2nd International Conference on Engineering Optimization
September 6-9, 2010, Lisbon, Portugal

Mixed-Integer Nonlinear Problems in Transportation Applications

Armin Fügenschuh, Stefan Vigerske
Konrad Zuse Zentrum für Informationstechnik Berlin (ZIB)
Takustraße 7, D-14195 Berlin, Germany
{fuegenschuh,vigerske}@zib.de

Henning Homfeld, Hanno Schülldorf
Technische Universität Darmstadt, Fachbereich Mathematik
Dolivostraße 15, D-64293 Darmstadt, Germany
{homfeld,schuelldorf}@mathematik.tu-darmstadt.de

Abstract
We give a short introduction to mixed-integer nonlinear programming from a linear programming perspective. In this approach, all nonlinearities are approximated by linear inequalities and spatial branching, such that in the end only linear programs remain that can be solved efficiently. We present two transportation applications from rather different fields: Railway freight on one side, and natural gas on the other. Both deal with service network design and naturally lead to MINLP formulations. We demonstrate how the nonlinear part can be approximated in a linear way.
Keywords: Mixed-Integer Nonlinear Programming; Linear Programming; Service Network Design; Railway Freight Transportation Networks; Natural Gas Networks.

1 Introduction

Many real-life problems contain a mixture of discrete decisions and continuous phenomena. In the sequel we present two different examples from the area of transportation network design and show how they can be modeled as mixed-integer nonlinear programs. For their numerical solution we apply techniques from mixed-integer linear programming (MILP). Hereto the nonlinear relationships are approximated by using only linear constraints and discrete decision variables. The problems are then solved with linear programming based branch-and-bound or branch-and-cut algorithms. In this way, modern numerical solvers are today able to handle large-scale nonlinear problem instances with thousands of variables and constraints. A further advantage, in particular in comparison to pure heuristic methods or gradient-based methods, is that these methods compute a certificate of optimality. That is, in every step an estimation towards the global optimum is given, so that the user can decide whether an intermediate solution is satisfactory or if some more time should be invested for searching a potentially better one.

One application concerns a service network design problem arising at Deutsche Bahn, one of the largest European railway companies. Their rail freight cars follow prescribed routes from origins via intermediate shunting yards to destinations. The goal in designing such routes is to reduce the number of trains and their travel distances. Various additional real-world hard constraints make the problem difficult to formulate and also to solve. The discrete decisions in this application represent the design of the service network. The nonlinearities arise from the fact that the waiting time of the cars in a shunting station decreases inversely proportional with the number of departing trains. We present different strategies how to re-formulate this nonlinear constraint as a linear one, such that numerical MILP solvers can be applied. Computational results using test- and real-world data show that this problem can be solved for instance sizes that are of practical interest.

The other application deals with the optimal extension of an existing gas transportation network. The physics of natural gas in pipelines is best modeled by partial differential equation systems. On a coarser level there exist reasonable approximation formulas. However, these formulas are still nonlinear equations. Besides the pipelines, the network contains active elements which can be controlled by dispatchers, such as valves and compressor stations to decrease or increase the pressure. Some elements can be controlled in a continuous fashion, others can only be on or off, hence they are discrete. In order to increase the networks capacity, new pipelines or new active elements can be built. Since these investments are very cost intensive, the pipeline operator is interested in the most economic network extension. Each potential extension gives rise to a further discrete decision variable. Since the number of potential extensions is astronomic, we give first some ideas how to reduce the complexity, before a numerical MILP solver can be applied on linearized versions of that problem.

2 Mixed-Integer Linear and Nonlinear Programming

We first describe mixed-integer linear programs (MILPs) and solution methods for them, and later discuss how mixed-integer nonlinear programs (MINLPs) fit into this framework.

MILPs are mathematical optimization problems that have the following structure:

$$\begin{aligned} z^{\text{MILP}} = \quad & \min \quad & c^T x, && \text{(1a)}\\ & \text{subject to} \quad & A\,x \le b, && \text{(1b)}\\ & & x \in \mathbb{Z}^p \times \mathbb{R}^{n-p}, && \text{(1c)} \end{aligned}$$

where $m, n, p \in \mathbb{N}, p \le n, A \in \mathbb{Q}^{m\times n}, b \in \mathbb{Q}^m$, and $c \in \mathbb{Q}^n$ is the input data describing the instance. Special cases include $p = n$, where all variables are integers, and $p = 0$, where all variables are continuous. Any vector $x \in \mathbb{Z}^p \times \mathbb{R}^{n-p}$ with $A\,x \le b$ is called a feasible solution for (1). Solving problem (1) means to compute a feasible solution that either minimizes the objective function $c^T x$, such that for any other feasible solution y we have $c^T x \le c^T y$. If this is computationally not possible within a reasonable time bound, we want to know at least how far away from global optimality the solution x is, with respect to some distance measure we will discuss in the sequel.

