Mathematical Programming Study 11 (1979) 65–84.
North-Holland Publishing Company

OPTIMIZATION IN WATER DISTRIBUTION SYSTEMS ENGINEERING

Uri SHAMIR*

Technion, Haifa, Israel

Received 14 September 1977
Revised manuscript received 13 February 1979

The structure and operation of water distribution systems is explained briefly, and the basic mathematical tools used to analyze their physical behaviour are outlined. Methods for optimal planning, design and operation are presented and discussed. They are grouped according to the engineering problem which they address and to the method of solution. This survey is given at a level of detail which should suffice to understand each method and appreciate its potential and deficiencies. For more detail, or for actual implementation one has to refer back to the original works, which are listed at the end of the paper.

Key words: Water Distribution Systems, Optimal Design of Networks, Optimal Operation of Networks.

1. Introduction

This paper is a survey of optimization methods applied to problems of water distribution systems engineering. For people engaged in this field of engineering it constitutes a catalog of tools presently at out disposal to reach optimal solutions for problems which have previously been dealt with by engineering judgement and the use of some analytical tools. For those interested in optimization methods and their applications in other areas, the methods which have been developed specifically for water distribution systems may have potential either for direct application to other areas or as sources of ideas for formulating and solving other problems.

The following section is devoted to a brief description of the basic mathematical tools used in the analysis of water distribution systems. Components of such systems and their physical laws are presented, and it is explained how steady state flow solutions are obtained. The next section outlines the types of problems faced by the engineer in planning, design and operation. This sets the stage for the main body of the paper, in which optimization methods are presented and discussed. Those methods which are considered to be most readily applicable to real world problems are described in greater detail, while others are included mainly for completeness. The distinction obviously reflects the author's opinions, which may not be shared by others.

* Visiting Professor, Department of Civil Engineering, University of Toronto, and Director, Charles Howard & Associates, Ltd. Toronto. On leave from the Faculty of Civil Engineering, Technion, Haifa, Israel.

The paper concludes with an analysis of certain problems of water distribution systems engineering for which approaches or operational optimization tools are still missing, to which future work should be directed.

The extensive list of references at the end of this paper should aid in directing the interested reader to the original works, which contain more detail about the methods and their application.

2. Mathematical models for steady-state analysis

Water distribution systems are designed to deliver water from sources to consumers through pipeline networks equipped with a variety of components—pumps, valves of various types, reservoirs, etc. The mathematical model of a network consists of *links* connected at *nodes*. Nodes are points at which two elements connect, or at which a flow enters or leaves the network. To maintain mathematical tractability it is customary to model only the major structure using some appropriate representation of the detailed structure of the actual system. For example, it is customary to include in the model only pipes above some minimum diameter, and also to lump withdrawals taken along the pipes and assign them to the nodes.

At each node we define the *head* and the *consumption*. The head is a measure of the energy per unit weight of water. It is measured by the elevation to which the water would rise in a stand-pipe at that point, and is the sum of the topographic elevation, the pressure and velocity heads (the latter is usually negligible). While there exists a head at any point along a pipe we shall refer by "heads" to the set of values at the nodes. At certain nodes there are *inputs* of water into the system. At others there are *consumptions*, which are viewed as negative inputs.

Each *link* is characterized by a physical law which relates the flow through it to the head difference between its ends. For pipes there are several empirical flow equations; one commonly used is the Hazen–Williams equation:

$$Q = \alpha C_{HW} D^{2.63} (\Delta H/L)^{0.54} \tag{1}$$

where Q is the discharge; C_{HW} a smoothness coefficient (the Hazen–Williams coefficient); D the diameter, ΔH the head difference between the two end nodes, and L the length. α is a numerical coefficient, whose value depends on the units used. (For Q in m^3/sec and D in cm $\alpha = 5.4 \times 10^{-3}$; for Q in cfs and D in inches $\alpha = 6.28 \times 10^{-4}$). $\Delta H/L$ is the hydraulic gradient, later denoted J.

For centrifugal pumps, the head added by the pump, ΔH, is usually approximated by a polynomial of the form

$$\Delta H = a + bQ + cQ^2. \tag{2}$$

Every other type of element has its own law. Reservoirs are connected at nodes; the water level in the reservoir is the head at the node.

The mathematical model of the network is a set of simultaneous non-linear algebraic equations, which correspond to a steady-state flow in the network under fixed boundary conditions: a set of inflows and consumptions at the nodes and a set of fixed heads at specified nodes. The solution of these equations is called a *steady-state flow solution*. There are two basic types of equations: node equations and path equations. The first express material continuity at a node. For node j:

$$\sum_i Q_{ij} + I_j = 0 \tag{3}$$

where Q_{ij} is the flow from node i to node j, and I_j is the inflow into node j. Consumptions (withdrawals) are denoted by C_j and appear as negative values of I_j.

Path equations equate the head difference, b_p, between the end nodes of path (p) in the network to the sum of head gains and losses in all links belonging to this path:

$$\sum_{(i,j)\in p} \Delta H_{ij} = b_p \tag{4}$$

A path may connect any two nodes. Usually path equations are formulated between pairs of nodes at which the heads are known. A special case are *loop equations*, where the two ends are at the same node, and then $b_p = 0$.

There are many different ways to construct the mathematical model. In a network with N nodes, a set of N node equations fully determines the flow solution, provided one head (a reference head) is given. When loop equations are used one needs as many equations as there are basic loops, i.e., loops which do not have pipes intersecting them. General path equations can be formulated in many ways and the rule is that they should be mutually independent. Any one of these formulations can be used to obtain the flow solution, i.e. the heads and flows throughout the network. The set of simultaneous non-linear algebraic equations can be solved by an appropriate iterative technique—Hardy–Cross, Newton–Raphson or linearization ([34, 35, 36]).

When the network is tree-shaped and has no loops—which is common in irrigation systems—once the inputs and withdrawals are fixed the flows in all links are known and the flow solution (i.e. the heads at the nodes) is trivial.

When there is internal storage in the system, i.e. there are operational reservoirs, then the sequence of flow solutions for varying withdrawals over time (e.g. a day) is of importance. Subsequent flow solutions are linked through the changes in water levels in the reservoirs. Simulation of such a system's operation is carried out by obtaining a flow solution for the initial boundary conditions, the resulting flows are used to update reservoir levels, then these levels are used as boundary conditions for the next flow solution, and so on.

