

**OPTIMAL COORDINATE SENSOR PLACEMENTS FOR ESTIMATING
MEAN AND VARIANCE COMPONENTS OF VARIATION SOURCES**

A Thesis

by

QINYAN LIU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2004

Major Subject: Industrial Engineering

**OPTIMAL COORDINATE SENSOR PLACEMENTS FOR ESTIMATING
MEAN AND VARIANCE COMPONENTS OF VARIATION SOURCES**

A Thesis

by

QINYAN LIU

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

Yu Ding
(Chair of Committee)

Amarnath Banerjee
(Member)

Jyhwen Wang
(Member)

Mark Spearman
(Head of Department)

May 2004

Major Subject: Industrial Engineering

ABSTRACT

Optimal Coordinate Sensor Placements for Estimating
Mean and Variance Components of Variation Sources.

(May 2004)

Qinyan Liu, B.S., Tsinghua University;

M.S., Tsinghua University

Chair of Advisory Committee: Dr. Yu Ding

In-process Optical Coordinate Measuring Machine (OCMM) offers the potential of diagnosing in a timely manner variation sources that are responsible for product quality defects. Such a sensor system can help manufacturers improve product quality and reduce process downtime. Effective use of sensory data in diagnosing variation sources depends on the optimal design of a sensor system, which is often known as the problem of sensor placements. This thesis addresses coordinate sensor placement in diagnosing dimensional variation sources in assembly processes. Sensitivity indices of detecting process mean and variance components are defined as the design criteria and are derived in terms of process layout and sensor deployment information. Exchange algorithms, originally developed in the research of optimal experiment design, are employed and revised to maximize the detection sensitivity. A sort-and-cut procedure is used, which remarkably improve the algorithm efficiency of the current exchange routine. The resulting optimal sensor layouts and its implications are illustrated in the specific context of a panel assembly process.

ACKNOWLEDGMENTS

I would like to express my very sincere gratitude to Dr. Yu Ding, the chairman of my advisory committee, for his guidance, support, and advice during my research work. This work could not have been completed without his generosity and patience as well as his profound knowledge. I would also like to express my appreciation to Dr. Amarnath Banerjee and Dr. Jyhwen Wang for serving on my committee.

I would like to show my deepest appreciation to my parents, Mr. Fangyuan Liu and Mrs. Hengying Wu, for their endless love and support during my study.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	vi
LIST OF TABLES.....	vii
CHAPTER	
I INTRODUCTION	1
II FORMULATION FOR ROOT CAUSE DIAGNOSIS.....	7
III DESIGN CRITERIA FOR ROOT CAUSE DIAGNOSIS	12
III.1 Diagnosability and Sensitivity Indices	12
III.2 Formulations for Sensor System Optimization	19
IV EXCHANGE ALGORITHM AND ITS REVISION	23
IV.1 Exchange Algorithms from Optimal Experimental Design	23
IV.2 Fast Exchange Algorithms with Sort-and-Cut	25
IV.3 Implementation and Comparison	29
V OPTIMAL SENSOR LAYOUT AND DISCUSSION	33
VI SUMMARY	37
REFERENCES	39
APPENDIX	45
VITA.....	46

LIST OF FIGURES

FIGURE		Page
1	Mechanism of CMM and OCMM.....	2
2	Illustrative example: variation sources in panel assembly processes	8
3	The computation time versus the number of sensors	32
4	Optimal sensor layouts	36

LIST OF TABLES

TABLE	Page
1 Comparisons of the resulting algorithms.....	30

CHAPTER I

INTRODUCTION

Dimensional integrity is a major quality concern in many discrete-part manufacturing processes. In the automotive and aviation industries, for instance, dimensional problems contribute to about two-thirds of the total quality problems during a new product launch (Shalon et al. 1992, Ceglarek and Shi 1995).

Coordinate Measuring Machine (CMM) is widely used in discrete-part industries to ensure the dimensional quality of a manufactured product. The mechanism of a CMM is illustrated in Fig 1(a). A CMM usually consists of a spatial frame that provides the coordinate reference (not shown in the figure), a mechanical arm that can move along guided tracks, and a probe that retrieves the coordinate information when its tip touches the surface of a manufactured work piece. A CMM has the limitation of being low throughput. A CMM with a single mechanical arm and touch probe has to perform the measurement job sequentially; it will then take hours to finish all measurements on a complicated product. For instance, CMM can measure only 6-8 automotive bodies per day in an automotive body shop with a daily throughput of 1,000 units. Meanwhile, high manufacturing cost of a CMM prohibits using multiple CMMs to perform measurement jobs in parallel.

Recent advancement in sensor technology introduces the optical CMM (known as OCMM). An OCMM replaces the mechanical arm and the touch probe in a CMM

This thesis follows the style of *IIE Transactions*.

with an optical sensor unit (Fig 1(b)), which consists of a laser source and two CCD (Charged-Couple Device) image sensors. The laser source will shed a beam on the surface of a work piece and the CCD sensors will detect the reflective laser beam. The sensor unit is also installed within a spatial frame and will calculate the coordinate of the measured point relative to the frame reference using triangulation.

The OCMM sensor unit is much less expensive than a touch probe. It is then much affordable to deploy multiple optical sensor units, performing parallel measurement jobs of multiple product characteristics. An OCMM station with multiple sensor units are capable of measuring as many as 150 product features on a car body within one minute (refer to Fig 1(c)). The high throughput capability enables OCMM to be built into the production process and obtain 100% inspection of dimensional quality characteristics (Apley and Shi 2001).

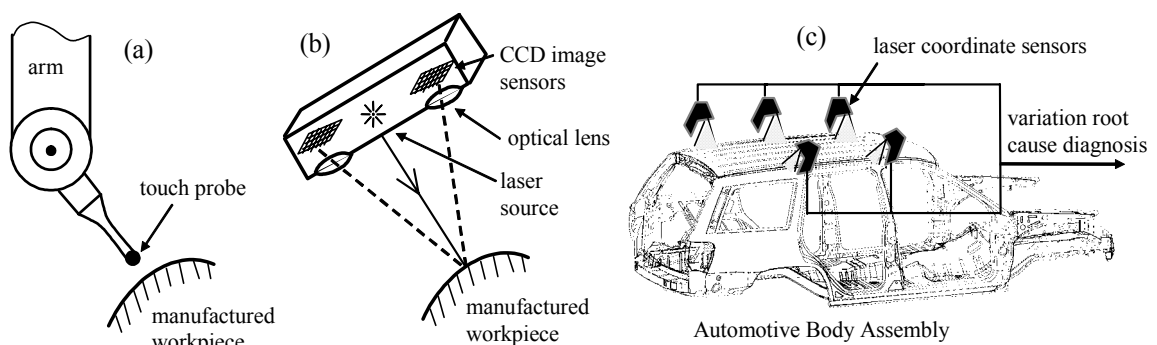


Figure 1. Mechanism of CMM and OCMM

Deployment of OCMMs results in a shift in quality control philosophy. With CMMs, dimensional measurements are taken offline and sampled from a large product population. In such a manner, CMMs are used to inspect the key dimensional product

features and ensure them to be statistically acceptable. Now the implementation of in-process OCMMs leads to continuous dimensional measurements of every manufactured product and provides the potential of determining the underlying process variation sources that are responsible for product defects; this process is known as “root cause diagnosis.” Root cause diagnosis is critical because the identification of variation sources will lead to corrective actions, restoring the manufacturing system to its normal condition in a timely manner.

Recent research efforts have advanced the state of the art of root cause diagnosis in complicated manufacturing systems -- one focused application area is the automotive assembly process. Statistical methods employed there include the estimation methods (Apley and Shi 1998, Lawless et al. 1999, Carlson et al. 2000), pattern recognition using principal components (Hu and Wu 1992, Ceglarek and Shi 1996, Ding et al. 2002a), and factor analysis (Apley and Shi 2001; Apley and Lee 2003).

Apparently, effective use of product measurements in root cause diagnosis depends to a great extent on the design of a sensor system. A poorly designed sensor system may not be able to provide the desired diagnosability or sensitivity in identifying variation sources. Design of a sensor system is realized through the design or determination of (i) an individual sensor unit; (ii) the number of sensors needed; (iii) sensor locations; and (iv) operation strategies such as how many and how often measurements will be taken. In this thesis, we assume the use of the commercially

available optical sensor unit as described in Fig 1(b) and will not get into the physical details of how a sensor unit is designed. Meanwhile, the high measurement throughput of the chosen optical sensor units makes their operations simple because in-process data are automatically collected from every product. Then, the sensor system design considered in this thesis boils down to the determination of the number and locations of coordinate sensors; it is usually called the problem of “sensor placement” in engineering practice. Please note that the “location of a sensor” in this thesis refers to the location of a product feature that a sensor measures instead of the place where a sensor is physically installed. For this reason, coordinate sensor placement is equivalent to the selection (of the number and locations) of dimensional measurement features on a product.

Relevant research on the selection of dimensional measurement features can be classified into two categories: inspection-oriented and diagnosis-oriented. Inspection-oriented feature selection is mainly based on the study of key product characteristics (KPCs) because the purpose of inspection is to make sure KPCs meet their designed tolerances. KPCs may be decided through an empirical analysis of the product/process flow (Ceglarek et al. 1994, Soderberg and Carlson 1999) or through a more quantitative sensitivity-based design evaluation (Whitney et al. 1994, Thornton 1999, Ding et al. 2002b).

