MANAGEMENT SCIENCE
Vol 35, No 5, May 1989
Printed in U.S.A.

GAUSSIAN INFLUENCE DIAGRAMS*

ROSS D. SHACHTER AND C. ROBERT KENLEY
Department of Engineering-Economic Systems, Stanford University, Stanford, California 94305
Tiburon Systems, Inc., 2085 *Hamilton Avenue, San Jose, California* 95125

An influence diagram is a network representation of probabilistic inference and decision analysis models. The nodes correspond to variables that can be either constants, uncertain quantities, decisions, or objectives. The arcs reveal probabilistic dependence of the uncertain quantities and information available at the time of the decisions. The influence diagram focuses attention on relationships among the variables. As a result, it is increasingly popular for eliciting and communicating the structure of a decision or probabilistic model.

This paper develops the framework for assessment and analysis of linear-quadratic-Gaussian models within the influence diagram representation. The "Gaussian influence diagram" exploits conditional independence in a model to simplify elicitation of parameters for the multivariate normal distribution. It is straightforward to assess and maintain a positive (semi-)definite covariance matrix. Problems of inference and decision making can be analyzed using simple transformations to the assessed model, and these procedures have attractive numerical properties. Algorithms are also provided to translate between the Gaussian influence diagram and covariance matrix representations for the normal distribution.
(INFLUENCE DIAGRAM; GAUSSIAN DECISION MODEL; MULTIVARIATE NORMAL ASSESSMENT)

0. Introduction

An influence diagram is a graphical representation for a decision problem under uncertainty. Since it was formalized by Howard and Matheson (1981) it has been used primarily to help decision makers and analysts structure models and communicate concepts of conditional independence and information flow. In recent years, methodology has been developed to analyze probabilistic inference and decision analytic models directly within the influence diagram representation (Olmsted 1983, Shachter 1986, 1988). In this way, assessment, evaluation, and sensitivity analysis can all be performed on the model in the form most natural to the decision maker. Until now, application of this methodology has been limited to discrete probability distributions.

The linear-quadratic-Gaussian model is a sensible application of influence diagram theory to continuous variables. Gaussian models are widely used for decision making and statistical inference. Examples range from optimal control (Bryson and Ho 1975) to multivariate linear regression (Graybill 1976). Much of the current research in the field relies heavily on matrix theory, and while it provides a unified perspective for the formulation and solution of diverse problems, it tends to separate the analysts from the decision makers. On the other hand, path analysis provides a graphical approach to Gaussian models (Wright 1934), but its analytic use has been limited. There also has been some work (Pearl 1985) applying a restricted class of influence diagrams to multivariate normal models.

This paper develops a "Gaussian influence diagram," which combines the analytical tractability of linear-quadratic-Gaussian models with the graphical structure of the influence diagram. Not only can earlier results on influence diagram models be applied to

* Accepted by Robert L. Winkler, former Departmental Editor; received January 27, 1987. This paper has been with the authors 6 months for 2 revisions.

0025-1909/89/3505/0527$01.25

Gaussian influence diagrams, but many of the standard operations on linear-quadratic-Gaussian models have attractive Gaussian influence diagram counterparts. In fact, some of those results were suggested by Yule (1907). For example, model assessment is simplified, since the parameters tend to have natural physical meanings, and the number of parameters is reduced when there is conditional independence in the model structure. It is easy to elicit and maintain a valid positive semidefinite covariance matrix within the Gaussian influence diagram, allowing stable real-time processing even when models are collinear. The basic operations are scalar with a graphical and intuitive interpretation that makes them easy to teach to students or computers.

§1 introduces some of the influence diagram concepts with a simple example, while §2 provides more formal definitions. §3 presents the basic influence diagram reductions on the pure multivariate normal model. Those reductions are applied to solve the general probabilistic inference problem in §4. §5 compares the covariance representation with the Gaussian influence diagram for assessment and provides algorithms for converting from one representation to the other. A quadratic value function and decision variables are introduced in §6 and §7, respectively. §8 develops and demonstrates the algorithm to solve decision analysis problems. §9 generalizes the algorithm for exponential utility over the quadratic value function. Finally, §10 presents conclusions and directions for future research. Appendix A provides a detailed proof of the arc reversal theorem, while Appendix B explores the relation between the Gaussian influence diagram and other familiar representations for the multivariate normal model.

1. Influence Diagram Example

In this paper a complete framework will be developed for performing decision analysis within the Gaussian influence diagram. In this section, some of the basic influence diagram concepts are introduced along with an example which illustrates the influence diagram structure.

There are some graph-theoretical terms which will be used frequently throughout the paper. A *node* is the basic element within the influence diagram, and a (*directed*) *arc* is a line with an arrow which points from one node to another. A *directed graph* consists of a finite set of nodes and directed arcs, and a *network* is a graph in which additional data are stored in the nodes and arcs. A *directed path* is a sequence of nodes which can be visited by moving along arcs in the direction of their arrows. A *directed cycle* is a directed path which starts and ends with the same node.

An *influence diagram* is a network consisting of a directed graph with no directed cycles. Each node in the graph represents a variable in the model. This variable can be either a constant, an uncertain quantity, a decision to be made, or an objective. It is convenient to refer interchangeably to a node in the diagram and the variable to which it corresponds. A *probabilistic node* represents an uncertain quantity (or a constant); an influence diagram with only probabilistic nodes is said to be *probabilistic*. The diagram is called *Gaussian* if the joint probability distribution for the variables in the model is multivariate normal. In this paper Gaussian probabilistic influence diagrams are developed first, and then results are generalized to diagrams with a quadratic objective function and decision variables.

Each variable in a Gaussian probabilistic influence diagram may take a scalar real value, its *outcome*. There is a *conditional probability distribution* for each variable over the real line. The *conditioning variables* for the distribution are indicated by arcs into a variable's node, coming from the conditioning variables' nodes. If there are no arcs going into a node then its variable has an unconditional probability distribution.

In the Gaussian influence diagram, the conditional probability distribution can be specified by an (*unconditional*) *mean*, a *conditional variance*, and, corresponding to each conditioning variable, a *linear coefficient*. It is convenient to draw the Gaussian diagram with a mean and conditional variance associated with each node, and a coefficient associated with each arc. (Hence they will sometimes be called *arc coefficients*.) If all of the nodes and arcs have been drawn, then the influence diagram is *partially specified*. If all of the data (means, variances, and arc coefficients) have also been supplied, then it is a *fully specified* diagram.

A classic example (Breiman 1980) helps to motivate the use of influence diagrams. An epidemiologist in the 1930's with the benefit of modern computational tools (but not modern medicine) is attempting to track down the cause of polio epidemics. During a regression run, he comes across a startling relationship: there appears to be a high correlation between the incidence of polio and the consumption of Coca-Cola. Checking the data with all of the measures available to him, this relationship appears to be highly significant. He obtains additional data from earlier years, and it confirms the correlation.

He can represent this relationship with the influence diagram shown in Figure 1a. Each of the nodes corresponds to an uncertain quantity, so it is a probabilistic node, which is drawn as an oval. The arc from *Coca-Cola Sales* to *Polio Cases* indicates that the two variables might be dependent, and that a probability distribution for *Polio Cases* would be conditional to the outcome for *Coca-Cola Sales*. To obtain the full joint distribution, he would also need a marginal distribution for *Coca-Cola Sales*. He could equivalently represent this relationship by the diagram in Figure 1b, in which the arc is pointed in the opposite direction. In that case, he would obtain the joint distribution from a marginal distribution for *Polio Cases* and a conditional distribution for *Coca-Cola Sales* given the number of polio cases. It is important to note that an arc in the influence diagram is not causal. Rather, it indicates (possible) probabilistic dependence, so there is nothing inconsistent with the two models in Figure 1.

Our epidemiologist is facing a dilemma: Coca-Cola is a popular beverage and yet he observes that its usage is linked to the outbreak of a major disease. The solution lies in another, more complicated influence diagram, shown in Figure 2. In this model *Polio Cases* is conditioned on *Swimming Pool Usage*. The distributions for *Swimming Pool Usage* and *Coca-Cola Sales* are each conditioned by both *Temperature* and *Income*. Even more interesting than which arcs are present is the conditional independence implied by the *missing* arcs. In this model, *Coca-Cola Sales* and *Polio Cases* are indeed dependent, but they are conditionally independent given *Temperature* and *Income*.

The preceding story is a parable and should not be taken literally. It is clear, nonetheless, that one must resist the temptation to imply "cause" from correlation, and that the influence diagram representation is a clear and powerful tool for building models with conditional independence (even when the diagram is only partially specified). It will be shown later that it is also a convenient representation for assessment and evaluation of Gaussian probabilistic and decision models.

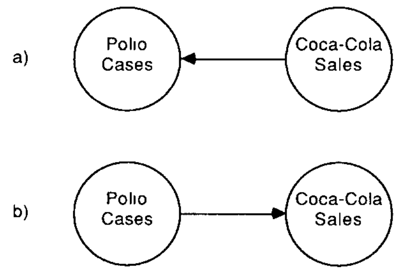

FIGURE 1. Epidemiologist's Simple Influence Diagram.

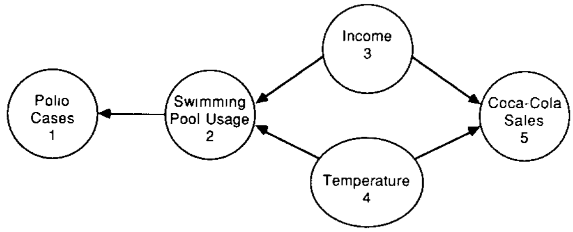

FIGURE 2. Epidemiologist's More Complicated Influence Diagram.

For more information, a complete and readable introduction to influence diagrams appears in Howard and Matheson (1981). A discussion of some of the issues on causal relationships and model building appears in Shachter and Heckerman (1987).

2. Gaussian Influence Diagram

In this section the concepts of influence diagrams are formalized. The basic graphical and probabilistic elements are defined for the Gaussian influence diagram.

Assign indices to the nodes and variables in the model, so that the nodes are given by $N = \{1, \ldots, n\}$ and correspond to variables $X_1, \ldots, X_n$. The conditioning variables for X_j have indices in the set of *conditional predecessors* $C(j)$, $N \supset C(j)$, and are indicated in the graph by arcs from the nodes $C(j)$ into node j. Likewise, the set of indices of all of the variables conditioned by X_i, the (*direct*) *successors* $S(i)$ of the node i, is given by $S(i) = \{j \in N: i \in C(j)\}$.

As a convention, a lower-case letter represents a single node in the graph and an upper-case letter represents a set of nodes. If J is a set of nodes then X_J denotes the vector of variables indexed by J. In this notation $X_{C(j)}$ are the conditioning variables for X_j. When the order of the nodes in a set is important it will be called a *sequence*, and written as a column vector, such as $\mathbf{s} = (1, \ldots, n)$.

