Global optimization of water distribution systems

Kaj-Mikael Björk
IAMSR, CIEM / Åbo Akademi University
Joukahainengatan 3-5A, FIN-20520 ÅBO
Finland
Email: kaj-mikael.bjork@abo.fi

Lazaros G. Papageorgiou
Dept. of Chemical Engineering
University College London
Torrington Place
London WC1E 7JE, U.K.

Abstract

Recently [1] presented a new model for the solution of the water distribution network problem. The model captured successfully the essence of the problem structure, but could not guarantee global convergence. This paper adds, to the current research track, a new model and method that can guarantee global convergence for the water network design problem. The contribution is important since there has not been any effort to find methods ensuring global convergence and the solutions found to several example problems in earlier contributions are not qualitatively established.

1. Introduction

Water distribution networks have been a well studied topic during the last decades. The main problem is to find an optimal structure for a network, when there is a water source (or several sources) and links between the nodes in the network (the distribution points). The use of optimization in the quest to find good network structures has been used since [2] and remains an interesting research track. Usually the network under study is supplied by gravity. This means that the optimization routine needs to find the optimal diameters for the pipelines between the nodes as well as the right flow rates in the network. The demand at each distribution point is fixed as well as the supply of water from the source node. The task to find the optimal solution may sound easier than it really is. The reason is found in the nonlinear relationship between the pipe diameters, flow rates and difference in the elevation (the network is supplied only by gravity). This relationship will result in a non-convex optimization task and with both discrete and continuous decision variables.

Earlier work on the subject is, for instance, [2], who used a method to linearize the nonlinear relationships and solved iteratively a simplified problem. However, they could not guarantee global convergence or take the non-linear costs into account in a correct way. Other contributions in this direction are [3] and [4]. These contributions also need good starting points (initial guesses) in order to be able to converge to a good solution. [5] used a nonlinear method based on Lagrangian multipliers instead of the linear methods used in earlier contributions. However, they could not guarantee global convergence. Other approaches using nonlinear optimization methods for the proposed problem are [6], [7] and [8], for instance. All suffer from not being able to guarantee global convergence. There are also contributions using genetic algorithms for the proposed problem, i.e. [9] and [10].

Recently [1] presented an MINLP (Mixed Integer NonLinear Programming) model that simultaneously solved the design problem (network lay-out) and the flow optimization. This model is very interesting in that sense and seems to be very promising. However, they could not guarantee global convergence because of the non-convex constraints in the model. The model was solved with a standard convex MINLP solver implemented in GAMS and it is thus, quite sensitive to the initial guesses provided to the optimization routine. The contribution of [1] is an excellent starting point in the quest to find a method guaranteeing global optimality for the proposed problem. We wanted also to find a method that would not require any initial guesses.

The structure of the water distribution network problems addressed in this paper is the same as in the earlier contributions. A small example is found in Figure 1.

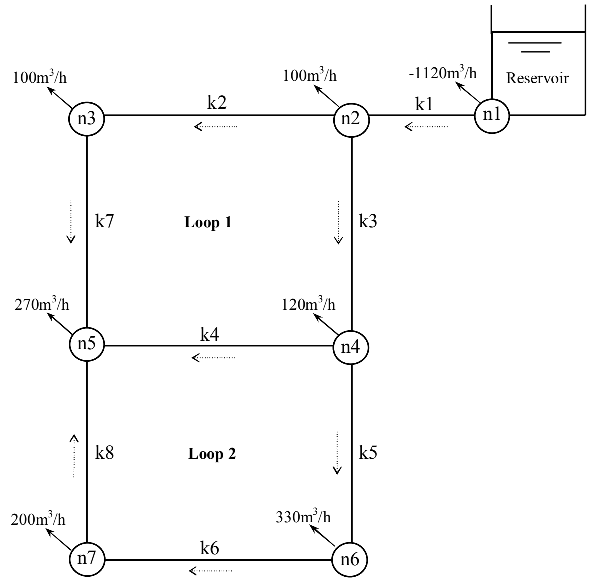


Figure 1. An example structure

The water distribution network includes nodes, n1-n7 in Figure 1 as well as links, k1-k8. The link has a head node and tail node, depending on the initial flow direction, i.e. link k2 has n3 as the head node and n2 as the tail node (in the example given in Figure 1). Each node represents a source or a sink, resulting in certain requirements on the elevation for the nodes (as the distribution network is supplied by gravity).

