

# 國立交通大學

統計學研究所



碩士論文

第二階段二維剖面監控之研究

A Study on Phase II Two-dimensional Profile Monitoring

研究生：劉俞伶

指導教授：洪志真 博士

中華民國一〇一年七月

第二階段二維剖面監控之研究

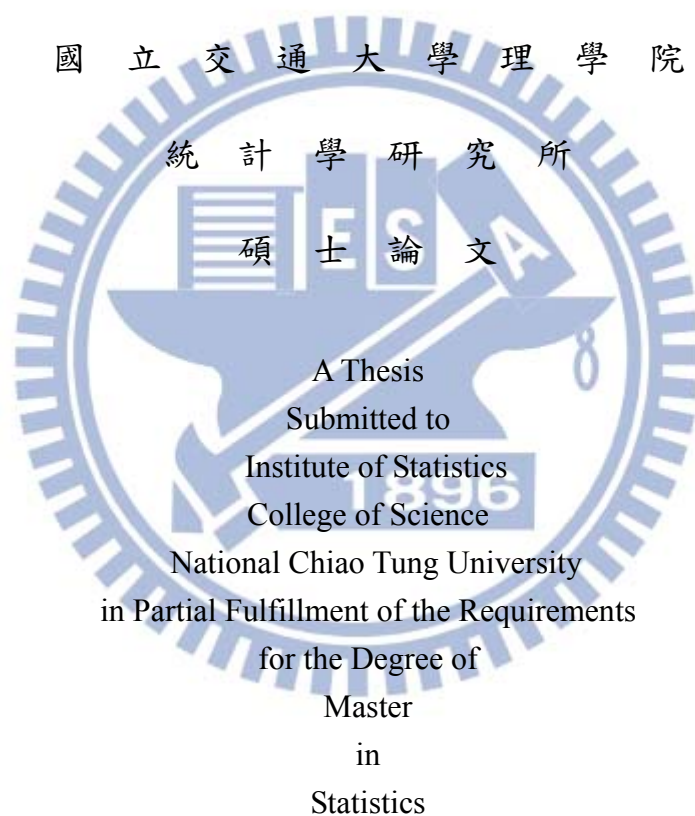
A Study on Phase II Two-dimensional Profile Monitoring

研究生：劉俞伶

Student：Yu-Ling Liu

指導教授：洪志真 博士

Advisor：Dr. Jyh-Jen Horng Shiau



July 2012

Hsinchu, Taiwan, Republic of China

中華民國一〇一年七月

## 第二階段二維剖面監控之研究

研究生：劉俞伶

指導教授：洪志真 博士

國立交通大學統計學研究所碩士班

### 摘要

近年來以統計觀點來監控制程或產品剖面資料的研究已有廣泛發展，大多數的研究皆針對一維剖面提出第一階段或第二階段監控策略。本篇文章是針對二維剖面並納入剖面與剖面之間的變異，運用 PCA 及 MPCA 提出第二階段的監控策略。對於運用 PCA 方面，將二維剖面矩陣的行向量拉成一維向量，並利用主成分分析的方法來分析此一維向量的共變異結構，進而對主成分投影量 (principal component scores) 提出監控策略。對於運用 MPCA 方面，先利用 Ye (2005) 提出的演算法來找尋分別為行方向和列方向的兩個基底矩陣，然後將新進的二維剖面同時對這兩個基底矩陣投影而獲得座標矩陣 (coordinate matrix)，進而對座標矩陣提出監控策略。兩方法皆採用常見的 Hotelling  $T^2$  管制圖來確認製程穩定性。評估兩種方法之績效的方式則是利用平均連串長度來比較偵測能力；整體而言，MPCA 表現較 PCA 來得好，且在執行時間方面，MPCA 較不會受到維度增大的影響。

