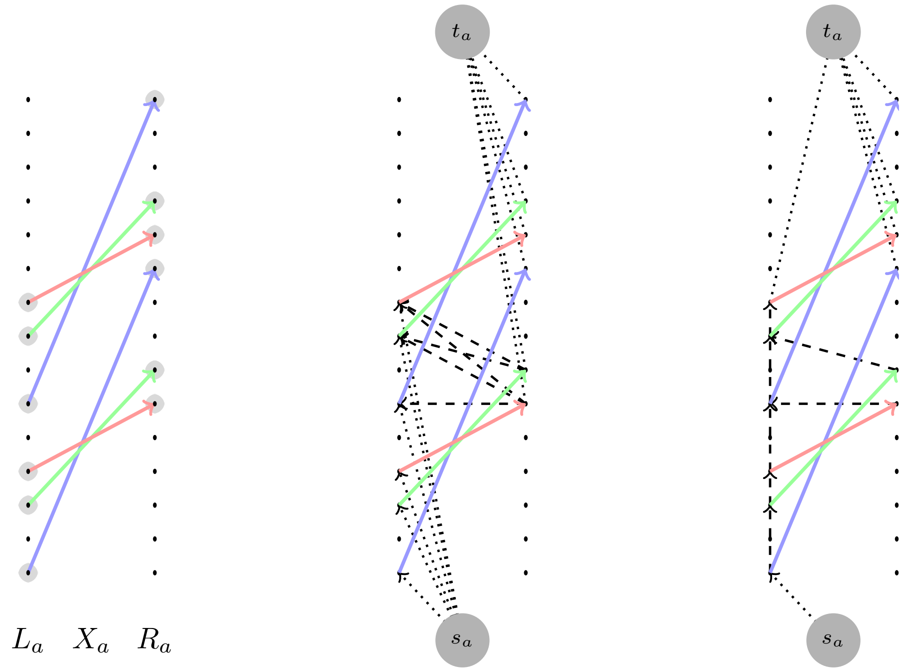

Fig. 1: Example for the construction of a track digraph.

configurations, i. e., selections of X_a, on track a if and only if the headway times are transitive. Thus, it is possible to solve $\max_{\tilde{x} \in \mathcal{H}} \langle p^r, \tilde{x} \rangle$ by longest s_a-t_a-path computations in G_a.

The idea of extended formulations is shown in Figure 2. On the left-hand side, the rough structure of the clique formulation can be seen, i. e., with appropriate binary matrices N and M. N is the classical network matrix, see Schrijver (2003), representing all feasible paths $\mathcal{P}^r$ in G_T^r. Furthermore, let $A = \bigcup_{r \in R} A^r$ and let $\delta \in \{-1, 0, 1\}^{|A|}$ be the classical right-hand side to ensure flows from the corresponding sinks to sources. The matrix M is simply the collection of M^r. On the right-hand side the structure of the extended model is shown. Matrix N_C denotes the network matrix of the configuration networks and C is the necessary coupling part with N. Note that matrices N and N_C are block diagonal matrices where each block corresponds to either a time expanded train network or a configuration network.

4 Solving the Relaxations

In this section we briefly describe our main algorithmic approach to solve the dual problems (LR-Clq) and (LR-Cfg). The standard approach is to employ a non-smooth first-order subgradient based method to solve the non-smooth convex optimization problems. In particular, bundle methods proved to be