A sequence of nodes $\mathbf{s}$ in the graph is said to be *ordered* if all of the conditional predecessors of any node in the graph precede it in $\mathbf{s}$. (This is equivalent to requiring that the nodes along any directed path in the graph appear in that same order in $\mathbf{s}$.) Because the influence diagram can have no directed cycles there is always at least one ordered sequence for the nodes in the graph. To create such a sequence, start with a node which has no predecessors. Now, add one node at a time to the end of the sequence, choosing any node whose predecessors, if any, are already in the sequence. In general, the order of the nodes is not unique, since there might be more than one node which could be selected at the same time. For example, there are six possible ordered sequences for the nodes in Figure 2: (3, 4, 2, 1, 5), (4, 3, 2, 1, 5), (3, 4, 2, 5, 1), (4, 3, 2, 5, 1), (3, 4, 5, 2, 1), and (4, 3, 5, 2, 1).

When an operation is performed on a set of nodes *in order* (*with respect to* $\mathbf{s}$), it means in the same order as the nodes appear in an ordered sequence $\mathbf{s}$; *in reverse order* (*with respect to* $\mathbf{s}$) means in the opposite order from which the nodes appear in $\mathbf{s}$.

The influence diagram graphical structure represents a particular factorization of a joint distribution into conditional distributions. In fact, for any sequence of nodes $\mathbf{s}$ there is an influence diagram which factors the joint distribution in that order:

$$\begin{aligned}\Pr\{X_N\} &= \Pr\{X_{s_1}\} \cdot \Pr\{X_{s_2} \mid X_{s_1}\} \cdot \;\cdots\; \cdot \Pr\{X_{s_n} \mid X_{s_1}, \ldots, X_{s_{n-1}}\} \\ &= \Pr\{X_{s_1} \mid X_{C(s_1)}\} \cdot \Pr\{X_{s_2} \mid X_{C(s_2)}\} \cdot \;\cdots\; \cdot \Pr\{X_{s_n} \mid X_{C(s_n)}\},\end{aligned}$$

where $\Pr\{X_{s_1} \mid X_{C(s_1)}\}$ is an unconditional distribution since $C(s_1)$ must be the null set. By the construction of the ordered sequence, $C(s_j)$ must be a subset of $\{s_1, \ldots, s_{j-1}\}$.

If it is a proper subset, then there is some conditional independence revealed in the diagram.

For the remainder of the paper, assume that the influence diagram is fully specified and that X_N has a multivariate normal distribution, so that the influence diagram is Gaussian. The joint distribution for X_N can be characterized by a set of unconditional means $\mu_N = E[X_N]$ and a covariance matrix

$$\Sigma_{NN} = \text{Var}\,[X_N] = E[X_N X_N^T] - E[X_N]E[X_N^T].$$

The influence diagram, however, is based on conditional probability distributions and they have a special form for a Gaussian model: each conditional variable $\{X_j | X_{C(j)}\}$ is normally distributed with mean $(\mu_j + \sum_{k \in C(j)} b_{kj}(X_k - \mu_k))$ and variance v_j (fixed for a given set of conditioning variables). Therefore, the conditional distribution is characterized by the mean μ_j, conditional variance v_j, and linear coefficients b_{kj}. When v_j is zero, X_j is simply a deterministic linear function of $X_{C(j)}$.

The conditional model can also be viewed as a set of regression equations:

$$X_j = \mu_j + \sum_{k \in C(j)} b_{kj}(X_k - \mu_k) + (v_j)^{1/2} Z_j \qquad \text{for} \qquad j = 1, \ldots, n, \tag{1}$$

in which $Z_1, \ldots, Z_n$ are independent standard normal random variables. The matrix $\mathbf{B} = [b_{ij}]$ can be thought of as regression coefficients, or expressed in terms of Yule's (1907) partial regression coefficient β,

$$b_{kj} = \beta_{jk \cdot \{1, \ldots, j-1\} \setminus \{k\}} = \beta_{jk \cdot C(j) \setminus \{k\}},$$

where "\" is the operator for set subtraction, $J \setminus K = \{x \in J \colon x \notin K\}$. (By definition, $b_{kj} = 0$ if $k \notin C(j)$.) The relationship between the conditional and unconditional models is examined in §5 and Appendix B.

In the standard influence diagram, or even the standard regression, the conditional distribution could be specified simply as:

$$X_j = b_{0j} + \sum_{k \in C(j)} b_{kj} X_k + (v_j)^{1/2} Z_j \qquad \text{for} \qquad j = 1, \ldots, n,$$

where b_{0j} is a specified constant. This has the desirable property that a change in one distribution would be localized within the diagram. However, the unconditional means are more useful and understandable statistics than the b_{0j}'s, and it is more convenient to think of b_{0j} in terms of the means as $[\mu_j - \sum_{k \in C(j)} b_{kj}\mu_k]$.

The cost of this convenience is some extra bookkeeping throughout the diagram whenever the mean for any variable changes. In a diagram with two nodes and ordered sequence [1 2], when the mean for X_1 changes from μ_1 to μ_1', the new value for μ_2 is

$$\mu_2' = E[X_2] = E[E[X_2 | X_1]] = E[\mu_2 + b_{12}(X_1 - \mu_1)] = \mu_2 + b_{12}(\mu_1' - \mu_1).$$

If X_2 had a successor X_3 then, of course, μ_3 would change as a result of the change to μ_2, and so on. In general, if the mean for X_i changes, the change must be *propagated* along every directed path which emanates from node i. Letting $\mathbf{U}_{ij}$ be the effect on μ_j of a change in μ_i, the new means are $\mu_j' = \mu_j + \mathbf{U}_{ij}(\mu_i' - \mu_i)$ for $j \in N$. $\mathbf{U}_{ij}$ is computed by summing, over all of the directed paths from i to j, the product of the arc coefficients on each path. (As a special case, $\mathbf{U}_{ii} = 1$.) There is an alternate computation for mean propagation, which is significantly more efficient as the diagram gets larger: for each node along an ordered sequence, accumulate the net effect to that node's mean resulting (directly and indirectly) from a change in μ_i. In this way, no arc in the diagram is visited more than once. Assuming that s is an ordered sequence for the diagram, the procedure visits each node j in order with respect to s:

$$\mathbf{U}_{jj} \leftarrow 1,$$

$$\mathbf{U}_{ij} \leftarrow 0 \text{ for all } i \neq j.$$

For each $k \in C(j)$,

$$\mathbf{U}_{ij} \leftarrow \mathbf{U}_{ij} + \mathbf{U}_{ik}\mathbf{B}_{kj} \text{ for all } i \text{ preceding } j \text{ in s}.$$

(For more information about U, please refer to Appendix B.)

3. Probabilistic Reductions

There are several "reductions" which transform one influence diagram into another, without changing the underlying joint distribution for the variables of interest. These transformations are the elimination of "barren" nodes (Shachter 1986), the reversal of arcs (Howard and Matheson 1981), and the reduction of probabilistic nodes (Shachter 1988). These reductions are the basic steps in algorithms for probabilistic inference and decision analysis, presented later in the paper.

The simplest influence diagram reduction is the "barren" node reduction. Suppose that a variable in the model is not of direct concern to the decision maker and its outcome will not be observed. This is the type of variable which statisticians term a *nuisance parameter*. (This definition will be refined after value functions and informational predecessors are introduced in §§6 and 7.) If the node for a nuisance parameter has no successors in the graph, then it is called a *barren* node, and it is irrelevant for this particular problem. Therefore, it can be *eliminated* from the influence diagram without affecting the solution. In Figure 2, if the decision maker were concerned with *Polio Cases* and will never get to observe *Coca-Cola Sales*, then *Coca-Cola Sales* would be a barren node. If, on the other hand, he can observe *Coca-Cola Sales* and not *Temperature*, then *Coca-Cola Sales* might provide valuable information. Likewise, even though *Polio Cases* has no successors, it would not be barren since it is of direct concern to the decision maker.

PROPOSITION 1. Barren Node Reduction. *If node i is barren for the problem at hand, then it can be eliminated from the influence diagram.*

When a node j is eliminated from the diagram it is dropped from the set of nodes: $N \leftarrow N \setminus \{j\}$. The information within the node should not be discarded but the node must be excluded from all future operations on the influence diagram and its associated X_N.

The most important reduction on the influence diagram is the reversal of an arc between probabilistic nodes. This operation is the influence diagram representation of Bayes' Theorem. Figure 3 illustrates the reversal of an arc between nodes 4 and 5. Before the operation 4 was a conditional predecessor for 5; afterwards 5 is a conditional predecessor for 4. In the process, 3 also becomes a conditional predecessor for 4, while 1 becomes a conditional predecessor for 5. One interpretation is that in order to apply Bayes' Theorem, both 4 and 5 needed to share a conditioning set, so arcs (3, 4) and (1, 5) were added with arc coefficients zero.

Of course underlying the influence diagram graph are the conditional probability distributions. Corresponding to the reversal of the arc in the diagram is the calculation of new conditional distributions for X_4 and X_5. Their new conditional variances and arc coefficients are also shown in Figure 3. These results are summarized in Theorem 1, which is proved in Appendix A.

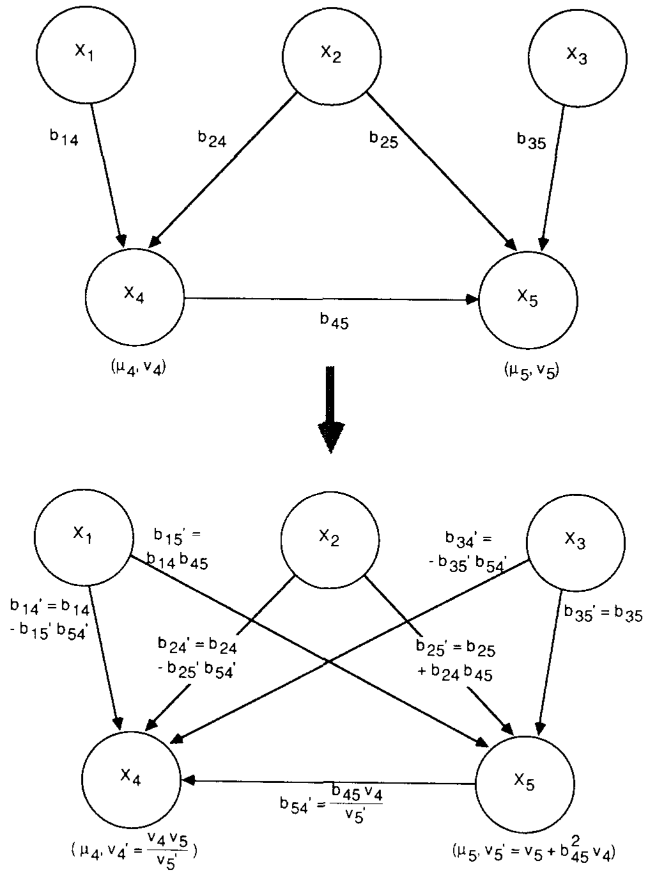

FIGURE 3 Reversing the Arc from Node 4 to Node 5.

