This article was downloaded by: [Georgia Tech Library]
On: 24 August 2014, At: 08:25
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

IIE Transactions

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uiie20

Objective-oriented optimal sensor allocation strategy for process monitoring and diagnosis by multivariate analysis in a Bayesian network

Kaibo Liu [a] & Jianjun Shi [a]

[a] H. Milton Stewart School of Industrial and Systems Engineering , Georgia Institute of Technology , Atlanta , GA , 30332 , USA

Accepted author version posted online: 17 Sep 2012.Published online: 07 Mar 2013.

To cite this article: Kaibo Liu & Jianjun Shi (2013) Objective-oriented optimal sensor allocation strategy for process monitoring and diagnosis by multivariate analysis in a Bayesian network, IIE Transactions, 45:6, 630-643, DOI: 10.1080/0740817X.2012.725505

To link to this article: http://dx.doi.org/10.1080/0740817X.2012.725505

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

IIE Transactions (2013) **45**, 630–643
Copyright © "IIE"
ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/0740817X.2012.725505

Objective-oriented optimal sensor allocation strategy for process monitoring and diagnosis by multivariate analysis in a Bayesian network

KAIBO LIU and JIANJUN SHI*

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
E-mail: jianjun.shi@isye.gatech.edu

Received November 2011 and accepted August 2012

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

Measurement strategy and sensor allocation have a direct impact on the product quality, productivity, and cost. This article studies the couplings or interactions between the optimal design of a sensor system and quality management in a manufacturing system, which can improve cost-effectiveness and production yield by considering sensor cost, process change detection speed, and fault diagnosis accuracy. Based on an established definition of sensor allocation in a Bayesian network, an algorithm named "Best Allocation Subsets by Intelligent Search" (BASIS) is developed in this article to obtain the optimal sensor allocation design at minimum cost under different specified Average Run Length (ARL) requirements. Unlike previous approaches reported in the literature, the BASIS algorithm is developed based on investigating a multivariate T^2 control chart when only partial observations are available. After implementing the derived optimal sensor solution, a diagnosis ranking method is proposed to find the root cause variables by ranking all of the identified potential faults. Two case studies are conducted on a hot forming process and a cap alignment process to illustrate and evaluate the developed methods.

Keywords: Bayesian network, partial observation, multiple mean shifts, optimal sensor allocation, diagnosis ranking

1. Introduction

Statistical advances in sensor technology have enabled the development of Distributed Sensor Networks (DSNs), which have found wide use in industrial and civilian applications such as industrial process monitoring and control, healthcare surveillance, and decision support systems.

One of the fundamental issues in a DSN is process control with the objectives of reducing variability and improving the final quality of products (Montgomery, 1997). A typical process control problem includes three steps: process monitoring, fault diagnosis, and process recovery. In process monitoring, various methods have been developed, such as univariate statistical process control (Shewhart, 1931), multivariate Hotelling T^2 statistical process control (Hotelling, 1947), MCUSUM and MEWMA control charts (Woodall and Ncube, 1985; Lowry *et al.*, 1992), and some principal component analysis–based techniques (Bakshi, 1998). Regarding the fault diagnosis procedure, one popular technique developed is the Mason–Tracy–Young (MTY) T^2 decomposition approach (Mason *et al.*, 1997). While the MTY approach is theoretically sound and appealing, it involves a significant number of computational issues for high-dimensional variables and has inherent deficiencies in identifying the root cause variables in a system. Therefore, a causation-based T^2 decomposition method was proposed by Li, Jin, and Shi (2008) to improve the diagnosability in a causal model. On the other hand, the diagnosis problem can be viewed as a task to correctly classify a fault into one of the pre-defined classes. New techniques have been developed based on the classification point of view, such as Fishers discriminant analysis (Duda *et al.*, 2000) and support vector machine (Vapnik, 1999).

Another critical concern about a DSN is the cost, which includes installation, maintenance, and operational costs (Edan and Nof, 2000). Even though a fully deployed sensor system can minimize information loss, the total cost associated with these sensors can be overwhelming. Moreover, the massive amount of data collected from different sensors can create dependent or redundant information. The efficiency of the sensor network can be benchmarked in terms of the sensing cost to achieve customer demands (Ding *et al.*, 2003). Therefore, it is desirable to quantify the dependencies via a statistical model in order to effectively and efficiently collect sufficient information at the minimum cost.

*Corresponding author

0740-817X © 2013 "IIE"

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

In the past decades, extensive efforts have been made to improve manufacturing system design and product quality. In most cases, manufacturing system design aims at increasing production throughput, reducing lead time, and improving customer demand satisfaction. On the other hand, product quality improvement focuses on quick abnormality detection and accurate fault diagnosis (Li *et al.*, 2007). Some effort has been expended on investigating the couplings or interactions between production system design and product quality (Li, Blumenfeld, and Marin, 2008; Li and Meerkov, 2009). In general, different system designs may result in different quality performance levels. In this article, we focus on developing an optimal sensor allocation strategy for detecting and diagnosing process variation sources in a timely manner, so that the customer demand can be satisfied economically and efficiently. The effective use of sensor data has provided unprecedented opportunities for quality and productivity improvement (Ding *et al.*, 2006). With a guaranteed level of customer satisfaction, implementing the minimum number of sensors in the system can significantly reduce the cost and time associated with sensor operation and maintenance, so that lead time and inventory level can be reduced. It has been shown that an optimal sensor system enables manufacturers to improve product quality and reduce production downtime as well as inventory level (Liu *et al.*, 2005).

The optimal sensor allocation problem is a decision-making process that involves the determination of (i) the minimum cost of the sensors required to achieve customer demand and (ii) which variables to measure. Most methodologies developed for this type of problem are based on the assumption that certain prior knowledge is sufficient to quantitatively describe the relationship between physical parameters (Khan *et al.*, 1998; Khan and Ceglarek, 2000). However, there is a lack of general formulation when such prior knowledge is unknown. Li and Jin (2010) proposed an optimal sensor allocation method by integrating causal models and set-covering algorithms for single mean shift detection using a univariate control chart, which addressed the problem of sensor allocation for process monitoring in causal models. Although the problem is well defined in their paper, the developed method does not take the Bonferroni correction into consideration when searching for the optimal solution. Therefore, the proposed allocation strategy in their paper cannot guarantee to satisfy the specified Average Run Length (ARL) requirement in certain detection cases. This point will be further elaborated in the case study section of this article. In addition, the single mean shift assumption in their paper leads to limited applications in practice, because there could be cases with multiple mean shifts having different signs and magnitudes in a complex DSN. Although it is possible to build a univariate control chart for each variable, the false alarm rate of that system will increase and the charts will be unable to describe the relationships between variables. Thus, a multivariate statistical control chart is preferred and adopted in this article.

The objectives of this article are to study the optimal sensor allocation strategy that can meet the customer requirement in terms of ARL with minimum sensing cost and then conduct a generic root cause diagnostic analysis with partial information after implementing the optimal strategy. This research focuses on mean shifts detection and diagnosis with an assumption that the variance of each variable is constant.

To conduct the tasks mentioned above, a Bayesian Network (BN) is adopted in this article to represent the causal relationships among a set of variables in a DSN. The application of a BN has recently been successfully demonstrated in fault detection and diagnosis (Verron *et al.*, 2008). The advantages of choosing a BN as a framework can be summarized as follows:

1. A BN can handle incomplete datasets since the dependency information is embedded visually and numerically.
2. A BN provides an effective tool to describe a causal relationship that is mathematically defined in terms of probabilistic independence statements. The causal interpretations facilitate the application of a BN to solve real-world problems (Pearl and Verma, 1991; Li and Shi, 2007).
3. A BN in conjunction with Bayesian statistical techniques facilitates the combination of domain knowledge and data (Heckerman, 1995).

Generally speaking, a BN can be obtained by integrating engineering knowledge and observational data (Buntine, 1994; Koller and Friedman, 2009). Li and Shi (2007) presented an example of using a BN to describe the causal relationships in a multistage rolling manufacturing process. In this article, we assume that the BN has already been correctly acquired, so that no more learning procedures are involved in the following discussions.

The rest of this article is organized as follows: Section 2 introduces the key terminologies of a BN. Section 3 presents a concrete formulation for the optimal sensor allocation problem and an intelligent searching algorithm for the optimal solution. Section 4 proposes a diagnosis ranking method to find the root cause variables, followed by diagnosability analysis, which includes discussions about diagnosability criteria, minimum diagnosable class, and diagnosable-ensured optimal sensor allocation strategy. Section 5 performs case studies based on a hot forming process and a cap alignment process to illustrate and evaluate the proposed methodologies under different mean shift detection and diagnosis scenarios. Finally, conclusions are drawn in Section 6.

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

2. BN representation and terminology

A BN is a probabilistic graphic model that can be used to represent causal relationships in a system of m physical variables $\mathbf{X} = \{X_1, X_2, \ldots, X_m\}$. A BN has two components, one qualitative and the other quantitative. The qualitative component is referred to as the *structure* of a BN, which is a directed acyclic graph. This structure conveys two pieces of information. First, each node corresponds to one physical variable of the system. Second, each directed arc connects two variables and represents their probabilistic dependence. In this article, we use an Arabic number to represent each node and use X_i to represent the physical variable or the numerical value of node i. The quantitative component, also called the *parameter* of a BN, is a conditional probability annotated on each directed arc of the structure. If there is a directed arc from node i to j, i is the direct cause (called the *parent*) of j and j is the direct effect (also called the *child*) of i, in which *direct* means that the causal influence from i to j is not mediated through any other nodes. A path exists from node i to j if there is one or several directed arcs linking together from i to j; i.e., $i \to \cdots \to j$. In this case, i is an *ancestor* of j and j is a *descendant* of i, respectively. If i does not have any ancestors (descendants), then it is called a *root* (*leaf*) of the system. In this article, the sets of *parents*, *children*, *descendants*, and *ancestors* of variable X_i or node i are denoted as $\mathbf{PA}(i)$, $\mathbf{CH}(i)$, $\mathbf{DS}(i)$, and $\mathbf{AN}(i)$, respectively. The sets of *root* and *leaf* nodes of the system are denoted by **RO** *and* **LF**, respectively. In addition, the kth component of each set is denoted as $PA_k(i)$, $CH_k(i)$, $DS_k(i)$, $AN_k(i)$, RO_k, and LF_k.