關鍵字：二維剖面監控、剖面與剖面之間的變異、平均連串長度、主成分分析、多線性主成分分析

# A Study on Phase II Two-dimensional Profile Monitoring

Student: Yu-Ling Liu

Advisor: Dr. Jyh-Jen Horng Shiau

**Institute of Statistics**

**National Chiao Tung University**

## **Abstract**

The study of monitoring process or product profiles by means of statistics already has extensive development. Most research works focus on one-dimensional profile monitoring schemes of Phase I or Phase II. This thesis is directed towards two-dimensional profiles with profile-to-profile variation. We utilize principal component analysis (PCA) and multilinear principal component analysis (MPCA) to propose monitoring schemes of Phase II. Since two-dimensional profile data are represented as matrices, we vectorize these matrix data for PCA to analyze the covariance structure of the resulting one-dimensional vectors. With that, we propose a control chart based on principal component scores. For using MPCA, we first utilize the algorithm proposed by Ye (2005) to search for two basis matrices, one for columns and one for rows, and then project an incoming two-dimensional profile onto the two basis matrices simultaneously to obtain a score matrix. We construct a monitoring scheme based on this score matrix. Both control charts use the Hotelling's  $T^2$  statistic to monitor the stability of the process. The performances of the two methods are evaluated and compared in terms of the average run length. Generally, MPCA performs better than PCA. Moreover, for the execution time, the MPCA method is not affected as much as the PCA method when the size of the profile matrix

increases.

**KEY WORDS:** Two-dimensional Profile Monitoring, Profile-to-Profile Variation, Average Run Length, Principal Component Analysis, Multilinear Principal Component Analysis



## 誌謝

本論文能順利完成，由衷地感謝老師 洪志真教授這兩年悉心指導並對於學生的諸多疑問給予耐心且專業的引導。尤其在修改論文階段，老師鉅細靡遺地將論文中的錯誤挑出來，並給予諸多的寶貴建議，讓學生學習到嚴謹的研究態度及多元的思考方向，使學生獲益良多，學生再一次的深表感謝。此外還要感謝口試委員 王秀瑛教授、黃榮臣教授和鄭少為教授對於本研究給予實際且可行的建議，使本論文的架構更為完整且清晰。

光陰似箭，在交大統計研究所的兩年歲月也將告一段落。在這段歲月裡，很感謝博班 清仁學長和博文學長在課業及論文上給予協助及指導。感謝研究夥伴源毅同學在研究上的協助，增廣我的見聞。感謝家榕與子賢在我趕論文期間給予我鼓勵及關心，使我更有動力完成論文。感謝其他同學與朋友的陪伴與幫助。我很喜歡在交大兩年的生活，認識到許多人也學習到許多事，這將會是我人生中美好記憶的一部分。

最後，我很感謝我的父母親，在這兩年除了提供我不虞匱乏的生活外，也給予我努力往前衝的勇氣。感謝我姐姐的打氣及我妹妹在我趕論文時幽默逗趣的講話方式讓我能舒壓。希望大家都能健康與順心。

劉俞伶 謹誌于

國立交通大學統計學研究所

中華民國一〇一年七月

# 目錄

中文摘要.....	i
英文摘要.....	ii
誌謝.....	iv
目錄.....	v
圖目錄.....	vi
表目錄.....	vi
第一章 緒論.....	1
1.1 研究動機與目的.....	1
1.2 研究架構.....	2
第二章 文獻回顧.....	3
2.1 剖面監控之文獻.....	3
2.2 高階張量資料降低維度之文獻.....	4
2.3 MPCA 概念簡介.....	5
2.3.1 樣本觀點.....	6
2.3.2 母體觀點.....	7
2.3.3 維度之選擇.....	9
第三章 監控方法.....	11
3.1 PCA 方法.....	11
3.2 母體觀點演算法.....	12
3.3 MPCA 方法.....	12
3.4 ARL.....	14
3.4.1 PCA 方法的 ARL 算法.....	14
3.4.2 MPCA 方法的 ARL 算法.....	15
第四章 模擬與比較研究.....	16
4.1 範例一.....	16
4.1.1 $X_1(s) + X_2(t)$ 為布朗橋+積分布朗運動之情況.....	18
4.1.2 $X_1(s) + Y_1(t)$ 為布朗運動+積分布朗運動之情況.....	20
4.2 範例二.....	22
4.3 PCA 與 MPCA 的執行時間.....	24
4.4 晶圓應用範例.....	25
第五章 結論和未來展望.....	30
5.1 結論.....	30
5.2 未來展望.....	30
參考文獻.....	31
附錄.....	33