THEOREM 1. Arc Reversal. *Given that node $i \in C(j)$ and no other directed path exists from i to j,*

$$E[X_j | X_K] = \mu_j + \sum_{k \in K} (b_{kj} + b_{ki} b_{ij})(X_k - \mu_k), \qquad \textit{and}$$

$$\operatorname{Var}[X_j | X_K] = v_j + b_{ij}^2 v_i,$$

where $K = (C(j) \cup C(i)) \setminus \{i\}$. If neither v_i nor v_j is zero, then

$$E[X_i | X_j, X_K] = \mu_i + \sum_{k \in K} (b_{ki} - (b_{kj} + b_{ki} b_{ij}) b_{ji})(X_k - \mu_k) + b_{ji}(X_j - \mu_j),$$

$$\operatorname{Var}[X_i | X_j, X_K] = v_i v_j / [v_j + b_{ij}^2 v_i], \qquad \textit{and}$$

$$b_{ji} = b_{ij} v_i / [v_j + b_{ij}^2 v_i].$$

Operationally, the reversal of the arc proceeds in two steps. First:
$K \leftarrow (C(j) \cup C(i) \setminus \{i\}$,
$v_j^{\text{old}} \leftarrow v_j$,
$v_j \leftarrow v_j^{\text{old}} + b_{ij}^2 v_i$,
$b_{kj} \leftarrow b_{kj} + b_{ki} b_{ij}$ *for* $k \in K$,
$C(j) \leftarrow \{k \in K\colon b_{kj} \neq 0\}$.

If the new v_j is zero, X_i and X_j are conditionally independent given X_K. In that case, it would be unnecessary to construct an arc from j to i. Therefore the second step is:

If $v_j > 0$ then
$v_i^{\text{old}} \leftarrow v_i$,
$v_i \leftarrow v_i^{\text{old}} v_j^{\text{old}} / v_j$,
$b_{ji} \leftarrow b_{ij} v_i^{\text{old}} / v_j$,
$b_{ki} \leftarrow b_{ki} - b_{kj} b_{ji}$ *for* $k \in K$,
$C(i) \leftarrow \{k \in K \cup \{j\}: b_{ki} \neq 0\}$,
$b_{ij} \leftarrow 0$.

An important feature of the Gaussian arc reversal is that the new value for v_i is computed via a quotient and product of nonnegative numbers. Thus, roundoff errors cannot produce negative conditional variances. Traditional formulae for applying Bayes' rule to the multivariate normal distribution can produce negative variances from roundoff errors (Bierman 1977). This makes influence diagram processing attractive for real-time decision systems, where correcting for negative variances can be costly.

The condition that there be no other directed path from i to j before reversal prevents the creation of a directed cycle. It is equivalent to requiring that there be an ordered sequence for the original diagram in which node i immediately precedes node j. If the order of nodes i and j were switched in that sequence, it would become a valid ordered sequence for the diagram resulting from the arc reversal.

The arc reversal operation can be explained intuitively in terms of "flows" of variance. For a given ordering of the nodes in the diagram, each variable introduces uncertainty into the model through its conditional variance and propagates it to its successor variables based on the coefficients on its outgoing arcs. It thus contributes to (or "explains" some of) the unconditional variance of any variable which is "downstream" from it in the graph. When the arc from i to j is reversed, although there is no change to the underlying joint distribution, the conditional variances and arc coefficients must be revised. First, X_i will no longer be explaining some of the variance in X_j, so v_j must be incremented by the variance which had been contributed by X_i. Conversely, X_j will now be explaining some of the variance in X_i so v_i must be discounted to adjust for that. Finally, the total flow of variance to both i and j from their common predecessors must remain unchanged, despite the reversal of the arc. This is accomplished through the addition of arcs and the adjustment of arc coefficients.

This flow of variance can be misleading, however, because it depends on the ordering of the variables in the diagram. It is the joint distribution which is preserved during the arc reversal operation. Consider two nodes i and j, where $i \in C(j)$ but there is no other path between i and j. In that case, their conditional covariance is

$$\begin{aligned}\text{Cov}\,[X_i, X_j | X_{C(i)}] &= E[X_i X_j | X_{C(i)}] - E[X_i | X_{C(i)}] E[X_j | X_{C(i)}] \\ &= E[b_{ij} X_i^2 | X_{C(i)}] - E[X_i | X_{C(i)}] E[b_{ij} X_i | X_{C(i)}] \\ &= b_{ij} \,\text{Var}\,[X_i | X_{C(i)}] = b_{ij} v_i,\end{aligned}$$

the conditional variance of X_i weighted by the arc coefficient b_{ij}. (Some offsetting terms are omitted in the second step.) In general, there might be multiple directed paths from node i to node j, and the conditional covariance is given by

$$\text{Cov}\,[X_i, X_j | X_{C(i)}] = \mathbf{U}_{ij} v_i,$$

where $\mathbf{U}_{ij}$ is the coefficient introduced in §2 to propagate changes in the means. Since the joint distribution among the variables is not changed by the reversal operation, the conditional covariances do not change either. However, the flow is not preserved from any node i whose conditional predecessors $C(i)$ change as a result of the arc reversal. For example, when arc (i, j) is reversed, the predecessor sets $C(i)$ and $C(j)$ are revised, and the flows from the nodes i and j to their successors change in two ways: the conditional variances v_i and v_j are adjusted and the available paths from i and j change when the arc between them is reversed. Nonetheless, the flow is maintained from their predecessors during the reversal operation, since their predecessors' conditional predecessors are unaffected by the operation.

The two basic reductions, arc reversal and barren node reduction, can be combined into a compound reduction, called the *probabilistic node reduction*. Suppose a nuisance parameter in the model has successors. If the arcs to its successors were reversed then it would become a barren node and could be eliminated from the diagram. This is accomplished by repeatedly applying Bayes' Theorem until all of the relevant information has been extracted from the node.

PROPOSITION 2. Probabilistic Node Reduction. *Any probabilistic node i (corresponding to a nuisance parameter) can be reduced from the diagram. Reverse the arcs from node i into each successor in order (with respect to an ordered sequence). At that point, node i has no successors and is barren. Therefore it can be eliminated from the diagram.*

By reversing the arcs into the successors in the order of the sequence, each reversal will be performed with the first successor in the sequence, so the conditions of Theorem 1 are satisfied. One can think of the process as reordering the nodes in the diagram through arc reversal until the nuisance parameter is the last node in an ordered sequence.

4. Probabilistic Inference

A general problem of interest is to determine $\Pr\{X_J | X_K\}$, where J and K are arbitrary subsets of N. This will be called the *probabilistic inference problem*. Reductions are organized into an algorithm to solve this problem within the influence diagram. An extension of this algorithm allows the computation of the posterior distribution for the influence diagram given that some of the variables have been observed.

The probabilistic inference problem is solved by transforming the influence diagram through reductions so that only nodes in $J \cup K$ remain, and all of the nodes in K precede the nodes in J in an ordered sequence (with nodes in $J \cap K$ in the middle). With respect to the inference problem, the nodes in $N \backslash (J \cup K)$ are nuisance parameters, so they should be reduced (or at least reordered so they follow the nodes in $J \cup K$). Thus, our goal is a partial ordering of the nodes in the diagram. A simple algorithm, related to a "bubble sort" using the arc reversal operation, can achieve the desired order: starting with an ordered sequence, while there are two adjacent nodes in the sequence which are not in the desired order, switch them (after reversing the arc between them if there is one). When these conditions are no longer satisfied, the diagram must be in the proper order, and the process will terminate in a finite number of steps. The following is a more structured version of this algorithm, designed to minimize unnecessary arc reversals:

1. Construct an ordered sequence $\mathbf{s}$.
2. Reduce each node in $N \backslash (J \cup K)$, in reverse order (with respect to $\mathbf{s}$).
3. Visiting each node $i \in (J \backslash K)$ in reverse order (with respect to $\mathbf{s}$), reverse the arcs from i to K in order (with respect to $\mathbf{s}$).
4. Visiting each node $i \in (J \cap K)$ in reverse order (with respect to $\mathbf{s}$), reverse the arcs from i to $(K \backslash J)$ in order (with respect to $\mathbf{s}$).
5. If the variables X_K have been observed, instantiate them in order, following the steps given at the end of the section.

As an example, consider the model shown in Figure 2. Suppose that the goal is to obtain the probability distribution for *Swimming Pool Usage*, *Polio Cases*, and *Temperature* given *Polio Cases* and *Coca-Cola Sales*. Thus $J = \{1, 2, 4\}$ and $K = \{1, 5\}$.

First, create an ordered sequence for the nodes: (3, 4, 2, 1, 5). *Income* is the only nuisance parameter and it is reduced by first reversing the arc from *Income* to *Swimming*

Pool Usage, and then the arc to *Coca-Cola Sales*, to obtain the diagram shown in Figure 4a. Notice the arcs that were added from *Temperature* to *Income* and from *Swimming Pool Usage* to *Coca-Cola Sales* during the arc reversals. *Income* is now a barren node and can be eliminated.

Visiting the nodes in $J \backslash K$, start with *Swimming Pool Usage*. Both of its successors are in K, so first reverse the arc into *Polio Cases* and then the one into *Coca-Cola Sales* to obtain the diagram in Figure 4b. Finally, *Temperature* has two successors in K, so first reverse the arc into *Polio Cases* and then the one into *Coca-Cola Sales*, to obtain the diagram in Figure 4c.

Visiting the nodes in $J \cap K$, there is only *Polio Cases*. It has one successor in $K \backslash J$, so reverse the arc into *Coca-Cola Sales*. The diagram in Figure 4d is the final result.

Suppose that the sales figures for Coca-Cola and the medical reports for polio actually become available. Starting with the diagram in Figure 4d, replace the prior mean and variance for *Coca-Cola Sales* with the observed outcome and zero, respectively. (Remember to propagate the change in the mean throughout the diagram, as explained in §2.) Similarly for *Polio Cases*, substitute the observed outcome and zero for the mean and variance and propagate the change in the mean.

