An algorithm for observability determination in water-system state estimation

A. Bargiela, B.Sc., M.Sc., Ph.D., M.I.P.E.E., M.Inst.M.C.

Indexing terms: *Control theory, Control systems, Control equipment and applications, Mathematical techniques*

Abstract: A new algorithm for efficient determination of topological observability in water-system state estimation has been proposed. The algorithm is based on observation that the search for a spanning tree of full rank can be performed as a sequence of maximum assignments. After giving a brief outline of the observability theory expressed in terms of water systems, the algorithm is described in full detail. Computational efficiency of the algorithm is evaluated on a 34-node water distribution system.

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 offline and in the online implementation of the estimator. In online 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 pseudomeasurements are added to the measurement set to enable 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 offline 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 decide 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.

The first observability algorithms making use of floating point computations [1, 2] were not well suited to online implementation, and their applicability was limited to offline meter placement studies. Other methods [3, 4] based on logical procedures were satisfactory from the computational point of view but proved to be conservative, labelling as unobservable some systems which in fact were observable. An important contribution to the solution of the observability problem was made by Krumpholz, Clements and Davis [5] who formulated necessary and sufficient conditions for observability in power-system state estimation in terms of meter location and network topology. According to their analysis a network is observable if and only if it contains a spanning tree of full rank.

This paper initially presents the basic theoretical results for topological observability in water-system state estimation and then proposes an original method to find an observable spanning tree of a water network. The observability problem is transformed into a sequence of maximum assignments in bipartite measurement/branch graphs. The method is computationally efficient and always gives a correct answer to the observability problem.

2 Theoretical background

The theory of topological observability has been initially developed in the context of power-system state estimation [5]. This Section gives a brief outline of the observability theory expressed in terms of water systems.

The nonlinear measurement model for the water system state estimation is given by

$$z = g(x) + \omega \tag{1}$$

where z is the $m \times 1$ measurement vector, x is the $n \times 1$ state vector, $g(\ )$ is an $m \times 1$ nonlinear vector function and ω is an $m \times 1$ random vector which models the measurement errors. The water system is said to be observable if the measurement set contains a subset of n linearly independent measurements. Disregarding higher-order terms in a Taylor-series expansion of the function $g(x)$, the above definition is equivalent to the requirement that the Jacobian matrix of $g(x)$ is of full rank throughout the iterative solution of eqn. 1.

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_c consumer loads and m_f pipe flows. The state vector is of the form $x^T = [h, q]$, where h is the vector of nodal heads and q is the vector of fixed-head-node flows. To achieve a one-to-one correspondence

Paper 4205D (C8, C9), first received 25th March and in revised form 30th September 1985

The author was formerly with the Department of Engineering, University of Durham, and is now with the Department of Computing, Trent Polytechnic, Burton Street, Nottingham NG1 4BU, United Kingdom

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

(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.

The network extended by auxiliary nodes and by a head reference node with links to the head-measured nodes is called an augmented network graph. This graph consists of $n = N + F$ nodes and $L = P + F + m_h$ links (Fig. 1*b*).

