On-line monitoring of water distribution networks

A. Bargiela

ON-LINE MONITORING OF WATER DISTRIBUTION NETWORKS

by

A. BARGIELA

A thesis submitted to the

Faculty of Science

University of Durham

for the degree of Doctor of Philosophy

May, 1984

ACKNOWLEDGEMENTS

This thesis is a presentation of the research done by the author in the Department of Engineering, University of Durham, during the period from December 1981 to November 1983 under the supervision of Professor M.J.H. Sterling.

I am indebted to Professor Sterling for drawing my attention to the vital problems in the water supply industry and for continuous stimulation throughout the course of the project. I am also grateful to him for the excellent research facilities dedicated to the project and for his comments on the final version of this thesis.

I would also like to thank my friends and colleagues in the Power Systems Research Group, particularly to Dr. M. Irving for many fruitful discussions.

To my wife Francesca I express my deep appreciation for understanding and moral support throughout the course of this work. Thanks are also due to my parents for their continuous encouragement.

The financial support of the Science Research Council, who financed my post of Senior Research Assistant at the University of Durham is gratefully acknowledged.

ABSTRACT

This thesis is concerned with the development of a computer-based, real-time monitoring scheme which is a prerequisite of any form of on-line control. A new concept, in the field of water distribution systems, of water system state estimation is introduced. Its function is to process redundant, noise-corrupted telemeasurements in order to supply a real-time data base with reliable estimates of the current state and structure of the network. The information provided by the estimator can then be used in a number of on-line programs.

In view of the strong nonlinearity of the network equations, two methods of state estimation, which have enhanced numerical stability, are examined in this thesis. The first method uses an augmented matrix formulation of a classical least-squares problem, and the second is based on a least absolute value solution of an overdetermined set of equations. Two water systems, one of which is a realistic 34-node network, are used to evaluate the performance of the proposed methods.

The problem of bad data processing and its extension to the validation of network topology and leakage detection is also examined. It is shown that the method based on least absolute values estimation provides a more immediate

indication of erroneous measurements. In addition, this method demonstrates the useful feature of eliminating the effects of gross errors on the final state estimate.

The important question of water system observability is then studied. Two original combinatorial methods are proposed to check topological observability. The first one is an indirect technique which searches for a maximum measurement-to-branch matching and then attempts to build a spanning tree of the network graph using only the branches with measurement assignment. The second method is a direct search for an observable spanning tree. A number of systems are used to test both techniques, including a 34-node water supply network and an IEEE 118-bus power system.

The problem of minimisation of distributed leakages is solved efficiently using a state estimation technique. Comparision of the head profile achieved for the calculated optimal valve controls with the standard operating conditions for a 25-node network indicates a major reduction of the volume of leakages.

In the final part of this thesis a software package, which simulates the real-time operation of a water distribution system, is described. The programs are designed in such a way that by replacing simulated measurements with live telemetry data they can be directly used for water network monitoring and control.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 EVOLUTION OF THE CONTROL OF WATER DISTRIBUTION SYSTEMS

During the past fifteen years there has been a considerable investment of research in the field of water distribution systems. The reasons for this are complex but perhaps the most important is the fact that a typical water network expanded to the point where the ability of the human operator to perceive and process information became a hindering factor in achieving optimal operational decisions.

In the first attempt to aleviate the problem the use of a hard-wired logic [240] later replaced by computer software [275], [43] was proposed. In these schemes selected telemetered variables were used for control purposes, and as they exceeded certain limits, supplied by the operator as set points, control action was triggered. This approach was found to be simple and easy to implement but it suffered from several major drawbacks. First, it provided no insight into the actual state of the network since it treated the system as a control 'black box'. Secondly, the controls were

devised for normal operating conditions and such events as pump failure or heavy industrial consumption could result in improper control action. And finally, the policy of a single variable feedback ignored the interdependence of different measurements and controls.

The application of a regression model in a computerised control scheme [126], [72] overcame the latter drawback of a heuristic approach, nevertheless its usefulness remained limited to standard operation of the network.

A natural progression from the heuristic approach was to develop a mathematical model of a water network and to compute flows and pressures for predicted consumptions. This aproach was introduced, with respect to water systems, by Gilman, Goodman and Metkowski [122] and then used by Gilman, Demoyer and Goodman [125] for controlling a pressure district in Philadelphia.

Better understanding of the relationships between operational parameters of the water network, which came from simulation studies, resulted in a more systematic approach to the problem of optimal network operation. Different researchers [106], [107], [250], [251] proposed new, more efficient methods of optimisation of the pumping cost which, from the point of view of their computational requirements, were applicable to on-line control. However, all of these

algorithms involved solving an exactly determined set of network equations. As a result, an erroneous estimate of a consumer load, or an inaccurate value of a pipe parameter could invalidate all computations. This fact was realised early and it stimulated extensive research aimed at improving accuracy of consumption prediction [243], [108], [201] and validation of static and dynamic network parameters [202], [203]. A complementary approach, where the calculated pipe flows are iteratively corrected so as to force an agreement between measured and calculated nodal heads, was proposed in [67], [68] and is known as an inverse network solution. Unfortunately, computational requirements of this method are prohibitively high due to the slow convergence.

The need for another, more flexible approach to water network monitoring is also emphasized by the fact that many of the water authorities have recently implemented computerised telemetry systems which apart from the usual pump head, pump flow and reservoir level indications can provide in real-time a whole range of additional measurement information such as control valve openings or head in selected network nodes.

Our present work introduces a technique which is computationally efficient and which processes all available measurements. This technique, new in the field of water systems, is known as state estimation and has been

successfully applied in the electrical power system control [220], [221], [222], [161]. In the context of the water supply industry this research demonstrates applicability of computer-based, advanced network monitoring which is a prerequisite of any form of on-line control. The problems discussed in this thesis complement other research concentrated on minimisation of pumping cost. A strong economical motivation for efficient identification of pipe fractures and suppression of distributed leakages comes from the fact that, according to the water authorities reports, up to 30% of the water available may be lost through leakages in the pipe network.

1.2 THE ROLE OF STATE ESTIMATION IN THE WATER SYSTEM OPERATION

A radical change in the philosophy of water system operation manifests itself by the emergence of computerised telemetry which can potentially give greater insight into the system state. However this potential can only be fully realised if the appropriate information processing software is employed in order to monitor the network in terms of variables which are convenient for use both by a human operator and control algorithms. It is now a common practice to describe a water network by defining heads at all network nodes and inflows at fixed-head nodes which are the components of the state vector of the system. Given this information and the static parameters of the network, all

other variables of interest, such as pipe flows or consumer loads, may be calculated immediately.

The most straightforward method of obtaining the values of the state variables is to directly measure them. However, taking into account the size of a real-life water network and the cost of instrumentation and associated telemetry, this possibility becomes impractical.

Another way of finding the system state is to solve a set of mass-balance equations which are defined using the network topology data, the measured or estimated consumer loads and the inflows into the system. This method, if well implemented, is computationally efficient and is now widely used in off-line water network simulation studies. However, its application to on-line control scheme encounters the following difficulties:

1) The algorithm has no systematic way of dealing with measurement inconsistency. If one measurement is incorrect or lost, the load-flow approach gives incorrect results or no result at all, respectively, since it processes a set of independent equations.

2) The load-flow method provides no means to assess the confidence on the final results.

3) The input data is limited to the system inflows and consumer loads which, as in the case of predicted

values, may carry considerable errors. At the same time readily available measurements of other variables are not used.

For real-time water network monitoring it is therefore more effective to determine the system state using a state estimation procedure. This method overcomes difficulties associated with the load-flow solution by processing all available measurements and formulating the problem in terms of redundant equations. In effect, even in the presence of bad data, or when pieces of data are lost, it is still possible to obtain a good state estimate. Unlike the load-flow method, the state estimator also provides an indication of the accuracy of the estimates it produces. Another important feature of the state estimator is its ability to detect the presence of bad data and to evaluate the residual errors of the measurements.

The main assumption for successful performance of a state estimation program is that the measurement set should possess a degree of redundancy which enables erroneous information to be filtered out. In water distribution systems this redundancy is usually low and can be achieved only if the measurement information is combined with the pseudomeasurements (representing consumption prediction or information about network topology). By increasing the number of measurements it is possible therefore to improve both the reliability and accuracy of state estimation. On

the other hand, if some measurements are lost or are found incorrect it is necessary to check whether the measurement set still renders the network observable.

The state estimation problem can be split into the following three distinct subproblems:

a) Observability - This subproblem deals with the determination of measurement sets which allow the estimation of the state variables. If the system is found unobservable with respect to the measurement set concerned an appropriate pseudomeasurement is generated;

b) Estimation - Process of computing the state estimates from the knowledge of the measurements, consumption predictions, network structure and parameters;

c) Bad Data Detection and Identification - Procedure to check the presence of structural errors and/or bad data, and to identify which measurements carry gross errors, or which part of the structure is not properly modelled.

The main functions of the state estimator can now be enumerated:

1) Computation of a reliable, real-time data base from redundant, noise corrupted telemeasurements. This

includes the estimation of nodal heads and fixed-head-node flows wherever they are not directly measured, or where telemeter failure temporarily makes the data unavailable;

2) Detection, identification and suppression of bad measurements due to telemetry or instrumentation malfunction, inaccuracy in network parameters and unreported status changes;

3) Provision of data for real-time monitoring and on-line control algorithms;

4) Provision of a log of system states which is then used for the prediction of consumer loads.

The overall configuration of the state estimator is given in Fig. 1.1

1.3 A REVIEW OF THE LITERATURE ON POWER SYSTEM STATE ESTIMATION

The problem of estimating the state of a power system was initially formulated and developed by Schweppe, Wildes and Rom in a series of three papers [220], [221], [222]. The first paper introduces the weighted least-squares method as the algorithm to obtain the estimates. The inverse of the noise covariance matrix is chosen as weighting matrix since this choice yields the minimum variance unbiased estimates when the measurement errors are assumed to be normally

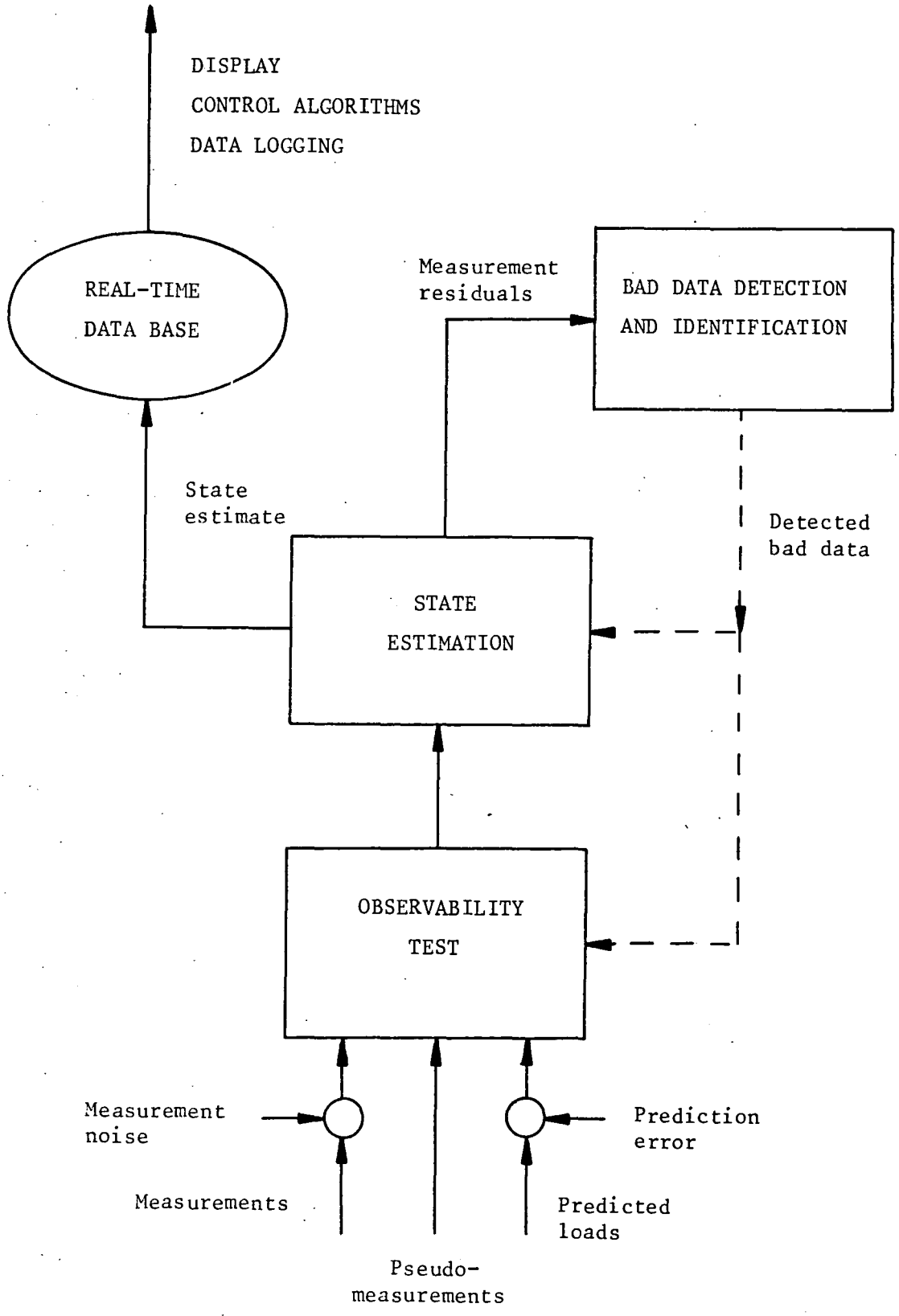


Figure 1.1 Overall configuration of the state estimation

distributed. The problem of detection and identification of errors is outlined, and the importance of factors such as redundancy and pseudomeasurements is also stressed. Reference [221] addresses itself to approximate measurement models for real-size power systems. Finally, reference [222] deals with the actual implementation of state estimators, especially with regard to time and storage requirements.

Stagg, Dopazo et al. compare the weighted least-squares technique and a method based on the solution of certain independent equations [238]. Their method consists of successive conventional load flow solutions and the use of a sensitivity transformation matrix to perform error analysis and to compute the unavailable bus values. The numerical examples presented in the paper suggest that the independent equations give better results. However, it would appear that the method requires the prior conditioning of measurements, which seems to be infeasible in practice. Numerical problems are encountered in solving the least-squares problem, as a consequence of ill-conditioning of the normal equation for the low level of redundancy used.

Reference [235] uses the weighted least-squares approach for state estimation with a weighting matrix depending on the relative importance of the data. As a result, it does not yield minimum variance estimates, which are desirable from a statistical point of view, but it gives a solution which is intuitively the most reliable.

The AEP 'lines-only' algorithm is introduced in reference [85], complemented in [86] and further discussed in reference [87]. This method is a modification of a weighted least-squares approach in which the voltage drop across a network element is related to bus voltages and line flows. The algorithm processes measurements of line power flows only. It has some numerical advantages, since the coefficient matrix of the linear system to be solved at each iteration is constant with respect to the states. Moreover, this matrix is as sparse as the bus admittance matrix, and is also symmetrical and real. This last property allows the separate computation of the real and imaginary parts of the states. The method's most frequently cited disadvantages are the lack of flexibility in choosing the measurements and difficulties in detecting and identifying bad data.

Reference [176] deals with tracking estimators, that is, estimators capable of following the time evolution of the power system. The estimator is viewed as a digital feedback loop, which uses new measurements to obtain new estimates by improving an old estimate via a feedback error signal operating through a gain matrix. The best gain matrix is the error covariance matrix updated at each time step. Several simplified gain matrices are studied in an attempt to develop methods to save computation time and storage while still providing good estimates.

Programming techniques for the implementation of the

weighted-least squares algorithm are suggested in [9].

In reference [10], a method is developed to accommodate buses with zero injections by considering equality constraints for each of them, instead of taking their (zero) injections as measurements of high accuracy. The results reported in [10] show that the method is computationally advantageous.

Another method of power system state estimation, based on the least absolute values optimisation technique, is proposed in reference [145]. The state estimates obtained are spanned by a minimal set of linearily independent equations with the least residual errors. The method proves to be very efficient in detection of gross measurement errors, while retaining a useful degree of noise filtering.

The matter of ill-conditioning of the normal equation approach used in the conventional solution of the state estimation problem is addressed in [208]. The Levenberg-Marquardt algorithm is suggested to cope with the ill-conditioning problem. Other techniques based on the Golub's orthogonal method and on the Peters-Wilkinson algorithm are investigated in [233] and [129], respectively.

Reference [223] presents a comprehensive survey of the power system state estimation problem, including the description of several estimation algorithms and their computational requirements.

1.4 THESIS ORGANISATION

Chapter II briefly reviews the conventional state estimation technique based on the normal equation approach and then introduces two methods which are designed to overcome numerical problems and to achieve a high computational efficiency. Both estimators are applied to the water systems, described in Appendix C, and the corresponding results are compared.

In Chapter III, two methods for detection and identification of bad data are described. The first method, based on hypothesis testing, is to be used in connection with the least-squares estimator. The second method takes advantage of the characteristics of the least absolute values estimator and is shown to be suitable for identification of both topological and gross measurement errors.

The observability problem is investigated in Chapter IV. By using an approximate linear measurement model the observability problem is proved to be equivalent to the search for an observable spanning tree in the augmented network graph. Two original methods to determine topological observability of a water network are then proposed. First, an algorithm based on the matching in the bipartite measurement-to-edge graph is described and applied to several different networks including a realistic 34-node

system. Next, a direct method of seeking an observable spanning tree, using the path property, is developed and tested on the same networks used to assess the performance of the matching method.

Chapter V deals with the problem of reduction of distributed leakages in the water network which complements the identification of bursts catered for by network monitoring. A new method of calculating the optimal valve controls, based on the application of the state estimation techniques, is presented. The economy of the optimal valve control policy is shown by comparison of the volume of leakages for uncontrolled, manually controlled and optimally controlled network.

A general description of the on-line software package for real-time monitoring of a water supply network is given in Chapter VI. The advantages of a modular structure of the package are discussed and block diagrams presenting the flow of information and means of coordination of the simultaneously executed tasks are given.

Finally, Chapter VII presents the main conclusions of this work and suggests areas of futher research work.

1.5 MAJOR CONTRIBUTIONS OF THIS THESIS

The main contributions of this thesis can be enumerated as follows:

a) Introduction of a state estimation technique for the purpose of real-time monitoring of a water distribution network;

b) Development of two state estimators, based on the least absolute values and least-squares optimisation techniques, respectively, implementation of which includes sparsity and numerical stability considerations;

c) An efficient method of calculating the residual sensitivity matrix to be used in connection with the augmented matrix state estimator;

d) Extension of a bad data identification procedure, using measurement residuals calculated by the least absolute values estimator, to the identification of topological errors;

e) Development of two different combinatorial methods to check topological observability of the water system;

f) Development of a method of reduction of distributed leakages by optimised control of valves installed

in the pipe network;

g) Production of a software package for on-line monitoring and control of a water distribution network which consists of:

- Network Simulation (SYSSYM)
- Telemetry Simulation (SYSTEL)
- Observability Test (OBSMATCH, OBSTREE)
- Estimation (SYSESTLP, SYSESTLS)
- Bad Data Processing (POSLEAK1, RESSENS)
- Valve Control (VALCON)
- Operators' Interface (OPERATOR)
- Graphical Display (NETDIS1)

CHAPTER II

WATER SYSTEM STATE ESTIMATION

2.1 INTRODUCTION

This chapter deals with state estimation methods which are suitable for real-time monitoring of a water system. In view of the strong nonlinearity of the network equations, low measurement redundancy and possibility of gross measurement errors, the state estimation algorithms have been selected with respect to both numerical stability and computational efficiency. The conventional approach via the Gauss normal equation is briefly reviewed in order to expose the fact that it is prone to numerical ill-conditioning. It is also interesting to note that the literature reports problems of ill-conditioning in least-squares solution via the normal equation [165], [238].

One method of circumventing the possibly poorly conditioned normal equation is to employ the orthogonal factorisation method as proposed by Golub [127]. This method performs an orthogonal transformation of the linearised measurement equations that enables the least-squares

solution to be calculated by a backward substitution. Application of the Golub method to power system state estimation [233] proved very successful from the numerical stability point-of-view, but at the same time the algorithm showed considerable reduction in computational efficiency. An alternative procedure, avoiding the ill-conditioning of the direct formation of the normal equation, has been proposed by Peters and Wilkinson [196]. Although their method enables the selection of a compromise between the conflicting requirements of numerical stability and matrix sparsity and, in general, it performs more efficiently than the Golub's method, it may occasionally give very poor results, as reported in reference [93]. The method of Peters and Wilkinson seems to be better suited to systems with high measurement redundancy [129], which is rarely, if ever, the case in water systems.

Another method of solving the least-squares problem, which has been adopted in this work, has been proposed by Siegel [232]. The initial measurement equations and the least-squares optimality conditions are incorporated into a supermatrix system which is then solved using an appropriate factorisation technique. The method is numerically stable since the formation of the normal equation, which augments the condition number, is avoided altogether. Also, the computational efficiency of this method is usually superior compared with the Golub's and Peters-Wilkinson's methods, particularly in the case of low redundancy levels [93].

The augmented matrix method performs well also in its alternative form proposed by Hachtel [93], implementation of which has been reported in reference [149].

Numerical examples, analysed in section 2.4, demonstrate that the augmented matrix state estimator converges in 4-5 iterations even for weak metering configurations. The method shows however that, as in all least-squares type of methods, gross measurement errors tend to have an exaggerated effect on the calculated system state, thus requiring reestimation of the state vector after the bad data has been found and removed.

The second state estimator, proposed in this thesis, overcomes difficulties associated with the presence of gross measurement errors by employing a least-absolute value estimation criterion. This method was first used for power system state estimation by Irving, Owen and Sterling [145], and proved very robust in presence of the error contaminated measurements. The state estimator rejects bad data during the estimation process so that the final estimate is defined only by the 'valid' measurements. Details of an efficient implementation of this algorithm exploiting sparsity considerations has already been published by the author [247].

This chapter is organised as follows. The weighted least-squares estimator is presented in section 2.2. Subsection 2.2.3 addresses itself to the classical normal

equation approach and the augmented matrix method is described in subsection 2.2.4. The least absolute values estimation algorithm and the details concerning its implementation are discussed in section 2.3. Finally, section 2.4 compares the results obtained by applying two estimation techniques to the water systems with different measurement configurations and different measurement noise.

2.2 WEIGHTED LEAST-SQUARES ESTIMATION

2.2.1 The Measurement Model

Consider an N-node water system with F fixed-head nodes for which m measurements are taken. It is assumed that the structure (topology) of the water network and element parameters are known. Under these conditions, the knowledge of the nodal heads and fixed-head-node flows provides enough information to calculate the pipe flows and consumer loads. For this reason the nodal heads and fixed-head-node flows are called the state variables of the water system.

The measurement model for the N-node water system with F fixed-head nodes is derived in Appendix A, and is given by

$$\underline{z} = \underline{g}(\underline{x}) + \underline{\omega} \qquad (2.1a)$$

$$E\{\underline{\omega}\} = \underline{0} \;\; ; \;\; E\{\underline{\omega}\,\underline{\omega}^{T}\} = R \qquad (2.1b)$$

where

m	:	number of measurements
$n = N+F$	:	number of state variables
$\underline{z}$	:	mx1 state measurement vector
$\underline{x}$	:	nx1 state vector
$\underline{\omega}$	:	mx1 zero-mean random vector relating the measured quantities and the state variables
$E\{\cdot\}$	:	expectation operation
R	:	covariance matrix of the measurement errors

The monitored variables whose measurements are the elements of vector $\underline{z}$ in Eq. (2.1a) are usually pump heads and flows, consumer loads, reservoir levels and their rates of change and selected nodal heads. Pipe flows may also be monitored but because of the cost of instrumentation they are less commonly used than the other types of measurements.

In addition to actual measurements it is common to use nontelemetered information about some of the variables of the water system which may be available through the analysis of the past data or implied by the network topology. These additional pieces of information are the so-called pseudo-measurements. Thus, the vector $\underline{z}$ is, in general, formed by actual telemeasurements and by pseudo-measurements.

In order to obtain a meaningful solution for the state estimation problem it is necessary that the number of measurements m is greater or equal to the number of state

variables n. In practice, m is usually greater than n, since this condition is also required for bad data detection and identification purposes, as indicated in Chapter 3.

The global measurement redundancy ρ is defined as

$$\rho = \frac{m}{n} \qquad (2.2)$$

Elements of the nonlinear vector function $\underline{g}(\underline{x})$ of Eq.(2.1a) represent empirical head-flow relationships for network elements and mass-balance for each network node. Rigorously, this function should be denoted as $\underline{g}(\underline{x},\underline{r})$, where $\underline{r}$ is the vector of network parameters. However, it is assumed throughout this work that the network parameters are determined before the on-line state estimation is attempted, so that the use of the simplified notation $\underline{g}(\underline{x})$ is justified.

The measurement errors are represented in the measurement model by the zero-mean random vector $\underline{\omega}$. These errors come from a variety of sources, such as meter inaccuracies, communication errors, effects of analog-to-digital conversions, etc. The measurement errors are usually assumed to be uncorrelated, thus the covariance matrix R is taken as diagonal. The elements of matrix R are the variances of the measured errors, and are in general considered as a function of meter's reading (see section A.4, Appendix A). In case of pseudo-measurements their variances are defined so as to reflect the degree of

uncertainty about the information carried by them. If a pseudomeasurement represents the predicted consumer load its uncertainty will usually be high, and if it represents the topological information about the network it will be very low.

2.2.2 The Least-Squares Objective Function

Given the measurement model of Eq. (2.1), the weighted least-squares formulation of the water system state estimation problem is based on the minimisation of the objective function

$$C(\underline{\hat{x}}) = [\underline{z}-\underline{g}(\underline{\hat{x}})]^T R^{-1} [\underline{z}-\underline{g}(\underline{\hat{x}})] \qquad (2.3)$$

with respect to the vector of state estimates, $\hat{x}$.
The weighting matrix in Eq. (2.3) is the inverse of the covariance matrix of the measurement errors. This choice of weighting matrix is important for the statistical properties of the estimator. For instance, if $\underline{\omega}$ is normally distributed, the estimator based on the minimisation of $C(\underline{\hat{x}})$ yields the minimum variance unbiased estimates.

2.2.3 Solution via the Normal Equation Approach

The minimisation of $C(\underline{\hat{x}})$ presents an unconstrained optimisation problem which can be solved using any optimisation method applicable to this class of problem. However, it turns out that the characteristics of the

nonlinear function in Eq. (2.3) provide a good approximation for its Hessian matrix near the solution, so that it is advantageous to use a second-order algorithm, such as Newton-Raphson method. The method linearises the weighted-least squares problem with respect to a current estimate $\hat{\underline{x}}_k$ at iteration number k, such that an improved estimate of the state vector is obtained from the iterative relationship

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k + \alpha \ \Delta\hat{\underline{x}} \qquad (2.4)$$

where α is a parameter modifying the step length in order to improve the convergence of the method. In particular, the Newton-Raphson method can be made norm reducing by selecting α so as to minimise in each iteration $C(\hat{\underline{x}}_k+\alpha\Delta\hat{\underline{x}})$ In practice however, it is usually sufficient if α is kept constant at $0.5 \leqslant \alpha < 1$.

The correction vector $\Delta\hat{\underline{x}}$ is computed by solving at each iteration

$$H\Delta\hat{\underline{x}} = -G \qquad (2.5)$$

where H is the Hessian matrix and G is the gradient vector of C i.e.

$$G = \frac{\delta C(\underline{x})}{\delta \underline{x}} \qquad (2.6a)$$

$$H = \frac{\delta^2 C(\underline{x})}{\delta \underline{x}^2} \qquad (2.6b)$$

For the function $C(\hat{\underline{x}})$ given by Eq. (2.3), the gradient vector is easily computed as

$$G = -2\, J^T(\hat{\underline{x}}_k)\, R^{-1}\, [\underline{z}-\underline{g}(\hat{\underline{x}}_k)] \qquad (2.7)$$

where

$$J(\hat{\underline{x}}_k) = \left.\frac{\delta \underline{g}(\underline{x})}{\delta \underline{x}}\right|_{\underline{x}=\hat{\underline{x}}_k} \qquad (2.8)$$

: mxn Jacobian matrix, elements of which are given in Appendix C

The least-squares procedure assumes that the Hessian matrix can be approximated near the solution by

$$H = 2\, J^T(\hat{\underline{x}}_k)\, R^{-1}\, J(\hat{\underline{x}}_k) \qquad (2.9)$$

Using Eqs. (2.7) and (2.9) in Eq. (2.5), the correction vector $\Delta\hat{\underline{x}}_k$ can be obtained by solving at each iteration

$$[J^T(\hat{\underline{x}}_k)\, R^{-1}\, J(\hat{\underline{x}}_k)]\, \Delta\hat{\underline{x}} = J^T(\hat{\underline{x}}_k)\, R^{-1} \Delta\underline{z} \qquad (2.10)$$

where

$$\Delta\hat{\underline{x}} = \underline{x}-\hat{\underline{x}}_k \qquad (2.10a)$$

$$\Delta\underline{z} = \underline{z}-\underline{g}(\hat{\underline{x}}_k) \qquad (2.10b)$$

It is easy to see that Eq. (2.10) can also be obtained by applying a linear least-squares method to the linearised measurement model derived in section A.2 of Appendix A.

One suitable stopping criterion for the iterations is

$$\max_i |\Delta\hat{x}_i| \leq \epsilon \qquad (2.11)$$

where ϵ is some pre-specified tolerance.

Equation (2.10) is usually solved by means of Cholesky's factorisation of $A = J^T R^{-1} J$ into LL^T, where L denotes a lower triangular matrix. Such a factorisation is possible because A is symmetric and positive definite since the meter configuration should ensure that it is also non singular. Sparsity and optimal ordering techniques are routinely employed in the process of factorising and solving Eq. (2.10). However, it must be noted that the sparsity of the matrix A is much less than that of the Jacobian matrix J due to the second order fill-ins in the process of forming $J^T R^{-1} J$. Apart from the computational effort involved in calculating a product of the matrices it is also laborious to establish a suitable indexing of the non-zero terms of the normal matrix.

An additional problem associated with the normal equation (2.10) is its inherent tendency to ill-conditioning which comes from the fact that the singular values of $J^T R^{-1} J$ are the squares of those of J. Thus, the solution $\Delta\underline{\hat{x}}$ is more sensitive to perturbations in the normal equation than to perturbations in the linearised measurement equation. Consequently, the normal equation method used for the water system state estimation may prove to be numerically unstable

for the following reasons:

i) The Jacobian matrix is not always well conditioned due to the low global measurement redundancy and/or weak measurement configuration;

ii) The strong nonlinearity of the network equations implies that the computation of the state vector requires several iterations during which the round-off error may be considerably amplified since the condition number of the normal matrix is big.

Next, we present another approach to solve least-squares problem, whose numerical characteristics are superior to those of the straightforward solution via the normal equation. The procedure is capable of significantly reducing the ill-conditioning, thereby assuring a higher level of accuracy and reliability for the iterative solutions.

2.2.4 Solution via the Augmented Matrix Approach

Minimisation of the weighted sum of squared measurement residuals Eq.(2.3) has been shown to result in the following adjustment procedure

$$J^T R^{-1} J \cdot \Delta\hat{\underline{x}} = J^T R^{-1} \cdot \Delta\underline{z} \qquad (2.12)$$

where J is an abbreviated notation for an mxn Jacobian matrix $J(\hat{\underline{x}}_k)$.

In order to overcome the numerical difficulties arising from the direct formation of the normal matrix $J^T R^{-1} J$, the adjustment procedure (2.12) can be formally written as a system of three simultaneous equations

$$\underline{r} = \Delta\underline{z} - J\,\Delta\hat{\underline{x}} \qquad (2.13a)$$

$$\underline{\lambda} = R^{-1}\,\underline{r} \qquad (2.13b)$$

$$J^T\underline{\lambda} = \underline{0} \qquad (2.13c)$$

where $\underline{r}$ and $\underline{\lambda}$ are the mx1 auxiliary vectors which do not have to be calculated explicitly. These equations can now be assembled into a supermatrix structure

$$\begin{bmatrix} 0 & I & J \\ -I & R^{-1} & 0 \\ J^T & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \underline{\lambda} \\ \underline{r} \\ \Delta\hat{\underline{x}} \end{bmatrix} = \begin{bmatrix} \Delta\underline{z} \\ 0 \\ 0 \end{bmatrix} \qquad (2.14)$$

Although the dimension of the augmented matrix is now (m+m+n), compared to n of the normal matrix, the equation (2.14) can be solved very efficiently using a sparse linear equation solution technique. The implementation of the state estimator, used here is a variant of the sparse Bartels-Golub decomposition algorithm [210], available as the Harwell subroutine LA05A. This algorithm takes advantage of the enhanced sparsity of the augmented matrix by performing computationally inexpensive row and column permutations before any eliminations are attempted. In fact, it is often

possible to permute to triangular form and to avoid time consuming pivot operations, with consequent fill-ins, altogether. The augmented matrix formulation of the least-squares problem is also attractive from the point of view of handling the data. The matrix can be constructed instantly from the original Jacobian and error covariance matrices avoiding any arithmetical operations. Unlike the normal equation method, indexing of the nonzero elements of the augmented matrix is also straightforward since it is essentially repetition of the indexing of the Jacobian matrix. An analysis of the conditioning of the system (2.14) carried out by Bjork [32] demonstrates that if the singular values of J are $\gamma_1 \leq \gamma_2 \leq \ldots \leq \gamma_n$ then the matrix in Eq(2.14) has eigenvalues

$$(\sigma \pm \sqrt{\sigma+4\gamma_i^2})/2 \qquad i=1,2,\ldots,n \quad \text{and} \quad \sigma$$

with the eigenvalue σ having multiplicity (2m-n). Thus system (2.14) shows essentially no deterioration in condition number as compared with the original measurement equation. The assumption that the standard deviation of measurement errors σ is the same for all readings does not affect the generality of the analysis.

The performance of the weighted least squares estimator based on the augmented matrix method is evaluated in section 2.4.

2.3 LEAST ABSOLUTE VALUES ESTIMATION

2.3.1 Formulation fo the LAV Estimation Problem

The least-squares method, described in the previous section, is the most common method of estimation of the state vector from an overdetermined set of measurements which gives a minimum-variance, unbiased estimate provided that the measurements are affected solely by Gaussian noise. Unfortunately, this is rarely the case in on-line computer control systems where the measurement inaccuracies are far from Gaussian distribution but in fact contain gross errors such as reversed sign of measurements, large systematic errors or zero readings due to telemetry or instrumentation malfunction. In all these situations the least-squares estimator gives poor results since the erroneous measurement is weighted according to the square of its residual and therefore has an exaggerated effect on the state estimate. Geometrically, the state estimate achieved with the least-squares criterion represents a point in the space of feasible solutions minimising the sum of squares of distances between itself and measurement hyperplanes. Thus, a gross measurement error, corresponding to the hyperplane which is distant from the true solution point will disproportionately affect the final solution.

In order to overcome this problem, various modifications of the original least-squares criterion, decreasing the weighting on large residuals, have been

proposed. Handschin,Schweppe et al. [131] discuss several such modifications and compare their relative performance. However, in general, modified cost functions prove to be less easily handled in computations.

A more effective approach based on minimisation of the modulus of measurement inconsistency,

$$C(\underline{\hat{x}}) = \underline{w}^T \, |\underline{z}-\underline{g}(\underline{\hat{x}})| \qquad (2.15)$$

was proposed in [145]. The solution point is defined here by the intersection of n hyperplanes with the smallest measurement noise. The erroneous measurement corresponding to the deviant hyperplanes, unlike the least-squares formulation, does not influence the solution point since this would increase distances to the remaining n-1 hyperplanes defining the state estimate. This gives rise to a potential for the total rejection of bad data from the measurement set providing that the number of gross errors does not exceed m-n. The penalty incurred in achieving the rejection of bad data is that n measurements spanning the solution are included with their associated measurement noise. However, in the water distribution systems field this effect is negligible.

Solution of the unconstrained, nonlinear optimisation problem (2.15) can be obtained by applying an iterative procedure analogous to (2.4) where the initial guess of the

state vector is sequentially improved until some convergence criterion is satisfied.

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k + \alpha\, \Delta\hat{\underline{x}} \qquad k = 1, 2, \ldots \qquad (2.16)$$

$$\max_i |\Delta\hat{x}_i| \leqslant \epsilon \qquad i = 1, 2, \ldots n \qquad (2.17)$$

where k is the iteration number.

Expanding the measurement vector function $\underline{g}(\hat{\underline{x}})$ by the latest state estimate $\hat{\underline{x}}_k$ using a first order Taylor series the linearised measurement model can be expressed as (see Appendix A.2)

$$\Delta\underline{z} = J(\hat{\underline{x}}_k)\cdot\Delta\hat{\underline{x}} + \underline{\omega} \qquad (2.18)$$

where $\Delta\underline{z} = \underline{z} - g(\hat{\underline{x}}_k)$

$J(\hat{\underline{x}}_k)$ is mxn Jacobian matrix

$\Delta\hat{\underline{x}}$ is nx1 vector of state increment

$\underline{\omega}$ is mx1 measurement noise vector same as in the measurement equation (2.1a)

The nonlinear optimisation problem (2.15) is now simplified to minimisation of the linearised cost function

$$\min_{\Delta\hat{\underline{x}}} \underline{w}^T |\Delta\underline{z} - J(\hat{\underline{x}}_k)\cdot\Delta\hat{\underline{x}}| \qquad (2.19)$$

where $\underline{w}$ is mx1 measurement weight vector elements of which are equal to the diagonal elements of the inverse of the measurement covariance matrix.

Because a direct solution of Eq. (2.19) is computationally inconvenient, due to the presence of a modulus operator, artificial variables $\underline{q}$ and $\underline{s}$ can be introduced with the following definitions

$$s_i = \begin{cases} \Delta\underline{z}_i - J_i(\hat{\underline{x}}_k)\Delta\hat{\underline{x}} & \text{if } \Delta z_i \geq J_i(\hat{\underline{x}}_k)\Delta\hat{\underline{x}} \\ 0 & \text{otherwise} \end{cases}$$

$$q_i = \begin{cases} -(\Delta z_i - J_i(\hat{\underline{x}}_k)\Delta\hat{\underline{x}}) & \text{if } \Delta z_i < J_i(\hat{\underline{x}}_k)\Delta\hat{\underline{x}} \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \ldots, m$

Problem (2.19) can then be represented as

$$\min_{r,s,\Delta\hat{\underline{x}}} \quad \underline{v}^T \begin{bmatrix} \Delta\hat{\underline{x}} \\ \underline{q} \\ \underline{s} \end{bmatrix} \tag{2.20a}$$

subject to

$$\Delta\underline{z} - J(\hat{\underline{x}}_k)\Delta\hat{\underline{x}} + \underline{q} - \underline{s} = 0 \tag{2.20b}$$

where $\underline{v}$ is $(n+m+m)\times 1$ weight vector

$$\underline{v}^T = [0 \,|\, \underline{w}^T \,|\, \underline{w}^T]$$

2.3.2 Solution via the Linear Programming Approach

The system of equations (2.20a) and (2.20b) developed in the previous subsection presents essentially a linear programming problem which can be written in standard form

$$\min \underline{v}^T \underline{y} \qquad (2.21a)$$

subject to

$$E\underline{y} = \underline{b} \quad \text{with} \quad \underline{y} \geq 0 \qquad (2.21b)$$

where

$$E = [J(\hat{\underline{x}}_k) \mid I \mid -I]$$

$$\underline{b} = \Delta\underline{z} + J(\hat{\underline{x}}_k)\,\underline{d}$$

$$\underline{y} = [(\Delta\hat{\underline{x}}+\underline{d})^T \mid \underline{g}^T \mid \underline{s}^T]^T$$

$\underline{d}$ is nx1 vector elements of which are sufficiently large to guarantee $\Delta\hat{x}_i + d_i \geq 0$, $i = 1,\ldots,n$

Solution of this problem is equivalent to selection of the (n+m+m)x1 vector $\underline{y}$ from [(m+n+n)!/(n+n)!m!] candidates such that the objective function (2.21a) is minimised. A solution is <u>feasible</u> if it satisfies (2.21b) and is <u>basic</u> if at least n+m components of $\underline{y}$ are zero. The remaining m components are called <u>basic variables</u> and can be assembled into the mx1 vector $\underline{y}_B$ that satisfies the equation

$$B\underline{y}_B = \underline{b} \qquad (2.22)$$

where B is an mxm <u>basic matrix</u> formed of columns of E corresponding to elements of $\underline{y}_B$.

The classical method of solving (2.21b) is to move from one basic feasible solution to another improving at each stage the objective function. This is known as the Simplex method. Such a process is guaranteed to terminate in

a finite number of steps since the number of corner points of the polytope defined by (2.21b) is limited and each consecutive solution for y_B gives a smaller value of the objective function (nondegenerate case).

Since the constraint matrix E is sparse being formed of the Jacobian matrix and two unit matrices, the basis matrix is also sparse and consequently Eq. (2.22) can be solved efficiently using the sparsity exploiting techniques. For reasons of numerical stability and sparsity preservation it is advantageous to use the elimination form of the inverse when solving the set of linear equations (2.22). The Bartels-Golub decomposition is one of the suitable techniques [21]. This technique is also amenable for further refinements which consist of row and column interchanges before any pivoting is performed [210].

The algorithm, as implemented by Reid in form of Harwell subroutine LA05A, begins with the application to the original basis of Gaussian elimination with row and column interchanges. It is convenient to express this elimination in the form of the equation

$$M_r M_{r-1} \ldots M_1 B = P\ U\ Q \qquad (2.23)$$

where each M_i is a matrix which differs from I in just one off-diagonal element (and therefore represents an elementary row operation), P and Q are permutation matrices and U is upper triangular. The solution of (2.22) can now be

easily obtained since the function (2.23) allows the basis inverse B^{-1} to be expressed in the form

$$B^{-1} = Q^T U^{-1} P^T M_r M_{r-1} \ldots M_1 \qquad (2.24)$$

After an iteration of the Simplex algorithm the new basis, say $\bar{B}$, differs from B in just one column and so satisfies the equation

$$M_r M_{r-1} \ldots {}_1\bar{B} = P S Q \qquad (2.25)$$

where S differs from U in this column in which B and $\bar{B}$ differ. Therefore S has in general the triangular form with a 'spike'. The algorithm attempts to restore the upper triangular form first by permuting columns so as to produce the upper Hessenberg form and then by including row interchanges to eliminate the subdiagonal elements. If these row operations are written as $M_{r+1}, \ldots, M_r$ then the new factorisation can be expressed as

$$M_{r+1} \ldots M_r \ldots M_1 \bar{B} = \bar{P} \bar{U} \bar{Q} \qquad (2.26)$$

which is exactly of the form (2.23).

Computational efficiency of the Simplex method clearly depends on the efficiency of updating the inverse of the basis matrix and on the number of Simplex iterations. It is the author's belief that the Reid's algorithm give the state-of-the-art solution to the former problem. The latter one has been solved by the author by proposing a procedure which enables the construction of a feasible basis which

needs very few Simplex iterations to produce an optimal solution [247]. The procedure takes advantage of the fact that in on-line state estimation a good approximation of the state vector is usually available as its value from the previous step. This procedure can be summarised as follows:

1. Construct matrix F consisting of the columns of the Jacobian matrix corresponding to the basic state variables $\underline{y}^A$ in the previous optimal solution.

2. Factorise matrix F into two factors F^A and F^R where

   F^A consists of the rows of F which represent measurements with zero slack variables in the optimal solution ($\dim F^A \leqslant n$)

   F^R consists of the remaining rows of F

$$F^T = \left[\begin{array}{c} F^A \\ \hline F^R \end{array}\right]$$

and the corresponding right-hand-side vector is

$$\underline{b} = \left[\begin{array}{c} \underline{b}^A \\ \hline \underline{b}^R \end{array}\right]$$

3. Solve the reduced set of linear equations

$$F^A \underline{y}^A = \underline{b}^A$$

4. Form an initial feasible basis for the Simplex algorithm following the scheme:

IF

$$y_j^A \geq 0 \quad \text{for all} \quad j=1,\ldots,\dim F^A$$

THEN

$$B = \left[\begin{array}{c|c} F^A & 0 \\ \hline F^R & U \end{array}\right]$$

$$\underline{b} = \left[\begin{array}{c} \underline{b}^A \\ \hline \underline{b}^R - F^R \underline{y}^A \end{array}\right]$$

and U is a diagonal matrix with elements

$$u_{ii} = \operatorname{sgn}(b_i^R - F_i^R \underline{y}^A), \quad i=1,\ldots,(m-\dim F^A)$$

ELSE

B is chosen in conventional way

The restarting algorithm also provides a means for bypassing phase one of the Simplex method provided the set of 'valid' measurements spans the state vector. The low measurement redundancy suggests that for the purpose of water system state estimation the primal version of the Simplex algorithm can be profitably used. Such a formulation enables the error contaminated measurements to be removed efficiently by zeroing the associated measurement weights, thus avoiding the need for refactorisation of the basis matrix. Also, the unit matrices appearing in the structure of the constraint matrix E can be directly incorporated into a pivotal strategy.

Numerical stability of the linear programme depends on the stability of the basis updating scheme. Reid's algorithm allows a compromise to be chosen between stability and sparsity considerations by constraining the size of the pivot element. In case the number of fill-ins exceeds a certain limit the algorithm automatically compresses the storage of the basis matrix.

2.4 NUMERICAL RESULTS

2.4.1 Introduction

Programs have been written in FORTRAN 77 to implement the augmented matrix and linear programming state estimators. Sparsity and numerical stability considerations have been taken into account in both implementations. The programs make use of the sparse matrix factorisation scheme developed by Reid and available as Harwell subroutine LA05A. All computational results have been obtained on a Perkin Elmer 3220 minicomputer with 32-bit word length and floating point arithmetic using the FORTRAN VII Z compiler.

Two test systems, whose network diagrams and parameters are presented in Appendix C, have been used to evaluate the relative performance of the least-squares and least absolute values estimators. Preliminary results were first obtained for the 11-node network described in section C.1 of Appendix C. The methods were then tested on a

realistic 34-node water distribution network presented in section C.2.

In every run of the state estimator the nodal heads and the fixed-head-node flows were initialized with values corresponding to constant pressure increases above ground level and average flows respectively thus giving a worst-case guess about the state variables. Measurements were simulated by superimposing the product of a normally distributed random number and a fixed percentage of the actual readings on the values obtained from load flow studies (see Appendix A, section A.4). The measured quantities are nodal heads, fixed-head-node flows, consumer loads and network element flows. The redundancy defined by Eq. (2.2) varies according to each case.

The convergence criterion given by (2.11) is used in all studies, with $\epsilon = 1 \times 10^{-4}$ [m^3/s] for all flows and $\epsilon = 1 \times 10^{-2}$ [m] for heads.

2.4.2 Performance Tests

The initial experiments with the Jacobian matrix kept constant after two or three Newton-Raphson iterations proved that because of the strong nonlinearity of network equations this may slow-down the convergence of the estimators. Consequently, in the current version of the programs the Jacobian matrix is updated at each iteration. The computing times displayed in Tables 2.1. 2.2, 2.3 and 2.4 give the

total run-time of the estimators in a real-time environment including the update of the Jacobian matrix but excluding the I/O operations.

Case 1: 11-node network

Table 2.1 shows some typical results obtained for the 11-node network with three fixed-head nodes. A total of 18 measurements and pseudo-measurements were taken, which amounts to a redundancy of 1.29. An error with standard deviation 0.001 was used in simulating the measurements.

It can be seen that the results obtained by both estimators are very similar but the linear programme is computationally more efficient.

