Hybrid Genetic Algorithms for Multi-objective Optimisation of Water Distribution Networks

Edward Keedwell and Soon-Thiam Khu

Centre for Water Systems, School of Engineering and Computer Science and Mathematics, University of Exeter, North Park Road, Exeter, UK
{E.C.Keedwell, S.T.Khu}@exeter.ac.uk

Abstract. Genetic algorithms have been a standard technique for engineers optimising water distribution networks for some time. However in recent years there has been an increasing interest in multi-objective genetic algorithms that allow engineers a set of choices when implementing a solution. A choice of solutions is vital to help engineers understand the problem and in real world scenarios where budgets and requirements are flexible. This paper discusses the use of a local search procedure to speed up the convergence of a multi-objective algorithm and reports results on a real water distribution optimisation problems. This increase in efficiency is especially important in the water network optimisation field as the simulation of networks can be prohibitively expensive in computational terms.

1 Introduction

1.1 Water Distribution Network (WDN) Modelling

Water distribution network modelling is used for a variety of purposes such as: strategic and master planning; system design; energy management and daily operations; fire protection studies and emergency responses; water quality investigations; and many others. The computer modelling of water distribution networks continues apace in the water industry as computers become increasingly powerful, more complex systems are available to be modelled. Open source software such as EPANET (Rossman, 1993), which is the water distribution modelling software used in this paper, allows easy simulation of these complex systems. In recent years, computational methods such as non-linear programming, dynamic programming and search techniques have been used to optimise the design, operation and rehabilitation of these networks

K. Deb et al. (Eds.): GECCO 2004, LNCS 3103, pp. 1042–1053, 2004.

1.2 Network Optimisation

The optimisation of WDN models can primarily be used for three purposes: the design of networks for new supply areas (design problems); modifying existing designs to meet new demands or other factors (rehabilitation problems) and modifying network parameters to ensure that they are accurate with respect to the real world (calibration problems). This paper is concerned with the problems of rehabilitation and design although the methods described herein can theoretically be used for any of these problems. A standard rehabilitation problem is concerned with taking an existing network and modifying its parameters or components to satisfy new constraints in the performance of the network. Therefore any process designed to optimise a rehabilitation problem must consider both the performance of the network (to be maximised) and the cost of the proposed solution (to be minimised). A design problem is similar except that instead of deciding on pipe sizes to meet some new criteria, the pipe sizes of the entire network must be decided from scratch. The results in later sections show that the proposed hybrid method is applicable to both rehabilitation and design problems.

1.3 Optimisation Algorithms

As described in 1.2, a rehabilitation problem is concerned with taking an existing network and modifying its components, however, in the actual WDN there exist a large number of components which act in combination. This combinatorial effect means that even a network with a modest number of components will contain a huge number of potential solutions and therefore that it is not possible to evaluate every possible solution within the timeframe of the project. Therefore there exist a number of algorithms which are designed to find near optimal solutions from the overall set in real-time. Notably genetic algorithms (GAs) have found considerable success in this domain (Savic and Walters, 1997) and are the subject of much of the rest of this paper.

Many engineering problems involve conflicting objectives, and the optimisation of WDNs is no exception. For instance in a rehabilitation problem, the optimal monetary solution is to leave the system as it is with no modifications, equally though, investing a huge amount of money in the system will give superior network performance. Clearly there is a trade-off here between investment and network performance. It is this trade-off, which, by using the same evolutionary processes as a standard GA, can be optimised. The goal of a multi-objective genetic algorithm (MOGA) therefore is to yield a set of solutions which represent the optimal trade-off between two or more conflicting aspects of a system. In this paper we describe the use of MOGAs on a standard rehabilitation problem.