Readers interested in methods and programs for flow solutions and simulation over time are referred to [9, 30, 34, 35].

3. Problems of planning, design and operation

The layout of the small-diameter pipes in a distribution network is essentially fixed by the land development. Water has to be delivered to every building, and pipes usually follow streets, so the layout of the finer grid is fixed to a large extent. On the other hand, the engineer has to select the pipe materials and diameters, consider the possibility of breaking the network into pressure zones (separated by special pressure regulating devices), allocate the loads to the various sources, determine the layout and sizing of the feeder mains, and fix the locations and designs of the pumps, reservoirs and other facilities. *Planning* is the phase of selecting the layout and main features of the system. *Design* is the phase of fixing the sizes and characteristics of the various components. The two are complementary and should be carried out simultaneously. Still, in some instances the layout and location of facilities has already been fixed, and the engineer is then faced with a design problem—that of sizing the components.

Distribution systems have to operate under time-varying conditions. Considering long range changes, on the scale of years, there is the need to meet increasing demands, i.e. sizing and timing of new facilities needed to meet these demands. This *capacity expansion* problem is common to many areas of engineering, and although for water distribution systems it may have some special characteristics it will not be dealt with here. The scale of time-varying conditions which is of specific concern is associated with the *operation* of the system. Similar to power systems, water distribution systems operate under loads which change over the day, the week and the seasons. Sources, pumps, valves and reservoirs are operated to meet these varying demands. Setting the operating policy of the system is an integral part of the design process, since sizing the components depends on their operation—and vice versa. When the system already exists, the problem is how to operate it under normal and exceptional (breakdown, emergencies) conditions.

In the design one would like to consider explicitly the detailed operation of the system, that is, the hour by hour position of the pumps, valves and reservoirs. No way has been found as yet to formulate and solve the design and operation optimization problem in this form. Instead, one can include in the formulation of the design problem specification of the operation under one or several *loadings* (sets of consumptions). These are "typical" or "critical" loadings — for example, average daily consumptions, maximum demands during the day, low demands (which usually occur at night, during which reservoirs can be filled), high demands for fire-fighting (usually concentrated in the highly congested business district), etc. The solution of this design and operation problem provides both sizing of the various components and their operation under these typical loadings.

This solution does not specify how the system is to be operated over time (day, week, season). The resulting design may, therefore, not be feasible from the operational standpoint, and if it is feasible it may not be optimal. Specifically, the optimal balance of pipe, pump and reservoir sizes which will ensure reliable supply throughout the day and week at least cost can be arrived at only through a model which simulates the operation over time. Still, if a reasonable number of loadings is considered within the design formulation, and if each is assigned its proper weight in the objective function, the resulting design can be said to reflect operational consideration. Temporal operation of an existing system is dealt with by separate methods.

Planning, design and operation are thus three aspects of a single problem, but under certain circumstances the task may be narrowed to design and operation, or operation alone. In the following section which deals with optimization, we shall treat planning and design as a single problem. The method is as follows: one specifies facilities (pipes, pumps, valves, reservoirs) wherever they seem reasonable; the optimization procedure is allowed to set any design variable to zero, thereby eliminating the element from the solution. This procedure will not create a facility where one was not specified, and is therefore limited to those configurations stipulated by the designer. Still, this procedure allows for selection among alternatives.

4. Formulation of network optimization models

We first discuss the general structure of optimization problems for planning, design and operation of water distribution systems and then go on to describe and discuss the methods which have been developed to solve them.

4.1. Decision variables

As mentioned above, planning is treated as part of the design problem—by allowing design variables to take on zero values one may eliminate components from the solution, thereby selecting between alternative layouts and locations of facilities.

In selecting pipes one has to decide on their material, diameter and wall thickness (pressure bearing capacity). In the optimization, it is usually assumed that material (i.e. smoothness) and wall thickness have been fixed, and the remaining decision is on the diameter. In selecting pumps there are various considerations, but for the optimization only the head vs. discharge characteristic of each pump will be considered as the decision variable. Similar considerations hold for valves and reservoirs.

Decision variables of the planning-and-design problem include:

(a) Pipe diameters,

(b) Pump locations and characteristics,

(c) Valve locations, and

(d) Reservoir locations and sizes.

When operation of the system under a set of typical or critical loadings is included in the formulation, the following decision variables are added:

(e) Pumps to be operated (ON/OFF) under each loading, and

(f) Valves to be operated (amount of pressure loss provided by the valve) under each loading.

For optimization of the operation over time, the decision variables are the times at which the controlled components are switched on or off. Surrogates for these may be states of the system at which the components are to be switched, for example, reservoir levels which are setpoints for controlling the activation and de-activation of pumps.

4.2. *Objective functions*

Minimum cost is the criterion most often used in optimization of water distribution systems. Total cost is made of capital plus operating costs. The latter usually reflect only energy costs, since operation and maintenance may be included in the capital cost. Practical difficulties are encountered in formulating cost functions for distribution systems. Actual costs are often quite different from what was assumed in the design phase, due to price changes and unforseen circumstances during construction. More reliable cost models are needed.

Performance indicators, such as minimum pressures at supply nodes, are usually treated as constraints in the optimization. Some work has been done in which certain performance criteria were used in the objective functions in a multiple-objective formulation, [10].

4.3. *Constraints*

Several types of constraints appear in the optimization models. First, the physical laws of flow in the network have to be satisfied. These are equality constraints for continuity of mass and/or of hydraulic head lines-equations (3) and/or (4). The consumptions are usually treated as fixed externally, and so they appear as constraints in the node continuity equations. Limits are normally set for the heads and/or pressures at some nodes. Minimum pressures are to be guaranteed under all loadings, to meet service standards to domestic and industrial consumers and to ensure sufficient operating conditions for fire fighting. Maximum pressures are specified when there is a danger of pipes bursting or equipment being damaged under excessive pressure.

Sensivity analysis may be used to test the effect of certain changes in the values of the constraining parameters on the cost, since the values given by law, regulation or convention are generally not based on economic analysis, and are therefore not necessarily optimal.