One of today's most successful ways to solve problem (1) is to tackle it from two sides, the primal and the dual. On the primal side we aim at finding feasible solutions quickly. Here one can resort to meta-heuristics such as genetic algorithms, tabu search, or simulated annealing, which have to be adapted to the problem at hand. These methods have demonstrated their ability to find solutions in short time for various optimization problems. However, it is in principle not possible to guarantee their optimality when these heuristics are terminated (usually when a certain amount of time has passed). If a feasible solution is returned by such heuristic we denote its objective function value by z^{primal}. Hence it is also necessary to approach the problem from the dual side.

The dual side deals with relaxations of (1) to something simpler that can be solved efficiently. Relaxation means that some parts of the problem are dropped, typically those that prevent us from solving the problem to optimality in the first place. Thereafter the relaxed parts are iteratively re-introduced into the problem formulation, in such a way that it remains solvable.

In general mixed-integer programming one typically relaxes the integrality condition in (1c). Then one obtains

$$\begin{aligned} z^{\text{dual}} = \quad & \min \quad & c^T x, && \text{(2a)}\\ & \text{subject to} \quad & A\,x \le b, && \text{(2b)}\\ & & x \in \mathbb{R}^n, && \text{(2c)} \end{aligned}$$

a problem called linear programming (LP) relaxation of (1). The set of feasible solutions to (2) is a polyhedron, that is, a convex subset of $\mathbb{R}^n$ that by a theorem of Weyl and Minkowski can alternatively be described as an intersection of a finite set of linear hyperplanes or as the sum of a (bounded) convex hull H of finitely many points and a (unbounded) cone C of finitely many rays. Problem (2) can now be solved efficiently with any linear programming algorithm, such as Dantzig's simplex algorithm [11]. Solving here means that these algorithms either return a global optimal solution (which is not necessarily unique), together with a certificate (that is, a proof) of optimality, or they verify that the instance is either unbounded (that means, the objective function points into the direction of the cone, so that one can find better and better solutions by following this direction), or they verify that the instance is infeasible, i.e., the polyhedron is in fact the empty set. Note that due to the convexity of the set of feasible solutions to (2) and the linearity of the objective function an optimal solution to (2), if it exists, will always be on the boundary of the polyhedron. By the nature of Dantzig's simplex algorithm, it return a basic solution, i.e., a solution in a "corner" of the polyhedron, that is, an extremal point of the convex hull H. Since all feasible solutions of (1) are also feasible solutions for (2), we have that $z^{\text{dual}} \le z^{\text{MILP}}$. Hence the solution value of the LP relaxation is a lower bound that can easily be obtained on the optimal solution of the MILP, which is difficult to obtain. When also a primal solution is available, one can define the optimality gap as follows:

$$\text{gap} := \frac{z^{\text{primal}} - z^{\text{dual}}}{z^{\text{dual}}}. \tag{3}$$

The gap is a common measure to specify how far away both bounds are. However, it gives no idea which of the two bounds should be improved – it can well be that one of them already equals the global optimum.

In order to improve the dual bound one has to re-introduce the integrality condition back into the formulation. One way to do this is to use cutting planes. It is a well known fact in convex analysis that two disjoint convex subsets in $\mathbb{R}^n$ can be separated by a hyperplane, such that one subset is on one side and the other subset is on the other side of the hyperplane. In our case, one of the subsets is the convex hull of the set of feasible solutions to (1) (one can prove that this is indeed a convex set, and, since we assumed rational data for A and b, even a polyhedron). The other convex subset is the optimal solution to the LP relaxation (2). If this point is already part of the first convex set, we are done and an optimal solution to (1) is found. Indeed this happens for some special structured problems. Several important graph problems, such as the maximum flow problem, the minimum cost flow problem, or the shortest path problem belong to this class. However, for such problems there exist much more efficient combinatorial algorithms, such that one does not have to use linear programming, which is a more general approach and thus computationally much slower. In general one expects that both convex sets are in fact disjoint, and thus a separating hyperplane does exist.

Finding such separating hyperplane, also called cutting planes or cuts, can be very demanding from a computational point-of-view. Over the past 50 years there were many approaches described in literature. Some of them apply only to problems having a special structure (such as comb inequalities for the traveling salesman problem [13]), others can be used for general purpose integer programming (such as mixed-integer rounding cuts [16], Gomory's mixed-integer cuts [12], or lift-and-project cuts [4]). Typical for cuts from the first category is that they are hard to generate, and sometimes finding such cut is almost as difficult as solving the underlying problem. Typical for cuts from the second category is that they are very easy to generate, but are too weak to improve the LP relaxation significantly. Moreover, they tend to have nasty algebraic structures – many non-zero and very large or very small coefficients – which causes numerical problems on finite precision floating point arithmetic hardware [20]. When a cutting plane is found, it is added to the LP relaxation, which gives an improved LP relaxation of the MILP. Hence one can solve this problem again, and check if the new LP optimal solution is now feasible for the MILP, or otherwise try to identify another separating hyperplane that cuts it off. This again is either time-consuming or increases the numerical problems, or both. The bottom line is that cutting planes must be handled with care, and one cannot expect to solve general mixed-integer problems by just adding a sufficiently large number of cuts to the LP relaxation.