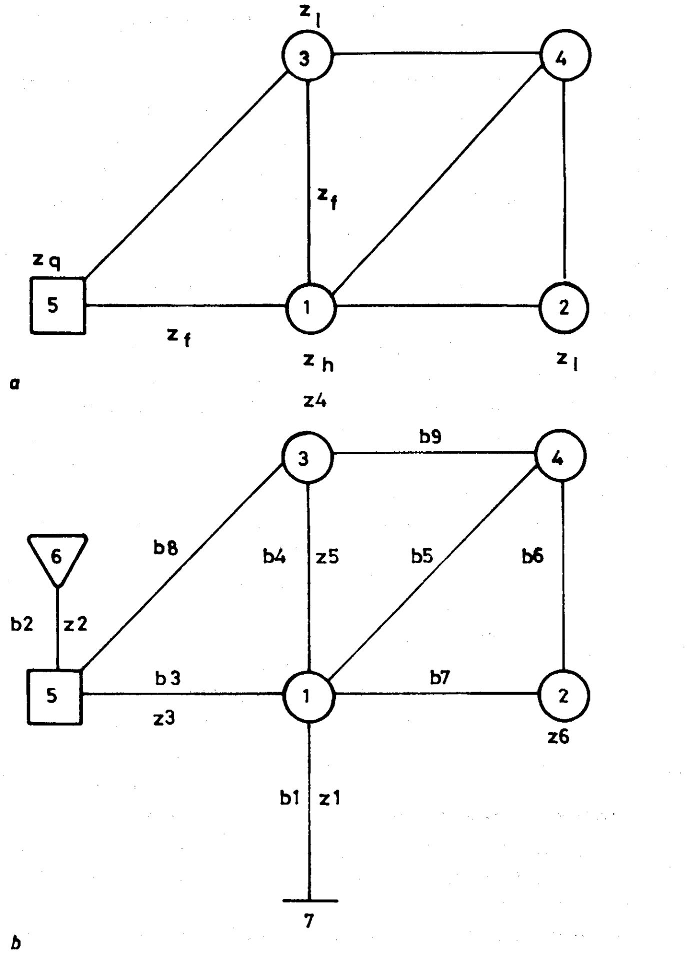

Fig. 1 *Original network graph (a) and augmented network graph (b)*

○ variable-head node — z_h = head measurement
□ fixed-head node — z_q = fixed-head-node flow measurement
▽ auxiliary node — z_c = load measurement
⊥ reference node — z_f = pipe flow measurement

Each link contributes one element to the diagonal $L \times L$ hydraulic conductivity matrix $\boldsymbol{Y}$. The P elements of the matrix $\boldsymbol{Y}$ 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 remaining $F + m_h$ elements of $\boldsymbol{Y}$ represent the sensitivity of fixed-head flows and the sensitivity of nodal heads to their own changes and are therefore equal to 1.

Having the required definitions, the measurement model can be expressed as

$$\begin{bmatrix} z_h \\ z_q \\ z_c \\ z_f \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}_h \boldsymbol{Y} \boldsymbol{A}_r \\ \boldsymbol{M}_q \boldsymbol{Y} \boldsymbol{A}_r \\ \boldsymbol{M}_c \boldsymbol{Y} \boldsymbol{A}_r \\ \boldsymbol{M}_f \boldsymbol{Y} \boldsymbol{A}_r \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{h} \\ q \end{bmatrix} + \begin{bmatrix} \omega_h \\ \omega_q \\ \omega_c \\ \omega_f \end{bmatrix} \tag{2}$$

or simply as

$$z = \boldsymbol{G}_m \boldsymbol{x} + \boldsymbol{\omega} \tag{2a}$$

where

z_h, z_q, z_c, z_f are $m_h \times 1$, $m_q \times 1$, $m_c \times 1$ and $m_f \times 1$ measurement vectors of head, fixed-head-node flow, consumer load and pipe flow respectively;

$\omega_h, \omega_q, \omega_c, \omega_f$ are $m_h \times 1$, $m_q \times 1$, $m_c \times 1$ and $m_f \times 1$ measurement noise vectors corresponding to vectors z_h, z_q, z_c, z_f;

$\boldsymbol{M}_h, \boldsymbol{M}_q, \boldsymbol{M}_c, \boldsymbol{M}_f$ are $m_h \times L$, $m_q \times L$, $m_c \times L$ and $m_f \times L$ meter placement matrices for head, fixed-head-node flow, consumer load and pipe flow, respectively.

$\boldsymbol{A}_r$ is $L \times n$ reduced node-to-branch incidence matrix obtained from the full incidence matrix by deleting the row corresponding to the reference node.