1.4 Hybrid Algorithms

The above scenario shows that GAs, and in particular, MOGAs can be used to present realistic solutions to the problems of optimising WDN performance. However, these population-based techniques suffer from difficulties when applied on today's complex models. The problem exists because WDN modelling, especially for large networks and extended period simulation can incur large computation time. Typically, GAs and MOGAs use a population size of 50-200 and 100-10,000 generations. This can therefore lead to anywhere between 5,000 and 2 million model evaluations for an optimisation run. This is evidently not feasible even if each model simulation requires 1 minute to run on a standard machine – 1 x 2,000,000 = 2 million minutes = 3.8 years. Even for the shortest runs (5,000 evaluations), if the objective function takes 1 minute to evaluate, then running times of 3 ½ days can be expected. Thus, it is imperative that for simulation of large WDNs that the number of evalutions for any optimisation algorithm to be reduced to a manageable number.

The main work in this paper therefore is to investigate the possibility that a MOGA-Local Search hybrid algorithm can exploit the behaviour of the GA whilst reducing the overall number of evaluations required to obtain near-optimal solutions. In particular we explore a combined method of neighbourhood search and NSGA-II, the current best performer in WDN MOGA optimisation.

2 Method

2.1 Trade Off Surfaces for Multiple Objectives

In this paper, the version of MOGA used is the Non-Dominated Sorting Algorithm-II (NSGA-II) (Deb et al, 2001), which is currently considered to be one of the best algorithm for water distribution system optimisation (Farmani *et al*, 2003). The hybrid approach uses NSGA-II unchanged except for the fact that the local search element is run periodically within the NSGA-II execution and the results from the local search are returned to the population and the optimisation continues. Both the local search and all MOGAs (including NSGA-II) rely on the concept of domination. With single objective GAs, solutions can be compared with respect to their fitness and a selection made. However, MOGAs need to find a trade-off of solutions which is, ideally, both optimal and well-spread over the objective space. Figure 1 shows such a comparison between trade-off surfaces.

Domination

One solution is said to dominate another if it is better (what constitutes better is determined by the problem, maximisation or minimisation) or equal in every objective.

Non-Domination
One solution can be said to be non-dominated if there is no other solution in the solution set which dominates it.

Figure 1 shows the concepts of domination and non-domination. Each of the solution sets marked by a "circle" are said to be non-dominated as they have at least one better objective value. Head deficit is a measure of the ability of the system to meet the original pressure requirements of the WDN. The cross in Figure 1 however is dominated by "C" by virtue of the fact that it has higher cost and head deficit.

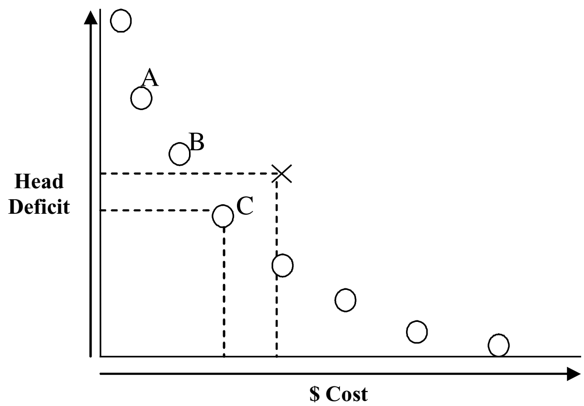

Fig. 1. Dominated and non-dominated solutions.

2.2 Multi-objective Area Measurement

The problem of comparing one or more pareto-curves in an experiment is non-trivial and a variety of methods have been put forward to evaluate these curves. One of the most obvious methods is to depict the curves on a chart and visually inspect which curve is closer to the optimum. However, whilst this can be effective, it is neither very scientific nor suitable for determining optimality during the course of an optimisation. The most suitable measure for this in our opinion is the hyper-volume or S-metric (Zitzler, 2000). Essentially this computes the area of the objective space that the pareto-curve covers, therefore meaning that the greater area covered, the more points that are dominated by the current curve. The method of computing the metric is shown in Figure 2 where each of the points has its area computed against the worst solution, W. The shaded area represents the area that is covered by the metric and as can be seen when minimizing both objectives for this method, the closer the solution set is to the optimum, the larger the area will be.