Another way to re-introduce the relaxed integrality constraints on the variables is branching. If the optimal solution x^* to the LP relaxation (2) is not feasible, there exists an index i with $1 \leq i \leq p$ such that $x_i^* \in \mathbb{R}\backslash\mathbb{Z}$. Hence we can create two new LP relaxations (sons) from the previous (father) LP relaxation (2). In the one we add the constraint $x_i \leq \lfloor x_i^* \rfloor$, and in the other we add $x_i \geq \lceil x_i^* \rceil$ to (2). Thus we also get rid of the solution vector x^*, but at the price of creating two new problems out of one. When solving these two problems, it can happen again that some integer variable has a fractional value. In this case, branching can recursively be applied. This leads to a rapidly growing list of open problems: Every time we obtain a fractional solution, we split the problem into two subproblems and add them two the list. One can think of this list as a tree (similar to an ancestor tree), where the very first LP relaxation is the root node, and all other subproblems are its descendants. In order to keep this list as small as possible, one considers several reasons when further branching in a node is not necessary, and hence this part of the tree can be pruned: The node LP is infeasible, the node LP is a feasible solution for (1), or the objective function value of the node LP is greater-or-equal to the objective function value of the current best feasible solution to (1). The latter is the reason why having good feasible solutions early, for example from a heuristic, is important when solving large problems with this method.

Finally we remark that both methods, the branching and the cutting, can be combined in a procedure known as branch-and-cut. Here the LP relaxations in the tree are first improved by cutting planes, and branching is only carried out, if no more "useful" cuts are found. Today this approach is the foundation of almost any successful numerical MILP solver. As one can imagine it requires a careful implementation to find the right balance between cutting and branching. However, modern implementations of this method are able to solve problem instances having several thousands of variables and constraints.

At the end of our survey we want to sketch how nonlinear constraints can be handled within this framework. For further details we refer to [21, 22, 23]. Imagine a nonlinear constraint of the general form $f(x) \leq 0$, where $f : \mathbb{R}^n \to \mathbb{R}$, is part of the problem formulation. In order to set up an LP relaxation one replaces f by a linear approximation $D\,x \leq d$, where $D \in \mathbb{Q}^{q\times n}, d \in \mathbb{Q}^q$ and $\{x : f(x) \leq 0\} \subseteq \{x :$

$D\,x \leq d\}$. After solving this LP relaxation one can locally improve the linear approximation by either adding cutting planes that separate the LP relaxation solution vector from the convex hull of the set $\{x : f(x) \leq 0\}$. If this is not possible then the solution vector is part of the convex hull but not part of $\{x : f(x) \leq 0\}$. In this case one has to apply a procedure known as spatial branching, which is branching on a continuous variable. Denote by x^* the solution to the LP relaxation. Then one subproblem consists of $\{x : f(x) \leq 0, x \leq x^*\}$, and the other of $\{x : f(x) \leq 0, x \geq x^*\}$. Now one can try to find better linear approximations for these two sets.

3 Railway Freight Service Network Design

In 2006, around 500,000 Mil. ton kilometers freight was transported in Germany. The vast majority of freight is transported on the street (75%), a small share of 5% by ships, and the remaining 20% by train. The latter corresponds to 100,000 Mil. ton kilometers, where inland and cross-border traffic have an equal share of 45% each, leaving 10% for transit traffic. Deutsche Bahn, the largest German railway company, offers mainly two products to commercial and industrial customers that want to transport goods via rail.

Typically large customers order whole trains of about 20 to 40 cars. In this case, Deutsche Bahn as the operator can pull such a train by a locomotive from origin to destination. Small customers on the other hand order only 1 to 5 cars. In such case it is too expensive to pull this group of cars by a single locomotive through the network. Instead the cars are only pulled to the next classification yard. There they are grouped with the cars from other customers, and then as new trains pulled to the next classification yards. There the trains are disassembled, and the cars are again re-grouped with others. This procedure is now repeated until each car has reached its final destination. The central question emerging here for our research is the following: What are the “best” paths for all cars that fulfill certain operative restrictions. Whatever “best” means in this context will be specified in the sequel.

This problem is challenging from a computational point of view due to the enormous complexity inherent in a large railway company. Deutsche Bahn operates a network of a total length of about 38,000 km. This network is traversed each day by approximately 5,000 freight trains with together 150,000 cars. The origins and destinations of the customers’ orders are 2,200 terminal stations. Re-classification is carried out at intermediate classification yards of three different sizes: 200 are considered as small (“Satellitenbahnhöfe”), 30 as medium (“Knotenbahnhöfe”), and 11 are large (“Rangierbahnhöfe”). The differences among these will be explained in the sequel.

Classification yards schematically are made of three parts: entry tracks, sorting tracks, and exit tracks. When a train enters the yard, it is parked on an entry track. There the train is disassembled, and the individual cars are pushed over the hump, entering the sorting tracks behind. Each sorting track is assigned to a unique station (another yard or a terminal). Only by gravitational force the cars then roll into the right sorting tracks. As soon as enough cars are gathered on one sorting track, these cars are connected. This new train is then pulled into the exit group, where it waits, until it can leave the yard and re-enter the network. The operations within a classification yard alone give rise to interesting and difficult optimization problems. For our model, however, we can treat each yard as a black-box, and only have to take the respective capacity restrictions into account.