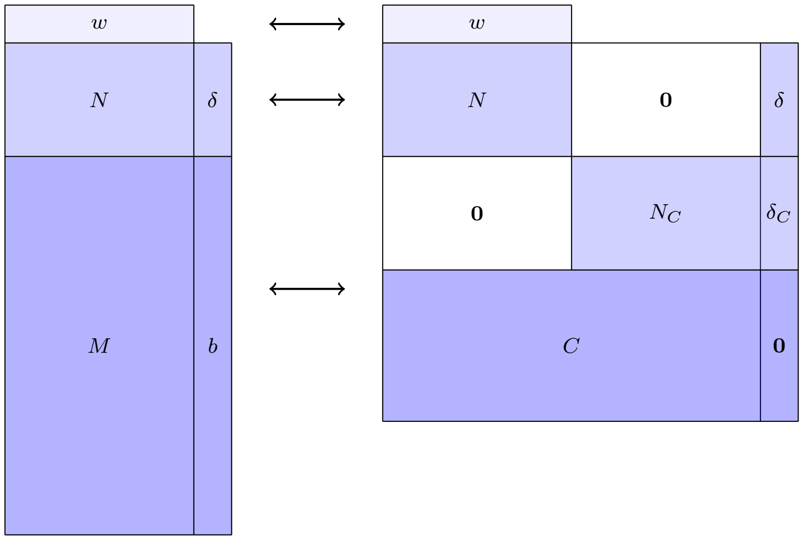

Fig. 2: Comparison of the matrix structure of the clique formulation and the extended formulation.The left picture shows the structure of the constraint matrix for clique constraints with (N, δ) being the flow constraints in the train networks, (M, b) the clique constraints. The right picture shows the configuration networks formulation with (N_C, δ_C) the additional flow constraints for the configuration networks and $(C, 0)$ the coupling between train and configuration networks.

very valuable tools. They are described in great detail in Bonnans et al (2003); Hiriart-Urruty and Lemaréchal (1993); Lemaréchal et al (1995).

A bundle method solving a non-smooth convex optimization problem

$$\min_{y \in \mathbb{R}^m} f(y),$$

which is given by a first order oracle computing for each $y \in \mathbb{R}^m$ the function value $f(y)$ and a subgradient $g(y) \in \partial f(y)$, works as follows. Given a current iterator $\hat{y}$ the algorithm forms a cutting plane model

$$\hat{f}(y) = \max\{\ell_i + \langle g_i, y \rangle : i \in N\}$$

of f around a $\hat{y}$. Then it determines a new candidate $\bar{y} \in \mathbb{R}^m$ by solving an auxiliary problem (for some *weight parameter* $u > 0$)

$$\bar{y} = \arg\min\left\{\hat{f}(y) + \frac{u}{2}\|y - \hat{y}\|^2\right\}. \tag{5}$$

The term $\frac{u}{2}\|y - \hat{y}\|^2$ can be thought of as some kind of trust region term preventing the new iterate to be too far away from $\bar{y}$. Afterwards the function f is evaluated at $\bar{y}$ and the function value $f(y)$ is compared with the model

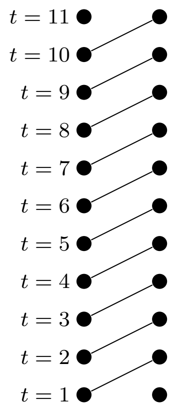

Fig. 3: Example train graph for bad convergence of (LD_2).

value $\hat{f}(\bar{y})$. If the real progress is reasonably large compared with the progress promised by the model, i. e., if

$$f(\hat{y}) - f(\bar{y}) \geq \varrho \cdot \left(f(\hat{y}) - \hat{f}(\bar{y}) \right),$$

for some *descent parameter* $\varrho \in (0, 1)$, the algorithm accepts the $\bar{y}$ as the new iterate (a so called "serious step"). Otherwise the model $\hat{f}$ is a bad approximation of f in $\bar{y}$, so the candidate is dropped and the algorithm stays at $\hat{y}$ but improves the model at $\bar{y}$ using subgradient information $g(\bar{y}) \in \partial f(\bar{y})$ (the algorithm does a "null step").

5 Comparing both Approaches

Next we compare the approaches presented in Sections 3.1 and 3.2 from a computational point of view. Theoretically both are equivalent in the sense that the optimal solutions correspond to each other, see Schlechte (2012). We need to separate inequalities (3) for (LR-Clq), and to solve problem (4) for (LR-Cfg). Both tasks are computationally hard in practice, so they are usually only done approximately (i. e., only some inequalities are separated and the optimization problem is relaxed itself). In practical applications the second approach turned out to provide stronger bounds (Borndörfer and Schlechte, 2007). However, when employing a first order method, e. g., a bundle method, to solve the Lagrangean dual problem, both models behave quite differently (also see Fischer (2013), Chapter 6).