For example, Fig. 1 shows a linear Gaussian BN to represent the causal relationships involved in a hot forming process (Li and Jin, 2010). It has four process variables (X_4: temperature; X_3: material flow stress; X_2: tension in workpiece; X_5: Blank Holding Force (BHF)) and one quality variable (X_1: final dimension of workpiece). A two-dimensional (2D) physical illustration of the hot forming process is shown in Fig. 2. According to the definition, nodes 5 and 4 are parents of node 2, and they are the roots of the system; nodes 2 and 3 are the children of node 4; {2, 3, 4, 5} are ancestors of node 1; and {1, 2, 3} are descendants of node 4. There are two distinct paths from node 4 to 1, 4→2→1 and 4→3→1.

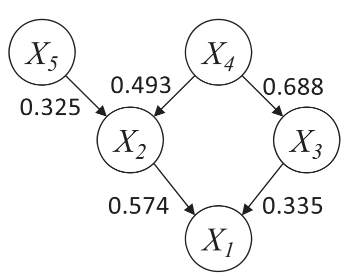

Fig. 1. BN structure of a hot forming process (Li and Jin, 2010).

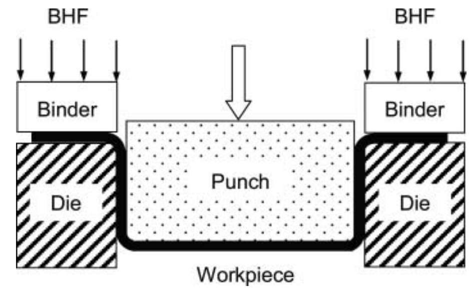

Fig. 2. A 2D illustration of the hot forming process (Li and Jin, 2010).

In this article, it is assumed that when the system is under normal operational condition, all variables will follow standard normal distributions, which can be achieved by standardizing all variables. Then, a linear Gaussian parameterization given the BN is represented as (Lauritzen and Wermuth, 1989; Li and Jin, 2010):

$$X_j = \sum_{k=1}^{card(\mathbf{PA}(j))} p(PA_k(j), j) X_{PA_k(j)} + V_j, \tag{1}$$

where $p(PA_k(j), j) \in (0, 1)$ is called the *path coefficient*, which resembles the conditional probability distribution $p(X_j | X_{PA_k(j)})$ to represent the influence from the kth parent of j to j; $card(\mathbf{PA}(j))$ is the cardinality of the set $\mathbf{PA}(j)$, the number of parents of j. In addition, we assume that $V_j \sim N(0, \sigma_j^2)$ is independent of $X_{PA_k(j)}$ and V_i $(i \neq j)$, which represents the randomness of variable X_j itself, or the random noise that cannot be described by the linear part of the model. It is used to maintain the unit variance of each variable X_i. Moreover, the *path effect* from i to j is defined to be the product of all path coefficients on that path connecting i to j. Since there may be several paths between i and j, $\tilde{\gamma}(i, j)$ is defined to be the *total effect* of i on j, which equals to the sum of all the path effects from i to j, so that:

$$\tilde{\gamma}(i, j) = \begin{cases} \in (0, 1) & \text{if } i \in \mathbf{AN}(j) \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

For example in Fig. 1, $\tilde{\gamma}(4, 1) = 0.493 \times 0.574 + 0.688 \times 0.335 = 0.5135$ and $\tilde{\gamma}(1, 4) = 0$. Since each variable follows standard normal distributions under regular operational condition, X_5 is equal to V_5 (i.e., $V_5 \sim N(0, 1)$), which is not influenced by any other variables, while X_2 is equal to $0.325X_5 + 0.493X_4 + V_2$ (i.e., $V_2 \sim N(0, 0.651)$), which is influenced by both variables X_5 and X_4.

To obtain the variance of V_j, we can recursively represent the term $X_{PA_k(j)}$ by Equation (1) until parameterization on the root nodes of the system. It can be readily shown that Equation (1) can be transformed into:

$$X_j = \sum_{k=1}^{card(\mathbf{AN}(j))} \tilde{\gamma}(AN_k(j), j) V_{AN_k(j)} + V_j. \tag{3}$$

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

Since X_j follows a standard normal distribution under normal operational condition and the V_j is independent of each other, the variance σ_j^2 can be obtained by

$$\sigma_j^2 = 1 - \sum_{k=1}^{card(\mathbf{AN}(j))} [\tilde{\gamma}(AN_k(j), j)]^2 \sigma^2_{AN_k(j)}. \tag{4}$$

At each step, we acquire the variance σ_j^2 if all $\sigma^2_{AN_k(j)}(k = 1, \ldots, card(\mathbf{AN}(j)))$ are obtained. Then in the end, all variances, $\sigma_j^2 (j \in m)$ can be known sequentially.

3. Optimal sensor allocation for quick mean shifts detection

3.1. *Problem formulation*

The optimal sensor allocation strategy is a mission-specific task under given objectives. In this article, we investigate the optimal allocation strategy under the requirement that at most p variables in the system may possibly have a mean shift with at least magnitude τ.

Since the graphical representation of a BN has already embedded the numerical relationship among variables, it is possible to monitor the propagation of multiple mean shifts by investigating a multivariate distribution. Although most linear Gaussian BNs are represented by Equation (1) to show the causal influences from parents (Koller and Friedman, 2009; Li and Jin, 2010), we find that a matrix representation is easier and more compact to illustrate mean shifts propagation within a BN. Specifically, $\mathbf{X}$ can be represented by the following forms:

$$\mathbf{X} = [\mathbf{PC}]\mathbf{X} + \mathbf{V} + \Delta, \tag{5}$$

where $\mathbf{X} \sim \mathbf{N}_{\mathrm{m}}(\mathbf{u}, \mathbf{\Sigma_X})$ is the measurement value of each physical variable with mean $\mathbf{u}$ and the diagonal elements of $\mathbf{\Sigma_X}$ are all ones; $[\mathbf{PC}]$ is the path coefficient matrix of the BN; $\mathbf{V} \sim \mathbf{N}_{\mathrm{m}}(\mathbf{0}, \mathbf{\Sigma_V})$ represents the randomness of $\mathbf{X}$ needed to maintain the unit variance of each variable; and Δ indicates the expected mean shift of each variable. For compactness, we define $\mathbf{A} = \mathbf{I}_{\mathrm{m}} - \mathbf{PC}$. According to the definition of $\mathbf{PC}$, $\mathbf{A}$ is a triangular matrix whose diagonal elements are all ones. Thus, $\mathbf{A}$ is an invertible matrix. The measurement value of a physical variable is available if and only if a sensor is deployed on it, so that only partial elements of $\mathbf{X}$ can be obtained during online measurement. Denote the sensor allocation set by $\mathbf{s}$, then:

$$(\mathbf{X} - \mathbf{A}^{-1}\Delta)_{\mathbf{s}} = (\mathbf{A}^{-1}\mathbf{V})_{\mathbf{s}}, \tag{6}$$

where $(\mathbf{X} - \mathbf{A}^{-1}\Delta)_{\mathbf{s}}$ represents the sub-vector of $\mathbf{X} - \mathbf{A}^{-1}\Delta$ with rows indexed by $\mathbf{s}$. From the definition of $\mathbf{V}$, its covariance matrix $\mathbf{\Sigma_V}$ is a diagonal matrix, where the diagonal elements can be obtained by Equation (4). $(\mathbf{X} - \mathbf{A}^{-1}\Delta)_{\mathbf{s}} \sim \mathbf{N}_{card(\mathbf{s})}(\mathbf{0}, [\mathbf{A}^{-1}\mathbf{\Sigma_V}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}})$, where $[\mathbf{A}^{-1}\mathbf{\Sigma_V}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}$ is a sub-block matrix of $\mathbf{A}^{-1}\mathbf{\Sigma_V}(\mathbf{A}^{-1})^{\mathrm{T}}$ with rows and columns indexed by $\mathbf{s}$.

Considering the null hypothesis $H_0 : \Delta = \mathbf{0}$ against the alternative hypothesis $H_a : \Delta \neq \mathbf{0}$, the chi-square control chart based on individual observation rejects H_0 if

$$\mathbf{X_s}^{\mathrm{T}}\{[\mathbf{A}^{-1}\mathbf{\Sigma_V}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}\mathbf{X_s} > \mathcal{X}^2_{card(\mathbf{s}),\alpha}, \tag{7}$$

where $\mathcal{X}^2_{card(\mathbf{s}),\alpha}$ is the upper control limit, the $(1-\alpha)$th quantile of the chi-square distribution with degree of freedom $card(\mathbf{s})$. When the system becomes abnormal and has a mean shift, the Type 2 error β of the test is

$$\beta = \Pr\{\mathbf{X_s}^{\mathrm{T}}\{[\mathbf{A}^{-1}\mathbf{\Sigma_V}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}\mathbf{X_s} \leq \mathcal{X}^2_{card(\mathbf{s}),\alpha} | \Delta \neq 0\}. \tag{8}$$

Thus, $ARL_{\mathbf{s}}(\Delta) = 1/(1-\beta)$ denotes the ARL needed to detect the expected mean shift Δ corresponding to the sensor allocation set $\mathbf{s}$. Denote $ARL_U(\tau)$ as the upper bound of the ARL specified by a practitioner to detect the expected mean shift Δ with at least magnitude τ in its non-zero elements. Accordingly, we give a definition of *detectable* as follows.