The problem of routing single cars through a railway network with intermediate classification operations emerged in the OR literature in the 1960s [2]. Computational studies mainly cover scenarios from US and Canadian railway freight companies [1, 3, 5, 6, 9, 10, 14, 17, 18], where different operational rules compared to Deutsche Bahn are applied. There the problem is known as the *blocking problem*, since it involves the decision which cars should stay together as a block (more than a car, less than a train) during their journey through the network. We will explain these differences in the next section.

3.1 The Model

There are mainly two aspects that distinguish the single car scheduling at Deutsche Bahn from some of the approaches found in the literature. First, the main cost factor are trains and the sum of all train travel distances (also called train kilometers, for short), the second most important is the use of infrastructure, and only the least important are car kilometers. Hence cars can make detours, as long as they do not violate some maximum travel time constraints. That means that one is interested in those car paths that bundle as much orders as possible. We give a formulation of the problem as a nonlinear mixed-integer programming problem.

Denote by V the set of all stations (yards and terminals), and by $A \subset V \times V$ the set of precedence relations. The set of customer orders is denoted by K.

We introduce three families of integer variables. For the routes of the cars we introduce decision variables $x^k_{i,j} \in \mathbb{B}$ for all $(i,j) \in A$ and $k \in K$. If $x^k_{i,j} = 1$ then station j is directly after station i in the path for the cars belonging to order k. Another decision of the model concerns the number of trains from station i to j, for which we introduce the integer variable $y_{i,j} \in \mathbb{N}$. Finally the model has to decide on the number of sorting tracks $n_{i,j} \in \mathbb{N}$ at station i on which trains are assembled exclusively in direction of station j.

Similar to the variables, there are also three main groups constraints: one for the orders, one for the trains, one for the yards.

The central constraints for the orders are the usual multi-commodity flow conservation constraints:

$$\forall i \in V, k \in K : \sum_{j:(i,j)\in A} x^k_{i,j} - \sum_{j:(j,i)\in A} x^k_{j,i} = \begin{cases} 1, & \text{if } i = o(k), \\ -1, & \text{if } i = d(k), \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

stating that each order k has to start at its origin $o(k)$ and end at its destination $d(k)$, and is not lost in between.

There is an upper bound restriction T_k on the time that the cars of order k are allowed to travel in the network. Typically the freight company offers so-called A-C-relations (delivery within 48 hours) and A-B-relations (express delivery within 24 hours). We denote by parameter $t_{i,j}$ the travel time from station i to j and introduce a variable $u^k_{i,j}$ to model the time that order k spends in station i, when heading to station j. The time limit constraint is the following:

$$\forall k \in K : \sum_{(i,j)\in A} (u^k_{i,j} + t_{i,j} \cdot x^k_{i,j}) \le T_k. \tag{5}$$

Waiting times on entry and exit tracks, sorting times, and times for coupling and decoupling the trains, can all be well estimated for each station i from historical data. One component however cannot be taken as a constant, namely the waiting time on a sorting track, until enough cars are gathered to start the assembly of a new train. This time depends on the number of trains per time period that are created, and which is the variable $n_{i,j}$ in our model. Hence a more precise model would not take $u^k_{i,j}$ as a constant, as we did in (5), but rather $u^k_{i,j}(n_{i,j}, x^k_{i,j})$ as a variable (or a function) depending on the number of trains and the routing decision:

$$\forall (i,j) \in A, k \in K : u^k_{i,j} = \frac{24}{n_{i,j}} \cdot x^k_{i,j}, \tag{6}$$

see also Figure 1. Due to these constraints our problem has turned into a nonlinear (nonconvex) mixed-integer problem. Since it is a combinatorial problem with otherwise linear structure we rather try to reformulate (6) as a linear constraint instead of using nonlinear solution techniques that are faster on the nonlinear part but then struggle with the integer decisions.

Each yard i has a maximal capacity of cars H_i that it can handle per time period. For large yards with a hump the limiting factor is the respective hump capacity. For small yards without a hump, the sorting is carried out by small locomotives, which is even more a limiting factor. If we denote by v_k the number of cars belonging to order k we can formulate this capacity restriction as the following constraint:

$$\forall i \in V : \sum_{k\in K, j:(i,j)\in A} v_k \cdot x^k_{i,j} \le H_i. \tag{7}$$

The number of trains that can be assembled per time period on a single sorting track is a station dependent parameter N_i for yard i. Typically this value is between 3 and 6. We thus have the constraint

$$\forall (i,j) \in A : n_{i,j} \le N_i \cdot y_{i,j}. \tag{8}$$

This means, the number of tracks in station i on which trains in direction of j can be assembled equals the number of trains from i to j divided by the number trains per track and time period.