Product inspection is to perform a statistical inference on measured product features alone. Root cause diagnosis, on the other hand, is to make inference on

process variation sources that are correlated to product measurements. For this reason, diagnosis-oriented measurement feature selection aims at choosing the features that could lead to certain optimal condition (e.g., maximum separation) for identifying variation sources. In so doing, certain criterion to characterize the distinction between variation sources should be defined and then an optimization routine will be employed to optimize the chosen criterion. Prior research for this objective includes the papers by Khan et al. (1998), Khan et al. (1999), Wang and Nagarkar (1999), and Khan and Ceglarek (2000). The methods developed by Khan and his colleagues are to maximize the minimum distance between variation patterns that are computed as the eigenvectors of a measurement covariance matrix. However, due to the fact that their diagnostic procedure assumes the occurring of a single variation source at a time, their sensor placement strategy only ensures each single variation source to be optimally distinguished from the others. Another limitation is that their strategy is based on the specific way of how the variation patterns are defined and constructed. Their results may no longer be optimal if a different type of variation pattern is defined and used. Wang and Nagarkar (1999) developed a sensor placement strategy for a more generic situation when multiple simultaneous variation sources may exist. A D-optimal criterion is used, the same as that in the optimal experimental design (Fedorov 1972, Atkinson and Donev 1992), and Powell's direct search (Powell 1992) is employed to find the optimal sensor placement.

Different from Wang and Nagarkar (1999), this thesis will investigate the use of E-optimality instead of D-optimality as the design criterion for sensor placements. The

E-optimality characterizes the minimal sensitivity of a coordinate sensor system in detecting the mean and variance components of variation sources. Optimization of such a sensitivity criterion is equivalent to ensuring a maximum separation of variation sources. This thesis will develop an inequality relationship so that a unified criterion can be found for both mean-detection and variance-detection sensitivities. Exchange algorithms, initially developed for optimal experimental designs, were employed in optimizing the chosen sensitivity criterion. Following an idea proposed by Lam et al. (2002) in a molecule selection application, we devise a sort-and-cut procedure to address specific problems relevant to sensor system design, which considerably improves the algorithm efficiency.

This thesis is structured as follows. In Chapter II, we use an automotive assembly process as a case in point to explain the procedure for root cause diagnosis and present the linear diagnostic model that links the product dimensional measurements to the process variation sources. Design criterion, i.e., the mean and variance detection sensitivity and their relationship, is discussed in Chapter III. The optimization algorithm and the suggested revision are presented in Chapter IV. We discuss the resulting optimal sensor layouts as well as a few practical issues in Chapter V. Finally, we conclude the thesis in Chapter VI.

CHAPTER II

FORMULATION FOR ROOT CAUSE DIAGNOSIS

In this chapter, we consider the dimensional measurements taken in auto-body assembly processes with the presence of fixture-related process variation sources.

Figure 2(a) shows the side panel assembly of the Sport Utility Vehicle (SUV) in Fig 1(c). This side panel assembly is made of four components: A-pillar panel, B-pillar panel, rail roof side panel, and rear quarter panel. We simplify each panel component in the assembly as a 2-dimensional polygonal work piece (Fig 2(a)). In a 2D panel assembly process, each work piece is usually held by a set of fixture during assembly operation, illustrated in Fig 2(b). A set of fixture constitutes a 4-way locator (P_1) that controls motion in both x and z directions ($\delta P_1(x, z)$), and a 2-way locator (P_2) that controls motion only in z direction ($\delta P_2(z)$). The fixture location for the whole assembly is indicated in Fig 2(a) as $P_1 - P_8$. Optic coordinate sensors are used to monitor the dimensional integrity of the eventual assembly.

The work piece could have small positional perturbation if all locator-contacts function properly within their designed tolerances. If there is damage on the pinholes or wear on the locators, the work piece will then undergo large random deviation from its nominal position and thus cause a mean shift or excessive dimensional variability on the final assembly. In Fig 2(b), for instance, sensors M_1 to M_3 could detect a mean shift or excessive panel variability caused by a large z-direction deviation $\delta P_2(z)$. In

this case, the process variation source is the low locator positioning repeatability. After detecting a large mean shift and variability in product, we would like to identify the malfunctioning fixture locator responsible for the dimensional quality defect.

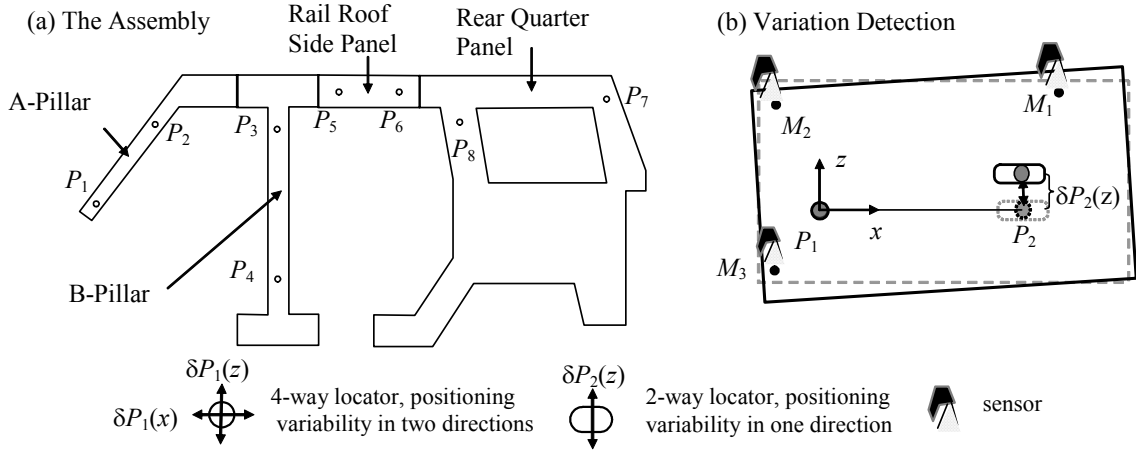


Figure 2. Illustrative example: variation sources in panel assembly processes

In order to perform root cause diagnosis, the first step is to establish a diagnostic model that links the product measurements to the process variance sources. We here use the simple example in Fig 2(b) to get a flavor of how the model is developed.

Denote by \mathbf{u} the deviations at fixture points such as P_1 and P_2 and by \mathbf{y} the deviations measured by coordinate sensors. In Fig 2(b), $\mathbf{u} = [\delta P_1(x) \quad \delta P_1(z) \quad \delta P_2(z)]^T$, where δ is perturbation operator. When a coordinate sensor measures the position of a point, it generally returns three coordinate values of that point in space. In the 2D assembly process discussed above, only two of the three coordinates, i.e., those of x and z directions, are of interests to us, and thus, the product measurements are $\mathbf{y} = [\delta M_1(x) \quad \delta M_1(z) \quad \delta M_2(x) \quad \delta M_2(z) \quad \delta M_3(x) \quad \delta M_3(z)]^T$. Meanwhile, whenever

one more sensor is installed, two more measurements will be augmented to the vector of \mathbf{y} in the 2D assembly process.

For dimensional control, the relationship between \mathbf{y} and \mathbf{u} can be obtained by a standard kinematics analysis (Paul 1981). In quality control, the deviations of product/process features are usually much smaller than nominal values. Hence, a linear model structure or a linearization of originally nonlinear systems is often acceptable for discrete-part manufacturing processes. The linear diagnostic model can be generally expressed as

$$\mathbf{y} = \mathbf{A}\mathbf{u} + \mathbf{v}, \quad (1)$$

where \mathbf{v} is the additive sensor noise. Matrix \mathbf{A} constitutes multiple row blocks, i.e.,

$\mathbf{A}^T = [\mathbf{A}_1^T \quad \dots \quad \mathbf{A}_s^T]$ with the i^{th} block as

$$\mathbf{A}_i = \begin{bmatrix} 1 & \frac{P_1(z) - M_i(z)}{P_2(x) - P_1(x)} & \frac{M_i(z) - P_1(z)}{P_2(x) - P_1(x)} \\ 0 & \frac{M_i(x) - P_1(x)}{P_2(x) - P_1(x)} & \frac{P_2(x) - M_i(x)}{P_2(x) - P_1(x)} \end{bmatrix}, \quad i=1, \dots, s, \quad (2)$$

where s is the number of sensor, $P_1(x)$, $M_i(z)$ etc are the nominal coordinates of the locators and the sensors, respectively. The number of variation sources (i.e., the dimension of \mathbf{u}) is denoted by p and the number of measurements (i.e., the dimension of \mathbf{y}) is denoted by n . In this particular process $p=3$ and $n=2s$ that is the product of the number of sensors and the measurements retrieved per sensor. For a real assembly in Fig 2(a), when more parts with complicated shapes and different orientations are involved, the \mathbf{A} matrix will look a lot messier. However, it still follows the same

spirit as we did for the simple case to develop a linear diagnostic model. For a detailed derivation of relevant kinematic models, please refer to Jin and Shi (1999), Mantripragada and Whitney (1999), Ding et al. (2000) for assembly processes or Djurdjanovic and Ni (2000) and Zhou et al. (2003b) for machining processes.