Definition 1. A mean shift Δ is *detectable* as long as $ARL_{\mathbf{s}}(\Delta) \leq ARL_U(\tau)$.

According to this definition, the optimal sensor allocation problem can be formulated as a searching problem for the best set of nodes, $\mathbf{s}$, that satisfies:

$$\min_{\mathbf{s}} C = \sum_{i \in \mathbf{s}} C_i, \quad \text{subject to} \quad ARL_{\mathbf{s}}(\Delta) \leq ARL_U(\tau), \tag{9}$$

where C_i is the cost of the sensor on node i.

3.2. *Best allocation subsets by intelligent search algorithm*

For a mean shift $\Delta \neq \mathbf{0}$, $\mathbf{X_s}^{\mathrm{T}}\{[\mathbf{A}^{-1}\mathbf{\Sigma_V}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}\mathbf{X_s} \sim \mathcal{X}^2_{card(\mathbf{s})}(\delta)$, where $\mathcal{X}^2_{card(\mathbf{s})}(\delta)$ represents a non-central chi-square distribution with non-centrality parameter $\delta = (\mathbf{A}^{-1}\Delta)^{\mathrm{T}}_{\mathbf{s}}\ \{[\mathbf{A}^{-1}\mathbf{\Sigma_V}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}(\mathbf{A}^{-1}\Delta)_{\mathbf{s}}$ and degree of freedom $card(\mathbf{s})$. Since δ is a function of $\mathbf{s}$, we can express it as $\delta(\mathbf{s})$. According to the analysis in Section 3.1, Δ is detectable if and only if

$$\beta = ncx2cdf(\mathcal{X}^2_{card(\mathbf{s}),\alpha}, card(\mathbf{s}), \delta(\mathbf{s})) \leq 1 - \frac{1}{ARL_U(\tau)}, \tag{10}$$

where $ncx2cdf(\mathcal{X}^2_{card(\mathbf{s}),\alpha}, card(\mathbf{s}), \delta(\mathbf{s}))$ is a built-in function in MATLAB for computing the non-central chi-square cumulative probability at the value $\mathcal{X}^2_{card(\mathbf{s}),\alpha}$ with the corresponding degrees of freedom $card(\mathbf{s})$ and non-centrality parameter $\delta(\mathbf{s})$. Assume that Equation (10) achieves equality at $\delta(\mathbf{s}) = \tilde{\delta}(card(\mathbf{s}))$ given the cardinality of the set $\mathbf{s}$. Since $ncx2cdf$ is a decreasing function in terms of the

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

non-centrality parameter given the degree of freedom parameter, a definition of *feasible solution* is presented as follows.

Definition 2. An allocation set **s** is a *feasible solution* to detect the mean shift Δ if and only if $\delta(\mathbf{s}) \geq \tilde{\delta}(card(\mathbf{s}))$.

In other words, Equation (9) is simplified into the following problem:

$$\min_{\mathbf{s}} C = \sum_{i \in \mathbf{s}} C_i,$$
$$\text{subject to } \delta(\mathbf{s}) \geq \tilde{\delta}(card(\mathbf{s})). \quad (11)$$

Since the optimal solution depends on the cardinality of the set **s**, in this article we propose a Best Allocation Subsets by Intelligent Search (BASIS) algorithm in terms of *card*(**s**) to intelligently search for an optimal sensor allocation solution which resembles the idea of the "best" subsets algorithm for model selection in regression (Hocking and Leslie, 1967). In the best subsets algorithm, several "good" candidate sets are identified for each possible number of predictors, which provide additional information for the investigator in making the final selection of predictors for the regression model. Similarly, several feasible solution sets can be identified in terms of the different number of sensors allocated in the system. Here *intelligent* means that the optimal sensor allocation solution can be derived exhaustively without conducting complex simulation studies. The BASIS algorithm is illustrated in Fig. 3 and all of the feasible solutions are recorded in the set "**Opt.**" Accordingly, the optimal solution can be achieved by choosing the one from **Opt** with the lowest sensor cost. The computation intensity of the BASIS algorithm is shown in Appendix A.

3.3. *Solution to optimal sensor allocation for quick mean shifts detection*

3.3.1. *Single mean shift case*

We start by investigating the simplest case in which our detection scenario of interest is that at most one variable in the system may possibly have a mean shift with at least magnitude τ. Denote all of our interested detection scenarios of Δ in Fig. 3 as $\bar{\Delta}$, which form a space as R_1 (i.e., $\Delta \in R_1$). In addition, denote the space specified by the kth mean shift scenario as R_1^k. Let $\Delta(k) \in R_1^k$ be the single mean shift case with the kth element equal to $\tau (k = 1, 2, \ldots, m)$. Then $\bar{\Delta}$ can be summarized as a countable set that only contains m elements, $\bar{\Delta} = \{\Delta(1), \Delta(2), \ldots, \Delta(m)\}$. Modifying $\delta(\mathbf{s}_i(j))$ in Fig. 3 to be

$$\min_{\Delta(k)} \left\{ (\mathbf{A}^{-1}\Delta(k))^{\mathrm{T}}_{\mathbf{s}_i(j)} \{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}_i(j)}\}^{-1} (\mathbf{A}^{-1}\Delta(k))_{\mathbf{s}_i(j)} \right\}, k = 1, 2, \ldots, m, \quad (12)$$

we get the following proposition.

Proposition 1. *For single mean shift detection, if* **s** *is a feasible solution obtained from the BASIS algorithm, then any* $\tilde{\Delta} \in R_1$ *is detectable.*

Proof. See Appendix B. ■

According to Appendix A, it can be shown that the running time of the BASIS algorithm for the single mean shift case is bounded by $\Theta\{(2^{m-card(\mathbf{LF})}) \times m\}$.

1. *For i=m:*-1:1*,*

   Compute $\tilde{\delta}(i)$*, where* $ncx2cdf\left(\chi^2_{i,\alpha}, i, \tilde{\delta}(i)\right) = 1 - \frac{1}{ARL_U(\tau)}$.

2. *List all searching sets with cardinality i. The number of sets is* $n_i = \binom{m}{i}$ *and each set is denoted as* $\mathbf{s}_i(j), j = 1,2,\ldots,n_i$. *Denote flagsol*=0.

   a) *For j*=1*:* n_i*,*

   b) *Compare* $\delta(\mathbf{s}_i(j)) = (\mathbf{A}^{-1}\Delta)^{\mathrm{T}}_{\mathbf{s}_i(j)}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}_i(j)}\}^{-1}(\mathbf{A}^{-1}\Delta)_{\mathbf{s}_i(j)}$ *with* $\tilde{\delta}(i)$.

   c) *If* $\delta(\mathbf{s}_i(j)) \geq \tilde{\delta}(i)$*,*

   d) $\mathbf{s}_i(j)$ *is a feasible solution and record* $\mathbf{s}_i(j)$ *into set* **Opt**. *flagsol*=1*;*

   e) *End.*

   *If (flagsol==*0*), break.*

3. *End.*

Fig. 3. BASIS algorithm for optimal sensor allocation solution.

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

3.3.2. *Multiple mean shifts case*

The BASIS algorithm proposed in Fig. 3 can also be generalized for detection of multiple mean shifts where at most p variables in the system may possibly have a mean shift with at least magnitude τ. Denote the space of our interested detection scenarios by R_p. Given that at most p variables may possibly have a mean shift, the total number of faulty scenarios is $\sum_{l=1}^{p}\binom{m}{l}$. Denote the space defined by the kth mean shift scenario as R_p^k, so that:

$$R_p = \bigcup_{k=1}^{\sum_{l=1}^{p}\binom{m}{l}} R_p^k.$$

In the single mean shift case, all of our interested detection scenarios $\bar{\Delta}$ can be treated as a finite countable set, of which the cardinality equals the total number of variables in the system. In addition, each element of the set $\bar{\Delta}$ has exactly one non-zero entry that is equal to τ. However, in the multiple mean shifts case, since positive and negative mean shifts mitigate each other, the cumulative mean shift propagated to the descendant node can be indiscernible though the magnitudes of mean shifts from ancestors are significant. Thus, $\bar{\Delta}$ cannot be easily expressed as a countable set as in the single mean shift case.

In order to address the problem of optimal sensor allocation for multiple mean shifts detection by the BASIS algorithm in Fig. 3, we rewrite $\delta(\mathbf{s}_i(j))$ to be a quadric optimization function with the following form:

$$\delta(\mathbf{s}_i(j)) = \min_{k}\left\{\min_{\Delta(k)}\left\{\Delta(k)^{\mathrm{T}}(\mathbf{A}^{-1})^{\mathrm{T}}_{\mathbf{s}_i^{\mathrm{r}}(j)}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}} \times (\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}_i(j)}\}^{-1}(\mathbf{A}^{-1})_{\mathbf{s}_i^{\mathrm{r}}(j)}\Delta(k)\right\}\right\}, \quad (13)$$

$$\text{subject to} \quad \Delta(k) \in R_p^k,\ k = 1, 2, \ldots, \sum_{l=1}^{p}\binom{m}{l},$$

where $(\mathbf{A}^{-1})_{\mathbf{s}_i^{\mathrm{r}}(j)}$ is the sub-block matrix of $\mathbf{A}^{-1}$ with only rows indexed by $\mathbf{s}_i(j)$; $\Delta(k) \in R_p^k$ is any mean shift case that belongs to the kth faulty scenario ($k = 1, 2, \ldots, \sum_{l=1}^{p}\binom{m}{l}$). Since $(\mathbf{A}^{-1})^{\mathrm{T}}_{\mathbf{s}_i^{\mathrm{r}}(j)}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}_i(j)}\}^{-1}(\mathbf{A}^{-1})_{\mathbf{s}_i^{\mathrm{r}}(j)}$ is shown to be a Positive SemiDefinite (PSD) matrix (in Appendix B), the objective function is bounded below on the feasible region and the global minimizer exists as long as the problem has a feasible solution. Based on Appendix A, it can be shown that the running time of the BASIS algorithm for the multiple mean shifts case is bounded by $\Theta\{(2^{m-card(\mathbf{LF})}) \times \sum_{l=1}^{p}\binom{m}{l}\}$.