Each yard i has a total number of sorting tracks Y_i. Hence the total number of assignments must not exceed this capacity restriction:

$$\forall i \in V : \sum_{j:(i,j)\in A} y_{i,j} \le Y_i. \tag{9}$$

The trains have an upper limit on the total length and the weight of the cars. It is not allowed to assemble trains longer than 700m. This is because freight trains have to wait in the network at certain bypass tracks so that faster passenger trains can overtake them. Since these bypasses are 700m long (speaking of the German railway network, in Denmark, for example, this is 850m), no longer freight train is allowed to travel in the network. If l_k denotes the length in meters of all cars belonging to order k and $L_{i,j}$ the maximum length of trains between i and j, then

$$\forall (i,j) \in A: \sum_{k \in K} l_k \cdot x_{i,j}^k \leq L_{i,j} \cdot n_{i,j}. \tag{10}$$

There is an upper bound weight restriction $W_{i,j}$ that reflects the strength of DB's locomotives and the track properties between i and j. If w_k denotes the weight in tons of all cars belonging to order k, then

$$\forall (i,j) \in A: \sum_{k \in K} w_k \cdot x_{i,j}^k \leq W_{i,j} \cdot n_{i,j}. \tag{11}$$

Finally we have the "Leitweg" unique successor constraint: The successive station only depends on the destination, not on the origin, and also not on the order. That means, if two different orders with the same destination meet at some station in the network, it is not allowed to send them further to different stations, which is modeled by the following inequalities:

$$\forall k, l \in K, l \neq k, d(k) = d(l), i \in V, (i, j_1), (i, j_2) \in A, j_1 \neq j_2 : x_{i,j_1}^k + x_{i,j_2}^l \leq 1. \tag{12}$$

The overall objective is to find the most economical car paths. The most crucial cost component is the number of train kilometers, second is the amount of used infrastructure, which we assume is proportional to the number of sorting tracks. Of least importance is the number of car kilometers. Let $\delta_{i,j}$ denote the distance (in the network) between station i and j, then we have the following objective function:

$$\alpha_1 \cdot \sum_{(i,j) \in A} \delta_{i,j} \cdot n_{i,j} + \alpha_2 \cdot \sum_{(i,j) \in A} y_{i,j} + \alpha_3 \cdot \sum_{k \in K, (i,j) \in A} \delta_{i,j} \cdot x_{i,j}^k \rightarrow \min, \tag{13}$$

where $\alpha_1, \alpha_2, \alpha_3$ are suitably scaled weight parameters.

3.2 Linearizing the Turnover Time Constraints

Since we are only interested in the integer points in $x_{i,j}^k$ and $n_{i,j}$, we are able to derive a complete description of the convex hull of (6). If the number of trains $n_{i,j}$ can be bounded from above, this description is compact in a way that it does not need too much additional inequalities.

For the linearization we approximate (6) by the following family of linear equations, cf. Figure 2:

$$\begin{aligned} T(2a+1)x_{i,j}^k - Tn_{i,j} &\leq a(a+1)w_{i,j}^k, \quad \forall\ a = 1, \ldots, N_{i,j} - 1, && \text{(14a)} \\ Tx_{i,j}^k &\leq N_{i,j} w_{i,j}^k, && \text{(14b)} \\ w_{i,j}^k &\leq Tx_{i,j}^k, && \text{(14c)} \\ N_{i,j} w_{i,j}^k &\leq TN_{i,j} + Tx_{i,j}^k - Tn_{i,j}, && \text{(14d)} \\ x_{i,j}^k &\leq 1. && \text{(14e)} \end{aligned}$$

Here $N_{i,j}$ denotes an upper bound on the number of trains that can at most travel from i to j. The better this bound, the tighter the formulation will be.

Theorem 1 *The equations* (14) *describe the convex hull of the set of points*

$$\mathcal{P}_{i,j}^k := \{(w_{i,j}^k, n_{i,j}, x_{i,j}^k) \in \mathbb{R}_+ \times \mathbb{Z}_+ \times \{0,1\} : w_{i,j}^k = \frac{T}{n_{i,j}} x_{i,j}^k, n_{i,j} \leq N_{i,j}\}. \tag{15}$$

After replacing (6) by (14) we transformed the MINLP into a pure MILP, which is ready to be solved by standard software. Note that neither spatial branching nor additional cuts are needed here.

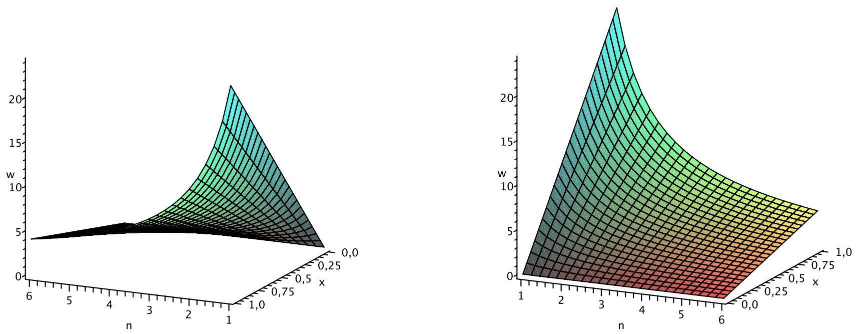

Figure 1: Nonlinear waiting time function (6) from two different angles.