The paper is organized as follows; first we describe the non-convex MINLP model presented in [1] as well as discuss the problem occurring in the global optimization. Then we present a new model that will be convex and provide a lower bound solution to the original problem. Afterwards we outline the global optimization procedure as well as solve an example problem. Finally, we discuss our results and give an outline for further research.

2. Non-convex Problem Statement (Model M1)

In order to provide the reader with a full description, the model by [1] is presented here in a complete manner. The problem formulation can be stated as follows: The pipe layout and connectivity of the links as well as the pipe lengths, the demand at each node (assumed to be constant) and the minimum head requirement at each node are given. Then, the optimal flow rates and directions, heads at each node as well as the diameter of each pipe (chosen from a set of predefined diameters) need to be determined.

The indices, parameters and sets needed in the non-convex MINLP model (Model M1) are:

Indices
k link
l loop
m diameter
n node

Sets
d commercial diameters
I_n links k entering node n
O_n links k leaving node n

Parameters
T_k total length of link k
c_{km} unit cost per length of link k with the diameter m
D_n demand at node n

The formulation is based on the following key variables:

Binary Variables
Y_{km} 1 if diameter m is present in stream k; 0 otherwise
Z_k 1 if stream k is selected; 0 otherwise

Continuous Variables
Q_k pipe flow in link k, positive if the direction is unchanged (from the original lay-out) and negative if the flow is reversed
ΔH_k head loss in link k, positive if the direction is unchanged (from the original lay-out) and negative if the flow is reversed
H_n head elevation at node n, non-negative

Given these premises, the model (M1) is presented as follows:

The objective function that minimizes the investment costs for the pipelines in the water distribution network is simply given by the linear expression

$$\min \sum_k T_k \sum_m c_{km} Y_{km} \ . \tag{1}$$

Note that the diameters belong to a set of predefined commercial ones, not letting the optimization procedure choose a customized diameter. There is also a need for some linear constraints (c.f. [1], for further details). These are

$$\sum_m Y_{km} = Z_k \ , \qquad \forall k \tag{2}$$

to ensure that at most one diameter is chosen.

$$\sum_{k \in I_n} Q_k - \sum_{k \in O_n} Q_k = D_n \ , \qquad \forall n \tag{3}$$

to ensure that the demand at each distribution node is met and

$$\sum_{k \in Loopl} \Delta H_k = 0 \ , \ \forall l \tag{4}$$

to ensure the difference in altitude has the right sign according to the flow direction. The following constraints will ensure that the right difference in elevation will be obtained when there is a connecting link between the nodes.

$$\Delta H_k = H_{i_k} - H_{j_k} + B_k, \quad \forall k \qquad (5)$$

$$\Delta H_k^{LO} \cdot (1 - Z_k) \leq B_k \leq \Delta H_k^{UP} \cdot (1 - Z_k), \forall k \ (6)$$

Finally, we need to ensure that the elevation level for each node is high enough and that there is no flow if a connecting link is removed by the optimization procedure.

$$H_n \geq H_{n,\min}, \qquad \forall n \qquad (7)$$

$$Q_k^{LO} \cdot Z_k \leq Q_k \leq Q_k^{UP} \cdot Z_k, \ \forall k, \qquad (8)$$

where Q_k^{UP} is a valid upper bound on the flow in link *k*. A good upper bound is found as the total inflow into the network. A valid lower bound, Q_k^{LO}, is then set as -Q_k^{UP}.