To succeed, the worst possible solution (W) has to be determined as a reference point to compute the bounds of the search space. In the experiments below we determine the worst hydraulic solution to be where every pipe diameter is minimal and the worst

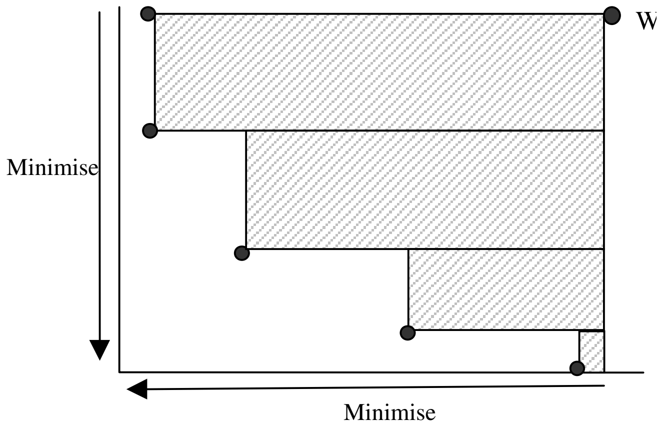

Fig. 2. Computation of the hyper-volume metric.

cost solution to be where every pipe diameter is maximal. Also, in the following experiments, we have normalised the calculations such that the maximum theoretical area is 1.0. For a precise definition of the S-metric, readers are referred to Zitzler (2000).

2.3 Hybrid Algorithm

The hybrid algorithm is used to drive the MOGA solutions towards the optimum during the optimisation to reduce the need for extended runs which require more model simulations. It uses a simplistic local search for a set number of iterations to find new, dominant solutions which are then entered back into the population of the MOGA. The standard local search is a neighbourhood search, it alters the diameters of two pipes at a time (one increment and one decrement) until it finds a solution which dominates the current solution. This is then selected and the process is repeated until some predetermined number of iterations has been completed.

This domination search ensures that only solutions which dominate the current best set are included in the search and therefore the insertion into the MOGA population. Once the solutions have been discovered, they are added back into the population, replacing the lowest ranked (and therefore most-optimal) solutions in the MOGA. Therefore the algorithm drives the MOGA to a more optimal solution set, but potentially at the price of some diversity in the population.

In following experimentation, we considered three separate local search components for the hybrid algorithm. The previous paragraph describes the operation of the simplest local search component that is used in the standard hybrid. However, because the problem is well-known a greater amount of knowledge can be built into the local search execution. This information takes the form of knowing which parts of the current solution are not meeting the requirements of the optimisation in one of the objectives. In this example, we know that some nodes will have a pressure deficit and some a pressure surplus, therefore the advanced heuristics utilise this knowledge.

Each of the local search techniques must increment and decrement a pipe diameter simultaneously at each iteration. This is because to achieve a solution that dominates the current starting point, the solution must attain both lower cost and improved hydraulic performance. Simply incrementing a pipe diameter could yield better hydraulic performance, but cost will be increased, and by decrementing one, the reverse is true. The following hybrids use this combined increment/decrement behaviour:

1. **The standard hybrid** uses a simple neighbourhood search, taking the current solution and incrementing or decrementing the decision variables blindly. In this search, a neighbourhood of solutions is considered which are one increment and decrement away from the current solution. This hybrid moves along the chromosome incrementing and decrementing until such time as it finds a solution that dominates the current one whereupon it moves to this solution and restarts.
2. **The list heuristic** builds a list of those nodes that do not have their pressure requirements met (deficits) and those which have too much pressure (surpluses). These two lists are sorted to discover the largest deficits and surpluses and the algorithm proceeds down the list, incrementing the immediately upstream pipes of the deficit list and decrementing the immediately upstream pipes of the surplus list. If there are only deficits or surpluses in the network, then the algorithm uses opposite ends of the same list (i.e. it will attempt to increment the greatest deficit and decrement the smallest).
3. **The upstream heuristic** finds the nodes in the network with the greatest deficit and surplus. The search algorithm then proceeds upstream of each of these nodes, and repeats. Therefore after discovering the node with greatest deficit, the algorithm increments the upstream pipe adjacent to that node, and then the pipe further upstream, and then further upstream and so on, until a dominating solution is discovered.