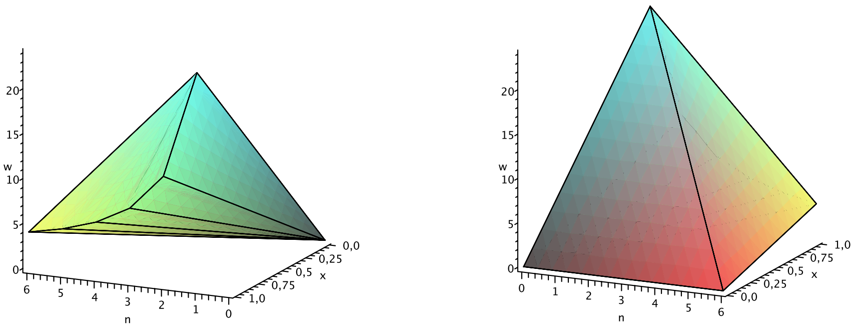

Figure 2: Polyhedral approximation (14) of the nonlinear waiting time, same perspective as in Fig. 1.

4 Extension Planning for Natural Gas Networks

Natural gas is a nontoxic, odorless, transparent, and flammable gas that originates from underground deposits. It is often found together with crude oil or coal since it is formed by similar biological and geological processes. The main ingredients of this fossil source of energy are methane (75-99%), ethane (1-15%), propane (1-10%), and small fractions of butane and ethene. Today natural gas is mainly used for heating private houses and office buildings, for the generation of electrical power, and as fuel for vehicles. It is also used for several reactions in chemical process engineering. The world's joint resources in natural gas are assumed to last for the next 60 to 500 years, if conveyance and consumption remain on the current level. In many regions natural gas was found and excavated. The largest producers are the USA, Russia, and Canada. With natural gas around one quarter of the world's energy demand is covered. Since natural gas is the "greenest" energy source among the fossil ones, its market share is estimated to grow to 50% towards the end of this century.

The natural gas must be transported from the deposits to the customers, sometimes over distances of several thousand kilometers. For very long distances (more than 4000km) it is more economic to cool down the gas to $-160°$C such that is becomes liquid and to transport it via ships. For shorter distances or for the delivery to the end customers large pipeline systems are used. The long distance steel pipelines have a diameter of up to 1.5m and the gas has a pressure of up to 100 bar. Every 100 to 150km a compressor station is needed to maintain this pressure level. Their pumps are consuming parts of the gas they should transport. For a 4000km transport, around 8% of the gas is consumed. For the delivery of the gas to cities a fine grid of low pressure (16 bar) pipelines is used. A third grid delivers gas to end customers, where the pressure is only slightly higher (20 mbar) than the atmospheric pressure.

In the past the gas network operator and gas vendor were the same company. Hence the operator had full control on the entry side of the network. He could then decide where and when to fill or empty gas storages or how to route the gas through the pipeline in the most convenient way. Due to European deregulation laws this is no longer the case. Operators and vendors were forced to split up into totally separate and independent companies (but sometimes still connected in a holding). Now the operator does not know and cannot control where gas is entering the network, but still he is responsible that all nominated quantities reach their destinations.

The increasing importance of gas as a fossil energy source and the currently emerging higher flexibilities in a deregulated gas market are challenges for the network operators. It is a realistic assumption that the current gas networks will soon be reaching their capacity limitations, and hence they are in demand for an extension. Building new pipelines or compressor stations are huge investments, typically in the range of hundreds of million Euros. Hence a careful planning is necessary, which creates a current demand for mathematical optimization methods [7, 8, 15, 19, 24].

4.1 The Model

Given is a gas pipeline network as a directed graph $D = (V, A)$, where the set of nodes V is the disjoint union of entry nodes V^{entry}, exit nodes V^{exit}, and intermediate nodes V^{inter}. The set of arcs is the disjoint union of passive arcs, or pipelines, A^{pipe}, and active arcs, which are compressor stations A^{cs}, or valves A^{valve}. Finally, we have a set of potential pipeline arcs A^{new}, from which some can be built in addition to the existing network.

For each node $i \in V$ the amount of gas entering or leaving the network is given by $s_i \in \mathbb{R}$. We have $s_i > 0$ for all $i \in V^{\text{entry}}$, $s_i < 0$ for all $i \in V^{\text{exit}}$, and $s_i = 0$ for all $i \in V^{\text{inter}}$. The vector $s \in \mathbb{R}^V$ specifies a nomination scenario of gas buyers and sellers, which the gas transportation company needs to deliver through their network. If this is possible, the scenario is said to be feasible, otherwise it is infeasible. We are here interested in infeasible scenarios. The question is, if an infeasible scenario can be turned into a feasible one by extending the network. If so, one might additionally ask for the minimum cost extension.