In root cause diagnosis, one will make inferences about \mathbf{u} based on a sample of measurements of \mathbf{y} . The following assumptions are usually made for this kind of problem: (A1) The p variation sources are independent so that \mathbf{u} has a diagonal variance-covariance matrix; (A2) Sensor noise \mathbf{v} is independent of \mathbf{u} . It is of zero-mean and has the variance-covariance matrix $\sigma_v^2 \mathbf{I}_n$, where \mathbf{I}_n is an $n \times n$ identity matrix and σ_v^2 is the variance of sensor noise. The sensor system satisfying (A2) is known as a *homogeneous* sensor system. A coordinate sensor system using sensor units from the same manufacturers and in the same stage of its service life qualifies as a homogeneous sensor system. On the other hand, a sensor system constituted by different types of sensors or the same type of sensors but with different accuracies is a *heterogeneous* sensor system. In this thesis, we limit our discussion to the design of a homogenous sensor system.

Model (1) appears to be similar to a linear regression model, e.g., $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The differences are as follows. In regression, \mathbf{X} is a data matrix containing measurements of predictors, oftentimes determined through a designed experiment. In model (1), \mathbf{A} is not a data matrix of predictors. Instead it is determined by system design parameters such as locator and sensor locations. The design determining \mathbf{A} is not to design an

experiment but an engineering system. The \mathbf{A} matrix is called *system matrix* in engineering system designs. Meanwhile, \mathbf{u} is not the vector of parameters but a vector of unknown *random* inputs. By contrast, $\boldsymbol{\beta}$ in the above regression model is often a *constant* unknown vector equivalent to the fixed effect in statistical inference. In fact, our model (1) fits better to a mixed linear model with both fixed effect and random effect (please refer to McCulloch and Searle 2000 for mixed linear models). If we write model (1) as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\mu} + \mathbf{A}\tilde{\mathbf{u}} + \mathbf{v}, \quad (3)$$

where $\boldsymbol{\mu}$ is the mean vector of \mathbf{u} and $\tilde{\mathbf{u}}$ is its zero-mean random part, then $\boldsymbol{\mu}$ corresponds to the fixed effects and $\tilde{\mathbf{u}}$ corresponds to the random effects. For root cause diagnosis, one needs to detect the mean components of $\boldsymbol{\mu} \equiv [\mu_1 \ \cdots \ \mu_p]^T$ and the variance components of $\tilde{\mathbf{u}} \ \boldsymbol{\theta} \equiv [\sigma_1^2 \ \sigma_2^2 \ \cdots \ \sigma_p^2]^T$. Please note that we will assume in this thesis that the knowledge about sensor noise variance, σ_v^2 , is available from the sensor vendor's calibration and specification.

CHAPTER III

DESIGN CRITERIA FOR ROOT CAUSE DIAGNOSIS

III.1 Diagnosability and Sensitivity Indices

In the theory of mixed linear models, Rao and Kleffe (1988) defined parameter *identifiability* as to verify whether or not the differences between input parameters lead to differences of the corresponding distributions of the observations \mathbf{y} . Following this concept and also noting that the distribution of \mathbf{y} is usually specified up to the first and second order moments in many engineering applications, Zhou et al. (2003a) defined the *diagnosability* for mean and variance components as: a linear parametric function $\mathbf{p}^T \boldsymbol{\mu}$ is said to be mean-diagnosable if

$$\mathbf{p}^T \boldsymbol{\mu}_1 \neq \mathbf{p}^T \boldsymbol{\mu}_2 \Rightarrow \mathbf{m}_y |_{\boldsymbol{\mu}=\boldsymbol{\mu}_1} \neq \mathbf{m}_y |_{\boldsymbol{\mu}=\boldsymbol{\mu}_2}, \forall \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \quad (4)$$

and a linear parametric function $\mathbf{f}^T \boldsymbol{\theta}$ is said to be variance-diagnosable if

$$\mathbf{f}^T \boldsymbol{\theta}_1 \neq \mathbf{f}^T \boldsymbol{\theta}_2 \Rightarrow \text{vec}(\boldsymbol{\Sigma}_y) |_{\boldsymbol{\theta}=\boldsymbol{\theta}_1} \neq \text{vec}(\boldsymbol{\Sigma}_y) |_{\boldsymbol{\theta}=\boldsymbol{\theta}_2}, \forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \quad (5)$$

where \mathbf{m}_y and $\boldsymbol{\Sigma}_y$ are the mean vector and covariance matrix of \mathbf{y} , respectively;

$\text{vec}(\cdot)$ is an operator to stack the columns of a matrix on top of one another, e.g.,

$$\text{vec}(\boldsymbol{\Sigma}) = [\sigma_{11} \quad \sigma_{21} \quad \sigma_{12} \quad \sigma_{22}]^T \text{ for a } 2 \times 2 \boldsymbol{\Sigma}.$$

The above definition characterizes whether or not the sensor system provides enough information to ensure that the mean and variance components of variation

sources can be separated. We can view it as -- if a variation source is diagnosable, no matter how small a change it undergoes, we can theoretically find an algorithm to estimate it provided that we have large enough amount of samples. If a sensor system does not ensure the diagnosability, no matter how large a variation source changes, we cannot uniquely pinpoint which variation source undergoes the change.

Diagnosability can be used as a design criterion for sensor system design. In fact, a sensor placement strategy leading to a full diagnosability was studied by Ding et al. (2003). On the other hand, however, the diagnosability condition does not make any distinction among diagnosable systems even though some sensor systems may have a superior performance than the others in the sense that it could easily detect a small change in the variation sources. This difference of detection capability is characterized by the concept of “sensitivity,” which may be interpreted as follows: a sensor system that has a zero sensitivity to any one of the variation sources provides no diagnosability, whereas a sensor system with a non-zero sensitivity to all variation sources possesses certain level of diagnosability. It is desirable that a sensor system not only has a full diagnosability but also is sensitive to the underlying changes of variation sources. This thesis will go beyond diagnosability, aiming to achieve a maximum separation of variation sources via the maximization of sensitivity indices defined below.

Based on model (3), we can have

$$\mathbf{m}_y = \mathbf{A} \cdot \boldsymbol{\mu} \tag{6}$$

$$\text{vec}(\Sigma_y) = \pi(\mathbf{A}) \cdot \boldsymbol{\theta} + \sigma_v^2 \cdot \text{vec}(\mathbf{I}_n), \quad (7)$$

where $\pi(\cdot)$ is a matrix transform defined as

$$\pi(\mathbf{A}) = \left[(\mathbf{a}^1 * \mathbf{a}^1)^T \quad \dots \quad (\mathbf{a}^1 * \mathbf{a}^n)^T \quad \vdots \quad \dots \quad \vdots \quad (\mathbf{a}^n * \mathbf{a}^1)^T \quad \dots \quad (\mathbf{a}^n * \mathbf{a}^n)^T \right]^T, \quad (8)$$

and \mathbf{a}^j is the j^{th} row vector of \mathbf{A} and $*$ represents the Hadamard product (Schott 1997).

Following the same spirit in defining diagnosability, the sensitivity for detecting changes in mean and variance components can be defined as the ratio of change in the mean or variance of \mathbf{y} over a perturbation of the mean or variance of the input sources. We define the detecting sensitivity of mean and variance components as follows.

Definition 1. Given measurement \mathbf{y} , the mean-detecting sensitivity, denoted as S_m , is defined as

$$S_m \equiv \min_{\delta \boldsymbol{\mu} \neq \mathbf{0}} \frac{(\delta \mathbf{m}_y)^T (\delta \mathbf{m}_y)}{(\delta \boldsymbol{\mu})^T (\delta \boldsymbol{\mu})}, \quad (9)$$

and the variance-detecting sensitivity, denoted as S_v , is defined as

$$S_v \equiv \min_{\delta \boldsymbol{\theta} \neq \mathbf{0}} \frac{\text{tr}(\delta \tilde{\Sigma}_y \cdot \delta \tilde{\Sigma}_y)}{(\delta \boldsymbol{\theta})^T (\delta \boldsymbol{\theta})}, \quad (10)$$

where $\tilde{\Sigma}_y$ is the covariance matrix contributed from the process variation sources, i.e., $\text{vec}(\tilde{\Sigma}_y) = \pi(\mathbf{A}) \cdot \boldsymbol{\theta}$.

Given the linear relation in equations (6) and (7) and utilizing the eigenvalue

property of a symmetric matrix (Schott 1997, pp. 105), we can express the above-defined sensitivity indices in terms of the eigenvalue of $\mathbf{A}^T \mathbf{A}$ as follows (the proof is fairly straightforward and is thus omitted):

$$S_m = \lambda_{\min}(\mathbf{A}^T \mathbf{A}) \text{ and } S_v = \lambda_{\min}(\pi(\mathbf{A})^T \pi(\mathbf{A})), \quad (11)$$

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix. In deriving S_v , the relation that $tr(\delta\tilde{\Sigma}_y \cdot \delta\tilde{\Sigma}_y) = \text{vec}(\delta\tilde{\Sigma}_y)^T \cdot \text{vec}(\delta\tilde{\Sigma}_y)$ is used.

Remarks:

1. The squared summation of elements in input/output vectors is used in the above definition so that we can have a scalar sensitivity index that is easy to interpret. The squared summations are equivalent to the Euclidean norm of the corresponding vector/matrix; $tr(\delta\tilde{\Sigma}_y \cdot \delta\tilde{\Sigma}_y)$ is the Euclidean norm of matrix $\delta\tilde{\Sigma}_y$.
2. In the variance sensitivity definition, we use $\tilde{\Sigma}_y$ rather than Σ_y because it is assumed that sensor noise variance σ_v^2 is known.
3. Without the minimum, the ratios in equations (9) and (10) are input dependent. Using input-dependent indices, we will have to design a sensor system for individual changes of input variation source and it would then be inconvenient. The minimum operator defines the sensitivity indices to be the smallest ratios given all possible combinations of input changes. Equation (11) shows that the above defined sensitivity indices are in fact input independent; they are solely

determined by the system matrix \mathbf{A} .

