Hybrid Method for Optimal Design of Water Distribution System

Milan Čistý, Zbyněk Bajtek

Slovak University of Technology, Faculty of Civil Engineering, Department of Land and Water Resources Management, Bratislava, Slovak Republic, e-mail: milan.cisty@stuba.sk.

Abstract. Optimal design of water distribution networks belong to large combinatorial optimization problems. Various methods have been developed and applied for the design of water distribution systems. Still there is some uncertainty about finding a generally reliable method. In the proposed method two algorithmic techniques are used – linear programming (LP) and genetic algorithm (GA). A combination of these two methods with the aim of eliminating the limitations of LP (which is not suitable for networks with loops) is used. It was investigated that proposed method provides results more reliable in terms of closeness to a global minimum. The method may be applied for use in the design and rehabilitation of both drinking water and pressurized irrigation systems.

Keywords: Water distribution system, optimization, heuristic method, linear programming

1 Introduction

The optimal design problem is to find the water distribution systems component characteristics (e.g., pipe diameters, pump heads, reservoir volumes) that minimize the total system cost. Various algorithms ranging from artificial intelligence to the optimization domain were applied. Alperovits and Shamir [1] presented a linear programming gradient for optimizing a water distribution network. Kessler and Shamir [12] used the linear programming gradient method as an extension of this method. It consists of two stages: an LP problem is solved for a given flow distribution, and then a search is conducted in the space of the flow variables. Later, Fujiwara and Khang [9] used a two-phase decomposition method extending the method of Alperovits and Shamir to non-linear modeling. Also, Eiger, et al. [6] used the same formulation as
Kessler and Shamir. Nevertheless, these methods fail when solving problems of large looped systems. The application of stochastic or so-called heuristic optimization methods to WDN optimization can be traced back to the late 1990s and they still remains as a vital tool for WDN optimization. Simpson, et al. [20] used a simple genetic algorithm in which each individual solution from the population of solutions is represented by a string of bits with identical lengths. The simple GA was then improved by Dandy, et al. [5] using the concept of the variable power scaling of the fitness function, an adjacency mutation operator, and gray codes. Savic and Walters [18] also used a simple GA in conjunction with an EPANET network solver. Many different heuristic techniques have been applied to the optimization of a water distribution system. Methods as simulated annealing [13], [4]; an ant colony optimization algorithm [14]; a shuffled frog leaping algorithm [8] and a harmony search [10]. Reca, et al [17] evaluated the performance of several meta-heuristic techniques. He compared these techniques by applying them to medium-sized benchmark networks. The results which he obtained for the Hanoi network (after ten different runs with five heuristic search techniques) varied from 6,173,421 to 6,352,526. These results differ by 1.5 – 4.5% from the known global optimum for this task, which is a relatively large deviation for such a small problem. It should be noted that real life problems never get the privileges afforded the Hanoi or other benchmark network.

How closely any solution from the computation of the optimization in a particular case could achieve a global optimum depends on the complexity of the problem. Generally, a more complex problem is a problem for which a larger or more complicated search space must be explored. The reduction of the search space is the main approach of this paper on how to design a method which has a greater potential to find results which are really close to a global optimum. A new hybrid GA-LP approach using a genetic algorithm and linear programming is proposed in this study for determining the least-cost design of a water distribution system. It is built on the advantages of both the deterministic and heuristic methods. The GA method is used in the outer loop of the proposed algorithm, which is intended for decomposing a complex looped network in a group of equivalent branched networks. LP is then used in an inner loop to solve each branched network and provide a minimum-cost design. After evaluating a high number of possible branched networks, an optimal solution is found for the original looped network.

The advantage of using this hybrid method consists in the fact that a GA in this case has a much smaller searching space than in a case when usually formulation of heuristic search (for example GA methodology) is used alone, which has a great impact when trying to achieve better results. How this is achieved will be explained later after introducing the details of the methodology.
2 Methodologies