The hybrids discussed here all attempt to discover a solution which dominates the starting solution. However, each of the techniques must be computationally efficient to improve GA performance so that they require less time than the standard GA. To attempt to achieve this, the heuristics include knowledge of the network nature of the problem to pinpoint first those areas where the largest improvements can be made. By modifying pipe diameters close to those nodes which are maximally different from optimal, it is hoped that both heuristics will find solutions more quickly. The upstream heuristic extends this concept by attempting to ensure that those least optimal nodes are satisfied no matter how far upstream the problem pipe is.

3 Experimentation

3.1 New York City Tunnels Problem

The work carried out in the following sections is based on the New York City tunnel (NYT) rehabilitation problem (Schaake and Lai, 1969). This is a well known problem which has been studied extensively in the literature (Bhave and Sonak, 1985; Murphy et al., 1993; Savic and Walters, 1997; Farmani, 2003 etc.) and involves replacing old trunk pipes with new pipes of larger diameters or putting in new trunk pipes alongside old ones within the network to meet new demands for the City of New York. Figure 3 shows the existing network configuration.

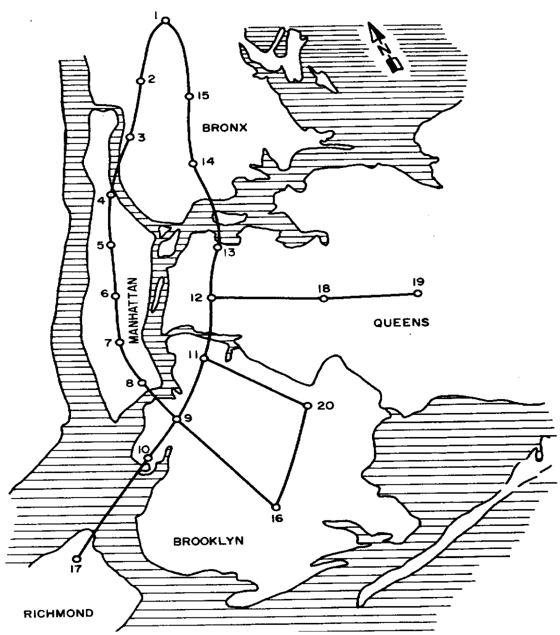

Fig. 3. Pictorial representation of the New York City Tunnels problem

There are a maximum of 21 new pipes to be laid, with the option of doing nothing. The new pipes vary in diameters, ranging between 60 – 204 inches. There is one reservoir which supplies the water to the network which contains 20 nodes which have various demands. A full enumeration of all possibilities would require 21^{16} ($1.43 * 10^{21}$) Epanet runs.

From an optimisation perspective, the objective of the NYT problem is to modify the rehabilitated pipe diameters to meet the demands at the nodes. The current optimal solution for this is 38.64 million dollars and no pressure deficit although this can vary slightly depending on the modelling software and parameters used.

3.2 Parameters

During experimentation with the algorithm we have discovered a number of parameters that gives the algorithm flexibility to discover the near- optimal sets. They include the following:

- *Local search frequency* – determines the number of times the local search is used throughout the optimisation.
- *Local search iterations* – determines the computational effort given to the local search when it is used
- *Replacement strategy* – determines which individuals in the GA population to replace, the best? Or the worst?
- *MOGA Parameters* – Those that can effect hybrid optimisation, such as population size.

Generally speaking we have found that low frequency, low iteration searches have produced the best results in conjunction with an optimal-replacement strategy which replaces the best-ranked solutions in the current GA population.

3.3 Experimental Procedure

Several experiments have been run on the New York Tunnels optimisation problem to determine the efficacy of the hybrid techniques described above.

Experiment 1 shows graphically the S-metric performance of each of the algorithms on a single optimisation run.

Experiment 2 evaluates the performance of each of the hybrids over 22 different random seeds and discovers how many model evaluations each algorithm requires to achieve a specific S-metric value.

These experiments are designed to compare the use of a hybrid optimisation procedure with NSGA-II, which is one of the best current multi-objective genetic algorithms.

4 Results

4.1 Experiment 1

In this experiment, each of the algorithms was run with the same random seed, and the S-Metric values at various points in the optimisation recorded. Each algorithm was run with a population size of 100, mutation and crossover rates of 0.9. The local search intervened once (at 50 generations) for 4 iterations. The resulting graph (Figure 4) gives an impression of the speed of convergence for all of the algorithms.

