IOP Conf. Series: Materials Science and Engineering **324** (2018) 012019 doi:10.1088/1757-899X/324/1/012019

Optimization of the graph model of the water conduit network, based on the approach of search space reducing

Iakov S. Korovin[1] and Maxim G. Tkachenko[1]

1 Southern Federal University, Taganrog, Russia
E-mail: Korovin_yakov@mail.ru

Abstract. In this paper we present a heuristic approach, improving the efficiency of methods, used for creation of efficient architecture of water distribution networks. The essence of the approach is a procedure of search space reduction the by limiting the range of available pipe diameters that can be used for each edge of the network graph. In order to proceed the reduction, two opposite boundary scenarios for the distribution of flows are analysed, after which the resulting range is further narrowed by applying a flow rate limitation for each edge of the network. The first boundary scenario provides the most uniform distribution of the flow in the network, the opposite scenario created the net with the highest possible flow level. The parameters of both distributions are calculated by optimizing systems of quadratic functions in a confined space, which can be effectively performed with small time costs. This approach was used to modify the genetic algorithm (GA). The proposed GA provides a variable number of variants of each gene, according to the number of diameters in list, taking into account flow restrictions. The proposed approach was implemented to the evaluation of a well-known test network - the Hanoi water distribution network [1], the results of research were compared with a classical GA with an unlimited search space. On the test data, the proposed trip significantly reduced the search space and provided faster and more obvious convergence in comparison with the classical version of GA.

1. Introduction

Optimization of the structure of loop water distribution networks is a complex combinatorial problem, related to NP-hard (Non-defintistic Polynomial-time) class, since it is nonlinear, bounded, nonsmooth, nonconvex, and therefore multi-modal [2]. Although methods of mathematical programming such as linear and nonlinear ones are known to be applied to solve this problem [2-5], heuristic methods are currently relevant because of their ability to obtain quasi-optimal solutions for solving large-dimensional problems in conditions of limited working time. Genetic algorithms (GA) and a number of other algorithms are among the most common heuristic approaches used to design waterway networks. GA are widely used to create an effective net of conduits [2,3]. And also it is generally known that GA are based on the rules of evolution and natural selection.

Although heuristic approaches can cope with global optimization problems, they do not always provide finding the optimal solution. In addition to the apparently low possibility of task solved, another disadvantage is time, required for convergence of the method. Despite a significant number of works on optimization of convergence processes, heuristic methods are still relatively inefficient and take a lot of time when working with large graph models, including modeling the water distribution network. The low search efficiency is due apparently to the wide range of possible solutions that the algorithm should test. Since the search space is very large, general-purpose heuristic algorithms spend a significant amount of time evaluating unrealizable solutions. Consequently, the probability of finding the optimal

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

IOP Conf. Series: Materials Science and Engineering **324** (2018) 012019 doi:10.1088/1757-899X/324/1/012019

solution decreases and the rate of convergence increases as the size of the search space increases. Thus, the actual task in this area is to develop methods for reducing the search space.
In this paper, we propose an approach to improve the efficiency of heuristic methods, used to optimize in the graph representation of the task of designing water distribution networks. The proposed approach is based on the restriction of the search space on the basis of the analysis of two opposite scenarios, the formation of a flow in the network with subsequent consideration of the constraints imposed on the flow velocity in the edges of the graph of the simulated network. The proposed approach is compared with the classical formulation of GA with an unlimited search space when analyzing data in the model of the Hanoi water distribution network (HWDN).