Consider the following example with one track (u, v), two trains $R = \{A, B\}$, $\forall\, r, r' \in R$: $h_{(u,v)}(r, r') = 10$. Train A has a higher priority, so an optimal solution runs Train A at $t = 1$ and Train B at $t = 11$, see Figure 3 for an illustration.

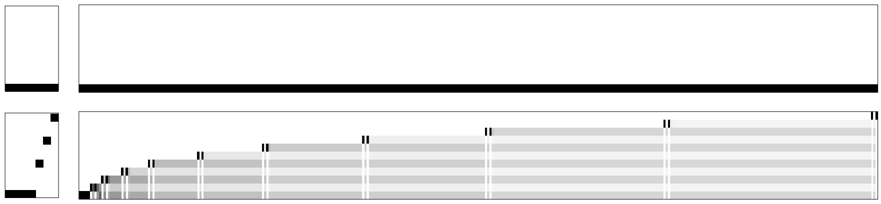

Fig. 4: Fractional solutions of train A (first row) and train B (second row) for (LD_1) (left pictures) and (LD_2) (right pictures), see Fischer (2013).

Figure 4 shows the primal solutions and their development after a certain number of iterations for (LR-Clq) and (LR-Cfg). We illustrate the fractional solution values with gray-scale values, from white representing 0.0 to black for 1.0. The upper row shows the active flow variables from train A and the lower one from train B, respectively. On the left-hand side, the algorithm increases the augmented costs on all arcs for $t = 1, \ldots, 10$ quickly for (LR-Clq). Thus, after few iterations Train B is forced to start at $t = 11$.

In contrast, for (LR-Cfg) the algorithm increases the augmented costs for Train B only for one arc in each iteration. After a few iterations $t = 1$ gets too expensive for Train B, so it uses $t = 2$. Then the costs for $t = 2$ are increased until B uses $t = 3$, but $t = 2$ is still used with some fractions as you can see from the gray color and so on. Eventually train B is allocated at $t = 11$ and the optimal primal solution is found, but obviously it needs several iterations until the gray areas are removed from the primal solution. During solving (LR-Cfg), the flow of train B smeared across the whole interval of potential departure times. Note, a finer discretization does not affect the convergence of (LR-Clq), but has an even worse effect on (LR-Cfg).

Figure 5 shows the objective value of a large instance after a certain number of iterations of the bundle method. The convergence for (LR-Clq) is *much* faster than for (LR-Cfg), although (LR-Cfg) might lead to better bounds eventually.

The reason is as follows. In each iteration the bundle method only gets information about violated constraints. For (LR-Clq) a single constraint couples train arcs of several time steps, whereas in (LR-Cfg) each constraint only couples one train arc (with its corresponding arc in the configuration network). Therefore much more iterations are required for (LR-Cfg) until the same information is accumulated.

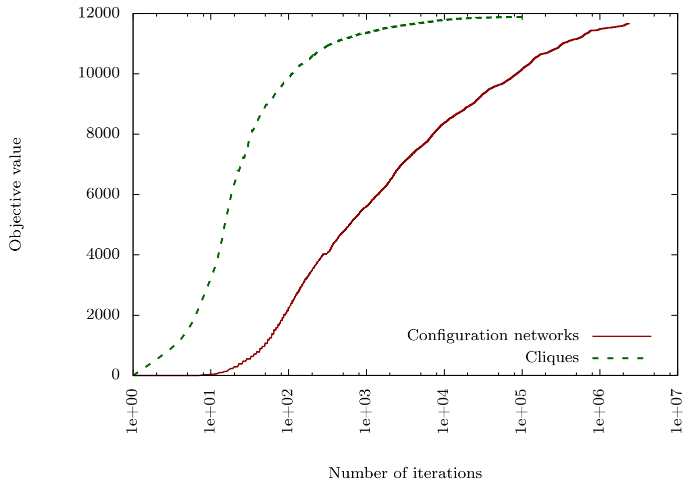

Fig. 5: Objective function after a certain number of iterations for (LR-Clq) (dashed line) and (LR-Cfg) (solid line), see Fischer (2013).