In order to be able to give answers to this question we have to model the behavior of natural gas in pipelines, i.e., the physics of gas. To this end, we introduce two families of variables. One are flow-per-arc variables $f_{i,j} \in \mathbb{R}$ for all $(i, j) \in A$. If $f_{i,j} > 0$ then gas flows from i to j, and for $f_{i,j} < 0$ the gas flows in the opposite direction. The other are pressure-per-node variables $p_i \in \mathbb{R}_+$, which are bounded by minimum and maximum node pressures:

$$\forall i \in V : \underline{p}_i \le p_i \le \overline{p}_i. \tag{16}$$

The upper bound relates to physical limitations (if violated the network might be damaged), the lower bound is due to contracts with the customers. A final family of binary decision variables $x_{i,j} \in \{0, 1\}$ for all $(i, j) \in A^{\text{new}}$ is introduced to model the decision if pipeline (i, j) is build ($x_{i,j} = 1$) or not ($x_{i,j} = 0$).

Since we study a network flow problem, we have the usual flow conservation or continuity constraint in our model:

$$\forall i \in V : \sum_{j:(i,j)\in A} f_{i,j} - \sum_{j:(j,i)\in A} f_{j,i} = s_i, \tag{17}$$

which means that all gas that enters a node has to leave the node, only plus or minus the gas that is nominated at this node. This constraints occur in any network flow model. In addition to that we have special constraints that model the physical properties of the flow of gas. Here we use a quadratic pressure loss equation of the form

$$\forall (i, j) \in A^{\text{pipe}} : f_{i,j}\,|f_{i,j}| = k_{i,j}(p_i^2 - p_j^2). \tag{18}$$

Here $k_{i,j}$ is a pipeline dependent constant that models all physical properties of the gas, such as its compressibility and its (estimated average) Reynolds number, and the pipe, such as its length, diameter, and roughness.

For potential new pipelines we have to multiply the right-hand side of (18) by the building decision:

$$\forall (i, j) \in A^{\text{new}} : f_{i,j}\,|f_{i,j}| = k_{i,j}(p_i^2 - p_j^2) \cdot x_{i,j}. \tag{19}$$

Compressor stations can increase the pressure up to a certain level (for example, the output pressure can be increased by at most 60% of the input pressure):

$$\forall (i, j) \in A^{\text{cs}} : p_i \le p_j \le 1.6\, p_i. \tag{20}$$

Valves are working in the opposite direction. They decrease the pressure level:

$$\forall (i, j) \in A^{\text{valve}} : p_i \ge p_j. \tag{21}$$

The objective is to minimize the network extension costs, which are the sum of costs for individual decisions:

$$\sum_{(i,j)\in A^{\text{new}}} c_{i,j}\, x_{i,j} \to \min, \tag{22}$$

where $c_{i,j} \in \mathbb{R}_+$ are the construction cost for pipeline $(i, j) \in A^{\text{new}}$.

4.2 Linearizing the Pressure Loss Constraints

The above model is a mixed-integer nonlinear problem. We solve it by applying MILP techniques as described in Section 2. A first step is to decrease the degree of nonlinearity in the model as far as possible. Note that the pressure variables occur only as squared variables. Thus it is possible to replace them by new variables $\pi_i \in \mathbb{R}_+$ for all $i \in V$ with $\pi_i = p_i^2$. Then (18) reads as follows:

$$\forall (i,j) \in A^{\text{pipe}} : f_{i,j}\,|f_{i,j}| = k_{i,j}(\pi_i - \pi_j), \tag{23}$$

and the right-hand side is already linear.

Since $x_{i,j}$ is a zero-one variable, the right-hand side of constraints (19) can be linearized by replacing the variable multiplication by an addition and a suitably big value of M:

$$\forall (i,j) \in A^{\text{new}} : k_{i,j}(\pi_i - \pi_j) - M \cdot (1 - x_{i,j}) \leq f_{i,j}\,|f_{i,j}| \leq k_{i,j}(\pi_i - \pi_j) + M \cdot (1 - x_{i,j}). \tag{24}$$

Furthermore it pays off to add cutting planes to improve the LP relaxation, such as

$$\forall (i,j) \in A^{\text{new}} : -M\,x_{i,j} \leq f_{i,j} \leq M\,x_{i,j}. \tag{25}$$

The only remaining nonlinear function is the monotone parabola $f_{i,j} \mapsto f_{i,j}\,|f_{i,j}|$. This function is replaced by a linear approximation of the convex hull of its epigraph. This approximation is then refined by cutting planes and spatial branching, see Figure 3.

a) b) c)

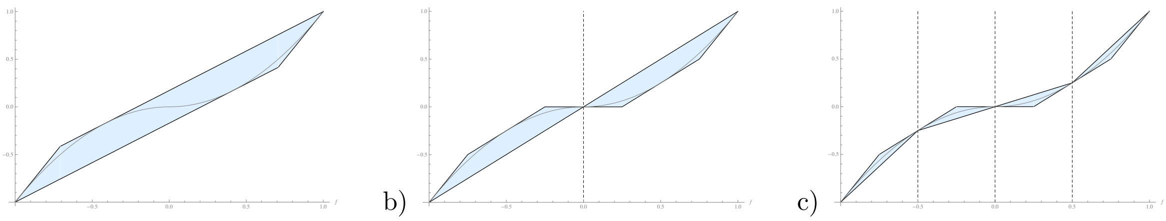

Figure 3: a) Polyhedral outer approximation of $f \mapsto f\,|f|$, b) initial spatial branching on zero, c) further spatial branching.