Special types of constraints may arise from specific formulations of the optimization models. These will be discussed specifically for each model to be presented.

5. Branching network models

5.1. Linear programming model

Consider first a branching network, supplied from one or more sources by gravity, and designed for a single loading. At the supply nodes a specified consumption has to be satisfied, C_j at node j. At some or all of the nodes the head, H_j, is to be within a given range, HMIN_j to HMAX_j. The layout is given, and the length of the link (pipe) connecting nodes i and j is L_{ij}.

The linear programming design procedure [16, 17, 19, 24] is based on a special selection of the decision variables: for each link allow a set of "candidate diameters", the decision variables being the lengths of the segments of these diameters within the link. Denoting by X_{ijm} the length of the pipe segment of the m-th diameter in the link between nodes i and j, then

$$\sum_m X_{ijm} = L_{ij} \quad \text{for all } (i, j) \tag{5}$$

where each link may have a different set of candidate diameters. For a branching network in which the consumptions are known, the discharges in all links, Q_{ij} are fixed. The head loss in segment m of the link is:

$$\Delta H_{ijm} = J_{ijm} X_{ijm} \quad \text{for all } (i, j, m) \tag{6}$$

where J is the hydraulic gradient ($\Delta H/L$ in eq. (1)) which is a function of the discharge and of the diameter (if the smoothness is assumed to be selected in advance, and therefore fixed).

Starting from any node in the system, s, at which the head is fixed (e.g. a reservoir), and selecting any path from it to node n, at which the head has to be within a given range, one may formulate the constraint

$$\mathrm{HMIN}_n \leq H_s \pm \sum_{(i,j)} \sum_m J_{ijm} X_{ijm} \leq \mathrm{HMAX}_n \tag{7}$$

The first summation is over all links along the selected path, and the second over all segments of the link. The signs of the terms depend on the direction of flow in the link. In order to reduce the number of constraints and improve computational efficiency, head constrains may be formulated only for part of the nodes, provided they suffice to ensure that pressures throughout the network are within their acceptable ranges. If this method is used one has to examine the solution to ascertain that all heads are satisfactory; wherever they are not, a new constraint has to be added and the problem resolved.

The cost of a pipeline with a fixed diameter can reasonably be taken as linearly proportional to its length. Thus, the total cost of the pipeline network is

$$\sum_{(i,j)} \sum_m C_{ijm} X_{ijm} \tag{8}$$

Minimization of (8), subject to (5), (7) and non-negativity of the X is a linear program. Preselection of the candidate diameters actually constitutes a constraint as well. A reasonably narrow list is to be preferred for each link, to keep the number of decision variables low. However, unless limitation of the list reflects actual availability of only certain pipe diameters care should be taken that the implicit constraint due to the limited list should not be binding in the optimal solution. After an LP solution has been reached with a particular set of candidate diameters one examines it to see which diameters have been selected for each link. Wherever a link is made entirely of a single diameter which is at an end of its list, the list for that link has to be expanded in the proper direction and the problem solved again, until this constraint is not binding for all links. This procedure can be incorporated into the algorithm and performed within the computer program itself.

It can be shown that if the cost of a pipe is a convex function of the diameter (as it normally is) then in the optimal solution of the LP each link will contain at most two segments, their diameters being adjacent on the candidate list for that link.

The LP formulation can be extended to include the cost of pumps, and their operation, using linear or linearized cost functions [2, 24]. Reservoirs can also be included, using the head in the reservoir as the decision variable, and fitting it with a linear cost function. More than one loading can be considered. Each loading results in a set of constraints of type (7), possibly with different bounds on the heads for each loading, and the entire set is solved simultaneously in the LP. If energy costs are included the objective function contains a weighted sum of the energy costs for operating under the different loadings.

5.2. Dynamic programming model

Optimal design of a branching network, with or without pumps and reservoirs, can easily be formulated as a dynamic programming (DP) problem. The solution requires more computer time than the LP method, but the formulation is free of certain shortcomings present in the LP.

Kally [21] used DP to optimize the diameter and wall thickness (which determines the pressure bearing capacity) of the segments of a pipeline, as well as the heads added by the pumps located along it. The objective function included capital plus operating costs. Similarly, Liang [28] used DP to optimize the diameters of segments between takeoff points to consumers along a pipeline fed by a pump at its upstream end. The objective was to minimize capital cost, subject to minimum pressure constraints at the takeoffs.

Probably because of its computational inferiority, the use of DP for branching networks has not been developed to an operational stage. Still, for completeness we shall present the structure of the DP formulation.

The network is divided into segments. A segment is defined between adjacent takeoff points, so that the discharge remains the same along the segment, or, if these segments are too long, a finer division may be used, allowing pipe properties to change from one segment to the next. Takeoffs are given, and minimum pressures are to be satisfied at each takeoff node. Decision variables may include the diameters, material (roughness and strength) and class (pressure bearing capacity) of pipes, and the capacities of pumps. For clarity of the presentation we shall assume that material and class of the pipes have been fixed, so that diameters are the only decision variables. The objective function may include capital cost of pipes and pumps, energy costs, and any benefits, costs or penalties which are functions of the heads at the nodes. No assumptions, such as continuity or convexity, have to be made about these functions. The quantities to be supplied at the nodes are assumed fixed, and constraints may be imposed on the minimum and/or maximum heads at the nodes.

The *state variables* are the heads at the nodes, and the nodes are the *stages*. Computation proceeds upstream, starting from the downstream end of each branch. The *recursive equation* of the DP is:

$$F^*_{j+1}(H_{j+1}) = \underset{D_k}{\mathrm{Min}}[g(D_k) + f(H_{j+1}, H_j) + F^*_j(H_j)] \tag{9}$$

D_k is the diameter of segment k which connects node $(j+1)$ to its downstream neighbor, node j. $g(D_k)$ is the cost of this segment. H_{j+1} is the head at node $(j+1)$; H_j is the head at node j *given* H_{j+1} and the diameter D_k, and can easily be computed since the discharge in the segment is known. $f(H_{j+1}, H_j)$ is the cost (or benefit) associated with the link, for the given heads at its two ends. $F^*_j(H_j)$ is the optimal value for the portion of the system downstream of node j, given the head at node j. The minimization is over all admissible values of D_k, and is performed for each of a set of discretized values of H_{j+1} over its admissible range. Whenever $H_j = H_j(H_{j+1}, D_k, Q_k)$ is outside its admissible range, the examined D_k is disallowed.