4. The above definition is also consistent with the intuitive relation between sensitivity and diagnosability, about which we expressed before the definition. The diagnosability conditions obtained in Zhou et al. (2003a) are: the mean components are diagnosable if $\mathbf{A}^T\mathbf{A}$ is of full rank and the variance components are diagnosable if the matrix $\{(\mathbf{a}_i^T\mathbf{a}_j)^2\}_{i,j=1}^p$ is of full rank, where \mathbf{a}_i is the i^{th} column vector of \mathbf{A} and $\{ \}_{i,j=1}^p$ is a $p \times p$ matrix. It can be shown that $\{(\mathbf{a}_i^T\mathbf{a}_j)^2\}_{i,j=1}^p = \pi(\mathbf{A})^T\pi(\mathbf{A})$; the proof is included in Appendix I. It is then apparent that a full diagnosability is guaranteed if the corresponding sensitivity index is nonzero and a system with a zero sensitivity is equivalent to the one that is not fully diagnosable.

Since S_m and S_v are different functions of \mathbf{A} , a sensor system design may end up with different results, depending on which one of the objectives is chosen, either achieving the maximum mean detection sensitivity or the maximum variance detection sensitivity. However, further investigation found an inequality relationship between S_m and S_v ; S_m^2 is a lower bound for S_v for the same \mathbf{A} . The result is stated in Lemma 1 and its proof can be found in Appendix II.

Lemma 1. For the same \mathbf{A} , $S_v \geq (S_m)^2$

There is another angle to view the relationship between S_v and S_m . For equation (7), if the variance components in $\boldsymbol{\theta}$ are estimated using a maximum likelihood

estimator (MLE), the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is approximated by the inverse of Fisher Information Matrix as

$$\text{cov}(\hat{\boldsymbol{\theta}}) \propto [\{\text{tr}(\boldsymbol{\Sigma}_y^{-1}(\mathbf{a}_i \mathbf{a}_i^T) \boldsymbol{\Sigma}_y^{-1}(\mathbf{a}_j \mathbf{a}_j^T))\}_{i,j}^p]^{-1}. \quad (12)$$

Some constant is omitted from the right hand side in (12), thus we used “ \propto ” instead of “ $=$ ”. This expression suggests that $\text{cov}(\hat{\boldsymbol{\theta}})$ depends on the true value of $\boldsymbol{\theta}$ because $\text{vec}(\boldsymbol{\Sigma}_y) = \boldsymbol{\pi}(\mathbf{A}) \cdot \boldsymbol{\theta} + \sigma_v^2 \cdot \text{vec}(\mathbf{I}_n)$. Under a normal process condition when there is no outstanding variation sources, we can assume that $\boldsymbol{\theta} = \mathbf{0}$ and then $\boldsymbol{\Sigma}_y = \sigma_v^2 \mathbf{I}_n$. As such, equation (12) becomes

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\theta}}) |_{\boldsymbol{\theta}=\mathbf{0}} &\propto \sigma_v^4 \cdot [\{\text{tr}(\mathbf{a}_i \mathbf{a}_i^T \mathbf{a}_j \mathbf{a}_j^T)\}_{i,j}^p]^{-1} \\ &= \sigma_v^4 \cdot [\{(\mathbf{a}_i^T \mathbf{a}_j)^2\}_{i,j}^p]^{-1} = \sigma_v^4 \cdot [\boldsymbol{\pi}(\mathbf{A})^T \boldsymbol{\pi}(\mathbf{A})]^{-1}. \end{aligned} \quad (13)$$

Thus, the variance of linear parametric function $\mathbf{f}^T \hat{\boldsymbol{\theta}}$ under a normal process condition is

$$\text{cov}(\mathbf{f}^T \hat{\boldsymbol{\theta}}) |_{\boldsymbol{\theta}=\mathbf{0}} \propto \sigma_v^4 \cdot \mathbf{f}^T [\boldsymbol{\pi}(\mathbf{A})^T \boldsymbol{\pi}(\mathbf{A})]^{-1} \mathbf{f}. \quad (14)$$

Then, the maximum variance of $\mathbf{f}^T \hat{\boldsymbol{\theta}}$ for any unit vector \mathbf{f} is the maximum eigenvalue of $[\boldsymbol{\pi}(\mathbf{A})^T \boldsymbol{\pi}(\mathbf{A})]^{-1}$. In other words, S_v , the smallest eigenvalue of $[\boldsymbol{\pi}(\mathbf{A})^T \boldsymbol{\pi}(\mathbf{A})]$, represents the maximum variance of $\mathbf{f}^T \hat{\boldsymbol{\theta}}$, $\forall \|\mathbf{f}\|=1$, under a normal process condition. The criterion to maximize S_v is then equivalent to selecting an \mathbf{A} to minimize the maximum variance of $\mathbf{f}^T \hat{\boldsymbol{\theta}}$ under a normal process condition. Similarly, it is not difficult to show that maximizing S_m is equivalent to minimizing the maximum

variance of $\mathbf{p}^T \hat{\boldsymbol{\mu}}$, $\forall \|\mathbf{p}\|=1$, under a normal process condition.

On the other hand, given model (1), we can estimate \mathbf{u} by $\hat{\mathbf{u}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$ and use $\hat{\mathbf{u}}$ as if it was directly observed. Then, the variance-covariance matrix of $\hat{\mathbf{u}}$ is

$$\text{cov}(\hat{\mathbf{u}}) = \mathbf{A}^\# (\mathbf{A} \cdot \text{diag}(\boldsymbol{\theta}) \cdot \mathbf{A}^T + \sigma_v^2 \cdot \mathbf{I}_n) (\mathbf{A}^\#)^T. \quad (15)$$

where $\mathbf{A}^\# \equiv (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ and $\text{diag}(\cdot)$ converts a vector into a matrix of which diagonal elements are the vector elements. Under a normal process condition where $\boldsymbol{\theta}$ is assumed $\mathbf{0}$, we have

$$\text{cov}(\hat{\mathbf{u}}) |_{\boldsymbol{\theta}=\mathbf{0}} = \sigma_v^2 \cdot (\mathbf{A}^T \mathbf{A})^{-1}. \quad (16)$$

Equation (16) suggests that maximizing S_m is also equivalent to minimizing the maximum prediction variance of $\hat{y} = \mathbf{a}^T \hat{\mathbf{u}}$, $\forall \|\mathbf{a}\|=1$, under a normal process condition.

Equation (16) and Lemma 1 indicate that minimizing the maximum prediction variance of $\hat{y} = \mathbf{a}^T \hat{\mathbf{u}}$ is the same as to minimize the maximum variance of $\mathbf{p}^T \hat{\boldsymbol{\mu}}$ and simultaneously provides a lower bound for the maximum variance of $\mathbf{f}^T \hat{\boldsymbol{\theta}}$. This is not surprising since $\hat{\mathbf{u}}$ contains richer information than either $\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\theta}}$; it can be used to simultaneously calculate both the mean components in $\hat{\boldsymbol{\mu}}$ and the variance components in $\hat{\boldsymbol{\theta}}$, while, on the other hand, knowing $\hat{\boldsymbol{\mu}}$ or $\hat{\boldsymbol{\theta}}$ does not lead to the reconstruction of $\hat{\mathbf{u}}$.

Based on Lemma 1, we choose to use S_m , i.e., $\lambda_{\min}(\mathbf{A}^T \mathbf{A})$, as the unified criterion

for optimal sensor placement in order to simplify the design process. Nonetheless, the optimization routines presented in the subsequent chapter should be equally applicable to the maximization of S_v .

III.2 Formulations for Sensor System Optimization

With the unified sensitivity index $S_m = \lambda_{\min}(\mathbf{A}^T \mathbf{A})$, the problem of optimal sensor placement will be formulated as follows. The design parameters are the number and locations of sensors, denoted by $\boldsymbol{\varphi}(s) \equiv [X_1 \quad Z_1 \quad \cdots \quad X_s \quad Z_s]^T$. Moreover, certain constraints should be satisfied. One constraint is that a sensor location has to be a point on the product (geometrical constraint), represented by $G(\cdot) > 0$, where $G(\cdot)$ represents the appropriate geometry function of a manufactured product. Meanwhile, in order for OCMM to perform parallel measurements, we exert the second constraint to avoid any possible optical interference among laser beams when taking measurements. Our engineering knowledge indicates that it would be sufficient for enabling parallel measurements if we keep the sensor locations at least 100-mm apart from each other. For a given number of sensors, we try to find the optimal sensor locations that maximize S_m , namely

$$\begin{aligned} \max_{\boldsymbol{\varphi}(s)} S_m &\equiv \lambda_{\min}(\mathbf{A}^T \mathbf{A}) \\ \text{subject to } &G(\boldsymbol{\varphi}(s)) > 0 \quad \text{and} \quad \text{the 100-mm apart rule} \end{aligned} \quad (17)$$

The criterion of maximizing $\lambda_{\min}(\mathbf{A}^T \mathbf{A})$ is the same as the E-optimality in optimal experimental design, initially proposed by Ehrenfeld (1955), when \mathbf{A} matrix is considered as the mathematical equivalence of \mathbf{X} in a regression model. Other

optimality criteria were also proposed in optimal experimental design, such as the D-optimality ($\max \det(\mathbf{A}^T \mathbf{A})$), the A-optimality ($\max \text{tr}(\mathbf{A}^T \mathbf{A})$), where $\text{tr}(\cdot)$ and $\det(\cdot)$ are the trace and the determinant of a matrix, respectively. These three measures are related to each other through eigenvalues of $\mathbf{A}^T \mathbf{A}$ $\{\lambda_i\}_{i=1}^p$:

$$D_{opt} : \det(\mathbf{A}^T \mathbf{A}) = \prod_{i=1}^p \lambda_i ; A_{opt} : \text{tr}(\mathbf{A}^T \mathbf{A}) = \sum_{i=1}^p \lambda_i ; \text{ and } E_{opt} : \lambda_{\min}(\mathbf{A}^T \mathbf{A}). \quad (18)$$

The D-optimality criterion is the most widely used in experimental designs mainly due to its attractive mathematical properties (Fedorov 1972, pp. 138, Atkinson and Donev 1992, pp. 107). It possesses an invariant property under scaling, i.e., optimal experiments can be designed using a group of standardized dimensionless variables instead of the original physical variables.