TABLE 2.1

Computational results for Case 1

Method	WSSR	Errors in the Estimates				Computing Time	N-R Iter.
		Average		Variance			
		Heads	Flows	Heads	Flows		
	10^{-6}	$[m]*10^{-3}$	$[m^3/s]*10^{-3}$	$[m]*10^{-3}$	$[m^3/s]*10^{-6}$	[s]	
L-S	17.1	1.46	0.0101	0.336	0.0033	1.254	4
LAV	72.0	1.12	0.0104	0.306	0.0005	0.953	4

Case 2A: 34-node network. Low level of Gaussian measurement noise

The 34-node water distribution network used here represents Doncaster Eastern and Thorne Zones of the Yorkshire Water Authority network. The total number of measurements and pseudo-measurements (54) gives a redundancy of 1.29. The standard deviation for Gaussian measurement noise was selected as 0.0002 while the pseudo-measurements were assumed to be correct.

The results for case 2A are presented in Table 2.2, part A. It can be seen that the performance of the two state estimators is again very similar but this time the least-squares estimator gave more accurate values of the state variables. Both methods converged in 4 iterations; however, the linear programme proved to be computationally more efficient than the augmented matrix method.

Case 2B: 34-node network. High level of Gaussian measurement noise

For the same meter configuration used in case 2A, the effect of a higher measurement noise level of 0.001 has been investigated. Both estimators converged in 5 iterations. The computing time for the least-squares estimator remained practically unchanged while for the linear programme it actually decreased. This is due to the fact that in the presence of greater measurement noise the simplex defined by

the measurement hyperplanes becomes less 'flat' in the neighbourhood of the optimal solution thus saving some basis interchanges. At the same time the average errors in the estimates obtained with the linear programme were increased compared to Case 2A, while for the augmented matrix method they remained unaffected. It may be argued, however, that for the purpose of on-line control both estimates are equally valid so preference could be given to a more efficient least absolute values estimator.

TABLE 2.2

Computational results for Cases 2A and 2B

Method	WSSR	Errors in the Estimates				Computing Time	N-R Iter.
		Average		Variance			
		Heads	Flows	Heads	Flows		
	10^{-3}	$[m]*10^{-3}$	$[m^3/s]*10^{-3}$	$[m]*10^{-3}$	$[m^3/s]*10^{-6}$	[s]	
A) L-S	0.64	6.81	0.039	0.076	0.0088	5.571	4
A) LAV	19.0	8.24	0.038	0.112	0.0085	4.812	4
B) L-S	17.2	8.30	0.033	0.101	0.0027	5.582	5
B) LAV	110.0	16.04	0.043	0.692	0.0040	4.392	5

Case 3: 34-node network. Measurement data with a single gross error

To further assess the performance of the state estimators the measurement data of case A was corrupted by a gross measurement error. Specifically the reading of the load

measurement at node 8 was reduced to 50% of its true value.

Table 2.3 shows that despite a smaller value of the weighted sum of squared residuals (WSSR) the least-squares estimator gives significantly worse estimates. The average residual error for heads is approximately 50 times higher, and its variance is 25 times higher than the corresponding values for the LAV estimator. The augmented matrix method also converges slower (8 iterations) and requires more computational time.

TABLE 2.3

Computational results for Case 3

Method	WSSR	Errors in the Estimates				Computing Time	N-R Iter.
		Average		Variance			
		Heads	Flows	Heads	Flows		
		$[m]*10^{-3}$	$[m^3/s]*10^{-3}$	$[m]*10^{-3}$	$[m^3/s]*10^{-6}$	[s]	
L-S	1.52	193.1	0.046	2.97	0.0093	9.744	8
LAV	4.12	3.63	0.015	0.026	0.0014	4.450	5

Case 4: 34-node network. Weak measurement configuration

In order to test the numerical stability of the state estimators a weak measurement configuration has been purposely devised. The measured quantities have been limited to flows at all fixed-head nodes, one head measurement at

node 30, one pipe flow measurement between the nodes 12 and 30, and measurements or pseudo-measurements of loads at all network nodes except node 1. This implies that the calculation of the state variables in the subnetwork consisting of nodes 1, 26, 29, 33 and 34 critically depends on calculation of the flow between the nodes 29 and 18. The redundancy for this meter configuration is 1.024. The standard deviation of the Gaussian measurement noise is selected as 0.001.

Results for this case are presented in Table 2.4. The least-squares estimator converges in 4 iterations and requires only 5.782 s to produce the estimates while the least absolute values estimator needs 6 iterations and 8.521 s of computing time. Additionally, the least-squares estimator gives considerably smaller errors in the estimates thus emphasizing a usefulness of this approach if the system is structurally ill-conditioned.

TABLE 2.4

Computational results for Case 4

Method	WSSR	Errors in the Estimates				Computing Time	N-R Iter.
		Average		Variance			
		Heads	Flows	Heads	Flows		
	10^{-3}	$[m]*10^{-3}$	$[m^3/s]*10^{-3}$	$[m]*10^{-3}$	$[m^3/s]*10^{-6}$	[s]	
L-S	2.29	4.87	0.0410	0.109	0.0089	5.782	4
LAV	49.10	107.1	0.0204	28.64	0.0008	8.521	6

2.5 CONCLUDING REMARKS

In this chapter the use of the weighted least-squares and least absolute values estimators for the purpose of real-time water network monitoring has been studied. The least-squares estimator implementation in its augmented matrix formulation is computationally efficient and exhibits very good numerical stability characteristics, especially in the case of structurally ill-conditioned systems. However, the least-squares approach is intrinsically sensitive to gross measurement errors thus requiring further bad data processing followed by reestimation of the state variables. In contrast, the least absolute values estimator is robust in the sense that it is not greatly affected by the presence of bad data which is automatically rejected, so that the state estimate is defined by the 'valid' measurements only. The algorithm based on the revised Simplex method proved also to be computationally efficient and numerically stable. The errors in the state estimates obtained with the linear programme and the augmented matrix method in normal operating conditions are very similar and are well within the limits defined by on-line control requirements.

Experience based on simulation studies lends support to the conclusion that the linear programme is somewhat better suited for the purpose of on-line network monitoring. In the case of a weak measurement configuration a hybrid approach amalgamating the features of the least-squares and

least absolute values estimators could be profitably developed.

A possible drawback of the LAV estimator is that the numerical complexity of the linear programme increases quadraticaly with the problem size. However, an application of the restarting algorithm in on-line operation of the state estimator circumvents the problem.

CHAPTER III

BAD DATA ANALYSIS

3.1 INTRODUCTION

Since the presence of bad data can in general be detrimental to the estimator performance, there is a need to develop procedures to detect whether abnormally erroneous measurements are present in the measurement set. This being the case, it is also necessary to identify the faulty observations so that they can be either eliminated from the measurement set or replaced by pseudo-measurements.

In practice, the bad data are caused by a variety of reasons, such as failures of communication links, defective meters or transducers, errors in modelling pseudo-measurements etc. If there is a high enough local measurement redundancy it is sometimes possible to reject erroneous data by prefiltering the measurements. This procedure consists of simple checks to determine if the measurements are within certain limits and plausibility tests based on comparisons of redundant measurements. However, the prefiltering tests are usually not effective if bad data is either corrupted by less than a certain

percentage of the meter reading or the neighbouring measurements are not directly comparable with the faulty one. For the purpose of this work we assume that such a pre-processing of measurements has already been performed but still some gross errors creep into the measurement set.

This chapter addresses itself to the investigation of bad data detection and identification methods to be used in connection with the least-squares and least absolute values estimators described in Chapter II. The main objective is to devise efficient computational procedures for bad data processing.

The literature on the subject of bad data detection and identification is reviewed in section 3.2. Section 3.3 presents a view on the foundations of the methods employed to solve these problems with respect to the least-squares estimator. A new method of computation of the residual covariance matrix, based on the augmented matrix approach is described in subsection 3.3.3. Section 3.4 is concerned with the bad data identification procedure to be used in connection with the least absolute values estimator. The numerical results for several tests on the detection and identification of bad data are presented and discussed in section 3.5. Finally, section 3.6 summarizes the main results of this chapter.

3.2 A REVIEW OF PREVIOUS WORK

A number of papers have been devoted to the study of the bad data processing problem especially in connection with the on-line power system state estimation. In general terms, the problem was first described by Schweppe et al. in [220], [221], [222]. For bad data detection, monitoring of the weighted sum of squared residuals was suggested, while for the identification of bad data points the largest normalised residuals were sought. These methods became subsequently known as $C(\hat{\underline{x}})$-test and $\underline{r}_N$-test respectively.

Another approach to the bad data detection/ identification problem consists in penalizing the largest residuals so that the potential bad data have a reduced influence in the final estimates [178], [131]. This method implies the use of a non-quadratic cost function. The most frequently used combinations of the cost functions are: quadratic-straight, quadratic-square root and quadratic-constant. When successful, the method seems to allow a direct identification of the bad data point through the examination of the residuals. However, the choice of the most suitable cost criterion, the possibility of local minima and the increase of computational complexity of the estimator are the problems that one has to take into account when using such a method.

Irving, Owen and Sterling [145] have proposed a method

which has the advantages of the nonquadratic estimators and which avoids their difficulties. By using the weighted sum of moduli of the measurement residuals the method gives an automatic rejection of bad data so that the state estimate is not affected by erroneous measurements.

A deterministic technique for bad data suppression has been presented by Debs, Larson and Hajdu [66]. The comparison of measurements at successive time steps is utilized to detect and identify bad data. It seems however that such a method would be unable to detect a single bad measurement caused by a meter with a slowly increasing error, as well as multiple interacting bad data.

Dopazo et al. [85] proposed a technique based on hypothesis testing theory. The $C(\hat{\underline{x}})$-test is used for detecting bad data, and identification is performed through a Student's t - test for the estimated value of the measured quantity. The use of confidence limits is also suggested to assess the estimates.

From the computational point of view, Broussolle [38] suggests the use of the sparse inverse method to compute the covariance matrix of residuals, which is needed in the $\underline{r}_N$-identification test.

More recently, Clements et al. [156] have attacked the problem by using a combinatorial approach to investigate the bad data detectability and how the bad data spread on the

residuals.

3.3 BAD DATA PROCESSING IN LEAST-SQUARES ESTIMATION

3.3.1 Foundations of Bad Data Detection and Identification Methods

The bad data detection and identification methods developed in connection with the least-squares estimators are based on examination of both the measurement residuals and a function of them. The reason for doing so is that the residuals convey combined information about possible violations of the assumptions about meter accuracy and distribution of measurement noise which have been made during the construction of the measurement model.

For a measurement model which considers m measurements, the residual vector is an mx1 vector defined by

$$\underline{r} = \underline{z} - \underline{\hat{z}} \qquad (3.1)$$

where $\underline{z}$ is the mx1 vector of measurements and $\underline{\hat{z}}$ is the vector of estimates for the measured quantities.

Equation (3.1) clearly shows that a residual is a mismatch between the actual measurements and the value of the measured quantity as computed by the least-squares algorithm. Therefore the mismatch can be seen as the amount that the measurement model cannot account for. If it is assumed that the model represents the system with the

expected accuracy, which implies that the values of the parameters in the model (hydraulic resistances of pipes, pump characteristics, etc.) and the model structure are accurately known, then one can think of the residuals as estimates of the measurement errors. Since certain assumptions about the measurement errors are made when the model is established, it should be expected that, if no bad data are present, the residuals will tend to behave in a manner that confirms those assumptions. If, on the other hand, some residuals or functions of the residuals clearly violate the assumptions, one can infer that erroneous measurements are present. Furthermore, by investigating the residuals individually, it should be possible to locate the faulty measurements i.e. to identify the bad data.

Consider the measurement model for water system state estimation after linearization with respect to a point $\hat{\underline{x}}_k$

$$\Delta\underline{z} = J(\hat{\underline{x}}_k)\cdot\Delta\hat{\underline{x}}+\underline{\omega} \qquad (3.2a)$$

$$E\{\underline{\omega}\} = \underline{0}\ ; \quad E\{\underline{\omega}\,\underline{\omega}^T\} = R \qquad (3.2b)$$

where

$$\Delta\underline{z} = \underline{z}-g(\hat{\underline{x}}_k) \qquad (3.3a)$$

$$\Delta\hat{\underline{x}} = \underline{x}-\hat{\underline{x}}_k \qquad (3.3b)$$

$$J(\hat{\underline{x}}_k) = \left.\frac{\delta \underline{g}(\underline{x})}{\delta \underline{x}}\right|_{\underline{x}=\hat{\underline{x}}_k} \quad : \text{Jacobian matrix} \qquad (3.3c)$$

ω : random vector which models measurement errors.

In addition, assume that $\hat{\underline{x}}_k$ is close enough to the solution point such that Eq. (3.2a) is a good approximation to the nonlinear model. For example, $\hat{\underline{x}}_k$ can be taken as the last linearisation point before convergence. Under this condition, one can apply the methods for residual analysis, usually employed for the linear least-squares case. The Jacobian matrix computed for such a point will hereafter be denoted simply by J.

The weighted least-squares solution is obtained by minimizing the cost function

$$C(\hat{\underline{x}}) = [\Delta\underline{z} - J\,\Delta\hat{\underline{x}}]^T R^{-1} [\Delta\underline{z} - J\,\Delta\hat{\underline{x}}] \qquad (3.4)$$

$C(\hat{\underline{x}})$ is merely the weighted sum of squares of the residuals for the linearised model, where the weighting matrix is the inverse of the covariance matrix of the measurement errors.

3.3.2 Bad Data Detection

The vector of measurement errors $\underline{\omega}$ has been partially characterised by Eq. (3.2b). To apply the bad data detection procedure which will be described next, an

additional assumption concerning the distribution of $\underline{\omega}$ is required, namely, that $\underline{\omega}$ is normally distributed. The notation

$$\underline{\omega} \sim N(O,R) \qquad (3.5)$$

indicates that the measurement error is considered to be normally distributed, with zero mean and covariance matrix R

Having the complete statistical characterization of the measurement errors in the absence of bad data, it is possible now to undertake the search for a procedure for bad data detection using the residuals. A natural candidate for such a procedure would be an individual test on the residuals to find out whether any of them violates the assumptions made for the measurement errors. However, this technique would require the use of the covariance matrix of the residuals whose computation is costly. Considering the fact that a bad data detection routine is to be employed on-line after each state estimation, it may be concluded that the individual examination of residuals would not be an efficient technique to simply detect the presence of bad data. Nevertheless, the technique is needed to identify the faulty measurement after its presence is detected as discussed in the next subsection.

Because of the computational difficulty of individually examining the residuals it is a common practice to monitor a weighted sum of squared residuals which is readily

available as the objective function in the least-squares problem and which exhibits clearly distinct behaviour under bad data-free situations and under the presence of bad data.

To decide whether bad data are present or not, a test must be performed on $C(\hat{\underline{x}})$. The residuals are random variables normally distributed which implies that $C(\hat{\underline{x}})$ is a random variable which has a chi-square distribution with $m-n$ degrees of freedom, where m is the number of measurements and n is the dimension of the state vector. If however a gross measurement error is present the normality assumption of $\underline{\omega}$ is violated and consequently $C(\hat{\underline{x}})$ is no longer chi-square distributed.

In view of the aforementioned facts, the bad data detection procedure can be seen as a testing of the hypothesis about the distribution of $C(\hat{\underline{x}})$. This can be formulated as follows:

H_o : $C(\hat{\underline{x}})$ is chi-square distributed

H_1 : H_o is false

The significance level of the test, also called false alarm probability, is the probability of rejecting H_o when it is actually true, and is normally denoted by α. Using this definition, it is possible to determine a detection threshold level K corresponding to a false alarm probability α such that

$$P(C(\hat{\underline{x}}) > K \mid C(\hat{\underline{x}}) \text{ is chi-square}) = \alpha \qquad (3.6)$$

where $P(\ |\ )$ denotes conditional probability.

From Eq. (3.6), the detection threshold level K can be computed as

$$K = \chi^2(m-n);\beta \qquad (3.7)$$

where $\chi^2(m-n);\beta$ denotes the $\beta=(1-\alpha)$ quantile of the chi-square distribution with $m-n$ degrees of freedom. When the number of degrees of freedom is large (in practice, greater than 30), the chi-square distribution approaches a Gaussian distribution with mean $m-n$ and variance $2(m-n)$ [131]. In this case K can be computed as a quantile of appropriate normal distribution.

The $C(\hat{\underline{x}})$-test can now be performed by comparing the value of $C(\hat{\underline{x}})$ with a threshold K. If for a given α the value $C(\hat{\underline{x}})$ is greater than K, then this fact is taken as an evidence that the null hypothesis is false, i.e. that bad data are present.

The implementation of the $C(\hat{\underline{x}})$ detection test can be outlined as follows: After each state estimation run, the weighted sum of squared residuals $C(\hat{\underline{x}})$ is computed. Then, $C(\hat{\underline{x}})$ is compared with the threshold K corresponding to a given level of false alarm probability α. If $C(\hat{\underline{x}}) > K$, one concludes that bad data are present, and an identification procedure can be invoked to find out which measurements are wildly erroneous. Otherwise, the state estimates are

accepted on the grounds that there is not enough evidence to decide on the presence of bad data.

3.3.3 Bad Data Identification

After the presence of an abnormally inaccurate measurement is detected, the next step is to locate the faulty meter. This requires the individual examination of the measurement residuals. In the case of a single bad measurement, a possible identification strategy would be to find the maximum residual and then expect that the corresponding measurement is the faulty one. However, this is not necessarily true, for two reasons: i) Meters for different quantities have different accuracies, so that the variances of the corresponding measurements can be significantly different; and ii) The residuals are in general correlated among themselves, so that an error associated with a measurement can spread over other residuals.

In order to allow a fair comparison of the residuals it is mandatory to normalize them with respect to their standard deviation. After this normalization is performed, the measurement corresponding to the maximum normalized residual is most likely to be bad data. [131]

Defining the residual vector $\underline{r}$ as

$$\underline{r} = \Delta\underline{z} - \Delta\hat{\underline{z}} \qquad (3.8)$$

where

$$\Delta\hat{\underline{z}} = J \cdot \Delta\hat{\underline{x}}$$

and calculating $\Delta\hat{\underline{x}}$ using the normal equation approach presented in section 2.2.3

$$\Delta\hat{\underline{x}} = (J^T R^{-1} J)^{-1} J^T R^{-1} \Delta\underline{z}$$

equation (3.8) can be written as

$$\underline{r} = W \Delta\underline{z} \qquad (3.9)$$

where W is the residual sensitivity matrix

$$W = [I - J(J^T R^{-1} J)^{-1} J^T R^{-1}] \qquad (3.9b)$$

Using the above expression the residual covariance matrix $D = E\{\underline{r}\,\underline{r}^T\}$ can be calculated as

$$D = E\{\Delta\underline{z}\,\Delta\underline{z}^T\} - 2J(J^T R^{-1} J)^{-1} J^T R^{-1} E\{\Delta\underline{z}\} + J(J^T R^{-1} J)^{-1} J^T R^{-1} E\{\Delta\underline{z}\,\Delta\underline{z}^T\} R^{-1} J (J^T R^{-1} J)^{-1} J^T \qquad (3.10)$$

which taking into account Eq. (3.2b) gives

$$D = R - J(J^T R^{-1} J)^{-1} J^T \qquad (3.11)$$

The normalized residuals can now be obtained as

$$\underline{r}_N = \sqrt{(\text{diag } D)^{-1}}\,\underline{r} \qquad (3.12)$$

The drawback of the identification method based on the search for the maximum normalised residual is the amount of computation required to obtain the variances of the residuals. Even though only the diagonal elements of D are needed it is still necessary to perform the inversion of the normal matrix $J^T R^{-1} J$. This problem has received some attention in the literature and methods have been proposed to reduce the number of required operations. One of such methods makes use of the fact that actually only the elements of the inverse of the normal matrix that correspond to nonzeros of $J^T J$ need to be computed [38]. The method approximately halves the computational burden but in the same time it requires much more complicated programming. Additionally, the method of Broussolle is not free from the problem posed by structural ill-conditioning of the right hand side of Eq.(3.11). This fact may lead to inaccuracies which in turn may invalidate the whole computations of the matrix D.

In this work an alternative method for calculating the residual covariance matrix which avoids numerical difficulties associated with the standard approach is proposed. The method is based on the observation that by using the augmented matrix state estimator, which apart from a state estimate gives an estimate of the residual error, the residual sensitivity matrix W can be easily found as $\delta \underline{r} / \delta \Delta \underline{z}$. Consequently, the elements of the matrix D can be found as

$$D = W\ R \qquad (3.13)$$

The consecutive columns of W are obtained by solving the following equation

$$\begin{bmatrix} 0 & I & J \\ -I & R^{-1} & 0 \\ J^T & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \underline{u} \\ \underline{w}_i \\ \underline{v} \end{bmatrix} = \begin{bmatrix} \underline{e}_i \\ \underline{0} \\ \underline{0} \end{bmatrix} \qquad i=1,\ldots,m \qquad (3.14)$$

where $\underline{w}_i$ is mx1 vector of measurement residuals corresponding to the unit change of the i-th measurement (i-th column of the matrix W)

$\underline{e}_i$ is mx1 unit vector

$\underline{u}$, $\underline{v}$ are mx1 and nx1 auxiliary vectors

Since the factors of the augmented matrix are readily available from the state estimator the computation of the residual sensitivity matrix involves only m back-substitutions, thus giving a computationally efficient algorithm. Another property favouring the approach proposed here is its numerical stability since the condition number of the augmented matrix remains unchanged compared to the condition number of the linearised measurement matrix (see section 2.2.4)

In the case of multiple non-interacting bad data, i.e. when the covariances between any pairs of residuals which correspond to the bad data are small as compared to the respective variances, an effective identification can be obtained by using a trivial extension of the procedures

discussed above. It consists simply of looking for the maximum normalised residual and eliminating the measurement corresponding to it. If, after the following estimation, bad data is still detected the next measurement with the largest residual is eliminated, and so on until $C(\hat{\underline{x}})$-test does not indicate the presence of bad data.

Although the same approach can be used for multiple interacting bad data, the results are usually not as good. In fact, examples discussed in section 3.5 seem to confirm the findings of Handschin, Schweppe et al. [131] that, in such situations, the performance of non-quadratic estimator is better than the weighted least-squares estimator.

3.4 BAD DATA PROCESSING IN LEAST ABSOLUTE VALUES ESTIMATION

3.4.1 Identification of Gross Measurement Errors

The characteristic feature of the least absolute values (LAV) estimator is that it attempts to calculate state variables using a minimal observable set of measurements. In the system with n state variables and m measurements/pseudo-measurements the LAV estimator always accepts n measurements and rejects the remaining m-n readings. Since the criterion according to which the measurements are accepted or rejected is the magnitude of the measurement noise, it is reasonable to expect that if

any gross error is present in the initial data it will become the first candidate for rejection. Consequently, the procedure to identify bad data in connection with the least absolute values estimator is essentially concerned with examining the magnitude of the measurement residuals.

The reason for the effectiveness of such an approach is that the least absolute values estimator prevents 'smearing' of the residual errors, provided that there exists an error-free observable set of measurements. In this case, each measurement residual remains unaffected by the magnitude and location of the other bad data. Thus, it is possible to avoid time-consuming computations of the residual covariance matrix, which is the main computational burden in the least-squares-based approach, and to calculate only the weighted residuals.

The implementation of the bad data identification method consists essentially in weighting the residuals with respect to their standard deviation and comparing them to the prespecified threshold. As a result, the method has very small computational requirements. The next section extends this approach for the case of topological errors which can be seen as multiple interacting bad data.

3.4.2 Identification of Topological Errors

It is usual to make a distinction between erroneous measurements and topological errors since they represent

different physical phenomena. However, from the point of view of the state estimator both errors substantiate bad data points.

In the context of the water distribution systems two main types of topological errors can be specified:

i) incorrect status of control valves which gives a wrong image of actual network connectivity, and

ii) leakages which effectively extend the network by additional load nodes.

The presence of either of these errors is equivalent to neglecting a part of the actual network structure thus producing an imbalance at the network nodes incident to the questionable pipe. One can now think of the topological error as a pair of erroneous load measurements for which the error terms, devised by the state estimator, are carrying information about a topology misspecification. As a further consequence the conditions of detectability of the topological errors can be determined by superimposing the detectability conditions for two gross measurement errors, i.e. the network must be observable after the erroneous mass-balance equations are rejected.

Figures 3.1 a - c give a graphical representation of the topological errors.

Type 1 (Fig.3.1 a). If in the real network the pipe between the nodes i and j is closed, the actual flow q_{ij} is zero regardless of the pressure difference between i and j. On the other hand, if the network model incorrectly assumes that the pipe is open the calculated flow q_{ij} will in general, have a nonzero value q. In order to compensate this mismatch, the least absolute values estimator calculates the system state not using the model of the pipe i-j and rejects the mass-balance measurements in the end-nodes of this pipe. The error terms associated with the rejected measurements represent a flow which has an equal magnitude and opposite direction to q_{ij} so that the average flow between the nodes i and j is equal zero as it is in the real network.

Type 2 (Fig. 3.1 b). Similarily as in the case 1 the valve status in the pipe i-j is incorrect but this time the pipe is open in the real network while it is assumed to be closed in the network model. The error terms devised by the state estimator r_i and r_j represent the actual flow between the nodes i and j ($r_i=-r_j=q$) which cannot be allowed through the pipe i-j since it is modelled as closed.

Type 3 (Fig. 3.1 c). If there exists a leakage between the nodes i and j it forms an additional load node in the actual network. The magnitude of the leakage q_l is determined by the pressure at the nodes i, l and j. At the same time, the network model has no representation of

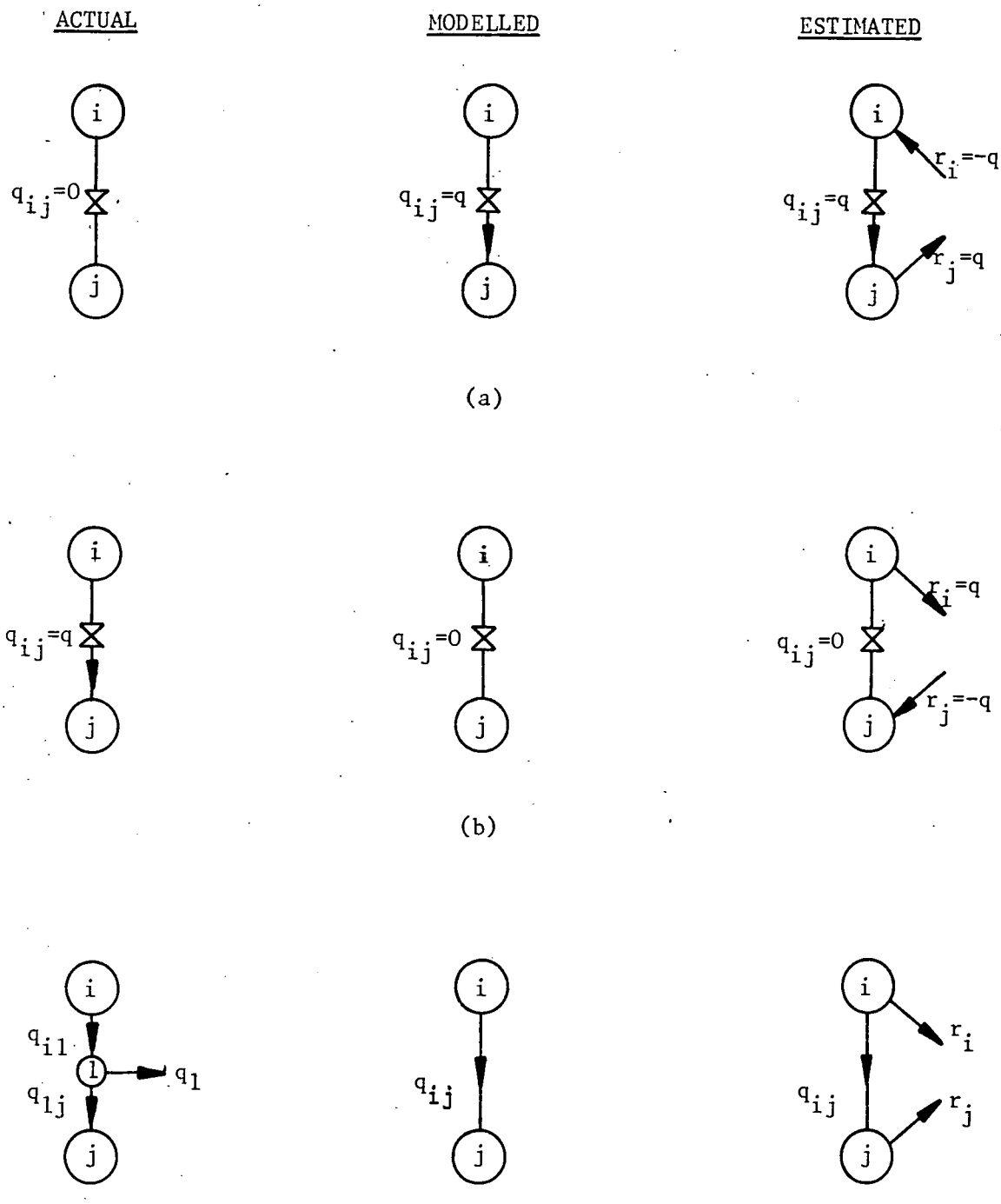


Figure 3.1 Identification of topological errors

(a) closed valve monitored open;
(b) opened valve monitored closed;
(c) leakage.

the leak-node and the flow q_{ij} is calculated as a function of the pressure difference between the nodes i and j. The magnitude of the flow q_{ij} is therefore less than q_{il} and greater than q_{lj}. In order to balance the whole network the state estimator rejects the mass-balance equations in the end-nodes of the leaking pipe giving the error terms r_i and r_j. The sum of these residuals represents the value of the leakage q_l.

Given the estimation results and a list of the measurements rejected by the least absolute values estimator, the identification of topological errors can be implemented as a following post-processing procedure.

1. Perform observability tests to find out in which part of the network topological errors are detectable using the current measurement configuration (see Chapter IV).

2. Calculate the weighted residuals of the rejected measurements/pseudo-measurements and identify the bad data

3. Test the residuals of the mass-balance equations:
   - if the residuals in the end-nodes of a pipe with a valve have approximately equal magnitude and opposite signs then the valve status is incorrectly monitored;
   - if the residuals in the end-nodes of a pipe are

negative then there exists a leak in the pipe;

- otherwise the residual represents a gross measurement error.

3.5 NUMERICAL RESULTS

This section presents the numerical results for the bad data processing methods presented in sections 3.3 and 3.4. The methods are labelled as L-S (least-squares based) and LAV (least absolute values based) respectively. The comparison is concentrated on the ability of the methods to correctly identify bad data and on their computational efficiency.

The computer programs have been written in FORTRAN 77 and run under FORTRAN VII Z compiler on a Perkin-Elmer 3220 minicomputer with 32-bit word length and floating point arithmetic.

All the test cases refer to the 34-node water distribution system described in Appendix C. The measurement configuration is identical to the one used to assess the performance of the estimators (Case 2A, 2B and 3). A total of 54 measurements and pseudo-measurements is used, yielding a global redundancy of 1.29. The methods were tested on many sets of measurement data containing topological and gross measurement errors and a representative sample is given below.

Case 1: Gross Measurement Error of Consumer Load at Node 8

Table 3.1 shows some typical results obtained for the measurement data contaminated with a gross measurement error. The magnitude of the error is equal to 35 standard deviations of the measurement concerned and the false alarm probability is selected as 0.05.

As far as the estimation time is concerned, the performance of both methods is very similar; however, the least-squares based method requires additional time for computation of the residual covariance matrix followed by reestimation of the state vector. In contrast, the identification of bad data by the least absolute values estimator is done in the course of the estimation and only an inexpensive search for maximum weighted residual is required. Both methods correctly identify the bad data point.

TABLE 3.1

Computational results for Case 1

Method	Estimation		Identification and Reestimation*		Bad Data Identified	Total Time [s]
	N-R Iter.	Time [s]	Cycles	Time [s]		
L-S	4	4.587	1	3.355	YES	7.942
LAV	5	4.387	-	0.027	YES	4.414

* only for L-S method

Case 2: A) Leakage in the Pipe 4-20

B) Leakage in the Pipe 12-20

The effect of the presence of a topological error in the form of a leakage has been studied on two examples. First, a leakage in the pipe 4-20, for which a flow in normal operating conditions is low ($0.0013 m^3/s$), is simulated. Next, an identical in magnitude ($0.01 m^3/s$) leakage is simulated in the pipe 20-12 which has a flow of $0.0271\ m^3/s$.

As in Case 1, computational efficiency of the least-squares approach is grossly affected by the necessity of calculating the normalised residuals and subsequent reestimation of the state vector. Moreover, the $C(\hat{\underline{x}})/\underline{r}_N$-test proved to be less sensitive in the case where the relative magnitude of the leakage, as compared to the flow in the

TABLE 3.2

Computational results for Case 2

Method	Estimation		Identification and Reestimation*		Bad Data Identified	Total Time [s]
	N-R Iter.	Time [s]	Cycles	Time [s]		
A) L-S	1	1.137	2	5.595	YES	6.732
A) LAV	1	2.950	-	0.027	YES	2.997
B) L-S	1	1.248	-	-	NO	-
B) LAV	1	2.990	-	0.027	YES	3.017

* only for L-S method

leaking pipe, is smaller.

Case 3: Incorrect Valve Status

For the same meter configuration as in Case 1 and 2 another type of topological error has been introduced. Control valve in the pipe 17-9 is monitored as closed while in the real network it remains open. The magnitude of the flow through the control valve is relatively small (0.0063 m^3/s), taking into account the standard deviation of the flow measurements (0.0006 m^3/s); however, due to the strong local measurement configuration both methods correctly identify bad data. The estimators converge in 4 iterations requiring 4.525 s and 4.412 s of computation time for L-S and LAV method respectively. In this case the least-squares based method requires two identification/reestimation cycles since the topological error implies two gross measurement errors and the $\underline{r}_N$-test guarantees the correct identification of a single bad data at a time.

TABLE 3.3

Computational results for Case 3

Method	Estimation		Identification and Reestimation*		Bad Data Identified	Total Time [s]
	N-R Iter.	Time [s]	Cycles	Time [s]		
L-S	4	4.525	2	5.595	YES	10.120
LAV	4	4.412	-	0.027	YES	4.439

* only for L-S method

Case 4: Leakage in the Pipe 4-20 and Gross Measurement Error of Consumer Load at Node 6

In order to check the limitations of the bad data identification procedures a measurement set with multiple interacting bad data has been devised. According to the observability criterion it is not possible to identify bad data points since the network becomes unobservable after removing the erroneous load measurements at nodes 4, 20 and 31. In fact, both methods fail to give a correct answer. The state estimates converge in 6 and 4 iterations (6.625 s and 4.841 s) for L-S and LAV method respectively but in both cases they are affected by the presence of gross errors.

3.6 CONCLUDING REMARKS

This chapter has been devoted to the investigation of the bad data detection and identification methods in water system state estimation. The main objective has been to develop techniques to be used in connection with the real-time estimators described in Chapter II.

For the least-squares estimator, using the augmented matrix formulation, a new technique of calculating the residual sensitivity matrix has been proposed. Apart from the computational efficiency this technique is also numerically more stable, compared to the corresponding methods reported in the literature, since it avoids the formation of the normal matrix with consequent squaring of

the condition number. The bad data identification technique based on the combined $C(\hat{\underline{x}})/\underline{r}_N$-test is shown to perform well in the case of a single bad data. However, as the number of gross measurement errors increases the method becomes computationally less efficient due to the repeated calculations of the residual sensitivity matrix followed by the reestimation of the state vector.

By exploiting the fact that the least absolute values estimator avoids 'smearing' of the measurement residuals, it has been possible to develop an efficient procedure for identification of the topological errors. The procedure merely checks the magnitude and sign of the weighted measurement residuals since the topological errors are shown to be equivalent to a pair of gross mass-balance errors in the end-nodes of the pipe concerned. The computational overhead associated with this post-processing is negligible so the time required to identify bad data is determined by the efficiency of the state estimator.

A number of tests have been carried out on the realistic 34-node system and the corresponding results are reported in section 3.5. The results confirm the requirement that in order to identify bad data it is necessary to have available a local measurement redundancy. Consequently, by performing the observability test it is possible to determine in which areas of the network bad data is detectable using the current measurement set.

CHAPTER IV

METHODS TO DETERMINE WATER SYSTEM OBSERVABILITY

4.1 INTRODUCTION

The observability problem in water system state estimation consists essentially in determining whether the measurements currently available to the state estimator provide sufficient information to allow the computation of the estimates.

Observability tests are important both as a design tool in meter placement studies performed off-line, and in the on-line implementation of the estimator.

In on-line operation, the availability of a routine to check whether the water system is observable or not is very important for the efficiency of the estimation process. Before the state estimation, the observability routine determines whether the current measurement set renders the system observable. If this is the case, the state estimation proceeds. Otherwise, the system is unobservable, and the estimator will not be able to calculate the states for the whole network using the available measurements.

This situation may arise as a result of meter or telemetry failure, changes of network topology by means of valve controls, and also as a consequence of the elimination of measurements previously identified as bad data. In these cases, the observability routine should identify the observable subsystems so that, in a subsequent step, either the state estimation is applied to the subnetworks of the original system, or appropriate pseudo-measurements are added to the measurement set to allow the estimation of the states for the whole system.

Observability considerations also have relevance in the planning stage of a metering system to be used for state estimation. In these off-line studies, the objective is to achieve a metering system design which will guarantee reliable estimates even in the event of meter and telemetry failures. To take into account the possibility of bad data elimination, measurements can be omitted singly, in pairs etc. The observability test is then used to assess the resulting metering systems and to indicate where they should be reinforced by the addition of further measurements.

There are some other questions that are related to the observability problem. For example, the determination of a minimal measurement set which makes the system observable may be used as a first step to determine how redundancy should be added to enhance the estimates accuracy and the performance of bad data detection and identification. Other

related problems are the determination of detectability conditions for leakages in the network, limitation of the spread of the residual error, or considerations of financial aspects of the telemetry system and meter placement designs.

This chapter initially presents the basic theoretical results for topological observability in water system state estimation. Two original methods are then proposed which essentially aim to find an observable spanning tree of the water network. The first technique transforms the observability problem into a matching problem in bipartite measurement-to-branch graph. Branches with measurement assignment are then used to build a spanning tree of the network. The second method undertakes a direct search of the observable spanning tree. The method starts from an arbitrary node. The equivalence of the preservation of the path property and the existence of the observable spanning tree has been exploited in order to devise a procedure to correct possible misassignments during the first stage of the algorithm.

This chapter is organised as follows. Section 3.2 reviews the literature on power system observability for state estimation since practical methods for solving the observability problem originated from this area of application. The equivalent observability considerations with respect to the water network, to the author's knowledge, have not been reported in the literature. Section

4.3 is devoted to the introduction of the observability definitions in the context of water distribution networks. The basic conditions for topological observability are presented in section 4.4. Sections 4.5 and 4.6 contain descriptions of two new methods to determine topological observability.

4.2 A REVIEW OF PREVIOUS WORK

The importance and complexity of the observability problem has been recognized since the very early stages of research on power system state estimation. Schweppe and Wildes [220], in the first of the three papers which originally proposed the use of state estimation techniques for power systems, acknowledged the difficulties associated with the meter placement problem and how it could affect the performance of the estimator. They employed the covariance matrix of the estimation errors as a tool for selecting the type and location of meters, on a trial-and-error basis.

Observability questions, however, started to receive more extensive attention in the literature only after 1973. Some proposed methods are still based on trial-and-error procedures, using different criteria. Thus, Edelman [97] assesses the metering system from the point of view of the condition number of the information matrix, and Ariatti et al. [8] use reliability and quality of the estimates as the criteria to compare distinct metering schemes. A different

approach is used by Koglin [154], which starts with the set of all possible measurements and sequentially eliminates the measurements which do not significantly affect the quality of the estimates. This is decided by comparing the expectation of the quadratic errors against given limits. Fetzer and Anderson [110] formulate the problem by using the concept of observability from linear control theory. The proposed method for measurement selection starts with a given measurement set which is then sequentially augmented by one measurement at a time. At each step, the new measurement is the one whose component orthogonal to the measurement hyperspace is the largest. This can actually be seen as an application of principal component analysis. The computational requirements of the method seem to be an obstacle to its practical application.

The measurement selection problem is formulated using information theory and non-linear programming by Phua and Dillon [197]. The aim is to maximize the information about the state vector in the measurement set. Considerations about measurement accuracy and financial costs are modelled as constraints in the optimisation problem.

All the above methods make use of floating point calculations and are actually intended for off-line meter placement studies. Other authors have sought methods to be used in both on-line and off-line studies. These methods are usually based on logical procedures. Handshin and Bongers

[130], for example, proposed an observability test which consists of checking the connectivity of the Jacobian matrix. However, this test only gives a weak necessary condition for observability.

Clements and Wollenberg [58] investigate minimum observability conditions using network topology and the Kirchhoff Laws. They also introduce the concept of observable islands. The proposed algorithm is a heuristic procedure which first considers line flow measurements and then processes injection measurements, one at a time. Although it has been shown that the algorithm may give conservative results [4], [156], the Clements and Wollenberg paper brought about some ideas which were pursued in subsequent works. Allemong et al. [4], who detected that Clements and Wollenberg method could provide conservative results, suggested a new algorithm based on the same ideas to correct the problem. Basically, the algorithm searches for an observable tree in the network by using the principles of generation of trees. Krumpholz, Clements and Davis [156] use network topology and an algorithm for the flow problem in transportation networks to devise a method for solving the observability problem. The concepts of algebraic, numerical and topological observabilities are introduced, and the theoretical framework for the method is developed considering a linear approximation for the measurement model. More recently the same authors [57] have published an enhanced version of their algorithm which avoids possible

misassignements of measurements to branches in its attempt to build the largest forest of full rank.

Quintana, Simoes-Costa and Mandel [199] proposed a method which directly searches for an observable spanning tree in the measurement graph using an algorithm for matroid intersection.

For the networks with a big proportion of flow measurements Van Cutsem and Gailly [273], [274] proposed an enumerative procedure which examines all possible measurement assignments. The algorithm is simple but has limited applicability since in general case it can be very expensive computationally.

4.3 OBSERVABILITY DEFINITIONS

4.3.1 Observability and Numerical Observability

Consider a set of M measurements taken in water distribution system. It is assumed that the measurable quantities are nodal heads, fixed-head-node flows, consumer loads and pipe flows.

A water system is said to be observable or solvable in the static state estimation sense with respect to a given measurement set M, if the fixed-head-node flows and the nodal heads throughout the system can be determined by processing the measurements in M by a static state

estimator. Otherwise, the water system is said to be unobservable with respect to M.

From this definition it can be immediately concluded that a necessary condition for water system observability is that the Jacobian matrix in Eq. (2.10) must be of full rank.

It should also be noticed that observability depends, to a certain extent, on the operating point used for the linearisation of the measured model. This is so because, for a given measurement set, the numerical values of the entries of the Jacobian matrix vary according to the operating point. Theoretically, this might affect the rank of the Jacobian matrix. Also, it may happen that the Jacobian matrix is of full rank when, computed with respect to a certain operating point but, in the course of the iterative solution, numerical problems may develop such that the final estimates cannot be obtained. To take into account these factors, the definition of numerical observability, analogous to the one by [156], can be introduced.

A water system is said to be numerically observable in the static state estimation sense, with respect to a given measurement set M if the estimates can be obtained using the flat start (i.e. nodal heads equal to a fixed pressure increase above the corresponding ground level and fixed-head-node flows equal to average flows) as the initial guess for the estimation algorithm.

The use of the flat start in the definition of numerical observability may seem arbitrary. However, it reflects the fact that the flat start is usually the most severe initial guess to start the estimation algorithm, as it is used only when no better point to initialize the iterations is known. It should also be remarked that, apart from possible rank deficiency and numerical problems, a water system is also considered as numerically unobservable if the flat start is too far away from the actual state so that convergence cannot be achieved [156].

Testing numerical observability amounts to solving the static state estimation problem from the flat start for the given measurement set. This procedure cannot be considered as a feasible candidate for a practical observability test for at least two reasons: first, observability would be decided by solving the problem instead of being an 'a priori' result, and second, the method gives no clue as to where in the system the problem resides.

Another possible way to test observability would be the floating point calculation of the rank of the Jacobian matrix. But, in spite of the fact that efficient algorithms for computing the rank of a matrix are currently available, these methods are still too time-consuming for on-line applications. Besides, such methods would also be unable to provide indications about the location of the problem and about the observable susbsystems.

4.3.2 Measurement Model for Water System State Estimation

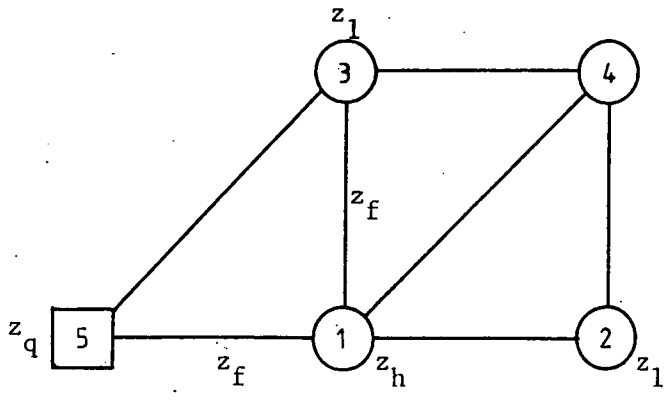
The derivation of topological observability conditions described in the next section is based on the approximate measurement model for water system state estimation. The non-linear measurement model has been discussed in Appendix A is given by

$$\underline{z} = \underline{g}(\underline{x}) + \underline{\omega} \qquad (4.1)$$

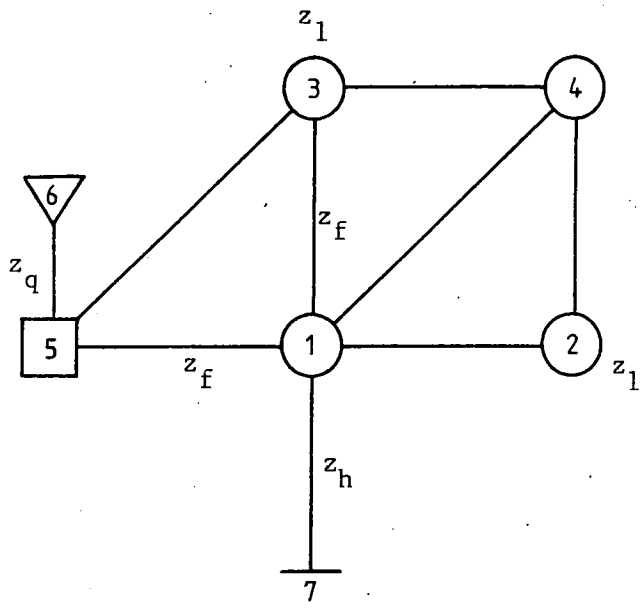
where $\underline{z}$ is the mx1 measurement vector, $\underline{x}$ is the nx1 state vector, $\underline{g}(\cdot)$ is an mx1 non-linear vector function, and $\underline{\omega}$ is an mx1 random vector which models the measurement errors.