For a pump located between nodes $(j+1)$ and j one uses:

$$F^*_{j+1}(H_{j+1}) = \underset{H_j}{\mathrm{Min}}[g(H_{j+1}, H_j) + f(H_{j+1}, H_j) + F^*_j(H_j)] \tag{10}$$

$g(H_{j+1}, H_j)$ here is the capital plus operating cost for a pump designed to deliver the known discharge, Q_k, from head H_{j+1} at its intake to head $H_j(>H_{j+1})$ at its discharge. The minimization is over a set of discrete values of H_j.

At every branching node of the network one adds up the values of F^* for the downstream branches which connect at it. At such a node j, $F^*_j(H_j)$ is still the optimal cost of the part of the network downstream from it, except that now it is a sum of the optimal costs for all the branches originating at node j. Thus, one follows the single-line procedure outlined above, starting from all downstream extremities of the network. Whenever two or more branches join at a node their contributions are added up for each value of the head at the branching node. The computational effort for a branching network is not much greater than for a single line with the same number of nodes, but the logic of the program is considerably more involved—if it is to be a general purpose program rather than one designed to optimize a particular network.

6. Looped network models

6.1. Linear and separable programming models

The basic difference between branching and looped networks is that in the former the flows in the pipes are fully determined by the loading, whereas in the latter the flow distribution also depends on the pipe properties (diameters, smoothness, lengths). This is the main difficulty which has to be overcome in applying linear or separable programming to looped networks.

Lai and Schaake [26] have addressed this problem by making the assumption that the heads at all nodes, as well as the demands, are given in advance. The solution thus gives the optimal diameters for the assumed pressure pattern. Since the heads are fixed the flow through each link is a function of the link properties only, and if the length and pipe material are given, the flow is a function only of the diameter. For node j at which a demand C_j has to be satisfied, the following constraint is a statement of continuity at the node:

$$\sum_i Q_{ij} = \sum_i K_{ij} D_{ij}^{2.63} = C_j \qquad (11)$$

where K_i is a coefficient whose value is determined by the given data: the heads at nodes i and j, and the length and smoothness of the pipe connecting them (see eq. (1)). The objective function Lai and Schaake used was:

$$\sum_{(i,j)} aL_{ij}D_{ij}^{e} + b\left[\sum_{(i,j)} Q_{ij}\Delta H_{ij} + \sum_j C_j H_j\right] \qquad (12)$$

a and b are constants which account for unit conversion, a present value factor, etc. e is a coefficient whose value is determined through analysis of pipeline cost data. The first sum is the capital cost of pipelines, the second is the cost of energy lost in the flow through the pipes, and the third is the cost of the residual energy at the supply nodes. There are no pumps within the network, and the energy terms represent the cost of supplying water to the network from some external source by a pump. By making the substitution $Y_{ij} = D_{ij}^{2.63}$ the constraints (11) become linear. Using equation (1) for the flows, the objective function (12) becomes:

$$\sum_{(i,j)} (aL_{ij} Y_{ij}^{e/2.63} + b_{ij} Y_{ij}) \qquad (13)$$

where the coefficients b_{ij} are based on the heads and flows. Minimization of (13) subject to constraints of type (11) for all supply nodes and to non-negativity of the y_{ij} is a separable program (which Lai and Schaake solved by a self-developed iterative LP program). The method was used in a study of New York City's primary distribution system [10] in which several performance criteria were also introduced as objectives (eg. sum of the residual pressures at all the supply nodes, the residual pressure at the farthest supply point, etc.).

Kally [23] approached the problem by extending the LP formulation for branching systems. His reasoning is as follows. In a branching system, if one changes the length of a pipe segment (which has some fixed diameter) the resulting changes in the heads at nodes are linearly proportional to the magnitude of this change. In a looped network the same effect is non-linear, due to the re-distribution of flows once the design is changed. Still, if the change in length of a segment is small enough the resulting change in heads is approximately linearly proportional to the change in length. The ultimate decision variables in Kally's formulation are the lengths of the segments, the same as was for a branching network, and the objective function is also the same. The problem is solved through a sequence of linear programs; the decision variables in each LP are the lengths in each link along which the diameter is to be changed from one diameter to another, i.e. the length to be taken away from one segment and given to another. Denoting by s_{ijm} the length in link (i, j) of change from the present diameter to diameter m (of a candidate list), these changes have to satisfy

$$\sum_m s_{ijm} \leqslant L_{ij} \tag{14}$$

that is, the sum of all changes cannot exceed the total length of the link. Also, considering the minimum heads required at certain nodes, HMIN_j, the changes in diameters are limited by

$$\sum_{(i,j)} \sum_m \left(\frac{\Delta H_k}{\Delta s_{ijm}}\right) X_{ijm} \leqslant (H_k - \mathrm{HMIN}_k) \tag{15}$$

where H_k is the head at node k in the present iteration, and $(H_k - \mathrm{HMIN}_k)$ is therefore the extra head which can still be eliminated (if feasible from other considerations). The coefficients $(\Delta H_k/\Delta s_{ijm})$ are the linear approximations for the rate of change of head with respect to changes in segment lengths. Kally's approach was to obtain these values as the differences in heads at the nodes between two flow solutions—one with the existing segments, and another with one $\Delta s_{ijm} = 1$ and the others $=0$. For each $\Delta s_{ijm} = 1$ one has to run a network solver and obtain a flow solution, then one computes the head differences Δh_k between it and the "basic" flow solution–the one with all $\Delta s_{ijm} = 0$.

The objective function for each iteration's LP is

$$\mathrm{Min} \sum_{(i,j)} \sum_m C_{ijm} s_{ijm} \tag{16}$$

where C_{ijm} is the cost of changing one unit length of pipe in link (i, j) from its present diameter to diameter m.

Upon solution of the LP the changes of diameters over lengths s_{ijm} are introduced and a new iteration is begun: a "basic" flow solution is computed, a series of flow solutions for all $\Delta s_{ijm} = 1$ are obtained and (16) is minimized, subject to (14), (15) and non-negativity of the s.