## 圖目錄

圖一、晶圓氧化厚度在(a)管制內及在(b)管制外之二模擬剖面 .....	2
圖二、範例一之假設模型在情況(i)布朗橋+積分布朗運動下所模擬出來兩個管制內的二個維剖面.....	18
圖三、範例一之假設模型在情況(i)下以(a)方式偏移時，PCA 及 MPCA 兩方法的 ARL 曲線.....	19
圖四、範例一之假設模型在情況(i)下以(b)方式偏移時，PCA 及 MPCA 兩方法的 ARL 曲線.....	19
圖五、範例一之假設模型在情況(ii)布朗運動+積分布朗運動下所模擬出來兩個管制內的二維剖面.....	20
圖六、範例一之假設模型在情況(ii)下以(a)方式偏移時，PCA 及 MPCA 兩方法的 ARL 曲線.....	21
圖七、範例一之假設模型在情況(ii)下以(b)方式偏移時，PCA 及 MPCA 兩方法的 ARL 曲線.....	21
圖八、由 4.2 小節之假設模型所模擬出來兩個管制內的二維剖面.....	23
圖九、PCA 在 $k=4$ 及 MPCA 在 $\tilde{p}=3$ 及 $\tilde{q}=3$ 時個別的 ARL 曲線，兩者完全重合.....	23
圖十、PCA(解釋比例約為 85%)和 MPCA(解釋比例約為 70%)在 $X_1$ 和 $X_2$ 的平均向量皆有偏移及只有 $X_1$ 的平均向量有偏移兩種情形的 ARL 曲線 .....	24
圖十一、晶圓氧化的目標厚度在 4.1.1 小節之情形下所模擬出來的兩個管制內的二維剖面.....	27
圖十二、晶圓氧化的目標厚度在 4.1.2 小節之情形下所模擬出來的兩個管制內的二維剖面.....	27
圖十三、在情況(I)下之 150 筆測試資料用 PCA 方法監控的管制圖 .....	28
圖十四、在情況(I)下之 150 筆測試資料用 MPCA 方法監控的管制圖 .....	28
圖十五、在情況(II)下之 150 筆測試資料用 PCA 方法監控的管制圖 .....	29
圖十六、在情況(II)下之 150 筆測試資料用 MPCA 方法監控的管制圖 .....	29

## 表目錄

表一、PCA 及 MPCA 在範例一的 CPU 執行時間 (單位：秒).....	25
表二、PCA 及 MPCA 在範例二的 CPU 執行時間 (單位：秒).....	25



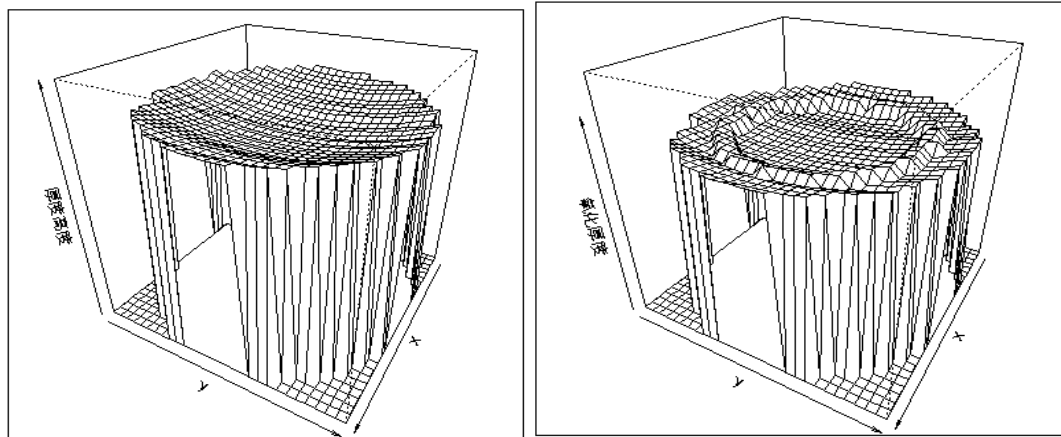
# 第一章 緒論

## 1.1 研究動機與目的

統計製程管制被廣泛運用在很多領域，對於產品或製程上單一或多個可量測的品質特性之監控技術其研究已近乎成熟。近年來，某些產品製程所感興趣的品質特性已不再是傳統的單維或多維之變量而是一些變量間之函數。品質特性常被稱為反應變數，當反應變數為一個單維變量之函數，則在幾何上的意義是曲線。像這樣的函數在先前的文獻是以剖面 (profile) 稱之。關於監控一維剖面在第一和第二階段的研究近年來愈來愈受到重視並已有相當程度的發展。