Consider that the water system comprises N nodes, F fixed-head-nodes and P pipes. A total of m measurements are taken namely: m_h head magnitudes, m_q fixed-head-node flows, m_l consumer loads and m_f pipe flows. Since the heads are usually measured with respect to the equalised ground level the dimension of the state vector is n=N+F. The state vector is of the form $\underline{x}^T = [\underline{h},\underline{q}]$, where $\underline{h}$ is the vector of nodal heads and $\underline{q}$ is the vector of fixed-head-node flows. In order to achieve a one-to-one correspondence between state variables and network nodes a concept of auxiliary nodes is introduced. The following properties are inherent to the auxiliary nodes:

i) each fixed-head-node in the network has an auxiliary node corresponding to it;



(a)



(b)

variable-head node

fixed-head node

auxilary node

reference node

z_h - head measurement

z_q - fixed-head-node flow measurement

z_l - load measurement

z_f - pipe flow measurement

Figure 4.1 (a) Original network graph

(b) Augmented network graph

ii) the auxiliary node can only be connected to its fixed-head-node;

iii) the nodal pressure is not defined in the auxiliary node,

iv) the flow between the auxiliary node and the fixed-head-node is defined by the fixed-head-node flow.

It is also convenient, for the purpose of analysis of the measurement configuration, to include in the network diagram a head reference node with links to the head-measured nodes. The network digram now consists of n+1 nodes and $l=P+F+m_h$ links (Fig.4.1 a and 4.1 b). Each link contributes one element to the diagonal 1x1 hydraulic conductivity matrix Y. The P elements of the matrix Y corresponding to the links between the n nodes of the original network represent the sensitivity of the flow to the changes of the nodal heads. The value of these elements vary according to the operating conditions of the network, and can be calculated from the network element equations given in Appendix A. The remaining $F+m_h$ elements of Y represent the sensitivity of fixed-head flows and the sensitivity of nodal heads to their own changes and are therefore equal to 1.

A direction is assigned to each link of the augmented network diagram, arbitrarily, so that the network can be represented as a directed connected graph. The structure of this graph is described by its (n+1)x1 incidence matrix A,

as defined in Appendix E, with the exception that property (iii) and (iv) of the auxiliary nodes implies that the link between a fixed-head-node and the auxiliary node does not appear in the incidence list of the fixed-head-node. Furthermore, let A_r denote the nxl reduced incidence matrix obtained from A by deleting the row corresponding to the reference node.

Having the required definitions the measurement model can be expressed as

$$\begin{bmatrix} \underline{z}_h \\ \underline{z}_q \\ \underline{z}_l \\ \underline{z}_f \end{bmatrix} = \begin{bmatrix} M_h \, Y \, A_r \\ M_q \, Y \, A_r \\ M_L \, Y \, A_r \\ M_f \, Y \, A_r \end{bmatrix} \cdot \begin{bmatrix} \underline{h} \\ \underline{q} \end{bmatrix} + \begin{bmatrix} \underline{\omega}_h \\ \underline{\omega}_q \\ \underline{\omega}_l \\ \underline{\omega}_f \end{bmatrix} \qquad (4.2)$$

where

$\underline{z}_h$, $\underline{z}_q$, $\underline{z}_l$, $\underline{z}_f$: are m_hx1, m_qx1, m_lx1 and m_fx1 measurement vectors of head, fixed-head-node flow, consumer load and pipe flow respectively;

$\underline{\omega}_h$, $\underline{\omega}_q$, $\underline{\omega}_l$, $\underline{\omega}_f$: are m_hx1, m_qx1, m_lx1 and m_fx1 measurement noise vectors corresponding to vectors $\underline{z}_h$, $\underline{z}_q$, $\underline{z}_l$, $\underline{z}_f$;

M_h, M_q, M_l, M_f : are m_hx1, m_qx1, m_lx1 and m_fx1 meter placement matrices for head, fixed-head-node flow, consumer load and pipe flow respectively.

Construction of the matrices M_h, M_q and M_f is straightforward since the head, fixed-head-flow and pipe flow measurements

can be readily associated with the links of the augmented network diagram. Each row of these matrices has only one element, corresponding to the measured link, equal to 1 and the remaining elements of the row equal to zero. A consumer load at any node can be calculated as a sum of the flows in the pipes connected to this node. Thus, construction of the matrix M_1 involves examination of the network links incident to the measured nodes. If the link is directed 'to the node' the coresponding entry in the matrix is +1, otherwise the entry is -1. All remaining elements in the row are equal to zero.

4.3.3 Topological Observability

In subsection 4.3.1 it has been stated that the use of floating point calculation methods may be impractical for testing observability since they provide no insight into the location of the measurement deficiency. This question is related to the network topology and motivates a topologically based observability algorithm. In the present thesis we turn our attention to methods which do not depend on the numerical values of the Jacobian matrix entries, but rather investigate whether the measurement set provides enough information about the network topology to the state estimator. Observability determination, from this point of view, becomes the study of the topology of a graph derived from the original network according to the quality and quantity of the measurements in the metering scheme under

consideration.

Topological observability is derived from the condition that the Jacobian matrix must be of full rank [156]. To properly define this form of observability, consider the approximate measurement model presented in the previous section and summarised as

$$\underline{z} = M_m Y A_r^T \underline{x} + \underline{\omega} \qquad (4.3)$$

where

$\underline{z}^T = [\underline{z}_h^T | \underline{z}_q^T | \underline{z}_l^T | \underline{z}_f^T]$, mx1 measurement vector

$\underline{\omega}^T = [\underline{\omega}_h^T | \underline{\omega}_q^T | \underline{\omega}_l^T | \underline{\omega}_f^T]$, mx1 measurement noise vector

$M_m^T = [M_h^T | M_q^T | M_l^T | M_f^T]$, mx1 meter placement matrix

$\underline{x}$, nx1 state vector

Y, lxl hydraulic conductivity matrix

A_r, nxl reduced node-to-branch incidence matrix.

The definition of topological observability then follows from the measurement model given by Eq. (4.3).

'An n-node water system is <u>topologically observable</u> with respect to a given measurement set M if and only if the rank of the matrix $G=M_m Y A_r^T$ is equal to n'.

The structure of G does not depend on elements of Y, which change with operating point, but is determined by the meter placement and the network topology.

In the remainder of this chapter attention will be

focused on topological observability. Although the possible occurence of numerical problems while investigating observability from the topological point-of-view are not taken into account, it will be seen that methods using this approach can provide all the information required from an observability routine. The conditions necessary to achieve topological observability are discussed in the following section.

4.4 CONDITIONS FOR TOPOLOGICAL OBSERVABILITY

4.4.1 Preliminary Definitions

Some concepts and definitions have to be introduced before the topological observability conditions can be established. An appropriate starting point is an investigation how the measurements interrelate with the state variables.

By examining the structure of meter placement matrices M_h, M_q, M_l and M_f of Equation 4.2 the following remarks can be made:

a) A nodal head measurement carries information about the corresponding state variable;

b) A fixed-head-node flow measurement, similarly to the head measurement, also directly provides information about the value of the state variable;

c) The effect of a consumer load measurement is to interrelate the head of the measured node with the

heads of the nodes connected to it; and

d) A pipe flow measurement will produce an equation in the measurement model which interrelates the nodal heads corresponding to the ends of the monitored pipe.

Using the definition of an augmented network graph introduced in subsection 4.3.2, the concept of measurement assignment is a direct consequence of remarks a), b), c) and d) above.

A measurement z can be assigned to an edge e of the network graph if:

i) z is a head measurement and e is an edge connecting the measured node with the reference node; or

ii) z is a fixed-head-node flow measurement and e represents a link between the fixed-head node and the auxiliary node, or

iii) z is a consumer load measurement at either of the two ends of the pipe which corresponds to edge e; or

iv) z is a flow measurement taken at the pipe of the water distribution system which corresponds to edge e of the network graph.

If the measurement z is assigned to edge e, we also say that edge e is associated with measurement z.

Since the load measurement is the only one which can be assigned to different edges it is useful to introduce a concept of measured and unmeasured nodes. A node is said to be measured (unmeasured) if there is (is not) a load measurement available at this node.

The fundamental concept to be used in establishing the topological observability conditions is that of observable spanning trees.

Consider a set M of measurements taken in a water distribution system. A spanning tree of the augmented network graph G is an observable spanning tree if and only if it is possible to assign a measurement $z \epsilon M$ to each one of the edges of G such that no two edges are associated with the same measurement.

4.4.2 Conditions for Topological Observability

Conditions for topological observability have been derived by Krumpholz, Clements and Davis [156]. They have used the approximate measurement model, analogous to the one presented in Section 4.3.2, and applied a transformation which changes the problem from the nodal framework to the branch framework. The observability conditions are rephrased here to fit the definitions introduced in the previous sections.

Theorem 4.1

(Necessary Condition for Topological Observability)

If a water system is topologically observable with respect to a measurement set M, then there exists a spanning tree of its augmented network graph which is an observable spanning tree and whose branches are associated with measurements of M.

Proof: see [156]

Theorem 4.2

Suppose that there exists an observable tree in the augmented network graph whose branches are associated with measurements of a measurement set M. Then, if the vector formed by the diagonal hydraulic conductivities of the pipes does not lie on a certain (n-1) dimensional surface C, the water system is observable with respect to the measurement set M.

Proof: see [156]

Notice that, rigorously, Theorem 4.2 does not provide a sufficient condition for topological observability. However, those cases in which the existence of an observable tree does not imply topological observability are unlikely to appear in practice. In order for these cases to occur, the hydraulic conductivities of the pipes must combine themselves in such a way that they reduce the rank of G, as determined by the measurement set which corresponds to the

observable tree. This matrix would be of full rank for a different set of hydraulic conductivities of the pipes.

Thus, topological observability will be investigated here by seeking an observable spanning tree of the network graph. It should be remembered however that the existence of such a tree is only a necessary condition for topological observability. Situations in which such a tree exists, and yet, the system is topologically unobservable are mathematically possible, although unlikely to occur in practice.

Theorem 4.3 defines equivalent conditions of topological observability.

Theorem 4.3

Suppose that there exists a tree of the augmented network graph, then the following are equivalent:

i) the tree is of full rank

ii) the tree has a path property

iii) there exists a measurement assignment for the tree

A tree of the network graph is said to have the path property if every path of branches of the tree between two nodes with unmeasured load contains at least one branch whose flow is measured. The validity of the Theorem 4.3 follows immediately from the path property of trees of full rank.

4.5 OBSERVABILITY DETERMINATION BY THE MATCHING METHOD

4.5.1 The Measurement to Branch Assignment

This section proposes a new method to check if a water system is topologically observable with respect to a given measurement set M. This method does not attempt to directly find a spanning tree of the network graph with the required measurement assignment. Instead, it seeks a subset B_o of the set of branches B of the augmented network graph whose elements can be associated with elements of the measurement set M in one-to-one fashion. This one-to-one correspondence is referred as the assignment (M_o,B_o) and the number of its elements is called the assignment length β. The search for the observable spanning tree can now be restricted to the subgraph G_o of G formed by the branches B_o of the measurement assignment (M_o,B_o). If such a tree is found the network is topologically observable. Otherwise, an attempt is made to modify G_o by breaking loops and adding new, previously unassigned branches to enable construction of a tree of full rank. Failing that, the network is declared unobservable and the algorithm returns a maximal observable forest of G. Figure 4.2 schematically shows the basic steps required by the proposed method to test observability of a water distribution system with a given set of measurements M. The following subsections describe the details of each block of Fig. 4.2

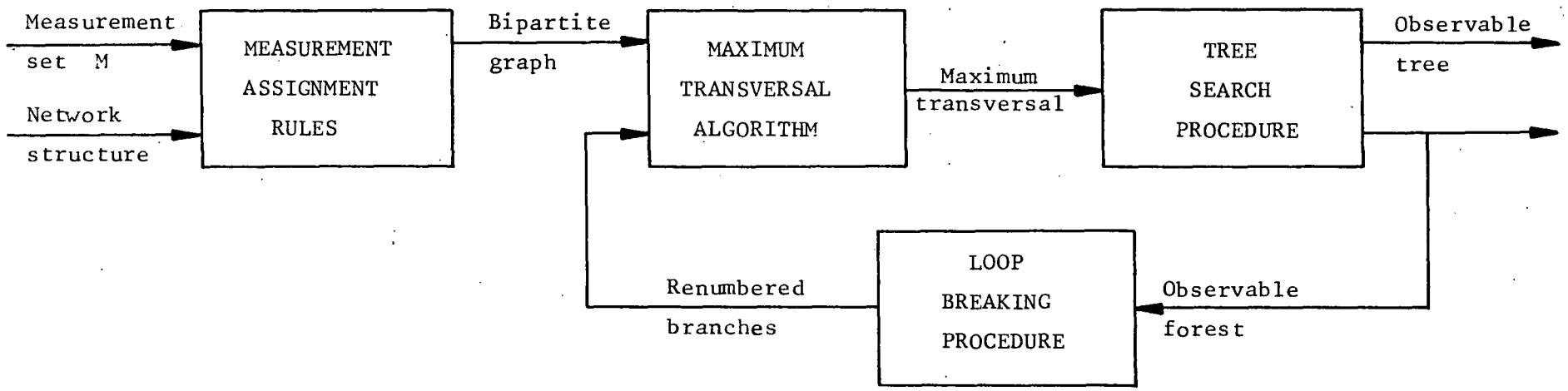


Figure 4.2 Basic steps in the Matching Method

4.5.2 Problem Formulation Using Bipartite Graphs and Matchings

The first step in the search for a measurement assignment (M_o, B_o) is the formulation of the problem using bipartite graphs.

A bipartite graph is a graph whose vertex set can be partitioned into two subsets, X and Y, so that each edge of the graph has one end in X and one end in Y; the partition (X,Y) is called a bipartition of the graph [75], [102].

In addition, the following definitions will be required in the sequel. The adjacency set Ω_i of a measurement z_i in a water distribution system is the set of all branches of the augmented network graph related to the measurement z_i through the approximate measurement model (4.2). Each element of Ω_i is said to be adjacent to a measurement z_i.
The adjacency set can be seen as another means of expressing information contained in the meter placement matrices M_h, M_q, M_l and M_f defined in the subsection 4.3.2.

To associate the elements of the measurement set M with the branch set B of the augmented network graph, a bipartite graph of the type (M,B) is constructed. The edges of this graph are determined by the following rules, derived from the remarks in subsection 4.3.2

a) If measurement z_i is a head measurement at the node i, the corresponding vertex $z_i \epsilon M$ is connected to the

vertex $b_i \epsilon B$ which represents itself a branch between the node i and the reference node;

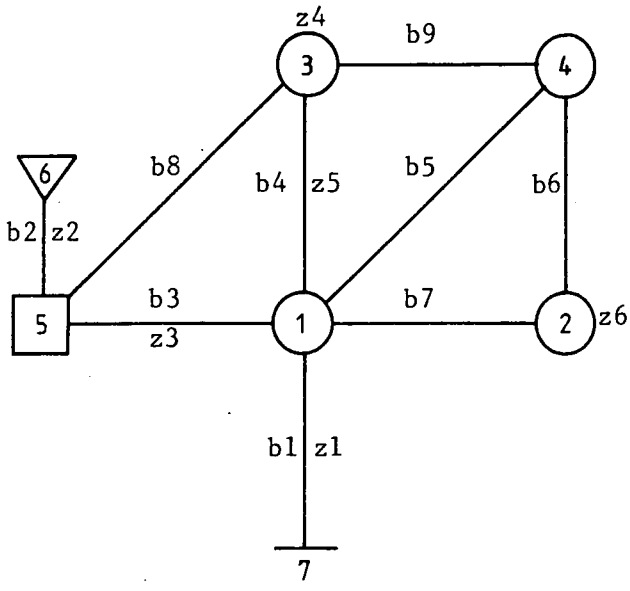
b) If measurement z_i is a fixed-head flow measurement the edge of the bipartite graph (M,B) connects z_i with a branch b_i which links the fixed-head node with the corresponding auxiliary node of the augmented network graph;

c) If measurement z_i is a measurement of flow in a pipe represented by $b_i \epsilon B$, the corresponding edge of the bipartite graph connects z_i and b_i;

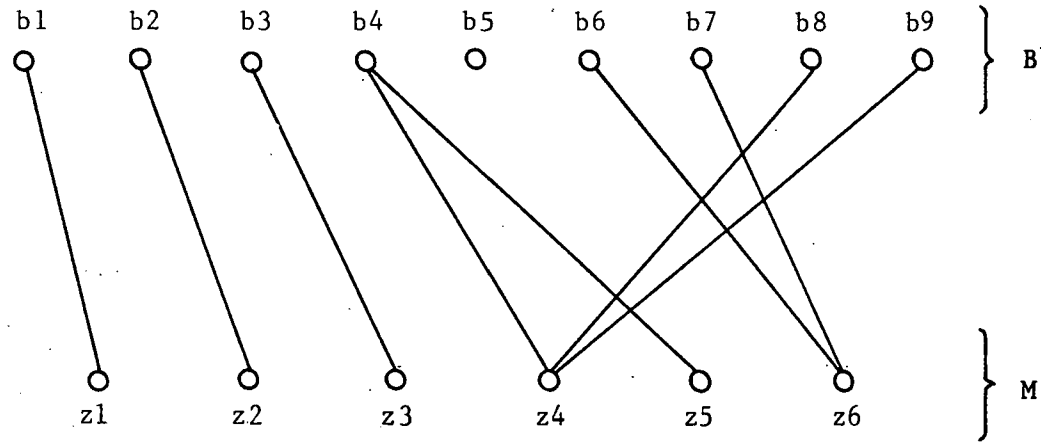
d) If measurement z_i is a consumer load measurement, and the adjacency set for this measurement is Ω_i then the vertex $z_i \epsilon M$ is connected to all vertices $b_i \epsilon \Omega_i$.

To illustrate the concept of the (M,B) bipartite graph, consider the augmented network graph with some measurements as indicated in Fig. 4.1 b. The corresponding (M,B) bipartite graph is presented in Fig. 4.3 b.

In the search for a (M_o, B_o) assignment in the bipartite graph the concept of matching in a graph is needed. A subset M of the set of edges of a graph G is a <u>matching</u> in G if its elements are edges with distinct ends, such that no two of them are incident to the same vertex. If a vertex v is incident to some edge of the matching M, then M is



(a)



(b)

Figure 4.3 (a) Augmented network graph

(b) Bipartite graph (M,B)

said to saturate v, and v is M-saturated [75].

We are particularly interested in the case where the graph G is a bipartite graph. For a bipartite graph with bipartition (M,B), a complete matching of the vertices in set M into those in set B is a matching in which there is one edge incident to every vertex in M [75]. Figure 4.4 shows a complete matching of M into B for the bipartite graph of Fig. 4.3 b.

A matching M is a maximum matching in a graph G if G has no matching M whose number of edges is greater than the number of edges in M. A complete matching is a maximum matching.

From the above definitions, it appears that an assignment (M_o, B_o) can be seen as a maximum matching of M into B.

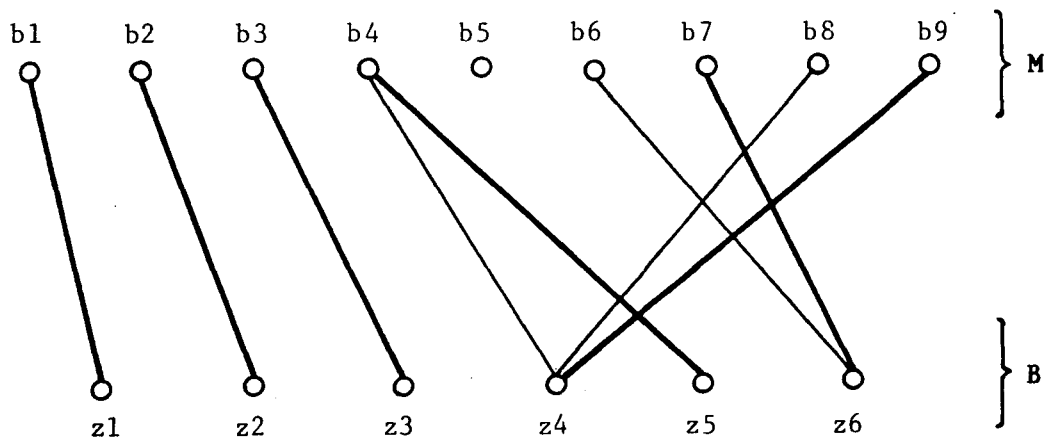


Figure 4.4 Maximum matching of M into B for the graph of Fig. 4.3 (b)

The condition for the existence of a maximum matching is stated in form of a theorem whose proof can be found in references [75] and [102]. Theorem 4.4 below requires the following definitions:

If **M** is a matching in a graph G, an M-alternating path in G is a path whose edges are alternately in E-**M** and in **M**, where E is the set of edges of G. An M-augmenting path is an **M**-alternating path whose endpoints are **M**-unsaturated.

Theorem 4.4

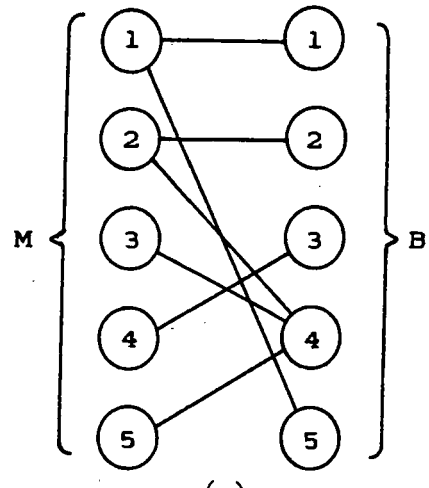
A matching **M** in a graph G is a maximum matching if and only if G contains no **M**-augmenting path.

It is important to note that the theorem is general in that it does not restrict the number of elements of the bipartite sets M and B. It also guarantees that a maximum matching can always be found for a connected graph G. The following corollary can be derived from the theorem 4.4.

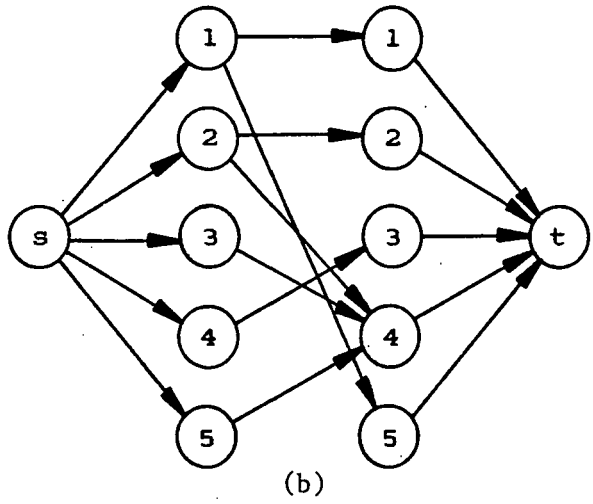
Corollary 4.1 [102]

The number of edges in a maximum matching of a bipartite graph G is constant and is equal to the maximum flow in the network built on G.

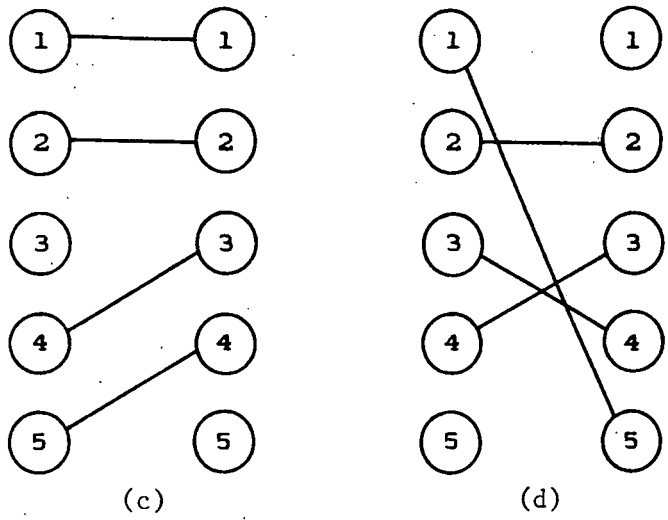
The corollary is illustrated in Fig. 4.5. There are



(a)



(b)



(c) (d)

Figure 4.5 (a) Bipartite graph G

(b) Matching network

(c), (d) Maximum matchings

several possible maximum matchings; each of them however has the same number of components.

4.5.3 An Algorithm to Find an (M_o, B_o) Assignment

The (M_o, B_o) assignment problem described in subsection 4.5.1 can now be solved using a bipartite graph formulation. The algorithm presented here bases on the work of Hopcroft and Karp who used the concept of a layered network [81] in order to devise a reassignment path. An efficient implementation of this algorithm, specialised for obtaining a maximum transversal of a square matrix, is given by Duff [91] and the FORTRAN code is available as a Harwell subroutine MC21A [90]. The author's generalisation of the Duff algorithm consists essentially of enabling a different number of elements in the bipartite sets M and B.

The maximum matching is constructed in m major steps, where m is the number of measurements. After the k-th step we have a maximum matching of the first k measurements into the set of branches B. The search for a larger matching, which includes an M-unsaturated vertex $u \epsilon M$ consists of forming a tree of a bipartite graph G called an M-alternating tree rooted at u. Such a tree has the following properties

a) The M-unsaturated vertex u is a vertex of a tree; and

b) For every vertex v of the tree, the edges of the unique path connecting v to u on the tree are edges which are alternately contained and not contained in M.

An example of the matching M in the bipartite graph G, presented in Fig. 4.5 a, after processing two elements of the set M, is given in Fig. 4.6 a, and the M-alternating tree rooted at the third element of M is shown in Fig.4.6b The reassignment corresponds to replacing the branch 2-4 with two branches 2-2 and 3-4 thus enlarging the current matching.

There are several different techniques available for finding an M-alternating path. Duff [91] has found that

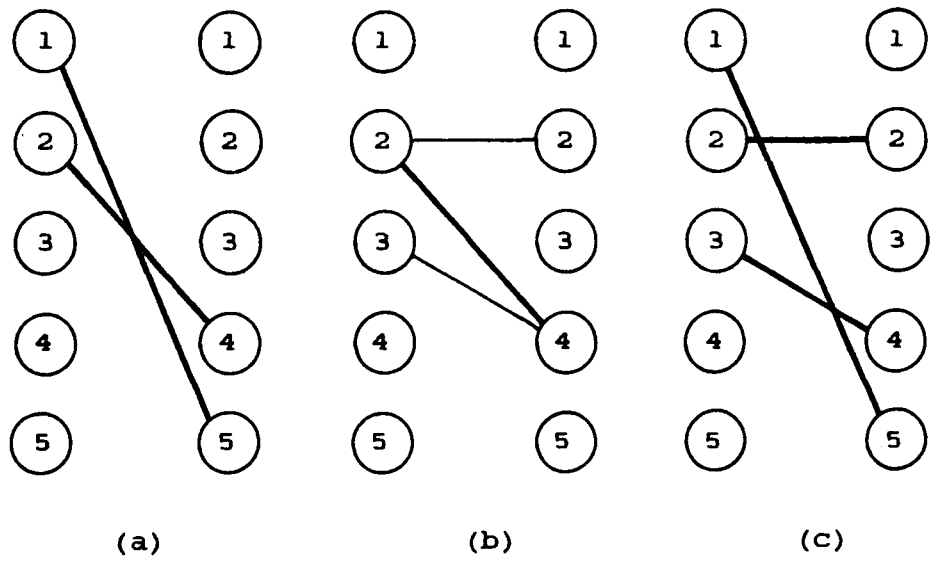


Figure 4.6 (a) Matching M after two steps

(b) M-alternating tree rooted at 3

(c) Matching M after three steps

application of a depth first search with a look-ahead technique gives the best practical results despite the fact that other algorithms give better theoretical bounds on computational complexity.

In accessing the vertices of a graph in a depth first search (DFS), we search edges from the current vertex and add to our path the first vertex encountered that we have not yet visited. This becomes the current vertex and we proceed from it as before. If all the vertices that can be reached from the current one at the end of the path are already visited, we backtrack to the vertex added to the path immediately before the present one, make that the current vertex and proceed as before. The depth first search algorithm is illustrated in Fig. 4.7, where heavy lines denote edges in the path and the vertices are numbered in the order in which they are visited.

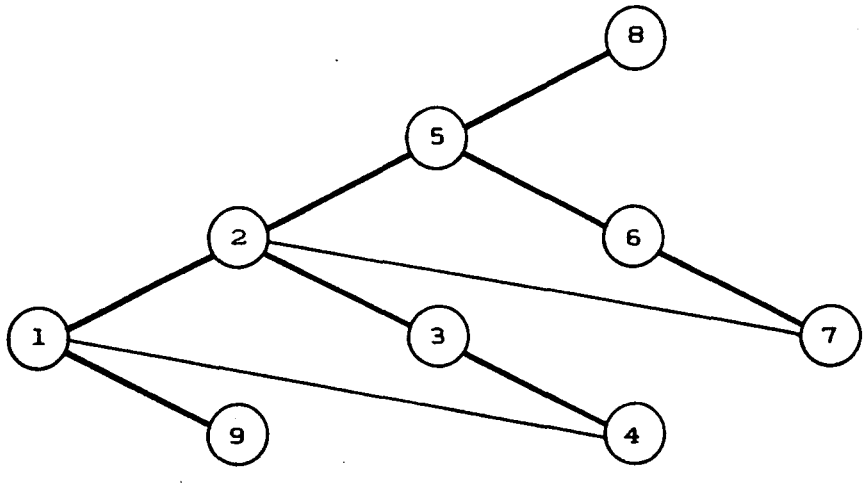


Figure 4.7 Graph indicating DFS ordering

In practice, the efficiency of a depth first search scheme can be enhanced by checking all unvisited vertices reached from the current one to see if any are free. If so, the M-alternating path has been found and the matching extends by one. This look-ahead technique of Duff saves visiting vertices 3 to 8, of Fig. 4.7, in the case that vertex 9 is found free.

The theoretical upper bound on computational complexity of the DFS algorithm with a look-ahead technique can be found as the product of the number of vertices and edges of the bipartite graph concerned. However, in most practical cases the algorithm performs as if its complexity was linearily dependent on the sum of the number of vertices and edges in the bipartite graph.

4.5.4 Determination of an Observable Spanning Tree in the (M_o, B_o) Assignment

Section 4.5 started by introducing the concept of (M_o, B_o) assignment. The observability problem was then formulated in terms of bipartite graphs and matchings. Finally, subsection 4.5.3 presented methods to find a maximum matching in a bipartite graph, and to solve the optimal assignment problem. This section will look into the connections between the observability conditions of section 4.4 and the (M_o, B_o) assignment.

To begin investigating how conclusions about observability can be drawn from (M_o, B_o) assignment, assume that the water distribution system is topologically observable with respect to a certain measurement set M. By Theorem 4.1, an observable spanning tree of the augmented network graph of the water network exists. Let M_1 be the subset of M whose measurements are assigned to the branches of the observable spanning tree. The length of such an assignment (M_1, B_1) equals n since the number of branches of the spanning tree of the graph with n+1 vertices is n. On the other hand, it is possible to find a measurement to branch assignment (M_o, B_o), corresponding to the maximum matching M, which by definition has the length at least n and which contains (M_1, B_1). Thus, the search for an observable spanning tree can be performed directly on the subgraph G_o of the network graph G formed by all the branches of the measurement assignment (M_o, B_o).

The important conclusion to be drawn from the above considerations is:

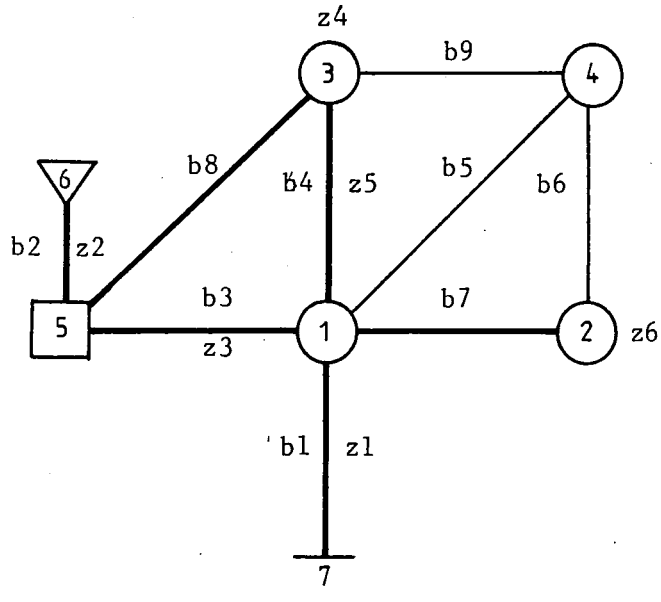
'If a water distribution system is topologically observable with respect to a measurement set M then there exists a measurement assignment (M_o, B_o) whose branches contain a spanning tree of the network graph. Conversely, if there is no assignment (M_o, B_o) which contains a spanning tree of G then the water system is topologically unobservable with respect to M.'

However, one cannot claim that for every assignment (M_o,B_o) in the observable network the spanning tree of G can be found. This is because the assignment can give rise to loops in the network graph, which has an effect of limiting the number of vertices incident to the branches of B_o. The possibility of creating such loops is apparent since the water network connectivity is not taken into account during construction of the maximum matching M. Figure 4.8 presents two possible measurement to branch assignments for the system of Fig. 4.3. The assigned branches are marked on the network graph with bold lines. In the first case the network graph G_o, formed by the branches B_o of the (M_o,B_o), does not contain a spanning tree of G and in the second case it does.

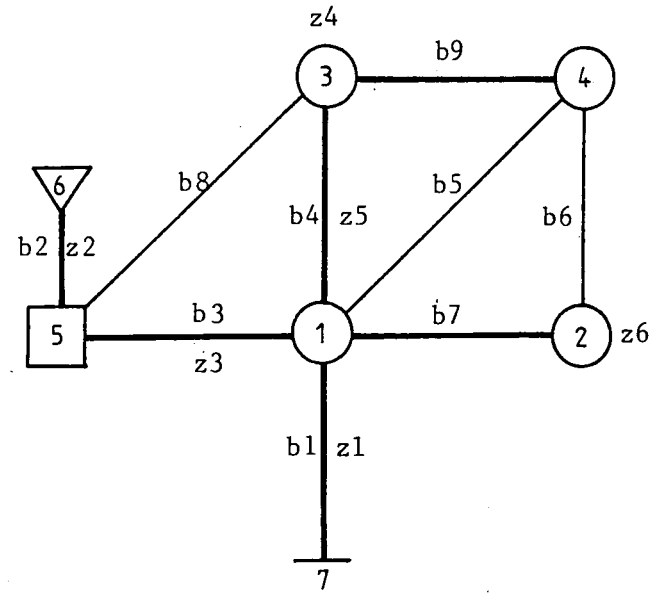
A procedure is then required which would aid the matching algorithm in maximising the number of vertices incident to the branches of B_o. Such a procedure will be described later in this subsection.

In some special cases the question of topological observability can be answered immediately by examination of the assignment length β,

a) If the (M_o,B_o) assignment has the length $\beta=l$, where l is the number of branches of the augmented network graph G, then the water network is observable with respect to the measurement set M.



(a)



(b)

Figure 4.8 (a), (b) Network graphs with different measurement assignments

Every spanning tree of the graph G is an observable spanning tree since B_o contains all the branches of G.

b) If the length of the (M_o, B_o) assignment $\beta < n$, then the water network is topologically unobservable due to the fact that B_o contains less branches than any spanning tree of G. Consequently, no observable spanning tree exists.

However, if $n \leq \beta < 1$ an attempt to find a spanning tree of G in G_o must be made. Because of the possibility of creating 'observable' loops by the branches of B_o, as indicated in Fig. 4.8, a sequence of measurement assignments is generated such that each contains no more loops than the previous one, and a tree search procedure is reinitialised for every new set of branches B_o. This process continues until an attempt has been made to disconnect every loop or a spanning tree of G has been found. If a spanning tree is not found for any of the consecutive sets B_o, the system is declared topologically unobservable, and a tree search procedure returns a maximal forest of G in G_o. This is a valuable piece of information since it can be used to add pseudo-measurements in order to make that portion of the system observable. Alternatively, the state estimation can be carried out only for the observable part of the system.

The procedure to maximize the number of vertices incident to the branches of the set B_o is outlined below.

Essentially, it consists of identifying 'observable loops' in a maximal forest and reassigning the measurements associated with the loop edges so as to include the forest linking edges in the updated measurement assignment (M_o, B_o).

Notice that the search for loops is equivalent to the search for biconnected components of the graph G_o since each loop is wholly contained in one of the bicomponents. We can therefore use an algorithm of Hopcroft and Tarjan [138] to mark the bicomponents which contain more than one edge. The algorithm performs a depth first search along the edges of the graph. Each new vertex reached is placed on a stack, and for each vertex a record is kept of the lowest vertex on the stack to which it is connected by a path of unstacked vertices. When a new vertex cannot be reached from the top of the stack, the top vertex is deleted, and the search continues from the next one on the stack. If the top vertex does not connect to a vertex lower than the second one on the stack, this second vertex is an articulation vertex of the graph. All edges examined during the search are placed on a further stack, so that when an articulation vertex is found the edges of the corresponding biconnected component may be retrived.

When the stack is exhausted, a complete search for a connected component has been performed. If the graph is connected, the process is complete. Otherwise, an unreached vertex is selected as a new starting point and the process

is repeated until all the graph has been exhausted. Isolated vertices are merely skipped, since they have no adjacent edges.

The reassignment procedure can now be solved efficiently by introducing an additional loop-measurement incident to all loop edges of G_o and performing one step of the maximum matching algorithm on the modified bipartite graph. An M-alternating path rooted at the loop-measurement is required to terminate at a forest linking edge thus diminishing the number of forest components of G_o. The procedure terminates if either an observable spanning tree of G is found or a forest cannot be linked.

The flow chart exhibited in Fig. 4.9 summarizes all the steps in the implementation of the matching method.

Computational complexity of this algorithm is given as $\max[O(A),O(B),O(C),O(i \cdot D),O(i \cdot E)]$ where A, B, C, D and E are the labels of those parts of the algorithm which are critical to an estimation of the complexity and i is a maximal number of repetitions of loop-breaking procedure D and E. The upper bound on i is equal n since this is the maximum number of forest components of a graph with n vertices. The number of operations performed by the algorithm is then proportional to $\max[n,n\tau,n,n\tau,nn] = n\tau$ where τ is the number of edges of the bipartite graph and is of order l (l is the number of edges of the agumented network graph). However, in most practical cases the maximum

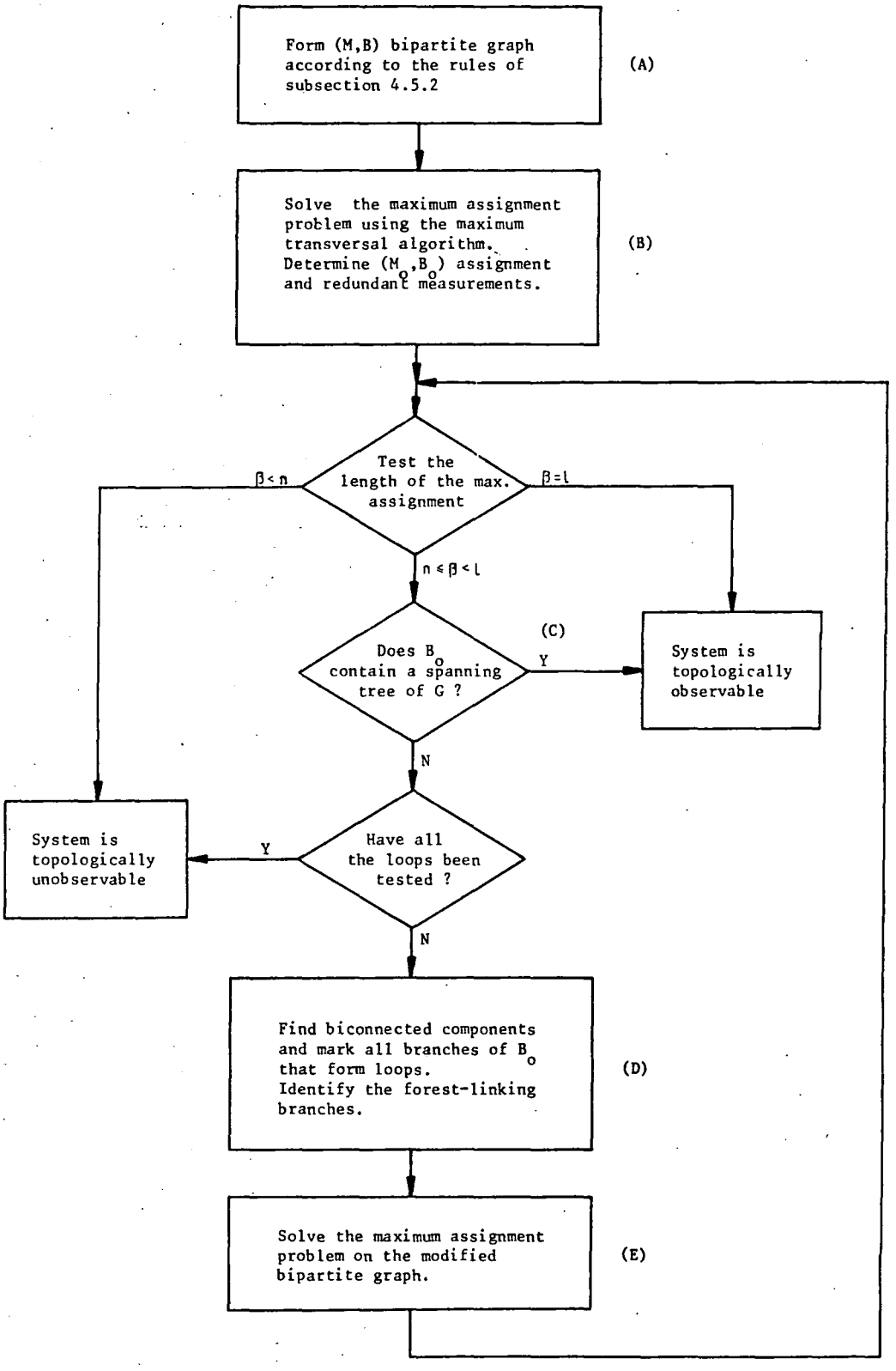


Figure 4.9 Flow chart for the Matching Method

matching algorithm shows a complexity which is far less than its theoretical upper bound and is approximately equal $n+\tau$.

4.5.5 Examples

Preliminary tests on the matching method were performed using the water systems whose augmented network graphs are shown in Fig. 4.10 along with the corresponding measurement sets. The system of Fig. 4.10 a to c is the one used to explain the observability algorithm and the system of Fig. 4.10 d to f has been taken from the literature on water system control [205]. Initial measurement assignments are given in Fig. 4.11 and the observable trees or forests are depicted in Fig. 4.12.

For the system in Fig. 4.10 d, a loop-breaking procedure needs not to be activated since the initial measurement assignment contains an observable tree. However, this largely depends on the vertex labelling and all the other examples refer to a more general case where the maximum matching produces only a forest.