Those optimality criteria were also used in observer design for control systems (Muller and Weber 1972; Patton 1989) as well as other engineering systems designs (e.g., the D-optimality was used by Wang and Nagarkar (1999) for sensor placement). In the sensor system design, D-optimality may be interpreted as that if an eigenvalue of $\mathbf{A}^T \mathbf{A}$ is considered as the sensitivity index for the canonical model with a diagonalized \mathbf{A} matrix, the D-optimality corresponds to the multiplication of sensitivities for all input/output pairs. Likewise, the A-optimality is the sum of those sensitivities. These two criteria attempt to optimize an aggregated (multiplication or summation) sensitivity when designing a sensor system. By contrast, the E-optimality is more conservative because it optimizes the smallest sensitivity. Our experience indicates that the E-optimality criterion is more easily to be accepted by practitioners.

Meanwhile, the invariant property of the D-optimality may be inapplicable to an engineering system design because engineering system designs are often accompanied by complex constraints, e.g., the geometric constraints imposed by the shape of the rear quarter panel (refer to Fig 2(a)). This complexity makes it almost impossible to design an engineering system based on a group of dimensionless standardized variables.

The optimization in equation (17) does not determine the number of sensors. Noticing that an increase in sensor number will generally result in a larger maximum of S_m , people usually try to determine the appropriate sensor number by trading off between the benefit gained from an increase in S_m and the cost for more sensors. However, in engineering practice, it is not easy to quantify the monetary saving associated with an increase in S_m . It is thus difficult to define an accurate cost function to attain this trade-off. Alternatively, we can specify a lower bound for S_m or equivalently, an upper bound for the maximum prediction variance based on (16). Then, the second optimization formulation is to minimize the sensor number, while satisfying a lower bound constraint on S_m in addition to other constraints previously specified, i.e.,

$$\begin{aligned} \min_{\boldsymbol{\varphi}} \quad & s \\ \text{subject to} \quad & S_m \geq c, G(\boldsymbol{\varphi}(s)) > 0, \text{ and the 100 – mm apart rule} \end{aligned} \tag{19}$$

where c is the lower bound, decided based on engineering requirements.

In the next chapter, we will mainly study the optimization in (17), which is

equivalent to the “exact” design problem in optimal experimental designs. Optimization (19) can be solved using the resulting exact design algorithm with a gradually increasing sensor number. In Chapter V, we will briefly discuss other considerations in solving (19) as well as how to select constant c .

CHAPTER IV

EXCHANGE ALGORITHM AND ITS REVISION

IV.1 Exchange Algorithms from Optimal Experimental Design

The optimization problems formulated in Chapter III are nonlinear in design parameters $\boldsymbol{\phi}$. Standard nonlinear programming approaches (such as quadratic programming) are usually based on a derivative calculation and they will be easily entrapped in a local optimum. The derivative-based approaches will be especially ineffective for a non-convex design space, imposed by the geometry of panels involved, not to mention those design spaces in the assembly that are not simply connected, for instance, the rear quarter panel, since the window-opening area is not a candidate area for sensor placement.

In the research of optimal experimental design, exchange algorithms were developed for optimizing those aforementioned design criteria, such as D-, A-, and E-optimality; please see Cook and Nachtsheim (1980) and Atkinson and Dover (1992) for reviews and comparisons of exchange algorithms. According to Meyer and Nachtsheim (1995), exchange algorithms have more freedom to maneuver on a complicated design space because each of its exchanges involves only a part of design parameters (associated with one design point). Then, exchange algorithms could be more effective in escaping a local optimum than the derivative-based nonlinear programming. Additionally, exchange algorithms have other advantages to be applied

to engineering system designs. Their procedures are intuitive and implementation is easy. The algorithms are flexible and can easily handle complicated constraints in engineering system design. It can also be used for a wide variety of design criteria.

To use exchange algorithms, we discretize the continuous design space first. We call the resulting discretized design space with N_c candidate sensor locations as the candidate space (denoted as D_c) and the space with s current sensor locations as the sensor space (denoted as D_s). The basic idea of an exchange algorithm is to start with a set of s design points (i.e. the sensor location) in D_s , usually randomly selected, and exchange the current design points with those points in the much larger D_c in order to improve the chosen design criterion. In exchange algorithms, however, the action of exchange is not carried out for every single point. Every point in D_s will be *tested* against a point in D_c , meaning that the improvements in design criterion is recorded, supposing the point in D_s had been exchanged with a point in D_c . There are different variants to the above basic idea, depending on how often the action of exchange is actually carried out. One option is to perform the exchange action after all points in D_s have been tested against the entire set of points in D_c . It exchanges the pair of points, one in D_s and one in D_c , which made the maximum improvement in design criterion. This option is actually the celebrated Fedorov exchange algorithm. Another option is to perform the exchange action for every point in D_s after that design has been tested against all points in D_c . In other words, point i in D_s will be exchanged with a point in D_c that maximizes the improvement in design criterion and the same action is repeated in a sequential order for $i=1, 2, \dots, s$. The second option is the modified Fedorov

exchange. In combinatorial optimization, these are two extreme cases of a general k -exchange algorithm, with $k=s$ for Fedorov exchange and $k=1$ for modified Fedorov exchange (Aarts and Lenstra 1997).

When applying them to the sensor placement problem, we notice that exchange algorithms, especially the Fedorov algorithm, could consume a great deal of CPU time for the cases with a moderate to large number of sensors (e.g., $s=8$ or larger). This is not surprising because the exchange algorithm was initially developed for experimental design with a relatively small number of factors and experiments (Cook and Nachtsheim 1980). In the subsequent chapter, we will suggest and implement a sort-and-cut procedure that will shorten the computation time without sacrificing much the optimal value it finds. Fedorov algorithm, which provides us a better optimal solution, will serve as a benchmark for comparison.

IV.2 Fast Exchange Algorithms with Sort-and-Cut

Let us first conceptually understand the factors affecting the algorithm's computation time. Define the process to pass over the entire D_c set once as a "loop." There are two major factors affecting the run time: the average number of loops and the size of the candidate space N_c . In order to reduce the computation time, we will have to reduce the average number of loops as well as the size of the candidate design space N_c . The following sort-and-cut procedure is employed to achieve both goals. The basic idea is to perform multiple exchanges in each loop to reduce the average number of loops, and after each loop, discard a subset of the candidate design points to

reduce N_c .

When designing a uniform coverage design in molecule selection, Lam et al. (2002) suggested that instead of exchanging one design point per loop, one may want to exchange multiple candidate points in the upper tail of the distribution of improvements in design criterion among all the candidates. In this way, the number of design points that will be changed during each loop will be more than one so that the average number of loops required to replace all random initial designs can be reduced.

In so doing, we should record the improvement in design criterion that a candidate location can make if the corresponding exchange is indeed carried out. Denote the Δ as the improvement in the S_m criterion, i.e., $\Delta \equiv S_m^{new} - S_m^{old}$. Record all Δ_j 's ($j=1, \dots, N_c$) when we loop through the N_c candidate locations. Sort the value of Δ_j 's in a descending order as $\Delta_{(1)} \geq \Delta_{(2)} \geq \dots$ and so on. The distribution of improvements is approximated by the sorted values $\Delta_{(j)}$. Then, we will set an integer number q so that the first q candidate locations in the upper tail of $\Delta_{(j)}$ will be exchanged in each loop.

On the other side, we can reduce the total number of candidate points N_c . The sorted values of design improvement $\Delta_{(j)}$ actually provide us valuable information about the potential of a candidate location. Those candidate locations with a low Δ value are less likely to be picked up by the exchange algorithm in next iterations. Thus, we could remove a proportion of candidate points after each iteration. Denote by α the proportion of candidate points that will be kept after a cut.

To implement this sort-and-cut procedure, there are two parameters to be determined: q and α . In our sensor system design problem, the sensor number in D_s usually is only a small percentage of that in D_c space (for instance, $D_s=10$ but $D_c = 10,000$). We recommend an aggressive choice of α , e.g., from 10% to 20% (For the above example, if $\alpha=10\%$ for the first two iterations, the remaining locations in D_c are still about 10 times of that in D_s).