In general, this method (which computer scientists call "instantiation") can be used to compute the posterior distribution for the entire model whenever evidence becomes observable for variables X_J. First solve the probabilistic inference problem for $\Pr\{X_N | X_J\}$. Then, for each observed variable in order, replace the prior means and variances with the observed outcomes and zeros, respectively, and propagate the changes in the means.

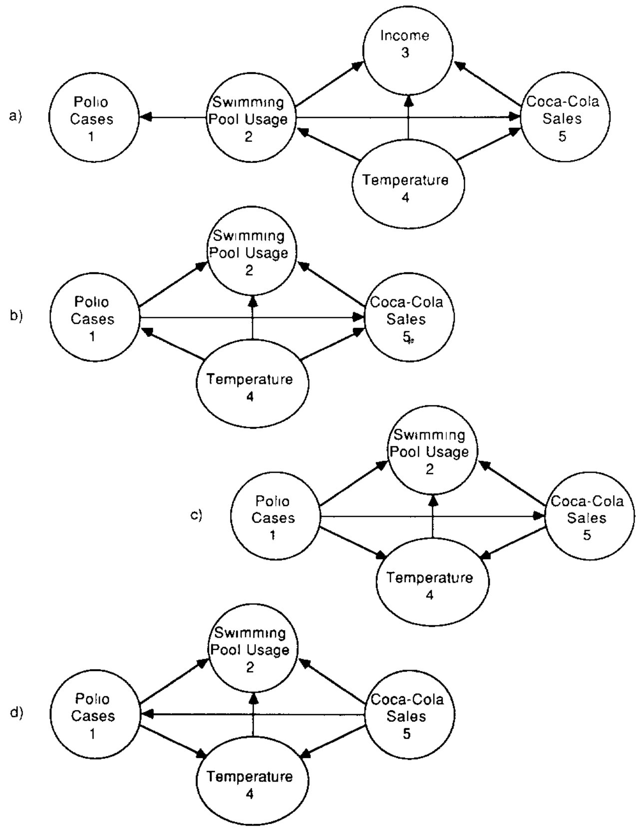

FIGURE 4. Reversing Arcs to Solve the Inference Problem.

While the propagation of the means is the step in which the model "learns" from the evidence, it is the application of Bayes' Theorem in the inference problem that requires computational effort and that makes that learning process possible.

5. Relationship to Covariance Representation

The standard representation of the multivariate normal distribution is in terms of its unconditional statistics, the mean vector and covariance matrix. In this section, the assessment of the Gaussian influence diagram is shown to be simpler than the direct assessment of the covariance matrix. In addition, algorithms are provided to transform a problem from one representation to the other.

The theoretical relationship between the two representations is captured in the proofs in Appendix B. Basically, the influence diagram is closely related to a special $\mathbf{U}^T\mathbf{D}\mathbf{U}$ decomposition of the covariance matrix. It is this decomposition, similar to one in Yule (1907), that is exploited in the transformation algorithms and leads to the following theorem:

THEOREM 2. Rank of Covariance Matrix. *Σ_{NN} is positive (semi-)definite if and only if $v_N > (\geq)\ 0$. Furthermore, the rank of Σ_{NN} is equal to the number of nonzero elements in v_N.*

This result is particularly important for the assessment and maintenance of a multivariate normal model, since covariances must be positive semidefinite (PSD) at all times. It shows how to verify whether a Gaussian influence diagram has a positive (semi-)definite covariance matrix, and how to determine its rank. These are much more difficult questions to answer in the covariance representation.

The assessment of a Gaussian influence diagram is straightforward, given the influence diagram graph. For each variable an unconditional mean and conditional variance must be elicited, along with a linear coefficient for each arc. In many problems these quantities have natural physical meanings for the decision maker. It would be absurd to assess a negative variance, so the corresponding covariance matrix will always be PSD. It will, in fact, be positive definite (PD) unless some variable is a deterministic function of other variables in the model.

The standard approach is more problematic, since a covariance matrix must be assessed along with the unconditional means. Covariances rarely have a natural physical meaning to the decision maker, and the assessed matrix is unlikely to be PSD. Complex procedures such as the conjugate gradient method in Oren (1981) are needed to restore the PSD structure.

The influence diagram also has an advantage in the number of quantities to assess. In the worst case, each method requires $n(n+3)/2$ quantities: a mean vector and triangular matrix for covariance; mean and variance vectors and a strictly triangular coefficient matrix for the influence diagram. Whenever quantities are zero in the covariance case, there is an influence diagram model with corresponding zeros. On the other hand, the influence diagram becomes much simpler when there is conditional independence, while the covariance matrix does not. For example, the diagram in Figure 2 would in general have a covariance matrix without any zero elements, so 20 quantities would have to be assessed in the covariance representation, and they will almost surely violate the conditional independence in the diagram. By comparison, the Gaussian influence diagram not only enforces that conditional independence, but exploits it to require the assessment of only 15 quantities. As the size of the model increases, the savings become more dramatic.

Once a Gaussian influence diagram has been assessed, the corresponding covariance matrix can be constructed using the following algorithm:

1. Create an ordered sequence **s** for the variables in the influence diagram.
2. For $i = 1, \ldots, n$

$$j \leftarrow \mathbf{s}_i,$$

$$\Sigma_{ij} \leftarrow \Sigma_{ji} \leftarrow \sum_{k \in C(j)} \Sigma_{ik} b_{kj} \qquad \text{for} \qquad i = \mathbf{s}_1, \ldots, \mathbf{s}_{i-1},$$

$$\Sigma_{jj} \leftarrow v_j + \sum_{k \in C(j)} \Sigma_{jk} b_{kj}.$$

The theory on which this algorithm is based is presented in Appendix B. It is closely related to the inversion of the upper triangular matrix $(\mathbf{I} - \mathbf{B})$.

A related algorithm constructs a fully specified Gaussian influence diagram from the covariance representation given an arbitrary sequence **s** of the nodes. It is just the inverse of the algorithm above, with some special care taken to maintain a linearly independent basis. Afterwards, **s** will be an ordered sequence for the constructed diagram:

1. $\mathbf{t} \leftarrow (\ )$ (a null sequence)
   $\Rightarrow \Sigma_{\mathbf{tt}}^{-1} \leftarrow \mathbf{I}_0$.
2. For $i = 1, \ldots, n$
   $j \leftarrow \mathbf{s}_i$,
   $\mathbf{B}_{\mathbf{t}j} \leftarrow \Sigma_{\mathbf{tt}}^{-1} \Sigma_{\mathbf{t}j}$,
   $b_{kj} \leftarrow 0 \qquad \text{for} \qquad k \notin \mathbf{t}$,
   $C(j) \leftarrow \{k \in \mathbf{t}: b_{kj} \neq 0\}$,
   $v_j \leftarrow \Sigma_{jj} - \sum_{k \in C(j)} \Sigma_{jk} b_{kj}$.
   If $v_j > 0$ then
   $\mathbf{t} \leftarrow (\mathbf{t}, j)$

$$\Rightarrow \Sigma_{\mathbf{tt}}^{-1} \leftarrow \begin{bmatrix} \Sigma_{\mathbf{tt}}^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} + \frac{1}{v_j} \begin{bmatrix} \mathbf{B}_{\mathbf{t}j} \\ -1 \end{bmatrix} [\mathbf{B}_{\mathbf{t}j}^T \ -1].$$

The constructed sequence **t** is a "stochastic basis" for the model: $X_{\mathbf{t}}$ explains all of the variance and $\Sigma_{\mathbf{tt}}$ is a maximal-dimension, linearly independent submatrix of the covariance matrix Σ_{NN}.

6. The Quadratic Value Function

In this paper, the variables in the model are assumed to have a multivariate normal distribution, and the influence diagram is specialized to take advantage of its properties. In this section, the previous results are generalized to allow for a quadratic value function, a standard extension of the multivariate normal model. The expected value of this value function is maintained as the influence diagram is transformed via reductions. In the next sections, this value function will serve as the criterion for making decisions.

The *value function* is the conditional expected value of a *criterion value*, X_0, given the values of the other variables in the model,

$$V(X_N) = E[X_0 | X_N] = (1/2) X_N^T \mathbf{Q} X_N + \mathbf{p}^T X_N + r,$$

where $\mathbf{Q}$ is assumed symmetric. There can only be one criterion variable, and it is not treated as a regular variable in the influence diagram, so $0 \notin N$. As the diagram is transformed during analysis and variables are reduced, the values and dimensions of $\mathbf{Q}$, $\mathbf{p}$, and r will be revised.

Although the criterion variable does not have a corresponding node in the Gaussian influence diagram, it is important to recognize its conditional predecessors when structuring the problem. Once the problem has been structured, however, there is little insight to be gained from including a criterion node in the diagram. At the time of assessment, the (quadratic) coefficients in the value function cannot be displayed in the diagram. Moreover, once the evaluation of the diagram commences, it is difficult to maintain a minimal conditional predecessor set for the criterion node; instead it is convenient to assume that all of the variables in the model condition the value function.

The criterion variable represents the decision maker's objective, so any (other) variable in the model which cannot be observed is a nuisance parameter and can be reduced from the diagram. On the other hand, since the value function depends on every variable in the model, nodes can no longer become barren. Instead, they must be reduced using *probabilistic node removal*, based on the following result, proved in Kenley (1986) and similar to operations in Olmsted (1983) and Shachter (1986).

THEOREM 3. Probabilistic Node Removal. *Given node $j \in N$ has no successors, let $K = N \setminus \{j\}$. If*

$$V(X_N) = E[X_0 | X_N] = (1/2) X_N^T \mathbf{Q} X_N + \mathbf{p}^T X_N + r,$$

where $\mathbf{Q}$ *is symmetric, then*

$$V_{new}(X_K) = E[X_0 | X_K] = E[V(X_N) | X_K]$$

$$= (1/2) X_K^T \mathbf{Q}_{new} X_K + \mathbf{p}_{new}^T X_K + r_{new}, \qquad \textit{with}$$

$$\mathbf{Q}_{new} \leftarrow \mathbf{Q}_{KK} + 2\mathbf{Q}_{Kj}\mathbf{B}_{Kj}^T + \mathbf{B}_{Kj}\mathbf{Q}_{jj}\mathbf{B}_{Kj}^T,$$

$$\mathbf{p}_{new} \leftarrow \mathbf{p}_K + \mathbf{B}_{Kj}\mathbf{p}_j + (\mathbf{Q}_{Kj} + \mathbf{B}_{Kj}\mathbf{Q}_{jj})(\mu_j - \mathbf{B}_{Kj}^T \mu_K),$$

$$r_{new} \leftarrow (1/2)\mathbf{Q}_{jj}[v_j + (\mu_j - \mathbf{B}_{Kj}^T \mu_K)^2] + \mathbf{p}_j(\mu_j - \mathbf{B}_{Kj}^T \mu_K) + r.$$

Operationally, the procedure for probabilistic node removal is:

1. Reverse the arcs from node j to each of its successors in order, just as in the probabilistic node reduction, so that node j will have no successors.
2. Take expectation of the value function with respect to $\{X_j | X_{N\setminus\{j\}}\}$, using the formulae in Theorem 3 to update $\mathbf{Q}$, $\mathbf{p}$, and r.
3. Eliminate node j from the influence diagram.