To further assess the performance of the matching method, it has also been used to investigate the observability of the realistic 34-node system described in Appendix B. A tree generation routine has been used to set up a base case to test the method. The measurement set has been devised such that a different measurement is assigned

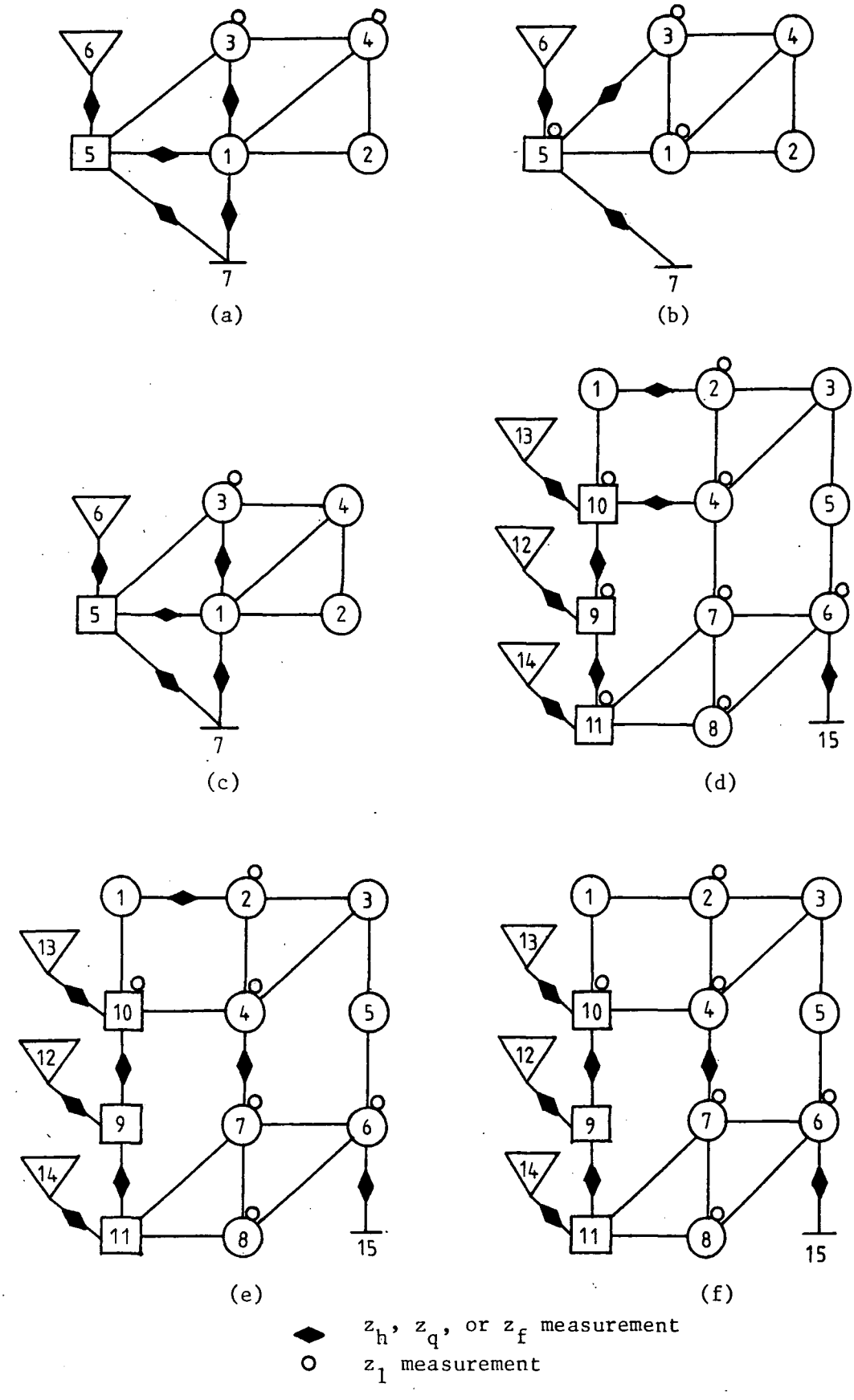


Figure 4.10 Augmented network graphs with corresponding measurement sets

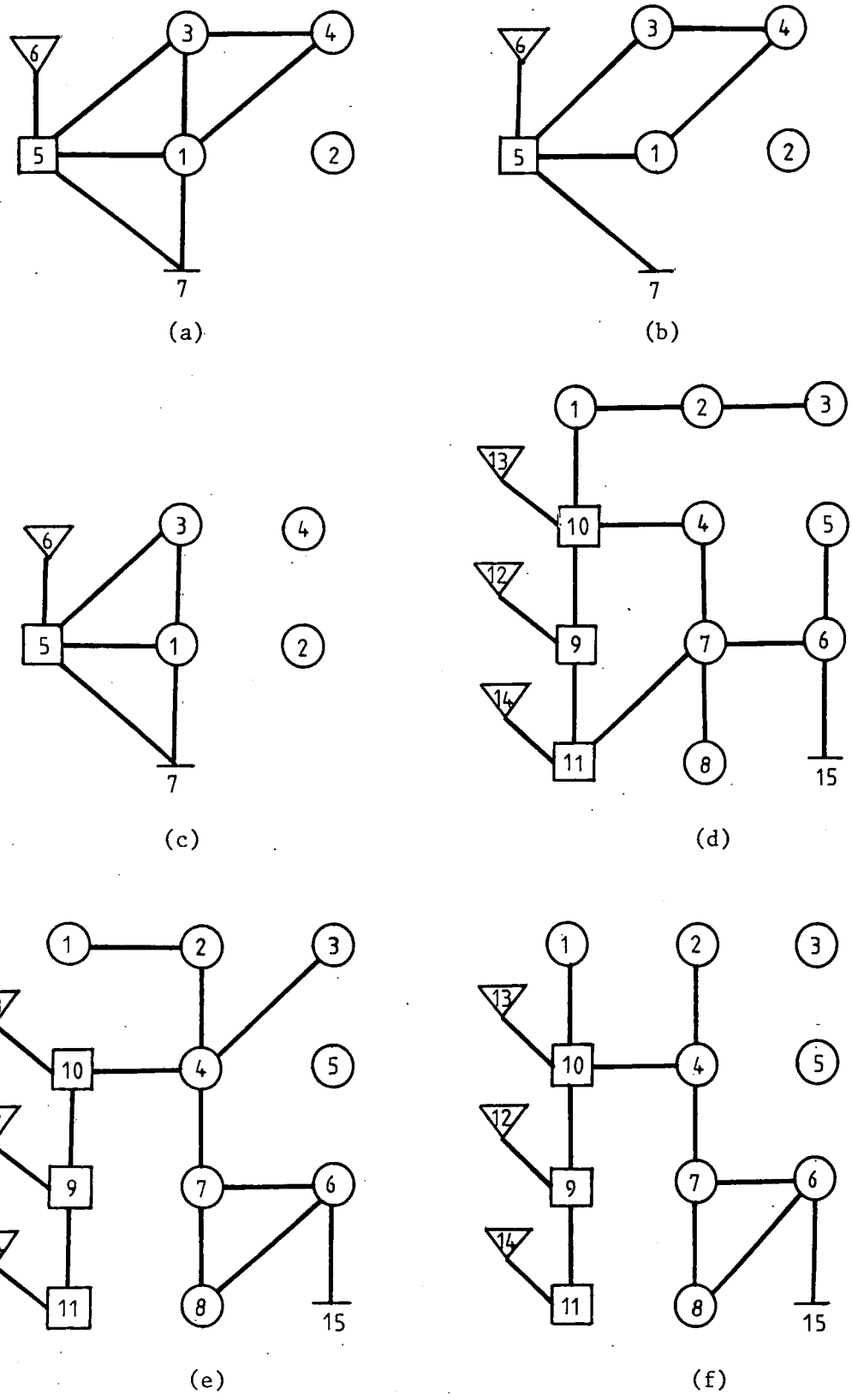


Figure 4.11 Initial measurement-to-branch assignment for the systems of Fig. 4.10 (a)-(f)

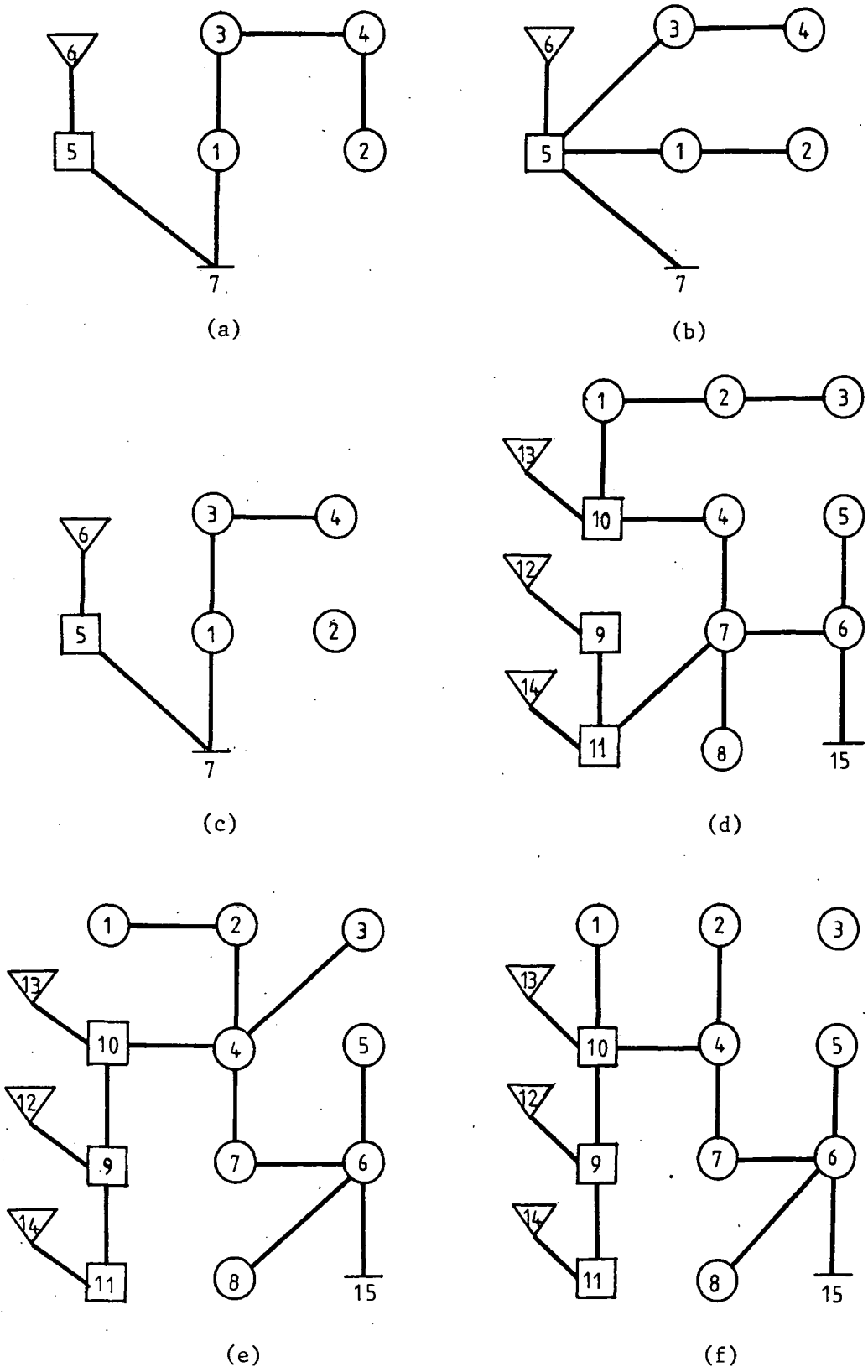


Figure 4.12 Observable trees/forests obtained through the Matching Method

to each branch of the tree. From this example five other cases are derived by changing the measurement set. The various cases are defined next.

Case 1 (base case): 42 measurements, namely 6 head measurements, 8 fixed-head-node flow measurements, 9 flow measurements and 19 load measurements, are taken throughout the system such that an observable spanning tree exists.

Case 2: The measurement set of the base case is reduced by one load measurement at node 8. The system is unobservable with respect to this reduced measurement set.

Case 3: One new load measurement is added at node 10. Despite the system now having the same number of measurements as in the base case, it remains unobservable and the added load measurement contributes only to the local redundancy.

Case 4: The measurement set of Case 3 is augmented by 5 measurements; however, this is done in such a way that the system remains unobservable with respect to the new measurement set.

Case 5: Two line flow measurements are replaced by two load measurements in the measurement set of Case 3. This allows the redundant measurement to be used rendering the system observable.

Case 6: Four new line flow measurements are added to the measurement set of Case 5. Since an observable spanning tree can be found for the previous measurement set the new

system is also observable.

The augmented network graphs for the six cases are presented in Figures 4.13-4.18. The results and computational times are summarized in Table 4.1. The computer runs were performed on a Perkin-Elmer 3220 minicomputer, using FORTRAN 77. The following comments apply to the data displayed in Table 4.1.

In all the cases the matching method correctly identified observable and unobservable systems. For Case 1, the observable spanning tree corresponding to the measurement set is retrieved after performing three reassignments. In cases 2, 3 and 4 the matching routine returns a maximum observable forest thus indicating parts of the network requiring meter reinforcement. In case 5, similarily to case 1, the observable spanning tree is found after 3 reassignments. The addition of new measurements to the measurement set of case 5 results in a maximum matching which contains an observable spanning tree. In effect the reassignment procedure does not have to be initialised.

The computing times displayed in Table 4.1 demonstrate strong dependence on the number of measurements regardless of whether the system is observable or unobservable. This result is expected since an increase in the number of measurements also increases the possibility that the maximal observable tree or forest is found directly at the stage of maximum matching. In practice, where the measurement

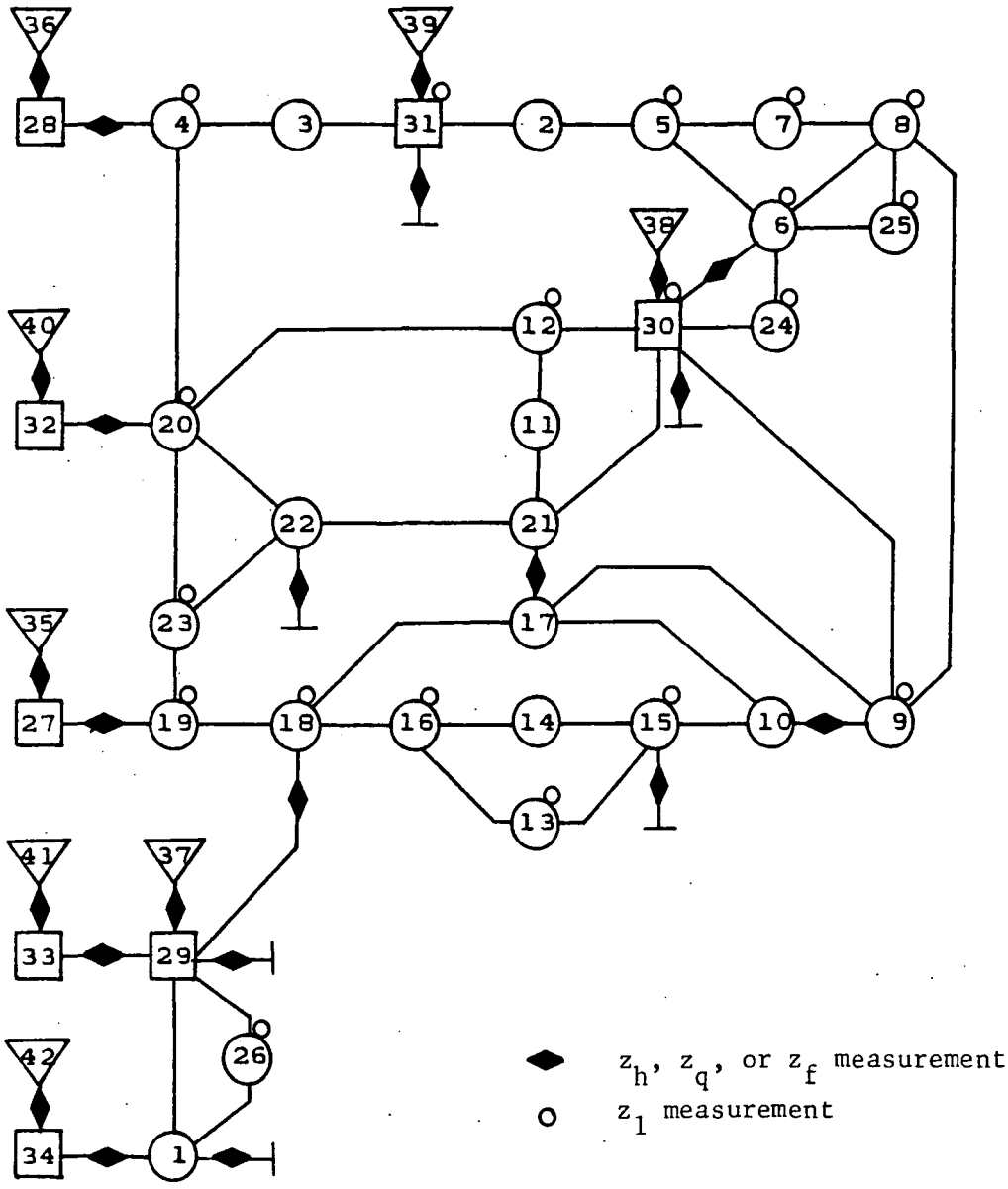


Figure 4.13 Augmented network graph for the measurement configuration of Case 1

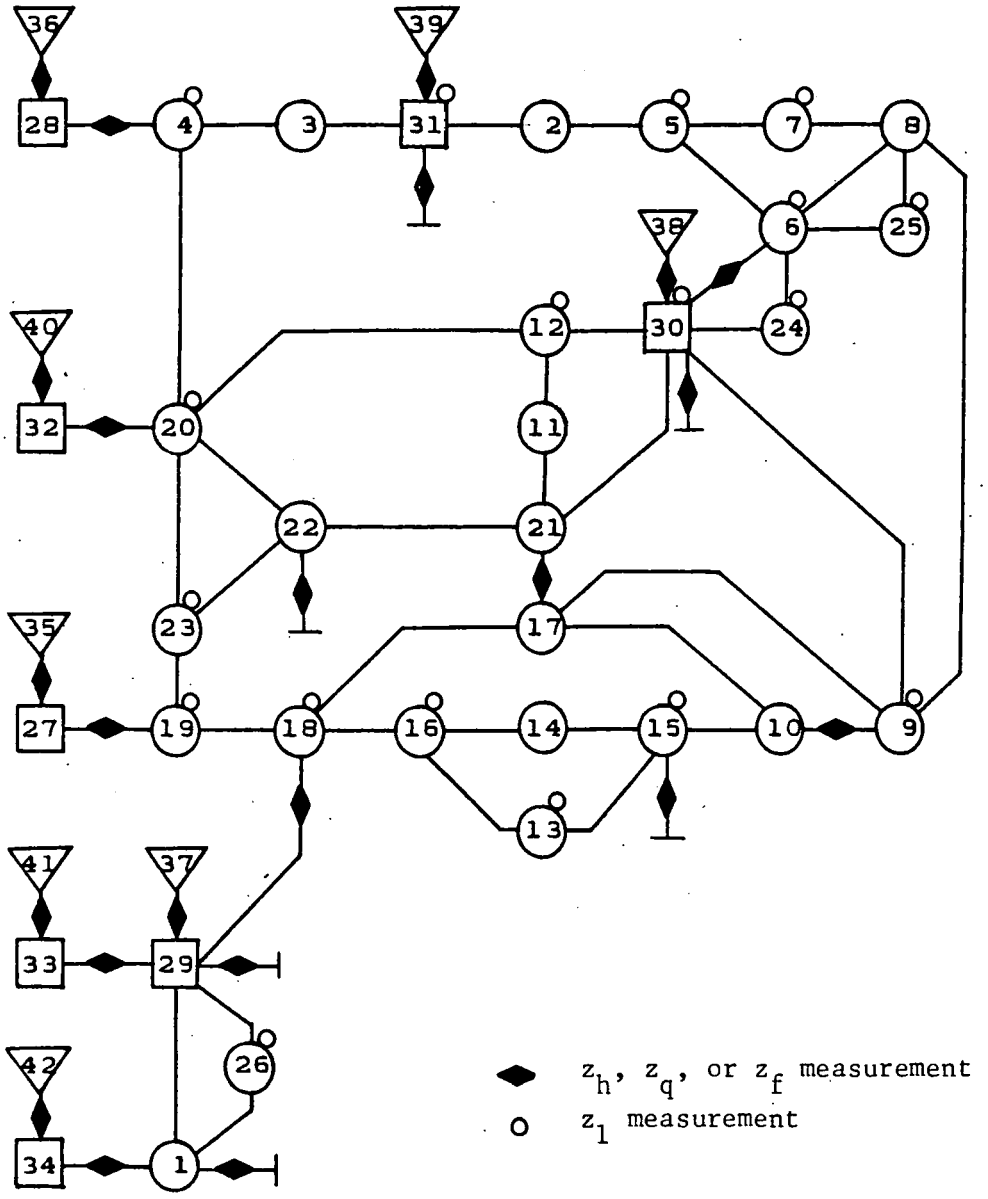


Figure 4.14 Augmented network graph for the measurement configuration of Case 2

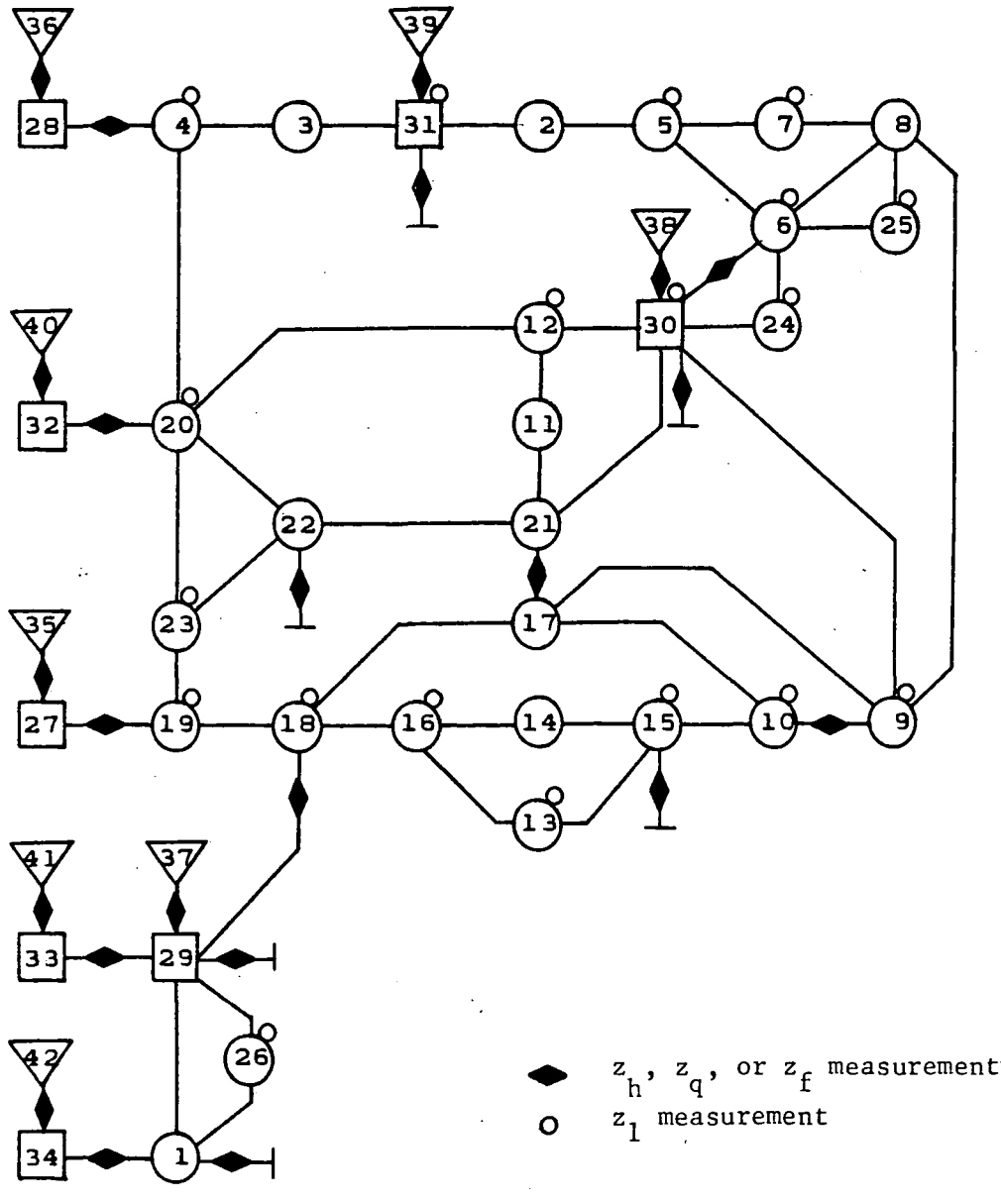


Figure 4.15 Augmented network graph for the measurement configuration of Case 3

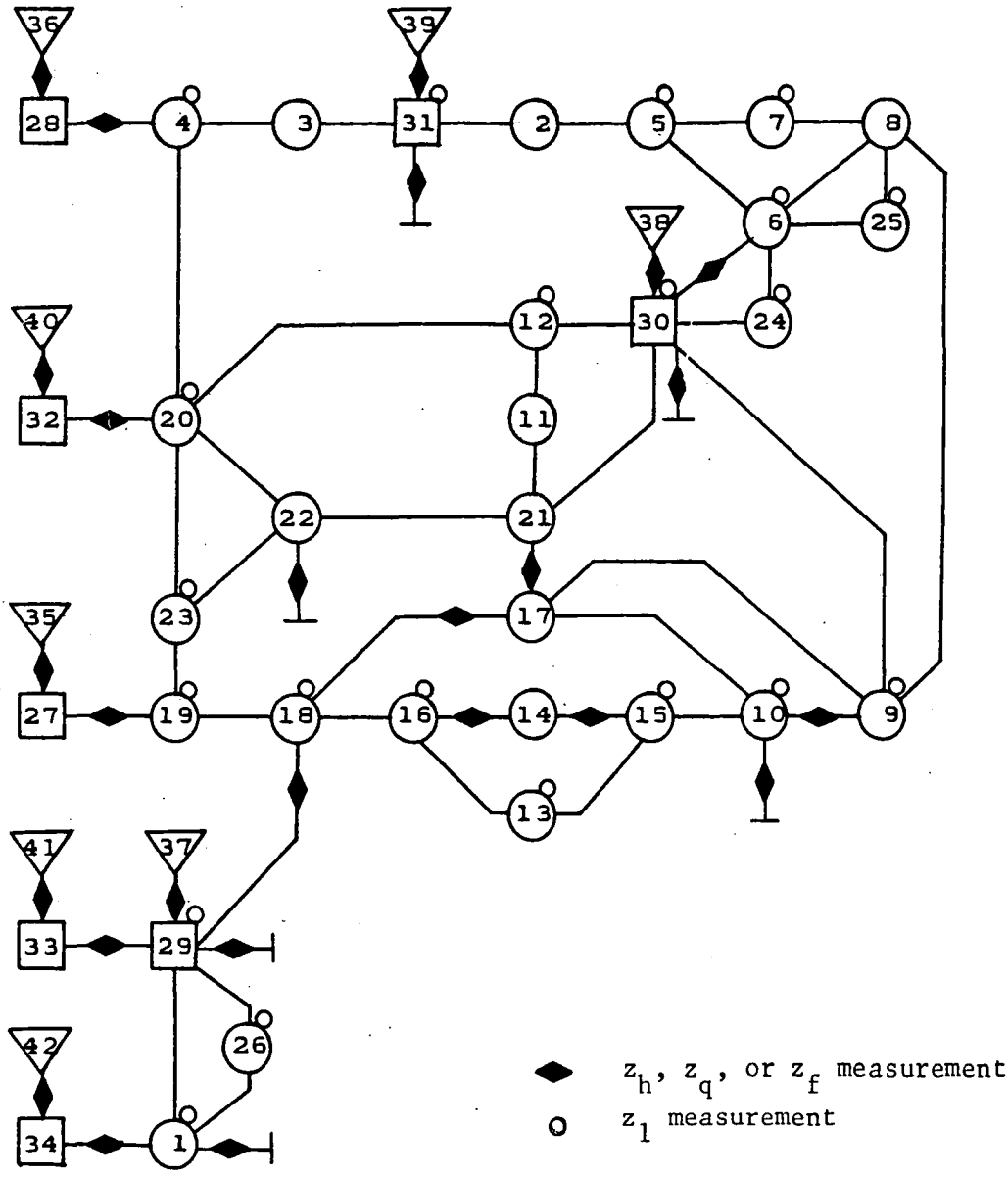


Figure 4.16 Augmented network graph for the measurement configuration of Case 4

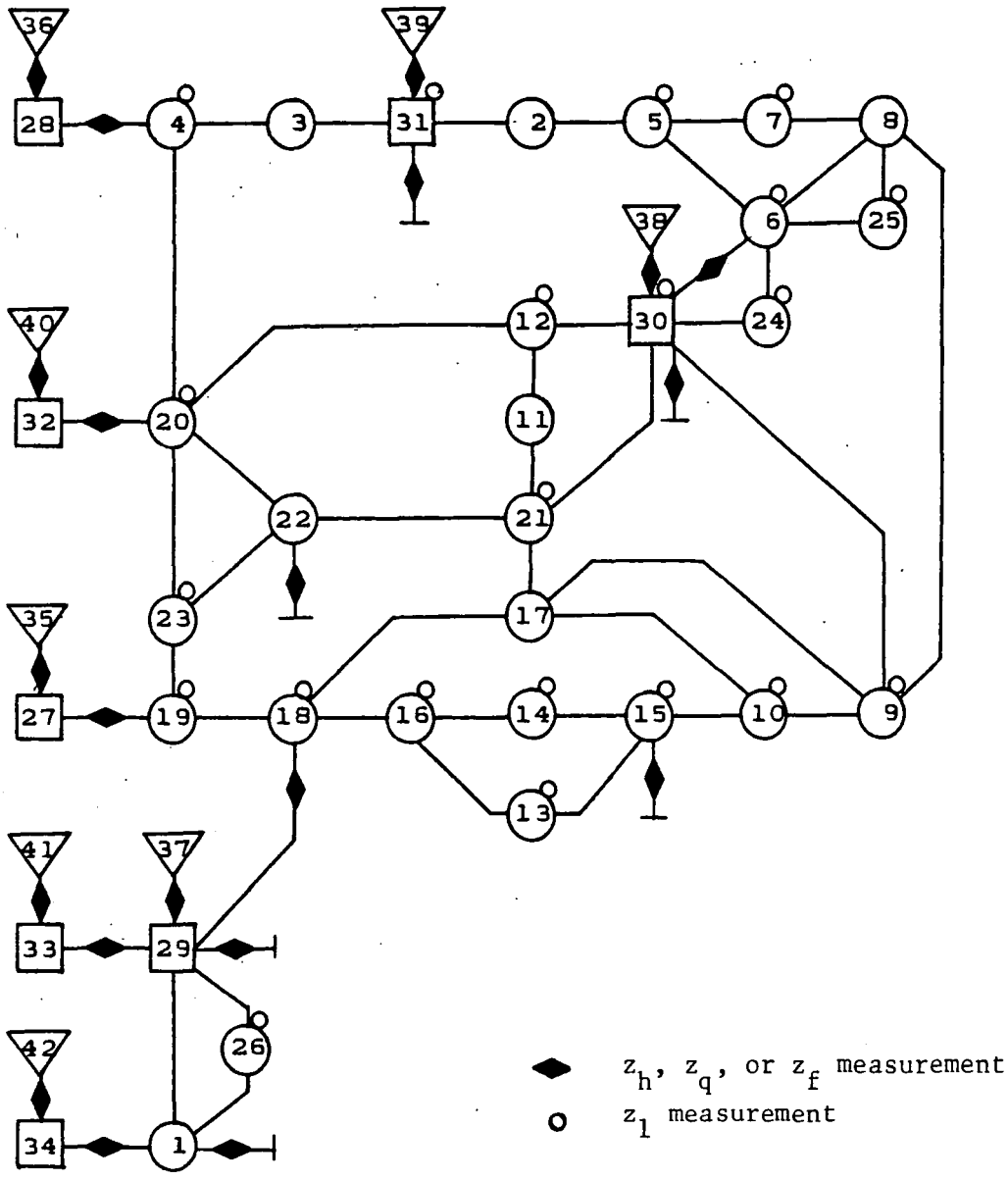


Figure 4.17 Augmented network graph for the measurement configuration of Case 5

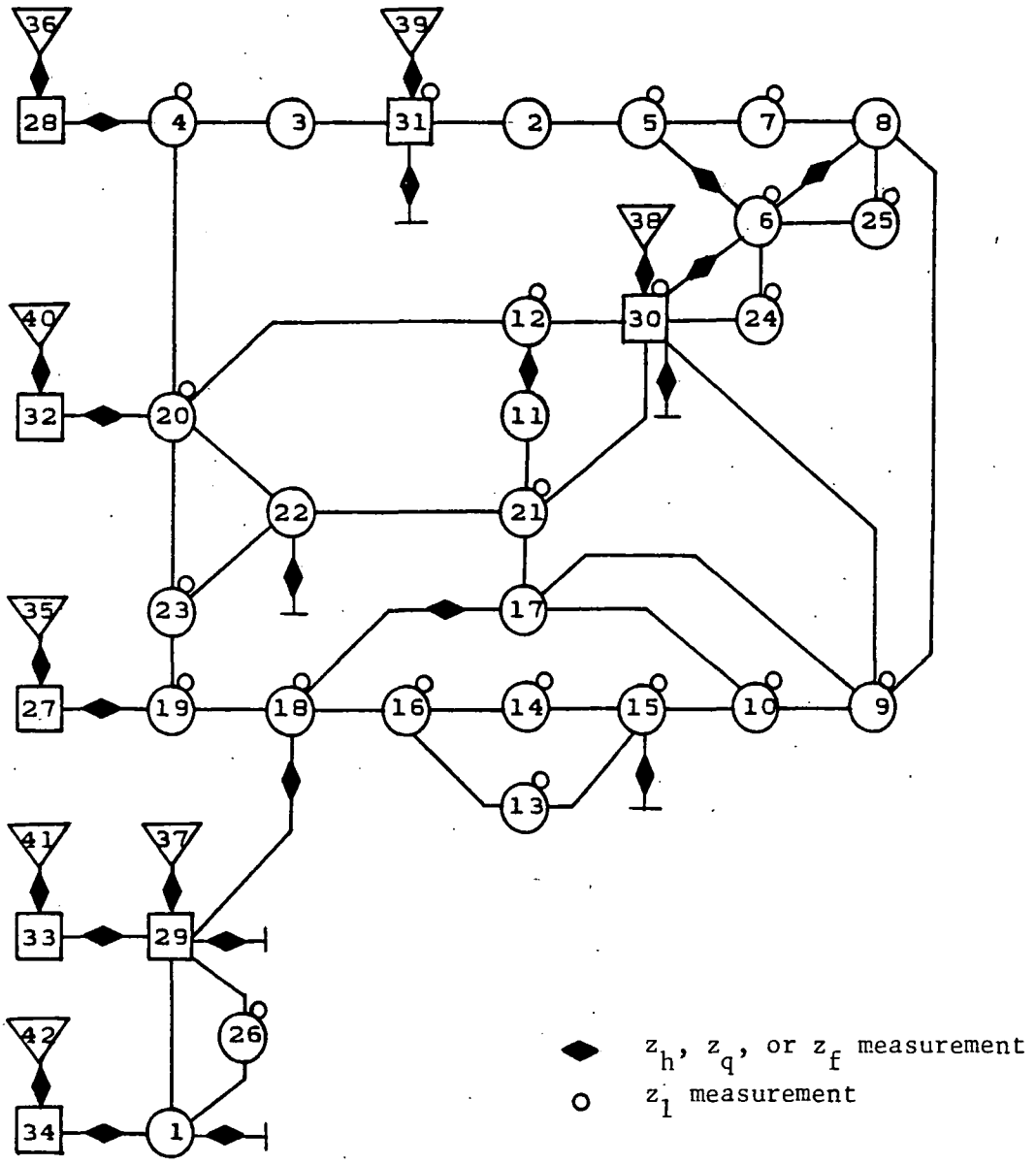


Figure 4.18 Augmented network graph for the measurement configuration of Case 6

TABLE 4.1

Computational results for the Matching Metchod

Case / Observability *	1 OBS.	2 UNOBS.	3 UNOBS.	4 UNOBS.	5 OBS.	6 OBS.
Number of measurements	42	41	42	47	42	46
Comp. time for bipartition [s]	0.016	0.016	0.016	0.017	0.016	0.016
Comp. time for max. matching [s]	0.008	0.008	0.008	0.008	0.008	0.008
Number of reassignments	3	3	2	1	3	0
Comp. time for ** tree search [s]	0.032	0.031	0.024	0.015	0.032	0.008
Comp. time for ** loop search [s]	0.060	0.079	0.040	0.020	0.058	-
Comp. time for ** reassignment [s]	0.017	0.022	0.011	0.005	0.017	-
Total time [s]	0.133	0.156	0.099	0.065	0.131	0.032
Result *	OBS.	UNOBS.	UNOBS.	UNOBS.	OBS.	OBS.

* OBS. = Observable
UNOBS. = Unobservable

** Sum for all reassignments

redundancy is an inherent feature of the measurement set, the matching method is likely to perform an observability test very efficiently. However, even in the worst case example the computing time of 0.156 s qualifies the method for on-line operation.

4.6 OBSERVABILITY DETERMINATION BY DIRECT SEARCH FOR AN OBSERVABLE SPANNING TREE

4.6.1 Problem Formulation

Unlike the matching algorithm, the method to be proposed in this section investigates observability by directly seeking a spanning tree in the augmented network graph defined in subsection 4.3.2. This graph is essentially a network graph extended by auxiliary nodes associated with the fixed-head-nodes of the water network and by one head-reference node. However, the number of edges of this graph varies according to the measurement set. Every measurement of a nodal head results in additional edge between a measured and a reference node.

The definition of an augmented network graph is compatible with the measurement assignment rules defined in subsection 4.4.1. The head, fixed-head-node flow and flow measurements are uniquely associated with the corresponding edges of the augmented network graph. Thus, according to theorems in subsection 4.4.2, the question of topological

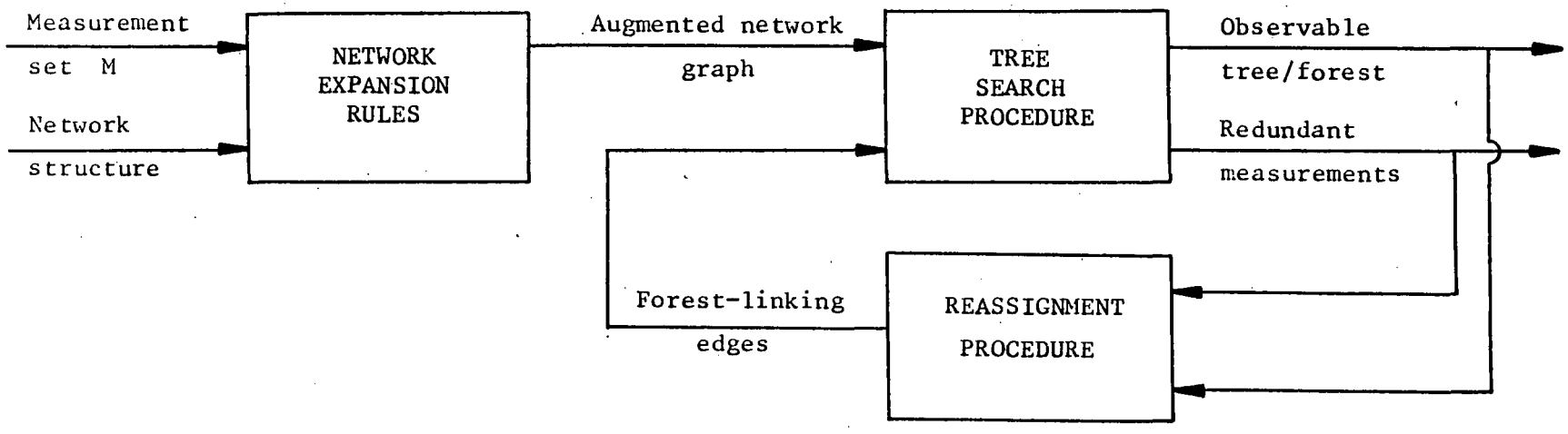


Figure 4.19 Basic steps in the Direct Tree Search Method

observability can be decided by seeking a spanning tree in the augmented network graph. In every vertex visited during construction of the tree, the edges which have a unique measurement assignment are used before those which are adjacent to the load measured nodes. The selection of an edge which can be assigned to a particular load measurement is critical in the sense that it can isolate some unmeasured nodes preventing the discovery of an observable spanning tree even if one exists. A tree search procedure must therefore be assisted by a routine which is able to correct such misassignments by checking if any of the redundant measurements can be used to expand the network tree.

The basic steps of our observability algorithm are schematically shown in Fig. 4.19 and are described in the next section.

4.6.2 An Algorithm to Identity an Observable Spanning Tree in the Augmented Network Graph

The observability algorithm presented in this section is based on the depth-first-search (DFS) procedure of Tarjan [261], modified to cater for constraints defined by the measurement assignment rules discussed in section 4.4. In order to build an observable tree over the largest portion of the network some rules of edge selection have been established:

i) Edges of the augmented network graph which can be associated with head, fixed-head-node flow or flow measurements are never assigned to load measurements even if such are available in their end-vertices;

ii) In every vertex, edges which have a unique measurement assignment are selected before those which can be associated only with a load measurement;

iii) If an unmeasured vertex is reached through an edge assigned to the load measurement the vertex is put on the stack for further consideration and we backtrack to the previous vertex.

Rules i) and ii) ensure that, if it is possible to reach some vertices via the edges which have a unique measurement assignment the algorithm will do so, and rule iii) minimises the number of edges connecting to a single unmeasured node, thus saving the load measurements for further assignments. However, there is no simple rule which can give guidance about how to assign load measurements [57]. In the proposed algorithm a strategy of devising an efficient reassignment policy instead of looking for an optimal assignment has been adopted. In addition to the vertex predecessor function an appropriate labelling of the edges of the observable tree has been devised.

The following definitions can now be introduced. A connected subgraph Z_1 of an observable tree T is called a tree zone if for every two vertices $v_i, v_j \epsilon Z_1$ a path between v_i and v_j contains only edges assigned to load measurements. If all the vertices of a tree zone are measured the zone is called active, otherwise the zone is called inactive. It follows immediately from the path property of observable trees that an inactive tree zone can have only one unmeasured vertex.

Once an observable tree of the augmented network graph has been found the observability question can be decided easily in the following cases:

a) If the observable tree spans the whole network graph then, according to Theorem 4.3, the network is topologically observable with respect to the measurement set concerned;

b) If the observable tree is not a spanning tree of the network graph and it does not contain any active tree zone, then the path property of trees of full rank ensures that the system is topologically unobservable.

The case in which the observable tree does not span the whole network graph but contains active tree zones requires investigation of the feasibility of amending the existing measurement assignment.

The measurement assignment rules adopted during construction of an observable tree of a network graph ensure that, on termination of a tree-building procedure, only unmeasured vertices remain unvisited and they are incident to inactive tree zones. Furthermore, if the network is topologically observable each inactive tree zone incident to an unvisited vertex is also incident to an active tree zone and can be reassigned such that the path property is preserved. Conversely, if the network is topologically unobservable the reassignment cannot be found.

The reassignment procedure checks whether it is possible to disconnect a path of tree edges between an unmeasured vertex of an inactive tree zone and a vertex of this zone which is incident to an univisited vertex. The constraint imposed on the algorithm is that the resulting forest must be linked into an observable tree by an edge connected to an active tree zone. This is illustrated in Fig. 4.20. An inactive tree zone 3-1-2 is incident both to an univisited node 4 and to an active tree zone formed of a single node 5. A path of edges 3-1-2 can be disconnected by removing an edge 1-3, and a resulting forest can be converted into observable spanning tree by adding edges 5-1 and 3-4 (Fig. 4.20 b). If the network graph did not include the edge 4-3 the network would be unobservable since the path 1-2 could not be disconnected.

The direct tree search observability algorithm is presented in more details in Fig. 4.21.

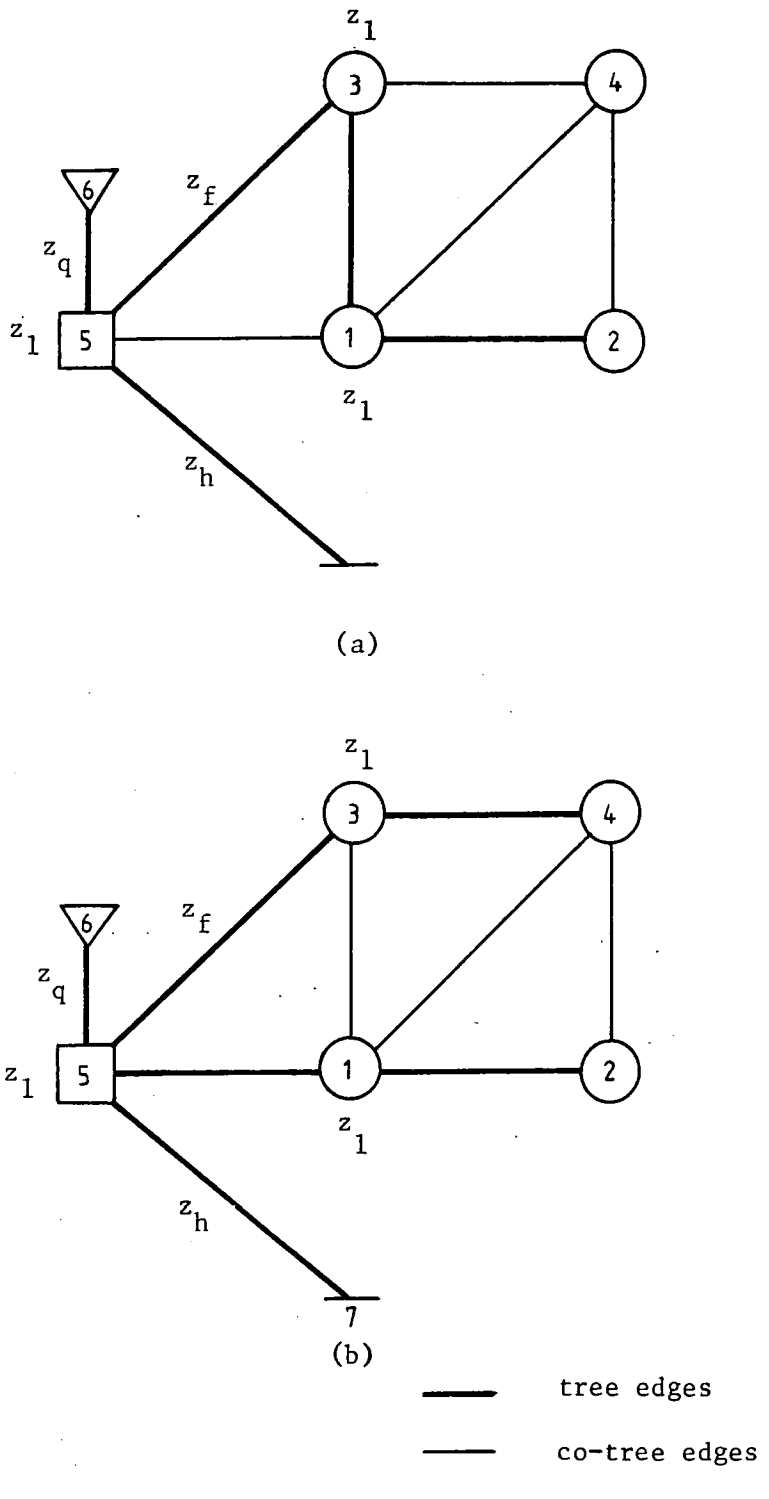


Figure 4.20 (a) Initial observable tree

(b) Observable spanning tree

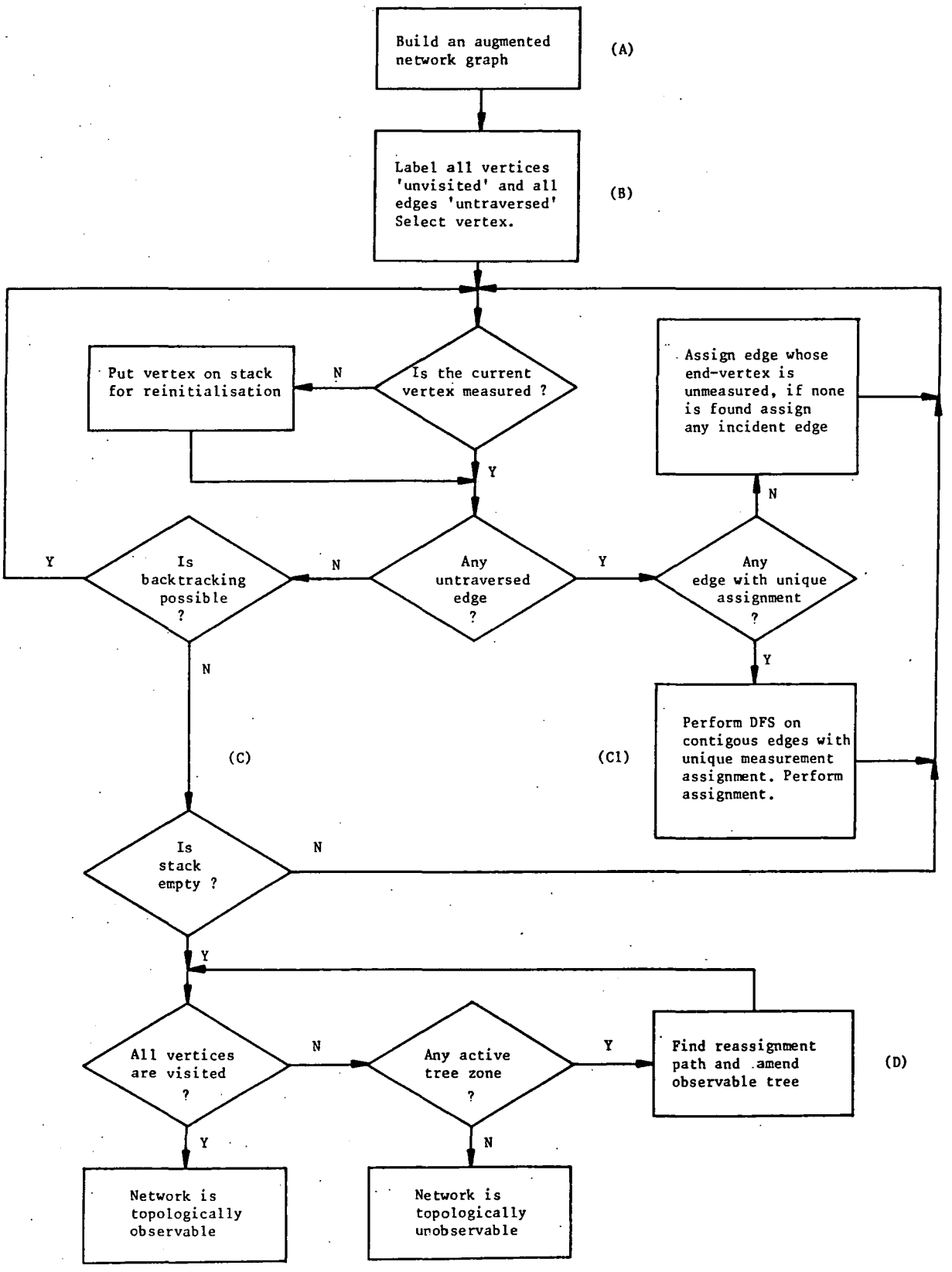


Figure 4.21 Flow chart for the Direct Tree Search Method

Computational complexity of this algorithm is equal to a maximal complexity of its main parts A, B, C(C1), and D allowing for the fact that D can be executed repeatedly. In parts A and B of the algorithm edges and vertices of an augmented network graph are processed a constant number of times, thus giving a computational complexity of l and n respectively (l is the number of edges and n is the number of vertices of the augmented network graph). Part C consists of a depth-first-search on a network graph. The fact that another depth-first-search (C1) on edges with unique measurement assignment is executed in its inner loop does not affect computational complexity of C since if an edge is marked as traversed in C1 it is not processed in C. The number of operations performed in C(C1) is therefore proportional to l. The upper bound on the length of the reassignment procedure D is equal to n (the number of tree edges) and the number of repetitions of D is limited to $l-n$ (the number of co-tree edges). Consequently, the maximum number of operations performed by the whole observability algorithm is proportional to $l \cdot n$. In most practical cases, however, the direct tree search algorithm shows a linear dependence on a problem size.