Construction of the matrices $\boldsymbol{M}_h$, $\boldsymbol{M}_q$ and $\boldsymbol{M}_f$ is straightforward because the head, fixed-head-node 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 $\boldsymbol{M}_c$ involves examination of the network links incident to the measured nodes. If the link is directed to the node, the corresponding entry in the matrix is $+1$, otherwise the entry is -1. All remaining elements in the row are equal to zero.

Eqn. 2*a* can be seen as a linearisation of the measurement eqn. 1. The structure of $\boldsymbol{G}_m$ does not depend on the elements of $\boldsymbol{Y}$, which change with operating point, but is determined by meter placement and network topology. Consequently, a concept of topological observability can be introduced:

Definition 1: An n-node water system is topologically observable with respect to a given measurement set, if, and only if, the rank of the matrix $\boldsymbol{G}_m$ is equal to n.

By transforming eqn. 2*a* from the nodal head to the pipe-flow frame of reference and assuming that the hydraulic conductivities of the pipes are such that they do not reduce the rank of the measurement matrix, it can be shown [5] that the topological observability of the water system is equivalent to the existence of a spanning tree of full rank. This result can be stated in the form of theorems 1 and 2.

Theorem 1: If a water system is topologically observable with respect to a measurement set $\boldsymbol{M}$, then there exists a spanning tree of its augmented network graph which is an observable spanning tree, and whose branches are associated one-to-one with measurements of $\boldsymbol{M}$.

Theorem 2: Suppose that there exists an observable tree in the augmented network graph whose branches are associated with measurements of a measurement set $\boldsymbol{M}$. 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 $\boldsymbol{M}$.

Rigorously, theorem 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 occur in practice. Thus, topological observability can be investigated by seeking an observable spanning tree of the network graph.

3 The observability algorithm

The observability algorithm proposed in this paper is based on the observation that the search for a spanning tree of full rank of an augmented network graph can be represented as two interacting subproblems. The objective of the first subproblem is to find a maximal subset $\boldsymbol{B}_0$ of the set of branches $\boldsymbol{B}$, whose elements can be associated with elements of the measurement set $\boldsymbol{M}$ in a one-to-one manner. This one-to-one correspondence is referred to as the assignment $(\boldsymbol{M}_0, \boldsymbol{B})$. The second subproblem is concerned with the search for the observable spanning tree in the subgraph $\boldsymbol{G}_0$ of $\boldsymbol{G}$ formed by the branches $\boldsymbol{B}_0$ of the measurement assignment $(\boldsymbol{M}_0, \boldsymbol{B}_0)$. If such a tree is found, the network is topologically observable. Otherwise, an attempt is made to modify $\boldsymbol{G}_0$ 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 $\boldsymbol{G}$. Fig. 2 schematically shows the basic steps of the observability algorithm.

3.1 Measurement-branch bipartition

The first step in the search for an $(\boldsymbol{M}_0, \boldsymbol{B}_0)$ measurement assignment is the formulation of the problem using bipartite graphs. A bipartite graph is a graph whose vertex set can be partitioned into two subsets, $\boldsymbol{X}$ and $\boldsymbol{Y}$, so that each edge of the graph has one end in $\boldsymbol{X}$ and one end in $\boldsymbol{Y}$; the partition $(\boldsymbol{X}, \boldsymbol{Y})$ is called a bipartition of the graph. To associate the elements of the measurement set $\boldsymbol{M}$ with the branch set $\boldsymbol{B}$ of the augmented network graph, a bipartite graph of the type $(\boldsymbol{M}, \boldsymbol{B})$ is constructed. The edges of this graph are determined by the following rules:

(*a*) If measurement z_i is a head measurement at the node i, the corresponding vertex $z_i \in \boldsymbol{M}$ is connected to the vertex $b_i \in \boldsymbol{B}$, which represents itself as 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 $(\boldsymbol{M}, \boldsymbol{B})$ 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 $b_i \in \boldsymbol{B}$, 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 $\boldsymbol{\Omega}_i$, then the vertex $z_i \in \boldsymbol{M}$ is connected to all vertices $b_i \in \boldsymbol{\Omega}_i$.