2. Formation of constraints in the graph model

In normalized graph model of water conduits (models without cycles), the flow distribution can be estimated by applying the equations of flow conservation in network nodes. From a practical point of view, the general procedure for this type of net is to determine speed limits of the flow. The limitations of the speed of the pipelines may vary depending on the material, diameter, and other considerations. Thus, sufficient flow rates can lead to increased pipe damage and significant energy losses. Low speeds, on the contrary, can lead to deposits and an increase in the total metal capacity and volume of the system. When speed limits are defined, the range of possible diameters that can be chosen for each edge of the graph is significantly reduced, which reduces the sophistication of net design. Nevertheless, in difference from normalized models, the flow distribution in loop networks is not known a priori and, as a consequence, this approach can not be applied.
The approach presented is built upon a reduction of the search space procedure by limiting the range of possible diameters that can be chosen for a specific edge of the net graph. The procedure consists of creating two opposite extreme flow propagation scenarios that satisfy the equations of nodal flow conservation and core requirements. The first scenario provides the most general distribution of the flow in the network that would satisfy the nodal requirements and flow restriction constraints. The result scheme will provide a network with high reliancy. The second (contrary) scenario is characterized by the highest accumulation of flow in several major pipes. This situation provides a thread distribution similar to that obtained for the spanning tree of the network graph.
The approach depicted in this paper, applied for calculating the distribution of extreme flows is the solution of the problem of optimization of quadratic functions (OQF) for each of them. The OQF problem is in general connected with the minimization or maximization of a quadratic function bounded by linear functions.
The objective function of the proposed OQP problem for the general distribution of the flow is to minimize the sum of the fluxes of squares of the fluxes in the edges of the network graph. These flow values must satisfy the law of flux conservation at network nodes. This set of constraints is linear. In general, the problem is presented by the equation below (1):

$$\sum_{i=1}^{n} Q_i^2 \to \min, \quad A \times Q = q, \quad Q \geq 0 \tag{1}$$

where: n is the number of edges in net graph, Q_i is the stream in the edge i, A is the matrix (m x n), and m is the number of vertices of the graph, and q is the constraint vector. The element of matrix $A\ [i, j]$ is equal to 1 if the stream from vertex i is directed to vertex j, -1 if it is directed in the opposite direction, and 0 if vertex j is not connected with vertex i.
One of the drawbacks of this approach is that the direction of the flows must be determined in advance to perform the calculation. For complex networks, the number of possible combinations of the flow direction can be high, and the search for the correct version is a complex procedure. To delete this limit, it is proposed to double the number of edges of the graph by adding a dummy channel for each network line in such a way that for each network line two pipes with opposite flows are considered. Using this

procedure, the number of edges in the graph increases to *2n*, and the matrix *A* will have dimension (*m x 2n*). The solving of the OQF minimization task provides the correct flow directions and values that minimize the amount of network flows. This gives the distribution of the flow with the maximum variance (MD). The second scenario with the maximum accumulation of the flow, also called the maximum concentration scenario (MC), is obtained by maximizing the objective function and solving the equivalent problem of maximizing the OQP. The solution of these tasks defines 2 vectors, which limit the range of possible flows within each edge of the network graph. Given the restrictions on the flow rate, the pair of vectors that determine the range of possible diameters between the minimum (D_{min}, *i*) and the maximum (D_{max}, *i*) for each edge *i* can be calculated as follows:

$$D_{\min,i} = \sqrt{\frac{4 \times Min(Q_{MD,i}, Q_{MC,i})}{\pi \times U_M}} \quad (2)$$

$$D_{\max,i} = \sqrt{\frac{4 \times Max(Q_{MD,i}, Q_{MC,i})}{\pi \times U_M}} \quad (3)$$

3. Genetic algorithm with limitations

The approach proposed in this paper can be used in conjunction with various heuristic methods. In this article, the GA approach was chosen to test the effectiveness of the proposed methodology. GA is a simulation model of optimization, which is oriented to work with large complex systems. The paper uses GA, based on the classical model proposed in [2]. To implement the proposed approach, the model was changed and implemented in the form of software in the programming language C ++, the development environment of Qt Creator.

Algorithm [2] uses a scheme with integer coding. Each solution is encoded by a vector of *n* discrete variables (the pipe diameter sizes coincides a single edge of the net graph). Each chromosome is encoded by an integer value, that ranges from 1 to *nd*, *i* (the last diameter possible). This approach provides many advantages, since there are no restrictions on the number of possible pipe diameters that can coincide to a defined graph edge. In the classical formulation of GA, the number of possible diameters was the same for each communication line, and this value was equal to the total number of diameters in the pipe database. The same coding scheme is adopted in the proposed algorithm, although changes have been made allowing to vary the number of possible diameters for each edge of the graph. The method that we propose has an integer coding scheme and a variable number of possible options for each edge. The number of options depends on the number of possible diameters involved in the speed limits of each edge of the graph.