4.6.3 Examples

The method based on the direct tree search algorithm has been applied to the same examples used to test the

matching method in subsection 4.5.5. The results are shown in Fig. 4.22. Although only in case (b) does the observable spanning tree coincide with the one given by the matching method, it is easy to confirm by looking at the measurement sets in Fig. 4.10 that all the results are viable alternatives for observable spanning trees or forests.

The method was subsequently applied to the 34-node system, which was also used to evaluate the performance of the matching method. The six measurement systems which have been investigated were defined as in subsection 4.5.5. The results and computing times are summarized in Table 4.2.

The maximum computing time among all six cases, not taking into account the time for reading data, is 0.151 s. This is an indication that also this method is suitable for on-line applications. The computing time displayed in Table 4.2 does not show any obvious relationship to the number of measurements since the redundant measurements do not influence the process of building an observable tree. It does however depend on the numbering of vertices of the augmented network graph and on the location of the measurement points.

In cases where the measurement set renders the system unobservable, the algorithm returns the maximum observable forest. By looking at its components, it is possible to find out which measurements should be added to the measurement set.

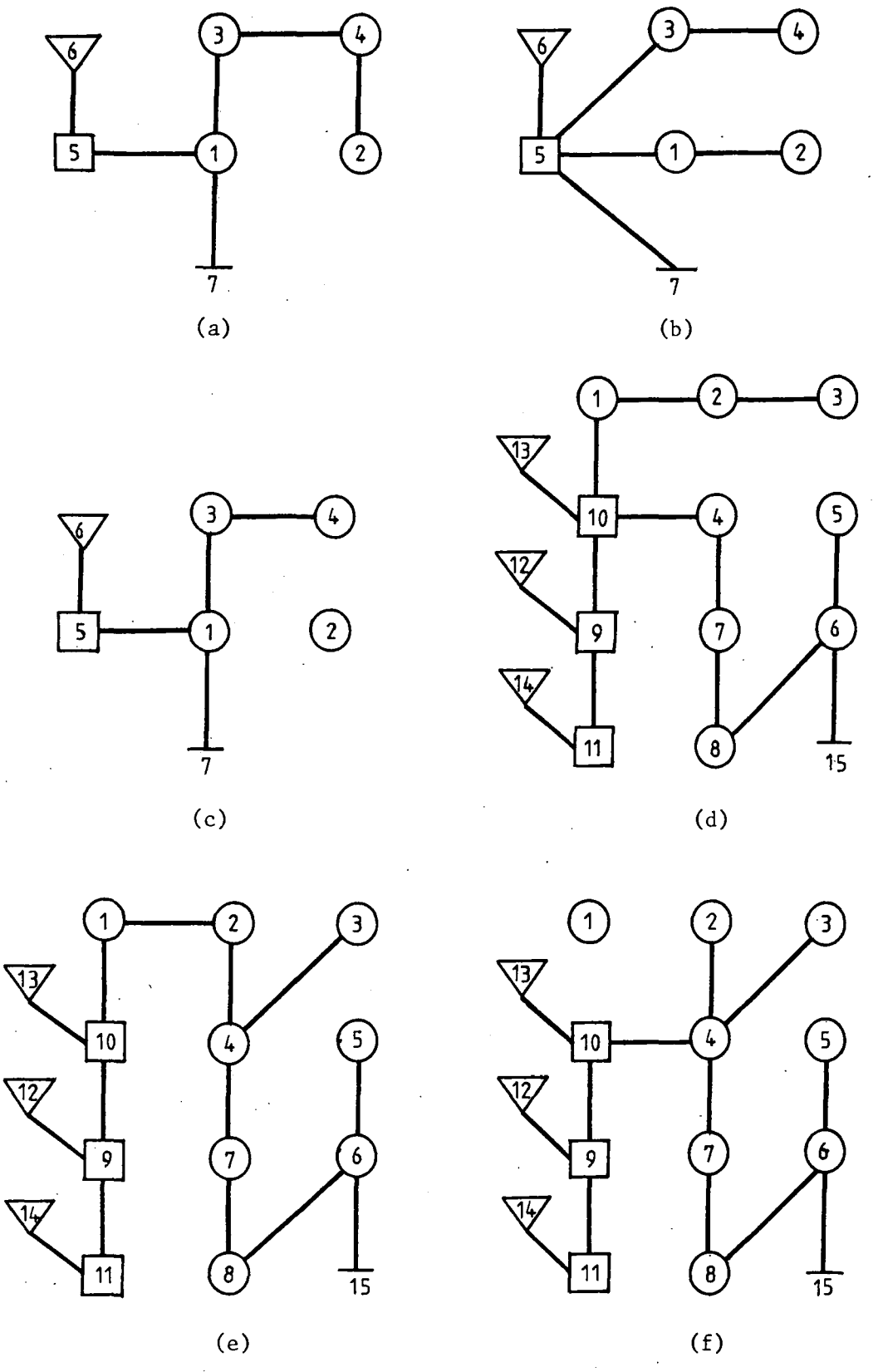


Figure 4.22 Observable trees/forests obtained through the Direct Tree Search Method

TABLE 4.2

Computational results for Direct Tree Search Method

Case / Observability *	1 OBS.	2 UNOBS.	3 UNOBS.	4 UNOBS.	5 OBS.	6 OBS.
Number of measurements	42	41	42	47	42	46
Comp. time for construction of augmented graph [s]	0.036	0.035	0.036	0.039	0.036	0.037
Comp. time for tree search [s]	0.046	0.049	0.046	0.045	0.046	0.046
Number of reassignments [s]	1	1	0	0	2	0
Comp. time for reassignments [s]	0.042	0.058	-	-	0.069	-
Total time [s]	0.124	0.142	0.082	0.084	0.151	0.083
Result	OBS.	UNOBS.	UNOBS.	UNOBS.	OBS.	OBS.

* OBS. = Observable
UNOBS. = Unobservable

** Sum for all reassignments

4.7 CONCLUDING REMARKS

Chapter 4 has been concerned with the investigation of the water system observability problem from the state estimation point of view. After introducing the various definitions of observability and presenting the conditions for topological observability, two methods have been proposed to solve the problem. Both methods use a new concept of an augmented network graph.

The first method formulates the observability problem using a measurement-to-edge bipartite graph and a maximum matching is sought in this graph. The edges which have a measurement assignment are subsequently used to build an observable spanning tree. If such a tree cannot be found directly a loop-breaking procedure, also based on the matching method, attempts to reassign measurements and link forest components. In the examples considered in subsection 4.5.5 the matching method proved to be very efficient especially for the systems with redundant measurements.

The second method proposed is based on the direct search for an observable spanning tree in the augmented network graph. During construction of the tree its edges are labelled so as to enable efficient reassignments in case the method identifies only an observable forest and at the same time some measurements remain unassigned. The method has been applied to the same systems and measurement configura-

tions used to test the matching method. In all cases, the results have been correct. Also, the computing times indicate that the method is feasible for on-line applications.

The direct tree search method has also been applied to investigate observability of an electrical power system [255]. An observability test on the IEEE-118 bus network gives a computation time of 0.300 s, confirming that in practice the algorithmic complexity of the technique grows linearily with problem size.

CHAPTER V

LEAKAGE REDUCTION BY OPTIMISED VALVE CONTROL

5.1 INTRODUCTION

The primary aim of a water distribution control is to maintain sufficient pressure to ensure that all demands, wherever and whenever they occur, can be met. The idealised requirement of system operation is to keep the pressure of the water in each individual node constant, relative to ground level. This is referred to as an optimal head profile. However, owing to the head/flow relationships in the network, the optimal head profile can only be maintained in a few nodes of the network while in the others the operational pressure remains higher. As the complexity of a distribution network grows, the task of achieving an optimum pressure becomes more and more difficult and the average overpressure tends to increase. This in turn results in an increased energy cost, increased volume of distributed leakages and higher risk of major bursts, particularly during the night period when the pressure additionally rises due to decrease of consumer demand. In complex networks the

volume of leakages can amount to approximately 25-30% of the total production and consequently represents the main potential for improvement of water distribution system economy.

Minimisation of the overpressures is possible by remote control of valves installed on the pipe network in accordance with the changing demand pattern. However, computation of the optimum valve settings is usually a relatively difficult task due to the high dimensionality of the optimisation problem and the nonlinearity of the network model. The application of conventional optimisation methods is consequently not realistic in view of the computational resources needed and the requirement for real time control of the water distribution system.

In this chapter, the optimisation problem is expressed in a form which enables application of linear programming optimisation techniques and in particular the sparse revised Simplex method is shown to be advantageous. This approach makes it possible to take full advantage of the sparse structure of the problem and to achieve low solution times. A highly sparse factorisation of the basis matrix is maintained using an algorithm proposed by Reid [210]. Studies on networks of different sizes give rise to an estimated computation time, for a network of 100 nodes with 10 control valves, of about 20 sec using a Perkin Elmer 3220 minicomputer. An application of the computed control policy

is shown to result in a substantial reduction of the distributed leakages in the system.

5.2 SUPPRESSION OF DISTRIBUTED LEAKAGES

5.2.1 Formulation of the problem

The task of minimisation of the volume of leakages can be seen as a minimisation of discrepancies between a current and an optimal head profile in the network subject to operational limits on the valve controls v_k.

$$\min_{v_k} \sum_i | h_i - h_i^o | \qquad (5.1)$$

$$\text{s.t.} \quad 0 < v_k < v_k^{max}$$

where i = 1, ..., N - is the number of network nodes

k = 1, ..., K - is the number of control valves

It is apparent however that because of the limited number of control valves ($K << N$), not all heads h_i can be controlled independently, therefore it is practical to consider only the subset of network nodes which impose the most severe requirements for the water supply system. These are usually the nodes which have locally the highest ground elevation or the biggest load. In effect, the optimisation problem (5.1) can be expressed as

$$\min_{v_k} \sum_j | h_j - h_j^o | \qquad (5.2)$$

$$\text{s.t.} \quad 0 < v_k < v_k^{max}$$

where j=1, ..., R is the number of reference nodes in the network.

To be able to perform the optimisation (5.2) it is necessary to find a functional relationship between nodal heads h_j and valve controls v_k. These can be expressed in many different forms depending on the chosen set of state variables. In the case of a water distribution system it is convenient to select the heads in all network nodes and inflows in fixed head nodes as state variables in order to enhance the preservation of sparsity in the mass balance equations

$$\sum_{j \in M_i} f_{ij}(\underline{h}) = b_i \qquad i = 1, \ldots, L \qquad (5.3)$$

$$\sum_{j \in M_i} f_{ij}(\underline{h}) + u_i = b_i \qquad i = L+1, \ldots, N \qquad (5.4)$$

$$u_i = u_i^o \qquad i = L+1, \ldots, N \qquad (5.5)$$

where $\underline{h} = [h_1, \ldots, h_N]^T$ is a vector of nodal pressures

u_i - is the inflow in the fixed-head node

f_{ij} - is a head/flow function of i-j network element

M_i - is a set of nodes incident to node i

b_i - is a nodal balance.

For a network containing control valves the state vector must be extended by the addition of variables representing valve openings. The head/flow function f_{ij} in equations (5.3) and (5.4) is then replaced by $f_{ij}(\underline{h},\underline{v})$ where $\underline{v} = [v_1, \ldots, v_k]^T$ and v_k is the k-th valve control. However v_k is the control variable and its value is not known in advance. The information about the value of v_k can only be expressed approximately in the following way,

$$v_k + \omega_k^v = v_k^o \qquad k = 1, \ldots, K \qquad (5.6)$$

$$v_k + \nu_k = v_k^{max} \qquad k = 1, \ldots, K \qquad (5.7)$$

where equation (5.6) represents uncertainity about a current approximation of valve control v_k^o, and equation (5.7) represents an operational limit of valve control ($\nu_k > 0$).
The problem described by equations (5.2) - (5.7) could be solved by a predictor-corrector type of procedure, however, this would involve a full load flow solution followed by a sensitivity analysis at every stage, implying unnecessarily high computational effort. In the present paper a formulation is proposed which allows for computation of optimal valve controls in a single stage. For this purpose equations (5.3), (5.4), (5.5) and (5.7) are complemented by variables

$$\omega_i^L : i = 1, \ldots, L,$$
$$\omega_i^L : i = L + 1, \ldots, N,$$

ω_i^u : i = 1, ..., N-L

ω_i^ν : i = 1, ..., K

respectively, in a similar form to equation (5.6) except that the values of these additional variables are kept zero. Additionally, equations for the head in the network reference nodes are written as

$$h_j + \omega_j^h = h_j^o \qquad j = 1, \ldots, R \qquad (5.8)$$

where ω_j^h is a discrepancy between current and optimal head profile.

Using the notation introduced above the optimisation problem (5.2) can now be written

$$\min_{\underline{x}} \underline{w}^T . \underline{\omega} \qquad (5.9)$$

s.t. $\underline{g}(\underline{x}) + \underline{\omega} = \underline{z}$

where $\underline{g}(.)$ is a nonlinear functional of $\underline{x}$.

$$\underline{x} = [h_1, \ldots, h_N, u_1, \ldots, u_{N-L}, v_1, \ldots, v_K, \nu_1, \ldots, \nu_K]^T$$

$$\underline{w} = [w_1^L, \ldots, w_N^L, w_1^u, \ldots, w_{N-L}^u, w_1^v, \ldots, w_K^v, w_1^\omega, \ldots w_K^\omega, w_1^h, \ldots, w_K^h]^T$$

$$\underline{\omega} = [\omega_1^L, \ldots, \omega_N^L, \omega_1^u, \ldots, \omega_{N-L}^u, \omega_1^v, \ldots, \omega_K^v, \omega_1^\nu, \ldots, \omega_K^\nu, \omega_1^h, \ldots, \omega_K^h]^T$$

$$\underline{z} = [d_1, \ldots, d_N, u_1, \ldots, u_{N-L}, v_1^o, \ldots, v_K^o, v_1^{max}, \ldots, v_K^{max}, h_1^o, \ldots, h_K^o]^T$$

The values of the elements of the weighting vector $\underline{w}$ are chosen in such a way as to reflect the requirements of the vector $\underline{\omega}$. Since the mass balance equations and the equations representing operational limits of the valves express physical relationships, the corresponding weights are high which has an effect of zeroing ω_i^L, ω_i^u and ω_i^v. Conversely, the weights corresponding to the equations for valve opening w_i^v are set to zero since the cost of the valve control is neglected in (5.2). The equations representing a discrepancy between the current and the optimal head profile are biased with some small positive weights and effectively are the only ones which contribute to the nonzero value of the performance index.

5.2.2 Solution via the Linear Programming Approach

To cope with the nonlinearity of the equations in (5.9) a method of iterative linearisation based on the Newton-Raphson process has been used. This can be summarised as follows

1) Expand $\underline{g}(\underline{x})$ to first order using a Taylor series about an initial guess of the state vector $\underline{x}^o$,

$$\underline{g}(\underline{x}) - \underline{g}(\underline{x}_k) = \underline{J} \cdot \Delta\underline{x} + \underline{\omega} \qquad (5.10)$$

where $\underline{J}$ is PxQ. Jacobian matrix

P is a number of equations $P = 2N-L+2K+R$

Q is a number of variables $Q = 2N-L+2K$

2) Solve the optimisation problem for the linearised constraints,

$$\min_{\Delta \underline{x}} \underline{w}^T \cdot \underline{\omega} \qquad (5.11)$$

$$\text{s.t. } \Delta \underline{z} = \underline{J}\, \Delta \underline{x} + \underline{\omega}$$

where $\Delta \underline{z} = \underline{g}(\underline{x}) - \underline{g}(\underline{x}^o)$

3) Update the estimate of the state vector x_k

$$\underline{x}^o_{k+1} = \underline{x}^o_k + \Delta \underline{x} \qquad (5.12)$$

4) If $\Delta \underline{x}$ satisfies a convergence test then stop, otherwise repeat iteration from 1^o.

The estimate of the state vector $\underline{x}_k$ computed according to the Newton-Raphson process generally converges even if the initial guess $\underline{x}_1$ is not good. In practice, the initial guess would be the result of the most recent state estimation, and convergence would be achieved in a few steps. The structure of the optimisation problem at stage 2) of the Newton-Raphson process facilitates an efficient solution using the sparse revised Simplex method. To satisfy requirements for nonnegativity of the variables and to allow a decrease of the state vector in 3) the following substitutions are introduced

$$\underline{\omega} = \underline{r} - \underline{s} \qquad (5.13)$$

where $\underline{r} = [r_1, \ldots, r_p]^T$, $\underline{s} = [s_1, \ldots, s_p]^T$

$r_i > 0$, $s_i > 0$, $r_i + s_i = \omega_i$, $r_i \cdot s_i = 0$, $i = 1, \ldots, P$

and

$$\Delta\underline{x}' = \Delta\underline{x} + \underline{d} \qquad (5.14)$$

$$\Delta\underline{z}' = \Delta\underline{z} + \underline{J}.\underline{d} \qquad (5.15)$$

where $\underline{d} = [d_1, \ldots, d_Q]^T$, and d_i is the maximum decrease of the state variable x_i in one iteration, thus

$$\Delta\underline{x}' = [\Delta x_1 + d_1, \ldots, \Delta x_Q + d_Q]^T$$

Now the linear programme can be written in a standard form,

$$\min_{\Delta\underline{x}'} \ \underline{w}^T \cdot (\underline{r} + \underline{s}) \qquad (5.16a)$$

s.t.

$$\Delta\underline{z}' = [J : I : -I] \cdot \begin{bmatrix} \Delta\underline{x}' \\ \underline{r} \\ \underline{s} \end{bmatrix} \qquad (5.16b)$$

where I is a unit matrix.

As the dimension of the basis in the primal Simplex method is determined by the number of equations (P) the introduction of variables $\underline{r}$ and $\underline{s}$ does not result in any increase of dimensionality of the problem. Also the computer memory requirements remain unchanged as the unit

matrices are incorporated implicitly in the Simplex algorithm.

In order to take full advantage of sparsity in the linear programme, the 'elimination' form of basis factorisation has been used. Reid [240] has proposed an algorithm for the elimination form which also applies a series of row and column permutations to give enhanced sparsity retention. An implementation of this basis handling mechanism is widely available as a routine LA05A in the Harwell subroutine library.

5.3 NUMERICAL RESULTS

The performance of the optimal valve control algorithm has been tested on several different size networks. The detailed results of a study of the effect of incorporation of the control valves and their operation are presented for the 25-node network shown in Figure 5.1. Parameters of the pipes are given in Table 5.1. The network contains 3 pumping stations which are controlled on an on/off basis but can accommodate ±15% variation of a flow without changing the water supply pressure. Since the variation of the load during the 24-hour period is over 200% (Figure 5.2) it is necessary to combine discrete and continuous control of the pumps. The pumping schedule, presented in Figure 5.3, has been devised so as to fully satisfy consumers demands subject to constraints on the magnitude of the

TABLE 5.1

Parameters of the 25-node system

LINE	LENGTH [m]	DIAMETER [m]	HAZEN-WILLIAMS COEFFICIENT
23- 1	606.	0.457	110
23-24	454.	0.457	110
24-14	2782.	0.229	105
25-14	304.	0.381	135
10-24	3383.	0.305	100
13-24	1767.	0.475	110
14-13	1014.	0.381	135
16-25	1097.	0.381	6
2- 1	1930.	0.457	110
3- 2	5150.	0.305	10
12-13	762.	0.457	110
15-16	914.	0.229	125
17-16	822.	0.305	140
18-17	411.	0.152	100
20-18	701.	0.229	110
19-17	1072.	0.229	135
20-19	864.	0.152	90
21-20	711.	0.152	90
21-15	832.	0.152	90
22-15	2334.	0.152	100
12-15	1996.	0.229	95
11-12	777.	0.229	90
10-11	542.	0.229	90
8-12	1600.	0.457	110
8-10	249.	0.305	105
9- 8	443.	0.229	90
6- 8	743.	0.381	110
22- 8	931.	0.229	125
22-21	2689.	0.152	100
4- 3	326.	0.152	100
5- 4	844.	0.229	110
6- 3	1274.	0.152	100
5- 6	1115.	0.229	90
7- 6	615.	0.381	110
5-22	1406.	0.152	100
5- 7	500.	0.381	110
6- 9	300.	0.229	90

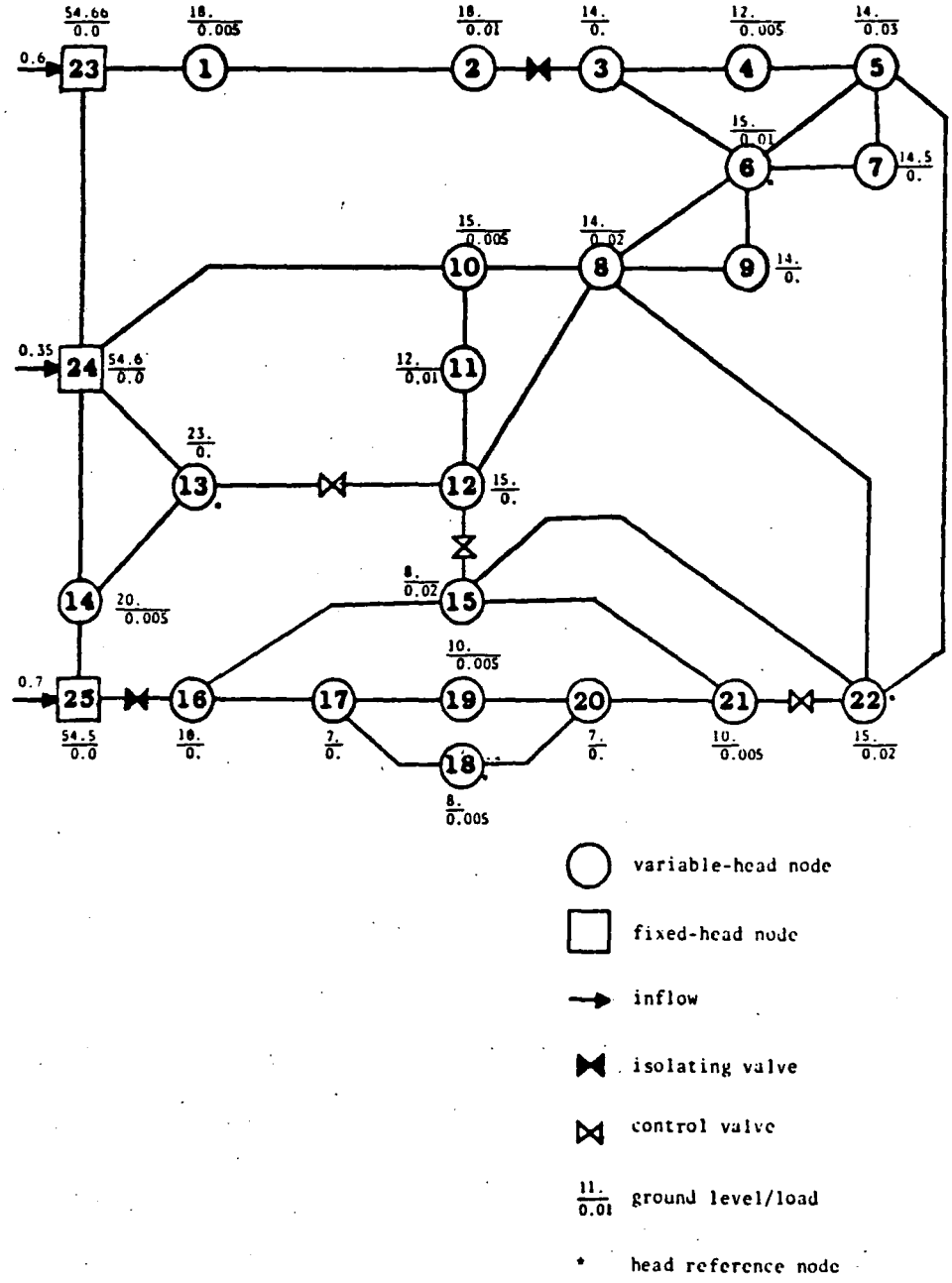


Figure 5.1 25-node system

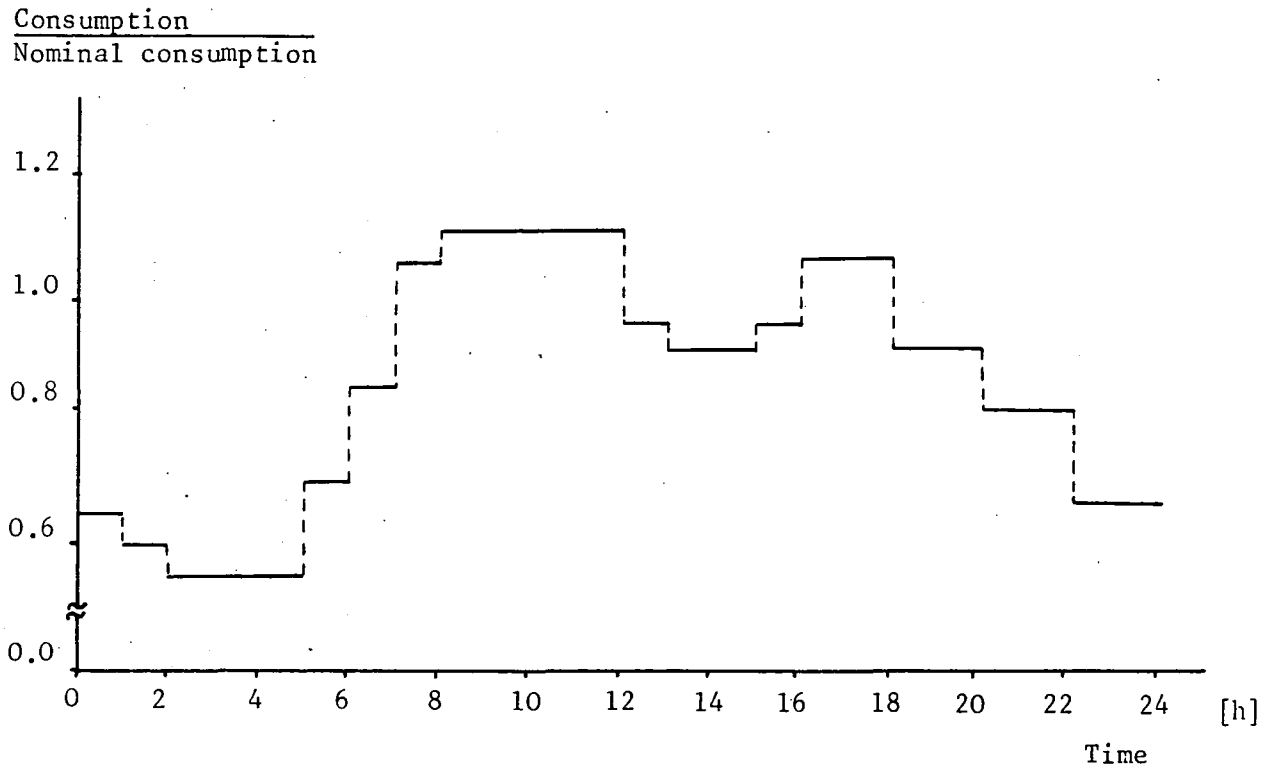


Figure 5.2 Normalised daily load pattern

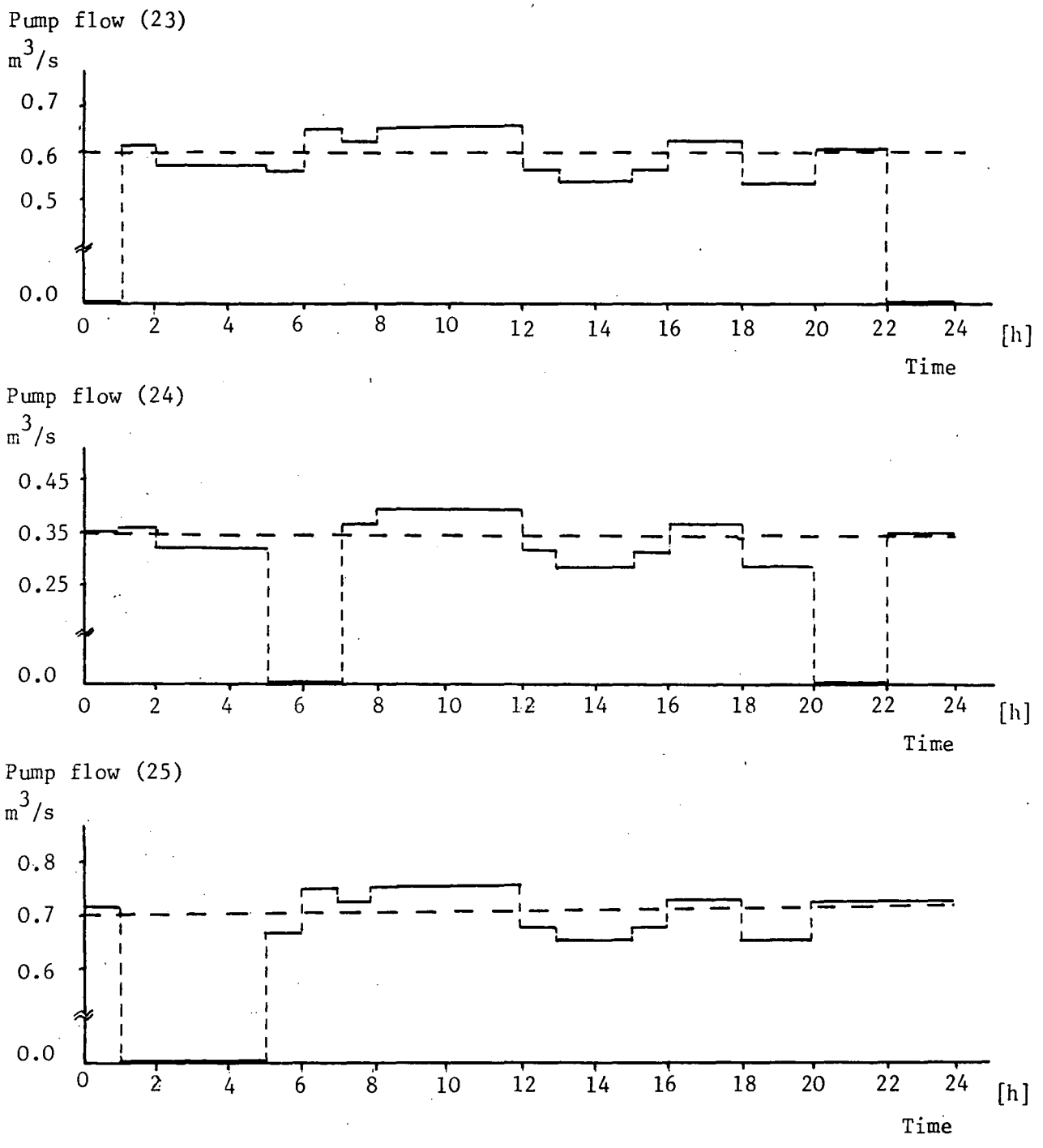


Figure 5.3 Pumping schedule

change of flow and the frequency of the on/off control of each individual pump. It is assumed that the hourly load in each of the 15 consumer supply nodes follows the pattern given in Figure 5.2. The nominal consumption corresponding to '1' in Figure 5.2 is given alongside the ground level of the node on the network diagram. In order to prevent an excessive service pressure in the nodes which have a low ground level, two isolating valves (constant throttling) and three control valves are used in pipes 2-3, 25-16 and 12-13, 13-15, 21-22 respectively. The current service pressure is measured at nodes 6, 13, 18 and 22 which have locally the highest ground level.

The volume of leakages V has been evaluated for each head profile based on the empirical relationship

$$V = C \cdot \sum_i (l_i \, h_{Ai}^{1.18}) \qquad i = 1, \ldots, S \qquad (5.17)$$

where S : is the number of pipes

C : is a constant depending on the network

l_i : is the length of the i-th pipe

h_{Ai} : is the average service pressure along the i-th pipe

Consequently for the optimal head profile (30mAq) the corresponding volume of leakages is $V_{30} = C.30^{1.18} \sum_i l_i$ and the water loss index can be introduced as

$$W = \frac{V - V_{30}}{V_{30}} \qquad (5.18)$$

Three schemes of operation of the network have been analysed. In the first case, the network with fully open control valves is considered. The head profile achieved during the operation of the network is identical to that for the network with no control valves. The limit of 30mAq for the service pressure is exceeded for all loads since the network has to maintain the capability of supplying some emergency loads. Decrease of the consumer load during the night period additionally increases overpressure in the network which is reflected by the high values of the water loss index given with a dotted line in Figure 5.5. In the second case the control valves are throttled in order to achieve the optimal head profile in the selected reference nodes during the highest daily consumption. This corresponds to the situation where the network has manually controlled valves which, except in emergency, have constant openings. The area between the dashed and dotted line in Figure 5.5 indicates 3.5% reduction of the total leak volume as a result of implementation of such a control policy.

In the third case the optimal valve controls, shown in Figure 5.4, have been applied. The discrepancy between the current and the optimal head profile is minimised for the whole range of the consumer loads giving an almost constant

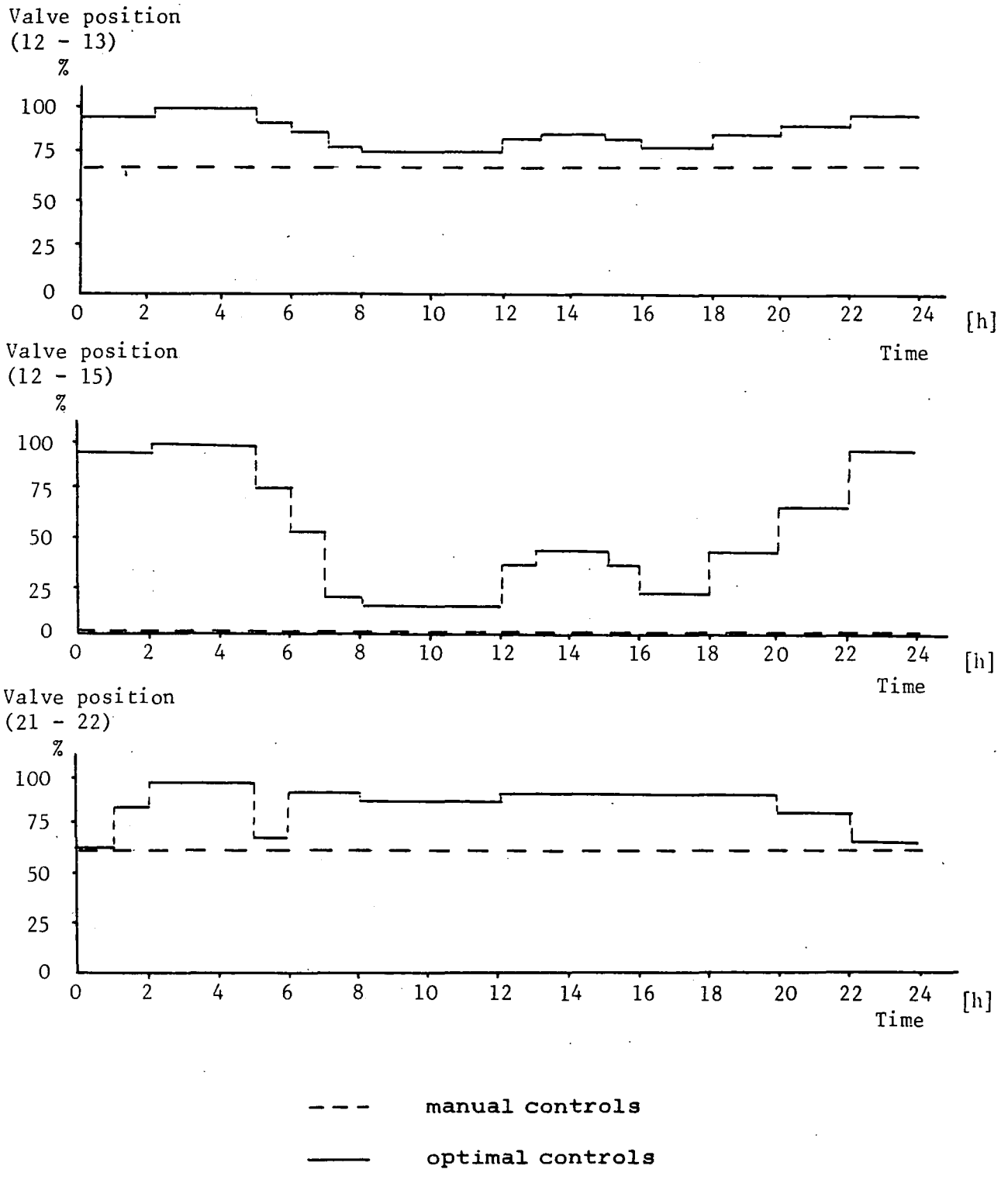


Figure 5.4 Valve control strategy for minimal leakage

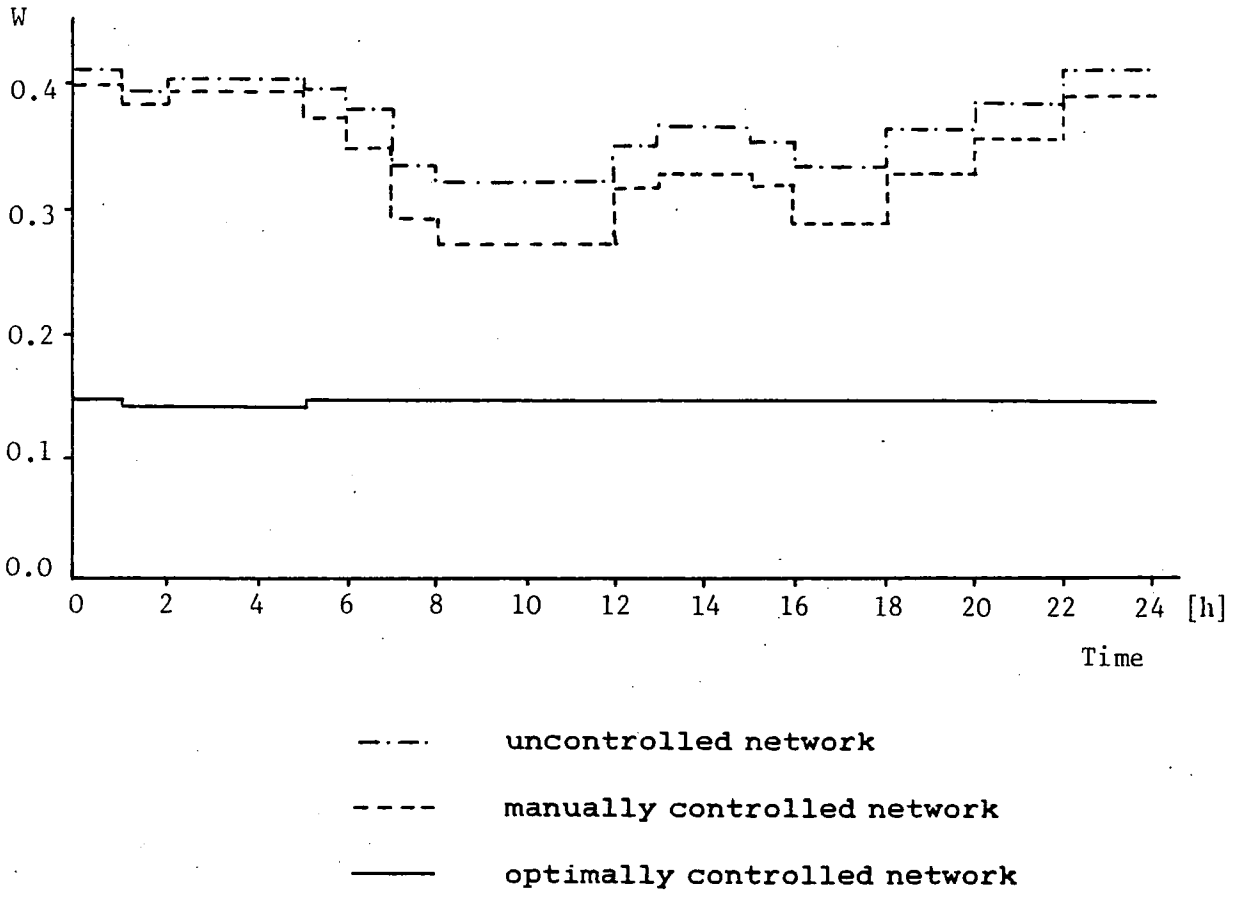


Figure 5.5 Hourly variation of the water loss index

TABLE 5.2

Variation of computation requirements with network size

No. of var.-head nodes	13	13	13	22	32	100
No. of fixed-head nodes	3	3	3	3	3	5
No. of control valves	1	3	6	3	3	10
No. of check points	4	4	4	4	4	10
No. of state variables	21	25	31	34	44	130
No. of equations	25	29	35	38	48	140
No. of N-R iterations	5-7	5-7	5-7	5-7	5-7	5-7*
Computation time for N-R iter. [s]	0.530	0.690	0.715	0.760	1.078	3.57*

* estimated value

value of the water loss index. Application of the optimal control policy results in 19.2% reduction of distributed leakages as compared with the network using manually controlled valves. The nonzero value of the water loss index indicates that by increasing the number of control valves a further reduction of leakages is possible, however the incremental saving achieved by adding one control valve to the network may be marginal.

The algorithm for computation of the optimal valve controls has been coded in FORTRAN 77 and implemented on a Perkin Elmer 3220 minicomputer with 32-bit word length and floating-point arithmetic. Comparisons of the execution time have been made for 16, 25 and 35-node networks having 1 to 6 control valves. The results are presented in Table 5.2 and include an estimated computation time for a 100-node network with 10 control valves.

5.4 CONCLUDING REMARKS

In this chapter a new algorithm for computation of the optimal valve controls in order to reduce distributed leakages in water supply network has been presented. Since the volume of water losses due to bursts of mains and distributed leakages can amount to approximately 30% of the total production, the on-line implementation of the leakage suppression algorithm together with an efficient network monitoring scheme has a strong economical motivation. By

controlling the network pressure profile it is also possible to reduce the risk of pipe ruptures, thus saving on the maintanance cost of the water supply system.

Simulation results, reported in section 5.3, indicate that it is possible to achieve a 20% reduction of the volume of leakages which amounts to 6% savings of the total water production cost. The algorithm proves to be computationally efficient, which makes it applicable to on-line operation using relatively inexpensive hardware.

The proposed method may also be used at the network planning stage to evaluate the economics of the installation of additional control valves in the network.

CHAPTER VI

SOFTWARE PACKAGE

6.1. INTRODUCTION

This chapter is concerned with description of the on-line software package for real-time monitoring of a water supply network. Taking into account the size of the FORTRAN code (approx. 16500 lines) and the complexity of interactions between program segments, a high level viewpoint has been adopted in describing the software. The lower level block-diagrams, explaining organisation of individual programs, are not included here since they can be easily obtained directly from the FORTRAN code by monitoring the CALL statements in their order of execution. Figures 6.1 and 6.9 essentially highlight a general concept of the organisation of the package by describing the flow of information and means of coordination of the simultaneously executed tasks.

There are three main groups of programs in the package (Figure 6.1). The programs of the first group simulate the behaviour of the real network and provide measurement information which in practice is retrieved using

some telemetry system. This data is effectively the only source of information for the second group of programs monitoring the network.

A major role of the monitoring programs is to supply information about the system state both for the human operator and control algorithms. Since the telemetered data is being updated without the intervention of a human intermediary the monitoring programs are said to be on-line to the process.

After checking topological observability of the system, with respect to the current set of valid measurements, the estimates of the state vector are calculated. This is followed by identification of bad data points which were not found during the pre-processing stage. Depending on the state estimation algorithm employed, the monitoring procedure involves either an iterative elimination of bad data from the set of valid measurements and recomputation of the state vector, or it simply marks erroneous measurements having rejected them in the course of the estimation. The results obtained with the monitoring programs are made available to the operator in the form of a print-out, graphical display and data file which is also used by control algorithms.

The third group of programs closes the control loop by devising and implementing control action. The flow of

information between the programs implies that the algorithmically calculated controls are off-line to the process since they are implemented by a human operator. Such a structure is natural at the initial stage of the computerised monitoring and control of a water network. However, it must be emphasized that the computer assisted control can be easily converted into a full on-line control scheme since the system is monitored on-line.

In order to achieve a degree of flexibility a highly modular structure of the software package has been adopted. Each task communicates with others via task common blocks and is therefore insensitive to the way in which the input data is being calculated. In particular, it is transparent for the monitoring programs whether the telemetered data is generated by a simulator or supplied by a telemetry computer. It is also possible, within this structure, to test alternative algorithms without affecting the integrity of the software.

The following sections of this chapter describe programs of the package in terms of their interactions with task common blocks and give software details.

Task sizes and execution times reported in this chapter refer to the 34-node water distribution network.

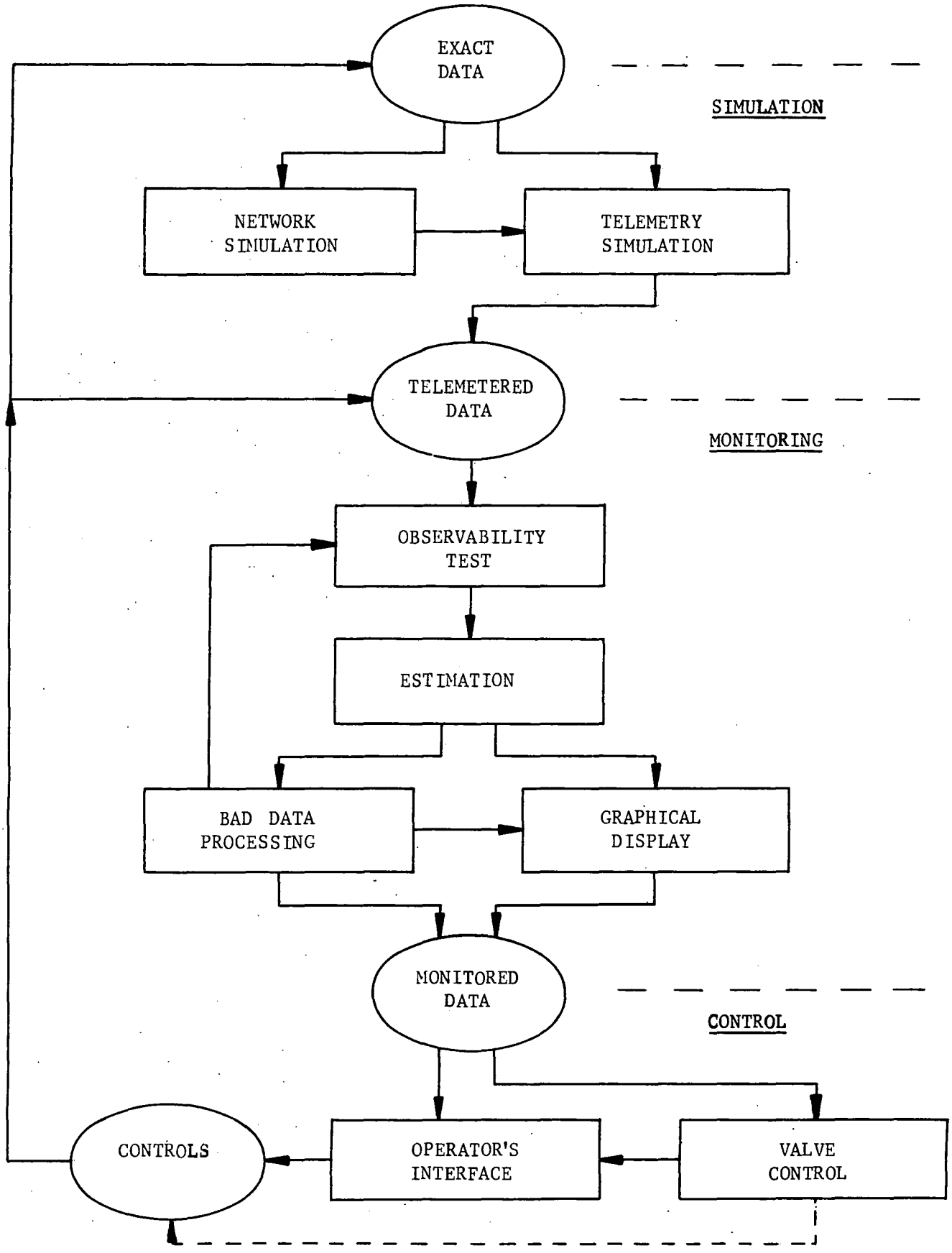


Figure 6.1 Software package

6.2 NETWORK SIMULATION PROGRAM

The water network simulation program (Figure 6.2) provides a facility to carry out on-line monitoring studies without recourse to a real-life telemetry system. The input data for the network simulator represents exact information about the system and, as such, are not available to the monitoring programs. They can only be modified by the control action of the operator.