對於某些特定的製程，品質特性為二維剖面。舉例來說，在半導體薄膜製程中 (Thin Film Process) 晶圓之氧化厚度 (氧化層上研磨的厚度) 是一個重要的品質特性。此類品質特性亦稱之為反應變數，其幾何上的意義是曲面。此例之反應變數是氧化厚度，而兩個解釋變數代表晶圓上量測厚度的位置。Gardner et al. (1997) 以空間特徵 (spatial signature) 一詞來描述此二維剖面，並提出利用 thin-plate splines 方法來估計氧化厚度，且在不考慮晶圓與晶圓之間有個體差異的模型下提出監控方法。

本文針對二維剖面之監控問題，在考慮剖面有個體差異性下，藉由傳統的主成分分析 (Principal Component Analysis, 簡稱 PCA) 及影像處理領域上的多線性主成分分析 (Multilinear Principal Component Analysis, 簡稱 MPCA)，提出第二階段製程監控方法。圖一仿 Gardner et al. (1997) 文章中的晶圓氧化厚度之例，以布朗運動 (Brownian motion) 來納入晶圓與晶圓之間的差異性，呈現兩個二維剖面；(a)圖、(b)圖分別為兩片晶圓之氧化厚度在管制內和管制外的情形，其中 (b)圖所呈現管制外的情形為晶圓由內往外研磨時，在某一圈研磨機器發生狀況而導致那一圈的厚度較厚。



圖一、晶圓氧化厚度在(a)管制內及在(b)管制外之二模擬剖面。

## 1.2 研究架構

第一章是敘述本篇論文的研究動機及目的。第二章是文獻回顧，包括回顧一維和二維剖面的研究發展，及在影像處理領域上對於高階張量資料降低維度的研究情形，最後回顧 Hung et al. (2012) 對於 MPCA 概念的介紹。第三章是在考慮剖面與剖面之間具有個體差異性下，對於二維剖面提出第二階段的監控策略，並評估所提方法的執行效率。第四章列舉三個模擬範例，我們藉由前兩個範例來說明並評估在第三章中所提的方法，藉由第三個範例來陳述實際的晶圓資料應如何處理。第五章是本文的結論及未來展望。最後的附錄包括第二章中之定理二及第三章中所提的性質和定理的證明。

## 第二章 文獻回顧

本篇論文主要探討二維剖面之監控問題，在考量剖面間存在個體差異性之情形下，運用 PCA 及 MPCA 提出第二階段製程監控方法。在影像處理領域方面，已有很多方法是在處理高階張量 (higher-order tensors) 資料降低維度的問題，MPCA 是其中之一。底下的文獻回顧分成三小節，2.1 小節是回顧剖面監控、2.2 小節是回顧影像處理領域中關於高階張量資料降低維度的研究發展及 2.3 小節是 MPCA 概念簡介。

### 2.1 剖面監控之文獻

1.1 小節提到，有些產品或製程的品質特性是一些解釋變數的函數，因此，第二階段剖面監控主要是抽取製程現場樣本來確認隨著時間的推移這函數關係是否仍維持穩定。近年來，以統計觀點對於第一、二階段剖面監控的研究皆有廣泛發展。最早的研究是針對線性剖面 (linear profiles) 提出第一、二階段的監控策略。例如，Kang and Albin (2000) 提出兩種策略來監控線性剖面；Kim et al. (2003) 提出利用指數加權移動平均 (EWMA) 管制圖個別監控截距、斜率及誤差項的變異數；Zou et al. (2006) 及 Mahmoud et al. (2007) 皆利用類似的線性模型來描述線性剖面且分別提出監控理論；Saghaei (2009) 提出藉由累積和 (CUSUM) 來監控線性剖面。Zou et al. (2007) 及 Kazemzadeh et al. (2008) 考慮利用多個 (multiple) 或多項式 (polynomial) 迴歸模型來描述線性剖面。之後有些學者考慮非線性剖面。例如，Williams et al. (2007) 對於第一階段的監控統計量  $T^2$  提出三種形式來估計共變異矩陣，同時亦考慮非參數迴歸模型藉以衡量新進剖面與基準剖面之間的差異。另外，Gardner et al. (1997) 以晶圓氧化厚度為例，針對二維剖面發展一套方法。首先，藉 thin-plate splines 來估計實際的晶圓氧化厚度曲面；其次，找尋基準曲面 (baseline surface) 並提出四種衡量標準來量測新進的曲面與