To test and compare the approach proposed with the classical GA, the initial population is formed randomly. This population evolves from generation to generation, conducting an iterative reproductive cycle. This cycle involves the use of three classical operators: selection, crossing-over and mutation. To use the selection operator, the suitability of each individual is estimated as the sum of the cost of the pipes making up the network, plus the penalty function used to account for the pressure deficit at the graph nodes, according to the formula

$$F(D) = \sum_{i=1}^{n} c_i \times L_i + p \sum_{j=1}^{N} (\max(hr_1 - h_1), 0) \quad (4)$$

where - c_i – the cost per meter of the pipe, depending on the diameter of D_i and the L_i length of the rib *i*, *p* - the coefficient of the penalty function, *N* - the number of nodes in the network, hr_j - the required pressure in node *j*, h_j - the operating pressure in node *j*, determined using the hydraulic model EPANET [6] .

The value of the coefficient of the penalty function can affect the accuracy of the solution, so it must be properly selected. To solve this problem, some researchers recommend various methods for processing constraints, such as the use of variable values or self-adaptive penalty functions [7]. However, in this paper, for simplicity and for comparing both approaches under the same conditions, the constant value of the penalty function coefficient is used. When simulating a penalty multiplier, a value was assigned to avoid decisions that violated pressure constraints at the nodes of the network. To calculate the effective values of nodal pressures for each individual, the EPANET hydraulic solver was used in the population, the source codes of which can be found in the public domain [7].

4. Modelling the operation of the algorithm

To simulate the GA and the proposed method, software was developed that implements these algorithms. As a result of the work, the application displays the parameters of the structure of the modeled waterway graph. For processing on the input of the application, both network information and a database of available pipe diameters are fed. Network information is imported from the EPANET input file. The database of pipes is stored in a text file of tabular format. The output module records the best solution found by the model (the resuting vector of pipe diameters and the cost of the net as a whole).
A block diagram of the software model is shown in Fig. 1.