The probabilistic inference problem becomes much simpler when there is a criterion variable. The problem in that case is to determine $E[X_0 | X_K]$ where $0 \notin K$. The algorithm given in §4 now simplifies to performing the probabilistic node removal on each of the nodes which are not in K, in reverse order.

7. Decisions in the Gaussian Influence Diagram

This section introduces the two remaining components of the Gaussian influence diagram: nodes representing decisions and arcs indicating the information available at the time those decisions must be made. An example is introduced to illustrate the Gaussian influence diagram with decisions.

A *decision node* j, drawn as a rectangle in the influence diagram, represents a variable which is under the control of the decision maker. The arcs into a decision node indicate which variables' outcomes will be observed by the decision maker prior to his selection of the value for X_j. The nodes corresponding to those variables are called *informational predecessors*, and are denoted by the set $I(j)$. In other words, the decision maker will know the realization of variables $X_{I(j)}$ at the time he must choose X_j.

The decision maker's choice for X_j will be that value that maximizes the expected value of the criterion variable X_0, introduced in the previous section. In the Gaussian influence diagram, that expected value is a quadratic function of the variables in the model.

There are several important restrictions on influence diagrams with decisions which apply to the Gaussian influence diagram as well (Howard and Matheson 1981). First, if there are multiple decisions in the diagram they must be totally ordered: there must be only one order in which they can appear in an ordered sequence. This condition is equivalent to there being a directed path which contains all of the decisions, and their order corresponds to the order in time in which the commitments are made. As a result, it is unambiguous which decision is the *latest* decision in the influence diagram. Second, there is the requirement of *no-forgetting arcs*: if a variable will be observed at the time of one decision, it (and the value chosen for the decision variable) must be observed at the time of all subsequent decisions in the diagram. Third, even though the arcs into decision nodes have a different meaning from those into probabilistic nodes, a directed cycle is still not permitted.

Gaussian influence diagrams with decisions can be illustrated by a simple example of a marketing problem (Kenley 1986). A consultant has purchased a powerful computer for use in her practice. She plans to bill her clients for computer usage at an hourly rate and expects that the consulting practice will leave the computer unused most of the time. Therefore, she believes that there is a good opportunity to sell some of these idle hours to time-sharing users. She must decide on a fee structure for her consulting clients and her time-sharing users so as to maximize her expected profit.

The influence diagram for the consultant's problem is shown in Figure 5. The decision maker is free to choose hourly rates for both services. She believes that the number of *Consulting Hours* sold will depend on *Consulting Price*, and that the cost of the computer facilities for consulting, *Consulting Cost*, will in turn depend on *Consulting Hours*. Before she commits to an hourly rate for time-sharing users, *Time Share Price*, her accountant will work up a *Consulting Estimate*, his best guess of the number of *Consulting Hours* she will bill to her clients. She believes that the number of *Time Share Hours* purchased will depend on *Time Share Price* and the number of hours the computer is not busy with consulting work, *Idle Hours*. She thinks that the cost of running the time-sharing service, *Time Share Cost*, will depend on the hours purchased, *Time Share Hours*, and the hours that it is available, *Idle Hours*. Her profit will be the difference between total revenues and costs. Although *Profit* is not really a node in the Gaussian influence diagram, it is drawn while the problem is structured.

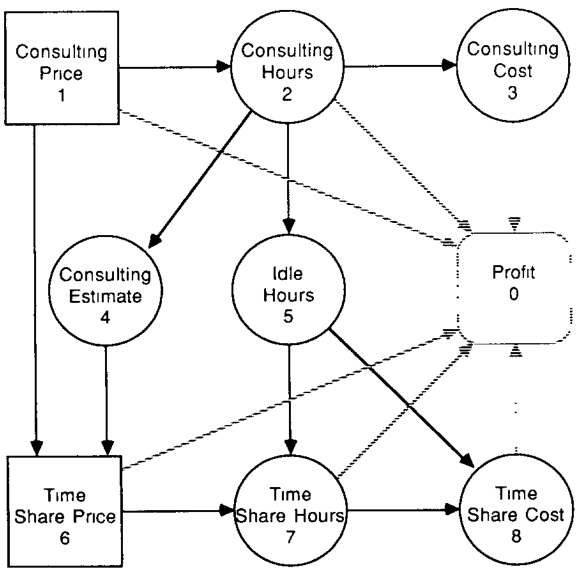

FIGURE 5. Partially Specified Diagram for the Consultant's Problem.

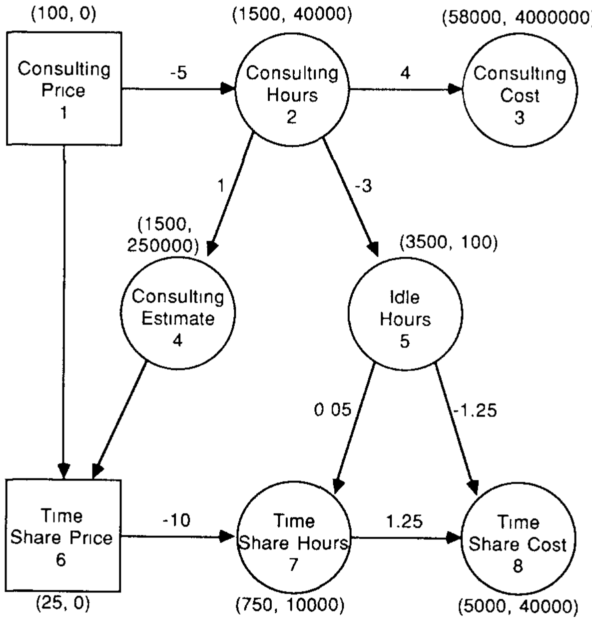

FIGURE 6. Fully Specified Diagram for the Consultant's Problem.

At this point the parameters in the influence diagram must be assessed. In the process, values are needed for the decision variables, in order to assess their successors. For example, it is impossible to estimate the expected number of *Consulting Hours* without assigning some *reference value* to the *Consulting Price*. This reference value might be the current operating policy or the result of a simple, deterministic optimization. In theory, it can have a conditional normal distribution just like a probabilistic variable. In practice, only its expected value will ever be examined, so it might just as well be treated as a constant value to be assessed. Therefore, the arc coefficients and conditional variance for the reference value can be set to zero.

In the consultant's problem, she is comfortable with \$100 and \$25 as reference values for *Consulting Price* and *Time Share Price*, respectively. She expects to sell 1500 *Consulting Hours* at that reference price. She believes that she will sell 5 fewer hours for every dollar she raises the price, and that given a price there is a standard deviation of 200 hours in her estimate for *Consulting Hours*. These values and her other assessments are shown in Figure 6. The mean and conditional variance are written next to each node, and the arc coefficient is next to each arc. She believes that the value function for profit is *Consulting Price* × *Consulting Hours* + *Time Share Price* × *Time Share Hours* − *Consulting Cost* − *Time Share Cost*, or

$$V(X_N) = X_1 \cdot X_2 + X_6 \cdot X_7 - X_3 - X_8 = (1/2)x_N^T \mathbf{Q} x_N + \mathbf{p}^T x_N + r, \qquad \text{where}$$

$$\mathbf{Q} \leftarrow \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{p} \leftarrow (0, 0, -1, 0, 0, 0, 0, -1), \qquad \text{and} \qquad r \leftarrow 0.$$

8. Evaluating the Gaussian Influence Diagram with Decisions

In this section, an algorithm is derived to evaluate the full Gaussian influence diagram with quadratic value function and decision variables. The algorithm reduces each node from the diagram in turn, until all that remains is the optimal expected value of the criterion as a constant in the value function. Each of the decision variables is replaced by a policy for selecting an optimal value given the information available at the time of the decision. The algorithm is demonstrated with the evaluation of the consultant's problem introduced in the previous section.

It is now possible to provide a more formal definition of a nuisance parameter in the context of decision analysis. Because there is a criterion variable to represent the direct concerns of the decision maker, a probabilistic node $j \in N$ is a nuisance parameter if it is not observed before any of the decisions in the diagram. This is true if it is not an informational predecessor for some decision node in N and, equivalently, due to the no-forgetting arcs, if it is not an informational predecessor for the latest decision node. In that case probabilistic node removal can be performed on node j to eliminate it from the diagram.

Suppose that all of the nuisance parameters have been reduced from the diagram, but that there are still nodes in N. By the discussion above, there must be at least one decision node (or all of the variables would be nuisance parameters), so let d be the latest decision node. Since there are no nuisance parameters, all of other nodes must be informational predecessors for d. These are the conditions under which the following result, proved in Kenley (1986) applies:

THEOREM 4. Optimal Policy for a Decision. *Given variable X_d is under the control of the decision maker, who will observe the realization of variables X_K, where $K = N \setminus \{d\} = I(d)$, prior to choosing X_d. Assume the decision maker seeks to maximize the value function $V(X_N) = (1/2)X_N^T\mathbf{Q}X_N + \mathbf{p}^TX_N + r$, where $\mathbf{Q}$ is symmetric and $\mathbf{Q}_{dd}$ is negative, so that an optimal choice exists and is unique. The preferred decision is*

$$X_{d^*}(X_K) = -(\mathbf{p}_d + \mathbf{Q}_{dK}X_K)/\mathbf{Q}_{dd}.$$

Thus, the optimal policy for X_d given X_k is linear with mean

$$\mu_d \leftarrow E[X_d] = -[\mathbf{p}_d + \mathbf{Q}_{dK}\mu_K]/\mathbf{Q}_{dd},$$

variance $v_d \leftarrow 0$, and linear coefficients

$$\mathbf{B}_{Kd} \leftarrow -\mathbf{Q}_{dK}/\mathbf{Q}_{dd}.$$

Using Theorem 4, the optimal policy for the decision variable X_d can now replace the reference value. (Since every other node had been an informational predecessor for node d, d cannot have any successors and there is no need to propagate the change in mean μ_d.) The decision variable X_d is turned into the optimal policy variable X_{d^*}, a probabilistic variable with conditional variance equal to zero. It is now a nuisance parameter and so are any of its former informational predecessors which are not observable at the time of an earlier decision.