There is also a set of non-convex constraints needed in order to establish the relationship between the altitude, the flow rates and the pipe diameters and the fixed pipe lengths. This is given by

$$\Delta H_k = 162.5 T_k \left(\frac{Q_k}{C}\right)^{1.85} \sum_m d_{km}^{-4.87} Y_{km}, \ \forall k \quad (9)$$

where C is the Hazen-Williams coefficient. The presented model (M1) is non-convex due the constraints in the Eq. (9). The most intuitive way to solve the model to global optimality would be to work on the Eq. (9) with some global optimization scheme. One possible way is to use the global optimization method by [11] (also found in [12]) for GGP-problems (Generalized Geometric Programming). This method can ensure global convergence, but it uses variable transformations and approximate relationships between the transformed and original variable introducing a quite large number of binary variables (in each iteration in the global optimization procedure). Due to the extent of variables in the non-convex constraints, this procedure will only solve extremely small instances of the problems according to model M1. There is a need of a method that can solve small to medium sized water distribution network problems to global optimality.

3. Convex Problem Statement (relaxed LB-model, Model M2)

There are two problems with model M1. First, the possibility of reverse flows; second, the non-convex relationship in Equations (9). Both impose non-convexity in the model. Removing these non-convexities requires new additional binary variables and several reformulations (in order to get to the convex M2). First we need some more variables as follows:

Additional Binary Variables
A_k, 1 if the stream *k* is reversed; 0 otherwise

Additional Continuous Variables
$Qabs_k$ absolute value of the pipe flow in link *k,* non-negative
$\Delta Habs_k$ absolute value of the head loss in link *k,* non-negative
ΔHap_k absolute value of the head loss in link *k raised to the power of 1/1.85,* non-negative
$QQabs_{km}$ absolute value of the flow in link *k* for pipe diameter *m,* non-negative

The objective function remains the same in M2 as in M1 (linear)

$$\min \sum_k T_k \sum_m c_{km} Y_{km} \qquad (10)$$

subject to the following constraints (these are also similar as in M1)

$$\sum_m Y_{km} = Z_k \,,\ \forall k \qquad (11)$$

$$\sum_{k \in I_n} Q_k - \sum_{k \in O_n} Q_k = D_n \,,\ \forall n \qquad (12)$$

$$\sum_{k \in Loopl} \Delta H_k = 0 \,, \forall l \qquad (13)$$

$$\Delta H_k = H_{i_k} - H_{j_k} + B_k \,,\quad \forall k \qquad (14)$$

$$\Delta H_k^{LO} \cdot (1 - Z_k) \leq B_k \leq \Delta H_k^{UP} \cdot (1 - Z_k)\ \forall k \qquad (15)$$

$$H_n \geq H_{n,\min} \,,\ \forall n \qquad (16)$$

but with the following additional linear constraints to model that the flow direction is correct

$$Q_k \geq Qabs_k - MA_k \,, \forall k \qquad (17)$$

$$Q_k \leq Qabs_k + MA_k \,, \forall k \qquad (18)$$

$$Q_k \geq -Qabs_k - M(1 - A_k) \,, \forall k \qquad (19)$$

$$Q_k \leq -Qabs_k + M(1 - A_k) \,, \forall k \qquad (20)$$

These constraints are needed to ensure that the variable takes negative values if the stream is reversed (from the initial flow). The same goes for the following constraints

$$\Delta H_k \geq \Delta Habs_k - MA_k \,, \forall k \qquad (21)$$

$$\Delta H_k \leq \Delta Habs_k + MA_k \,, \forall k \qquad (22)$$

$$\Delta H_k \geq -\Delta Habs_k - M(1 - A_k)\ \forall k \qquad (23)$$

$$\Delta H_k \leq -\Delta Habs_k + M(1 - A_k)\ \forall k \qquad (24)$$

In addition, the flows, $Qabs_k$, and the variables $\Delta Habs_k$ need to be set to zero if the link is removed, i.e.