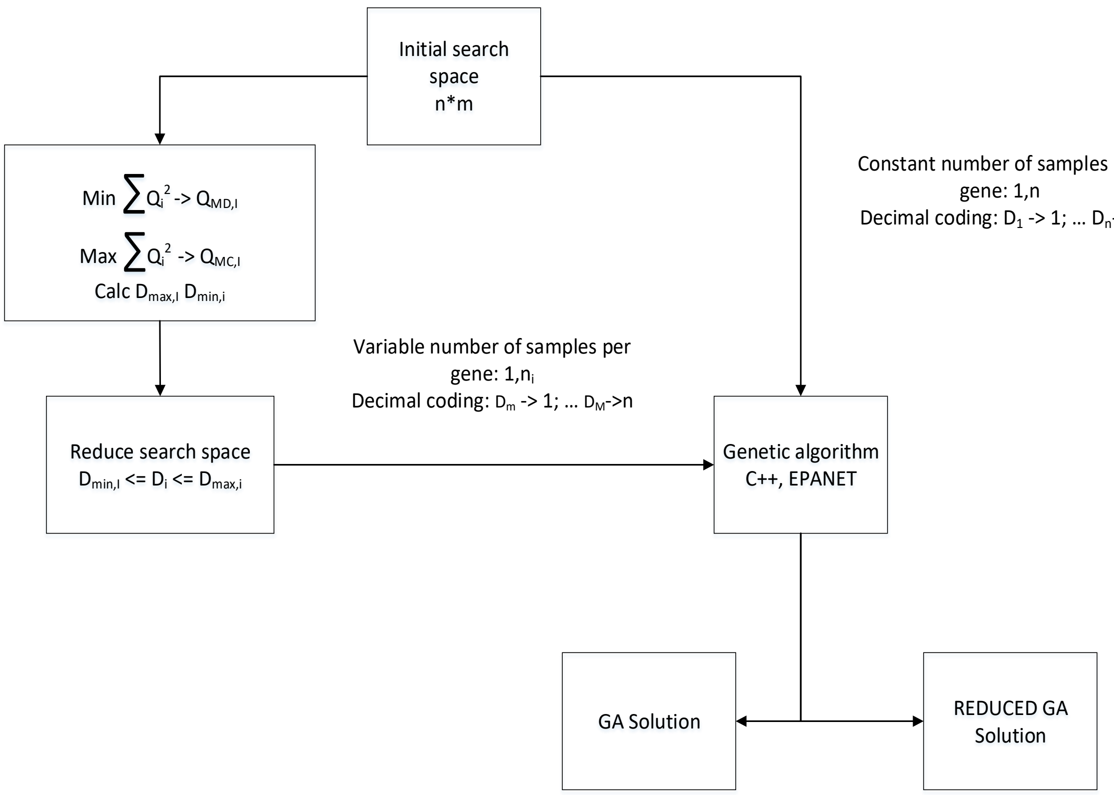

Figure 1. Block diagram of the software model.

The proposed approach was tested by a network test model to estimate the effectiveness of the classical GA approach with the limited GA model. For the test, the Hanoi waterway network was chosen, which was widely used in testing various algorithms for optimizing water conduit networks. The model can be considered as medium in size, with 32 nodes and 34 ribs and 3 loops. In the model, pumping units are not considered, since there is only one fixed source at a height of 100 m. The minimum pressure requirement for all nodes is fixed by a minimum height of 30 m. In this task, there is a set of 6 available diameters. The cost function is nonlinear. Pipe losses were calculated using the Hazen-Williams equation with a roughness factor C = 130. The values of the other parameters of the Hazen-Williams equation were established by default for the EPANET network analysis software [7].

IOP Conf. Series: Materials Science and Engineering **324** (2018) 012019 doi:10.1088/1757-899X/324/1/012019

To assess the effect of reducing the search space on the convergence rate and the accuracy of the solution, both algorithms were applied to solve the problem of designing a test network. For the subsequent comparison, in both algorithms the same input data and analysis parameters are selected. The size of the population was limited to 200 individuals. The number of generations was 300. The probability of crossing-over is Pcross = 0.85. The probability of mutation is Pmut = 0.08. The frequency of crossing the genes rcross = 0.3.
For simulation, the "leave-stable-delete-worst" reproduction plan is used, which leaves in the generation of the individual in the event that the value of their fitness function exceeds the level of the weakest suitable member of the population of ancestor. The least suitable member of the population is deleted and replaced by a descendant. The crossover operator implies that a pair of parent chromosomes exchange information to create a pair of chromosomes of the offspring that inherit their characteristics. Probability (the intersection of two chromosomes is determined by the input parameter Pcross.) In a homogeneous crossing-over, the chromosomes of the parents exchange their genetic information gene for a similar gene. The probability of gene exchange is determined by the frequency of crossing the genes (rcross) . To collect statistics, the application was run 20 times for both the proposed algorithm and for classical GA.

5. Simulation

The first stage in the calculation is determing the OQF, indicated in equation (1). The results of these calculations determine the range of flows for each channel of the network. Both flow rates represent the flow limits at each link in the network and, thus, reduce the search space. The results obtained are generated in Table 1. Table 1 contains the flow range, minimum and maximum diameters and the number of possible diameters compatible with the speed limits obtained as a result of calculating the OQF.

Table 1. Results of calculating OQF for the model.

Link	Q_{MD}	Q_{MC}	Dmin	Dmax (mm)
1	19,94	19,94	1016	1016
2	19,05	19,05	1016	1016
3	5326	6810	1016	1016
4	5196	6680	1016	1016
5	4471	5955	1016	1016
6	3466	4950	762	1016
7	2116	3600	609.6	1016
8	1566	3050	508	1016
9	1041	2525	406.4	1016
10	2000	2000	609.6	1016
11	1500	1500	508	1016
12	940	940	406.4	1016
13	1484	0	304.8	1016
14	2099	615	304.8	1016
15	2379	895	304.8	1016
16	2968	1205	508	1016
17	3833	2070	609.6	1016
18	5178	3415	762	1016
19	5238	3475	762	1016