The $(\boldsymbol{M}, \boldsymbol{B})$ bipartitite graph corresponding to the augmented network graph of Fig. 1*b* is given in Fig. 3.

3.2 Maximum assignment

The $(\boldsymbol{M}_0, \boldsymbol{B}_0)$ assignment problem can be solved in several different ways. The algorithm employed in the observability routine is based on the work of Hopcroft and Karp [10], who used the concept of a layered network of Dinic [6] to devise a reassignment path. An efficient implementation of this algorithm has been obtained by generalising a maximum transversal algorithm of Duff [7] to enable a different number of elements in the bipartite sets $\boldsymbol{M}$ and $\boldsymbol{B}$.

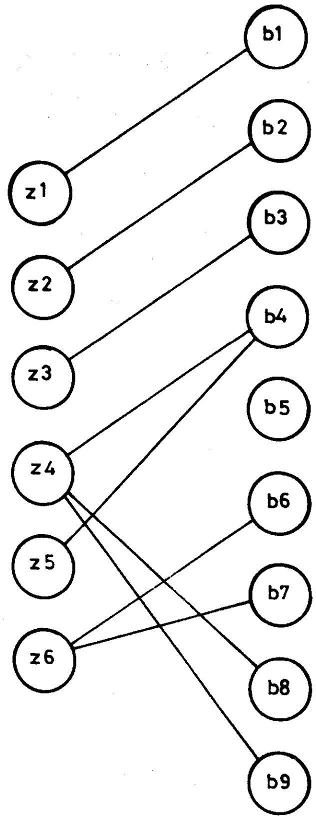

Fig. 3 *Bipartite graph* (M, B)

The maximum assignment $\boldsymbol{M}$ is constructed in m major steps, where m is the number of measurements. After the kth step there is available a maximum assignment of the first k measurements into the set of branches $\boldsymbol{B}$. The search for a larger assignment which includes an M-unsaturated vertex $u \in \boldsymbol{M}$, consists of forming a tree of a bipartite graph called an M-alternating tree rooted at u [8]. Such a tree has the following properties:

(*a*) The M-unsaturated vertex is a root of a tree

(*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 assignment M in the bipartite graph $\boldsymbol{G}$, after processing four elements of the set $\boldsymbol{M}$, is given in Fig. 4*a*, and the M-alternating tree rooted at the fifth element of $\boldsymbol{M}$ is shown in Fig. 4*b*. The reassignment corresponds to replacing the branch $z4$–$b4$ with two branches $z5$–$b4$ and $z4$–$b8$, thus enlarging the current assignment.

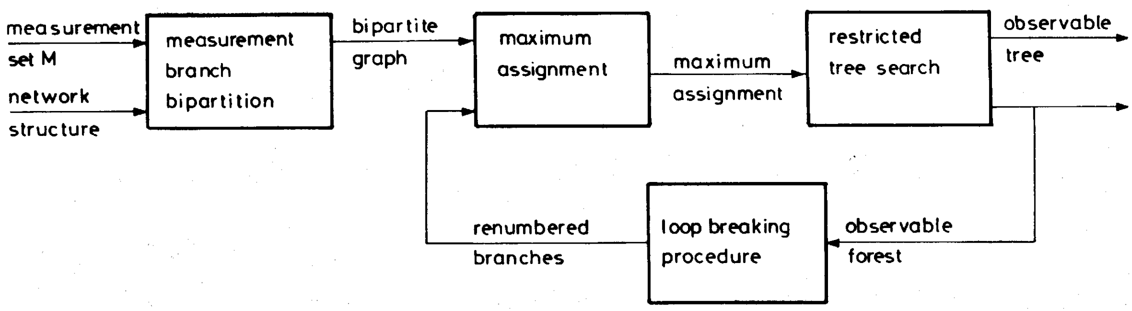

Fig. 2 *Basic steps in the observability algorithm*

The maximum assignment produced by the maximum transversal algorithm depends on the order in which the edges of the bipartite graph ($\boldsymbol{M}$, $\boldsymbol{B}$) are examined. In general, there are several possible maximum assignments; each of them, however, has the same number of components β (Fig. 5). In some special cases, the question of topological observability can be answered immediately by examination of the assignment length:

(*a*) If the ($\boldsymbol{M}_0$, $\boldsymbol{B}_0$) assignment has the length $\beta = L$, where L is the number of branches of the augmented network graph $\boldsymbol{G}$, then the water network is observable with respect to the measurement set $\boldsymbol{M}$. Every spanning tree of the graph $\boldsymbol{G}$ is an observable spanning tree since $\boldsymbol{B}_0$ contains all the branches of $\boldsymbol{G}$.

(*b*) If the length of the ($\boldsymbol{M}_0$, $\boldsymbol{B}_0$) assignment $\beta < n$, then the water network is topologically unobservable due to the fact that $\boldsymbol{B}_0$ contains less branches than any spanning tree of $\boldsymbol{G}$. (Consequently, no observable spanning tree exists.)

However, if $n \le \beta < L$, an attempt to find a spanning tree of $\boldsymbol{G}$ built of branches $\boldsymbol{B}_0$ must be made.

The theoretical upper bound on computational complexity of the maximum transversal algorithm can be found as the product of the number of vertices and edges

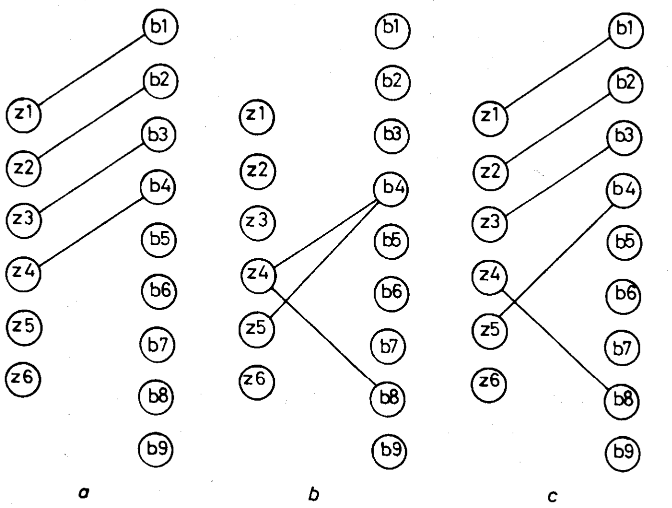

Fig. 4 *Assignment M after four steps (a), M-alternating tree rooted at z5 (b) and assignment after five steps (c)*

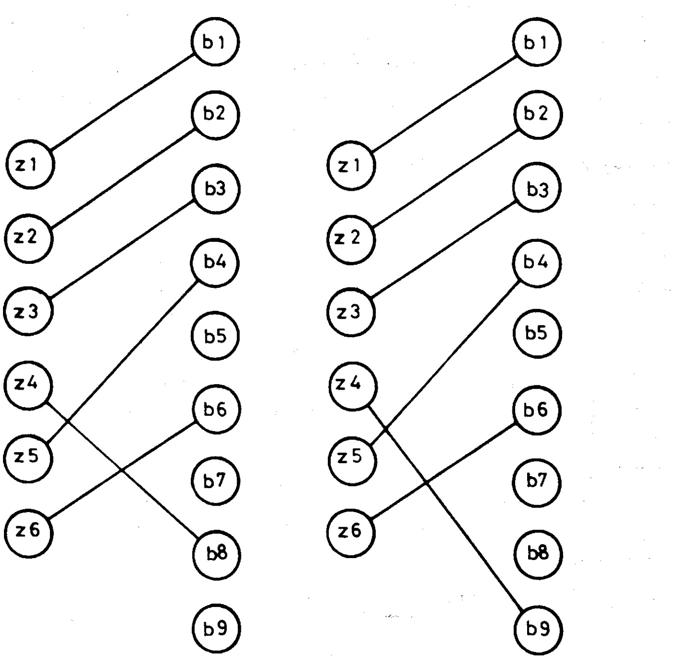

Fig. 5 *Maximum matchings*

of the bipartite graph concerned. However, in most practical cases, the algorithm performs as if its complexity was linearly dependent on the sum of the number of vertices and edges.