6 Combining both Approaches

In order to overcome the bad convergence properties of (LR-Cfg) we propose a different dualization approach. We consider

$$\min \left\{ \sum_{r \in R} \langle w^r, x^r \rangle + \imath_{\mathcal{H}}(\tilde{x}) \colon x^r \in \mathcal{P}^r, \sum_{r \in R} M^r x^r = \sum_{r \in R} M^r \tilde{x}^r \right\}, \qquad \text{(IP-Comb)}$$

being equivalent to (IP). Note, an optimal solution satisfies $\tilde{x} \in \mathcal{H}$, hence $\sum_{r \in R} M^r \tilde{x}^r \leq b$ and consequently $\sum_{r \in R} M^r x^r \leq b$. Lagrangean relaxation leads to

$$\min_y \varphi^{\text{cmb}}(y) := \left\{ \sum_{r \in R} \max_{x^r \in \mathcal{P}^r} \langle w^r - {M^r}^T y, x^r \rangle + \min_{\tilde{x} \in \mathcal{H}} \langle {M^r}^T y, \tilde{x} \rangle \right\}. \qquad \text{(LR-Comb)}$$

This formulation has the advantage that each coupling constraint couples several train arcs at the same time, hence we can hope for good convergence of the optimization algorithm. Furthermore, one can show that solving (LR-Comb) is equivalent to solving (LR-Cfg) by a proximal bundle method with an appropriately chosen scaling matrix (see (Bonnans et al, 2003)).

7 Combined Approach and Scaled Bundle Methods

One disadvantage of the combined approach is that during the run of the algorithm far more constraints are separated for (LR-Comb) than for (LR-Clq) or (LR-Cfg). The reason is as follows. In the case of (LR-Clq) a single clique constraint (1') for some $C \in \mathcal{C}$ is only separated if the left-hand side of the constraint is larger than one. This can only happen if at least two trains compete for the corresponding infrastructure arc at the same time. In contrast, the respective constraint for (LR-Comb)

$$\sum_{(r,a)\in C} x_a^r = \sum_{(r,a)\in C} \tilde{x}_a^r,$$

is violated, and therefore separated, as soon as some train uses one of the coupled arcs, even if no headway restriction is actually violated. Indeed, the number of separated constraints could be up to 100 times as large as for (LR-Cfg).

In order to overcome these difficulties, we use the following approach. Comparing the two dual functions φ^{cfg} and φ^{cmb} arising from (LR-Cfg) and (LR-Comb), respectively, we see the relation

$$\varphi^{\text{cfg}}(M^T y) = \varphi^{\text{cmb}}(y).$$

Indeed, given a current center $\hat{p} = M^T \hat{y}$, then solving the bundle subproblem (5) w. r. t. a model $\hat{\varphi}^{\text{cmb}}$ of φ^{cmb} is equivalent to solving the following scaled subproblem w. r. t. the model $\hat{\varphi}^{\text{cfg}}(p)$

$$\min_p \left\{ \hat{\varphi}^{\text{cfg}}(p) + \frac{u}{2} \|p - \hat{p}\|^2_{(M^T M)^{-1}} \right\}, \tag{6}$$

where $\|p\|^2_{(M^T M)^{-1}} := p^T (M^T M)^{-1} p$ denotes a scaled norm. Note that using scaled norms leads to convergent bundle methods under reasonable assumptions, see Bonnans et al (2003). Therefore, by only replacing the norm we get the same steps when solving (LR-Cfg) as we would get by solving (LR-Comb). Furthermore, storing the full matrix M to define the scaling term is not necessary. It is sufficient to use (and thus to separate) only few rows of M, so that solving (6) leads to steps that are good approximations of the steps defined by (5) for (LR-Comb). In particular, if we separate only an easily separable subset of all possible cliques $\mathcal{C}$ this only influences the scaling matrix, i. e. the algorithm, but not the quality of the model. In contrast, separating only few constraints in (LR-Clq) does influence the model quality, usually leading to weaker bounds.