2.1 Linear Programming - Deterministic Part of Proposed Method

The linear programming method has long been accepted as an approach for the optimal selection of diameters for pipes in branched networks, e.g., in the design of irrigation systems. The mathematical formulation of this problem is as follows:

\[ X_{11} + X_{12} + \ldots + X_{1n} = B_1 \]
\[ X_{21} + X_{22} + \ldots + X_{2n} = B_2 \]
\[ \text{etc.} \]
\[ X_{m1} + X_{m2} + \ldots + X_{mn} = B_m \] \hspace{1cm} (1)
\[ A_{11} X_{11} + A_{12} X_{12} + \ldots + A_{xy} X_{xy} \leq C_1 \] \hspace{1cm} (2)
\[ \text{etc.} \]
\[ D_1 X_{11} + D_2 X_{11} + \ldots + D_j X_{nn} = \min \] \hspace{1cm} (3)

The solution has to comply with inequalities:

\[ X_{11} > 0; X_{12} > 0 \text{ etc. up to } X_{nn} > 0 \] \hspace{1cm} (4)

where:
- \( X_{ij} \) unknown length of the selected diameter \( j \) on section \( i \). The allowable diameters must satisfy the velocity conditions on the section,
- \( B_i \) total length of the section,
- \( C_i \) allowable total loss for the constraint – described below,
- \( D_j \) unit price of a pipeline with the diameter number \( i \).

When linear programming is applied in order to solve the optimal design of pipeline networks, the unknown will be the lengths of the individual pipeline diameters on the section. In conditions (Eq. 1) the requirement that the sum of the unknown lengths of the individual diameters in each section has to be equal to its total length is enforced. The second type of equations - constraints (Eq. 2) - represents the condition that the total pressure losses in a hydraulic path between a pump station or tank and every critical node (the end of the pipe network; the extreme elevation inside the network) should be equal to or less than the known value. This constraint is based on the minimum network pressure requirements needed for the operation of the system. There should be the same number of these constraints as there are critical nodes in the network. Given the minimization requirement for the investment costs, the objective function (Eq. 3) is the sum of the products of the individual pipeline unit prices and their required lengths. As will be described later, this system of the equation should be compounded automatically in the fitness function of the GA.

When multi-demand conditions in the LP model are incorporated, there will be a system of constraint (Eq. 2) for every demand pattern. When pumps are also included in a model, the main input parameter for a pump is its pump curve. The right sides of the constraints (Eq. 2) vary according to the pump’s operating conditions. The head of the pump may be treated as an unknown variable.
2.2 Genetic Algorithms - Heuristic Part of the Proposed Method

In order to overcome the above-mentioned deficiencies of the linear programming techniques, heuristic optimization techniques have been introduced for solving the optimization of water distribution systems. Firstly, a genetic algorithms methodology, which is also used in this study, was applied. This is a search procedure inspired by the mechanics of natural genetics and natural selection. Its basic concepts are briefly summarized below; a good introduction to the subject is given by Goldberg [11].

The first step is to represent a solution to the problem by a string of genes that can take on some value from a specified finite range. This string of genes, which represents the solution, is known as a chromosome. Then an initial population of chromosomes is constructed at random. Genetic algorithms are implemented as a computer simulation in which a population of chromosomes evolves toward better solutions by means of genetic operators such as inheritance, mutation, selection, or crossover. At each generation, the fitness of each chromosome in the population is measured. The fitter chromosomes are more often selected probabilistically to produce offspring for the next generation. This process is repeated until some form of convergence in the fitness is achieved.

2.3 Hybrid GA–LP Approach to the Optimal Design of Water Distribution Systems

The proposed method is based on a combination of linear programming methodology and genetic algorithms. The main reason is that linear programming always finds the global optimum if it exists. But because LP is suitable only for solving branched networks, the GA method is used for decomposing a complex looped network into a group of branched networks. These branched networks are characterized by hydraulic behavior identical to a looped network on the condition of having identical diameters on the corresponding sections of the network. Identical hydraulic behavior means that there are identical flows in the corresponding pipes and identical pressures in the corresponding nodes in the original looped network and branched networks investigated. The decomposition of a looped network means that before the optimization, every loop is split in some demand node (or node in which a branch is connected) which is part of this loop. Linear programming is then applied for optimizing every branch network produced by GA from the original looped network, and GA is simultaneously applied for the evolution of the best splitting option.

The main aspect of the motivation to propose the new optimization algorithm presented is that the definition of a chromosome is in an obvious GA approach as long (in the sense of the number of genes) as many pipe sections are in the water distribution network. The definition of a chromosome proposed herein produces significantly shorter strings (which mean a smaller search space) for most networks because the number of genes will be the same as the number of loops in the network and not equal to the number of pipes. This means easier searching with better results.