3.3 Restricted tree search

Given the set of branches $\boldsymbol{B}_0$ of the maximum assignment M the restricted tree search has been implemented as a breadth-first-search procedure. If the procedure returns a spanning tree of $\boldsymbol{G}$, the observability algorithm terminates, otherwise it proceeds to the reassignment routine.

3.4 Loop breaking procedure

The maximum assignment rules guarantee that if a water system is topologically observable with respect to a measurement set $\boldsymbol{M}$, then there exists a measurement assignment ($\boldsymbol{M}_0$, $\boldsymbol{B}_0$) whose branches contain a spanning tree of the network graph. Conversely, if there is no assignment which contains a spanning tree of $\boldsymbol{G}$, then the water system is topologically unobservable with respect to $\boldsymbol{M}$. However, one cannot claim that, for every assignment ($\boldsymbol{M}_0$, $\boldsymbol{B}_0$) in the observable network, the spanning tree of $\boldsymbol{G}$ 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 $\boldsymbol{B}_0$. The possibility of creating such loops is apparent because the water network connectivity is not taken into account at this stage. A procedure is then required which would aid the maximum assignment algorithm in 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.

The first part of the loop breaking procedure, an identification of observable loops, has been implemented as a search for biconnected components of $\boldsymbol{G}_0$ using an algorithm of Hopcroft and Tarjan [9]. This is possible because each loop is contained in one of the bicomponents and each bicomponent which has more than one edge contains at least one loop. The second part, a reassignment of measurements, is solved by introducing an artificial loop-measurement incident to all loop edges of $\boldsymbol{G}_0$ and performing one step of the maximum assignment 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 $\boldsymbol{G}_0$. The procedure terminates if either an observable spanning tree of $\boldsymbol{G}$ is found or a forest cannot be linked. Fig. 6 illustrates the process of reassignment of the branch $z4$–$b8$ into $z4$–$b9$ using the loop measurement z_{L1}.

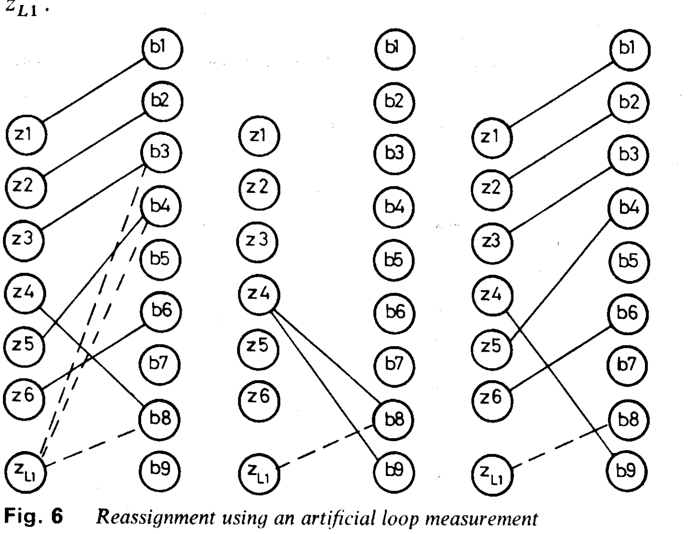

Fig. 6 *Reassignment using an artificial loop measurement*

The flowchart exhibited in Fig. 7 summarises all the steps in the implementation of the observability algorithm.