Kohlhass and Mattern (25) used separable programming in optimizing a looped network in which the heads are fixed in advance—a condition similar to the one stipulated by Lai and Schaake [26].

Recently, a more general method has been developed [2] and introduced into practice. The method—call LPG, for Linear Programming Gradient— is based on the following reasoning. If the flows throughout a looped network are known then its optimal design can be obtained by an LP formulation similar to that for a branching network. The optimization has therefore to find the optimal flow distribution, and for it the optimal design. This is achieved by a hierarchial approach: in the lower level of the hierarchy the optimal design for a particular flow distribution is obtained by LP; in the higher level, the flow distribution is modified, using certain results of the LP solution, towards an optimal flow distribution. This procedure is continued iteratively until some termination criterion is met.

The LP for a fixed flow distribution in a looped network operating under gravity for a single loading is: minimize (8), subject to (5) and (7), to non-negativity of the X, and to the additional constraint:

$$\sum_{(i,j)\in p} \sum_{m} J_{ijm} X_{ijm} = b_p \tag{17}$$

where p designates a path in the network, and b_p is the (known) head difference between its ends (all other notation is the same as in Section 5.1). Eq. (17) has to hold for all loops in the network with $b = 0$. When pumps, valves or reservoirs are to be included the objective function and constraints have to be augmented, as will be explained below.

Having added constraints (17), the LP can be solved, and the set of optimal segments will be such that the network is hydraulically balanced be virtue of the fact that the constraints (17) have been satisfied. Denoting by $\boldsymbol{Q}$ the vector of flows in all the paths, which may be any arbitrary set of flows as long as they satisfy continuity at all nodes, then the optimal cost of the network, F, for this $\boldsymbol{Q}$ can be written as

$$F = \mathrm{LP}\ (\boldsymbol{Q}) \tag{18}$$

where LP denotes that F is the outcome of a linear program. Next, $\boldsymbol{Q}$ is modified in a way which approaches optimality. Denoting by ΔQ_p the change of flow in path p, then

$$\frac{\partial F}{\partial(\Delta Q_p)} = \frac{\partial F}{\partial b_p} \cdot \frac{\partial b_p}{\partial(\Delta Q_p)} + \sum_{r\in R} \frac{\partial F}{\partial b_r} \frac{\partial b_r}{\partial(\Delta Q_p)} = W_p \frac{\partial b_p}{\partial(\Delta Q_p)} + \sum_{r\in R} W_r \frac{\partial b_r}{\partial(\Delta Q_p)} \tag{19}$$

where W_p and W_r are the dual variables of constraint (17) for the paths p and r, respectively, where R are the paths which share a link (or more than one link) with path p. Using eq. (17) and the definition of J from eq. (1):

$$\frac{\partial b_p}{\partial(\Delta Q_p)} = \frac{\partial b_p}{\partial(\Delta Q_p)} = \sum_{(i,i)\in p} \sum_m [1.852\,\alpha\, Q_{ij}^{0.852} C_{ijm}^{-1.852} D_{ijm}^{-4.87} X_{ijm}]$$
$$= 1.852 \sum_{(i,j)\in p} \frac{1}{Q_{ij}} \sum_m \Delta H_{ijm}. \qquad (20)$$

$\partial(\Delta Q_p) = \partial(Q_p)$ because both are incremental changes in the flow in the path. Q_{ij} and ΔH_{ijm} have already been used in setting up the LP, so once it has been solved and the duals W_p and W_r are known the components of the gradient

$$G_p = \frac{\partial F}{\partial(\Delta Q_p)} = 1.852 \Big[W_p \sum_{(i,j)\in p} \frac{1}{Q_{ij}} \sum_m \Delta H_{ijm}$$
$$\pm \sum_{r\in R} W_r \sum_{(i,j)\in r} \frac{1}{Q_{ij}} \sum_m \Delta H_{ijm} \Big] \qquad (21)$$

can easily be computed. The sign of the additional terms is positive when path r uses link (i, j) in the same direction as path p, and negative otherwise. With these components once can define a vector change in path flows, $\Delta \boldsymbol{Q}$ such that

$$\mathrm{LP}(\boldsymbol{Q} + \beta \Delta \boldsymbol{Q}) < \mathrm{LP}(\boldsymbol{Q}) \qquad (22)$$

β is a step size, which is selected by an appropriate one-dimensional search procedure.

Several loadings should be considered in the design. Maximum hourly flows during the day and fire fighting demands are normally used, but often the low demands, as may occur at night, have to be considered as well. This is particularly important when there are reservoirs in the system, which should fill during low demand periods and then empty during high demands, thereby reducing the peak loads on the sources. For each loading one specified in LPG an initial flow distribution which satisfies continuity at all nodes. A number of head constraints (7) are written for each loading, and the entire set is included in a single LP matrix. Once the LP has been solved, the gradient section modifies the flow distribution for each of the loadings, using the results of the LP.

Since the initial flows for each loading are quite arbitrary, there may not exist a set of segment diameters such that the head line constraints (17) are satisfied for all loadings. Therefore, two new variables are added in each of these constraints—one with a positive and the other with a negative sign, and both required to be non-negative. These variables are assigned a large penalty coefficient in the objective function and can therefore be viewed as artificial variables. They do, however, have a physical meaning in our case. Each such variable may be viewed as a valve in the path, able to take up the excess head for the specified flow in the path. These *dummy valves* are "operated" to satisfy head-line continuity. Because of the large penalty they invoke, the program attempts to eliminate them from the solution. If this is possible, i.e. there is a feasible set of segment diameters without valves, then their introduction has merely served the purpose of reaching this solution by use of the LP algorithm. If, on the other hand, a dummy valve does appear in the final optimal solution, this means that a real valve will have to be installed at the location specified, and operated accordingly. When an actual valve does exist in the system it is represented in the LP by the head loss it provides under each loading.

LPG is thus designed to reach a design which is hydraulically balanced for all loadings, and at the same time move the design towards optimality. The iterative LP-Gradient procedure is terminated when any one of several criteria is met (no significant improvement from one iteration to the next, specified number of flow-change iterations exceeded, etc.).