The simulator calculates an exact state vector, by applying a Newton-Raphson iterative procedure to the square set of nonlinear mass-balance equations, and passes it to the telemetry simulation program which calculates the values of the measurements.

Software Details (SYSSYM)

Size - 1570 lines of FORTRAN 77 code
96 k bytes

Speed - 0.8 s/iter (PE 3220)

6.3 TELEMETRY SIMULATION PROGRAM

Using an exact state vector, supplied by the network simulation program, and information about the meter positioning the telemetry simulation program (Figure 6.3) calculates the exact values of the measurements. In order to

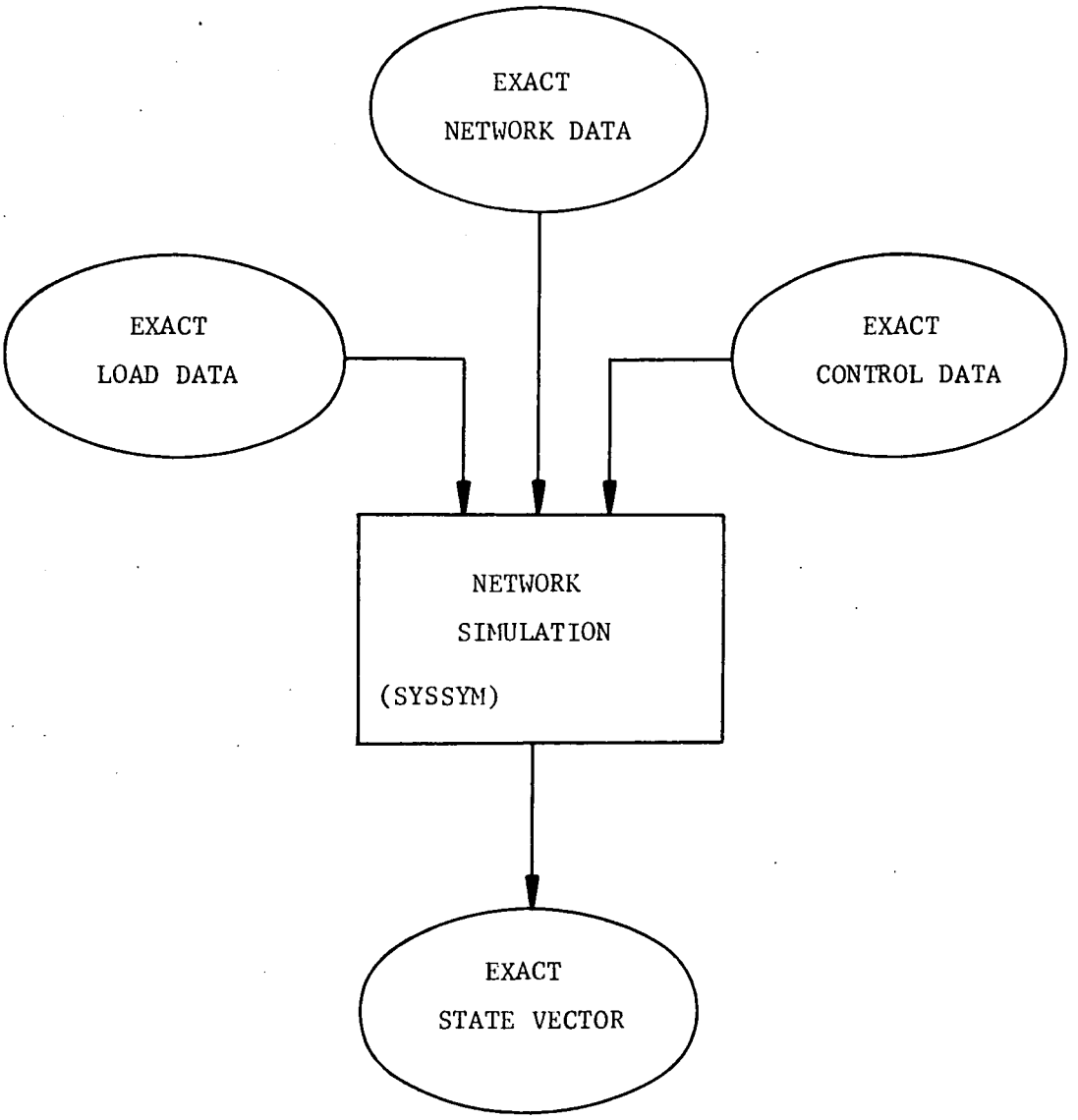


Figure 6.2 Interaction of network simulation program with task common blocks

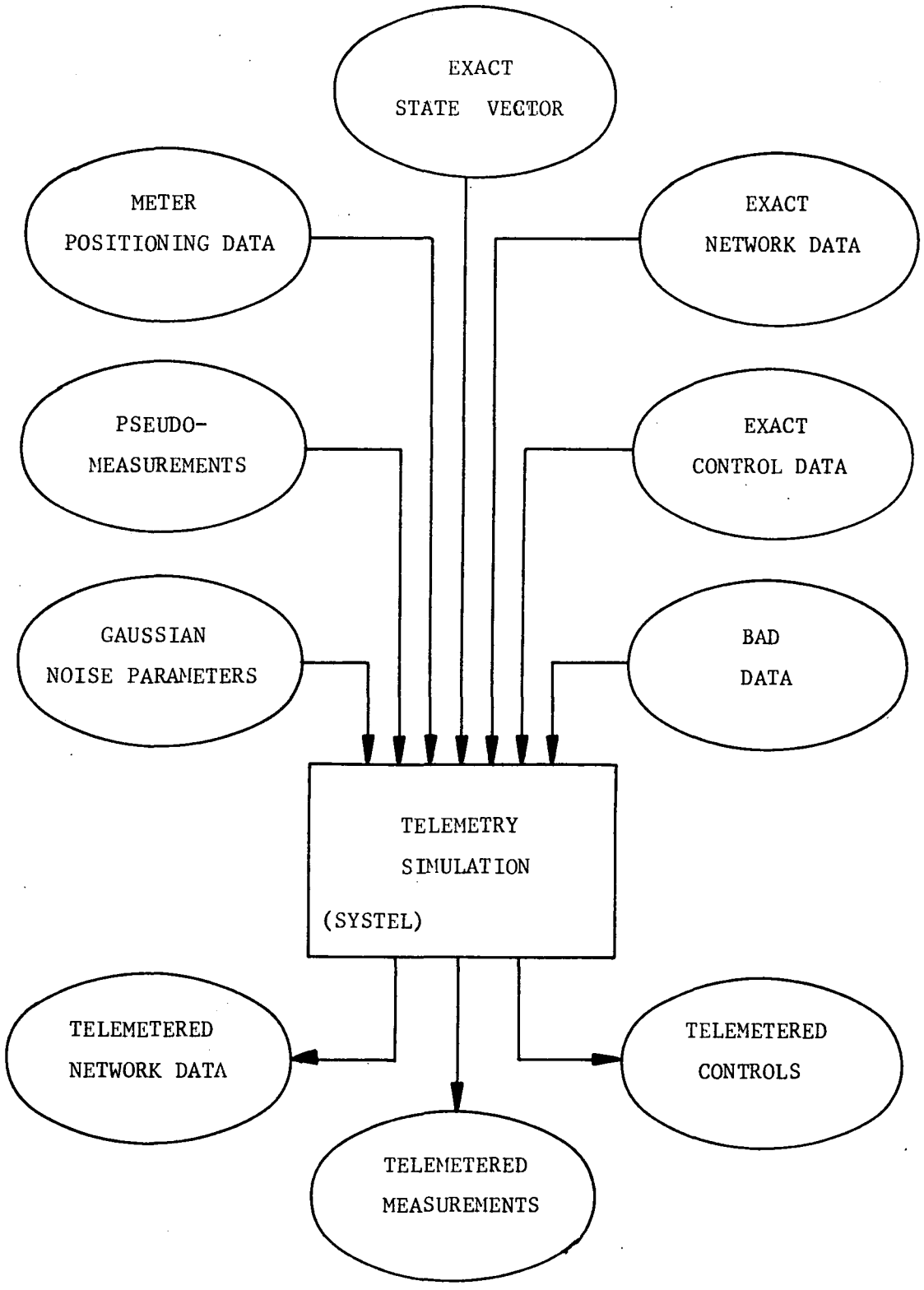


Figure 6.3 Interaction of telemetry simulation program with task common blocks

obtain a realistic set of telemeasurements pseudorandom measurement noise is then superimposed on the meter readings. The program also enables the simulation of telemetry or instrumentation malfunction by making provision for the corruption of the measurement set by gross measurement errors and/or topological errors.

Software details (SYSTEL)

Size - 570 lines of FORTRAN 77 code
38 k bytes

Speed - ~0.3 s (PE 3220)

6.4 OBSERVABILITY PROGRAMS

The observability routine (Figure 6.4) checks whether the current set of measurement points can provide sufficient information to allow the computation of the state estimates. If the system is found unobservable the program generates pseudo-measurements which restore the observability. The program is also used to determine the detectability of bad data points after suppression of some measurements in the intial telemetered data.

Two different programs to assess the topological observability of the network have been developed. The first program (OBSMATCH) uses the concept of maximum matching in

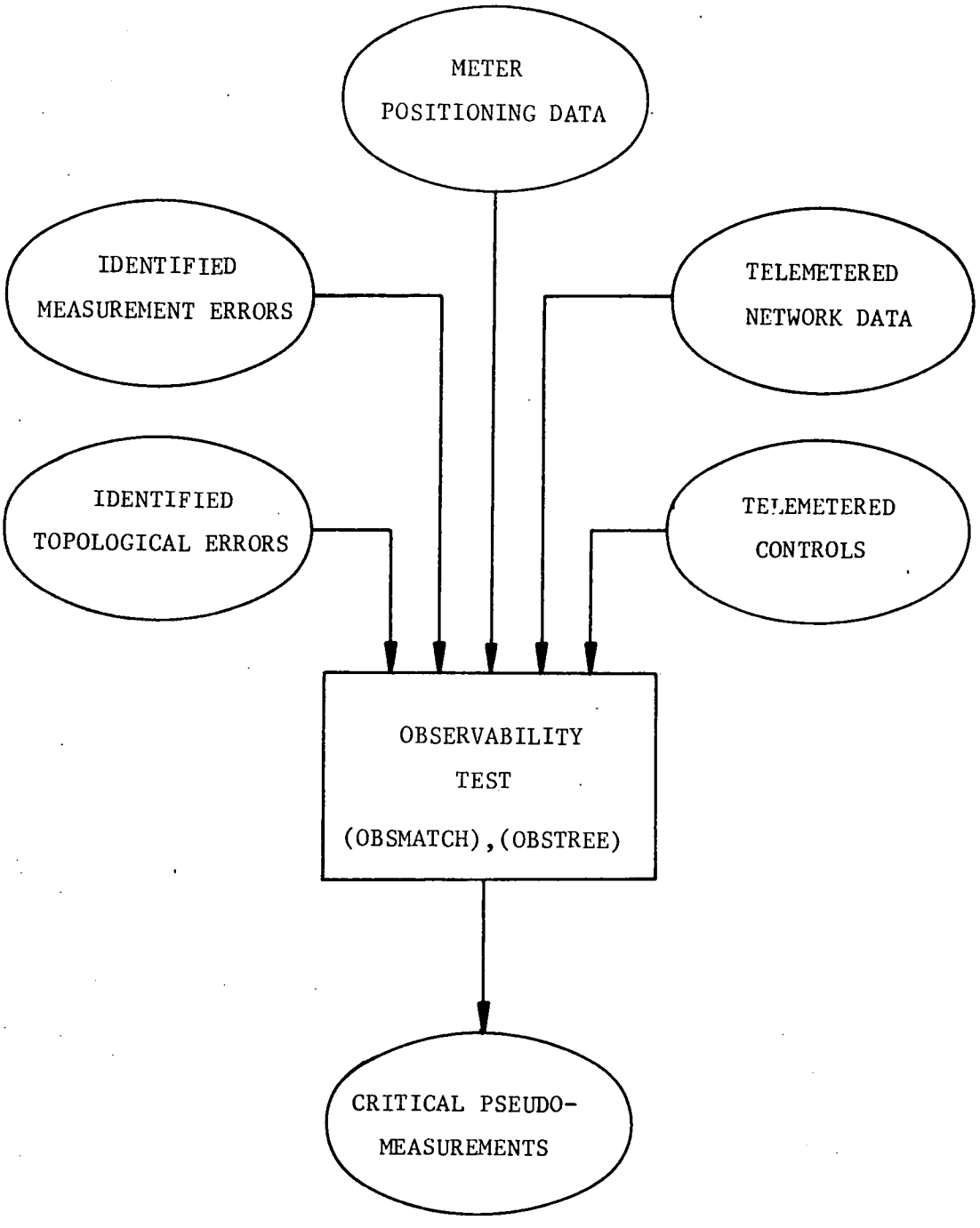


Figure 6.4 Interaction of observability programs with task common blocks

the bipartite measurement-to-branch graph and the second one (OBSTREE) is based on a direct search for an observable spanning tree of the network.

Software details (OBSMATCH)

Size - 820 lines of FORTRAN 77 code
37.75 k bytes

Speed - ~0.1 s (PE 3220)

Software details (OBSTREE)

Size - 1050 lines of FORTRAN 77 code
43.25 k bytes

Speed - ~0.1 s (PE 3220)

6.5 STATE ESTIMATION PROGRAMS

The state estimation program (Figure 6.5) plays a key role in the network monitoring package. It processes raw telemetered data, augmented by pseudo-measurements which are generated by the observability routine, and calculates an estimate of the state vector. The output of the state estimator also includes estimates of the measurement residuals, thus enabling detection and identification of bad data points.

Using the same structure of task common blocks two

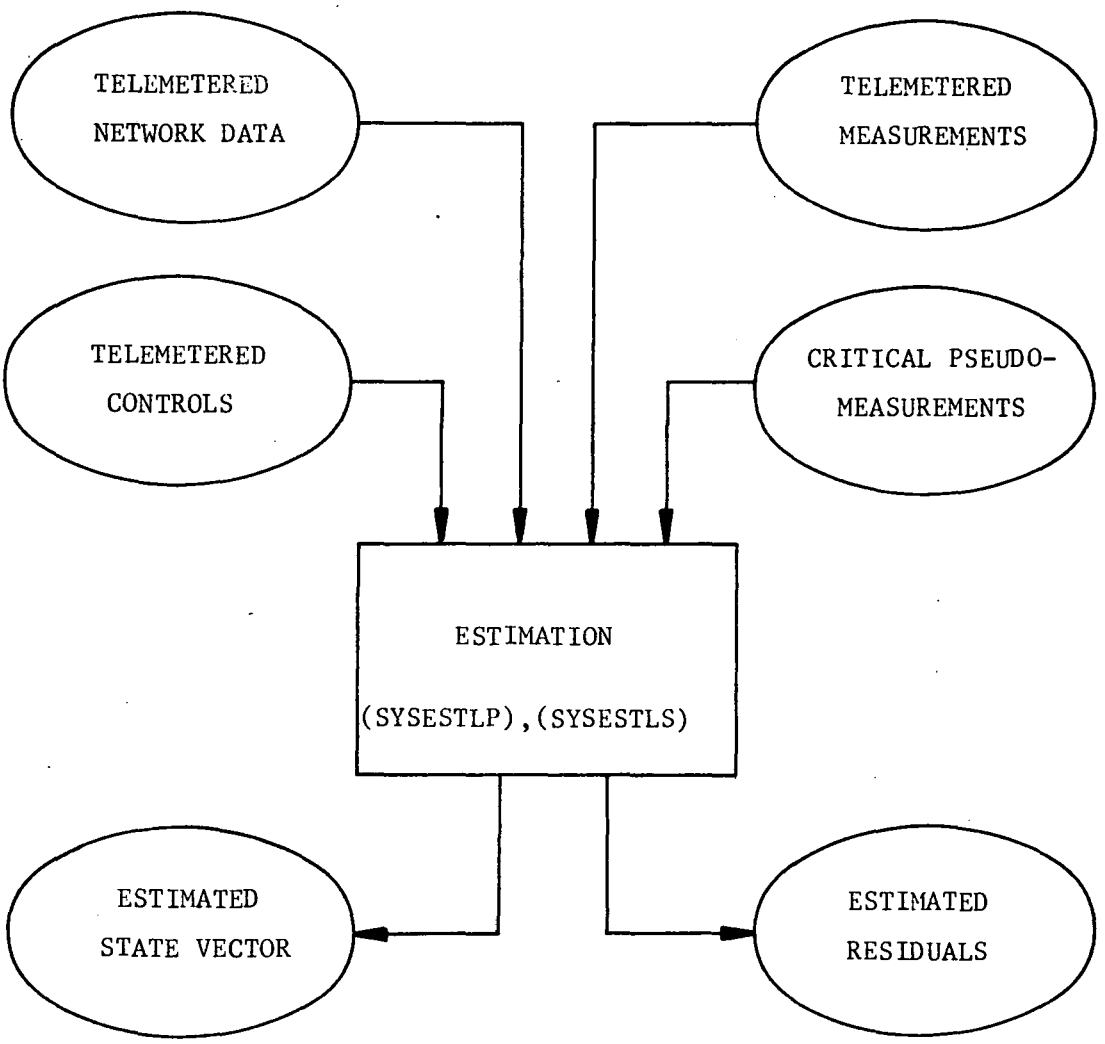


Figure 6.5 Interaction of state estimation programs with task common blocks

state estimators based on the augmented matrix method (SYSESTLS) and on the linear programming approach (SYSESTLP), respectively, have been developed and implemented.

Software details (SYSESTLS)

Size - 3280 lines of FORTRAN 77 code
164.75 k bytes

Speed - ~1.2 s/iter (PE 3220)

Software details (SYSESTLP)

Size - 3970 lines of FORTRAN 77 code
161 k bytes

Speed - ~1.1 s/iter (PE 3220)

6.6 BAD DATA PROCESSING PROGRAMS

The input data of the bad data detection and identification program (Figure 6.6) depends on the estimator used to calculate the state vector. In the case of the least-squares estimator the procedure requires information about the measurement residuals and nominal accuracy of the measurements. If the least absolute values estimator is employed the gross measurement errors can be identified by simply processing the measurement residuals. Additional information about meter positioning can be used by either method in order to improve the reliability of the

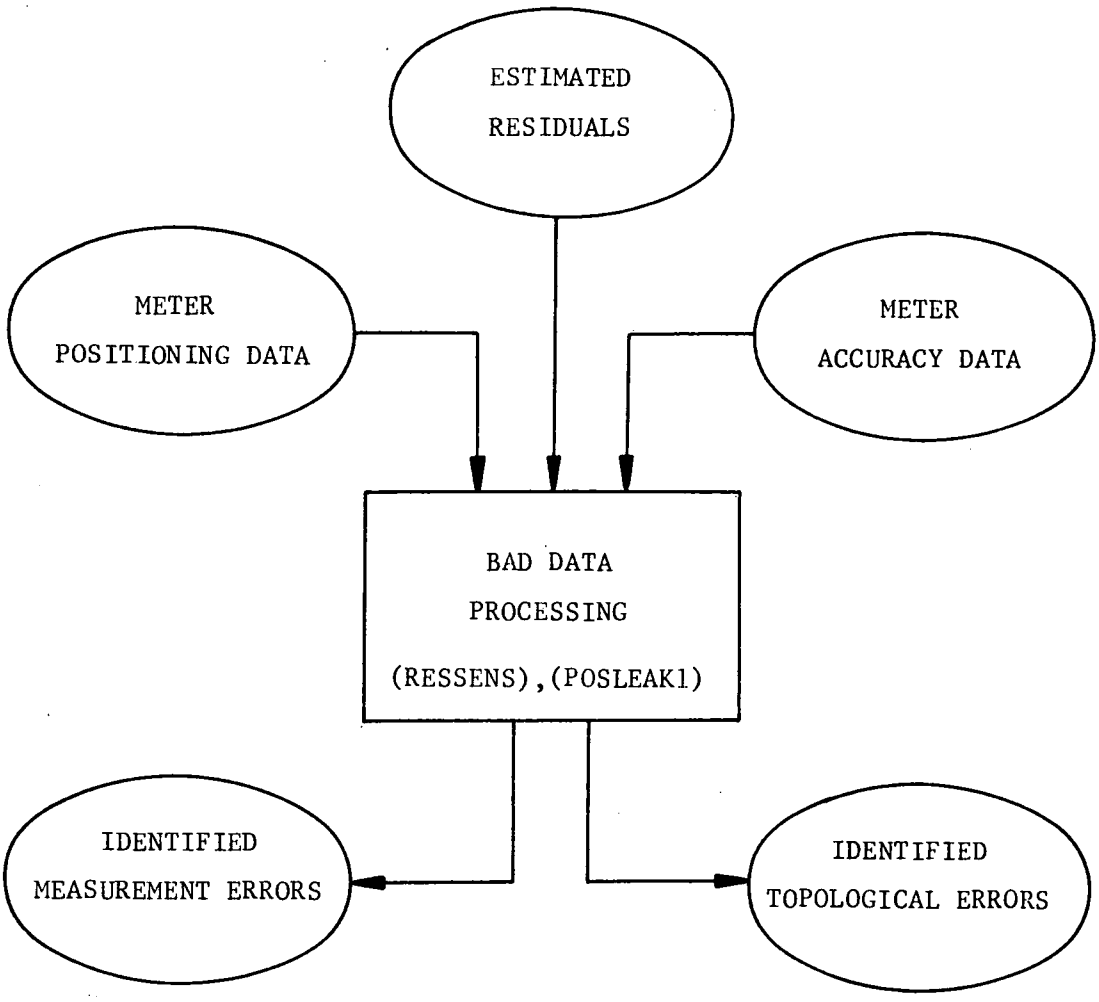


Figure 6.6 Interaction of bad data processing programs with task common blocks

identification of multiple interacting bad data.

Software details (POSLEAK 1 + MAIN)

Size - 300 lines of FORTRAN 77 code
24 k bytes

Speed - 0.027 s (PE 3220)

Software details (RESSENS + MAIN)

Size - 400 lines of FORTRAN 77 code
28 k bytes

Speed - 1.2 s/cycle (PE 3220)

6.7 VALVE CONTROL PROGRAM

The valve control program (Figure 6.7) uses a real-time data base created by the monitoring programs. At this stage the telemetered data is expected to be free from bad data points and the estimate of the state vector is assumed to reflect the actual state of the system. The output of the program is the set of optimal valve controls which can be implemented directly, in on-line mode, or indirectly via the human operator.

Software details (VALCON)

Size - 2900 lines of FORTRAN 77 code

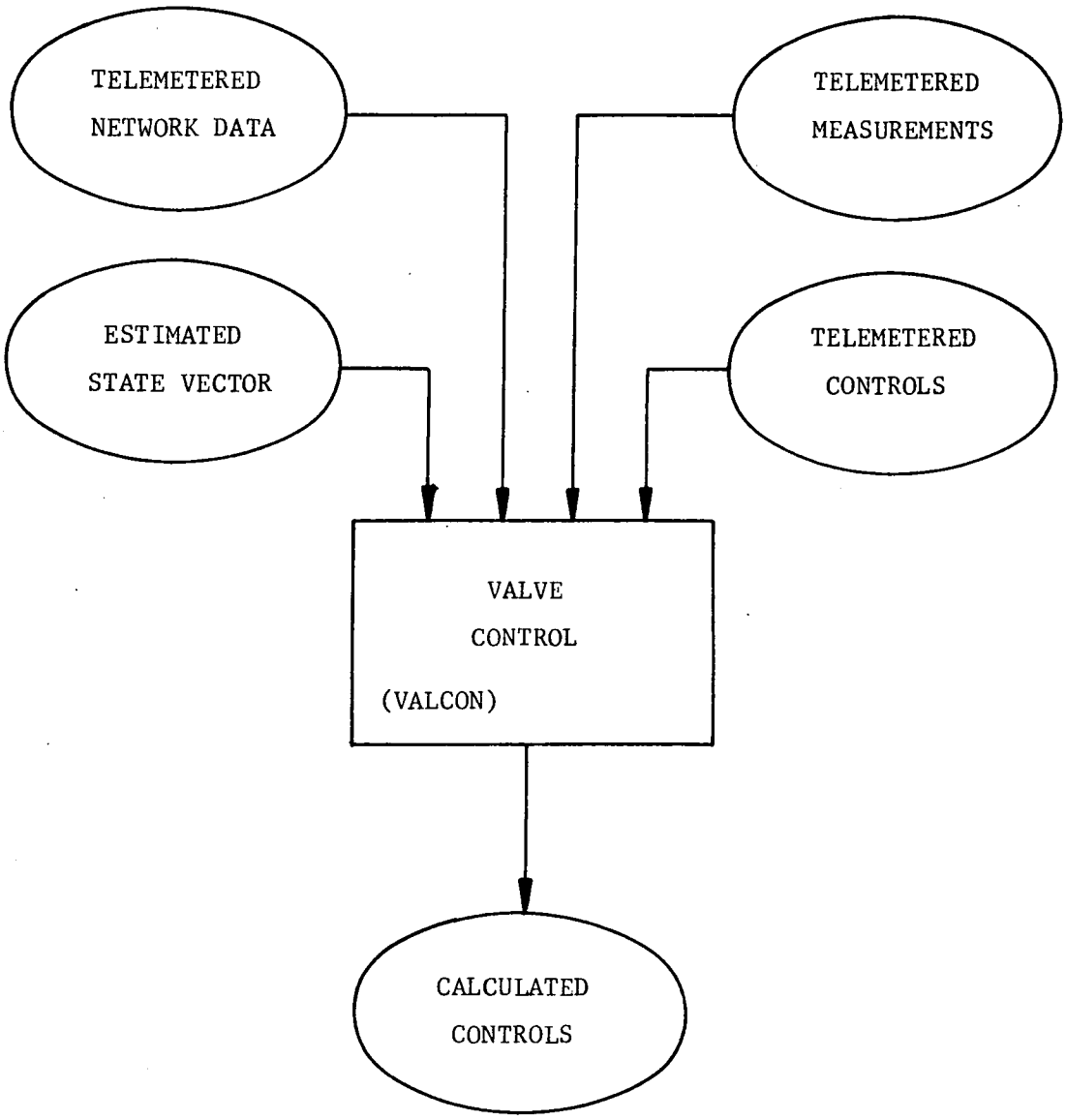


Figure 6.7 Interaction of valve control program with task common blocks

120 k bytes

Speed - ~1.1 s/iter (PE 3220)

6.8 GRAPHICAL DISPLAY PROGRAM

The software package facilitates presentation of the results of simulation, estimation and bad data processing routines both in tabular and graphical form. Consequently, the graphical display program (Figure 6.8) has access to the exact, telemetered and monitored data files. The network is represented in the form of a diagram illustrating the main hydrological elements. The program allows for continuous zooming onto any part of the network and, according to the chosen magnification coefficient, for a varying amount of detail about the network to be displayed. Emergency states of the network such as leakages or control valve failures, which need further attention of the operator, are monitored in the form of alarms.

Software details (NETDIS1)

Size - 1680 lines of FORTRAN 77 code
110 k bytes

Speed - 3 s CPU time
20 s transmission (SIGMA colour graphic with transmission line 9600 bit/sec)

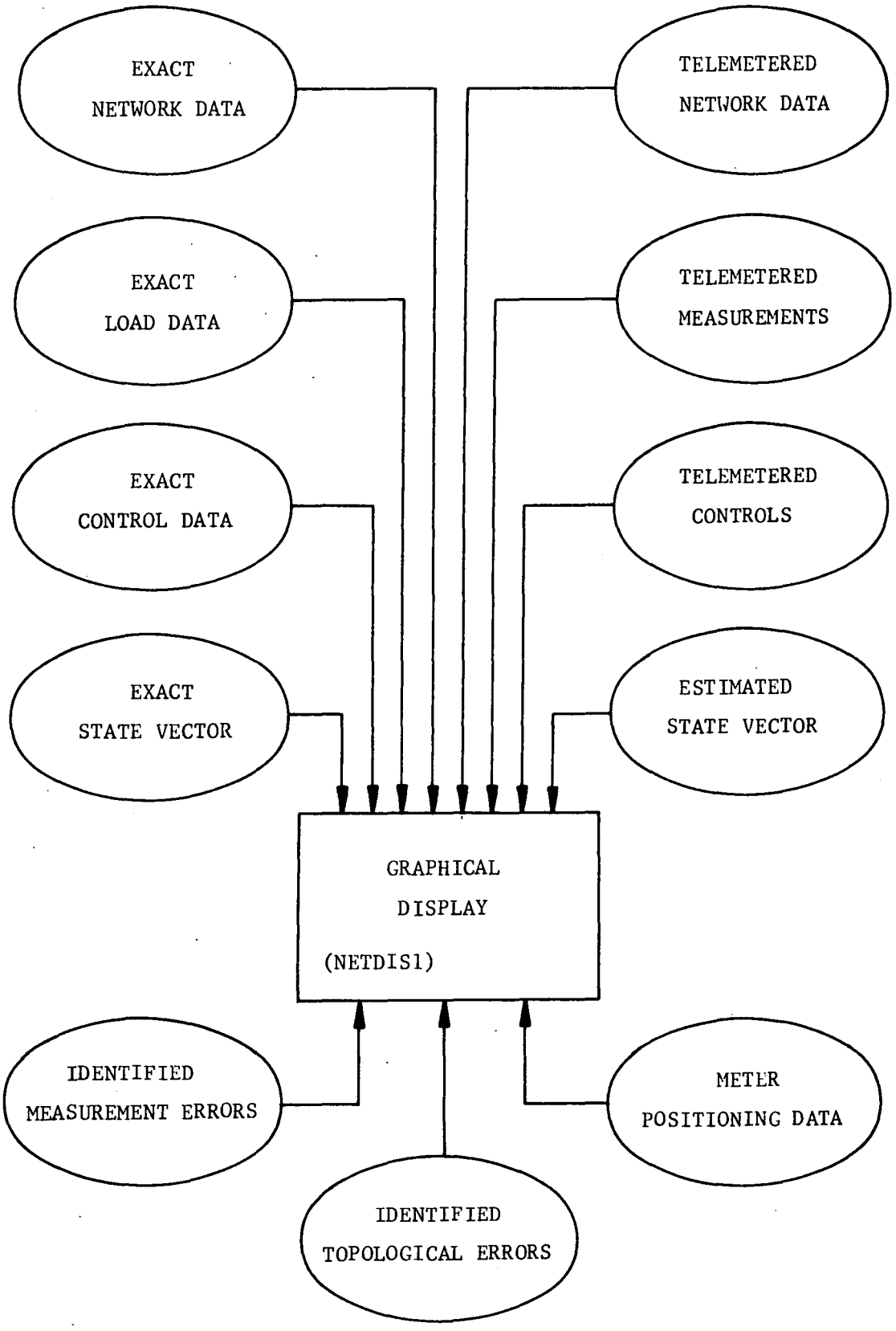


Figure 6.8 Interaction of graphical display program with task common blocks

6.9 OPERATORS' INTERFACE PROGRAM

An operator interface program (Figure 6.9) enables the operator to select and implement controls using information provided by the monitoring programs and optimal valve control algorithm. It also allows modification of the set of measurement points, the changing of Gaussian noise parameters and the simulation of the occurrence of bad data by corrupting the values of the telemeasurements.

Software details (OPERATOR)

Size - 700 lines of FORTRAN 77 code
48.25 k bytes

Speed - depends on the display required
(average: 3 s CPU time + 20 s
transmission time with 9600 bit/s
transmission line).

6.10 CONCLUDING REMARKS

In the previous sections of this chapter the general architecture of the software package for real-time water network monitoring has been described. By observing which task common blocks are made available to each individual program the data flow pattern within the package can be readily established. Since the number of interactions precludes a comprehensive discussion of the modes of

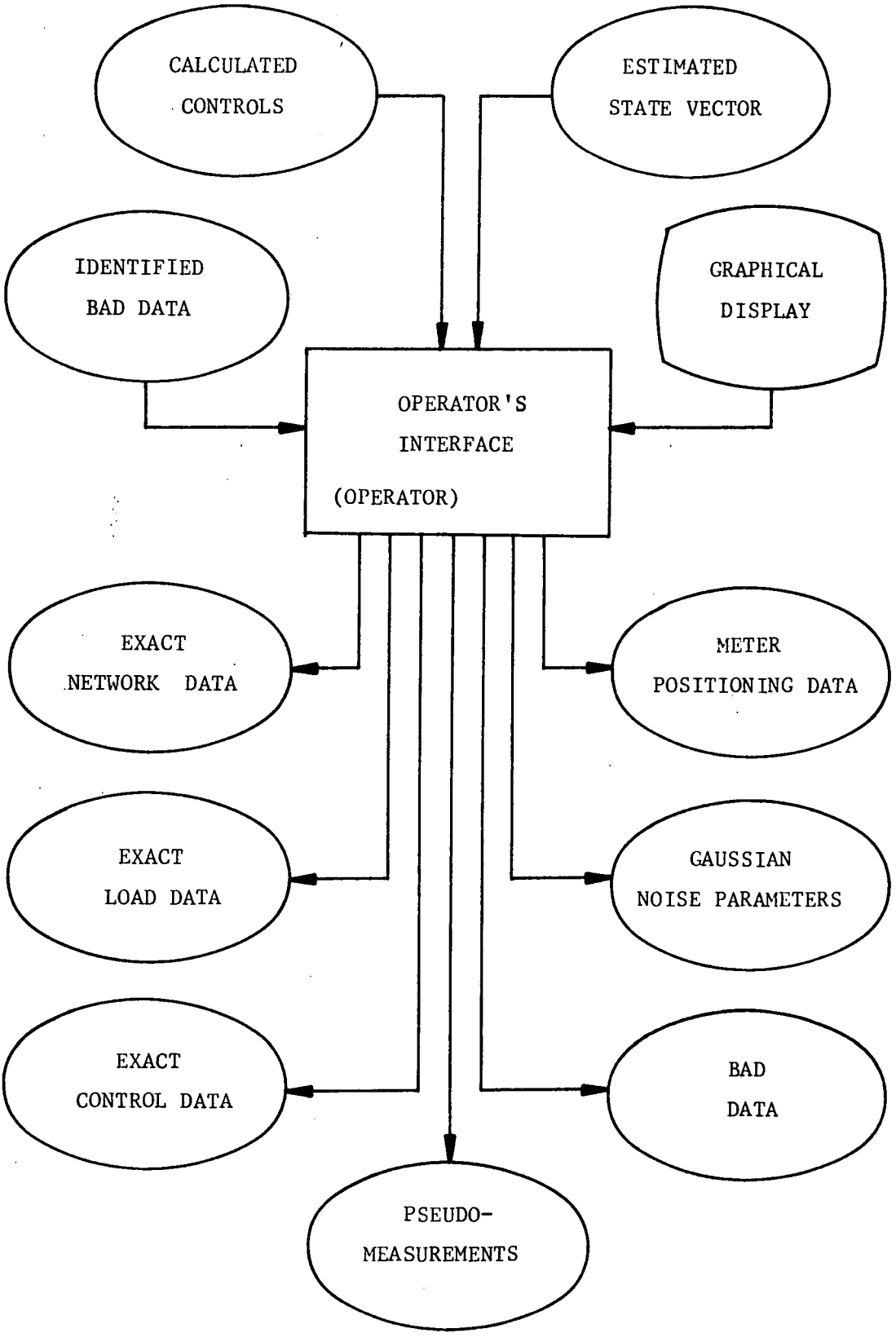


Figure 6.9 Interaction of operators' interface program with task common blocks

operation of the package within the format of this thesis, it is reasonable to focus the attention on one specific state of the system and to present only a sample of displays which are given to the operator. A detailed discussion of the results obtained with different programs being included in previous chapters.

The set of three graphical displays presented here refer to the situation where the 34-node water supply network contains a leakage in the pipe between nodes 4 and 20, control valves are monitored correctly, the telemetered data contains no gross measurement errors and the measurement configuration ensures a local measurement redundancy in every node of the network. Figure 6.10 gives the actual state of the system calculated by the network simulation program. The location of the metering points and the values of the corresponding telemeasurements are depicted in Figure 6.11 and the estimate of the system state is shown in Figure 6.12. The graphical form of display makes it easy to associate the numerical information with the network topology which is particularly useful in emergency operating conditions such as pipe ruptures or instrumentation malfunctions. While the Figures 6.10 to 6.12 give a good example of the form of an interface offered by the package they only represent a small fraction of the overall system and in particular they do not refer to the more specialised tasks like the valve control or observability testing.

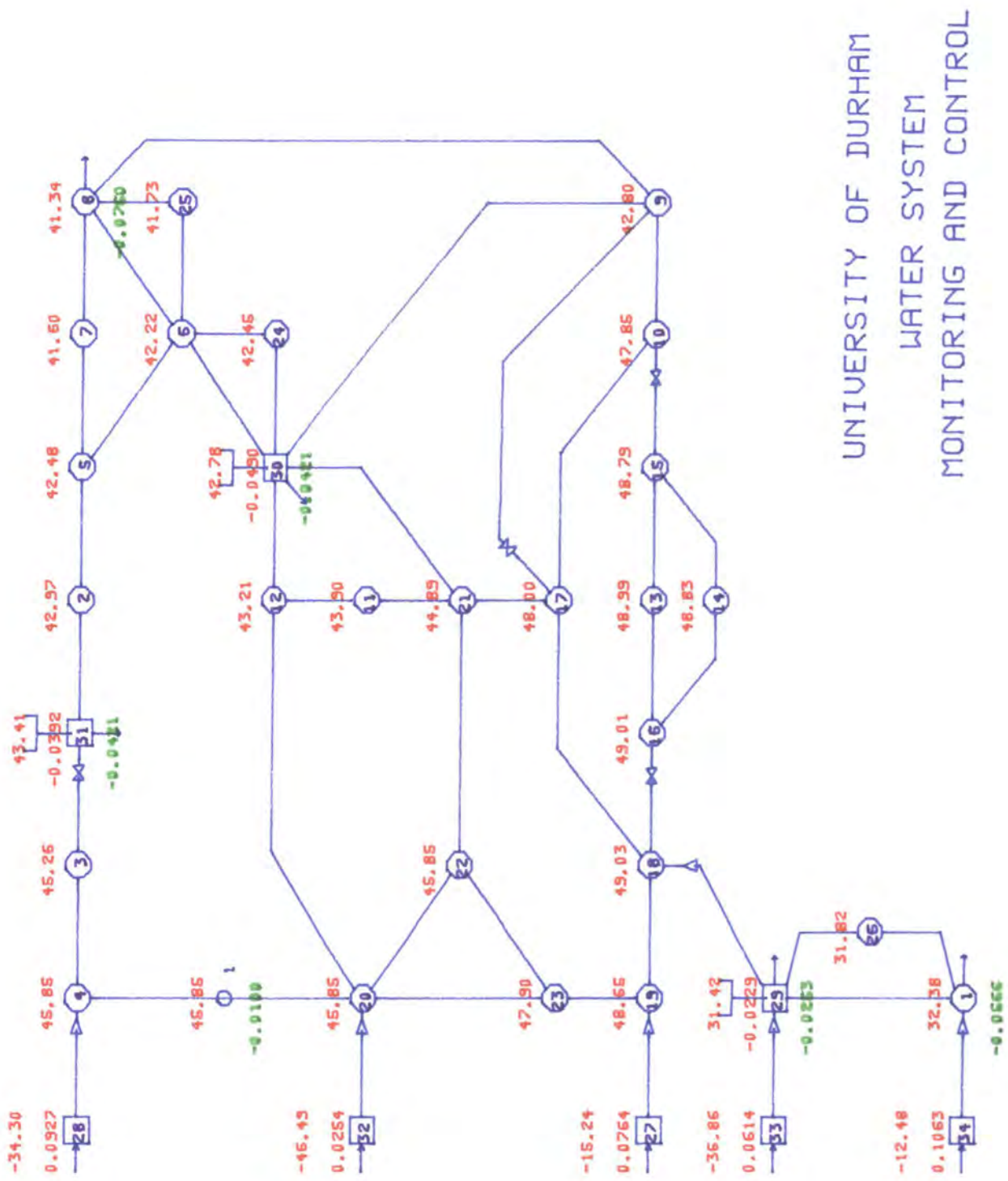


Figure 6.10 Actual state of the system

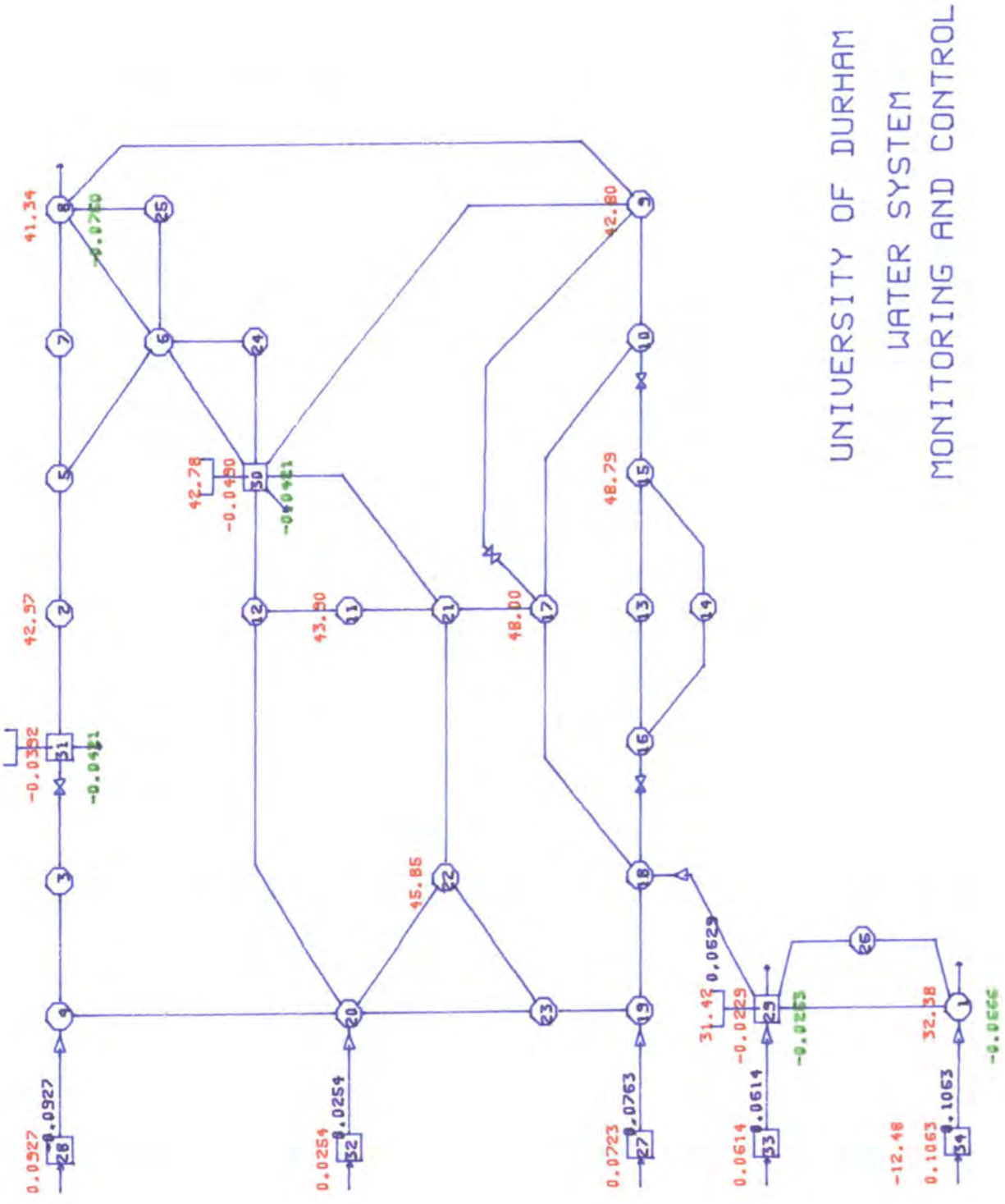
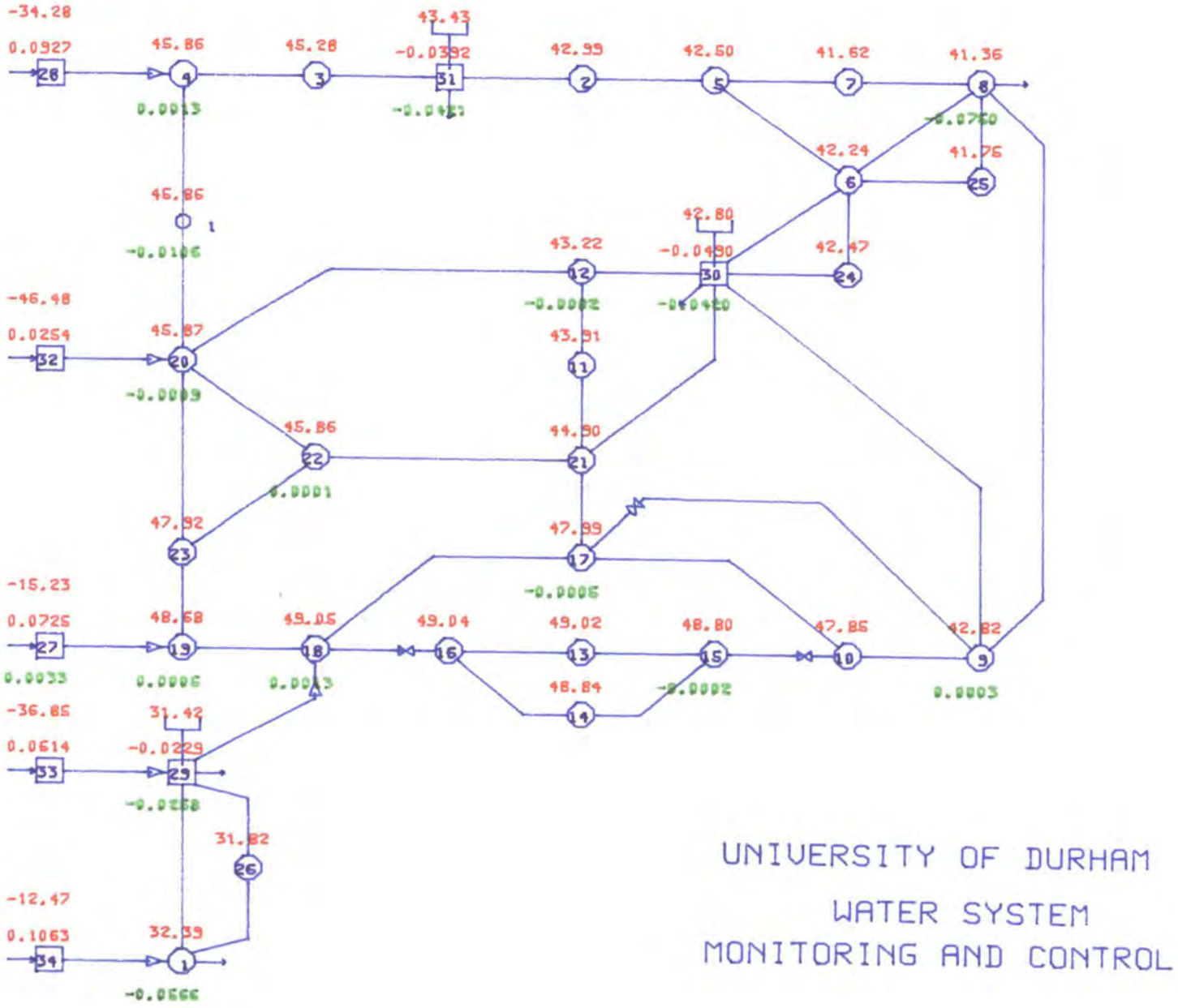


Figure 6.11 Telemeasurements



UNIVERSITY OF DURHAM

WATER SYSTEM

MONITORING AND CONTROL

Figure 6.12 Estimated state of the system

CHAPTER VII

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

7.1 CONCLUSIONS

In this work consideration has been given to a real-time water network monitoring scheme which is a prerequisite of any form of on-line control and which in itself gives a significant saving of operational cost by providing an indication of the leakages and information about the current head profile in the whole network.