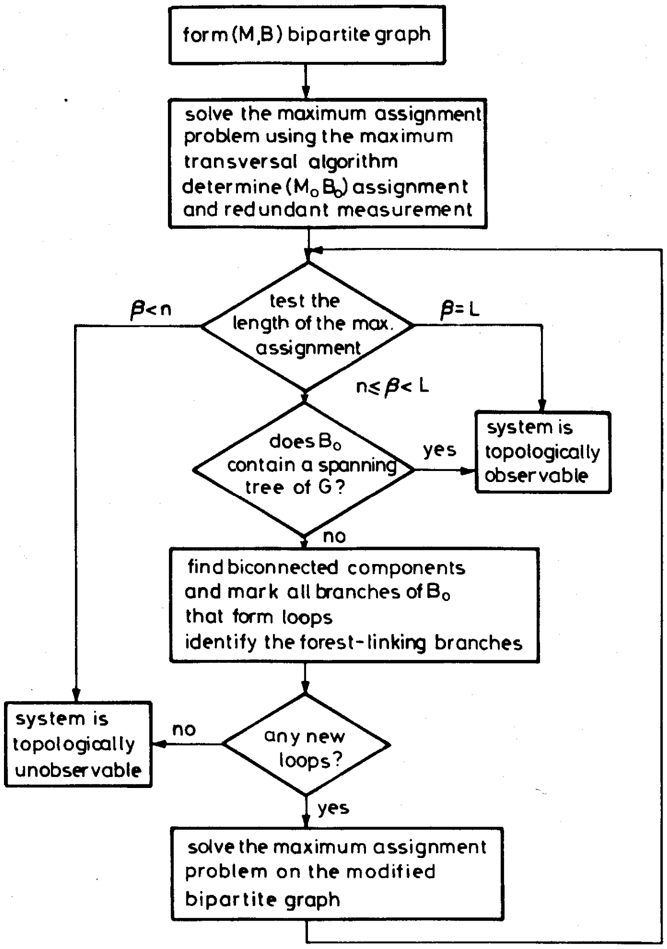

Fig. 7 *Flowchart for the observability algorithm*

4 Numerical results

To assess computational efficiency of the observability algorithm, several water networks with various measurement configurations have been studied. The algorithm was coded in Fortran 77 and run on a Perkin Elmer 3220 minicomputer. A representative sample of the results obtained for the 34-node water network is given in Table 1. The computing times demonstrate strong dependence on the number of measurements, regardless of whether the system is observable or unobservable. This is expected, as an increase in the number of measurements also increases the possibility of identifying a maximal observable tree or forest, without the need of initialising the reassignment procedure. Consequently, the observability algorithm presented in this paper is particularly well suited for the state estimation process where the measurement redundancy is an inherent feature of the measurement set.

It is worth noting that the computational efficiency of the algorithm is not affected by the large number of possible loops in the network because the algorithm searches for biconnected components of the augmented network graph, and these are at most $N + F - 1$, and in practice much less.

Table 1: Computational results for the 34-node system

Case/ observability*	1 obs.	2 obs.	3 obs.	4 unobs.	5 unobs.	6 unobs.
Number of measurements	42	43	46	41	42	47
Total time, s	0.133	0.095	0.032	0.156	0.099	0.065
Result*	obs.	obs.	obs.	unobs.	unobs.	unobs.

* obs. = observable
unobs. = unobservable

5 References

1 EDELMAN, H.: 'A universal assessment for the superior quality of distribution of measuring points for the state-estimate of high voltage network'. Proc. 5th PSC Conference, Cambridge, 1975

2 KOGLIN, H.J.: 'Optimal measuring system for state estimation', Proc. 5th PSC Conference, Cambridge, 1975

3 HANDSHIN, E. (Ed.), and BONGERS, C.: 'Theoretical and practical considerations in the design of state estimators for electric power systems' (Elsevier, 1972)

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

5 KRUMPHOLZ, G.R., CLEMENTS, K.A., and DAVIS, P.W.: 'Power system observability: a practical algorithm using network topology', *IEEE Trans.*, 1980, **PAS-99**, pp. 1534–1542