Intuitively, when the interested maximum number of mean shift variables p increases, or the interested smallest magnitude of mean shift τ reduces, or the interested maximum average run length requirement $ARL_U(\tau)$ decreases, the final number of sensors needed to satisfy the specified requirement will increase. In other words, the higher the uncertainty about the system, the higher the number of sensors required. Thus, the proposed methodology is more favorable for a relatively robust manufacturing production system in which it is rare to see too many variables simultaneously have a mean shift. Wang and Jiang (2009) showed that the likelihood of multiple sources shifting simultaneously decreases dramatically as the number of shifted sources increases and concluded that the likelihood of one source shifting dominates. Therefore, choosing the allocation strategy is a trade-off problem in practice, and the final solution depends on how to allocate priorities on cost and detection.

In reality, the practitioner can determine the interested minimum magnitude of mean shifts for detection, the maximum number of simultaneous mean shift variables, and the upper bound of the specified ARL based on engineering domain knowledge to ensure the production yield. The engineering knowledge includes, but is not limited to, the tolerance specifications for each variable specified in product/process design (related to the interested minimum magnitude of mean shifts, τ), the probability that a fault occurred at each variable (related to the maximum number of simultaneous mean shift variables, p), and the cost of defective products produced (related to the specified ARL requirement, $ARL_U(\tau)$).

4. Diagnosis ranking for multiple mean shifts

4.1. *Problem formulation*

After implementing the optimal allocation strategy, the next topic is how to diagnose the root cause variables if an out-of-control signal is detected by the control chart. The existing methodologies, such as MTY T^2 decomposition (Mason *et al.*, 1997) or causation-based T^2 decomposition (Li, Jin, and Shi, 2008), cannot be applied here since that require that the measurements of all variables are available. Therefore, it is desirable to develop a new diagnosis method for the case when only partial information is at hand. The difficulties of this problem arise from two aspects: (i) it is challenging to diagnose mean shifts from m nodes using only a few sensors, since only the variables with a sensor deployment are measurable; and (ii) it becomes more complicated when different combinations of the signs and magnitudes of the mean shifts are taken into consideration. To solve these problems, we propose a diagnosis ranking method that can be regarded as an extension to the LASSO-based variable selection method (Wang and Jiang, 2009). Recall that we investigate an optimal allocation strategy subjected to the requirements that at most p variables in the system may possibly have a mean shift before an out-of-control signal is triggered and the diagnosis procedure is conducted after implementing the optimal sensor allocation strategy. Therefore, we further assume that the number of true mean shift variables is bounded by p. As a result, the total number of faulty scenarios is $\sum_{l=1}^{p}\binom{m}{l}$ and the space defined by the kth mean shift scenario is denoted as R_p^k. Similarly, $\Delta(k) \in R_p^k$ is any mean shift case that belongs to the kth faulty scenario ($k = 1, 2, \ldots, \sum_{l=1}^{p}\binom{m}{l}$).

Constrained to the space defined by each mean shift scenario, the constrained maximum likelihood approach can be implemented to rank the possibility of each mean shift given the out-of-control sample. It is clear that the constrained maximum likelihood approach is equivalent to solving the following problem given the out-of-control signal $\mathbf{X_s}$:

$$\min_{\Delta(k)} (\mathbf{X}-\mathbf{A}^{-1}\Delta(k))_{\mathbf{s}}^{\mathrm{T}}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}(\mathbf{X}-\mathbf{A}^{-1}\Delta(k))_{\mathbf{s}},$$
$$\text{subject to } \Delta(k) \in R_p^k. \tag{14}$$

Again, Equation (14) can be transformed into a quadratic optimization problem as follows:

$$\min_{\Delta(k)} \left\{\Delta(k)^{\mathrm{T}}(\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{T}}\mathbf{M}(\mathbf{A}^{-1})_{\mathbf{s}^r}\Delta(k) - 2\mathbf{X}_{\mathbf{s}}^{\mathrm{T}}\mathbf{M}(\mathbf{A}^{-1})_{\mathbf{s}^r}\Delta(k)\right\}, \tag{15}$$
$$\text{subject to } \quad \Delta(k) \in R_p^k,$$

where $\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}$ is defined as $\mathbf{M}$ for compactness. If the minimized objective value is larger than or equal to the critical value $\mathcal{X}^2_{card(\mathbf{s}),\alpha} - \mathbf{X}_{\mathbf{s}}^{\mathrm{T}}\mathbf{M}\mathbf{X}_{\mathbf{s}}$, then we have at least $(1-\alpha)\%$ confidence to say that the kth faulty scenario is unlikely to be the true root cause.

The proposed diagnosis ranking method is illustrated in Fig. 4, where the final ranked solutions are recorded in the set $\widetilde{\mathbf{RD}}$. Assume that there are q ranked solutions included in $\widetilde{\mathbf{RD}}$ and the kth one is denoted as $\widetilde{\mathbf{RD}}^k$. Therefore, the final conclusion about the mean shift variables is identified in $\widetilde{\mathbf{RD}}^1$.

The idea behind Fig. 4 is inspired by the LASSO-based variable selection method (Wang and Jiang, 2009). That method assumes that not all variables will shift simultaneously and there is an upper bound on the total number of mean shift variables. This upper bound can be set to be p as discussed at the beginning of this section. The original LASSO-based variable selection algorithm assumes that all variables are measurable and we extend this diagnosis method to the case where only partial observations are available. Due to this extension, new issues (e.g., fault diagnosability and minimal diagnosable class) will be illustrated and discussed in the following sections. One problem associated with the proposed diagnosis ranking method is that the faulty scenarios with the number of p true mean shift variables always occur in the first few ranks in $\widetilde{\mathbf{RD}}$ since more degrees of freedom are available for those cases in the quadratic optimization function. Recall that the procedure suggested by the classic statistical process control design recommends the investigator to stop the machine and utilize the out-of-control action plan immediately when a certain out-of-control sample occurs. Therefore, it is more desirable that some in-control variables are false positives than the case that certain true mean shift variables are excluded from $\widetilde{\mathbf{RD}}^1$. Accordingly, the definition for *correct diagnosis* is given as follows.

Definition 3. A diagnosis result is called *correct* if all true mean shift variables are in $\widetilde{\mathbf{RD}}^1$.

4.2. *Fault diagnosability and minimal diagnosable class*

One problem for diagnosis with partial information is that a certain faulty scenario might be an alias of others. Given that the total number of faulty scenarios is $\sum_{l=1}^{p}\binom{m}{l}$ and the space defined by the kth fault is denoted as R_p^k, we get the following definition for *diagnosable*.

Definition 4. The diagnosis problem is said to be *diagnosable* if for any non-zero $\Delta_1 \in R_p^{k_i}, \Delta_2 \in R_p^{k_j}$, such that $(\mathbf{A}^{-1}\Delta_1)_{\mathbf{s}} = (\mathbf{A}^{-1}\Delta_2)_{\mathbf{s}}$, then $i = j (i, j \in \{1, 2, \ldots, \sum_{l=1}^{p}\binom{m}{l}\})$.

The intuition behind this definition is that if two mean shift cases are from different faulty scenarios, we can see distinct measurement values on sensors. In other words, we should be able to distinguish different mean shift scenarios

For k=1: $\sum_{l=1}^{p}\binom{m}{l}$,

Solve $obj(k) = \min_{\Delta(k)}\{\Delta(k)^{\mathrm{T}}(\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{T}}\mathbf{M}(\mathbf{A}^{-1})_{\mathbf{s}^r}\Delta(k) - 2\mathbf{X}_{\mathbf{s}}^{\mathrm{T}}\mathbf{M}(\mathbf{A}^{-1})_{\mathbf{s}^r}\Delta(k)\}$, *subject to*

$\Delta(k) \in R_p^k$.

If $obj(k) \le \mathcal{X}^2_{card(\mathbf{s}),\alpha} - \mathbf{X}_{\mathbf{s}}^{\mathrm{T}}\mathbf{M}\mathbf{X}_{\mathbf{s}}$,

Add $obj(k)$ *into* **RD**.

End.

Sort **RD** *according to* obj *values and record into* $\widetilde{\mathbf{RD}}$.

Fig. 4. Diagnosis ranking method for multiple mean shifts.

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

in order to uniquely identify the root cause variables. According to the definition of diagnosable, we introduce the definition of *minimal diagnosable class*. The concept of minimal diagnosable class was first discussed in Zhou *et al*. (2003) and Shi (2006) and the definition was given based on a state-space model. The minimal diagnosable class reveals the interrelationship among different faults. A minimal diagnosable class is a group of faults that cannot be further distinguished. In this article, the definition of minimal diagnosable class is represented as follows.

Definition 5. A set of faults $\{k_i, \ldots, k_j\}(i, j \in \{1, 2, \ldots, \sum_{l=1}^{p} \binom{m}{l}\})$ forms a *minimal diagnosable class* if there exists non-zero $\Delta_i \in R_p^{k_i}, \ldots, \Delta_j \in R_p^{k_j}$ such that $(\mathbf{A}^{-1}\Delta_i)_{\mathbf{s}} = \cdots = (\mathbf{A}^{-1}\Delta_j)_{\mathbf{s}}$.