基準曲面之間的差異，進而判定晶圓在研磨過程是否有狀況發生；其中四種衡量標準包括(i)2-階範數 (2-norm)、(ii)1-階範數(1-norm)、(iii)設定限制條件及(iv)對不同區域的厚度加不同的權重。以上是 Gardner et al. (1997) 所提方法的概要，文中並未納入剖面間存在個體差異性之考量也未針對所提方法研究執行效能。Shiau and Weng (2004) 以非參數迴歸來處理更一般化的剖面並提出監控策略。以上所提研究均假設固定效應之模型。Shiau et al. (2009) 將 Shiau and Weng (2004) 之非參數固定效應 (nonparametric fixed-effect) 模型推廣至剖面間有個體差異性 (亦即為隨機效應) 的情形，並運用 PCA 來獲得資料對前幾個主要的主成分個別投影的投影量 (即主成分分數 (score)) 且提出  $T^2$  統計量來監控之。關於剖面間存在個體差異性之一維剖面研究亦有其他學者提出不同的監控方法。例如，Ding et al. (2006) 對第一階段製程管制提出利用獨立成分分析 (Independent Components Analysis) 來監控剖面。Jensen et al. (2008) 及 Jensen and Birch (2009) 分別提出線性及非線性混合效應 (mixed-effect) 模型。Qiu et al. (2010) 以無母數迴歸模型提出監控策略。上述除了 Gardner et al. (1997) 的文章外，其餘皆是學者針對一維剖面監控所提出的監控方法。最近幾年，有些學者對於一維剖面模型有更進一步的推廣，他們考慮一次不只一個反應變數的情形。這裡的反應變數是函數，以幾何觀點來看，學者們一次考慮多條曲線。例如，Eyvazian et al. (2011) 考慮多變數管制圖 (multivariate control chart) 來監控模型中的所有參數、Zou et al. (2012) 提出 LASSO-based 多變數管制圖。

## 2.2 高階張量資料降低維度之文獻

在資料分析上將高維度資料降低維度的方法中 PCA 是常見的方法之一。其方法為：對  $m$  個變數得到  $n$  個  $m \times 1$  之觀測向量，計算此筆資料之  $m \times m$  共變異矩陣並對其做特徵根分解；目標是選擇較少的特徵向量當作新座標軸使得轉換後的新變數能夠盡可能保留原本資料的變異。然而，當資料是高階張量資料，如果將它拉成一個很長的一維向量，在運用傳統 PCA 來處理則會有一些問題產生。因

為此情形通常參數個數會遠大於樣本個數使得大多的統計方法都不適用。克服此困難的方法是利用資料本身的張量結構。De Lathauwer et al. (2002a) 對於一個給定的  $H^{\text{th}}$  階張量提出將一般的奇異值分解 (singular value decomposition, 簡稱 SVD) 推廣到高階奇異值分解 (HOSVD) ; De Lathauwer et al. (2002b) 有更進一步的研究 ; Yang et al. (2004) 對於影像資料提出二維度 PCA (two-dimensional PCA, 簡稱 2DPCA) ; Zhang and Zhou (2005) 提出二方向二維 PCA (two-directional two-dimensional PCA, 簡稱  $2D^2$  PCA) 的方法來改善 2DPCA, 藉由模擬顯示出  $2D^2$  PCA 比 2DPCA 在重建表現上較佳 ; Ye (2005) 制訂矩陣的一般低秩 (low rank) 近似的問題, 同時也提出演算法 ; Lu et al. (2008) 將 Ye (2005) 的問題一般化並對任意階的張量資料提出多線性主成分分析方法 ; Kolda and Bader (2009) 提出張量分解的一般觀點及應用 ; Li et al. (2010) 將張量分解方法應用於監督學習 (supervised learning) 方面之迴歸與分類上 ; Hung et al. (2012) 對於二階張量的 MPCCA 提出一些有用的統計性質。