An example of possibilities for splitting a loop is shown in Fig. 1. Let us suppose that all the parameters of the network in Fig. 1a) are already known (diameters, lengths, demands, flows, etc.), so we are not in the design stage for this network, but
we can suppose that it already exists. The loop can be transformed to a branch layout without affecting the hydraulic behavior of the network (the flows and pressures remain the same) if the loop is split in the demand node in which the flows are entering from the two pipes connected to this node. There is one such node for every loop in the network. This node could be a hydrant or branch connection. In Fig. 1a) the original loop with a node in which there is a demand 5 lps is shown. So this is the node in which it is possible to break the loop without affecting the hydraulic behavior of the network in the current conditions. A loop could be split in this node in such a way that a “twin” node is introduced to the network; the identical elevation on it is assumed as in the original node (and de facto identical position). The original demand has to be split between these two nodes: the upper node will have in the first case demand set to 4 lps and the lower node at 1 lps (Fig. 1b).

The original demand in every split node selected has to be divided between the original node and its “twin node” according to some rule with a rational number of alternatives (e.g., Fig. 1b and 1c). The number of alternatives depends on the demand rate – a high demand in a node requires more alternatives. For networks with more loops, a combination of alternatives for every loop should be considered. Because we are now talking about designing the network, it is necessary to propose diameters for every alternative of the branch network which we get by this procedure. This is accomplished by LP. The cheapest one is an optimum which an algorithm is searching for. The search for the best possibility of splitting the loops is guided by the GA, which proposes one possible branch network for every chromosome in which both the splitting nodes and parameters for splitting the demand in these nodes are coded. The branch network (or this chromosome) is evaluated by accomplishing its design by LP, which is nested in the fitness function of the GA. This represents single iteration of the algorithm. The formulating of the chromosome and establishing the so-called Loop Links matrix (LL) and Loop Splitting Options matrix (LSO), both of which help to organize searching for the best loop-splitting locations on the network, and finally the R and R_s parameters for splitting the demands in the nodes where the loops are divided. There is the same number of genes as the number of loops in a
network. Genes are coded as integer numbers, which indicate a row number in an LSO matrix. This matrix defines all the splitting possibilities for every loop. The second half of the genes in a chromosome is also coded as an integer value, and this value defines the ratio by which the demand in the original split node should be divided between the original node and its “twin node” after splitting the particular loop in its location. The proposed algorithm in its basic form is summarized in following diagram on

**Fig. 2.** The scheme of the GALP algorithm

In the last part of this section two remarks on the details of the proposed algorithm follow:

*Firstly*, one more condition should be added to the LP model in the case of using it in the context of solving looped networks.

The principle of the conservation of energy dictates that the difference in energy between two points must be the same regardless of the path that is taken. Thus, the difference in energy at any two points connected in a network is equal to the energy gains from the pumps and the energy losses in the pipes and fittings that occur in the path between them. This equation can be written for any open path between any two points. Of particular interest are paths around loops because the changes in energy must sum to zero. A linear programming model must also pay attention to this principle, so its formulation, which is expressed by formulas 1-4 for the branched network
only, have to be expanded by the loop condition in the proposed methodology; however each one of them was split into two branches by the described method. There should be the same pressure in the node in which the loop was breaking and in its twin node after proposing the diameters in the network. Therefore, while the LP model is being built, the following conditions should be added, which ensure that the same pressure must be in the original and corresponding dummy twin nodes:

\[ E_i = A_{11} X_{11} + A_{12} X_{12} + \ldots + A_{x y} X_{xy} \]  \hspace{1cm} (5)

\[ E_{ti} = A_{11} X_{11} + A_{12} X_{12} + \ldots + A_{xy} X_{x} \]  \hspace{1cm} (6)

\[ E_i - E_{ti} = 0 \]  \hspace{1cm} (7)

where:
- \( E_i \) and \( E_{ti} \) are energy losses from the source (pump, reservoir) to the split node on loop \( i \) and to its twin node,
- \( A_{mn} \) is hydraulic loss in section \( m \) and diameter \( n \), which belongs to the specified path,
- \( X_{ij} \) is unknown length of the selected diameter \( j \) on section \( i \).