4.3. *Solution and analysis to diagnosis ranking method and diagnosability*

4.3.1. *Single mean shift case*

We start by investigating the simplest case where a single variable in the system may possibly have a mean shift. By implementing the algorithm as shown in Fig. 4 where $p = 1$, the group of faults in a minimal diagnosable class will achieve the same objective value by the diagnosis ranking method. Therefore, if one of the faults in a minimal diagnosable class is identified in $\widetilde{\mathbf{RD}}^1$, the root cause cannot be further distinguished. Given a sensor allocation set $\mathbf{s}$, the minimal diagnosable class can be identified in the single mean shift scenario by the following proposition.

Proposition 2. *Given a single mean shift diagnosis case with m faults $\Theta = \{k_1, k_2, \ldots, k_m\}$ and k_i is the ith mean shift fault, the fault set $\Theta[\mathbf{v}] \subseteq \Theta$ is a minimal diagnosable class if the non-zero column vector $\boldsymbol{\tau}(i)$ of the Reduced Row Echelon Form (RREF) of $(\mathbf{A}^{-1})_{\mathbf{s}^r}$ can be expressed as $\boldsymbol{\tau}(i) = a\boldsymbol{\tau}(j)$ for $\forall v(i), v(j) \in \Theta[\mathbf{v}]$, where $v(i)$ is a column index for $\boldsymbol{\tau}(i)$ and a is a constant number.*

Proof. See Appendix C. ■

4.3.2. *Multiple mean shifts case*

To extend the single mean shift case to the multiple mean shifts case, we consider the framework where the number of true mean shift variables is bounded by p. Similar to the single mean shift case, the group of faults in a minimal diagnosable class will achieve the same objective value by implementing the diagnosis ranking method in Fig. 4. The minimal diagnosable class can be identified in terms of the number of out-of-control variables g from the following proposition, where $1 \le g \le p$.

Proposition 3. *Given $g(1 \le g \le p)$ out-of-control variables with $\binom{m}{g}$ faults $\Theta = \{k_1, k_2, \ldots, k_{\binom{m}{g}}\}$ and k_i is the ith mean shift fault, the fault set $\Theta[\mathbf{v}] \subseteq \Theta$ is a minimal diagnosable class if the non-zero column vector $\tau(i)$ of the RREF of $(\mathbf{A}^{-1})_{\mathbf{s}^r}$ can be expressed as $\sum_{l=1}^{g}[\boldsymbol{\tau}(i)]_l = \sum_{l=1}^{g} a_l[\boldsymbol{\tau}(j)]_l$ for $\forall \mathbf{v}(i), \mathbf{v}(j) \in \Theta[\mathbf{v}]$, where $\mathbf{v}(i)$ is the column index for $\boldsymbol{\tau}(i)$, $[\boldsymbol{\tau}(i)]_l$ is the lth element of $\boldsymbol{\tau}(i)$, and a_l is a constant number.*

Proof. The proof for Proposition 3 follows the same logic as Proposition 2 and thus is omitted here. ■

An example is provided here to illustrate this proposition. We consider a sensor allocation strategy for the hot forming problem in Fig. 1. Sensors are allocated on {1, 3, 5} to monitor the system where at most two variables ($p = 2$) have a mean shift until an out-of-control signal is detected. Then the RREF of $(\mathbf{A}^{-1})_{\mathbf{s}^r}$ is as follows:

$$\begin{pmatrix} 1.000 & 0.574 & 0 & 0.283 & 0 \\ 0 & 0 & 1.000 & 0.688 & 0 \\ 0 & 0 & 0 & 0 & 1.000 \end{pmatrix}$$

From Proposition 2, when only one mean shift occurs ($g = 1$), Θ ={{1}, {2}, {3}, {4}, {5}}. $\Theta[\mathbf{v}]$ ={{1}, {2}} forms a minimum diagnosable class and the total minimal diagnosable classes are {{1}, {2}}, {{3}}, {{4}}, {{5}}. Similarly, when two mean shifts occur ($g = 2$), the total minimal diagnosable classes are {{1, 3}, {2, 3}, {1, 4}, {2, 4}, {3, 4}}, {{1, 5}, {2, 5}}, {{1, 2}}, {{3, 5}}, {{4, 5}}.

4.4. *Optimal sensor allocation strategy by considering diagnosability*

In practice, when choosing an optimal allocation strategy, a decision maker will consider sensor cost, detection speed of the mean shifts, and fault diagnosis accuracy in an integrated manner. Generally speaking, the more sensors that are deployed in the system, the higher the accuracy of the diagnosis procedure and the quicker the detection of the mean shift, which is a trade-off with the higher cost of sensor implementation. Considering the example at the end of the last section, {{1, 3}, {2, 3}, {1, 4}, {2, 4}, {3, 4}} forms a minimum diagnosable class, which indicates that those faulty scenarios cannot be further distinguished. While two sensors are saved in this allocation strategy, this design has a very poor performance in terms of diagnosibility. Therefore, this section aims at discussing how to choose an optimal allocation strategy under the constraint that the problem is diagnosable. This task can be fulfilled by the proposed BASIS algorithm in Fig. 3 with minor revisions: adding a step to check whether the problem is diagnosable by Propositions 2 and 3 after 2a and modifying 2c to be "*If $\delta(\mathbf{s}_i(j)) \ge \tilde{\delta}(i)$ and the problem is diagnosable.*"

Although the added step to check whether the problem is diagnosable works adequately from a mathematical perspective, the procedure to derive an optimal solution is limited by the interpretation of the graph. In fact, it is a very challenging problem to find the optimal allocation solution directly from the graph. However, the following proposition can significantly narrow down the searching space in the modified BASIS algorithm to those nodes that have more than one child.

Proposition 4. *If it is a feasible sensor allocation solution given that the problem is diagnosable, then it contains all nodes that have at most one child.*

Proof. See Appendix D. ■

5. Case studies

In this section, we revisit the hot forming process introduced in Section 2 to evaluate the performance of the proposed BASIS algorithm in Fig. 3 and the diagnosis ranking method in Fig. 4 under different specified objectives for mean shift detection and diagnosis. The parameters selected in the case studies are within the typical ranges of actual systems and thus the case studies could represent the characteristics of the actual systems. In addition, in order to demonstrate the effectiveness of the BASIS algorithm, a cap alignment process is considered. The cost of each sensor is assumed to be same in the following case studies.

5.1. *Optimal sensor allocation strategy for quick mean shifts detection in the hot forming process*

5.1.1. *Single mean shift case*

Table 1 shows the feasible solutions obtained by the BASIS algorithm for a single mean shift fault subjected to different specified objectives (different combinations of τ and $ARL_U(\tau)$). These different objectives reflect the specific needs in practice. In the study, α is chosen to be 0.01. The optimal solution is the allocation strategy that has the smallest cardinality among all feasible solutions under each specified objective. As expected, the proposed BASIS algorithm is able to provide all feasible solutions in terms of the number of sensors needed in the system.

5.1.2. *Simulation evaluation and comparison with the integrated causal models and set-covering algorithms (Li and Jin, 2010)*

This section will evaluate the performance of the sensor allocation solutions proposed by the BASIS algorithm. There are five scenarios for a potential single mean shift, where each mean shift magnitude is equal to or greater than τ. However, it is sufficient for us to consider the case with mean shift magnitude τ, since a larger magnitude of mean shift is more noticeable using sensors under the single mean shift case. Moreover, it is sufficient to evaluate the boundary cases that either have a fewer number of sensors proposed as τ increases given the same $ARL_U(\tau)$ requirement or have a fewer number of sensors proposed as $ARL_U(\tau)$ increases given the same magnitude of τ. As a result, there are a total of five boundary cases identified (bold font) and the case number is denoted in superscript in Table 1. In fact, the algorithms (Li and Jin, 2010) choose to allocate sensors on the bold entries or the subsets of bold entries in the five boundary cases of Table 1. Recall that the greater the number of variables that are measured, the quicker the mean shift can be detected. Thus, comparing the ARL result with sensors deployed on the bold entries could provide a clear demonstration about the advantages of the BASIS algorithm over the integrated causal models and set-covering algorithms. A comparison of the performances of the five boundary cases by these two algorithms is shown in Table 2.

The evaluation process involves the following steps.

1. In each selected boundary case, a mean shift τ to each variable X_i ($i \in \{1, 2, 3, 4, 5\}$) is introduced.
2. A $M(= 5000)$ samples dataset of $\mathbf{X} = \{X_1, X_2, X_3, X_4, X_5\}$ in terms of their distributions with the introduced mean shift at the first sample of each X_i is simulated.
3. For each sample, testing statistic $\mathbf{X}_\mathbf{s}^{\mathrm{T}}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_\mathbf{V}(\mathbf{A}^{-1})^{\mathrm{T}}]_\mathbf{s}\}^{-1}\mathbf{X}_\mathbf{s}$ is plotted on a chi-square control chart with upper control limit $\mathcal{X}^2_{card(\mathbf{s}),\alpha}$ and lower control limit of zero. On the other hand, each univariate $X_j (j \in \mathbf{s})$ is plotted on a Shewhart control chart with control limits $\pm z_{\alpha/(2\times card(\mathbf{s}))}$, where the Bonferroni

Table 1. BASIS algorithm solutions for single mean shift case under different specified objectives

	τ			
$ARL_U(\tau)$	*1.5*	*2*	*2.5*	*3*
10	No solution	**{1, 2, 3, 4, 5}**[2]	**{1, 3, 5}**[4], {1, 2, 3, 5}, {1, 3, 4, 5}, {1, 2, 3, 4, 5}	{1, 3, 5}, {1, 2, 3, 5}, {1, 3, 4, 5}, {1, 2, 3, 4, 5}
15	No solution	{1, 2, 3, 4, 5}	{1, 3, 5}, {1, 2, 3, 5}, {1, 3, 4, 5}{1, 2, 3, 4, 5}	{1, 3, 5}, {1, 2, 3, 5}, {1, 3, 4, 5}, {1, 2, 3, 4, 5}
20	**{1, 2, 3, 4, 5}**[1]	**{1, 3, 5}**[3], {1, 2, 3, 5}, {1, 3, 4, 5}, {1, 2, 3, 4, 5}	{1, 3, 5}, {1, 2, 3, 5}, {1, 3, 4, 5}{1, 2, 3, 4, 5}	**{1, 2, 5}**[5],{1, 3, 5}, {1, 2, 3, 5}, {1, 2, 4, 5}, {1, 3, 4, 5}, {1, 2, 3, 4, 5}

1, 2, 3, 4, 5: Identified five boundary cases.