This process can be repeated for every decision node, until there are no nodes left in the influence diagram, $N = \varnothing$. At that point the value function is simply

$$v(X_N) = E[X_0|X_N] = E[X_0] = r,$$

where r is the expected value of the criterion given that all of the decisions have been made optimally. Even though the decision nodes are eliminated from N in the course of the algorithm, their optimal policies should be saved as part of the solution.

This process is the influence diagram analog for "averaging out and folding back" in decision trees. It is based on a similar algorithm in Shachter (1986) and is summarized as follows:

1. While there is at least one decision node in the diagram

 1a. Let d be the latest decision node.

 1b. Perform a probabilistic node removal on each node in $N \backslash (d \cup I(d))$, in reverse order.

 1c. Replace the reference value for decision X_d with the optimal policy d^* using the formulae in Theorem 4, and making X_d a probabilistic variable with zero variance.

2. Perform a probabilistic node removal on each node in N, in reverse order.

As an example, consider the influence diagram for the consultant's problem shown in Figure 6. An ordered sequence for the nodes is (1, 2, 3, 4, 5, 6, 7, 8).

In the first iteration, since the *Time Share Price* decision is made after *Consulting Price*, $d \leftarrow 6$. $I(6) = \{1, 4\}$, so probabilistic node removal is performed on nodes 8, 7, 5, 3, and 2, resulting in the diagram in Figure 7a. (In order to remove node 2, the arc (2, 4) has to be reversed.) At this point the value function is given by,

$$\mathbf{Q} = \begin{bmatrix} -8.62 & 0.138 & 0.647 \\ 0.138 & 0.0 & -0.021 \\ 0.647 & -0.021 & -20.0 \end{bmatrix},$$

$\mathbf{p} = (1757, -1.04\ 979)$, and $r = -65008$. The optimal policy for *Time Share Price* is

$$X_6^*(X_1, X_4) = 49 + 0.032X_1 - 0.001X_4,$$

shown in the diagram in Figure 7b. (The value function does not change in this step.)

In the second iteration, *Consulting Price* is the only decision, $d \leftarrow 1$. Since $I(1) = \varnothing$, probabilistic node removal is performed on nodes 6 and 4 to obtain the diagram in Figure 7c. The value function is now $\mathbf{Q} = [-9.97]$, $\mathbf{p} = (2073)$, and $r = -45118$. The optimal policy for *Consulting Price* is $X_1^* = 208$, as shown in Figure 7d.

There are no more decisions, so probabilistic node removal is performed on node 1 and there are no more nodes left in N. The optimal expected profit is $r = 170348$.

These results can be summarized in the *policy diagram* in Figure 8. This is just the starting diagram from Figure 6 with the reference values for the decisions replaced by the optimal policies. (Remember to propagate the changes in the means!) The policy diagram is a convenient tool for performing sensitivity analysis, since any combination

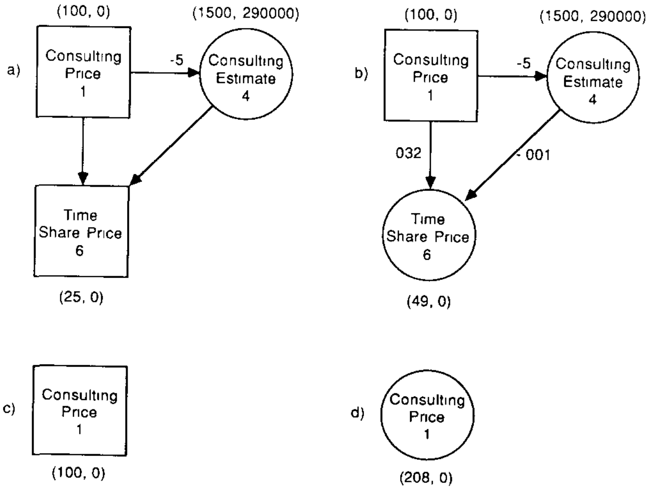

FIGURE 7. Choosing the Optimal Policies for the Consultant's Problem.

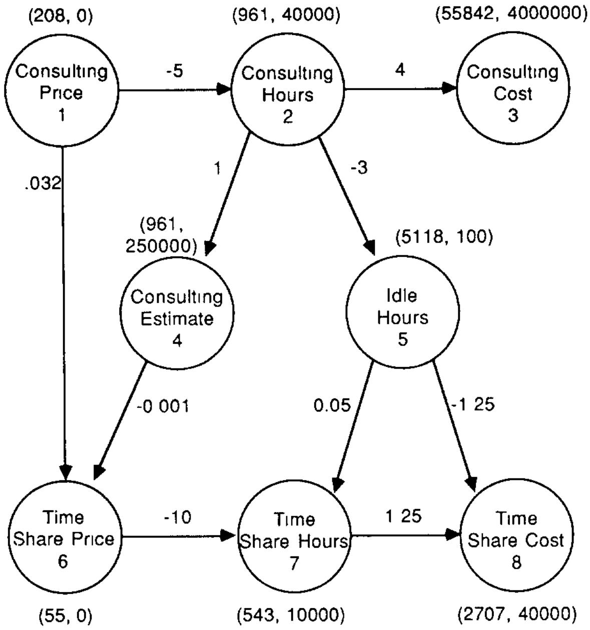

FIGURE 8. Policy Diagram for the Consultant's Problem.

of policies can be considered within such a diagram. The expected value of the criterion provides a standard for comparison for each set of policies. For example, the consultant's optimal prices are more than twice her reference values, yielding an increase in profit of over 60 percent over the reference case. If she fears that an exorbitant *Consulting Price* would hurt her consulting practice in the long run, she can still obtain the optimal *Time Share Price* given different values of *Consulting Price*. One last thing to note is that the optimal *Time Share Price* is not very sensitive to *Consulting Estimate*. She can calculate the value of that information by seeing how much the optimal expected value drops when she evaluates the diagram without the informational arc (4, 6). (This computes the "selling price" rather than the "buying price" for the information, but they are equal in this paper since the utility is either linear or exponential.) She would then know whether the accountant's services warrant his fee.

9. Quadratic Value Function with Exponential Utility

The previous results are generalized to allow an exponential utility over the quadratic value function. The decision maker is assumed to have a constant absolute risk aversion γ (Pratt 1964) when comparing distributions for the criterion variable.

Suppose that the criterion for making decisions is not the conditional expected value of the criterion $E[X_0 | X_N]$ but the conditional expected utility $E[u(X_0)| X_N]$ instead. Assume u is an exponential utility function, $u(t) = a + be^{-\gamma t}$, where $b\gamma < 0$. As before, $V(X_N) = (1/2)X_N^T\mathbf{Q}X_N + \mathbf{p}^TX_N + r$, but now define

$$U(X_N) = E[u(X_0)| X_N] = E[u(V(X_N))| X_N].$$

The following two theorems extend the inference and decision analysis results to this new value function. The proofs are in Kenley (1986).

THEOREM 5. Probabilistic Node Removal with Exponential Utility. *Given node $j \in N$ has no successors, let $K = N\setminus\{j\}$. If*

$$U(X_N) = E[u(X_0)| X_N] = a + b\exp[-\gamma((1/2)X_N^T\mathbf{Q}X_N + p^TX_N + r)],$$

where $\mathbf{Q}$ *is symmetric and* $b\gamma < 0$, *then*

$$\begin{aligned} U_{\text{new}}(X_K) &= E[u(X_0) \mid X_K] = E[U(X_N) \mid X_K] \\ &= a + b_{\text{new}} \exp[-\gamma((1/2)X_K^T \mathbf{Q}_{\text{new}} X_K + \mathbf{p}_{\text{new}}^T X_K + r_{\text{new}})]. \end{aligned}$$

If $v_j = 0$ *then* $\mathbf{Q}_{\text{new}}$, $\mathbf{p}_{\text{new}}$, *and* r_{new} *are updated as for the risk neutral case in Theorem* 3. *Otherwise* (*if* $v_j > 0$),

$$\mathbf{Q}_{\text{new}} \leftarrow \mathbf{Q}_{KK} + \mathbf{B}_{Kj}\mathbf{B}_{Kj}^T/(\gamma v_j) - [v_j/(\gamma + \gamma^2 v_j \mathbf{Q}_{jj})](\gamma \mathbf{Q}_{Kj} - \mathbf{B}_{Kj}/v_j)(\gamma \mathbf{Q}_{Kj} - \mathbf{B}_{Kj}/v_j)^T,$$

$$\mathbf{p}_{\text{new}} \leftarrow \mathbf{p}_K + (\mu_j - \mathbf{B}_{Kj}^T \mu_K)\mathbf{B}_{Kj}/(\gamma v_j) - [v_j/(\gamma + \gamma^2 v_j \mathbf{Q}_{jj})][\gamma \mathbf{p}_j - (\mu_j - \mathbf{B}_{Kj}^T \mu_K)/v_j] [\gamma \mathbf{Q}_{Kj} - \mathbf{B}_{Kj}/v_j],$$

$$r_{\text{new}} \leftarrow r + (\mu_j - \mathbf{B}_{Kj}^T \mu_K)^2/(2\gamma v_j) - (1/2)[v_j/(\gamma + \gamma^2 v_j \mathbf{Q}_{jj})][\gamma \mathbf{p}_j - (\mu_j - \mathbf{B}_{Kj}^T \mu_K)/v_j]^2,$$

and

$$b_{\text{new}} \leftarrow b/(1 + \gamma v_j/\mathbf{Q}_{jj})^{1/2}.$$

THEOREM 6. Optimal Policy for a Decision with Exponential Utility. *Given variable* X_d *is under the control of the decision maker, who will observe the realization of variables* X_K, *where* $K = N \setminus \{d\} = I(d)$, *prior to choosing* X_d. *Assume the decision maker seeks to maximize*

$$U(X_N) = a + b \exp[-\gamma((1/2)X_N^T \mathbf{Q} X_N + p^T X_N + r)],$$

where $\mathbf{Q}$ *is symmetric*, $\mathbf{Q}_{dd}$ *is negative, and* $b\gamma < 0$, *then the preferred decision has the same policy as was given for the risk neutral case in Theorem* 4.

10. Conclusions

This paper develops a Gaussian influence diagram to represent the linear-quadratic-Gaussian decision problem. It has all of the advantages of the general influence diagram for model structuring (Howard and Matheson 1981) and analysis (Shachter 1986, 1988). In addition, it fosters understanding of Gaussian processes by using the natural representation for correlation between random variables first suggested by Yule (1907). The matrix algebra employed in traditional representation and solution techniques has been replaced by simple scalar operations on the graph. Many technical problems with the traditional approach, such as the construction and maintenance of a positive semidefinite system, are handled easily in this framework.