$$Q_k \leq MZ_k \,, \forall k \qquad (25)$$

$$\Delta H_k \leq MZ_k \,, \forall k \qquad (26)$$

The absolute value of the ΔQ_k variable needs to be a sum of the following variables

$$Qabs_k = \sum_m QQabs_{km} \,, \forall k \qquad (27)$$

And each $QQabs_{km}$ needs to be set to zero if the pipe diameter is not used by the following constraints

$$QQabs_{km} \leq MY_{km} \,, \forall k \,, \forall m \qquad (28)$$

The non-convex constraints in M1, Eq. (9) is then rewritten and relaxed to a set of convex constraints according to

$$\begin{aligned} & 162.5 T_k \left(\frac{QQabs_{km}}{C} \right)^{1.85} d_{km}^{-4.87} \leq \Delta Hap_k^{1.85} \\ & \Rightarrow \qquad (29) \\ & (162.5 \cdot T_k)^{1/1.85} \cdot C^{-1} \cdot d_{km}^{-4.87/1.85} \cdot QQabs_{km} \\ & \leq \Delta Hap_k, \forall k, \forall m \end{aligned}$$

Note that the previous equation is now linear. However, we need to enforce the relationship between ΔHap_k and $\Delta Habs_k$ as follows:

$$\Delta Hap_k^{1.85} \leq \Delta Habs_k \,,\ \forall k \qquad (30)$$

,which is now the only non-linear (but convex) equation in the model. The model M2 is a convex MINLP-model that can be solved to global optimum with a standard MINLP-solver, like Alpha-ECP [13] or Dicopt [14], [15] using Outer Approximation [16], for instance. Model M2 is a relaxed version of model M1 due to the inequalities in Eqs (29-30). The relaxation results in a convex MINLP-model as well as a lower bound to the original problem according to M1. However, the lower bound is probably quite tight, since the relaxation is performed in the "right" direction, not leaving the relaxed model to result in very unlikely and poor solutions.

4. The Global Optimization Procedure

Finding the global optimum of the original non-convex MINLP problem (M1) requires, in general, more than just a convex LB (Lower Bound) solution. But the model M2 can be used in a global optimization procedure along with a UB (Upper Bound) solution.

Finding the UB solution is quite simple, when a LB solution is found from M2. In fact, given the solution from M2, fix the design lay-out and flow directions (i.e. fix all binary variables) according to the solution by M2 and then solve the problem M1. This remaining problem will have only a constant in the objective function, only continuous variables but non-convex constraints. However, the solution of this problem is quite simple, due to the simplified structure with the binary variables fixed. In fact, the remaining problem can also be solved as a system of non-linear equations given by Eqs. (3,4,5 and 9) and setting ΔH_k, Q_k and H_n as the unknown variables using the parameters given by the structure of the LB-solution.

If there is no feasible UB-solution, there is a need to start iterating. This iteration can be done, for instance, by putting an integer cut on the previously tested combination of binary variables (as in the Outer approximation algorithm, c.f. [16]) and resolving M2 along with the integer cut preventing the same combination of values of the binary variables to appear. An integer cut is given by (c.f. [16])

$$\sum_{i \in B^k} y_i - \sum_{i \in N^k} y_i \le \left|B^k\right| - 1, \quad k = 1,...K \qquad (31)$$

where $B^k = \{ i \mid y_i^k = 1\}$, $N^k = \{ i \mid y_i^k = 0\}$, $k=1,...K$. (Here k presents the previous iteration result and y all binary variables).

After the integer cut is added and a new LB-solution is obtained, the search for an upper bound solution is performed and so on. The global optimization procedure is outlined in Figure 2.