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

Table 2. Estimated ARL for the BASIS algorithm and the integrated causal models and set-covering algorithms for $\alpha = 0.01$

	BASIS algorithm (1)					*Integrated causal models and set-covering algorithm (2)*				
Boundary cases	X_1	X_2	X_3	X_4	X_5	X_1	X_2	X_3	X_4	X_5
Case 1: $\overline{RL}_{k,\cdot}$	5.65	10.14	7.49	17.44	17.47	16.21	15.59	16.04	12.86	15.29
Case 2: $\overline{RL}_{k,\cdot}$	2.48	4.45	3.30	8.29	8.28	7.03	6.73	6.84	6.03	6.92
Case 3: $\overline{RL}_{k,\cdot}$	3.68	15.42	5.66	14.90	6.06	5.62	23.38	5.55	12.44	5.57
Case 4: $\overline{RL}_{k,\cdot}$	2.04	8.99	3.08	8.31	3.31	2.99	13.87	2.93	7.14	3.02
Case 5: $\overline{RL}_{k,\cdot}$	1.18	1.80	15.61	9.45	2.10	1.92	1.89	31.45	8.14	1.89

correction method is implemented to maintain system Type 1 error as in the integrated causal models and set-covering algorithms.

4. Indices of the first out-of-control sample on the control charts proposed by these two methods are recorded as RL_{1,X_i} and RL_{2,X_i}, respectively.
5. Repeat steps 1 to 4 for $N(= 10\,000)$ times. The average of $RL_{k,X_i}(k = 1, 2)$, $\overline{RL}_{k,X_i}$ of the introduced mean shift on each variable X_i for these two methods is computed as presented in Table 2.

The following conclusions can be inferred from the evaluation studies reported in Table 2.

1. The BASIS algorithm outperforms the algorithm of Li and Jin (2010) in the sense of meeting ARL requirement. It is clear that all of the solutions proposed by the BASIS algorithm can satisfy the ARL requirement under each scenario, whereas the algorithm (Li and Jin, 2010) cannot guarantee that, as shown in cases 3, 4, and 5. The reason is that the latter algorithm does not take the Bonferroni correction method into consideration during the search procedure for the optimal solution. However, it is not a problem for the BASIS algorithm since it is conducted in terms of the number of sensors deployed.
2. It is guaranteed that there is no better allocation solution other than the ones indicated by the BASIS algorithm due to the exhaustive property of the proposed method. For example, when $\tau = 2$ and $ARL_U(\tau) = 15$, $\{1, 3, 5\}$ is not identified as a feasible solution, which can be verified by case 3 in Table 2 since the ARL needed to detect a mean shift occuring at X_2 is greater than 15.
3. The mean shift occurring at a variable with sensor deployment is straightforwardly shown to be able to be detected faster than the one without sensor deployment. This phenomenon is consistent with the one in Li and Jin's algorithms.

5.1.3. *Multiple mean shifts case*

As discussed in Section 3.3.2, the BASIS algorithm can be extended to the multiple mean shifts case where at most p variables in the system may have a mean shift with at least magnitude τ. In this case study, p is chosen to have a value of two. Similar to the single mean shift case, different specified objectives for detection (different combinations of τ and $ARL_U(\tau)$) are considered. By replacing the objective function in Fig. 3 (2b) with Equation (13) as discussed in Section 3.3.2, the feasible solutions are shown in Table 3. The evaluation process is similar to the single mean shift case as shown in Section 5.1.2 and thus it is omitted here.

5.2. *Diagnosis ranking method for true mean shifts in the hot forming process*

5.2.1. *Single mean shift case*

Since there are five variables in the hot forming process, five potential single mean shift faults are of interest. Specifically, we conduct the evaluation process based on the allocation solution $\{1, 3, 5\}$, which is a feasible solution under most of the specified objectives in Table 1. Furthermore, the minimum diagnosable class can be shown to be $\{\{1\},\{2\}\}, \{\{3\}\}, \{\{4\}\}, \{\{5\}\}$ by Proposition 2. Therefore, the mean shift scenarios from X_1 and X_2 are indiscernible and these two faults will achieve the same objective value by the diagnosis ranking method. Table 4 provides the correct diagnosis rate in terms of Definition 3 by the diagnosis ranking method in Fig. 4, where the single mean shift scenarios with different mean shift values are studied.

Table 3. BASIS algorithm solutions for multiple (two) mean shifts under different specified objectives

	τ			
$ARL_U(\tau)$	*1.5*	*2*	*2.5*	*3*
10	No solution	{1, 2, 3, 4, 5}	{1, 2, 3, 5}, {1, 2, 3, 4, 5}	{1, 2, 3, 5}, {1, 2, 3, 4, 5}
15	No solution	{1, 2, 3, 4, 5}	{1, 2, 3, 5}, {1, 2, 3, 4, 5}	{1, 2, 3, 5}, {1, 2, 3, 4, 5}
20	{1, 2, 3, 4, 5}	{1, 2, 3, 5}, {1, 2, 3, 4, 5}	{1, 2, 3, 5}, {1, 2, 3, 4, 5}	{1, 2, 3, 5}, {1, 2, 3, 4, 5}

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

Table 4. Single mean shift scenarios and correct diagnosis rates for solution {1, 3, 5}

	Mean shift in					$\tau = 1.5$	$\tau = 2.0$	$\tau = 2.5$	$\tau = 3.0$
Scenario	X_1	X_2	X_3	X_4	X_5	*Corr. prob*	*Corr. prob*	*Corr. prob*	*Corr. prob*
1	τ					0.852	0.923	0.962	0.981
2		τ				0.665	0.766	0.838	0.889
3			τ			0.600	0.684	0.739	0.780
4				τ		0.485	0.577	0.641	0.700
5					τ	0.832	0.913	0.947	0.974

It can be seen from Table 4 that the correct diagnosis rate increases as the mean shift magnitude gets larger, which is consistent with our intuition since the chi-square control chart is insensitive in detecting small mean shifts. Moreover, when the mean shift occurs at variable X_4, the correct diagnosis rate is the lowest. Since no sensor is allocated on X_4, the mean shift occurring at X_4 is likely to be mistakenly identified as a mean shift from either X_3 or X_5.

5.2.2. *Multiple mean shifts case*

As discussed in Section 4.3.2, the diagnosis ranking method can also be generalized for the multiple mean shifts case. We implement the sensor allocation solution {1, 2, 3, 5} based on the results in Table 3 and then continue the diagnosis procedure. Furthermore, p, the maximum number of true mean shift variables, is chosen to have a value of two. The minimum diagnosable classes are identified as {{1}}, {{2}}, {{3}}, {{4}}, {{5}}, {{1, 2}}, {{1, 3}}, {{1, 4}}, {{1, 5}}, {{2, 3}, {2, 4}, {3, 4}}, {{2, 5}}, {{3, 5}}, {{4, 5}} by Proposition 3. To make a thorough study of the diagnosis performance in response to different faults, a set of 25 experiments is designed to cover three general groups of possible mean shift patterns as shown in Table 5. The first group (cases 1 to 5) represents the single mean shift fault on each variable. The second and third groups (cases 6 to 15 and 16 to 25) represent the situations in which two variables simultaneously have an identical shift magnitude but in the same and opposite directions. Different magnitudes of mean shift in each fault scenario are also studied.

According to the results in Table 5, a similar pattern exists here as in the single mean shift diagnosis case: the larger the

Table 5. Multiple mean shifts scenarios and correct diagnosis rates for solution {1, 2, 3, 5}

	Mean shift in					$\tau = 1.5$	$\tau = 2.0$	$\tau = 2.5$	$\tau = 3.0$
Scenario	X_1	X_2	X_3	X_4	X_5	*Corr. prob*	*Corr. prob*	*Corr. prob*	*Corr. prob*
1	τ					0.987	0.997	0.999	1.000
2		τ				0.825	0.882	0.914	0.923
3			τ			0.631	0.688	0.729	0.762
4				τ		0.555	0.631	0.681	0.713
5					τ	0.942	0.978	0.992	0.998
6	τ	τ				0.632	0.778	0.999	0.931
7	τ		τ			0.546	0.667	0.751	0.837
8	τ			τ		0.428	0.507	0.587	0.670
9	τ				τ	0.662	0.778	0.875	0.937
10		τ	τ			0.301	0.377	0.426	0.465
11		τ		τ		0.440	0.566	0.673	0.768
12		τ			τ	0.576	0.719	0.827	0.907
13			τ	τ		0.273	0.355	0.442	0.550
14			τ		τ	0.501	0.640	0.740	0.818
15				τ	τ	0.408	0.507	0.605	0.673
16	τ	$-\tau$				0.646	0.774	0.869	0.933
17	τ		$-\tau$			0.538	0.651	0.746	0.834
18	τ			$-\tau$		0.410	0.507	0.580	0.662
19	τ				$-\tau$	0.665	0.788	0.870	0.935
20		τ	$-\tau$			0.547	0.693	0.814	0.904
21		τ		$-\tau$		0.339	0.430	0.526	0.622
22		τ			$-\tau$	0.589	0.727	0.831	0.907
23			τ	$-\tau$		0.242	0.315	0.364	0.422
24			τ		$-\tau$	0.505	0.634	0.732	0.822
25				τ	$-\tau$	0.408	0.508	0.601	0.673

mean shift magnitude, the higher the correct diagnosis rate. Moreover, there are no significant differences in the correct diagnosis rates comparing cases 6 to 15 and cases 16 to 25. Generally speaking, the diagnosis accuracy is poor when the mean shift variable contains X_4, since X_4 has no sensor allocated.