8 Computational Results

We tested the algorithm on the public available instances of the RAS Problem Solving Competition 2012, see RAS (2012). The instance consists of roughly 100 nodes forming a corridor with mostly one-directional tracks and a few so called sidings at which overtaking is possible. The sidings are modeled using nodes with capacity two, all other nodes have capacity one. The time horizon is 8 hours and the possible timetables of all trains are completely free in the sense that trains might wait arbitrarily long given that no capacity constraints are violated. In the instances an increasing number of trains from 12 up to 20 should be scheduled in the network (in this paper we focus only on the largest instances).

For the time expanded models we use a discretization step size of one minute. The large time horizon and the freedom of the schedule of the trains (in particular, there is no upper bound on how long a train is allowed to wait) cause the time expanded model to become very large. Therefore we employ a Dynamic Graph Generation technique of Fischer and Helmberg (2012) for dealing with the shortest path subproblems in the train graphs as well as in the configuration networks (in the models that use them).

We used two different, simple objective functions. The first objective function ($Obj1$) penalizes the delay at the final station, i. e., for an arc $((u,t_u),(v,t_v)) \in A_T^r$ with $u \neq v$ and v being the final station of train $r \in R$, the weight is

$$w^r_{((u,t_u),(v,t_v))} = \alpha \cdot (t_v - \underline{t}^r)^2,$$

where $\underline{t}^r$ denotes the earliest possible arrival time (without delay) at the final station. The weight of all other arcs is zero. This is a typical goal for freight trains, for which only the arrival time at the final destination is important.

The second objective function ($Obj2$) penalizes the delay at each intermediate station. An objective function of this kind is usually used for passenger trains, for which delays not only at the last but also at intermediate stations should be avoided. In both cases the delay is penalized quadratically. This emulates a typical goal that a large delay of a single train should be avoided in favor for small delays of several trains.

We solved the problem approximately for the three models (LR-Clq), (LR-Cfg) and (LR-Comb), where (LR-Comb) is solved using the scaled bundle method described in Section 7. All algorithms are implemented in the Nim programming language and compiled with the Nim compiler 0.11.2 and GCC 4.8. The quadratic bundle subproblems are solved using Cplex 12.5.1 CPLEX (2014), the dynamic graph generation shortest path problems in the train graphs used the Dyng library Fischer (2014). The configuration network subproblems are solved exactly using a dynamic programming approach. All tests were done on an Intel Core i7 CPU at 3.5 GHz, 8 cores and 16 GB of memory.

The results are shown in Figures 6 and 7 for ($Obj1$) and ($Obj2$), respectively. The left pictures show the development of the objective value (note that the dual problem is a maximization problem) after a certain number of iterations. The right pictures show the objective value after a certain running time.

The first observation is that model (LR-Clq) leads, as expected, to weaker relaxations than the configuration based models (LR-Cfg) and (LR-Comb). The advantage of the latter is not very large for ($Obj1$) but significant for ($Obj2$). This justifies why one is interested in these kind of models in the first place. Comparing the running times, one sees that (LR-Clq) converges faster than the configuration based models. However, for ($Obj2$) the advantage of the configuration based formulation is so large that (LR-Comb) quickly creates