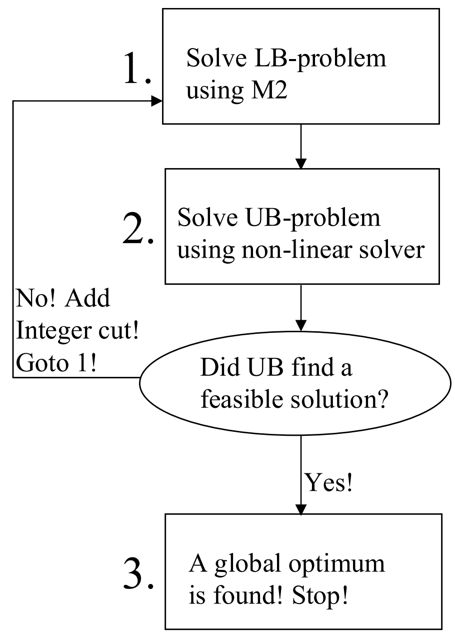


Figure 2. Global optimization procedure

The convergence of the global optimization procedure is evident, since there are a finite number of binary combinations that can be examined and cut off by the integer cuts. If only the UB-procedure is reliable, the global optimum will eventually be found. It is not evident that the method will converge very fast, however. But our experience with this method is that the LB-solution will be extremely tight. In our test example, for instance, we found that the LB-solution (M2) will give the global optimal objective value of M1.

However, the values on the flows and the difference in elevation (i.e. ΔH_k, Q_k and H_n) may not be accurate because of the relaxation of the non-convex constraints. But a possible UB-solution (given the lay-out from the LB-solution) is easy to find just with an Excel-worksheet using the goal seek function. In other words the global optimal solution may be found in the first iteration (in the global optimization procedure). Could it be so that the LB-model will always provide a feasible (and optimal) structure? This question is a future track of research, however.

In order to ensure that the global optimum is always found, the UB-procedure needs to be able to find the solution to the UB-problem. In our case example, this was done easily with any nonlinear solver. We used the solver implemented in Microsoft Excel. Since the UB problem is a non-convex NLP problem, it can be argued that the solver by Microsoft Excel will not find the global optimal solution to the UB-problem. However, the UB-problem can be solved with a global optimization method, such as the one presented in [11], [12] and [18]. The method presented in [11], [12] and [18] tackles Generalized Geometric Programming (GGP) problems with continuous and integer variables. This method will always find the global optimum to the UB-problem and thus, if this method will be used to solve the UB-problem, the global optimum will be guaranteed for the water distribution network problem presented in this paper. Since the global optimization method has convergence proof for all GGP problems, it could be argued that the original problem formulation by [1] could be solved directly with the actual global optimization method, presented here to solve only the UB-problem. General global optimization methods have often slow convergence and the initial test showed that the even the small test example in the next section could not be solved with the problem formulation by [1] and the method presented in [11], [12] and [18] in a decent time (probably not even in months). The UB-problem is much smaller than the entire problem (the UB-problem has no integer variables, except from possible integer cuts), and can be solved within decent time, and thus the global optimal solution to the water distribution network problems can be found. The UB-procedure is under further investigation. A more formal presentation and discussion would be adequate and interesting, but it is out of the scope of this paper.

5. Example

In this section, an example of a network consisting of 7 nodes is given. The problem is well known and used in [1] as well as in [8] and [10]. The problem originates from [2]. Even if the problem is small, the global optimal solution has not been established (to the authors' knowledge). The problem is often used as a test example for algorithm development and therefore suitable for being used here as the example to test the new global optimization procedure. In addition several solutions in the literature are given to *approximate* problems with relaxed constraints, and thus some solutions provided in the literature may be better than actual global optimum. A typical constraint that is often relaxed is the $H_n \geq H_{n,\min}$ constraint, where the head elevation is often assumed to be less than 30 m above the demand elevation. The minimum flow in a pipe is assumed to be -1120 m^3/h and the maximum flow in a pipe is assumed to be 1120 m^3/h (c.f. Figure 1).

The problem in [2] includes 7 nodes, 8 links and two loops. It has only one source node (node 1) of fixed head altitude of 210 m. The pipe lengths are all fixed at 1000 m have a coefficient, C, of 130. The demands and the head altitude at each node are given in the Appendix. The head has to be at least 30 m above the ground elevation, however. The costs for the commercially available diameters are given in the Appendix as well as the global optimal solution.