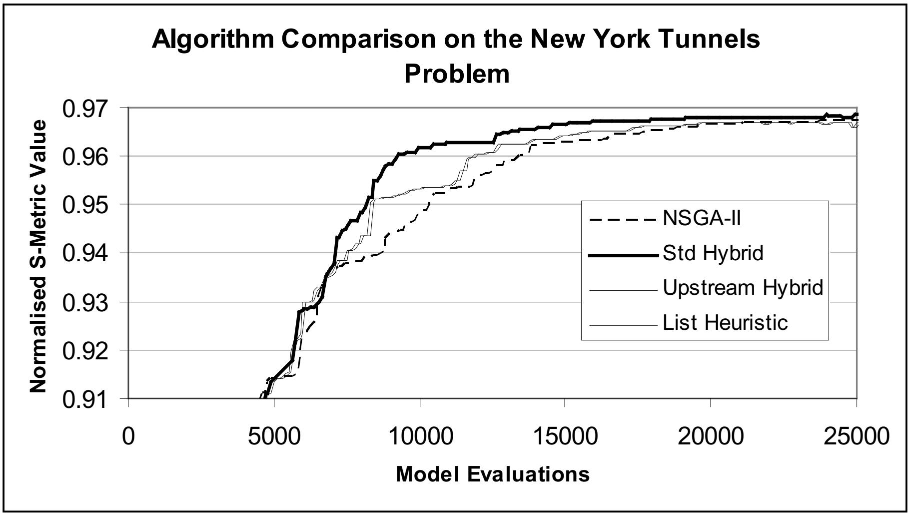

Fig. 4. Graph depicting the area under the Pareto curve against the computational cost, measured in model evaluations, for all algorithms

From Figure 4, it can be seen that the standard hybrid method outperforms all the other methods by some considerable margin in this case. It is also interesting to note that each of the hybrid methods outperforms the standard multi-objective GA on this problem but that the list and upstream heuristics are very close in performance. However, as this is only for one case the next experiment uses multiple trials to determine the performance of each algorithm.

4.2 Experiment 2

In this second experiment, NSGA-II and each of the hybrid algorithms was run with the same set of 22 different random seeds (12345 & 0-20). The number of model evaluations required to reach a fixed S-metric value of 0.95 is recorded. Table 1 shows the average, minimum, maximum and standard deviation of the number of model evaluations required to achieve this limit on the New York Tunnels problem. Each algorithm was run with 22 different random seeds and the local search run for 4 iterations. The table also shows these statistics for NSGA-II alone to achieve this area coverage.

Each of the hybrid techniques achieves some computational saving in comparison with the standard NSGA-II runs. The averages reveal a maximum of 7% model evaluation saving over the 22 runs when using the standard hybrid. The lowest performance of each algorithm is very similar and this suggests that the random seed is very important in the optimal execution of NSGA-II on this problem.

Table 1. Algorithm model evaluation comparison to 0.95 S-metric value

Algorithm	Mean	Min	Max	Std Dev
Standard Hybrid	6982	5845	8255	604.77
List Heuristic	7326.82	5900	8501	743.74
Upstream Heuristic	7326.82	5900	8501	743.74
NSGA-II	7495.46	5900	10000	911.03

Two further runs of this experiment with higher attainment targets of area coverage shows the extra computation required to achieve a marginally more optimal solution. Table 2 shows the extra model evaluations required to progress from 0.95 to 0.955 and 0.96 for each of the algorithms.

Table 2. Additional model evaluations required to achieve higher S-Metric values

Algorithm	0.955	0.96
Standard Hybrid	836.36	1981.82
List Heuristic	1204.55	2622.73
Upstream Heuristic	1204.55	2622.73
NSGA-II	1054.54	2609.10

This experiment shows again that by augmenting local search, better results can be achieved over the standard NSGA-II. In the best case (the standard hybrid), an S-metric value of 0.96 was achieved with an average of over 11% less model evaluations than the standard GA for the 22 random seeds. This indicates that the benefits of the local search procedure are still being felt later on in the optimisation.