It has been argued that, because of the possibility of the presence of bad data points among the measurements supplied by the telemetry system, it is essential that the measurement set possesses a degree of redundancy which enables rejection of spurious readings. As a consequence it has been necessary to depart from the load-flow solution of the network and to resort to the state estimation technique which can efficiently deal with an overdetermined set of measurement operations.

Two methods of state estimation have been proposed: a least-squares algorithm based on the augmented matrix

approach and a least absolute values algorithm which has been formulated as a linear programme. The theoretical background of both methods and their implementation taking into account sparsity and numerical stability considerations have been discussed in Chapter II. The results of tests conducted on two water systems have indicated that the least absolute values estimator is better suited for the purpose of on-line network monitoring. However, it is thought that in the case of a weak measurement configuration or for a large water network a hybrid approach amalgamating the features of the least-squares and least absolute values estimators could be profitably developed.

Bad data detection and identification problems have been investigated in Chapter III. For the least-squares estimator a new technique of calculating the residual sensitivity matrix has been proposed. It enables an efficient and numerically stable implementation of the bad data processing routine. The difficulty associated with this approach seems to be the fact that it is necessary to reestimate the state vector after identification of each bad data point since the method based on hypothesis testing guarantees identification of one measurement error at a time.

Another method of identification of bad data has been developed in connection with the least absolute values estimator. The procedure merely checks the magnitude and

sign of the weighted measurement residuals since the topological errors are shown to be equivalent to a pair of gross mass-balance errors in the end-nodes of the pipe concerned. The computational requirements of the procedure are very small, thus favouring the latter approach for on-line water network monitoring.

Numerical tests carried out on the realistic 34-node system using both routines have highlighted the relationship between the error detectability and measurement redundancy. It has been concluded that by performing the observability test it is possible to determine in which areas of the network bad data can be identified using the current measurement set.

The observability problem has been studied in Chapter IV. By analysing an approximate measurement model it has been shown that the observability test is equivalent to the search for a maximum spanning tree with measurement assignment in the augmented network graph. Consequently, the procedure does not require any floating-point computations. Two original combinatorial methods have been proposed to test topological observability of the water network. The first method searches for a maximum matching in a measurement-to-edge bipartite graph and attempts to build a spanning tree of the network using the branches with measurement assignment. The second method is based on the direct search for an observable spanning tree. The

equivalence of the preservation of the path property and the existence of the observable spanning tree has been exploited in order to devise a procedure to correct possible misassignments during the first stage of the algorithm. Both algorithms are computationally inexpensive, thus applicable to on-line observability checks on currently available measurement data.

Using a reliable data base created by the monitoring programs the problem of reduction of distributed leakages, which complements identification of pipe ruptures, has also been studied. The application of the state estimation technique proved to be an efficient method of calculating the optimal valve controls which minimize the overpressures in the network. The economy of the optimal valve control policy is shown by comparison of the volume of leakages for uncontrolled, manually controlled and optimally controlled network.

A general concept of the organisation of the water network monitoring software package has been described in Chapter VII. A useful degree of flexibility has been achieved by adopting a modular structure of the package in which the programs communicate with others via task common blocks.

7.2 SUGGESTIONS FOR FURTHER RESEARCH WORK

The following topics related to the problems discussed in this thesis deserve further research effort:

1) Development of the state estimation procedure which amalgamates the useful features of the least-squares and least absolute values estimators. The objective is to achieve an algorithm which has a linear computational complexity, good error detection features and is numerically stable;

2) Assessment of the advantages of treating the equations corresponding to zero load as equality constraints for the least-squares problem using the augmented matrix formulation. This approach has been proposed in reference [10] to be applied to the normal equation technique in order to save computing time and possibly reduce the number of iterations in Newton's method. The investigation of the possible effects of the use of equality constraints on the bad data identification procedures is also recommmended;

3) Study of an alternative approach to bad data identification problem in connection with the least-squares estimator;

4) Use of the reliability indices in observability studies;

5) Development of the combinatorial method of assessing the quality of the estimates calculated using the prespecified measurement set;

6) Research on the application of decomposition techniques to the water system state estimation problem;

7) Study of the problems of bad data detection, identification and observability related to two-level estimators.

REFERENCES AND BIBLIOGRAPHY

1. ABOYTES F., CORY B.J.: "Identification of measurement parameter and configuration errors in static state estimation", PICA Conf. Proc., 1975

2. ABRAMOV N.: "Methods of reducing power consumption in pumping water in water supply systems, including the use of booster stations", The Intern. Water Supply Assn., 8th Ann. Conf., M3-M20 (1969)

3. ALEXANDER S.M., GLENN N.L., BIRD D.W.: "Advanced techniques in the mathematical modelling of water distribution systems", Jour. AWWA, 67, 7, Jul. 1975, p.343

4. ALLEMONG J.J., IRISARI G.D., SASSON A.M.: "An examination of solvability for state estimation algorithms", IEEE PES Winter Meeting, 1980

5. ALVAREZ C., ALBERTOS P.:"On-line observability determination as a future result of state estimation algorithms", IEEE Trans. PAS, 101, 4, Apr. 1982, pp. 767-773

6. ARAFEH S.A.: "Real-time security assessment with fast optimum generation shift control", PICA Proc., Toronto, Ontario, 1977, pp. 360-368

7. ARAFEH S.A., SHINZINGER R.:"Estimation algorithms for large-scale systems", IEEE Trans. PAS, 98, 12, Dec. 1979, pp. 1968-1977

8. ARIATTI F., MARZIO L., RICCI P.: "Designing state estimation in view of reliability", Proc. 5th PSCC, Cambridge, 1975

9. ASCHMONEIT F.:"Optimal power system static-state estimation", Proc. 5th PSCC, Cambridge, 1975

10. ASCHMONEIT F., PETERSON N.M., ADRIAN E.C.:"State estimation with equality constraints", PICA Proc., Toronto, Ontario, May 1977, pp. 427-430

11. BABIC B.S.:"Decoupling load flow with variables in rectangular form", IEE Proc., 130, C, 3, May 1983, pp. 98-102

12. BALINSKI M.L., HELLERMAN E. (Eds.):Computational practice in mathematical programming, (Mathematical programming studies 4),North-Holland Publ.Co.,1975

13. BARGIELA A.:"Suboptimal dual control of discrète objects in presence of Gaussian noise with unknown parameters" (in Polish), Zeszyty Naukowe Pol. Sl., 47, 1978

14. BARGIELA A.: "A linear control algorithm of a multiexit reservoir in stochastic conditions" (in Polish), Zeszyty Naukowe Pol. Sl., 48, 1978

15. BARGIELA A., POCIASK U.:"Optimisation of the work of the chosen hydraulic object set", Proc. 2nd Int. Conf. Syst. Analysis and Simulation, Berlin, 1980

16. BARLOW J.F., MARKLAND E.:"Computer design of pipe networks", Proc. Inst. Civ. Engs., 51, Aug. 1972, pp. 225-235

17. BARLOW J.F., MARKLAND E.:"Computer analysis of pipe networks", Proc. Instn. Civ. Engrs., 43, May-Aug. 1969, pp. 249-259

18. BARRODALE I., ROBERTS F.D.K.:"An improved algorithm for discrete L_1 linear approximation", SIAM, Jour. Num. Anal., 10, 5, Oct. 1973, pp. 839-848

19. BARRODALE I., ROBERTS F.D.K.:"Algorithm 478: solution of an overdetermined system of equations in the L_1 norm", ACM, 17, 1974, pp. 319-320

20. BARRODALE I., ROBERTS F.D.K.:"Algorithms for restricted LAV estimation", Comm. Stat. Simul. Comp., 6, B, 1977, pp. 353-364

21. BARTELS R.H.:"A stabilization of the Simplex method", Num. Math., 16, 1971, pp. 414-434

22. BARTELS R.H., CONN A.R., SINCLAR J.W.:"A FORTRAN program for solving overdetermined systems of linear equations in the L_1 sense", Technical Report 236, Dept. Math. Science ,J. Hopkins University, Baltimore, 1976

23. BARTELS R.H., CONN A.R., SINCLAR J.W.:"Minimization techniques for piecewise differentiable functions: the L_1 solution to an overdetermined linear system" SIAM Jour. Num. Anal., 15, 2, Apr. 1978, pp.224-241

24. BAZARAA M.S., SHETTY C.M.: Nonlinear programming: theory and algorithms, John Wiley & Sons, 1979

25. BAZARAA M.S., JARVIS J.J.: Linear programming and network flows, John Wiley & Sons, 1977

26. BECHER A.E. Jr., BIZJAK G.J., SCHULZ J.W.:"Computer techniques for water distribution analysis", Jour. AWWA, 64, 7, Jul. 1972, p. 410

27. BELLAMY B.E.:"The analysis of networks of pipes and pumps", Jour. Instn. Engrs. Aust., 37, 4-5, pp. 111-116

28. BERTSEKAS D.P.:" Large-scale separable integer programming problems and the exponential metchods of multipliers", Tech. Report. Alphatech Inc., Contract No. DE-AC01-79-ET-29008

29. BERTSEKAS D.P.:"Projected Newton methods for optimization problems with simple constraints", Conf. Decision and Control, Dec. 1981

30. BERTSEKAS D.P., LAVER G.S., SANDELL N.R., POSBERGH T.A.: "Optimal short-term scheduling of large-scale power systems", Conf. Decision and Control, Dec. 1981

31. BIERMAN J.G.: Factorization methods for discrete sequential estimators, Academic Press, New York, 1977

32. BJORK A.: "Solving linear least squares problem by Gram-Schmidt orthogonalisation", BIT, 7, 1967, pp. 257-278

33. BONANOMI P., GRAMBERG G.:"Power system data validation and state calculation by network search techniques", IEEE, PES Summer Meeting, San Francisco, California, July 18-23, 1982

34. BONANSIGA A.P.: "Computer analysis of water distribution systems", Jour. AWWA, 67, 7, Jul. 1975, p.347

35. BRAMELLER A.,JOHN M.N.,SCOTT M.R.: Practical diakoptics for electrical networks, Chapman & Hall Ltd., London 1969

36. BRAMELLER A., HAMAM Y.M., ALLAN R.N.:Sparsity, its practical application to system analysis, Pitman, 1975

37. BRIGGS R.:"Instrumentation and control in the United kingdom water industry", AWWA Sem. Proc. Water Plant Instrum. and Autom., Public. 21034, Jun. 1976

38. BROUSSOLLE F.:"State estimation in power systems: detecting bad data through the sparse inverse matrix method", IEEE Trans., PAS, 97, 3, May-June 1978, pp. 678-682

39. BROYDEN C.G.:"The convergence of an algorithm for solving sparse nonlinear systems", Mathematics of Computation, 25, 144, Apr. 1971

40. BUNCH J.R., PARLETT B.N.:"Direct methods for solving symmetric indefinite systems of linear equations", SIAM Jour. Numer. Anal., 8, 4, Dec. 1971, pp. 639-655

41. BUNN D.W., PASCHENTIS S.N.:"A model-switching criterion for a class of stochastic linear programs", Appl. Mat'. Modelling, 5, Oct. 1981, pp. 336-340

42. BUNN D.W., PASCHENTIS S.N.:"Economic dispatch of electric power by stochastic linear programming", Research Paper 638, Graudate School of Business, Stanford University, Apr. 1982

43. BURCH R.H., MARLOW K.C.:"A computer based telecontrol and communication system for a water supply network", Jour. of Hydrology, 28, 1976, pp. 407-427

44. BURCHETT R.C., HAPP H.H., WIRGAU K.A.:" Large-scale optimal power flow", IEEE Trans., PAS, 101, 10, Oct. 1982, pp. 2995-2999

45. CAGNON C.R., JACOBY S.L.S.:"Computer simulation of water distribution networks", Jour. Trans. Eng. Div., ASCE, 101, TE 3, Aug. 1975, pp. 553-567

46. CARLSON C.E.C.:"The Denver system of controls", Jour. AWWA, 63, 8, Aug. 1971, p.513

47. CARPANETO G., TOTH P.:" Algorithm 548: solution of the assignment problem", ACM Trans. Math. Software, 6, 1, March 1980, pp. 104-111

48. CAVES J.L., EARL T.C.:"Computer applications: a tool for water distribution engineering ", Jour. AWWA, May 1979, pp. 231-235

49. CESARIO A.L.:" Computer modelling programs: tools for model operations", Jour. AWWA, Sept. 1980, pp. 508-513

50. CHAPMAN M.J., JONES R.P., PRITCHARD A.J.:"State observers for monitoring gas pipelines", 3rd IFAC Symposium Contr. Distr. Syst., Toulouse

51. CHEUNG T.Y.:"Computational comparison of eight methods for the maximum network flow problem", ACM Trans. Math. Software, 6, 1, March 1980, pp. 1-16

52. CHRISTIANSEE W.R.:"Short term load forecasting using general exponential smoothing", IEEE Trans., PAS, 90, Mar.-Apr. 1971, pp. 900-910

53. CLARKE D.A., McBEAN E.A., AL-NASSRI S.A.: "Uncertainties in water distribution systems", Jour. of Hydraulics Div. ASCE, 107, 10, 1981, pp. 1263-1267

54. CLEMENTS K.A., KRUMPHOLZ G.R., DAVIS P.W.:"Power system state estimation residual analysis: an algorithm using network topology", IEEE Trans., PAS, 100, 4, Apr. 1981, pp. 1779-1787

55. CLEMENTS K.A., KRUMPHOLZ G.R., DAVIS P.W.:"State estimator measurement system reliability evaluation. An efficient algorithm based on topological obervability theory", IEEE Trans., PAS, 101, 4, Apr. 1982, pp. 997-1004

56. CLEMENTS K.A., KRUMPHOLZ G.R., DAVIS P.W.:"Power system state estimation with measurement deficiency an algorithm that determines the maximal observable subnetwork", IEEE Trans., PAS, 101, 9, Sept. 1982, pp. 3044-3052

57. CLEMENTS K.A., KRUMPHOLZ G.R., DAVIS P.W.:"Power system state estimation with measurement deficiency an observability/measurement placement algorithm", IEEE Trans., PAS, 102, 7, Jul.1983, pp. 2012-2020

58. CLEMENTS K.A., WOLLENBERG B.F.: "An algorithm for observability determination in power system state estimation", IEEE PES Summer Meeting, 1975

59. COLLINS M., COOPER L., HELGASON R., KENNINGTONJ., LE BLANC L.: "Solving the pipe network analysis problem using optimization techniques", Management Science, 24, 7, March 1978, pp. 747-760

60. CONRAD V., WALLACH Y.:"Parallel optimally ordered factorization", PICA Conf. Proc., Toronto, Ontario, May 1977, pp. 302-306

61. COULBECK B.:Optimisation and modelling techniques in dynamic control of water distribution systems, PhD Thesis, Dept. Control Eng., University of Sheffield, July 1977

62. COULBECK B., STERLING M.J.H.:"Modelling techniques in dynamic control of water distribution systems", Jour. Instn. Measur. Contr., 11, Oct. 1978, pp. 385-389

63. COULBECK B., STERLING M.J.H.:"Optimised control of water distribution systems", IEE Proc., 125, 9, Oct. 1978, pp. 1039-1045

64. CROW W.B., EIDSNESS F.A.: "Savings through instrumentation and control in two water systems", Jour. AWWA, 57, 12, Dec. 1965, p. 1509

65. CURTIS A.R., REID J.K.:"On the automatic scaling of matrices for Gaussian elimination", Jour. Inst. Math. and its Appl., 10, 1972, pp. 118-124

66. DEBS A.S., LARSON R.E., HAJDU L.P.:"On-line sequential state estimation for power systems", 4th PSCC Conf., Grenoble 1972

67. DEMOYER R., GILMAN H.D., GOODMAN M.Y.:"Dynamic computer simulation and control methods for water distribution system", Gen. Elect. Doc. No. 73SD205, Philadelphia, PA, Feb. 1973

68. DEMOYER R.: A statistical approach to modelling and control of water distribution systems, PhD Dissertation, Polytechnic Inst. of New York, Brooklyn N.Y., 1973

69. DEMOYER R.:"Statistical approach to dynamic modelling and control of water distribution systems", NTIS, PB-226 604, June 1973

70. DEMOYER R., GILMAN H.D., GOODMAN M.Y.:"An application study in water distribution control", NTIS, PB- 227 336/74SD200, Jan. 1974

71. DEMOYER R., HORWITZ L.B.:"Macroscopic distribution system modelling", Jour. AWWA, 67, 7, Jul. 1975, pp. 377-380

72. DEMOYER R., HORWITZ L.B.:"Automatic control of water distribution", Jour. Dyn. Syst. Meas. and Cont., ASME Trans., June 1975, pp. 126-130

73. DEMOYER R., HORWITZ L.B.: A system approach to water distribution modelling and control, Lexington Books, Mass., 1975

74. DeNEUFVILLE R., SCHAAKE J.Jr., STAFFORD J.H.:"Systems analysis of water distribution networks", Jour. Sanit. Engrg. Div.-ASCE, 97, 6, Dec. 1971, p. 825

75. DEO N.:Graph theory with applications to engineering and computer science, Prentice-Hall Inc., Englewood Cliffs, N.J., 1974

76. DICKINSON G.C.:"Waterworks telemetry objectives", Inst. Water Engrs., Symposium at Sutton Coldfield, Feb. 1969

77. DIVENOT A.:"Une nouvelle methode de calcul des reseaux mailles", La houille blanche, 6, 1980, pp. 365-376

78. DIVENOT A.:"Localisation centralisee des ruptures dans un reseau de distribution d'eau", Techniques et Sciences Municipales, 76, 8-9, 1981, pp. 477-481

79. DILLINGHAM J.H.:"Automating distribution system operations", Jour. AWWA, March 1979, pp. 133-135

80. DILLINGHAM J.H.:"Computer analysis of water distribution systems, part 2", Water and Sewage Works, 114, 2, Feb. 1967, pp. 43-45

81. DINIC E.A.:"Algorithm for solution of a problem of maximum flow in a network with power estimation", Soviet Mat. Dokl., 11, 5, 1970, pp. 1277-1280

82. DODGE E.P., HOELLEIN H.R., TETMAJER L.:"The analysis of large, complex water networks with small computer systems", Jour. AWWA, 70, 7, July 1978, p. 366

83. DOMMEL H.W., TINNEY W.F.:"Optimal power flow solutions", IEEE Trans., PAS, 87, 10, Oct. 1968, pp. 1866-1874

84. DONACHIE R.P.:"Digital program for water network analysis", Jour. Hydraul. Div., ASCE, HY3, March 1974, pp. 393-403

85. DOPAZO J.F., KLITIN O.A., STAGG G.W., Van SLICK L.S.: "State calculation of power systems from line-flow measurements", IEEE Trans. PAS, 89, 7, Sept-Oct. 1970, pp. 1698-1708

86. DOPAZO J.F., KLITIN O.A., Van SLICK:"State calculation of power systems from line-flow measurements, part 2", IEEE Trans., PAS, 91, Jan-Feb. 1972, pp.145-151

87. DOPAZO J.F.,EHRMANN S.T., KLITIN O.A., SASSON A.M.: "Justification of the AEP real time load flow project", IEEE Trans., PAS, 92, Sept-Oct. 1973, pp. 1501-1509

88. DOPAZO J.F., KLITIN O.A., SASSON A.M.,"State estimation for power systems: detection and identification of gross measurement errors", PICA Conf. Proc., 1973

89. DUFF I.S.:"A survey of sparse matrix research", Proc. IEEE, 65, 4, Apr. 1977, pp. 501-535

90. DUFF I.S.:"Algorithm 575. Permutations for a zero-free diagonal", ACM Trans. on Math. Software, 7, 3, Sept. 1981, pp. 387-390

91. DUFF I.S.:"On algorithms for obtaining a maximum transversal", ACM Trans. on Math. Software, 7, 3, Sept. 1981, pp. 315-330

92. DUFF I.S., REID J.K.:"A comparison of sparsity ordering for obtaining a pivotal sequence in Gaussian elimination", Jour. IMA, 14, 1974, pp. 281-291

93. DUFF I.S., REID J.K.:"A Comparison of some methods for the solution of sparse overdetermined systems of linear equations", Jour. IMA, 17, 1976, pp. 267-280

94. DUFF I.S., REID J.K.:"An implementation of Tarjan's algorithm for the block triangularization of a matrix", ACM Trans. Math. Software, 4, 2, June 1978, pp. 137-147

95. DUFF I.S., REID J.K.:"Algorithm 529: permutations to block triangular form", ACM Trans. Math. Software, 4, 2, June 1978, pp. 189-192

96. DUFF I.S., REID J.K., MANSGARIAN J., NIELSEN I.: "Direct solution of linear equations whose matrix is sparse", Jour. IMA, 23, 2, 1978, pp. 235-250

97. EDELMANN H.:"A universal assessment for the superior quality of distribution of measuring points for the state-estimate of high voltage-network", PSC Conf. Proc., Cambridge, 1975

98. EDMONDS J.:"Optimum branching", Repr. from Jour. Res. Nat. Bur. Standards, Sect. B71, 1967, pp. 233-240

99. EDMONDS J.:"Matroid intersection", Annals of Discrete Mathematics, 4, 1979, pp. 39-49

100. EDMONDS J., KARP R.M.:"Theoretical improvements in algorithmic efficiency for network flow problems", Jour. Ass. Comp. Machinery, 19, 2, Apr. 1972, pp. 248-264

101. EPP R. FOWLER A.G.:"Efficient code for steady-state flows in networks", Jour. Hydraul. Div. ASCE, HY1, Jan. 1970, pp. 43-56

102. EVEN S.: Graph algorithms, Pitman, 1979

103. EVEN S., KARIV O.:"An $O(n^{2.5})$ algorithm for maximum matching in general graphs", 16th Annual Symp. on Foundations of Computer Science, IEEE, 1975, pp. 100-112

104. EVEN S., TARJAN R.E.:"Network flow and testing graph connectivity", SIAM Jour., 4, 1975, pp. 507-518

105. FALCAO D.M., KARAKI S.H., BRAMELLER A.:"Non-quadratic state estimation: a comparison of methods", 7th PSC Conf., Lausanne, 1981, pp. 1002-1006

106. FALLSIDE F., PERRY P.F.:"Hierarchical optimization of a water supply network", IEE Proc., 122, 2, Feb. 1975, pp. 202-208

107. FALLSIDE F., PERRY P.F.:"Hierarchical model for water-supply system control", IEE Proc., 122, 4, Apr. 1975, pp. 441-443

108. FALLSIDE F., PERRY P.F.:"On-line prediction of consumption for water supply control", Proc. 6th IFAC World Congress, Boston, Mass., Aug. 1975

109. FALLSIDE F., GRAY D.F., MOSS S.M., BURCH R.H.,JENKINS D., MARLOW K.C.: "On-line control of water supply system", Prog. Wat. Tech., 9, 5-6, 1977, pp. 289-298

110. FETZER E.E., ANDERSON P.M.:"Observability in the state estimation of power systems", IEEE Trans., PAS, 94, 6, Nov-Dec. 1975, pp. 1981-1988

111. FISHER M.L., NORTHUP W.D., SHAPIRO J.F.:" Using duality to solve discrete optimization problems: theory and computational experience", BALLINSKI M.L. (Ed), Mathematical programming study 3, North Holland Publ. Co., 1975, pp. 56-94

112. FINCHAM A.E., GOLDWATER M.H.:"Simulation models for gas transmission networks", Trans. Inst. MC, 1, 1, Jan/Mar. 1979

113. FISCHL R., HALPIN T., GUVENIS A.:"The application of decision theory to contingency selection", IEEE Trans., CAS, 29, 11, Nov. 1982

114. FLETCHER R.:"A new approach to variable metric algorithms", The Computer Jour., 13, 3, Aug. 1970, pp. 317-322

115. FORD L.R., FULKERSON D.R.:Flows in networks, Princeton University Press, 1962

116. FREESTON R.:"Feasibility of remote controls for distribution system", Jour. AWWA, 62, 11, Nov. 1970, p. 723

117. FRENZ C.M.:"Automated system control", Jour. AWWA, 63, 8, Aug. 1971, p. 508

118. FULKERSON D.R.:"An out-of-kilter method for minimal-cost flow problems", Jour. SIAM, 9, 1, March 1961, pp. 18-27

119. GAGNON C.R., JACOBY S.L.S.:"Computer simulation of water distribution networks", Transportation Engrg. Jour.-ASCE, 101, 3, Aug. 1975, p. 553

120. GENTLE J.E.:"Least absolute values estimation: an introduction", Comm. Statist.-Simul. Comput. (B), 6, 1977, pp. 313-328

121. GERLT J.L., HADDIX G.F.:"Distribution system operation analysis model", Jour. AWWA, 67, 7, Jul. 1975, p. 381

122. GILMAN H.D., GOODMAN M.Y., METKOWSKI R.:"Final report -mathematical modelling of water distribution systems", Gen. Elect. Doc. No. 71SD241, Philadelphia, Pa., July 1971

123. GILMAN H.D., GOODMAN M.Y., DEMOYER R.:"An experiment in computer-assisted supervisory control of a water distribution system", NTIS PB-213 645, June 1971

124. GILMAN H.D. et al.:"Mathematical modelling for status prediction and control of water distribution systems", NTIS, PB-227 919, 1973

125. GILMAN H.D., DEMOYER R., GOODMAN M.Y.:" Closed-loop control is coming", Water & Wastes Engrg., 10, June 1973, p. 42

126. GILMAN H.D., GOODMAN M.Y., DEMOYER R.:"Replication modelling for water-distribution control", Jour. AWWA, 65, 4, Apr. 1973, pp. 255-260

127. GOLUB G.H.: "Numerical methods for solving linear least-squares problems", Num. Math., 7, 1965, pp. 206-216

128. GU J.W., CLEMENTS K.A., KRUMPHOLZ G.R., DAVIS P.W.: "Topologically-partitioned power system state estimation", IEEE Trans., PAS, 102, 2, Feb. 1983, pp. 483-491

129. GU J.W.,CLEMENTS K.A., KRUMPHOLZ G.R., DAVIS P.W.: "The solution of ill-conditioned power system state estimation problems via the method of Peters and Wilkinson", IEEE Trans., PAS, 102, 10, Oct. 1983, pp. 3473-3480

130. HANDSHIN E., BONGERS C.: "Theoretical and practical considerations in the design of state estimators for electric power systems", HANDSHIN E. (Ed), Real-time control of electric power systems, Elsevier, 1972, pp. 104-136

131. HANDSCHIN E., SCHWEPPE F.C., KOHLAS J.,FIECHTER A: "Bad data analysis for power system state estimation", IEEE, PES, Summer Meeting & Energy Resources Conf., Anaheim, Cal., July 14-19, 1974,

132. HARRINGTON J.J.:"The role of computers in planning and managing water utilities", Jour. NEWWA, 81, 3, Sept. 1967, p. 231

133. HEIDARI M.,CHOW V.T., KOKOTOVIC P.V., MEREDITH D.D.: "Discrete differential dynamic programming approach to water resources system optimisation", Water Resources Research, 7, 1971, pp. 273-282

134. HEGENBART J.L.:"Use of telemetry and remote control in naturally occurring disasters", Jour. AWWA, 65, 2, Feb. 1973, p. 155

135. HENDERSON H.V., SEARLE S.R.:"On deriving the inverse of a sum of matrices", SIAM Rev., 23, 1, Jan. 1981, pp. 53-60

136. HOAG L.N., WEINBERG G.:"Pipeline network analysis by electronic digital computer", Jour. AWWA, 49, 1975, pp. 517-524

137. HOPCROFT J.E.,KARP R.M.:"An $n^{5/2}$ algorithm for maximum matching in bipartite graphs",IEEE Conf. Record 12th Annual Symp. on Switching and Automata Theory, 1971, pp. 122-125

138. HOPCROFT J.E., TARJAN R.:"Efficient algorithms for graph manipulation", Comm. ACM, 16, 6, June 1973, pp. 372-378

139. HORTON J.S., MASIELLO R.D.:"On-line decoupled observability processing", PICA Proc., Toronto, Ontario, May 1977, pp. 420-426

140. HOUSOS E.C., IRISARRI G.D.:"A sparse variable metric optimization method applied to the solution of power system problems", IEEE Trans., PAS, 101, 1, Jan. 1982, pp. 196-202

141. HUBBI W., REFSUM A.:"Starting algorithm and modification for Newton-Raphson load-flow methods", Jour. Elect. Power and Energy Syst., 5, 3, July 1983, pp. 166-172

142. HUNTINGTON R.:"Resurrection of water supply - New life for a faithful servant", Jour. Inst. Water Engrs. and Scientists, 33, 6, 1979, p. 497

143. IRISARRI G., HODGES S.F.:"An optimal ordering algorithm for sparse matrix applications", IEEE Trans., PAS, 97, 6, Nov-Dec. 1978, pp. 2253-2261

144. IRVING M.R.:"Decomposed estimation and control of electric power systems", PhD Thesis, Dept. of Contr. Engrg., University of Sheffield, Nov. 1977

145. IRVING M.R., OWEN R.C., STERLING M.J.H.:"Power-system state estimation using linear programming", IEE Proc., 125, 9, Sept. 1978, pp. 879-885

146. IRVING M.R., STERLING M.J.H.:"Substation data validation", IEE Proc., 129, 3, C, 1982, pp. 119-121

147. IRVING M.R., STERLING M.J.H.:"Economic dispatch of active power with constraint relaxation", IEE Proc., 130, C, 4, July 1983, pp. 172-176

148. IRVING M.R., STERLING M.J.H.:"Implementation of state estimation and data validation in electric power systems", BUNN D.W., FARMER E.D., (Eds), Short term forecasting, J. Wiley and Sons Ltd., 1984

149. IRVING M.R., STERLING M.J.H.: "Economic dispatch of active power by quadratic programming using sparse linear complementary algorithm", accepted for publication in Jour. Elect. Power and Energy Syst.

150. JOHNSON C.S.:" Telemetry data systems: parts 1 and 2", Instrument Technology, Aug. & Oct. 1976

151. KANBAYASHI T.,NAGAI M., YAMANAKA K., OKAMOTO M., SUZUKI T., "Computer control and operation information systems for the large water supply systems",Prog.Wat.Tech.,9, 5/6, 1977, pp.299-305

152. KENNINGTON J.L., HELGASON R.V.;Algorithms for network programming, J. Wiley and Sons Ltd, N.Y., 1980

153. KESAVAN H.K., CHANDRASHEKAR M.:"Graph-theoretic models for pipe network analysis", Jour. Hydraul. Div. ASCE, HY2, Feb. 1972, pp. 345-364

154. KOGLIN H.J.:"Optimal measuring system for state estimation", Proc. 5th PSC Conf., Cambridge, 1975

155. KOSONEN H., LINDSTROM K., NILSSON B., SCHUBERT J.: Advanced real-time modelling of the Finnish high-voltage network", Proc. 7th PSC Conf., Lousanne, July 1981

156. KRUMPHOLZ G.R., CLEMENTS K.A., DAVIS P.W.:"Power system observability: a practical algorithm using network topology", IEEE Trans. PAS, 99, 4, Jul-Aug. 1980, pp. 1534-1542

157. KUHN A.W.:"The Hungarian method for the assignment problem", Naval Res. Log. Quart., 2, 1955, p. 83

158. KURANZ J.H.: "Waukesha, Wis.", Jour. AWWA, 59, 1, Jan. 1967, p. 93

159. LAM C.F., WOLLA M.L.:"Computer analysis of water distribution systems: part 1 - Formulation of equations", Jour. Hydraul. Div. ASCE, HY2, Feb. 1972, pp. 335-344

160. LAM C.F., WOLLA M.L.:"Computer analysis of water distribution systems: part 2 - Numerical solution." Jour. Hydraul. Div. ASCE,HY3,Mar. 1972, pp. 447-460

161. LARSON R.E., TINNEY W.F., PESCHON J.:"State estimation in power systems: part 1 - Theory and feasibility", IEEE Trans. PAS, 89, 3, March 1970, pp. 345-352

162. LARSON R.E.,TINNEY W.F., HAJDU L.P., PIERCY D.S.: "State estimation in power systems: part 2 - Implementation and applications", IEEE Trans. PAS, 89, 3, March 1970, pp. 353-363

163. LAUER G.S.,BERTSEKAS D.P., SANDELL N.R., POSBERGH : "Solution of large-scale optimal unit commitment problems", IEEE Trans., PAS, 101, 1, Jan 1982

164. LEMIEUX P.F.:"Efficient algorithm for distribution networks", Jour. Hydraul. Div. ASCE, HY11, May 1972 pp. 1911-1920

165. LIU K.T.H.:"Pipe network analysis in integrated civil engineering systems (ICES)", Rep. 320-2031, IBM Cambridge Scientific Centre, Cambridge, Mass., Aug. 1968

166. LO K.L.,ONG P.S.,McCOLL R.D.,MOFFATT A.M.,SULLEY J. MOYES I.: "Power system state estimation: a method for metering reinforcement", IEEE Trans. PAS, 101, 9, Sept. 1982, pp. 2486-2491

167. LO K.L.,ONG P.S.,McCOLL R.D.,MOFFATT A.M.,SULLEY J. "Development of a static state estimator part 2: bad data replacement and generation of pseudo-measurements", IEEE Trans. PAS, 102, 8, Aug.1983, pp. 2492-2500

168. LUGTU R.L., HACKETT D.F., LIU K.C., MIGHT D.D.:"Power system state estimation: detection of topological errors", IEEE Trans. PAS, 99, 6, Nov/Dec. 1980, pp. 2406-2412

169. LUSCOMBE J.S.:"Computer-based process control", Jour. AWWA, March 1979, pp. 127-132

170. MAILANDER F., POTRZ W.J.:"Optimale Steuerung der Wasserverteilung in Verbundsystemen", GWF - Wasser/AbWasser, 122, 7, 1981, pp. 291-297

171. MAMANDUR K.R.C., BERG G.J.:"Efficient simulation of line and transformer outages in power systems", IEEE Trans. PAS, 101, 10, Oct. 1982, pp. 3733-3741

172. MARLOW K.C.,FALLSIDE F.:"Minicomputer, microprocessor and telecontrol applications to a water supply network", Jour. Inst. Wat. Engrs. and Scienticts, 34, 6, 1980, pp. 517-545

173. MARLOW K. C., FALLSIDE F.:"More on planned strategy for water telemetry", Contr. and Instrum., 1, Jan. 1981, pp. 47 & 51

174. MARLOW T.A., HARDISON R.L.,JACOBSON H., BIGGS G.E.: "Improved design of fluid network with computers", Jour. Hydraul. Div. ASCE, 4, HY4, July 1966, pp. 43-61

175. MARTIN D.W., PETERS G.,"The application of Newton's method to network analysis by digital computer", Jour. Inst. Wtr. Engrs., 17, 2, 1963, p. 115

176. MASIELLO R.D., SCHWEPPE F.C.:"A tracking static-state estimator", IEEE Trans. PAS, 90, Mar-Apr. 1971, pp. 1025-1033

177. MATSUMOTO K., HATAKEYAMA J.,SHIMAUCHI S.,TANIGUCHI H. MATSUDA T., HASEGAWA S.: "Control models for the computer operation of large water supply systems", Progr. Wtr. Tech., 9, 5/6, 1977, pp. 565-569

178. MERRILL H.M., SCHWEPPE F.C.: "Bad data suppression in power system static state estimation", IEEE Trans., PAS, 90, 12, Nov-Dec. 1971, ,pp. 2718-2725

179. McCORMICK M., BELLAMY B.E.:"A computer program for the analysis of networks of pipes and pumps", Jour. Instn. Engrs. Aust., 38, 3, March 1968, pp. 51-58

180. McGINNIS L.:"Implementation and testing of a primal-dual algorithm for the assignment problem", Operations Research, 31, 2, Mar-Apr. 1983, pp. 277-290

181. McPHERSON M.B.:"Generalised distribution network head-loss characteristics", Trans. ASCE, 126, pp. 1190-1206

182. McPHERSON M.B.: "System operation, control and monitoring: part 1 - Fundamentals", Water Supply Engrs. Conf., Univ. of Ill., College of Engrg., Urbana, Ill., Apr. 1975, pp. 49-58

183. MICHEL H.L., WOLFNER J.P.:"Know what's happening in your water system", The American City., 87, 6, June 1972, p. 80

184. MILLER B.M., FRIEDGEN R.D., KOCH W.E.:"Telemetering at Helix irrigation district, La Mesa, Calif.", Jour. AWWA, 63, 8, Aug. 1971, p. 501

185. MOSS S.M.: On-line optimal control of water supply network, PhD Thesis, Dept. of Engrg., University of Cambridge, Sept. 1979

186. MONTICELLI A., GARCIA A.: "Reliable bad data processing for real-time state estimation", IEEE Trans. PAS, 102, 5, May 1983, pp. 1126-1139

187. MUCKSTADT J., KOENIG S.A.: "An application of Lagrangian relaxation to scheduling in power-generation system",Operations Research,25, 3, May-June 1977, pp. 387-403

188. NAGENDRA RAO P.S., PRAKASA RAO K.S., NANDA J.: "An exact fast load flow method including second order terms in rectangular coordinates", IEEE Trans. PAS, 101, 9, Sept. 1982, pp. 3261-3268

189. NAHAVANDI A.N., CATANZARO G.V.: "Matrix methods for analysis of hydraulic networks", Jour. Hydraul. Div. ASCE, 99, HY1, Jan. 1973, pp. 47-63

190. NIESSNER H., REICHERT K.: "On computing the inverse of a sparse matrix", Int. Jour. for Numer. Methods in Engrg., 19, 1983, pp. 1513-1526

191. NISHIYA K.I., HASEGAWA J.: "Dynamic state estimation including identification of network configuration changes of electric power system", Elect. Engrg. in Japan, 101, 4, 1981, pp. 69-77

192. NISHIYA K., HASEGAWA J., KOIKE T.: "Dynamic state estimation including anomaly detection and identification for power systems", IEE Proc., 129, C, 5, Sept., 1982, pp. 192-198

193. PAGNOTTO V.A.: "Computer application in distribution system operations", Jour. AWWA, 65, 1, Jan. 1973, p. 134

194. PERRY P.F.: An on-line optimal control scheme for water supply networks, PhD Thesis, Dept. of Engrg., Univ. of Cambridge, July 1975

195. PESCHON J, BREE D.W., HAJDU L.P.: "Optimal power-flow solutions for power system planning", Proc. IEEE, 60, 1, Jan. 1972, 64-70

196. PETERS G., WILKINSON J.H.: "The least square problem and pseudo-inverses", The Computer Jour., 13, 3, Aug. 1970, pp. 309-316

197. PHUA K., DILLON T.S.: "Optimal choice of measurements for state estimation", PICA Conf. Proc., Toronto, Ontario, May 1977, pp. 431-441

198. PITCHAI R., A model for designing water distribution pipe networks, PhD Thesis, Harvard Univ., Cambridge, Mass., Jun. 1966

199. QUINTANA V.H.,SIMOES-COSTA A.S., MANDEL A.: "Power system topological observability using a direct graph-theoretic approach",IEEE Trans., PAS, 101, 3, March 1982, pp. 617-625

200. QUINTANA V.H., SIMOES-COSTA A., MIER M.: "Bad data detection and identification techniques using estimation orthogonal methods", IEEE Trans. PAS, 101, 9, Sept. 1982, pp. 3356-3364

201. RAHAL C.M.: Parameter estimation for water distribution network modelling and control, PhD Thesis, Dept. of Contr. Engrg., University of Sheffield, April 1980

202. RAHAL C.M., STERLING M.J.H.: "Dynamic parameter tuning for water distribution network models in an extended period simulation", Proc. Instn. Civ. Engrs., 71, 2, March 1981, pp. 151-164

203. RAHAL C.M., STERLING M.J.H., COULBECK B.: "Parameter tuning for simulation models of water distribution networks", Proc. Instn. Civ. Engrs., Sept. 1980, pp. 751-762

204. RAO H.S., BREE D.W.: "Extended period-simulation of water systems: part A", Jour. Hydr. Div., ASCE, 103,HY2, Feb. 1977, pp. 97-108

205. RAO H.S., BREE D.W., BENZVI R.: "Extended period simulation of water distribution networks", Doc. No. PB-230 148, Palo Alto Calif., Feb. 1974

206. RAO H.S., SEITLE R.A.: "Computer applications in urban water distribution system control", Jour. of Dynamic Syst. Measur. and Control, ASME, June 1975, pp. 117-125

207. RAO H.S., MARKEL L.C., BREE D.W.: "Extended-period simulation of water systems: part B", Jour. Hydr. Div., ASCE, 103, HY3, March 1977, pp. 281-294

208. RAO N.D., TRIPATHY S.C.: "Power system static state estimation by the Levenberg-Marquardt algorithm", IEEE Trans., PAS, 99, 2, Mar-Apr. 1980, pp. 695-702

209. RAO N.D., ROY L.: "A Cartesian coordinate algorithm for power system state estimation", IEEE Trans., PAS, 102, 5, May 1983, pp. 1070-1082

210. REID J.K.: "A sparsity-exploiting variant of the Bartels-Golub decomposition for linear programming bases", A.E.R.E., Harwell, England, Harwell Report CSS20, Aug. 1975

211. REID J.K. (Ed), "Large sparse sets of linear equations", IMA Conf., London, Apr. 1970

212. ROBERTS P.D., ROBERTS S.S.: "Algorithm 458: discrete linear L_1 approximation by interval linear programming", CACM, 16, 1973, pp. 629-631

213. ROBERTSON E.D.: "Applied organization and management in water supply development in Third World ", H_2O, 14, 21, 1981, pp. 501-508

214. ROBINSON P.M.: "Telemetering in San Jose, Calif.", Jour. AWWA, 63, 8, Aug. 1971, p. 499

215. ROSTRON J.T.: "Telemetering: Los Angeles county, Calif.", Jour. AWWA, 63, 8, Aug. 1971, p. 504

216. ROY L., RAO N.D.: "A new algorithm for real-time optimal dispatch of active and reactive power generation retaining nonlinearity", IEEE Trans., PAS, 102, 4, Apr. 1983, pp. 832-840

217. ROY L., RAO N.D.: "Real-time monitoring of power systems using fast second-order method", IEE Proc., 130, C, 3, May 1983, pp. 103-110

218. SAWICKI J., WILKOSZ K., KREMENS Z.: "Fast method for identification of bad data of active power flows in power network", IFAC Symp., Florence 1983

219. SCHUBERT L.K.: "Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian", Mathematics of Computation, 24, 1970

220. SCHWEPPE F.C., WILDES J.: "Power system static state estimation: part 1 - Exact model", IEEE Trans. PAS, 89, 1, Jan. 1970, pp. 120-124

221. SCHWEPPE F.C., ROM D.B.: "Power system static state estimation, part 2 - Approximate model", IEEE Trans. PAS, 89, 1, Jan. 1970, pp. 125-130

222. SCHWEPPE F.C.: "Power system state estimation: part 3 - Implementation", IEEE Trans. PAS, 89, 1, Jan. 1970, pp. 130-135

223. SCHWEPPE F.C., HANDSCHIN L.J.: "Static state estimation in electric power systems", Proc. IEEE, 62, 7, July 1974, pp. 972-982

224. SEIDLER J., BADACH A., MOLISZ W.: Methods of solving optimization problems, (in Polish), W.N.T., Warsaw, 1980

225. SEKINE Y., SUZUKI H.: "A method for estimating network configuration of power system", Proc. 5th PSC Conf., Cambridge, 1975

226. SHAMIR U., HOWARD C.D.D.: "Water distribution systems analysis", Jour. Hydraul. Div., ASCE, 94, HY1, 1968, pp. 219-234

227. SHAMIR U.: "Water distribution systems analysis", IBM Research Report RC4389, Yorktown Heights, N.Y., June 1973

228. SHAMIR U.: "Optimal design and operation of water distribution systems", Water Res. Research, 10, 1, Feb. 1974, pp. 27-36

229. SHAMIR U., HOWARD C.D.D.: "Engineering and analysis of water distribution systems", Jour. AWWA, 68, 9, Sept. 1977, p. 510

230. SHANNO D.F.: "On variable-metric methods for sparse Hessians", Mathematics of Computation, 34, 150, Apr. 1980, pp. 499-514

231. SHIRLEY W.O., BAILEY J.J.: "Use of digital computers in distribution system analyses", Jour. AWWA, 58, 2, 1966, pp. 1575-1584

232. SIEGEL I.G.: "Deferment of computation in the method of least squares", Mathematics of Computation, 19, 1965, pp. 329-331

233. SIMOES-COSTA A., QUINTANA V.H.: "A robust numerical technique for power system state estimation", IEEE PES, Summer Meeting, Minneapolis, July 1980

234. SIMOES-COSTA A., SALGADO R.: "Bada data recovery for orthogonal row processing state estimators", IFAC Symp., Florence 1983

235. SMITH O.J.: "Power system state estimation", IEEE Trans. PAS, 89, 3, March 1970, pp. 363-379

236. SMITH O.J., MAKANI K., KRISHNA L.: "Sparse solutions using hash storage", IEEE Trans. PAS, 91, Jul-Aug. 1972, pp. 1396-1404

237. SNYDER W.L., SASSON A.M.: "Security load flow employing a modified decoupled Hessian technique", PICA Conf., 1977

238. STAG G.W., DOPAZO J.F., KLITIN O.A., VAN SLYCK L.S.: "Techniques for the real-time monitoring of power system operations", IEEE Trans. PAS, 85, 4, Apr. 1970, pp. 545-555

239. SPOSITO V.A.: Linear and nonlinear programming, Iowa State University Press, 1975

240. SQUIERS K.M.: "Pumping and storage control in water supply systems", Proc. 7th Sanit. Engrg. Conf., Univ. of Illinois, Urbana, 1967

241. STEELE K.A.: "Optimization of energy in water distribution systems by effective measurement and control", Water (Australia), 8, 3, 1981, pp. 20-22

242. STERLING M.J.H.: Power system control, Peter Peregrinus Ltd., 1978

243. STERLING M.J.H., ANTCLIFFE D.J.: "A technique for the prediction of water demand from past consumption data", Jour. Inst. Wtr. Engrs., 28, 8, Nov. 1974, pp. 413-420

244. STERLING M.J.H., BARGIELA A.: "Detection of water leakage and measurement errors in distribution networks", Computer Appl. for Water Supply and Distrib. Conf., Leicester, March 1983

245. STERLING M.J.H., BARGIELA A., "An adaptive forecasting technique based on time series modelling", IEE Colloquium on Time Series Analysis and Forecasting Methods, London, Nov. 1983

246. STERLING M.J.H., BARGIELA A.: "The enhancement of time series modelling techniques for adaptive forecasting", BUNN D.W., FARMER E.D., (Eds), Short-term forecasting, J. Wiley and Sons Ltd., 1984

247. STERLING M.J.H., BARGIELA A.: "Minimum norm state estimation for computer control of water distribution systems", IEE Proc., 131, D, 2, March, 1984, pp. 57-63

248. STERLING M.J.H., BARGIELA A.: "Leakage reduction by optimised control of valves in water networks", submitted for publication in: Trans. Instn. Meas. and Contr.