This work extends influence diagram methodology into models with continuous random variables. It has a significant complexity advantage over discrete distribution influence diagrams, since all of the manipulations are polynomial in the number of variables. At the same time the influence diagram brings a powerful graphical representation of conditional independence to bear on multivariate normal problems. This simplifies assessment and analysis even when there is no sparsity in the (unconditional) covariance matrix, and provides simple, stable operations for real-time processing.

A generalization of the assumptions in this paper would allow conditional probability distributions such that

$$X_j = f_j(X_{C(j)}) + (v_j)^{1/2} Z_j \qquad \text{for} \qquad j = 1, \ldots, n,$$

in which $f_1, \ldots, f_n$ are arbitrary real-valued functions and $Z_1, \ldots, Z_n$ are independent random variables. In the multivariate normal, of course, the $\{f_j\}$ must be linear and the $\{Z_j\}$ must be normally distributed. If the $\{Z_j\}$ were not normal but the model were still linear, then all of the probabilistic inference results would still apply, except the conditional variances would (in general) no longer be independent of the conditioning variables' outcomes. If the functions were also nonlinear, the factorization of the covariance matrix could still be represented in the influence diagram. However, in that case the arc coefficients could not be used to compute conditional expected values.

In general, whenever the probabilistic model deviates from the multivariate normal the class of decision problems that can be solved in the Gaussian influence diagram is restricted: a probabilistic variable cannot depend on any decision variables, and it can only depend on probabilistic variables which will be observed at the same time as it. A general value function can be approximated by the second order Taylor series expansion about the mean values for probabilistic variables and the reference values for decisions. Essentially, decision analysis is limited to the pure "open-loop" or "closed-loop" cases of proximal analysis (Howard 1971, Kenley 1986).

There are many applications for the results in this paper. They provide an alternative representation and operations for Bayesian linear regression, business portfolio analysis, forecasting, causal modeling, path analysis, and discrete-time filtering (Kenley 1986). Even more important, they offer new explanations for known results in these areas and the potential for new insights. The tractability and polynomial-time nature of the manipulations suggest applications in knowledge acquisition and retrieval in expert systems (Shachter et al. 1987). The local nature of the computations (as opposed to traditional matrix operations) suggests significant opportunity to employ parallel processing techniques.[1]

[1] Most of this paper is based on research results presented in Kenley (1986), supported by Independent Research and Development funds from the Astronautics Division of Lockheed Missiles & Space Company, Inc. In particular, thanks go to Chuck MacQuiddy for approving Lockheed sponsorship of that research. We also wish to thank Jim Matheson, Ron Howard, Bob Nau, Dick Barlow, Dennis Lindley, Vic Hasselblad, the editors, and the referees for their helpful suggestions.

Appendix A. Proof of Arc Reversal Theorem

THEOREM 1. Arc Reversal. *Given that node $i \in C(j)$ and no other directed path exists from i to j,*

$$E[X_j|X_K] = \mu_j + \sum_{k\in K} (b_{kj} + b_{ki}b_{ij})(X_k - \mu_k), \qquad \text{and}$$

$$\mathrm{Var}\,[X_j|X_K] = v_j + b_{ij}^2 v_i,$$

where $K = (C(j) \cup C(i)) \setminus \{i\}$. If neither v_i nor v_j is zero, then

$$E[X_i|X_j, X_K] = \mu_i + \sum_{k\in K} (b_{ki} - (b_{kj} + b_{ki}b_{ij})b'_{ij})(X_k - \mu_k) + b'_{ij}(X_j - \mu_j),$$

$$\mathrm{Var}\,[X_i|X_j, X_K] = v_i v_j / [v_j + b_{ij}^2 v_i],$$

and, using primes to denote the coefficients revised during the arc reversal,

$$b'_{ji} = b_{ij}v_i / [v_j + b_{ij}^2 v_i].$$

PROOF. Since there is no other directed path from i to j and there are no directed cycles in the influence diagram, there is no directed path from either i or j to K. It follows from the conditional independence in the influence diagram that

$$\Pr\{X_i|X_K\} = \Pr\{X_i|X_{C(i)}\}, \qquad \text{and} \qquad \Pr\{X_j|X_i, X_K\} = \Pr\{X_j|X_{C(j)}\}.$$

Since $\mathbf{B}_{K\setminus C(i),i} = 0$ and $\mathbf{B}_{K\setminus C(j),j} = 0$,

$$\begin{aligned} E[X_j|X_K] &= E_{X_i}[E[X_j|X_i, X_K]|X_K] \\ &= E_{X_i}[\mu_j + \mathbf{B}_{Kj}^T(X_K - \mu_K) + b_{ij}(X_i - \mu_i)|X_K] \\ &= \mu_j + \mathbf{B}_{Kj}^T(X_K - \mu_K) + b_{ij}E_{X_i}[(X_i - \mu_i)|X_K] \end{aligned}$$

$$= \mu_j + (\mathbf{B}_{Kj}^T + \mathbf{B}_{Ki}^T b_{ij})(X_K - \mu_K) \qquad \text{and}$$

$$\begin{aligned}\operatorname{Var}[X_j|X_K] &= E_{X_i}[\operatorname{Var}[X_j|X_i, X_K]|X_K + \operatorname{Var}_{X_i}[E[X_j|X_i, X_K]|X_K] \\ &= E_{X_i}[v_j|X_K] + \operatorname{Var}_{X_i}[\mu_j + \mathbf{B}_{Kj}^T(X_K - \mu_K) + b_{ij}(X_i - \mu_i)|X_K] \\ &= v_j + \operatorname{Var}_{X_i}[b_{ij}X_i|X_{C(i)}] = v_j + b_{ij}^2 v_i.\end{aligned}$$

Because there is no other directed path from i to j, their conditional covariance can be expressed as

$$\operatorname{Cov}[X_i, X_j|X_K] = b_{ij}\operatorname{Var}[X_i|X_K] = b'_{ji}\operatorname{Var}[X_j|X_K].$$

Therefore,

$$b'_{ji} = b_{ij}\operatorname{Var}[X_i|X_K](\operatorname{Var}[X_j|X_K])^{-1} = b_{ij}v_i/[v_j + b_{ij}^2 v_i].$$

For Gaussian random variables, the conditional variances satisfy

$$\begin{aligned}\operatorname{Var}[X_i|X_j, X_K] &= \operatorname{Var}[X_i|X_K] - (\operatorname{Cov}[X_i, X_j|X_K])^2(\operatorname{Var}[X_j|X_K])^{-1} \\ &= v_i - (b_{ij}\operatorname{Var}[X_i|X_K])^2(\operatorname{Var}[X_j|X_K])^{-1} \\ &= v_i - (b_{ij}v_i)^2/[v_j + b_{ij}^2 v_i] \\ &= v_i v_j/[v_j + b_{ij}^2 v_i].\end{aligned}$$

For some $k \in K$ and $L = K \setminus \{k\}$,

$$\begin{aligned}E[X_i|X_k, X_L = \mu_L] &= E_{X_j}[E[X_i|X_j, X_k, X_L = \mu_L]|X_k, X_L = \mu_L] \\ &= E_{X_j}[\mu_i + b'_{ki}(X_k - \mu_k) + b'_{ji}(X_j - \mu_j)|X_k, X_L = \mu_L] \\ &= \mu_i + b'_{ki}(X_k - \mu_k) + b'_{ji}b'_{kj}(X_k - \mu_k) \\ &= \mu_i + (b'_{ki} + b'_{ji}b'_{kj})(X_k - \mu_k).\end{aligned}$$

Also,

$$E[X_i|X_k, X_L = \mu_L] = \mu_i + b_{ki}(X_k - \mu_k).$$

Hence,

$$b'_{ki} = b_{ki} - b'_{kj}b'_{ji} = b_{ki} - (b_{kj} + b_{ki}b_{ij})b'_{ji}. \qquad \#$$

Appendix B. Covariance Representation Proofs

This appendix formalizes the relationship among the influence diagram, covariance matrix, and some alternative representations for the multivariate normal distribution. It presents some of the results not derived in the body of the paper and tries to deepen a reader's understanding of the Gaussian influence diagram. None of the results in this section are really new, and some can be traced back as far as Yule (1907).

Most of the vectors and matrices in this section will be indexed by the full set of nodes, N. To improve readability, the subscripts will be omitted, whenever it is unambiguous, so X will represent X_N and $\mathbf{B}$ will denote $\mathbf{B}_{NN}$. Similarly, we will assume, without loss of generality, that the sequence of indices $(1, \ldots, n)$ is ordered, so that the matrix $\mathbf{B}$ is strictly upper triangular.

Let $\mathbf{D}$ and $\mathbf{S}$ be the diagonal matrices formed from the conditional variances and standard deviations respectively, that is, $\mathbf{D} = \operatorname{diag}(v) = \mathbf{S}^T\mathbf{S}$. Equation (1) which defined the random variables X in terms of standard normal random variables Z can now be written as

$$(X - \mu) = \mathbf{B}^T(X - \mu) + \mathbf{S}^T Z = \mathbf{B}^T(X - \mu) + E,$$

where random variables E, often called the *innovations* or *residuals*, are the differences between the values of X actually realized and those predicted by the linear model,

$$E = \mathbf{S}^T Z = (\mathbf{I} - \mathbf{B}^T)(X - \mu).$$

The relationships among the variables X, E, and Z are displayed in the Gaussian influence diagram in Figure 9. In fact, all of the expressions derived in this section can be computed by reversing arcs in this diagram.