6 DINIC, E.A.: 'Algorithm for solution of a problem of maximum flow in a network with power estimation', *Soviet Math. Docl.*, 1970, **11**, (5), pp. 1277–1280

7 DUFF, I.S.: 'On algorithms for obtaining maximum transversal', *ACM Trans.*, 1981, **7**, (3), pp. 315–330

8 DEO, N.: 'Graph theory with applications to engineering and computer science' (Prentice-Hall Inc., 1974)

9 HOPCROFT, J., and TARJAN, R.: 'Efficient algorithms for graph manipulation', *Commun. ACM*, 1973, **16**, (6), pp. 372–378

10 HOPCROFT, J., and KARP, R.M.: 'An $n^{2.5}$ algorithm for maximum matching in bipartite graphs'. IEEE Conf. Record 12th Annual Symposium on Switching and Automata Theory, 1971, pp. 122–125

6 Appendix

This Appendix details the measurement model obtained for the 5-node network of Fig. 1. Because of the assumption about the numerical values of hydraulic conductivities, expressed in theorem 2, only the pattern of nonzero elements in the measurement matrix $\boldsymbol{G}_m$ needs to be considered. Various stages of processing of this topological information are illustrated in Figs. 3–6.

$N = 5$ = number of network nodes

$F = 1$ = number of fixed-head nodes

$P = 7$ = number of pipes

$m_h = 1$ = number of head measurements

$n = N + F$ = number of augmented network nodes (without reference node) = number of state variables (without reference head)

$L = P + F + m_h$ = number of links in the augmented network

$$z_h^T = [z_1]$$

$$z_q^T = [z_6, z_4]$$

$$z_c^T = [z_2]$$

$$z_f^T = [z_3, z_5]$$

$$\boldsymbol{\omega}_h^T = [\omega_1]$$

$$\boldsymbol{\omega}_q^T = [\omega_6, \omega_4]$$

$$\boldsymbol{\omega}_c^T = [\omega_2]$$

$$\boldsymbol{\omega}_f^T = [\omega_3, \omega_5]$$

$$\boldsymbol{h}^T = [h_1, h_2, h_3, h_4, h_5]$$

$$\boldsymbol{q}^T = [q_5]$$

$$\boldsymbol{M}_h = [1, 0, 0, 0, 0, 0, 0, 0, 0]$$

$$\boldsymbol{M}_q = [0, 1, 0, 0, 0, 0, 0, 0, 0]$$

$$M_c = \begin{bmatrix} 0, 0, 0, 0, 0, & -1, 1, 0, & 0 \\ 0, 0, 0, 1, 0, & 0, 0, 1, & -1 \end{bmatrix}$$

$$M_f = \begin{bmatrix} 0, 0, 1, 0, 0, 0, 0, 0, 0 \\ 0, 0, 0, 0, 1, 0, 0, 0, 0 \end{bmatrix}$$

$$Y = \begin{bmatrix} 1 & & & & & & & & \\ & 1 & & & & & & 0 & \\ & & Y_{33} & & & & & & \\ & & & Y_{44} & & & & & \\ & & & & Y_{55} & & & & \\ & & & & & Y_{66} & & & \\ & & 0 & & & & Y_{77} & & \\ & & & & & & & Y_{88} & \\ & & & & & & & & Y_{99} \end{bmatrix}$$

$$A_r^T = \begin{bmatrix} 1, & 0, & 1, & -1, & -1, & 0, & -1, & 0, & 0 \\ 0, & 0, & 0, & 0, & 0, & -1, & 1, & 0, & 0 \\ 0, & 0, & 0, & 1, & 0, & 0, & 0, & 1, & -1 \\ 0, & 0, & 0, & 0, & 1, & 1, & 0, & 0, & 1 \\ 0, & 1, & -1, & 0, & 0, & 0, & 0, & -1, & 0 \\ 0, & -1, & 0, & 0, & 0, & 0, & 0, & 0, & 0 \end{bmatrix}$$