20	7637	7915	1016	1016
21	1415	1415	508	1016
22	485	485	304.8	1016
23	4947	5225	1016	1016
24	2890	3065	609.6	1016
25	2070	2245	609.6	1016
26	992	1270	406.4	1016
27	92	370	304.8	762
28	278	0	304.8	762
29	1011	1115	406.4	1016
30	721	825	306.4	1016
31	361	465	304.8	1016
32	1	105	304.8	1016
33	104	0	304.8	1016
34	909	805	304.8	1016

To assess the efficiency of the proposed method with the classical GA algorithm, 20 iterations were performed for each method with the same input data and analysis process. The simulation performed gave the following results:

- GAL min cost 6 204 922;
- GA - max cost 6 573 272;
- GA - avr.cost 6,392,721;
- The proposed algorithm - min cost 6 122 136;
- The proposed algorithm - max cost 6 412 528;
- The proposed algorithm is avr.cost 6 241 235.

With the proposed methodology, an obvious reduction in search space was achieved. In the test network, there are six possible diameters in the database, and the number of edges is 34. The resulting number of alternative designs is 22.87 x 1026, while the search space in the restricted task is 4.35 x 1016. The speed limits also play an important role as Reduction of search space increases with decreasing speed range.

Regarding the rate of convergence, both algorithms converged quickly enough, although the proposed algorithm outperformed the classical GA (the algorithm converged on the 127th generation for the proposed algorithm and for the 312th in the case of GA). In the proposed algorithm, it was required to calculate slightly less than 25 thousand estimates, for the GA - more than 60 thousand. Both these quantities cover an insignificant part of the entire search space.

The proposed methods significantly reduced the search space and made it real to achieve much faster and more accurate convergence than the classical GA formulation. Both algorithms have found solutions close to the global optimum. Speaking on the test network, the best solution found by the proposed algorithm was 6 122 136, the average cost of the simulations was close to optimal. For GA, minimum solution found was 6,204,922, and also the average cost higher than the analogous parameters of the proposed algorithm. The solution found by the algorithm is comparable to the optimal solution found in the literature (6 056 000 [5]).

6. Conclusions

In this paper we presented a novel approach, based on limiting and reducing the overall search space in the design of water conduits using graph models. This approach minimizes search space by analysing two opposite scenarios for the architecture of extreme flows, and then applying restrictions on flow rates in the pipes. The proposed approach was used in conjunction with GA in order to improve its effectiveness.

Research evaluation process revealed that the proposed algorithm significantly reduced the search space and ensured faster and more accurate convergence than the classical GA formulation for the test net. It is forseen that for more complicated nets, the benefits, due to the new approach may be more significant. Extra vital advantage of the presented reduction is that it can be associated with other heuristic algorithms in the given subject area.

7. References

[1] *Hanoi water distribution network* 2012 Centre for Water Systems University of Exeter http://emps.exeter.ac.uk/engineering/research/cws/resources/benchmarks/layout/hanoi.php
[2] Reca J and Martmez J 2006 Genetic algorithms for the design of looped irrigation water distribution networks Water Resour. Res 42 W05416.
[3] Varma K and Narasimhan S 1997 *Optimal design of water distribution systems using NLP method* J Environ Eng 123 p 381-388.
[4] Lansey K E and Eusuff M M et al 2003 *Optimization of water distribution network design using the Shuffled Frog Leaping Algorithm* J. Water Resour. Plan. Manag 129 p 10-25.
[5] Montalvo I and Izquierdo J et al 2008 *Particle Swarm Optimization applied to the design of water supply systems* Comput. Math. Appl 56 p 769-776.
[6] EPANET 2017 *Software That Models the Hydraulic and Water Quality Behavior of Water Distribution Piping Systems* United States Environmental Protection Agency https://www.epa.gov/water-research/epanet
[7] Geem Z W and Kim J H et al 2001 *A new heuristic optimisation algorithm: Harmony search* Simulation 76 p 60-68.

Acknowledgments
This paper is published due the financial support of the Russian Science Foundation (RSF) via the grant No 15-19-00196.