When a pump is to be designed, the head it has to add for each of the loadings is the decision variable. These variables are introduced into the constraints of types (7) and (17) with the proper sign. An iterative procedure is used in dealing with the non-linear cost vs. head function for pumps. For reservoirs, the decision variable is the elevation at which it is to be located. This elevation appears in all constraints for paths ending at the reservoir. A linear cost vs. elevation relation is used in the objective function.

6.2. Non-linear programming models

An early effort [33] was made to use a gradient-like technique in optimizing the design of a pipe network under one loading. The original work, which was not published, constituted the basis for Lemieux's thesis [27]. Pipe diameters are changed, one pipe at a time, according to the derivative of the objective function with respect to pipe diameters. After each change in a diameter the new network is solved, using a Newton–Raphson method, and the last Jacobian of this solution appears in the computation of the derivatives.

Pitchai [29] formulated a non-linear integer programming problem in seeking the optimal diameters of a network, and solved it by combining random search and examination of adjacent design points. Jacoby [20] used a gradient-approximation method to seek minimum of a merit function which combined the objective function and penalties for violation of head and continuity constraints for a single loading. An approximation of the gradient was computed by perturbing the design variables around their present values, and a move is made in the gradient direction. The sequence of steps is guided by several heuristic rules, and is terminated according to several rules. Since the search is conducted in the region which is hydraulically infeasible (it is an exterior-point method), if the search terminates prematurely one may not have a feasible hydraulic design. A more detailed analysis of Jacoby's paper may be found in [34].

Cembrowicz and Harrington [4] have dealt with minimization of the capital cost of a pipe network designed to operate under one loading. Using graph theory they claim to decompose the problem such that the non-convex objective function is broken into subsets of convex functions. Each function relates to a pipe or a loop and is minimized separately, using a method of feasible directions. The number of optimizations may be very large and they have to be scanned to locate the global optimim, so that computationally the method does not seem practical.

Watanatada [39] has developed a method for optimal design of pipeline networks supplied at a number of nodes, and applied it to real networks of moderate size. For a network with P pipes and M nodes, of which MS are supply nodes, the problem is

$$\text{Min} \quad \left[C_T(\boldsymbol{D}, \boldsymbol{H}, \boldsymbol{Q}) = \sum_{p=1}^{P} U_p L_P + \sum_{k=1}^{MS} S_k\right], \tag{23}$$

$$\text{subject to} \quad QR_k\ (\boldsymbol{D}, \boldsymbol{H}, \boldsymbol{Q}) = 0, \quad k = 1,\dots, M, \tag{24}$$

$$D_p \geqslant \text{DMIN}, \quad p = 1,\dots, P, \tag{25}$$

$$H_k \geqslant \text{HMIN}_k, \quad k = 1,\dots, M, \tag{26}$$

$$-QN_k \geqslant 0, \quad k = 1,\dots, MS, \tag{27}$$

where:

$\boldsymbol{D}$ = a vector of P pipe diameters,

$\boldsymbol{H}$ = a vector of M node heads,

$\boldsymbol{Q}$ = a vector of MS supply rates at supply nodes

$U_p = U_p(D_p)$, the cost per unit length of pipe as a function of its diameter.

$S_k = S_k(QN_k, H_k)$, the cost of supplying QN_k at a head H_k.

QR_k = algebraic sum of flows leaving the node.

This constrained optimization formulation was converted into an unconstrained one by using a variable transformation due to Box [3]. Utility variables Z_i, $i = 1,\dots, (P + M + MS)$ are defined by

$$D_p = \text{DMIN} + Z_p^2, \quad p = 1,\dots, P, \tag{28}$$

$$H_k = \text{HMIN}_k + Z_{P+k}^2, \quad k = 1,\dots, M, \tag{29}$$

$$QN_k = -Z_{P+M+k}^2, \quad k = 1,\dots, MS, \tag{30}$$

and the problem now becomes

$$\text{Min} \quad [C_T(Z)], \tag{31}$$

$$\text{subject to} \quad QR_k(Z) = 0, \quad k = 1,\dots, M. \tag{32}$$

A new function is now defined by combining constraints (32) with the objective function (34) according to the method suggested by Haarhoff and Buys [18]:

$$\text{Min} \left[F^r(Z) = C^T(Z) + \sum_{k=1}^{M} E_k^r QR_k(Z) + W \sum_{k=1}^{M} QR_k^2(Z)\right] \tag{33}$$

The superscript r is an iteration counter; E^r and W are penalty multipliers— the first is updated at each iteration and the second is a preassigned fixed constant. The variable metric method of Fletcher and Powell [14] was used to minimize (33), and was found to be superior to the Fletcher and Reeves method of conjugate directions[15], even though the former required more computer memory. Once the optimization is terminated, the diameters have to be rounded to commercially available values, and this has to be done in a way which will maintain hydraulic feasibility. Watanatada's method has the advantage that the flow solution is incorporated directly into the optimization, and one therefore does not need a network solver as a separate program. The danger is, however, that if the procedure terminates prematurely the solution may not be feasible hydraulically.

A method based on Abadie's [1] GRG method was developed by the author [35], in which the design and operation under a number of loadings are to be optimized. For a network with N nodes, operating under L loadings the problem is:

$$\text{Min} \quad \left[F(d, u, x, s) = f(d) + \sum_{l=1}^{L} w^l c^l(d, u^l, x^l, s^l)\right], \tag{34}$$

$$\text{subject to} \quad d \in D \tag{35}$$

$$u^l \in U^l, \quad \forall l \tag{36}$$

$$[G^l(d, u_,^l, x^l, s^l)] = \mathbf{0}, \quad \forall l \tag{37}$$

$$x^l = \{x \mid [G^l(d, u^l, x^l, s^l)] = \mathbf{0}\} \in X^l. \quad \forall l \tag{38}$$

where:

d = the design variables (pipe diameters, pump capacities), which have to belong to the set D.

u^l = the operation variables (valves and pumps ON/OFF) for the l-th loading, which have to belong to the set U^l.

x^l = the dependent variables (heads, consumptions) of the l-th flow solution, which have to be within ranges given by X^l.

s^l = the independent (fixed) variable in the l-th flow solution.

f = cost function of the design.

c^l = cost of operation for the l-th loading.

w^l = weights.