Because $\mathbf{B}$ is strictly upper triangular, $(\mathbf{I} - \mathbf{B}^T)$ is invertible and X can be expressed in terms of E and Z as

$$X - \mu = (\mathbf{I} - \mathbf{B}^T)^{-1}E = \mathbf{U}^T E = \mathbf{U}^T\mathbf{S}^T Z = \mathbf{A}^T Z,$$

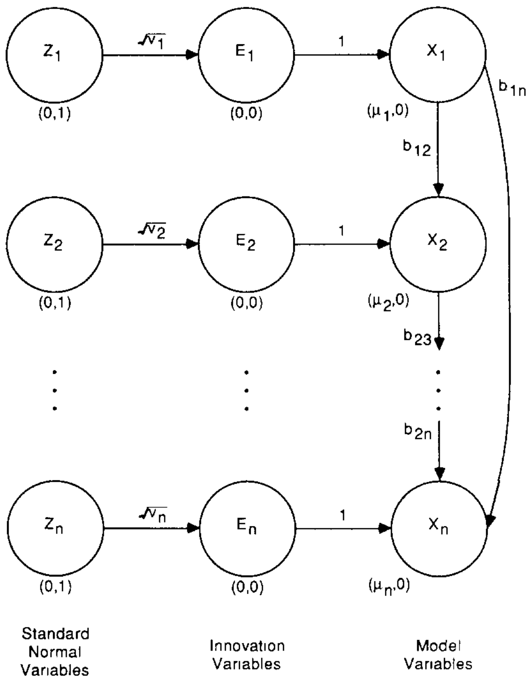

FIGURE 9. General Gaussian Influence Diagram with Innovations and Standard Normal Sources of Variance.

where $\mathbf{U} = (\mathbf{I} - \mathbf{B})^{-1}$ and $\mathbf{A} = \mathbf{SU}$. $\mathbf{U}$ summarizes the propagation of changes to the means, and was introduced for that purpose in §2. $\mathbf{A}$ is called the *Cholesky decomposition*. Although $\mathbf{U}$ and $\mathbf{A}$ are both upper triangular, $\mathbf{A}$ can be singular while $\mathbf{U}$ is always nonsingular. In fact, $\mathbf{A}$ is a symmetric square root of the covariance matrix Σ while $\mathbf{U}$ and the diagonal matrix of variances $\mathbf{D}$ form a convenient factorization of the covariance matrix:

$$\begin{aligned}\Sigma &= \text{Var}\,\{X\} = \text{Var}\,\{X - \mu\}\\ &= \text{Var}\,\{\mathbf{A}^T Z\} = \mathbf{A}^T\ \text{Var}\,\{Z\}\mathbf{A} = \mathbf{A}^T\mathbf{A}\\ &= \mathbf{U}^T\mathbf{S}^T\mathbf{S}\mathbf{U} = \mathbf{U}^T\mathbf{D}\mathbf{U}.\end{aligned}$$

If the covariance matrix is nonsingular, then its inverse, the *precision* matrix, is

$$\Sigma^{-1} = \mathbf{U}^{-T}\mathbf{D}^{-1}\mathbf{U}^{-1} = (\mathbf{I} - \mathbf{B})^T\mathbf{D}^{-1}(\mathbf{I} - \mathbf{B}).$$

Otherwise, the precision is a generalized inverse of the covariance, and it can be defined in terms of a generalized inverse of $\mathbf{D}$, written as $\mathbf{D}^-$, a diagonal matrix whose entries are v_j^{-1} for nonzero v_j and zero otherwise:

$$\Sigma^- = \mathbf{U}^{-T}\mathbf{D}^-\mathbf{U}^{-1} = (\mathbf{I} - \mathbf{B})^T\mathbf{D}^-(\mathbf{I} - \mathbf{B}).$$

The factorization of the covariance, $\Sigma = \mathbf{U}^T\mathbf{D}\mathbf{U}$, leads immediately to the following theorem, which provides a simple test within the influence diagram to confirm whether a symmetric Σ is positive (semi-)definite. By contrast, although the Cholesky decomposition $\mathbf{A}$ has the same rank as Σ it is well-defined only when Σ is PSD.

THEOREM 2. Rank of Covariance Matrix. *Σ is positive (semi-)definite if and only if $v > (\geq)\, 0$. Furthermore, the rank of Σ is equal to the number of nonzero elements in v.*

PROOF. Because $\mathbf{U}$ is nonsingular, $\Sigma = \mathbf{U}^T\mathbf{D}\mathbf{U}$ is a congruence transformation (Strang 1980, p. 259). Therefore, Σ has the same number of positive (negative, zero) eigenvalues as $\mathbf{D} = \text{diag}\,(v)$. #

The remainder of this section presents transformations between the covariance matrix Σ and the influence diagram coefficients $\mathbf{B}$ and v (or $\mathbf{D}$). We have already seen that everything else can be expressed in terms of $\mathbf{B}$ and $\mathbf{D}$, so the more difficult transformation is from the covariance to the influence diagram. This could be accomplished using standard techniques by first computing the Cholesky decomposition $\mathbf{A}$ to obtain $\mathbf{U}$ and $\mathbf{D}$, and then inverting $\mathbf{U}$ to compute $\mathbf{B} = \mathbf{I} - \mathbf{U}^{-1}$. Instead we derive an alternative (but related) procedure which iteratively transforms from one representation to the other. This is the procedure presented in §5.

For the remainder of the section, corresponding to an arbitrary $j \in N$, let $\mathbf{s} = (1, \ldots, j)$ and $\mathbf{t} = (j + 1, \ldots, n)$.

THEOREM 7. Stepwise Decomposition of Covariance.

$$\Sigma = \begin{bmatrix} \mathbf{0}_j \\ \mathbf{U}_{tt}^T \end{bmatrix} \mathbf{D}_{tt} [\mathbf{0}_j \ \mathbf{U}_{tt}] + \begin{bmatrix} \mathbf{I}_j \\ \mathbf{U}_{st}^T \mathbf{U}_{ss}^{-T} \end{bmatrix} \Sigma_{ss} [\mathbf{I}_j \ \mathbf{U}_{ss}^{-1} \mathbf{U}_{st}]$$

$$= \mathbf{W}_j^T \begin{bmatrix} \Sigma_{ss} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{tt} \end{bmatrix} \mathbf{W}_j,$$

where $\mathbf{I}_j$ *is the* $j \times j$ *identity matrix and*

$$\mathbf{W}_j = \begin{bmatrix} \mathbf{I}_j & \mathbf{U}_{ss}^{-1} \mathbf{U}_{st} \\ \mathbf{0} & \mathbf{U}_{tt} \end{bmatrix}.$$

PROOF. In order to condition on X_s, first express X_t in terms of X_s,

$$X_t = \mu_t + U_{st}^T E_s + U_{tt}^T E_t = \mu_t + U_{st}^T U_{ss}^{-T} (X_s - \mu_s) + U_{tt}^T E_t.$$

Now, obtain the covariance by conditioning,

$$\Sigma = \text{Var}\{X\} = E[\text{Var}[X|X_s]] + \text{Var}[E[X|X_s]]$$

$$= \text{Var} \begin{bmatrix} \mathbf{0}_j \\ \mathbf{U}_{tt}^T E_t \end{bmatrix} + \text{Var} \begin{bmatrix} X_s \\ \mathbf{U}_{st}^T \mathbf{U}_{ss}^{-T} X_s \end{bmatrix}. \qquad \#$$

This result is most useful when $\mathbf{t}$ contains only one element. In that case, $\mathbf{B}_{st} = \mathbf{U}_{ss}^{-1} \mathbf{U}_{st}$ and $v_t = \mathbf{U}_{tt}^T \mathbf{D}_{tt} \mathbf{U}_{tt}$. The proposition then states that

$$\Sigma_{st} = \Sigma_{ts}^T = \Sigma_{ss} \mathbf{B}_{st} \qquad \text{and} \qquad \Sigma_{tt} = v_t + \mathbf{B}_{st}^T \Sigma_{ss} \mathbf{B}_{st}.$$

If Σ_{ss} is also nonsingular, then invert these formulae to obtain

$$\mathbf{B}_{st} = \Sigma_{ss}^{-1} \Sigma_{st} \qquad \text{and} \qquad v_t = \Sigma_{tt} - \Sigma_{ts} \mathbf{B}_{st}.$$

Finally, let

$$\mathbf{U}_j = \begin{bmatrix} \mathbf{I}_{j-1} & \mathbf{B}_{sj} & \mathbf{0} \\ 0 & 1 & 0 \\ \mathbf{0} & 0 & \mathbf{I}_{n-j} \end{bmatrix}.$$

It can be verified that

$$\mathbf{W}_{j-1} = \mathbf{U}_j \mathbf{W}_j \qquad \text{for} \qquad j = 1, \ldots, n,$$

so one can compute $\mathbf{U}$ iteratively in terms of $\mathbf{B}$ as

$$\mathbf{U} = \mathbf{W}_0 = \mathbf{U}_1 \cdots \mathbf{U}_n \mathbf{W}_n = \mathbf{U}_1 \cdots \mathbf{U}_n.$$

This is precisely the product form of the inverse $\mathbf{U} = (\mathbf{I} - \mathbf{B})^{-1}$.

References

BIERMAN, G. J., *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977.

BREIMAN, L., "Lectures on Data Analysis," Univ. of California, Berkeley, 1980.

BRYSON, A. E. AND Y. HO, *Applied Optimal Control: Optimization, Estimation, and Control*, Revised Printing, Hemisphere Publishing, Washington, D.C., 1975.

GRAYBILL, F. A., *Theory and Application of the Linear Model*, Duxbury Press, North Scituate, MA, 1976.

HOWARD, R. A., "Proximal Decision Analysis," *Management Sci.*, 17, 9 (May 1971), 507–541; reprinted in Howard, R. A. and J. E. Matheson (Eds.), *The Principles and Applications of Decision Analysis*, Vol. II, Strategic Decisions Group, Menlo Park, CA, 1984.

——— AND J. E. MATHESON, "Influence Diagrams," 1981, in Howard, R. A. and J. E. Matheson (Eds.), *The Principles and Applications of Decision Analysis*, Vol. II, Strategic Decisions Group, Menlo Park, CA, 1984.

KENLEY, C. R., "Influence Diagram Models with Continuous Variables," Ph.D. Dissertation, EES Department, Stanford Unviersity, June 1986.

OLMSTED, S. M., "On Representing and Solving Decision Problems," Ph.D. Dissertation, EES Department, Stanford University, 1983.

OREN, S. S., "A Conjugate Directions Method for Subjective Assessment of Normal Distributions," *J. Optim. Theory Appl.*, 33, 1 (January 1981), 25–36.

PEARL, J., "Distributed Diagnosis in Causal Models With Continuous Variables," Technical Report R-54, CSD-860051, Computer Science Department, UCLA, December 1985.

PRATT, J., "Risk Aversion in the Small and in the Large," *Econometrica*, 32 (1964), 123–136.

SHACHTER, R., "Evaluating Influence Diagrams," *Oper. Res.*, 34 (November-December 1986), 871–882

———, "Probabilistic Inference and Influence Diagrams," *Oper. Res.*, 36, 4 (July/August 1988), 589–604.

———, D. EDDY, V. HASSELBLAD AND R. WOLPERT, "A Heuristic Bayesian Approach to Knowledge Acquisition: Application to Analysis of Tissue-Type Plasminogen Activator," in *Proc. Third Workshop on Uncertainty in Artificial Intelligence*, Seattle, WA, July 10–12, 1987.

———, AND D. HECKERMAN, "Thinking Backwards for Knowledge Acquisition," *AI Magazine*, 8 (Fall 1987), 55–61.

STRANG, G., *Linear Algebra and Its Applications*, Second Edition, Academic Press, New York, 1980.

WRIGHT, S., "The Method of Path Coefficients," *Ann. Math. Statist.*, 5, 3 (September 1934), 161–215.

YULE, G. U., "On the Theory of Correlation for Any Number of Variables, Treated by a New System of Notation," *Proc. Roy. Soc. London, Ser. A*, 79 (September 1907), 182–193.