The second remark on the details of the proposed algorithm relates to its compatibility with EPANET. In an obvious simulation-based GA approach a hydraulic network solver handles the pressure and velocity constraints and simultaneously evaluates the hydraulic performance of each trial solution. The most commonly used simulation model to analyze the network in such a manner is EPANET. In the proposed method simulation by EPANET (or another hydraulic engine) is not incorporated in the core of the objective function (or fitness function), but linear programming is used instead. But for the sake of compatibility with other optimization models, the computation of the friction head losses \( A_{mn} \) (Eq. 2) is executed by EPANET by calling its functions through the EPANET Toolkit [18].

### 3 Results

The performance of this model was evaluated by the optimization of the well-known Hanoi network and triple Hanoi water supply networks. The first problem is taken from the literature and the second problem was introduced by the authors for the sake of evaluating the proposed method for greater problem. The global optimum for the Hanoi network is known because the long usage of this benchmark in the optimization methods development community. The triple Hanoi network is derived from it in such a way that her global optimums could also be evaluated. On this basis it is possible to compare the results obtained in testing runs with the known global optimums. The water distribution trunk network in Hanoi Vietnam was first introduced by Fujiwara and Khang [9]. The parameters of this network are well known and can be found in the work of Fujiwara and Khang [9]. The network consists of three loops, so three \( LSO \) matrices were determined by the algorithm. Roulette wheel selection was used to choose the parents for the next generation. A one-point crossover was used because of the relatively short chromosome, and the probability of the selected pair of strings being subjected to the crossover operator was taken as \( p_c=0.9 \). The mutation
rate was set to be $p_m = 0.1$. A very simple penalty function was used: if the LP the final producer of every partial solution) does not find a solution (which could happen in some configurations), the algorithm gives this solution a significantly higher cost than the highest cost in the previous generation. That is why the run-time plots and other statistical evaluations of the computation will be described hereinafter only for the triple Hanoi network, which is larger (100 pipes).

In the triple Hanoi water distribution network all the corresponding parameters for the nodes and lines are the same as in the original Hanoi network on all three (single Hanoi network) parts except for four pipes, the head in the reservoir and the demand in one node. These changes were made for the sake of obtaining the same pressure in nodes 3, 33 and 63 (with a diameter of 1016 mm on pipes 1, 2, 35 and 68, which will certainly be proposed here by any optimization method because of the large flow in them) as in the original Hanoi network in node 3. In such conditions the same diameters should be the optimal solution for the corresponding pipes as in the original Hanoi network. These are the changes mentioned: the head in the reservoir is set to 105 m; the length of pipe 1 is 1 m; the length of pipe 2 is 1786.50 m; and the lengths of pipes 35 and 68 are 1641.69 m. In junction 3 the demand is equal to zero. Under these conditions the \textit{reference} (global) optimal solution of the triple Hanoi network could be evaluated as follows:

$$C_{TH} = 3 * C_H - 3 * L_1 * C_1 - 3 * L_2 * C_1 + (1 + 1786.5 + (2 * 1641.69)) * C_1.$$ (8)

where:
- $C_{TH}$ optimal cost of the triple Hanoi network,
- $C_H$ reference optimal cost of the Hanoi network ($6,057,744$),
- $L_1$ length of the first pipe on the original network (100m),
- $L_2$ length of the second pipe on the original network (100m),
- $C_1$ unit price of diameter 1016 mm ($278.28$).

For our solution, which is the best solution of the basic Hanoi network compared to the results known to the author, this means that the optimal solution of the triple Hanoi network should be $18,373,697.49$. A comparison of the results obtained by the proposed method (which can propose two diameters on some sections) with those published in the literature and obtained by a discrete diameter design (only one diameter is proposed for a section) is a bit problematic, and it is necessary to accomplish it comparatively to the corresponding global minimums for the discrete and split pipe designs. A discrete diameter design has somewhat limited possibilities in comparison with a split pipe design, which can propose two diameters for some sections of the network from the point of view of the resulting cost. Discrete diameter solutions for this reason always cost somewhat more than split pipe design solutions. That is why the \textit{percentual differences} were computed from the reference global minimum - in the case of GALP it is the reference global minimum for a split pipe design, and in the case of a discrete diameter design, another reference global minimum will be taken as the basis.