For a sort-and-cut procedure to work, the assumption is that the distribution of improvements approximated by data from the previous exchange routine can represent well enough the distribution in next exchange. However, whenever an exchange happens, the distribution cannot be exactly the same since the D_s space generating that distribution is no longer the same. The common ground is actually constituted by the design points that are not exchanged in this iteration. Thus, for the above assumption to hold, q should be smaller than s , namely, $1 \leq q < s$. When q is close to s , however, almost all sensors in D_s will be exchanged in one iteration and the distribution recorded in $\Delta_{(j)}$ from the previous loop does not truly represent the distribution for the new D_s space. A subsequent exchange based on the $\Delta_{(j)}$ could make a poor choice that has to be re-done in the following loops. On the other hand, too small a q will result in too few exchanges per loop and thus miss our original goal of having multiple exchanges to reduce the average number of loops. We therefore recommend selecting $q=s/2$ to strike a balance so that half of the sensors in D_s will not be exchanged to provide a common ground for distribution and half of the sensors will be exchanged to

reduce the average loop number.

Please note that we are different from Lam et al. (2002) that instead of using q as the direct control on the number of exchanges they set $\Delta_{(q)}$ as the threshold to control the exchange, i.e., if there is an improvement greater than $\Delta_{(q)}$, then carry out the exchange. We find using $\Delta_{(q)}$ in our application is not effective. The difference is due to our subsequent cut action, which is not used by Lam et al. (2002). The effect of the cut action, as explained in the above paragraph, requires us to have more direct control over the number of exchanges, which cannot be fulfilled by using $\Delta_{(q)}$.

The algorithm for an s -sensor exact design is summarized as follows.

Step 1. The candidate design space D_c is discretized and s locations are randomly selected to form D_s .

Step 2. In every iteration

(1) Rank the sensor locations in D_c in a descending order according to their Δ_j values;

(2) Cut off those locations with low Δ_j values and keep the top $\alpha \times 100\%$ of candidate points in D_c ;

(3) Exchange the top sensor locations in D_c , which satisfy the constraint condition, with the sensor locations in the current D_s space in a sequential order, starting from the one with the largest Δ_j . Each time, one sensor in D_c is added to D_s to generate a $(s+1)$ design, and then, remove one sensor from the $(s+1)$ design,

which makes the smallest decrease in the sensitivity index, so that the number of sensors in D_s will remain s . Repeat this exchange for q sensors.

Step 3. Repeat Step 2 until the improvement in design criterion for two successive designs is smaller than a pre-determined threshold (we used 0.1%).

IV.3 Implementation and Comparison

Algorithms described in the above chapters are coded in MATLAB and compared on the same computer. Their running times are longer than the cases if they were coded in C or FORTRAN. However, the relative performances should not be different due to the way of coding.

We measure algorithm efficiency by the time that it takes to find the optimal value. Meanwhile, we measure algorithm effectiveness by the average value of optimal solutions (i.e., average S_m) it finds when a group of random sensor layouts is used as the initial design. In the literature of algorithm comparison (e.g., Cook and Nachtsheim 1980), a relative effectiveness R was often used, which is defined as the ratio of the average optimal value over the best optimal solution found by all the algorithms in the comparison under the same setting. We use both measures in this study.

For the SUV side panel assembly in Figure 2(a), we discretize it with candidate points 10 millimeters apart. Our engineering experience indicates that this resolution of discretization is sufficient to generate a fine enough grid on a panel that has a size of over a thousand millimeters. The discretization results in a total of $N_c = 13,304$

candidate positions in D_c . As for the sort-and-cut procedure, we choose $\alpha=0.1$ and $q=s/2$. In this application, we only implement the cut action in the first iteration.

We perform numerical study to compare these algorithms for different choices of s . The minimal s is two because at least two sensors that are not collinear are needed to measure all variation sources associated with the two fixture locators (P_1 and P_8) that support the eventual side panel assembly. In the comparison, 50 trials with randomly generated initial designs are performed; the comparison results are included in Table 1. To save some space, only the results with an even s value are displayed, whereas the understanding can be generally extended to an odd s .

Table 1. Comparisons of the resulting algorithms

		Average Computer Time (seconds)	Average Maximal S_m	R
$s=2$	Fedorov	27.24	1.0044	0.9995
	Modified Fedorov	19.12	0.9975	0.9926
	Fast Exchange	2.47	0.9300	0.9255
$s=4$	Fedorov	98.12	2.0105	0.9980
	Modified Fedorov	50.61	2.0088	0.9972
	Fast Exchange	3.24	1.9794	0.9826
$s=6$	Fedorov	106.50	3.0145	0.9979
	Modified Fedorov	63.50	3.0150	0.9981
	Fast Exchange	4.21	2.9715	0.9837
$s=8$	Fedorov	187.95	4.0177	0.9972
	Modified Fedorov	87.55	4.0188	0.9975
	Fast Exchange	4.29	3.9913	0.9907
$s=10$	Fedorov	376.80	5.0214	0.9978
	Modified Fedorov	142.04	5.0207	0.9976
	Fast Exchange	5.11	4.9965	0.9928

We observed the following: (1) the sort-and-cut procedure indeed considerably improves the algorithm efficiency. The reduction in computation time ranges from 90% (for $s=2$) to 98% (for $s=10$). In fact, the reduction is more noteworthy when the sensor number is relatively large (e.g., $s=8$ or 10), which is desirable for engineering system design problem. We also apply this fast exchange algorithm to large sensor numbers such as $s=20, 30, 40, 50$, and 60 . The computation time versus the sensor number is shown in Fig. 3, where the value indicated below each mark is the average computation time. From this figure, one can find that the average time required to find an optimal 60-sensor design using this fast exchange algorithm is similar to what Fedorov needs for a 2-sensor design. More importantly, a linear trend is demonstrated in the computation time when the sensor number increases, which make this fast exchange algorithm an ideal tool for solving design problems with a large sensor number. (2) We also noticed that the R value increases as s increases. For $s=2$, $R=0.9255$, which is noticeably lower than Fedorov and modified Fedorov. This can be explained by the same reason we used to choose the q value in the previous chapter. When $s=2$, even if only one sensor has been exchanged, the previously recorded Δ_j can hardly represent the distribution for next loop because it is based on one single (usually randomly selected) sensor. When the sensor number increases, this problem is alleviated. Actually, for $s=4$ and onward, R is large enough compared with the other two algorithms, meaning that the fast exchange does not scarify much the optimal value. For $s=2$, since it is a small scale problem similar to those in experimental designs, Fedorov or modified Fedorov can be used directly.

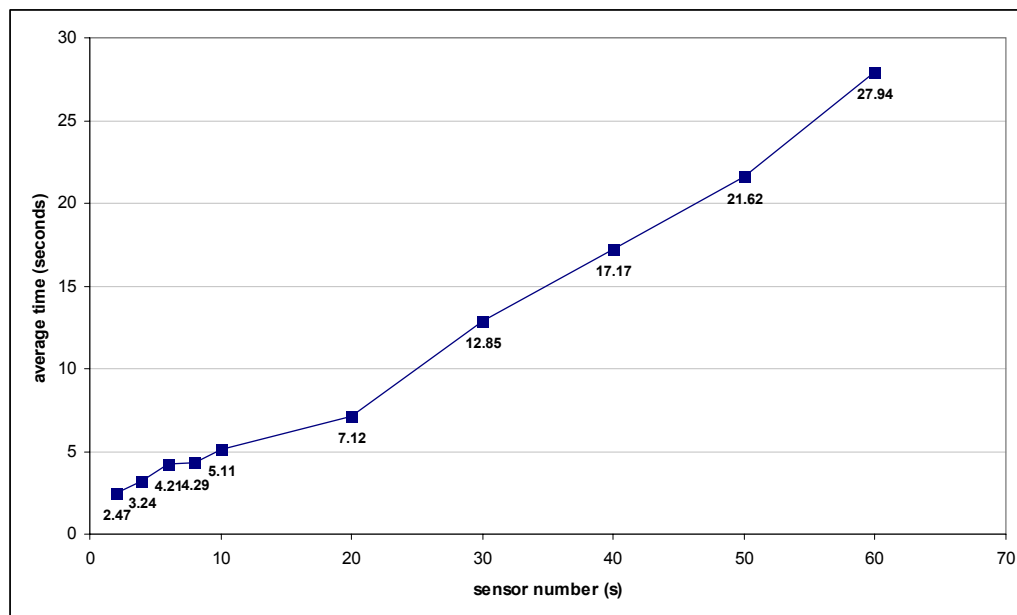


Figure 3. The computation time versus the number of sensors

CHAPTER V

OPTIMAL SENSOR LAYOUT AND DISCUSSION

The resulting optimal sensor layouts with an even sensor number ($s=2, \dots, 10$) are shown in Fig. 4, where a “*” mark indicates a sensor location. From the layouts, we observe that the sensors are located in the area close to the panel boundary and many of them are actually on the edge. It brings up the question if we can reduce our candidate locations by limiting our search to the geometry boundary of each part in the first hand. The answer is yes. However, we should also notice that not all the sensor locations are on the edge (refer to the cases for $s=8$ and $s=10$). Based on empirical knowledge alone, it is nontrivial to determine a search area that contains all the potential good sensor locations. In this study, we used the approximated distribution of design improvements in the sort-and-cut procedure, which provides more reliable information and quantitative evaluation to find the area of the potential good sensor locations. The algorithm is fairly general and it can be used together with a reduced candidate pool to further improve the algorithm efficiency, should the aforementioned intuitive rule was implemented before the search.