### 2.3 MPCCA 概念簡介

由於二維剖面的資料型態是矩陣, 因此以下所介紹之 MPCCA 基本概念乃針對二階張量 (two-order tensors) 資料。

將隨機矩陣  $X_{p \times q}$  視為二階張量, 令  $E(X) = \mu_{p \times q}$  及  $\text{cov}(\text{vec}(X)) = \sum_{pq \times pq}$ , 其中  $\text{vec}(\cdot)$  為將矩陣向量化之算子, 亦即將  $X$  的第二行放在第一行下面, 而後第三行放在第二行下面, 依此類推拉成一  $pq \times 1$  的向量。例如令  $X_{2 \times 2} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$ , 則  $\text{vec}(X) = (x_{11}, x_{21}, x_{12}, x_{22})^T$ 。MPCCA 主要目的是在找尋二基底矩陣

$\{A_0, B_0\}$  及座標矩陣 (coordinate matrix)  $U_{X, p_0 \times q_0}$  使得  $\mu + A_0 U_X B_0^T$  和  $X$  的差異度很小, 其中  $p_0 \leq p$  和  $q_0 \leq q$  皆為未知, 而  $A_0$  和  $B_0$  分別反應  $X$  行和列方向上的資訊, 並要求  $A_0^T A_0 = I_{p_0}$  及  $B_0^T B_0 = I_{q_0}$ 。故在影像處理應用上可藉由

$\mu + A_0 U_X B_0^T$  來重建  $X$ 。以下我們介紹  $A_0$ 、 $U_X$  和  $B_0$  的找法以及  $\tilde{p}$  和  $\tilde{q}$  的選擇準則，詳情見 Hung et al. (2012)。

### 2.3.1 樣本觀點

假設  $\{X_i\}_{i=1}^n$  是與  $X$  同分配的一組隨機樣本。在預先給定的  $\tilde{p}$  及  $\tilde{q}$  下，Ye (2005) 提出

$$\min_{\substack{A \in O_{p \times \tilde{p}} \\ B \in O_{q \times \tilde{q}} \\ \{U_i\}_{i=1}^n}} \frac{1}{n} \sum_{i=1}^n \|(X_i - \bar{X}) - AU_i B^T\|_F^2 \quad (1)$$

的準則，其中  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ， $\|\cdot\|_F$  為矩陣的佛賓尼斯範數 (Frobenius norm) 其定義為矩陣中的每個元素平方後相加最後將總和開根號， $O_{\ell \times \tilde{\ell}}$  為一個集合包括所有  $\ell \times \tilde{\ell}$  的矩陣  $M_{\ell \times \tilde{\ell}}$  滿足  $M^T M = I_{\tilde{\ell}}$ ，其中  $\tilde{\ell} \leq \ell$ 。底下的定理一為 Ye (2005) 對於式(1)的解所提出之一些有用的性質。

**定理一 (Ye, 2005)** 在預先給定的  $\tilde{p}$  及  $\tilde{q}$  下，令  $\hat{A}_{0, p \times \tilde{p}}$ 、 $\hat{B}_{0, q \times \tilde{q}}$  及  $\{\hat{U}_{i, \tilde{p} \times \tilde{q}}\}_{i=1}^n$  為式(1)的解。則

(a)  $\hat{U}_i = \hat{A}_0^T (X_i - \bar{X}) \hat{B}_0$  ;

(b)  $\{\hat{A}_0, \hat{B}_0\}$  為  $\max_{\substack{A \in O_{p \times \tilde{p}} \\ B \in O_{q \times \tilde{q}}} \frac{1}{n} \sum_{i=1}^n \|A^T (X_i - \bar{X}) B\|_F^2$  的解；

(c)  $\hat{A}_0$  的行向量是由  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) P_{\hat{B}_0} (X_i - \bar{X})^T$  的前  $\tilde{p}$  大的特徵值所對應的特徵向量依序排置而成， $\hat{B}_0$  的行向量是由  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^T P_{\hat{A}_0} (X_i - \bar{X})$  的前  $\tilde{q}$  大的特徵值所對應的特徵向量依序排置而成，其中  $P_{\hat{A}_0} \equiv \hat{A}_0 (\hat{A}_0^T \hat{A}_0)^{-1} \hat{A}_0^T$  為一正交投影矩陣，正交投影到  $\text{span}(\hat{A}_0)$  空間上， $P_{\hat{B}_0} \equiv \hat{B}_0 (\hat{B}_0^T \hat{B}_0)^{-1} \hat{B}_0^T$  為正交投影矩陣，正交