5.3. *Optimal sensor allocation strategy given problem is diagnosable in the hot forming process*

This section aims to study the optimal sensor allocation strategy given that the problem is diagnosable as discussed in Section 4.4. By implementing Proposition 4, $\{1, 2, 3, 5\}$ is required to allocate a sensor. In other words, $\{1, 2, 3, 5\}$ is a subset of the feasible solutions and it remains to be determined whether node 4 is also supposed to have a sensor deployed. Thus, the search space is significantly narrowed down by Proposition 4. By the modified BASIS algorithm in Section 4.4, we get the following conclusions about the optimal solution.

For a single mean shift detection and diagnosis, the problem is diagnosable if sensors are allocated on $\{1, 2, 3, 5\}$. Then, the feasible sensor allocation solutions given that the problem is diagnosable can be directly obtained from Table 1 by eliminating the ones that do not contain the subset $\{1, 2, 3, 5\}$. However, for the case with multiple mean shifts, the problem is not diagnosable if sensors are only deployed on $\{1, 2, 3, 5\}$ by Proposition 3. Thus, $\{4\}$ is also required to allocate a sensor. If the number of interested shift variables is bounded by two, the corresponding optimal solution can be achieved from Table 3, which is $\{1, 2, 3, 4, 5\}$.

5.4. *Computational time analysis of the BASIS algorithm for a large-scale manufacturing process*

To demonstrate the effectiveness of the proposed BASIS algorithm, we further conduct the optimal sensor allocation study in a large-scale manufacturing process under the same combinations of τ and $ARL_U(\tau)$ requirements as shown in Table 1. Specifically, a BN with 35 variables involved in a cap alignment process at Hewlett Packard is considered (see Figure 2 in Wolbrecht *et al*. (2000)). Due to the page limits, we only focus on the computational time of the BASIS algorithm in this case study.

All experimental studies were performed using MATLAB V7.9 in the Windows 7 operating system with two Intel Core i7-2820QM 2.30 GHz processors and 8 GB RAM. The average computational time to derive the optimal sensor allocation solution in a single mean shift case was about 75 seconds, and the average computational time in a multiple mean shifts case ($p = 2$) was about 495 seconds.

When the manufacturing workstation has hundreds or thousands of variables, directly implementing the BASIS algorithm can be inefficient. In this situation, we suggest sequentially decomposing a large BN into clustered sub-graphs if the path coefficient between these sub-graphs is insignificantly small. Then, the BASIS algorithm can be implemented separately within each clustered subsystem. If the number of variables within each cluster is still large, then it needs further improvement in the searching algorithm to find the optimal solution.

6. Conclusions

Mean shift monitoring and diagnosis are fundamental problems in multivariate process control. While numerous methodologies have been developed with adequate performance when all of the variables are measurable, it usually requires excessive cost for sensors in practice. Therefore, this article studies the problems of how to (i) optimally allocate sensors with minimum cost under different detection requirements and (ii) identify the process root cause with only partial information measurable.

In this article, a BN model is assumed to be available to represent the causal relationships among variables. The BASIS algorithm is proposed to find the optimal solution for process monitoring and then a diagnosis ranking method is developed to identify the root cause. The developed methodologies are successfully demonstrated on a hot forming process and a cap alignment process. The proposed BASIS algorithm is conducted in a systematic and intelligent way, such that for any potential solution $\mathbf{s}$ we only need to compare its non-centrality parameter $\delta(\mathbf{s})$ with the benchmark value $\tilde{\delta}(card(\mathbf{s}))$ in order to determine its feasibility. Without the BASIS algorithm, in order to derive the optimal solution, expensive Monte Carlo simulations have to be implemented to evaluate the ARL performance of each potential sensor allocation solution under each objective, which is nearly inapplicable in real practice. We would also like to point out that the sensor allocation problem is an off-line effort made in the system design stage. Thus, it is critically important to generate the best sensor placement strategy under given requirements in the design stage. As a result, the computation time spent in searching for the optimal solution is of less concern in this situation compared with real-time applications.

This study has revealed that a trade-off problem occurs when detection speed, fault diagnosis accuracy, and cost savings are taken into consideration. The monitoring and diagnosis capabilities that a DSN is able to provide depend strongly on the sensor deployment strategy in the system. Although an allocation design with fewer sensors can lower the running cost, fewer sensors can also prolong detection time and cause a loss in diagnosability. In contrast, although an allocation plan with many sensors is able to expedite the detection for a mean shift and identify faulty variables more accurately, the cost associated with these sensors can be overwhelming. The final optimal strategy requires integrating the analyses of cost and product quality, which therefore is contingent and objective oriented and will ultimately depend on how the practitioner allocates

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

the priorities toward cost, detection, and diagnosis. With guaranteed customer satisfaction by optimally designing sensor allocation strategy, the average cycle time and cost associated with sensors and inventory can be eventually cut down.

Acknowledgement

The authors gratefully acknowledge the support of the National Science Foundation under grant NSF CMMI-0927574.

References

Bakshi, B.R. (1998) Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal*, **44**(7), 1596–1610.

Buntine, W.L. (1994) Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, **2**, 159–225.

Ding, Y., Elsayed, E.A., Kumara, S., Lu, J.C., Niu, F. and Shi, J. (2006) Distributed sensing for quality and productivity improvements. *IEEE Transactions on Automation Science and Engineering*, **3**, 344–359.

Ding, Y., Kim, P., Ceglarek, D. and Jin, J. (2003) Optimal sensor distribution for variation diagnosis in multistation assembly processes. *IEEE Transactions on Robotics and Automation*, **19**, 543–556.

Duda, R.O., Hart, P.E. and Stork, D.G. (2000) *Pattern Classification*, second edition, Wiley, New York, NY.

Edan, Y. and Nof, S.Y. (2000) Sensor economy principles and selection procedures. *IIE Transactions*, **32**(3), 195–203.

Heckerman, D. (1995) A tutorial on learning with Bayesian networks, in *Learning in Graphical Models*, Jordan, M. (ed), MIT Press, Cambridge, MA, pp. 2.

Hocking, R.R. and Leslie, R.N. (1967) Selection of the best subset in regression analysis. *Technometrics*, **9**(4), 531–540.

Hotelling, H. (1947) Multivariate quality control, in *Techniques of Statistical Analysis*, Eisenhart, C., Hastay, M. and Wallis, W.A. (eds), McGraw-Hill, New York, NY, pp. 111–184.

Khan, A. and Ceglarek, D. (2000) Sensor optimization for fault diagnosis in multi-fixture assembly systems with distributed sensing. *ASME Journal of Manufacturing Science and Engineering*, **122**(1), 215–226.

Khan, A., Ceglarek, D. and Ni, J. (1998) Sensor location optimization for fault diagnosis in multi-fixture assembly systems. *ASME Journal of Manufacturing Science and Engineering*, **120**(4), 781–791.

Koller, D. and Friedman, N. (2009) *Probabilistic Graphical Models Principles and Techniques (*ì*Adaptive Computation and Machine Learning)*, The MIT Press, Cambridge, MA.

Lauritzen, S.L. and Wermuth, N. (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, **17**(1), 31–57.

Li, J., Blumenfeld, D.E. and Marin, S.P. (2007) Manufacturing system design to improve quality buy rate: an automotive paint shop application study. *IEEE Transactions on Automation Science and Engineering*, **4**(1), 75–79.

Li, J., Blumenfeld, D.E. and Marin, S.P. (2008) Production system design for quality robustness. *IIE Transactions*, **40**(3), 162–176.

Li, J. and Jin, J. (2010) Optimal sensor allocation by integrating causal models and set-covering algorithms. *IIE Transactions*, **42**(8), 564–576.

Li, J., Jin, J. and Shi, J. (2008) Causation-based T^2 decomposition for multivariate process monitoring and diagnosis. *Journal of Quality Technology*, **40**(1), 46–58.

Li, J. and Meerkov, S.M. (2009) *Production Systems Engineering*, Springer, New York, NY.

Li, J. and Shi, J. (2007) Knowledge discovery from observational data for process control using causal Bayesian networks. *IIE Transactions*, **39**(6), 681–690.

Liu, C., Ding, Y. and Chen, Y. (2005) Optimal coordinate sensor placements for estimating mean and variance components of variation sources. *IIE Transactions*, **37**(9), 877–889.

Lowry, C.A., Woodall, W.H., Champ, C.W. and Rigdon, S.E. (1992) A multivariate EWMA control chart. *Technometrics*, **34**(1), 46–53.

Mason, R., Tracy, N. and Young, J. (1997) A practical approach for interpreting multivariate T^2 control chart signals. *Journal of Quality Technology*, **29**(4), 396–406.

Montgomery, D.C. (1997) *Introduction to Statistical Quality Control*, third edition, John Wiley & Sons, New York, NY.

Pearl, J. and Verma, T.S. (1991) A theory of inferred causation. *Studies in Logic and the Foundations of Mathematics*, **134**, 789–811.