249. STERLING M.J.H., BARGIELA A.: "Detection of topological errors in water distribution systems", accepted for publication, Jour. Instn. Meas. and Contr.

250. STERLING M.J.H., COULBECK B.: "A dynamic programming solution to optimisation of pumping costs", Proc. Instn. Civ. Engrs., 59, 2, Dec. 1975, pp. 813-818

251. STERLING M.J.H., COULBECK B.: "Optimisation of water pumping costs by hierarchical methods", Proc. Instn. Civ. Engrs., 59, 2, Dec. 1975, pp. 789-797

252. STERLING M.J.H., COULBECK B.: "Linear dynamic models in water systems", Proc. Instn. Civ. Engrs., 61, 2, Sept. 1976, p. 582

253. STERLING M.J.H., COULBECK B.: "Analysis of pump interaction and efficiency in control of water supply", Trans. Inst. Meas. and Contr., 1, 2, Apr-June 1979, pp. 79-84

254. STERLING M.J.H., IRVING M.R.: "Constrained dispatch of active power by linear decomposition", IEE Proc., 124, 3, March 1977, pp. 247-251

255. STERLING M.J.H., IRVING M.R., "Integrated software for power system control", Internal Report, Dept. Engrg., Univ. of Durham, 1983

256. STIMSON K.R.: "Water distribution network analysis", Water System Control Conf., Leicester, 1981

257. STIMSON K.R., BRAMELLER A.: "An integrated mesh-nodal method for steady state water distribution network analysis", Jour. Inst. Wtr. Engrs. and Scientists, 35, 2, 1981, p. 186

258. STIMSON K.R., BRAMELLER A.: "Water distribution network analysis using interactive graphics", Jour. IWES, 36, 2, Apr. 1982, pp. 154-157

259. SYDER W.L., SASSON A.M.: "Security flow employing a modified decoupled Hessian technique", PICA Proc., Toronto, Ontario, May 1977, pp. 442-450

260. TARJAN R.E.: "Testing graph connectivity", Proc. 6th Annual Symp. on Theory of Comp., Apr. 30- May 2, 1974, pp. 185-193

261. TARJAN R.E.: "Depth-first search and linear graph algorithms", in: Large-scale networks: theory and design, IEEE Press, New York, 1976

262. TART J.S.: "Value of telemetering and recording", Jour. AWWA, 65, 2, Feb. 1973, p. 123

263. THACKRAY J.E., ARCHIBALD G.G.: "The Severn-Trent studies of industrial water use", Proc. Inst. Civ. Engrs., 70, 1, Aug. 1981, pp. 403-432

264. TINNEY W.F., WALKER J.W.: "Direct solution of sparse network equations by optimally ordered triangular factorisation", Proc. IEEE, 55, 11, Nov. 1967, pp. 1801-1809

265. TINNEY W.F., PETERSON M.M.: "Steady state security monitoring", HANDSCHIN E. (Ed), Real time control of electric power systems, Elsevier, 1972

266. TINNEY W.F., MEYER W.S.: "Solution of large sparse systems by ordered triangular factorisation", IEEE Trans. Aut. Contr., 18, 4, Aug. 1973, pp. 333-346

267. TINNEY W.F., POWELL W.L.: "The REI approach to power network equivalents", PICA Conf. Proc., Toronto, Ontario, May 1977, pp. 314-324

268. TOMLIN J.A.: "On scaling linear programming problems", BALINSKI (Ed), Mathematical programming study 4, North Holland Publ. Co., 1975, pp. 146-166

269. TO-YAT C.: "Computational comparison of eight methods for the maximum network flow problem", ACM Trans. on Mathematical Software, 6, 1, March 1980, pp. 1-16

270. TOYODA J., CHEN M., INOUE Y.: "An application of state estimation to short term forecasting", IEEE Trans. PAS, 89, Sept-Oct. 1970, pp. 1678-1682 and 1683-1688

271. TROTT W.J., YEH W.W.G.: "Optimisation of multiple reservoir system", Jour. Hydraul. Div. ASCE, HY10, Oct. 1973, pp. 1865-1884

272. TURGEON A.: "Optimal short-term hydro scheduling from the principle of progressive optimality", WRR, 17, 3, June 1981, pp. 481-486

273. VAN CUTSEM Th.: "Estimation d'etat statique dans les reseaux electriques: analyse de l'observabilite", (in French), Internal Report, University of Liege, July 1982

274. VAN CUTSEM Th., GAILLY P.J.: "A simple algorithm for power system observability analysis and related functions", IFAC Symp., Florence, 1983

275. VAN DYKE R.P.:"Computerized operation of distribution systems", Jour.AWWA, 62, 4, Apr. 1970, p. 234

276. VAN LOON R.E.: "Telemetering in Sunnyvale, Calif.", Jour. AWWA, 63, 8, Aug. 1971, p. 491

277. VAN OLINDA P., WRUBEL J.N., WOLLENBERG B.F., WODZELL G.W.:"Installation and start up of a on-line estimator", IEEE Trans. PAS, 100, 11, Nov. 1981

278. VON RADER J.E.: "Telemetric control of water distribution systems", Jour. AWWA, Nov. 1983, pp. 542-549

279. WALLACH Y., HANDSCHIN E.: "An efficient parallel processing method for power system state estimation IEEE Trans. PAS, 100, 11, Nov. 1981, pp. 4402-4406

280. WALSH J.(Ed), Numerical analysis: an introduction, Academic Press, London, 1966

281. WATANATACA T.: "Least-cost design of water distribution systems", Jour. Hydr. Div. ASCE, 99, HY9, Sept. 1973, p. 1974

282. WILLIAMS J.D., PESTKOWSKI T.: "Distribution system analysis by digital computer", Jour. Instn. Wtr. Engrs., 26, 7, Oct. 1972, pp. 381-391

283. WILSON R.J., BEINEKE L.W.: Applications of graph theory, Academic Press, London, 1979

284. ZARGHAMEE M.S.: "Mathematical model for water distribution systems", Jour. Hydr. Div., ASCE, 97, HYL, Jan, 1971, pp. 1-14

285. ZOLLENKOPF K.: "Bi-factorisation. Basic computational algorithm and programming techniques", REID J.K. (Ed), Large sparse sets of linear equations, Academic Press, London, 1971

APPENDIX A

MEASUREMENT AND NETWORK MODEL

A.1 THE NONLINEAR MEASUREMENT MODEL

Strictly speaking, a water distribution system model should represent the full dynamic behaviour since consumer loads, system supplies and reservoir volumes are continuously varying in time. However, even under normal operating conditions the formulation of a valid and feasible dynamic model for a water system is a very difficult task. To overcome this problem, one makes use of the fact that under normal operating conditions the variations in the system state are slow. Therefore, it is reasonable to consider that the water system is in steady state for a limited time interval. With this assumption, a static model for the water system can be formulated.

The quantities which are usually monitored in a water system are: i) nodal heads; ii) fixed-head-node flows; iii) consumer loads and iv) pipe flows. Let $\underline{z}$ denote the mx1 vector of measurements taken in a N-node water system with F fixed-head-nodes, and assume that $\underline{z}^o$ is the vector of the actual values for the respective measured quantities. In practice, the elements of the vector $\underline{z}$ differ from the

corresponding elements of $\underline{z}^o$. Meter inaccuracies, instrumentation malfunction, communication errors in transmitting the measurements to the control centre, effects of analog to digital conversions, etc. make up the difference between $\underline{z}$ and $\underline{z}^o$. This difference is of random nature, and it is modelled as a zero-mean random variable ω with covariance matrix R. Hence, the measurement model can be written as

$$\underline{z} = \underline{z}^o + \underline{\omega} \qquad (A.1)$$

$$E\{\underline{\omega}\} = \underline{0} \; ; \quad E\{\underline{\omega}\,\underline{\omega}^T\} = R \qquad (A.2)$$

where $E\{\ \}$ is the expectation operator.

Assuming that the pipe and other network element parameters are known, all the measurable quantities can be expressed as non-linear functions of the nodal heads and fixed-head-node flows, which are the <u>state variables</u> of the system. Define n = N + F. The <u>state vector</u> is an nx1 vector $\underline{x}$ whose first N elements are the nodal heads and the remaining F elements are the fixed-head-node flows. The non-linear vector function $\underline{g}(\ )$ which relates $\underline{z}^o$ to $\underline{x}$ is based on Kirchhoff's laws, and depends on the network structure and parameters. The non-linear measurement model is then expressed as

$$\underline{z} = \underline{g}(\underline{x}) + \underline{\omega} \qquad (A.3)$$

$$E\{\underline{\omega}\} = \underline{0} \; ; \quad E\{\underline{\omega}\,\underline{\omega}^T\} = R \qquad (A.4)$$

where

$\underline{z}$: mx1 vector of measurements

$\underline{g}(.)$:mx1 nonlinear vector function relating the measured quantities and the states

$\underline{x}$:nx1 state vector, formed by N nodal heads and F fixed-head-node flows

$\underline{\omega}$:mx1 zero-mean random vector which models the measurement errors

R : mxm covariance matrix of the measurement errors

A.2 THE LINEARISED MODEL

Consider a given point $\underline{x}_k$ of the state space around which the linearisation of the model given by Eq. (A.3) is to be performed. A first-order Taylor series approximation for the nonlinear vector function $\underline{g}(\ )$ is given by

$$\underline{g}(\underline{x}) = \underline{g}(\underline{x}_k) + \left.\frac{\delta \underline{g}(\underline{x})}{\delta \underline{x}}\right|_{\underline{x}=\underline{x}_k} \cdot (\underline{x}-\underline{x}_k) \qquad (A.5)$$

In addition, define

$$J(\underline{x}_k) = \left.\frac{\delta \underline{g}(\underline{x})}{\delta \underline{x}}\right|_{\underline{x}=\underline{x}_k} \qquad (A.6)$$

mxn Jacobian matrix and

$$\Delta \underline{x} = \underline{x} - \underline{x}_k \qquad (A.7)$$

Therefore, Eq. (A.5) may be written as

$$\underline{g}(\underline{x}) = \underline{g}(\underline{x}_k) + J(\underline{x}_k)\cdot\Delta\underline{x} \qquad (A.8)$$

By substituting Eq (A.8) to Eq (A.3) and rewriting Eq (A.4), the linearised model is finally expressed

$$\Delta\underline{z} = J(\underline{x}_k)\cdot\Delta\underline{x} + \underline{\omega} \qquad (A.9)$$

$$E\{\underline{\omega}\} = \underline{0} \;\; ; \;\; E\{\underline{\omega}\,\underline{\omega}^T\} = R \qquad (A.10)$$

where

$$\Delta\underline{z} = \underline{z} - \underline{g}(\underline{x}_k) \qquad (A.11)$$

A.3 HEAD-FLOW RELATIONSHIPS OF NETWORK ELEMENTS

There is a wide variety of network elements existing in a water distribution system. However, for the purpose of the network modelling it is sufficient to consider only several types of elements which have distinctly different characteristics. For example, there is no need to devise separate models for different control valves since, for the purpose of network flow calculations, the adjustment of a valve resistance gives sufficiently good approximation.

The following elements are usually employed to build up a network model [205].

A.3.1 Pipes

The head loss characteristics of a pipe between nodes i and j depends on the hydraulic resistance between the nodes and can be modelled as follows:

$$q_{ij} = 0.27746\ CHW_{ij}\ D_{ij}^{2.63} \frac{|h_j - h_i|^{0.54}}{L_{ij}} \qquad (A.12)$$

where q_{ij} : flow from node j to node i (m^3/s)

CHW_{ij} : Hazen-Williams coefficient for pipe

D_{ij} : diameter of pipe (cm)

L_{ij} : length of pipe (m)

h_j : head at node j (mAq)

h_i : head at node i (mAq)

This is usually used in the following form to give a consistent sign for flow as:

$$q_{ij} = r_{ij}^{-0.54} (h_j - h_i)\ |h_j - h_i|^{-0.46} \qquad (A.13)$$

where r_{ij} is the resistance between nodes i and j given by

$$r_{ij} = 10.742\ CHW_{ij}^{-1.85}\ L_{ij}\ D_{ij}^{-4.87} \qquad (A.14)$$

A.3.2 Parabolic Pumps

The pump characteristic can be approximated by a second order polynomial

$$h_i - h_j = a\, q_{ij}^{\,2} + b\, q_{ij} + c \qquad (A.15)$$

where a, b and c are empirically determined constants. Thus

$$q_{ij} = \frac{-b \pm \{b^2 - 4a(c - |h_i - h_j|)\}^{0.5}}{2a}$$

taking the positive root for constant 'a' positive and vice versa, and setting q_{ij} to zero for $h_i - h_j > c$ or for negative or equal to zero value of $\{b^2 - 4a(c - |h_i - h_j|)\}$

A.3.3 Pressure Reducing Valves

These may be modelled by assuming that between nodes i and j there is a valve with a setting equal to H_{PRV}. If $h_j \geq H_{PRV} \geq h_i$ the valve reduces the head to H_{PRV} to give a head drop of $(H_{PRV} - h_i)$ and flow takes place from j to i given by:

$$q_{ij} = r_{ij}^{-0.54} |H_{PRV} - h_i|^{0.54} \qquad (A.16)$$

If $h_i > H_{PRV}$ the valve shuts off, no reverse flow takes place and hence $q_{ij} = 0$

If $h_j < H_{PRV}$ and $h_i < H_{PRV}$ the valve acts as a pipe with a head drop of (h_j-h_i).

A.3.4 Non-Return Valves

In a pipe fitted with a non-return valve, the loss in head due to the valve itself is usually small and may be either neglected or included in the pipe resistance, thus

$$q_{ij} = \begin{cases} r_{ij}^{-0.54} (h_j-h_i) |h_j-h_i|^{-0.46} & h_j > h_i \\ 0 & h_j \leq h_i \end{cases} \quad (A.17)$$

A.3.5 Control Valves

These may be manually or automatically controlled and are currently modelled for both non-return and two-way valves by assuming that control varies the resistance of the equivalent pipe to give

$$q_{ij} = r_{ij}^{-0.54} (h_j-h_i) |h_j-h_i|^{-0.46} \quad (A.18)$$

where r_{ij} is now the independent valve control parameter which can vary from $r_{ij\ MIN}$ for a valve fully open to $r_{ij\ MAX}$ for a valve fully closed

A.3.6. Fixed-Head Nodes

These nodes correspond to reservoirs or boreholes (feeding head dependent pumps, valves, etc) with fixed or known head variation. The magnitude of in/out-flow at these nodes q_i does not depend on the nodal head h_i and is considered as a separate state variable.

A.4 MEASUREMENT SIMULATION

In section A.1, the measurement errors have been modelled as a zero-mean random variable with covariance matrix R. The measurement errors are usually assumed to be uncorrelated, which is equivalent to saying that matrix R is considered diagonal. Each diagonal element of R corresponds to the variance of the respective measurement error.

In this work, the measurement errors are simulated according to the formula

$$z_i = z_i^o + \alpha \; k_i \; z_i^o \qquad \text{(A.19)}$$

where

z_i : i-th measurement;

z_i^o : true value of the i-th measured quantity; obtained from a simulator;

α : standard normal random variable N(0,1),

k_i : constant representing meter accuracy.

The variance of the i-th measurements σ_i^2, is given by

$$\sigma_i^2 = (\alpha \cdot k_i \cdot z_i^o)^2 \qquad \text{(A.20)}$$

APPENDIX B

JACOBIAN MATRIX TERMS

B.1 MEASURABLE QUANTITIES

This appendix describes the equations to compute the elements of the Jacobian matrix used in the water system state estimation problem. The measurable quantities are the nodal heads, H_i, the fixed-head-node flows, Q_i, the consumer loads, L_i, and the pipe flows F_{ij}. Using the notation introduced in Appendix A these quantities may be calculated as

$$H_i = h_i \tag{B.1}$$

$$Q_i = q_i \tag{B.2}$$

$$L_i = q_i + \sum_{j \in M_i} q_{ij}(h_i, h_j) \tag{B.3}$$

$$F_{ij} = q_{ij}(h_i, h_j) \tag{B.4}$$

where M_i denotes the set of nodes which are incident to the node i.

B.2 JACOBIAN MATRIX TERMS

The Jacobian matrix used in the linearised measurement model (Eq. A.9) can be partitioned according to the type of measurements

$$J = \begin{bmatrix} \frac{\delta \underline{H}}{\delta \underline{h}} & 0 \\ 0 & \frac{\delta \underline{Q}}{\delta \underline{q}} \\ \frac{\delta \underline{L}}{\delta \underline{h}} & \frac{\delta \underline{L}}{\delta \underline{q}} \\ \frac{\delta \underline{F}}{\delta \underline{h}} & 0 \end{bmatrix} \qquad \text{(B.5)}$$

where $\underline{H}$, $\underline{Q}$, $\underline{L}$, $\underline{F}$ are the vectors of measurable quantities, and $\underline{h}, \underline{q}$ are the components of the state vector

The elements of matrix J in Eq (B.5) can be expressed in terms of the quantities defined in Eqs. (B.1) - (B.4)

$$\frac{\delta H_i}{\delta h_i} = 1 \qquad \text{(B.6)}$$

$$\frac{\delta Q_i}{\delta q_i} = 1 \qquad \text{(B.7)}$$

$$\frac{\delta L_i}{\delta h_i} = \sum_{j \epsilon M_i} \frac{\delta q_{ij}(h_i, h_j)}{\delta h_i} \qquad \text{(B.8)}$$

$$\frac{\delta L_i}{\delta h_j} = -\frac{\delta q_{ij}(h_i,h_j)}{\delta h_j} \qquad (B.9)$$

$$\frac{\delta L_i}{\delta q_j} = \begin{cases} 1 & \text{if } i=j \text{ is a fixed-head node} \\ 0 & \text{otherwise} \end{cases} \qquad (B.10)$$

$$\frac{\delta F_{ij}}{\delta h_i} = \frac{\delta q_{ij}(h_i,h_j)}{\delta h_i} \qquad (B.11)$$

$$\frac{\delta F_{ij}}{\delta h_j} = -\frac{\delta q_{ij}(h_i,h_j)}{\delta h_j} \qquad (B.12)$$

Taking into account Eqs.(A.12) - (A.18) the partial derivatives of elements flow can be calculated

- for pipes and control valves

$$\frac{\delta q_{ij}(h_i,h_j)}{\delta h_i} = -0.54\, r_{ij}^{-0.54} |h_j-h_i|^{-0.46} \qquad (B.14)$$

- for parabolic pumps

$$\frac{\delta q_{ij}(h_i,h_j)}{\delta h_i} = \pm \left\{ b^2-4a(c-|h_i-h_j|) \right\}^{-0.5} \qquad (B.14)$$

- for pressure reducing valves (if $h_j \geq H_{PRV} \geq h_i$)

$$\frac{\delta q_{ij}(h_i,h_j)}{\delta h_i} = -0.54\, r_{ij}^{-0.54} |H_{PRV}-h_i|^{-0.46} \qquad (B.15a)$$

- for pressure reducing valves (otherwise)

$$\frac{\delta q_{ij}(h_i,h_j)}{\delta h_i} = 0 \qquad \text{(B.15b)}$$

- for non-return valves (if $h_j > h_i$)

$$\frac{\delta q_{ij}(h_i,h_j)}{\delta h_i} = -0.54\, r_{ij}^{-0.54} |h_j - h_i|^{-0.46} \qquad \text{(B.16a)}$$

- for non-return valves (if $h_j \leq h_i$)

$$\frac{\delta q_{ij}(h_i,h_j)}{\delta h_i} = 0 \qquad \text{(B.16b)}$$

Equations (B.13) - (B.16) show that the elements of the Jacobian matrix can be obtained using the same expressions that appear in the computation of the measurable quantities. By recognizing this fact, one can get noticeable savings of computing time in the implementation of the state estimators.

APPENDIX C

TEST NETWORK PARAMETERS

C.1 11-NODE TEST WATER SYSTEM

The 11-node water distribution system of Fig. C.1. with 13 pipe-lines, two parabolic pumps, one fixed-head pump, and three fixed-head-nodes has been used to perform preliminary tests on estimation and bad data processing methods developed in this work. The system has been taken from reference [204], and its parameters are displayed in Table C.1.

C.2 34-NODE SYSTEM (DONCASTER AND THORNE ZONES)

The realistic 34-node system representing part of the Yorkshire Water Authority network is depicted in Fig. C.2. This system has been previously used in water system control studies [63], [202]. Its pipe and pump parameters are given in Table C.2.

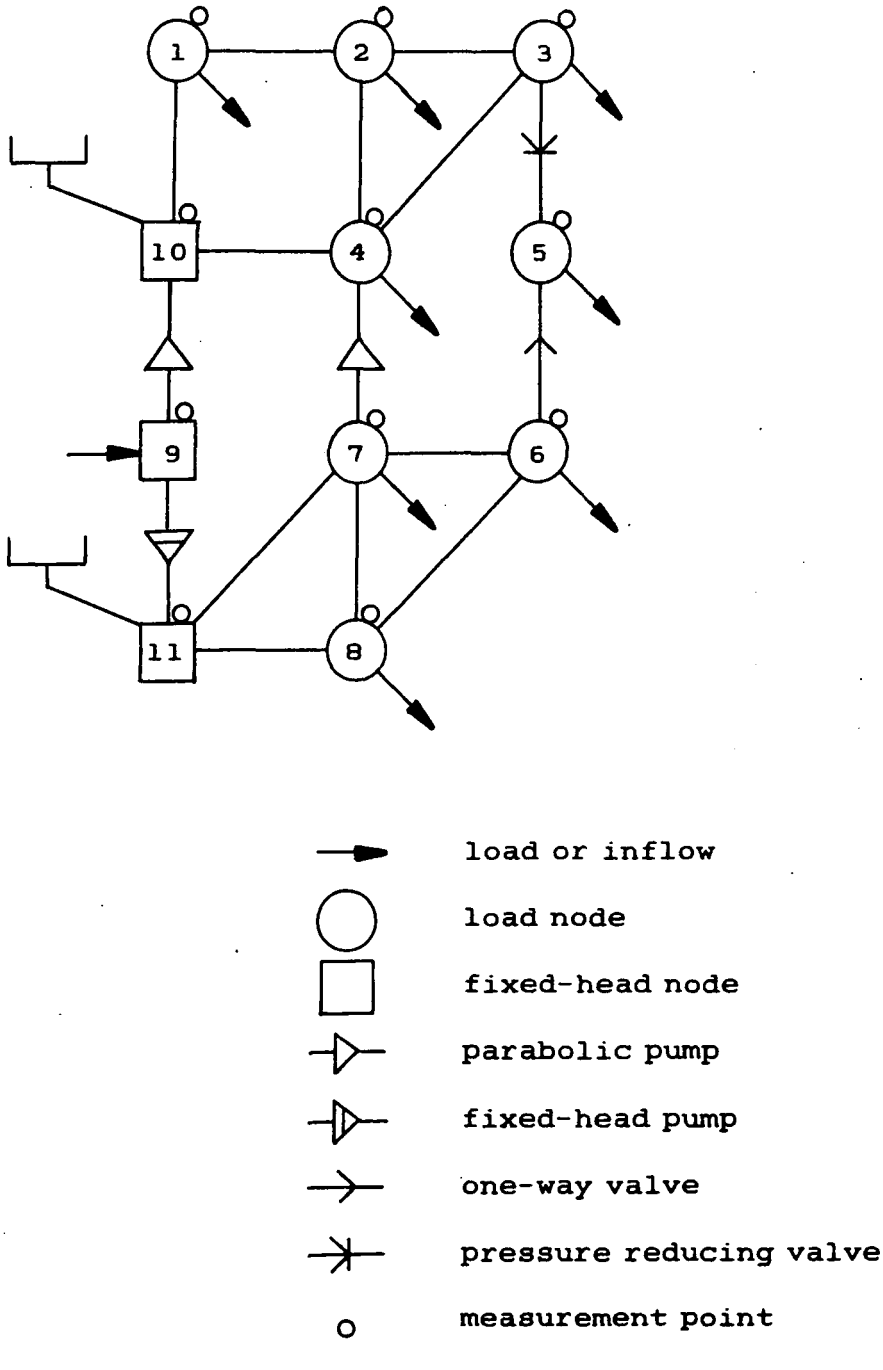


Figure C.1 11-node system

TABLE C.1

Parameters of the 11-node system

LINES	LENGTH [m]	DIAMETER [m]	HAZEN-WILLIAMS COEFFICIENT	HEAD-LIMIT [mAq]
10- 1	914.	0.406	100	
2- 1	914.	0.305	120	
3- 2	610.	0.254	110	
4- 2	610.	0.305	115	
10- 4	610.	0.305	110	
4- 3	610.	0.254	100	
6- 7	610.	0.254	110	
8- 7	610.	0.203	100	
11- 8	610.	0.305	110	
7-11	1219.	0.254	100	
8- 6	610.	0.254	120	
6- 5	1219.	0.203	100	
3- 5	1219.	0.203	110	109.73
9-11	305.	0.406	65	65.52

PARABOLIC PUMPS	a [s^2/m^5]	b [s/m^2]	c [m]
7- 4	-190.1	0.0	60.96
9-10	-1140.1	0.0	85.39

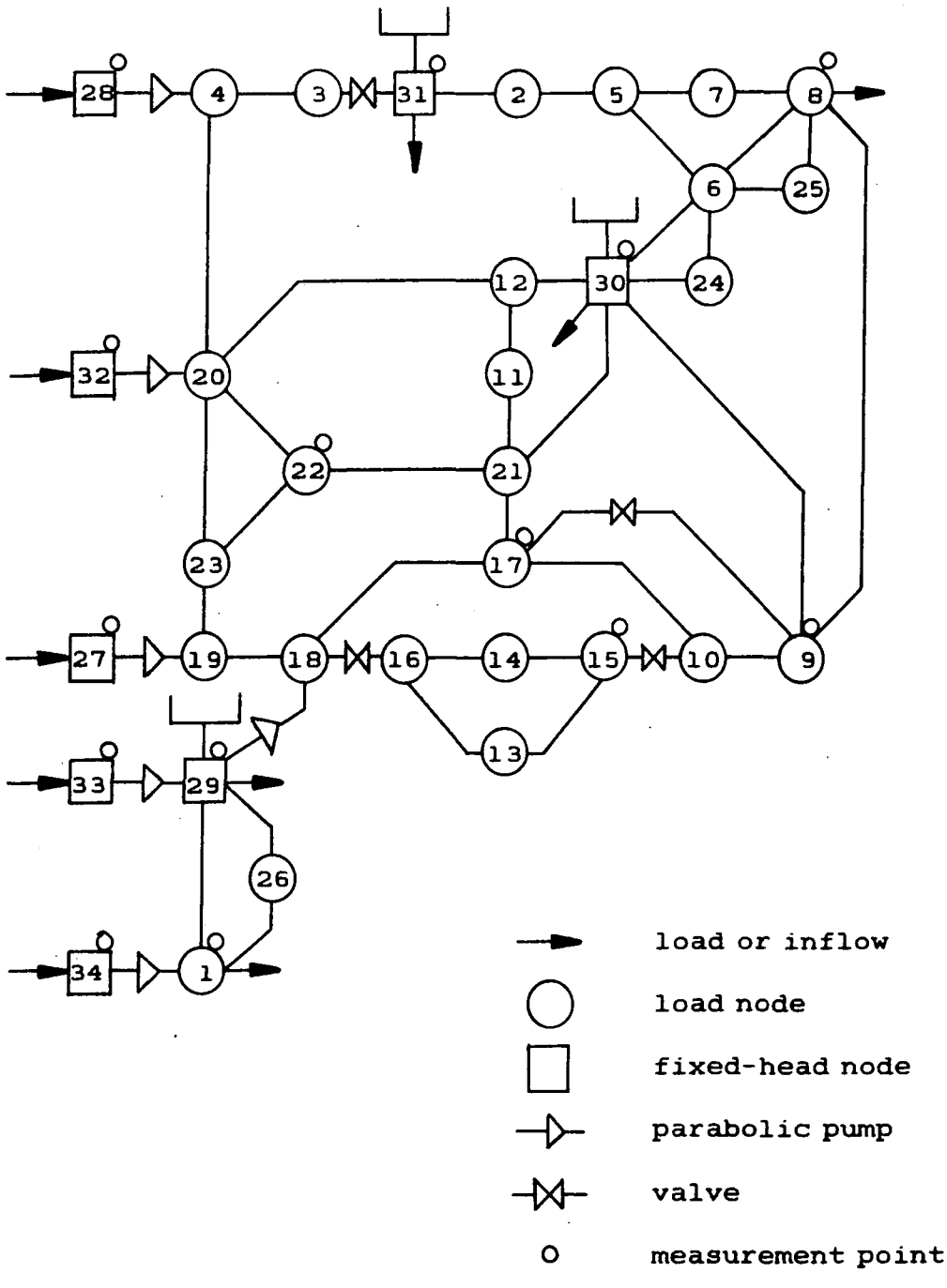


Figure C.2 34-node system

TABLE C.2

Parameters of the 34-node system

LINES	LENGTH [m]	DIAMETER [m]	HAZEN-WILLIAMS COEFFICIENT
4- 3	607.	0.457	110
20- 4	454.	0.457	110
23-20	2783.	0.229	105
23-19	304.	0.381	135
20-12	3383.	0.305	105
20-22	1768.	0.457	110
22-23	1015.	0.381	135
19-18	1097.	0.381	135
3-31	1930.	0.457	110
31- 2	3151.	0.305	100
22-21	762.	0.457	110
18-17	914.	0.229	125
18-16	823.	0.305	140
16-14	411.	0.152	100
14-15	701.	0.229	110
16-13	1072.	0.229	135
13-15	864.	0.152	90
15-10	711.	0.152	90
10-17	832.	0.152	90
17- 9	2334.	0.152	100
17-21	1969.	0.229	95
21-11	777.	0.229	90
11-12	542.	0.229	90
21-30	1600.	0.457	110
30-12	250.	0.305	105
2- 5	1028.	0.229	110
30-24	444.	0.229	90
30- 6	743.	0.381	100
30- 9	931.	0.229	125
9-10	2689.	0.152	100
5- 7	326.	0.152	100
7- 8	844.	0.229	110
5- 6	1274.	0.152	100
6- 8	1116.	0.229	90
6-25	615.	0.381	110
8- 9	1407.	0.152	100
1-29	427.	0.254	100
1-26	2098.	0.355	100

TABLE C.2 (cont.)

LINES	LENGTH [m]	DIAMETER [m]	HAZEN-WILLIAMS COEFFICIENT
25- 8	500.	0.381	110
24- 6	300.	0.229	90
26-29	1500.	0.355	100

PARABOLIC PUMPS	a [s^2/m^5]	b [s/m^2]	c [m]
28- 4	-4921.8	0.0	122.44
32-20	-444.4	-385.4	102.42
27-19	-812.3	89.7	62.03
29-18	-318.6	0.0	18.89
34- 1	-812.3	89.7	44.50
33-29	-4162.6	138.4	75.47

APPENDIX D

STORAGE SCHEMES IN SPARSITY PROGRAMMING [89]

D.1. INTRODUCTION

Sparsity techniques are used in the implementation of virtually all computer programs developed in connection with this thesis. One of the most important factors in the processing of sparse matrices is the selection of the appropriate storage schemes. Basically, two types of storage schemes are used, depending on whether the sparsity pattern of the matrix is expected to change or not. If the structure of the matrix under consideration is known 'a priori', a static storage scheme may be used. For instance, this is the scheme employed to store the Jacobian matrix in state estimation calculations. On the other hand, there are cases where the processing of a sparse matrix changes its structure; as a result, the storage scheme for those matrices should make provision for the changes. Such schemes are called dynamic storage schemes. Both techniques are described in the following sections.

According to how a matrix is formed, it can be stored row-by-row or column-by-column. Throughout this appendix it is assumed that the matrices are formed in a row-by-row

fashion. The extension for columnwise storage is straightforward.

D.2. A STATIC STORAGE SCHEME

Static schemes are employed when the whole structure of the sparse matrix to be dealt with is known. The static storage scheme that is used in this work requires the use of three different arrays to store an mxn matrix.

1) An integer array JCOL, containing the column indices of the nonzeros as encountered when scanning the matrix by rows, from row 1 to m;

2) An integer array XROW of length n+1, containing the pointers to the beginning of each row in JCOL. The last element of XROW points to the next available storage location in JCOL; and

3) A real (double precision) array VAL of the same length as JCOL, containing the numerical values of the nonzero elements, in the same order as their column indices appear in JCOL.

This method is called a row-pointer/column-index scheme.

As an example, consider the matrix A. The row-pointer/ column-index scheme for this matrix is given in Table D.1.

$$A = \begin{bmatrix} -8 & 0 & 1 & -3 \\ 0 & 0 & 4 & 2 \\ 0 & -3 & 0 & 0 \\ 1 & 0 & 0 & -2 \\ 0 & 6 & 1 & 0 \end{bmatrix} \qquad (D.1)$$

TABLE D.1

	XROW
1	1
2	4
3	6
4	7
5	9
6	11

	JCOL	VAL
1	1	-8
2	3	1
3	4	-3
4	3	4
5	4	2
6	2	-3
7	1	1
8	4	-2
9	2	6
10	3	1
11	-	-

The extra element in XROW is included for programming purposes.

The elements of row I of a sparse matrix stored as above can be examined by using the flow chart of Fig. D.1.

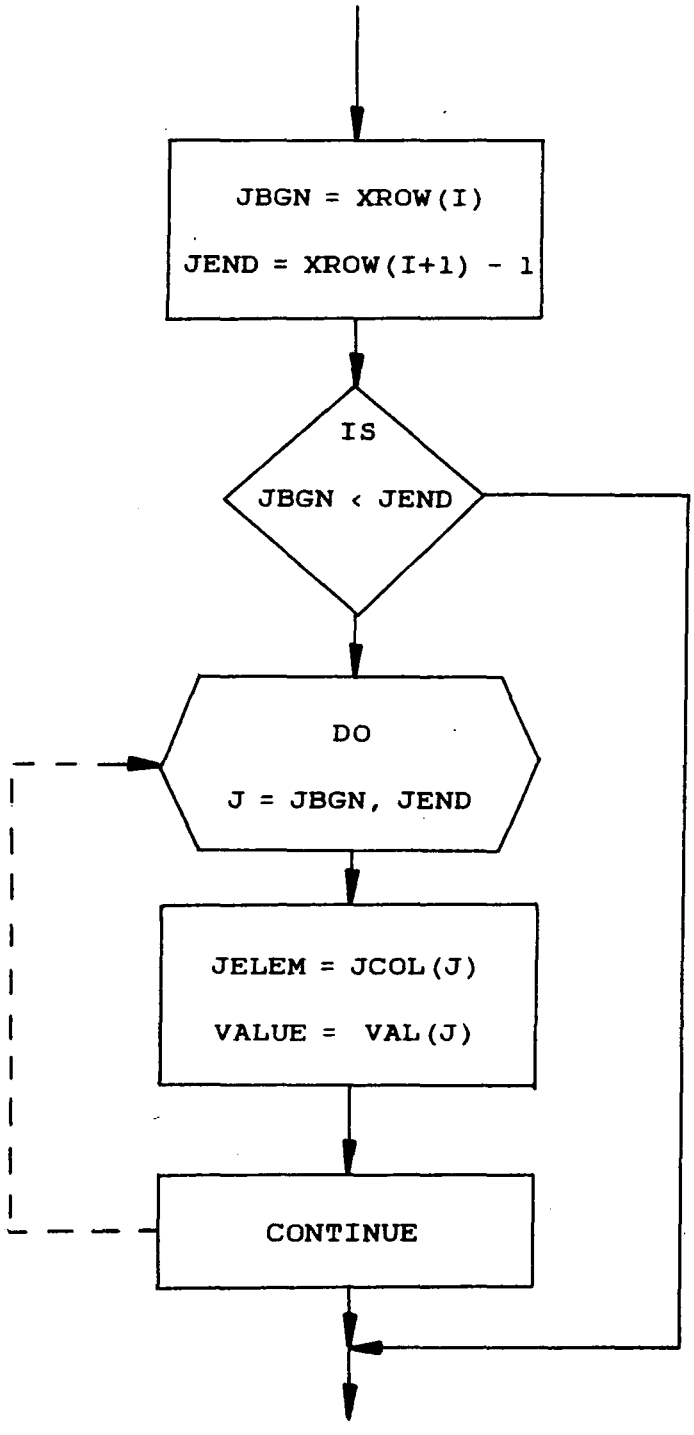


Figure D.1 Examination of elements of row I of a sparse matrix stored accordind to the row-pointer/column-index scheme.

D.3. A DYNAMIC STORAGE SCHEME

When the operations performed on a sparse matrix make it possible that zero elements become nonzeros, the use of the storage scheme described in section D.2 is not convenient. To allow the storage of new nonzeros, one should resort to a dynamic scheme. The most used type of dynamic scheme employs linked lists. A simple dynamic scheme for row-oriented storage using a one-way linked list requires four arrays:

1) An array of pointers, denoted by HEAD. HEAD(I) is the address in array JCOL which starts the list of nonzeros for row i;

2) An integer array JCOL containing the column indices of the nonzeros of the sparse matrix, which are not arranged in any particular order;

3) An integer array LINK of the same length as JCOL. The element LINK(k) give the position in JCOL where the next nonzero which is in the same row as JCOL(k) can be found. If LINK(k) = -i, it indicates the end of the list of nonzeros for row i; and

4) A real (double precision) array VAL of the same length as JCOL, with the numerical values for the corresponding elements of JCOL.

Matrix A of Eq. D.1 could be stored by using a

dynamic scheme as shown in Table D.2.

TABLE D.2

	HEAD
1	7
2	5
3	6
4	2
5	3

	JCOL	VAL	LINK
1	4	-2	-4
2	1	1	1
3	2	6	10
4	3	1	8
5	3	4	9
6	2	-3	-3
7	1	-8	4
8	4	-3	-1
9	4	2	-2
10	3	1	-5

To examine the elements of row I of a matrix stored as a linked list, the flow chart of Fig. D.2 can be used.

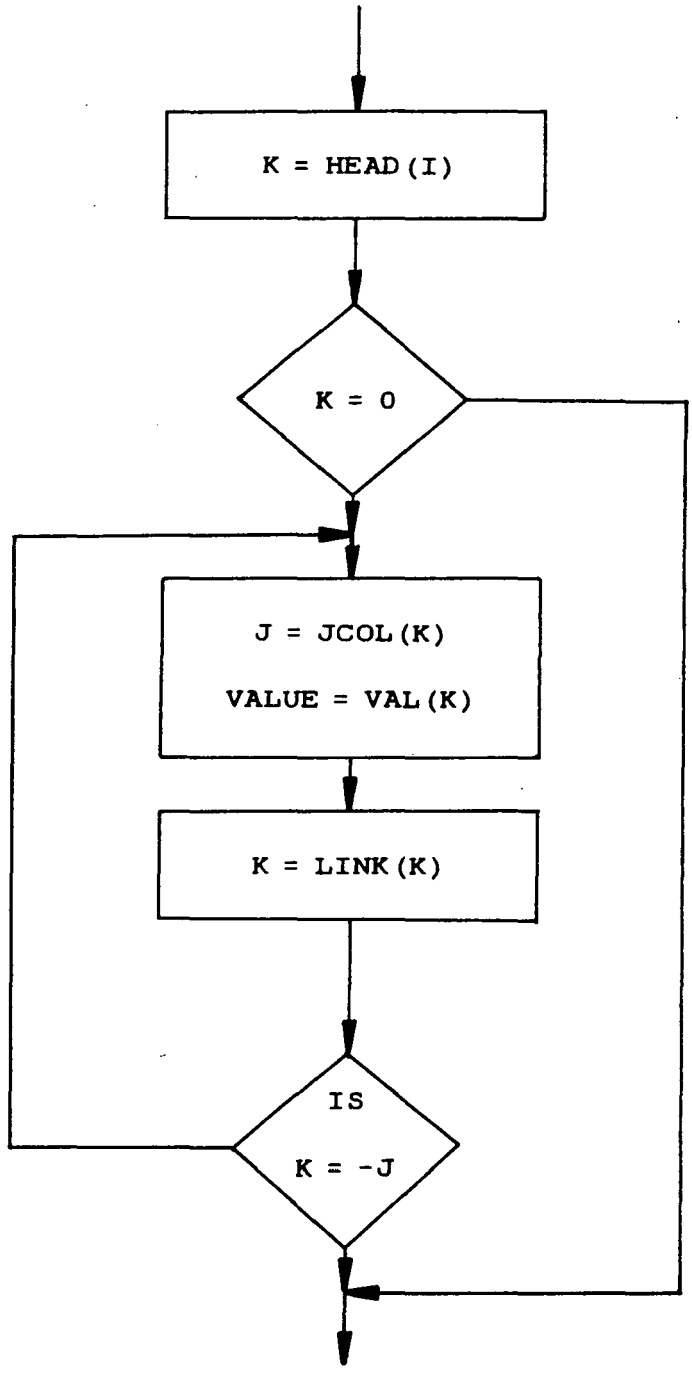


Figure D.2 Examination of row I of a sparse matrix stored according to the dynamic scheme using column links.

APPENDIX E

GRAPH THEORY

E.1 GRAPHS AND SUBGRAPHS [75], [102]

A graph $G = (V,E)$ consists of a set of objects $V = \{v_1, v_2, \ldots\}$ called vertices, and another set $E = \{e_1, e_2, \ldots\}$ whose elements are called edges, such that each edge e_k is identified with an unordered pair (v_i, v_j) of vertices. The vertices v_i, v_j associated with edge e_k are called the end vertices of e_k. The ends of an edge are said to be incident with the edge, and vice-versa. Graphs are usually represented by diagrams in which the vertices are indicated by points and each edge by a line connecting its end vertices.

An edge can have identical end vertices, in which case it is called self-loop. If the end vertices of an edge are distinct, it is called a link. It is possible to have more than one edge associated with the same pair of vertices. Such edges are said to be parallel. Two non-parallel edges are adjacent if they are incident to a common vertex; and two vertices are adjacent if they are the end vertices of the same edge. The number of edges incident on vertex v, with self-loops counted twice, is called the degree of

vertex v.

A graph is simple if it has neither self-loops nor parallel edges. A bipartite graph is one whose vertex set can be partitioned into two subsets, X and Y, so that each edge has one end in X and one end in Y. The partition (X,Y) is called a bipartition of the graph.

A graph g is a subgraph of the graph G if all the vertices and all the edges of g are in G, and each edge of g has the same end vertices in g as in G. A spanning subgraph of G is a subgraph of G whose vertex set is equal to the vertex set of G.

E.2 PATHS AND CONNECTION [75]

A walk is a finite alternating sequence of vertices and edges, beginning and ending with vertices, such that each edge is incident with the vertices preceding and following it. No edge appear more than once in a walk, but a vertex may appear more than once.

Vertices with which a walk begins and ends are called its terminal vertices. If a walk begins and ends at the same vertex, it is called a closed walk. Otherwise, it is an open walk.

A path is a walk in which no vertex appears more than once. A number of vertices in a path is called the length of the path.

A closed walk in which no vertex, except the initial and final vertex, appears more than once is called a circuit or a loop (not to be confused with a self-loop).

Two vertices u and v of a graph G are said to be connected if there is a path connecting them in G. A graph G is connected if there is at least one path between every pair of vertices in G. Otherwise, G is disconnected. A disconnected graph consists of two or more connected graphs, each of them is called a component of the graph G.

E.3 TREES AND FORESTS

A tree is a connected graph with no loops. The following are some of the properties of trees, whose proofs can be found in [75]:

a) There is one and only one path between every two vertices in a tree;
b) A tree with n vertices has n-1 edges;
c) Any connected graph with n vertices and n-1 edges is a tree; and
d) Every tree with two or more vertices has at least two vertices of degree one.

The distance between two vertices in a tree is the number of edges in the (unique) path connecting them.

A tree in which one vertex, called root, is

distinguished from all the others is called a rooted tree.

A tree T is said to be a spanning tree of a connected graph G if T is a spanning subgraph of G.

A forest is a graph with no loops. A disconnected graph with k components has a spanning forest, which is the collection of k spanning trees, one for each component. Every connected graph has at least one spanning tree.

An edge of a spanning tree T is called a branch of T. The edges of G that are not in T are called chords. A connected graph of n vertices has n-1 tree branches and e-n+1 chords which form a cotree.

E.4 INCIDENCE MATRIX

Let G be a graph with n vertices, e edges and no self-loops. Define the elements of an nxe matrix A whose n rows correspond to the n vertices, and e columns correspond to the e edges, as follows:

$a_{ij} = 1$, if the j-th edge e_j is incident on the i-th vertex v_i;

$a_{ij} = 0$, otherwise

Such a matrix is called the incidence matrix A for the graph G.

The following are some characteristics of incidence

matrices.

a) Since every vertex is incident on exactly two vertices, each column of A has exactly two 1's;

b) The number of 1's in each row equals the degree of the corresponding vertex.

It can easily be proved that the rank of an incidence matrix of a connected graph G with n vertices is n-1. Therefore, by removing any one row from the incidence matrix of a connected graph, the remaining (n-1)xe matrix is of rank n-1. Such an (n-1)xe submatrix of A, A_r, is called a reduced incidence matrix, and the vertex corresponding to the deleted row in A is the reference vertex.

Given that a tree is a connected graph with n vertices and n-1 edges, the reduced incidence matrix A_r for a tree is nonsingular.

The following theorem expresses a very important property of the submatrices of A.

Theorem E.1

Let A be an incidence matrix of a connected graph G with n vertices. An (n-1)x(n-1) submatrix of A is nonsingular if and only if the n-1 edges corresponding to the n-1 columns of this matrix constitue a spanning tree in G.

E.5 DIRECTED GRAPHS

A directed graph G consists of a set of vertices $V = \{v_1, v_2, \ldots\}$, a set of edges $E = \{e_1, e_2, \ldots\}$ and a mapping Ψ that maps every edge onto some ordered pair of vertices (v_i, v_j). A directed graph can be represented by a diagram where a vertex is indicated by a point and an edge by a line segment between v_i and v_j, with an arrow whose tail is v_i and whose head is v_j.

The indegree of a vertex v in a directed graph is the number of edges with head v, whereas the outdegree of v is the number of edges with tail v.

A direct walk from the vertex v_i to a vertex v_j in a directed graph is an alternating sequence of vertices and edges beginning with v_i and ending with v_j such that each edge is oriented from the vertex preceding it to the vertex following it. A directed path in a directed graph is a directed walk in which no vertex appears more than once.

E.6 INCIDENCE MATRIX FOR DIRECTED GRAPHS

The incidence matrix of a directed graph with n vertices, e edges and no self-loops is an nxe matrix A whose rows correspond to vertices, and columns correspond to edges of the directed graph, such that

$a_{ij} = 1$, if the i-th vertex is the tail of edge j

$a_{ij} = -1$, if the i-th vertex is the head of edge j

a_{ij} = 0, if the i-th edge is not incident to i-th vertex

The rank of the incidence matrix of a connected directed graph of n vertices is n-1. Deleting any one row from A, the (n-1)xe reduced incidence matrix A_r is obtained. The vertex corresponding to the deleted row is called reference vertex.

As in the case of nondirected graphs, the nonsingular matrices of order n-1 of A are in one-to-one correspondence with the spanning trees of the connected directed graph G of n vertices represented be A.