So we will take as a reference (the best) discrete diameter design solution obtained by [10]. This author used the harmony search method and found a feasible solution at a cost of $6,081,087$, and it is the best feasible solution for a discrete diameter design which has been published in the literature. Only solutions which are feasible in terms of the allowable nodal pressures computed by the EPANET network solver are taken into consideration. When this is taken as the reference (optimal) discrete diameter
design of the Hanoi network, the reference (global) optimal discrete diameter solution for the triple Hanoi network should be $18,443,867.49.

Fig. 3. Triple Hanoi network

The best results obtained with GALP, the GA optimization model OptiDesigner and the HSNet model based on the Harmony Search methodology are summarized in Table 1. The results in Table 1 demonstrate that the proposed GALP methodology gives significantly better results in terms of closeness to the global minimum. This is a consequence of the reduction of search space for GA in a GALP context. There are three loops in the basic Hanoi network. This means that the chromosome for the Hanoi network consists of 6 genes. There are 14 possibilities for splitting the first loop, 8 for the second and 7 for the third. The R_s factor used was set to be 0.1, so there are 9 possibilities for splitting the demand in the split node in every loop (01; 0.2; ...; 0.9). This means that there are $14 \times 8 \times 7 \times 9^3 = 571,536$ possibilities of which the search space of GALP for the basic Hanoi network consists. The search space of the GA for the same problem is $6^{10} = 2.86512 \times 10^{26}$, which is significantly greater. In the case of the triple Hanoi network there are $571,536^3 = 1.86694 \times 10^{17}$ possibilities for GALP and $6^{100} = 6.53319 \times 10^{77}$ possibilities for GA. In this case it can be seen that a three times greater problem has a smaller search space when using GALP in comparison to when GA is used alone.
GALP is the method proposed by the authors of this paper; references to other methods are GA [2], Harmony search [10] and OptiDesigner – http://www.optiwater.com/optidesigner.html. Other methods were also tested and compared to GALP, but they provide similar or worst results than those in the Table 1.

Table 1. Comparison of the best results when applying various methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Hanoi</th>
<th>triple Hanoi</th>
<th>Deviation from reference global optimum [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>GALP</td>
<td>6,057,697*</td>
<td>18,394,255</td>
<td>0.04</td>
</tr>
<tr>
<td>GA</td>
<td>6,081,087</td>
<td>19,269,160</td>
<td>4.47</td>
</tr>
<tr>
<td>OptiDesigner</td>
<td>6,115,055</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Harmony search</td>
<td>6,081,087*</td>
<td>18,839,302</td>
<td>2.67</td>
</tr>
</tbody>
</table>

*reference global solution
** a feasible solution was not found

4 Conclusions

The design of an optimal water distribution network is a complex task. Various deterministic and heuristic algorithms have been proposed and attempted for solving this problem.

Authors have proposed a method in which heuristic algorithms (the genetic algorithm) are incorporated, but the final solution is produced by linear programming. This method is described in the paper and was successfully tested on the benchmark networks. It was determined that the method gives results that are more trustworthy in terms of closeness to a global minimum. This is because of the fact that involving LP in an algorithm reduces the search space for heuristics very dramatically on most network configurations. Although one iteration of the proposed method is a little more computationally intensive than in the case of using only heuristic algorithm, this reduction is finally more important for the effectiveness of the algorithm.

The method is proposed as an alternative to existing methods, mainly when a least-cost design is to be solved. Its extension to solving multi-objective tasks should be evaluated in future research. Nevertheless, the authors suppose it should be tested too, if it is not better to use a more precise method for this single objective task and then refine the output from the least-cost design according to other criteria (or with other demand pattern) by applying a simulation model. The design is always multi-objective, but that does not mean that this multi-objectivity must be covered in one computation with too many objectives (and less precision). That is why the author believes that this method can also be practically useful also in the stage described and tested in this paper.

In this work only the basic GA method was used in the stochastic part of the algorithm. It gives good results, but there is a possibility open to replace it with some of the other and more effective heuristic methods which are available in the optimization.
community. The author expects that this can even refine the method in the future. The effect of such a refinement will mainly be revealed when significantly larger networks than those tested here will be solved.

Acknowledgement

This study was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, Grant No. 1/0496/08 and 1/0585/08 and by the Slovak Research and Development Agency, Grant No. APVV-0443-07.

References