Shewhart, W.A. (1931) *Economic Control of Quality of Manufactured Product*, American Society for Quality, New York, NY.

Shi, J. (2006) *Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*, CRC Press, Boca Raton, FL.

Vapnik, V.N. (1999) *The Nature of Statistical Learning Theory*, second edition, Springer-Verlag, New York, NY.

Verron, S., Tiplica, T. and Kobi, A. (2008) Fault detection and identification with a new feature selection based on mutual information. *Journal of Process Control*, **18**(5), 479–490.

Wang, K. and Jiang, W. (2009) High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology*, **41**(3), 247–258.

Wolbrecht, E., D'Ambrosio, B., Paasch, R. and Kirby, D. (2000) Monitoring and diagnosis of a multistage manufacturing process using Bayesian networks. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **14**(1), 53–67.

Woodall, W.H. and Ncube, M.M. (1985) Multivariate CUSUM quality control procedures. *Technometrics*, **27**(3), 285–292.

Zhou, S., Ding, Y., Chen, Y. and Shi, J. (2003) Diagnosability study of multistage manufacturing processes based on linear mixed-effects models. *Technometrics*, **45**(4), 312–325.

Appendix A: Computational intensity of the BASIS algorithm for different number of nodes

Given a system of m physical variables, the computational complexity of the BASIS algorithm is elaborated in Table A1. The total running time is the summation of the running time for each line executed,

Table A1. Computation complexity analysis for the BASIS algorithm

BASIS algorithm with input $\mathbf{A}^{-1}$, $\mathbf{\Sigma_V}$, α, $ARL_U(\tau)$, m	*Running time for each execution*	*Times of execution (worst case)*
For $i = m$:-1:1	t_1	m
Compute $\tilde{\delta}(i)$	t_2	m
List all searching set with cardinality i	t_3	$2^m - 1$
For $j = 1{:}n_i$	t_4	$2^m - 1$
Compare $\delta(\mathbf{s}_i(j))$ *with* $\tilde{\delta}(i)$	t_5	$2^m - 1$
If $\delta(\mathbf{s}_i(j)) \geq \tilde{\delta}(i)$, *record* $\mathbf{s}_i(j)$ *into set* **Opt**	t_6	$2^m - 1$
If $(flagsol == 0)$, *break*	t_7	m

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014

$t_{total} = t_1 \times m + t_2 \times m + t_3 \times (2^m - 1) + t_4 \times (2^m - 1) + t_5 \times (2^m - 1) + t_6 \times (2^m - 1) + t_7 \times m$, which is $\Theta(2^m)$.

In fact, it is impossible to detect the mean shift that occurs at *leaf* nodes without allocating sensors on them. Therefore, the total number of sensors needed in the system is at least *card*(**LF**) and all of the *leaf* nodes are required to deploy sensors. After taking this fact into consideration, the running time of the proposed algorithm is shortened to be $\Theta\{2^{m-card(\mathbf{LF})}\}$.

Appendix B: Proof of Proposition 1

For any $\tilde{\Delta} \in R_1$, there exists index i, so that $\tilde{\Delta} = a\Delta(i)$, where $|a| \geq 1$. If **s** is a feasible solution from the BASIS algorithm, then $\delta(\mathbf{s}) = (A^{-1}\Delta(i))_{\mathbf{s}}^{\mathrm{T}}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}(\mathbf{A}^{-1}\Delta(i))_{\mathbf{s}} \geq \tilde{\delta}(card(\mathbf{s}))$ based on Definition 2. $\delta(\mathbf{s})$ can be rewritten as $\Delta(i)_{\mathbf{s}}^{\mathrm{T}}(\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{T}}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathbf{T}}]_{\mathbf{s}}\}^{-1}(\mathbf{A}^{-1})_{\mathbf{s}^r}\Delta(i)_{\mathbf{s}}$, where $(\mathbf{A}^{-1})_{\mathbf{s}^r}$ is the matrix $\mathbf{A}^{-1}$ with only rows indexed by **s**.

Thus, it remains to prove that $(\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{T}}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}(\mathbf{A}^{-1})_{\mathbf{s}^r}$ is a positive PSD matrix, so that $(\mathbf{A}^{-1}\tilde{\Delta})_{\mathbf{s}}^{\mathrm{T}}\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}(\mathbf{A}^{-1}\tilde{\Delta})_{\mathbf{s}} \geq \delta(\mathbf{s}) \geq \tilde{\delta}(card(\mathbf{s}))$. Since **A** is a triangular matrix with positive diagonal elements, **A** and $\mathbf{A}^{-1}$ are Positive Definite (PD) matrices. Moreover, $\boldsymbol{\Sigma}_{\mathbf{V}}$ is also PD, since it is a diagonal covariance matrix for **V**. Therefore, $\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}$ is a symmetric PD matrix. Furthermore, $\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1}$ can be expressed by Cholesky decomposition as $\{[\mathbf{A}^{-1}\boldsymbol{\Sigma}_{\mathbf{V}}(\mathbf{A}^{-1})^{\mathrm{T}}]_{\mathbf{s}}\}^{-1} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$, where **L** is a lower triangular matrix with strictly positive diagonal entries. Thus, $(\mathbf{L}^{\mathrm{T}}(\mathbf{A}^{-1})_{\mathbf{s}^r})^{\mathrm{T}}(\mathbf{L}^{\mathrm{T}}(\mathbf{A}^{-1})_{\mathbf{s}^r})$ is a symmetric PSD matrix. As a result, any $\tilde{\Delta} \in R_1$ is detectable if **s** is a feasible solution obtained from the BASIS algorithm. ■

Appendix C: Proof of Proposition 2

Denote the row space of a matrix by $Row(\cdot)$ and the RREF of $(\mathbf{A}^{-1})_{\mathbf{s}^r}$ by $(\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{rref}}$. Note that $(\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{rref}}$ is unique and $Row((\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{rref}}) = Row((\mathbf{A}^{-1})_{\mathbf{s}^r})$. Therefore, if $\exists\Delta_1 \in R_p^{k_i}, \Delta_2 \in R_p^{k_j}$ (where $i \neq j$ and $i, j = 1, 2, \ldots, m$), and $(\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{rref}}\Delta_1 = (\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{rref}}\Delta_2 = \mathbf{b}$, where **b** is a constant vector, then $\exists\tilde{\Delta}_1 \in R_p^{k_i}, \tilde{\Delta}_2 \in R_p^{k_j}$, such that $(\mathbf{A}^{-1})_{\mathbf{s}^r}\tilde{\Delta}_1 = (\mathbf{A}^{-1})_{\mathbf{s}^r}\tilde{\Delta}_2 = \mathbf{b}$. In Proposition 2, if $\boldsymbol{\tau}(i)$ can be expressed as $\boldsymbol{\tau}(i) = a\boldsymbol{\tau}(j)$ for $\forall v(i), v(j) \in \boldsymbol{\theta}[\mathbf{v}]$, then $\exists\Delta_1 \in R_p^{k_i}, \Delta_2 \in R_p^{k_j}$, where $ae_{v(i)}^{\mathrm{T}}\Delta_1 = e_{v(j)}^{\mathrm{T}}\Delta_2$ and $\mathbf{e}_i = [0 \ldots 1 \ldots 0]^{\mathrm{T}}$ with the ith element equal to one, such that $(\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{rref}}\Delta_1 = (\mathbf{A}^{-1})_{\mathbf{s}^r}^{\mathrm{rref}}\Delta_2$. Thus, from Definition 5, $\boldsymbol{\theta}[\mathbf{v}]$ is a minimal diagnosable class. ■

Appendix D: Proof of Proposition 4

It is equivalent to prove that if a sensor allocation solution does not contain all of those nodes that have at most one child, then it is not feasible given that the problem is diagnosable. First, it can be seen that the sensor allocation solution cannot satisfy monitoring requirements if it does not deploy sensors on *leaf* nodes. Second, if there is one variable that has only one child but has no sensor deployed, then the mean shifts from this variable or from its child are aliases of each other. Thus, if the problem is required to be diagnosable, a necessary condition for a sensor allocation solution to be feasible is to allocate sensors on all of those nodes with at most one child. ■

Biographies

Kaibo Liu received a B.S. degree in Industrial Engineering and Engineering Management from the Hong Kong University of Science and Technology, Hong Kong, China, in 2009, and an M.S. degree in Statistics from the Georgia Institute of Technology, Atlanta, in 2011. Currently, he is a Ph.D. student at the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta. His research interests are focused on engineering-driven data fusion for system performance modeling, assessment, and improvement. He is a member of ASME, INFORMS, and IIE.

Jianjun Shi is the Carolyn J. Stewart Chair and Professor at the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology. Before he joined Georgia Tech in 2008, he was the G. Lawton and Louise G. Johnson Professor of Engineering at the University of Michigan. He was awarded his B.S. and M.S. in Electrical Engineering by the Beijing Institute of Technology in 1984 and 1987, respectively, and his Ph.D. in Mechanical Engineering by the University of Michigan in 1992. His research interests focus on the fusion of advanced statistics, signal processing, control theory, and domain knowledge to develop methodologies for modeling, monitoring, diagnosis, and control of complex systems in a data-rich environment. He is the founding chairperson of the Quality, Statistics and Reliability Subdivision at INFORMS. He currently serves as the Focus Issue Editor of *IIE Transactions on Quality and Reliability Engineering*. He is a Fellow of the Institute of Industrial Engineering, a Fellow of the American Society of Mechanical Engineering, and a Fellow of the Institute of Operations Research and Management Science; an elected member of the International Statistical Institute; and a life member of the ASA.

Downloaded by [Georgia Tech Library] at 08:25 24 August 2014