5 Conclusion

Evolutionary multi-objective optimisation is currently one of the most useful tools available for the optimisation of water distribution systems because it allows the decision maker a choice of options. However, as with other generational evolutionary algorithms, the number of model evaluations incurred by such a system is so great that often they are difficult to apply to real-world systems. The hybrid approaches detailed in this paper show that local search can be used as an effective method of speeding up the search for a Pareto front over the current state-of-the-art multi-objective optimiser, NSGA-II. Both experiments show that the standard hybrid performed better than the heuristic approaches however the relatively poor performance of the heuristics in this application can potentially be explained by a single factor.

In this WDN, the largest head deficits tend to occur at the farthest end of the network from the reservoir, and therefore making a change locally to these nodes relies heavily on the construction of the rest of the network. Put simply, if the pressures upstream of the node are too low, increasing the local pipe diameters is going to have a minimal effect on the head at that node whilst significantly increasing costs. Each of the heuristics is designed to target these high deficit nodes first and therefore wastes a number of model evaluations searching these options that are only of minor hydraulic benefit whilst incurring high cost, and therefore are non-dominant.

This paper therefore shows that each of the hybrid optimisation procedures described here can improve on the performance of the standard NSGA-II. This algorithm is currently one of the best for multi-objective optimisation and therefore this represents a significant result. However, these results also highlight the need for careful selection of heuristics in local search as intuitive heuristics to the problem did not perform as well as expected.

In summary, the domain of water distribution network optimisation is notoriously computationally expensive and the computational savings made by a hybrid technique such as this could mean that a wider variety of complex real-world systems can be optimised using multi-objective evolutionary techniques.

References

Bhave P. R. and Sonak V. V. (1992). A critical study of the linear programming gradient method of optimal design of water supply networks. Water Resour. Res., 28(6),_1577 – 1584.

Deb, K. and Goel, T. (2001) " Controlled elitist non-dominated sorting genetic algorithms for better convergence ," Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization (EMO-2001), 7--9 March. (Zurich, Switzerland), (pp. 67--81).

Deb, K., Agrawal, S., Pratap, A., Meyarivan, T. (2000). A Fast Elitist Non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. Proceedings of the Parallel Problem Solving from Nature VI Conference , 16-20 September. (Paris, France), (pp. 849--858).

Farmani, R., Savic D. and Walters G., (2003) “Multi-objective optimisation of water systems: A comparative study” Pumps Electromagnetic Devices and Systems Applied to Urban Water Management (Cabrera, Cabrera Jr. Eds) AA Balkema Publishers pp 247-255

Goldberg, D.E., (1989) *Genetic Algorithms in Search, Optimisation and Machine Learning*. Addison Wesley Publishing Company Inc.

Keedwell, E.C. and Khu, S.T. (2003) “More choices in water system design through hybrid optimisation” Advances in Water Supply Management: Proceedings of the International Conference on Computing and Control in the Water Industry pp 257-265

Murphy L. J., Simpson A. R. and Dandy G. C. (1993). Pipe network optimisation using an improved genetic algorithm. *Res. Rep. No. R109. Dept. of Civil and Envir. Engrg.,* Univ. of Adelaide, Australia.

Rossman L., (1993). “EPANET”, U. S. Environment Protection Agency.

Savic D. and Walters G., (1997) “Genetic Algorithms for Least-Cost Design of Water Distribution Networks” J. Water Resources Planning and Mgt., 123(2), 67-75

Schaake J. and Lai D. (1969). Linear programming and dynamic programming applications to water distribution network design. *Rep. 116, Dept. of Civil Engrg., Massachusetts Inst. of Technol.,* Cambridge, Mass.

Vairavamoorthy K. and Ali M., (2000). Optimal design of water distribution systems using genetic algorithms. Computer Aided Civil and Infrastructure Engineering, *15*, 374-382.

Zitzler E., Deb K. and Thiele L., (2000). Comparison of multiobjective evolutionary algorithms: Empirical results. Evolutionary Computation, *8(2),* 173-195.