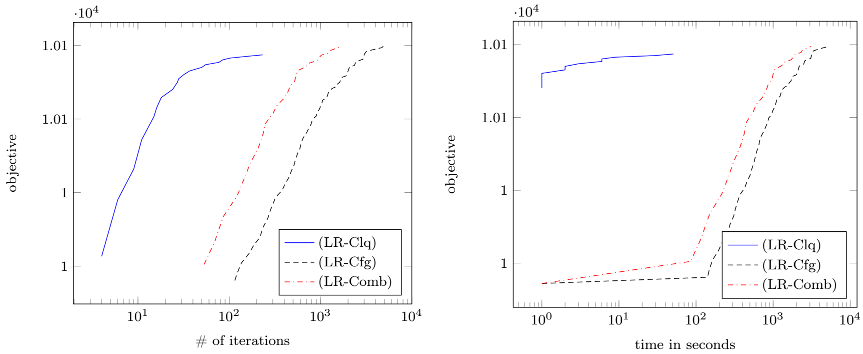

Fig. 6: Objective value after some iterations/time for all three relaxations with objective ($Obj1$).

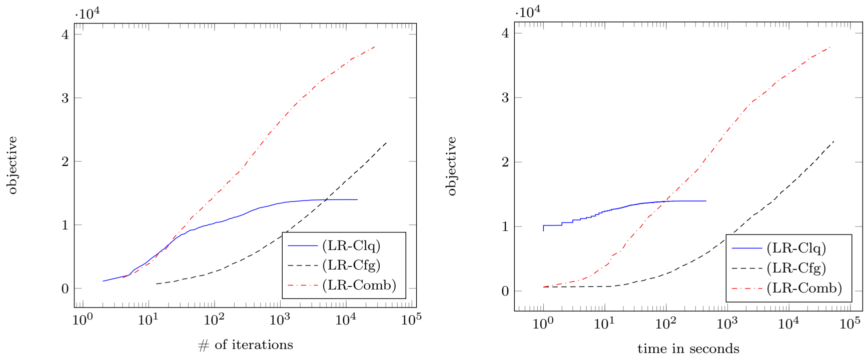

Fig. 7: Objective value after some iterations/time for all three relaxations with objective ($Obj2$).

better solutions. It is also apparent that the single iterations are much more costly for the configuration based models. This is mostly due to the fact that solving the configuration subproblems *exactly* takes a significant amount of time.

Comparing the traditional configuration based model (LR-Cfg) and the new combined model (LR-Comb), the latter performs much better. In particular considering ($Obj2$), the scaled model converges fast enough to outperform the clique based model after a short time, whereas (LR-Cfg) converges slower by more than one order of magnitude. However, it can also been seen, especially for ($Obj1$), that the advantage w. r. t. the number of iterations is larger than w. r. t. the running time. A single iteration for the scaled model is even more expensive for the new combined model. One reason is that the combined model must deal with both, configuration networks and cliques. Another reason is that the scaled quadratic subproblem (6) is more difficult to solve than the traditional subproblem.

9 Conclusions and Future Work

In this paper we proposed a new dualization approach for the train timetabling problem. Based on a time expanded formulation, the new approach combines the classical conflict clique based formulation with an extended formulation using configuration networks. The advantage of the new formulation is that it provides the good bounds of the configuration network formulation but at the same time has much better convergence properties.

While the computational tests are promising, they also show that a lot has to be done to improve the performance of the new approach. The additional steps required for the combined approach increase the computational effort and lead to more expensive iterations. Hence, further developments are necessary. One possibility would be to use the scaling only for configuration networks of some important infrastructure arcs, possibly by identifying these arcs during the solution process automatically.

Another important fact is that the new approach might allow to use the configuration based models for large instances in reasonable time. This allows the development of models that take advantage of the additional variables introduced for the extended formulation.