The most recent solution obtained for this problem by [8] and [10] as well as [1] had an objective value of 419000. This solution seems to be a good solution, but not proven to be globally optimal. Both [17] and [7] reported objective values less than 403000, but with approximate models. There is obviously not clarity over the best solution in this example. Therefore we solved this problem with the new method presented in this paper to global optimality for $\Delta H\min$ =30, 25, 20, 15, and 10 respectively. The requirements on the $\Delta H\min$ parameter are sometimes relaxed in the literature in order to obtain better solutions. The solutions for the different values of the $\Delta H\min$ parameter are given in Table 1 (all test runs terminated in the first iteration in the global optimization procedure). The LB-problems (model M2) has been solved with the αECP-method [13], whereas the UB problems has been solved with the goal seek function in Microsoft Excel.

$\Delta H\min$	Global optimal objective value
30	419000
25	376000
20	336000
15	306000
10	290000

Table 1. Global optimal objective values for the test examples for different $\Delta H\min$

It is worth noticing that for $\Delta H\min = 30$ (the basic assumption for the problem) the actual global optimal solution has an objective value of 419000 as given by [8] and [10] as well as [1]. The global optimal solution is also given in detail in the Appendix (for $\Delta H\min = 30$). It is worth pointing out that the global optimum decreases quite rapidly when the $\Delta H\min$ parameter is changing. This is interesting for actual applications in this field; the designers of the water distribution networks need to be motivated to find the smallest possible $\Delta H\min$ as well as investigate the possible savings of reducing the $\Delta H\min$. Of course, the designers need to evaluate the trade-off between the savings in the piping costs and extra costs for decreasing $\Delta H\min$.

6. Discussion

In this paper, a new global optimization method has been presented for the water distribution layout problem supplied by gravity. The new proposed method uses a tight LB-model, which can be solved to global optimality using standard MINLP-methods (such as the αECP-method). This LB-solution seems to be very tight and may provide the actual global optimal objective value for the original non-convex problem (model M1). This is not evident, however, and therefore a global optimization procedure is outlined in section 4. The authors of this paper have not been forced to iterate in the global optimization procedure in order to solve the example problem. An UB-procedure is also discussed and used to obtain the right variable values (for the global optimum) and verify the solution of the LB-procedure.

The example solved in this paper is well-known and a main theme of research in several contributions. Despite the focus this specific problem has obtained, no one has proven a global optimal solution to the problem. The method proposed in this paper was used to solve it to global optimality with several values on the minimum elevation difference, the parameter $\Delta H \min$. Interestingly, the global optimal value has been proposed by several authors (without the knowledge of it being the global optimal solution). Other authors that have presented objective values less than the global optimal value have used approximate methods or relaxed the $\Delta H \min = 30$ constraint. Therefore a sensitivity study of the $\Delta H \min$ was made to show that this parameter affects the global optimal value significantly.

Further research in this area would be to investigate whether the LB-solution will always provide a feasible network-structure. This would need further attention and a proof would be required. Also bigger water distribution networks would need to be examined.

7. References

[1] Sharma R. (2001) "Optimisation in the Water Industry" MEng Thesis, University College London

[2] Alperovits, E., and Shamir, U. (1977). "Design of optimal water distribution systems." *Water Resour. Res.*, **13**, 885-900.

[3] Goulter, I. C. and Morgan, D. R. (1985). "An integrated approach to the layout and design of water distribution networks." *Civ. Engrg. Syst.,* **2**, 104-113.

[4] Kessler, A., and Shamir, U. (1989). "Analysis of the linear programming gradient method for optimal design of water supply networks." *Water Resour. Res.*, **25**, 1469-1480.

[5] Fujiwara, O., and Khang, D. B. (1990). " A two-phase decomposition method for optimal design of looped water distribution networks." *Water Resour. Res.*, **26**, 539-549.

[6] Sonak, V. V., and Bhave, P. R. (1993). "Global optimum tree solution for single-source looped water distribution networks subjected to a single loading pattern", *Water Resour. Res.*, **29**, 2437- 2443.

[7] Eiger, G. *et al.* (1994). "Optimal design of water distribution networks." *Water Resources Research*, **30**, 2637-2646.

[8] Cunha, M. C., and Sousa, J. (1999). "Water distribution network design optimization: simulated annealing approach." *J. Water Resour. Plng. And Mgmt.*, ASCE, **125**, 215-221.