5 Conclusion

We gave a short introduction to mixed-integer nonlinear programming and demonstrated how mixed-integer linear programming techniques, in particular linear programming relaxations and branch-and-cut methods can be used to numerically obtain optimal or near-optimal solutions for given MINLP instances. We gave two applications from different fields of transportation engineering, railway freight and natural gas transportation. Both problems can be modeled as mixed-integer linear problems to large extends. However, in order to be able to handle particular and important subtleties of the application, one has to take also nonlinear constraints into account which make to problems much harder to solve. We demonstrated that an analysis of the problems can lead to good linear approximations such that numerical MILP solvers are able to handle the problem. Doing so we were able to compute numerical solutions for problem instances that are of practical relevance.

We believe that our approach is not limited to the two applications described here, but can also be applied to other nonlinear mixed-integer problems in general. With the current state of standard nonlinear solvers, which is still behind the state of linear solvers, it is clear that nonlinear solvers cannot handle large problem instances out of the box. Hence a careful analysis of any problem under consideration is necessary to be able to find optimal or near-optimal solutions.

References

[1] R. Ahuja. Solving real-life railroad blocking problems. *Interfaces*, 2007.

[2] A.A. Assad. Analytical models in rail transportation: An annotated bibliography. *INFOR*, 19:59 – 80, 1981.

[3] A.A. Assad. Analysis of rail classication policies. *INFOR*, 21:293 – 314, 1983.

[4] E. Balas, S. Ceria, and G. Cornuéjols. A lift-and-project cutting plane algorithm for mixed $0-1$ programs. *Mathematical Programming*, 58:295–324, 1993.

[5] C. Barnhart, J. Hong, and P.H. Vance. Railroad blocking: A network design application. *Operations Research*, 48:603 – 614, 2000.

[6] L.D. Bodin, B.L. Golden, and A.D. Schuster. A model for the blocking of trains. *Transportation Science B*, 14:115 – 120, 1980.

[7] Ian D. Boyd, Patrick D. Surry, and Nicholas J. Radcliffe. Constrained gas network pipe sizing with genetic algorithms. Technical Report EPCC-TR94-11, Edinburgh Parallel Computing Centre, 1994.

[8] Luis Castillo and Antonio Gonzáleza. Distribution network optimization: Finding the most economic solution by using genetic algorithms. *European Journal of Operational Research*, 108(3):527–537, 1998.

[9] T. Crainic, J.-A. Ferland, and J.-M. Rousseau. A tactical planning model for rail freight transportation. *Transportation Science*, 18(2):165 – 184, 1984.

[10] T. Crainic and J.-M. Rousseau. Multicommodity, multimode freight transportation: a general modeling and algorithmic framework for the service network design problem. *Transportation Research B*, 20(3):225 – 242, 1986.

[11] G.B. Dantzig. Maximization of a linear function of variables subject to linear inequalities. In T.C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 330 – 335. Wiley, New York, 1951.

[12] R.E. Gomory. An algorithm for the mixed integer problem. Technical Report RM-2597, The RAND Cooperation, 1960.

[13] M. Grötschel and M. Padberg. On the symmetric travelling salesman problem I: Inequalities. *Mathematical Programming*, 16:265 – 280, 1979.

[14] M.H. Keaton. Designing optimal railroad operating plans: Lagrangian relaxation and heuristic approaches. *Transportation Research B*, 23:330 – 345, 1992.

[15] O. Mariani, F. Ancillai, and E. Donati. Design of a gas pipeline: Optimal configuration. Technical Report PSIG 9706, Pipeline Simulation Interest Group, 1997.

[16] G.L. Nemhauser and L.A. Wolsey. A recursive procedure to generate all cuts for $0-1$ mixed integer programs. *Mathematical Programming*, 46:379 – 390, 1990.

[17] H. Newton. *Network Design Under Budget Constraints with Application to the Railroad Blocking Problem.* PhD thesis, Auburn University, Auburn, Alabama, USA, January 1997.

[18] H. Newton, C. Barnhart, and P.H. Vance. Constructing railroad blocking plans to minimize handling costs. *Transportation Science*, 32:330 – 345, 1998.

[19] A.J. Osiadacz and M. Górecki. Optimization of pipe sizes for distribution gas network design. Technical Report PSIG 9511, Pipeline Simulation Interest Group, 1995.

[20] M.W. Padberg. Classical cuts for mixed-integer programming and branch-and-cut. *Mathematical Methods of OR*, 53(2):173 – 203, 2001.

[21] E.M.B. Smith and C.C. Pantelides. A symbolic reformulation/spatial branch-and-bound algorithm for the global optimization of nonconvex MINLPs. *Computers & Chemical Engineering*, 23:457–478, 1999.

[22] M. Tawarmalani and N.V. Sahinidis. *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications.* Kluwer Academic Publishers, 2002.

[23] M. Tawarmalani and N.V. Sahinidis. Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. *Mathematical Programming*, 99:563–591, 2004.

[24] D. De Wolf and Y. Smeers. Optimal dimensioning of pipe networks with application to gas transmission networks. *Operations Research*, 44(4):596 – 608, July-August 1996.