$[G^l] = \mathbf{0}$ = a set of simultaneous node continuity equations (eq.(1)) for the l-th loading. The individual equations are $G^l = 0$, $j = 1, \ldots, N$.

It is assumed that the sets D, U^l and X^l simply specify a range of values for the corresponding variables. The L sets of equations (37) are combined with the objective function (34) to form the Lagrangian:

$$\mathscr{L}(d, u, x, s, \lambda) = F(d, u, x, s) + \sum_{l=1}^{L}\sum_{j=1}^{N} \lambda_j^l G_j^l(d, u^l, x^l, s^l)$$

$$= F + [G]^{\mathrm{T}} \cdot [\lambda] \tag{39}$$

where T stands for the transpose. At any point (d, u) the following has to hold

$$\left[\frac{\partial \mathscr{L}}{\partial x}\right] = \mathbf{0} = \left[\frac{\partial F}{\partial x}\right] + \left[\frac{\partial G}{\partial x}\right]^{\mathrm{T}} [\lambda]. \tag{40}$$

Since G and x can be separated into the independent flow problems, (40) can be decomposed into

$$\left[\frac{\partial \mathscr{L}}{\partial x^l}\right] = \mathbf{0} = \left[\frac{\partial F}{\partial x^l}\right] + \left[\frac{\partial G^l}{\partial x^l}\right]^{\mathrm{T}} [\lambda^l], \quad l = 1, \ldots, L. \tag{41}$$

The Lagrange multipliers, λ, are therefore solved in L groups of N values each from equation (41). The matrix $[\partial G^l / \partial x^l]$ is the last Jacobian of the l-th flow solution by the Newton–Raphson method, and is therefore available directly as a by-product of the flow solution. Once the λ's have been obtained, the components of the (reduced) gradient are computed from

$$[\nabla F] = \begin{bmatrix} \nabla d \\ \nabla u \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \mathscr{L}}{\partial d} \\ \dfrac{\partial \mathscr{L}}{\partial u} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial F}{\partial d} \\ \dfrac{\partial F}{\partial u} \end{bmatrix} + \left[\frac{\partial G}{\partial d}; \frac{\partial G}{\partial u}\right]^{\mathrm{T}} [\lambda]. \tag{42}$$

At a (local) optimum $[\nabla F] = \mathbf{0}$. At any other point $[-\nabla F]$ points in the direction of the steepest descent of F, while changes in the flow solutions due to a move in this direction are already taken into account. A move in this direction is now made, using a one-dimensional search procedure. At the new point the L flow problems are solved and the reduced gradient calculation repeated. The constraints on d, u^l, and x^l are used in selecting the step size in each move. The search is terminated by given criteria for improvement between interations, value of the gradient, number of iterations, etc.

Because flow solutions are computed at each step, and because the constraints on the decision variables are not violated during the moves, the current solution is feasible (this is essentially an interior point method), and if it terminates prematurely one at least has a feasible solution—which is better than the one having started with.

7. Other methods

Deb and Sarkar [7] based their method on the concept of equivalent diameters, for least cost design of a network operating under a single loading and in which the heads are assumed to be known. All pipes in the network are replaced by pipes of a fixed length and a diameter which makes the "equivalent" to the actual pipes. Combining the pipe flow equation with its cost function, then differentiating the cost with respect to diameter and setting equal to zero, and optimality condition for this special type of network is reached. The results are therefore quite limited in application.

Deb [6] again used a similar combination of the flow and cost equation, and developed a method for optimal design of a system consisting of a pumping station, an elevated reservoir and a pipe network fed from it by gravity. The results are further restricted by the fact that the shape of the pressure surface over the network is assumed to have a specific form. A series of runs with the reservoir at different locations and with different parameters in the pressure surface equations was used to test the sensitivity of the optimal solution to changes in these factors.

8. Operation over time

Some of the methods described above can be used to reach optimal operating decisions for existing networks. This is done by fixing all design variables at their actual values and carrying out the optimization for the operational variables. It should be borne in mind, however, that the loadings which are considered in this analysis represent critical or typical conditions, and in no way do they reflect the time sequence of operating conditions over the day or the week. Special methods have therefore to be developed for optimization of the operation over time, in which the only decision variables are the operation of pumps and the setting of valves during the specified time horizon.

Dreizin et al. [11] used a hydraulic simulator of a particular water system as the basic building block in a program which attempted to improve operating policies. The decision variables were those water levels (called set-points) in specified reservoirs at which pumps are to be switched on or off. No algorithmic optimization method was found for solving the problem, and a sequence of simulations with response surface analysis (a gradient-like search) was used.

Some work in the City of Philadelphia [8] resulted in selection of operating policies over a day based on a comparison of costs for several proposed policies. No optimization was attempted.

Sterling and Coulbeck [37, 38] optimized pumping costs in a water system, using dynamic programming and a two-level hierarchical approach. A group at the University of Cambridge, England, have been engaged in development of on-line control and optimization of the operation of regional water systems [12, 13] which are being implemented and tested in the field.

9. Summary and conclusions

Much has been done in developing methods for optimal design of water distribution systems. Some of the more recent work has already been proven in practical applications, and what is needed now is the transfer of this technology to a broad sector of the engineering profession. Much less has been achieved in optimal operation, and as energy costs are increasing more attention should be given to this area. But least of all has been done in the area of planning. Future efforts should be concentrated on screening models, which attempt to specify the basic layout and component location. Such models should address questions such as:

(a) The capacities of the sources, i.e. the development of the various intakes, supply reservoirs, wells and/or treatment plants, and how much each of them should supply to the system.

(b) Balancing of source and pump station capacities vs. storage within the system. This phase deals with the overall amount of storage and pumping capacity, and not necessarily with the location and sizing of individual components.

(c) The layout and capacity of the feeder main grid. This is closely linked with planning of storage and pumping.

The results of such screening models will then have to be handed over for design optimization, during which more insight may be gained, and the need may arise to re-run the screening model with new data.

References

[1] J. Abadie, "Application of the GRG algorithm to optimal control problems" in: J. Abadie ed., *Integer and Nonlinear Programming* (North-Holland, Amsterdam, 1970) 191–211.

[2] E. Alperovits and U. Shamir, "Design of optimal water distribution systems", *Water Resources Research* 13 (6) (1977) 885–900.