[9] Gupta, I. *et al.* (1999). "Genetic algorithm for optimization of water distribution systems." *Environmental Modelling & Software*, **14**, 437-446.

[10] Savic, D. A., and Walters, G. A. (1997). "Genetic algorithms for least-cost design of water distribution networks." *J. Water Resour. Plng. Mgmt.*, ASCE, **123**, 67- 77.

[11] Pörn R., Björk K-M. & Westerlund T., (2007). Global Solution of Optimization Problems with Signomial Parts. Accepted to *Discrete Optimization.*

[12] Björk K-M., (2002). A Global Optimization Method with some Heat Exchanger Network Applications. Dr-thesis (Tech.), Abo Akademi University, ISBN 952-12-1067-2.

[13] Westerlund, T. and F. Pettersson, (1995) "A Cutting Plane Method for Solving Convex MINLP Problems", *Computers and Chemical Engineering*, **19**, S131-S136.

[14] Brooke, A., Kendrick, D. and Meeraus, A., (1992) "GAMS - A User's Guide", Scientific Press, Palo Alto.

[15] Brooke, A., Kendrick, D. & Meeraus, A. (1998). *GAMS: A user's guide.*; San Francisco: The Scientific Press.

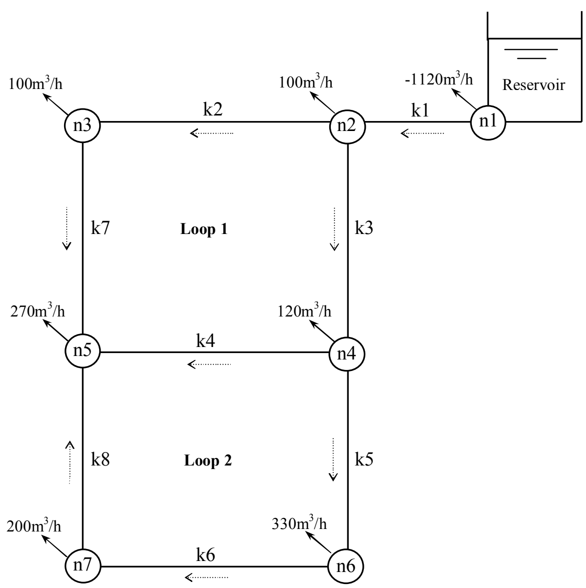
[16] Duran, M.A. and I.E. Grossmann, (1986) "An Outer-Approximation Algorithm for a Class of Mixed-integer Nonlinear Programs," *Math Programming* **36**, 307.

[17] Kessler, A., and Shamir, U. (1991). "Decomposition technique for optimal design of water supply networks." *Engrg. Optimization*, **17**, 1-19.

[18] Björk K-M., Lindberg P.O. and Westerlund T. (2002). "Some Convexifications in Global Optimization of Problems Containing Signomial Terms" *Computers and Chemical Engineering* **27**, 669-679.

Appendix

The data for the problem in [2] is given in this Appendix. The structure and the demand of each node are given in the following figure. Note that the source demand is negative.



The cost data and the dimension data for all available commercial pipes are given in the following table:

Diameter (*in*)	Cost (units)
1	2
2	5
3	8
4	11
6	16
8	23
10	32
12	50
14	60
16	90
18	130
20	170
22	300
24	550

The demand and elevation at each node is given in the following table

Node	Demand (m^3/h)	Elevation (m)
1 (source)	-1120.0	210.0
2	100.0	150.0
3	100.0	160.0
4	120.0	155.0
5	270.0	150.0
6	330.0	165.0
7	200.0	160.0

The global optimal solution for the problem is given in the following table

Flow (m^3/h)	ΔH (m)	Stream	Diameter (*in*)
1120.0	6.7	1	18
336.9	12.8	2	10
683.1	4.8	3	16
32.5	14.6	4	4
530.6	3.0	5	16
200.6	4.9	6	10
236.9	6.7	7	10
0.6	6.7	8	1