References

Belloni A, Sagastizábal C (2009) Dynamic bundle methods. Mathematical Programming 120:289–311, DOI 10.1007/s10107-008-0215-z, URL http://dx.doi.org/10.1007/s10107-008-0215-z

Boţ RI (2010) Conjugate Duality in Convex Optimization, Lecture Notes in Economics and Mathematical Systems, vol 637. Springer Berlin Heidelberg, DOI 10.1007/978-3-642-04900-2

Bonnans JF, Gilbert JC, Lemaréchal C, Sagastizábal CA (2003) Numerical Optimization. Springer

Borndörfer R, Schlechte T (2007) Models for railway track allocation. In: Liebchen C, Ahuja RK, Mesa JA (eds) ATMOS 2007 - 7th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, Dagstuhl Seminar Proceedings, http://drops.dagstuhl.de/opus/volltexte/2007/1170

Borndörfer R, Schelten U, Schlechte T, Weider S (2005) A column generation approach to airline crew scheduling. In: OR, Springer, pp 343–348

Borndörfer R, Löbel A, Weider S (2008) A bundle method for integrated multi-depot vehicle and duty scheduling in public transit. In: Hickman M, Mirchandani P, Voß S (eds) Computer-aided Systems in Public Transport, Springer-Verlag, Lecture Notes in Economics and Mathematical Systems, vol 600, pp 3–24, ZIB Report 04-14 at http://opus.kobv.de/zib/volltexte/2004/790/

Borndörfer R, Löbel A, Reuther M, Schlechte T, Weider S (2013) Rapid branching. Public Transport 5(1-2):1–21, DOI 10.1007/s12469-013-0066-8, URL http://dx.doi.org/10.1007/s12469-013-0066-8

Borndörfer R, Langenhan A, Löbel A, Schulz C, Weider S (2013) Duty scheduling templates. Public Transport 5(1-2):41–51, DOI 10.1007/s12469-013-0064-x, URL http://dx.doi.org/10.1007/s12469-013-0064-x

Caprara A, Fischetti M, Toth P (2002) Modeling and solving the train timetabling problem. Operations Research 50(5):851–861

CPLEX (2014) IBM ILOG CPLEX Optimization Studio V12.5.1 documentation. IBM ILOG CPLEX, URL http://www-01.ibm.com/support/knowledgecenter/SSSA5P_12.5.1

Fischer F (2013) Dynamic graph generation and an asynchronous parallel bundle method motivated by train timetabling. PhD thesis, Chemnitz University of Technology, URL http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-118358

Fischer F (2014) DynG Dynamic Graph Generation library. URL http://www.mathematik.uni-kassel.de/~fifr/fossils/dyng

Fischer F, Helmberg C (2012) Dynamic graph generation for the shortest path problem in time expanded networks. Mathematical Programming pp 1–41, DOI 10.1007/s10107-012-0610-3, URL http://dx.doi.org/10.1007/s10107-012-0610-3

Fischer F, Helmberg C (2014) A parallel bundle framework for asynchronous subspace optimization of nonsmooth convex functions. SIAM Journal on Optimization 24(2):795–822, DOI 10.1137/120865987, URL http://dx.doi.org/10.1137/120865987, http://dx.doi.org/10.1137/120865987

Helmberg C (2004) A cutting plane algorithm for large scale semidefinite relaxations. In: Grötschel M (ed) The Sharpest Cut, MPS-SIAM Series on Optimization, vol 4, SIAM/MPS, pp 233–256

Hiriart-Urruty JB, Lemaréchal C (1993) Convex Analysis and Minimization Algorithms I & II, Grundlehren der mathematischen Wissenschaften, vol 305, 306. Springer, Berlin, Heidelberg

Lemaréchal C, Nemirovski A, Nesterov Y (1995) New variants of bundle methods. Mathematical Programming 69:111–147, DOI 10.1007/BF01585555, URL http://dx.doi.org/10.1007/BF01585555

RAS (2012) RAS problem solving competition 2012. URL https://www.informs.org/Community/RAS/Problem-Solving-Competition/2012-RAS-Problem-Solving-Competition

Schlechte T (2012) Railway Track Allocation: Models and Algorithms. Doctoral thesis, Technische Universität Berlin

Schrijver A (2003) Combinatorial Optimization - Polyhedra and Efficiency. Springer

Weider S (2007) Integration of vehicle and duty scheduling in public transport. PhD thesis, TU Berlin, http://opus.kobv.de/tuberlin/volltexte/2007/1624/