[3] M.J. Box, "A Comparison of several current optimization methods, and the use of transformations in constrained problems", *The Computer Journal* 9 (1966) 67–77.

[4] R.G. Cembrowicz and J.J. Harrington, "Capital-cost minimization of hydraulic network", *Proceedings of the ASCE* 99 (HY3) (1973) 431–440.

[5] A.K. Deb, "Least cost design of branched pipe network systems", *Proceedings of the ASCE* 100 (EE4) (1974) 821–835.

[6] A.K. Deb, "Optimization of water distribution network systems", *Proceedings of the ASCE* 102 (EE4) (1976) 827–851.

[7] A.K. Deb and A.K. Sarkar, "Optimization in design of hydraulic networks", *Proceedings of the ASCE* 97 (SA2) (1971) 141–159.

[8] R. DeMoyer, H.D. Gilman and M.Y. Goodman, "Dynamic computer simulation and control methods for water distribution systems", Final Report by General Electric to the Office of Water Resources Research, Contract No. 14–31–0001-3734 (1973) 243.

[9] R. DeMoyer and L.B. Horwitz, *A systems approach to water distribution modeling and control* (Lexington Books, 1975) 143.

[10] R. de Neufville, J. Schaake and J.H. Stafford, "Systems analysis of water distribution networks", *Proceedings of the ASCE* 97 (SA6) (1971) 825–842.

[11] Y. Dreizin, N. Arad, U. Shamir and H. Matmon, "Evaluation of energy saving in regional water systems" (in Hebrew), Report 131. 105, Research and Desalination Div., Mekorot Water Co., Israel, (1971).

[12] F. Fallside and P.F. Perry, "Decentralized optimum control methods for water distribution systems optimization" CUED/B-ELEC/TR31, Department of Engineering, University of Cambridge (1974).

[13] F. Fallside and P.F. Perry, "Hierarchical optimization of a water supply Network", *Proceeding of the IEEE* 122 (2) (1975) 202–208.

[14] R. Fletcher and M.J.D. Powell, "A rapidly convergent descent method for minimization", *The Computer Journal* 6 (1963) 163–168.
[15] R. Fletcher and C.M. Reeves, "Function minimization by conjugate gradients, *The Computer Journal* 7 (1964) 149–154.
[16] I. Gupta, "Linear programming analysis of a water supply system", *AIIE Transactions* 1(1) (1969) 56–61.
[17] I. Gupta, M.Z. Hassan and J. Cook, "Linear programming analysis of a water supply system with multiple supply Points", *AIIE Transactions* 4(3) (1972) 200–204.
[18] P.C. Haarhoff and J.D. Buys, "A new method for the optimization of a nonlinear function subject to nonlinear constraints", *The Computer Journal* 13(2) (1970) 178–184.
[19] D. Hamberg, "Optimal locations of pumping stations in a braching network", (in Hebrew), M.Sc. Thesis, Faculty of Civil Engineering, Technion-Israel Inst. of Tech. (1974).
[20] S.L.S. Jacoby, "Design of Optimal Hydraulic Networks", *Proceedings of the ASCE* 94 (HY3) (1968) 641–661.
[21] E. Kally, "Pipeline planning by dynamic computer programming", *Journal AWWA* 61(3) (1969) 114–118.
[22] E. Kally, "Automatic planning of the least-cost water distribution network", *Water and Water Engineering* 75(902) (1971) 148–152.
[23] E. Kally, "Computerized planning of the least-cost water distribution network", *Water and Sewage Works* Reference 1972, R121 to R127.
[24] D. Karmeli, Y. Gadish and S. Meyers, "Design of optimal water distribution networks", *Proceedings of the ASCE* (94) (PL1) (1968).
[25] C. Kohlhass and D.E. Mattern, "An algorithm for obtaining optimal looped pipe distribution networks", In: Papers of the 6th Annual Symposium on the Application of Computers to the Problems of the Urban Society, Association of Computing Machinery, New York, (1971) 138–151.
[26] D. Lai and J.C. Schaake, "Linear programming and dynamic programming application to water distribution network design", Part 3 of: "Engineering Systems Analysis of the Primary Water Distribution Network of New York City", Dept. of Civil Engineering, M. I. T. (1969).
[27] P.F. Lemieux, "Minimum cost design of water-pipe networks", M.Sc. Thesis, Dept. of Civil Engineering, M. I. T. (1965).
[28] T. Liang, "Design conduit system by dynamic programming", *Proceedings of the ASCE*, 97 (HY3) (1971) 383–393.
[29] R. Pitchai, "A model for designing water distribution pipe networks", Ph. D. Thesis, Harvard University (1966).
[30] H.S. Rao, and D.W. Bree Jr., "Extended period simulation of water systems— part A", *Proceedings of the ASCE* 103 (HY2) (1977) 97–108.
[31] H.S. Rao, L.C. Markel and D.W. Bree Jr., "Extended period simulation of water systems—part B", *Proceedings of the ASCE* 103 (HY3) (1977) 281–294.
[32] H.J. Rasmusen, "Simplified optimization of water supply systems", *Proceedings of the* ASCE 102(EE2) (1976) 313–327.
[33] U. Shamir, "Minimum cost design of water distribution networks Unpublished Report, Dept. of Civil Engineering, M. I. T. (1964).
[34] U. Shamir, "Water distribution systems analysis", Report RC4389, IBM Watson Research Center, Yorktown Heights, N.Y. (1973).
[35] U. Shamir, "Optimal design and operation of water distribution systems", *Water Resources Research* 11(4) (1974) 27–36.
[36] U. Shamir and C.D.D. Howard, "Water distribution systems analysis", *Proceedings of the ASCE* 94 (HY1) (1968) 219–234.
[37] M.J.H. Sterling and B. Coulbeck, "Optimization of water pumping costs by hierarchial methods", *Proceedings of the Institution of Civil Engineers* 59(2) (1975) 789–797.
[38] M.J.H. Sterling and B. Coulbeck, "A dynamic programming solution to optimization of pumping costs", *Proceedings of the Institution of Civil Engineers* 59(2) (1975) 813–818.
[39] T. Watanatada, "Least cost design of water distribution systems" *Proceedings of the ASCE* 99 (HY9) (1973) 1487–1513.