The above exact design algorithm is to find an optimal sensor layout when the sensor number is specified. In order to solve for optimization (19), one may wonder if we can use a sequential routine, i.e., we start from one randomly generated point, and then sequentially add one more sensor from D_c to D_s which maximizes the resulting S_m ,

until eventually $S_m \geq c$. For such a sequential strategy to work well, the sensor layout for $(s-1)$ sensors should be a subset of the optimal layout for s sensors. This might be true when sensor number is small ($s < 4$) but it does not stay that way when sensor number is getting larger. Although we did not show the sensor layouts for odd sensor numbers, they actually agree with the phenomenon demonstrated by the displayed layouts in Fig. 4. Therefore, the sequential routine could often miss the optimal layout. Nonetheless, we can combine the sequential probing and the exact design. That is, first use sequential routine to probe and find a sensor number which can yield $S_m \geq c$ and then switch to an exact design routine to find the optimal sensor locations for sensor numbers around the one found by the sequential routine. This way, we can skip a number of time-consuming exact designs, especially when the resulting sensor number is relatively large.

In optimization (19), we specify a constant c to stop the algorithm. Usually the choice of c depends on engineering requirements and is specified under particular context. In this study, we can choose c based on the accuracy requirement. It is known that the OCMM, although more agile and faster, is not as accurate as the mechanical CMM -- the OCMM measurement repeatability is five to ten times lower than the CMM (Hu 1990). Let us be optimistic and consider an OCMM is five times less accurate than an CMM, namely, $\sigma_{v,OCMM}^2 = 5\sigma_{v,CMM}^2$. According to equation (16), under a normal process condition, the maximum variance in estimating process variation sources is $\frac{\sigma_{v,OCMM}^2}{S_m}$.

In order to achieve the same variance level as if a CMM was used to directly measure the process variation source, we hence require that $\sigma_{v,OCMM}^2 / S_m < \sigma_{v,CMM}^2$, which translates into $S_m > 5$. Then, we will choose $c=5$ in this study. Certainly, this value may change when the accuracy requirement is different. But the above logic can still be applied in determining an appropriate c . When choosing $c=5$, we find that 10 sensors will provide an equivalently well sensing capability as a CMM.

One may also wonder what happens if we have used S_v instead of S_m as our design criterion. Examples using S_v are shown in Fig. 4(f)-(h) for $s=2, 6, 10$. Interestingly, the sensor layouts using S_v bears a strong similarity as those using S_m , especially in terms of areas where the sensors will be located. Of course, the resulting layouts using S_v deviate to some extent from those using S_m and the deviation is less obvious for $s=2$ but is getting more appreciable for a larger s .

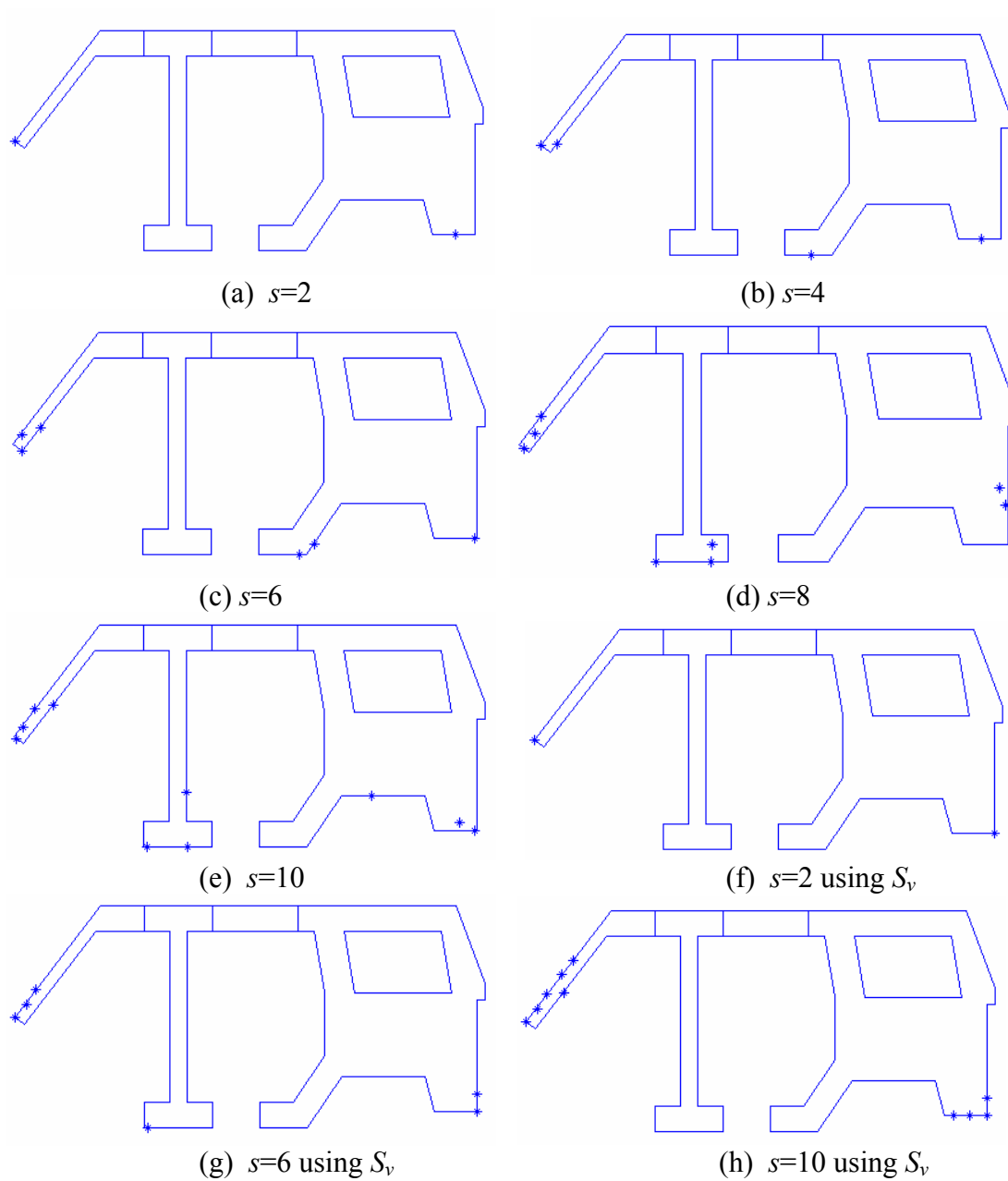


Figure 4. Optimal sensor layouts

CHAPTER VI

SUMMARY

This thesis presents some statistical and optimization methods for coordinate sensor placement for estimating the mean and variance components of variation root causes. Beyond previously defined diagnosability indices, sensitivity indices of a sensor system are defined to characterize the system capability of detecting the underlying process mean and variance changes. Mathematically, they are the same as the E-optimality criterion proposed in optimal experimental design. Optimization of the design criterion is then fulfilled by revising the exchange algorithm. In this application, we devised a fast exchange routine with a sort-and-cut procedure, which considerably reduces the algorithm's computation time while maintaining the optimal value it can find.

It is noted that the exchange algorithm, including its variants, has been intensively studied and broadly applied in experimental designs. However, its applications in engineering system design are not well explored. Given the flexibility of exchange algorithm and the empirical evidence that the presented fast exchange algorithm is capable of handling large scale engineering design problems, much more needs to be done to present theoretical justifications of the algorithm's performances.

Optimal design of sensor systems will surely make the task of variation root cause diagnosis more meaningful and efficient. However, criteria and methods presented in

this thesis will certainly find their applications beyond coordinate sensor placement since the approach is based on a general linear system model.

REFERENCES

- Aarts, E. and Lenstra, J.K., (1997) *Local Search in Combinatorial Optimization*, John Wiley and Sons, New York .
- Apley, D. W. and Lee, H. Y. (2003) Identifying Spatial Variation Patterns in Multivariate Manufacturing Processes: A Blind Separation Approach. *Technometrics*, **45**, 187-198.
- Apley, D.W., and Shi, J. (1998) Diagnosis of Multiple Fixture Faults in Panel Assembly. *ASME Journal of Manufacturing Science and Engineering*, **120**, 793-801.
- Apley, D.W., and Shi, J. (2001) A Factor-Analysis Methods for Diagnosing Variability in Multivariate Manufacturing Processes. *Technometrics*, **43**, 84-95.
- Atkinson, A.C. and Donev, A.N. (1992) *Optimum Experimental Designs*, Oxford University Press, New York.
- Carlson, J.S., Lindkvist, L., and Soderberg, R. (2000) Multi-Fixture Assembly System Diagnosis Based on Part and Subassembly Measurement Data. *Proceedings of the 2000 ASME Design Engineering Technical Conference*, September 10-13, Baltimore, MD.
- Ceglarek, D. and Shi, J. (1995) Dimensional Variation Reduction for Automotive Body Assembly. *Manufacturing Review*, **8**, 139 – 154.
- Ceglarek, D. and Shi. J. (1996) Fixture Failure Diagnosis for Autobody Assembly

- Using Pattern Recognition. *ASME Journal of Engineering for Industry*, **188**, 55-65.
- Ceglarek, D., Shi, J., and Wu, S.M. (1994) A Knowledge-Based Diagnostic Approach for the Launch of the Auto-Body Assembly Process. *ASME Journal of Engineering for Industry*, **116**, 491 - 499.
- Cook, R. D. and Nachtsheim, C. J. (1980) A Comparison of Algorithms for Constructing Exact D-Optimal Designs. *Technometrics*, **22**, 315-323.
- Ding, Y., Ceglarek, D., and Shi, J. (2002a) Fault Diagnosis of Multi-station Manufacturing Processes Using State Space Approach. *ASME Journal of Manufacturing Science and Engineering*, **124**, 313-322.
- Ding, Y., Ceglarek, D., and Shi, J. (2002b) Design Evaluation of Multi-station Manufacturing Processes Using State Space Approach. *ASME Journal of Mechanical Design*, **124**, 408-418.
- Ding, Y., Ceglarek, D., and Shi, J., (2000) Modeling and Diagnosis of Multi-station Manufacturing Processes: Part I State Space Model. *Proceedings of the 2000 Japan/USA Symposium on Flexible Automation*, July 23-26, Ann Arbor, MI, 2000JUSFA-13146.
- Ding, Y., Kim, P., Ceglarek, D., and Jin, J. (2003) Optimal Sensor Distribution for Variation Diagnosis for Multi-station Assembly Processes. *IEEE Transactions on Robotics and Automation*, **19**, 543-556.
- Djurdjanovic, D. and Ni, J., (2001) Linear State Space Modeling of Dimensional

- Machining Errors. *Transactions of NAMRI/SME*, XXIX, pp. 541-548.
- Ehrenfeld, S. (1955) On the Efficiency of Experimental Designs. *Annals of Mathematical Statistics*, **26**, 247-255.
- Fedorov, V. V. (1972) *Theory of Optimal Experiments*, Academic Press, New York.
- Hu, S.J. (1990) *Impact of 100% Measurement Data on Statistical Process Control (SPC) in Automobile Body Assembly*, Ph.D. Dissertation, The University of Michigan, Ann Arbor, MI.
- Hu, S.J. and Wu, S. (1992) Identifying Sources of Variation in Automobile Body Assembly Using Principal Component Analysis. *Transactions of NAMRI/SME*, XX, 311-316.
- Jin, J. and Shi, J., (1999) State Space Modeling of Sheet Metal Assembly for Dimensional Control. *ASME Journal of Manufacturing Science & Engineering*, **121**, pp. 756 – 762.
- Khan, A. and Ceglarek, D. (2000) Sensor Optimization for Fault Diagnosis in Multi-Fixture Assembly Systems With Distributed Sensing. *ASME Journal of Manufacturing Science & Engineering*, **122**, 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 & Engineering*, **120**, 781 – 792.
- Khan, A., Ceglarek, D., Shi, J., Ni, J., and Woo, T. C. (1999) Sensor Optimization for Fault Diagnosis in Single Fixture Systems: A Methodology. *ASME Journal of*

- Manufacturing Science and Engineering*, 121, 109 – 121.
- Lam, R.L.H., Welch, W.J., and Young, S.S. (2002) Uniform Coverage Designs for Molecule Selection. *Technometrics*, **44**, 99-109.
- Lawless, J.F., Mackay, R. J. and Robinson, J.A. (1999) Analysis of Variation Transmission in Manufacturing Processes – Part I & II. *Journal of Quality Technology*, **31**, 131 – 154.
- Mantripragada, R. and Whitney, D.E., (1999) Modeling and Controlling Variation Propagation in Mechanical Assemblies Using State Transition Models. *IEEE Transaction On Robotics and Automation*, **15**, pp. 124 – 140.
- McCulloch, C.E. and Searle, S.R. (2000) *Generalized Linear and Mixed Models*, New York: John Wiley & Sons.
- Meyer, R. K. and Nachtsheim, C. J. (1995) The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs. *Technometrics*, **37**, 60-69.
- Muller, P. C. and Weber, H. I. (1972) Analysis and Optimization of Certain Qualities of Controllability and Observability for Linear Dynamical Systems. *Automatica*, **8**, 237-246.
- Patton, R., Frank, P., and Clark, R., (1989) *Fault Diagnostics in Dynamic Systems: Theory and Applications*, Prentice Hall International Series.
- Paul, R.P. (1981) *Robot Manipulators: Mathematics, Programming, and Control*, MIT Press, Cambridge, MA 02139.
- Powell, M.J.D. (1992) A Direct Search Optimization Methods that Models the

- Objective and Constraint Functions by Linear Interpolation. *Numerical Analysis Reports*, The University of Cambridge. DAMTP 1992/NA5.
- Rao, C.R., and Kleffe, J. (1988) *Estimation of Variance Components and Applications*, Amsterdam: North-Holland.
- Schott, J. R. (1997) *Matrix Analysis for Statistics*, New York: John Wiley & Sons.
- Shalon, D., Gossard, D., Ulrich, K., and Fitzpatrick, D. (1992) Representing Geometric Variations in Complex Structural Assemblies on CAD Systems. *Proceedings of the 19th Annual ASME Advances in Design Automation Conference*, DE-vol.44-2, 121-132.
- Soderberg, R. and Carlson, J. S. (1999) Locating Scheme Analysis for Robust Assembly and Fixture Design. *Proceedings of the 1999 ASME Design Engineering Technical Conferences*, September 12-15, Las Vegas, Nevada.
- Thornton, A. C. (1999) A Mathematical Framework for the Key Characteristic Process. *Research in Engineering Design*, **11**, 145 – 157.
- Wang, Y. and Nagarkar, S. R. (1999) Locator and Sensor Placement for Automated Coordinate Checking Fixtures. *ASME Journal of Manufacturing Science and Engineering*, **121**, 709 – 719.
- Whitney, D.E., Gilbert, O., and Jastrzebski, M. (1994) Representation of Geometric Variations Using Matrix Transforms for Statistical Tolerance Analysis in Assemblies. *Research in Engineering Design*, **6**, 191 –210.
- Zhou, S., Ding, Y., Chen Y., and Shi, J. (2003a) Diagnosability Study of Multistage

Manufacturing Processes Based on Linear Mixed-Effects Models.
Technometrics, **45**, 312-325.

Zhou, S., Huang, Q., and Shi, J., (2003b) State Space Modeling for Dimensional Monitoring of Multistage Machining Process Using Differential Motion Vector.
IEEE Transactions on Robotics and Automation, **19**, pp. 296-308.

APPENDIX

A.1 Proof of $\{(\mathbf{a}_i^T \mathbf{a}_j)^2\}_{i,j=1}^p = \pi(\mathbf{A})^T \pi(\mathbf{A})$.

Proof. Recall that $(\mathbf{a}_i^T \mathbf{a}_j)^2 = \text{tr}(\mathbf{a}_i \mathbf{a}_i^T \mathbf{a}_j \mathbf{a}_j^T)$ and $\text{tr}(\mathbf{A}\mathbf{B}) = \text{vec}(\mathbf{A})^T \text{vec}(\mathbf{B})$ for any symmetric matrices \mathbf{A} and \mathbf{B} . The $(i, j)^{\text{th}}$ element in $\{\text{tr}(\mathbf{a}_i \mathbf{a}_i^T \mathbf{a}_j \mathbf{a}_j^T)\}_{i,j=1}^p$ is $(\text{vec}(\mathbf{a}_i \mathbf{a}_i^T))^T \text{vec}(\mathbf{a}_j \mathbf{a}_j^T)$. Actually, $\text{vec}(\mathbf{a}_i \mathbf{a}_i^T)$ is the i^{th} column vector in $\pi(\mathbf{A})$. That leads to the conclusion that $\pi(\mathbf{A})^T \pi(\mathbf{A}) = \{(\mathbf{a}_i^T \mathbf{a}_j)^2\}_{i,j=1}^p$. \diamond

A.2 Proof of Lemma 1.

Proof. We know that $\{\mathbf{a}_i^T \mathbf{a}_j\}_{i,j=1}^p = \mathbf{A}^T \mathbf{A}$. Then, $\{(\mathbf{a}_i^T \mathbf{a}_j)^2\}_{i,j=1}^p$ is actually $(\mathbf{A}^T \mathbf{A}) * (\mathbf{A}^T \mathbf{A})$. From the above proof, we know that $\{(\mathbf{a}_i^T \mathbf{a}_j)^2\}_{i,j=1}^p = \pi(\mathbf{A})^T \pi(\mathbf{A})$, which means $\pi(\mathbf{A})^T \pi(\mathbf{A}) = (\mathbf{A}^T \mathbf{A}) * (\mathbf{A}^T \mathbf{A})$. Theorem 7.28 in Schott (1997, p 276) stated that $\lambda_{\min}(\mathbf{A} * \mathbf{B}) \geq \lambda_{\min}(\mathbf{A}\mathbf{B})$ for any non-negative definite matrices \mathbf{A} and \mathbf{B} . The matrix $\mathbf{A}^T \mathbf{A}$ is a non-negative definite matrix so that we can have $\lambda_{\min}((\mathbf{A}^T \mathbf{A}) * (\mathbf{A}^T \mathbf{A})) \geq \lambda_{\min}((\mathbf{A}^T \mathbf{A})(\mathbf{A}^T \mathbf{A}))$. Since $\lambda_{\min}((\mathbf{A}^T \mathbf{A})(\mathbf{A}^T \mathbf{A})) = \lambda_{\min}^2((\mathbf{A}^T \mathbf{A}))$, the above inequality is equivalent to $\lambda_{\min}(\pi(\mathbf{A})^T \pi(\mathbf{A})) \geq \lambda_{\min}^2(\mathbf{A}^T \mathbf{A})$. \diamond

VITA

Name: Qinyan Liu

Permanent Address: Jiada Zhineng Xiaoqu, First, South 5-2-5A

Chengdu, Sichuan Province 610000, P. R. China

Educational Background: B.S. 2000. Department of Precision Instruments

Tsinghua University

M.S. 2002. Department of Precision Instruments

Tsinghua University