投影到  $\text{span}(\hat{B}_0)$  空間上。

下面是 Ye (2005) 所提計算  $\{\hat{A}_0, \hat{B}_0\}$  的演算法，其將之稱為 GLRAM 演算法。

**GLRAM (Ye, 2005)** 給定一隨機初始值  $A^{(0)} \in O_{p \times \tilde{p}}$ ，對於  $N = 0, 1, 2, \dots$

步驟一：  $B^{(N+1)}$  為  $\max_{B \in O_{q \times \tilde{q}}} \frac{1}{n} \sum_{i=1}^n \left\| A^{(N)T} (X_i - \bar{X}) B \right\|_F^2$  的解。

步驟二：  $A^{(N+1)}$  為  $\max_{A \in O_{p \times \tilde{p}}} \frac{1}{n} \sum_{i=1}^n \left\| A^T (X_i - \bar{X}) B^{(N+1)} \right\|_F^2$  的解。

步驟三： 重複步驟一及步驟二，直到  $\frac{1}{n} \sum_{i=1}^n \left\| A^{(N)T} (X_i - \bar{X}) B^{(N)} \right\|_F^2$  和

$\frac{1}{n} \sum_{i=1}^n \left\| A^{(N+1)T} (X_i - \bar{X}) B^{(N+1)} \right\|_F^2$  之間沒有顯著差異，即可令  $\hat{A}_0 = A^{(N+1)}$  及  $\hat{B}_0 = B^{(N+1)}$ 。

在 Hung et al. (2012) 的文章提到，GLRAM 演算法確保當  $N$  遞增時

$\frac{1}{n} \sum_{i=1}^n \left\| A^{(N)T} (X_i - \bar{X}) B^{(N)} \right\|_F^2$  會單調遞增 (指的是不嚴格遞增)，並且  $\frac{1}{n} \sum_{i=1}^n \|X_i - \bar{X}\|_F^2$

是其上界，故此演算法必會收斂。但 GLRAM 演算法有可能只找到局部最大值，

這和所選擇的隨機初始值  $A^{(0)}$  有關。Ye (2005) 提出重複選擇隨機初始值  $A^{(0)}$  來確

保得到整體最大值。相對的，Hung et al. (2012) 提出利用  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$  的

前  $\tilde{p}$  大的特徵值所對應的特徵向量來當作隨機初始值  $A^{(0)}$ 。因此，藉由定理一及

GLRAM 演算法就可獲得  $\{\hat{A}_0, \hat{B}_0\}$  及  $\{\hat{U}_i\}_{i=1}^n$ 。

### 2.3.2 母體觀點

相對於定理一之樣本觀點，我們有如下之母體觀點。

在預先給定的  $\tilde{p}$  及  $\tilde{q}$  下，將式(1)修改為

$$\min_{\substack{A \in O_{p \times \tilde{p}} \\ B \in O_{q \times \tilde{q}} \\ U}} E \left\| (X - \mu) - AUB^T \right\|_F^2 \quad (2)$$

**定理二** 在預先給定的  $\tilde{p}$  及  $\tilde{q}$  下，令  $\tilde{A}_{0, p \times \tilde{p}}$ 、 $\tilde{B}_{0, q \times \tilde{q}}$  及  $\tilde{U}_{0, \tilde{p} \times \tilde{q}}$  為式(2)的解。則

(a')  $Prob(\tilde{U}_0 = \tilde{A}_0^T (X - \mu) \tilde{B}_0) = 1$  ;

(b')  $\{\tilde{A}_0, \tilde{B}_0\}$  為  $\max_{\substack{A \in O_{p \times \tilde{p}} \\ B \in O_{q \times \tilde{q}}} E \|A^T (X - \mu) B\|_F^2$  的解 ;

(c')  $\tilde{A}_0$  的行向量是由  $E[(X - \mu) P_{\tilde{B}_0} (X - \mu)^T]$  的前  $\tilde{p}$  大的特徵值所對應的特徵向量依序排置而成， $\tilde{B}_0$  的行向量由  $E[(X - \mu)^T P_{\tilde{A}_0} (X - \mu)]$  的前  $\tilde{q}$  大的特徵值所對應的特徵向量依序排置而成，其中  $P_{\tilde{A}_0} \equiv \tilde{A}_0 (\tilde{A}_0^T \tilde{A}_0)^{-1} \tilde{A}_0^T$  及  $P_{\tilde{B}_0} \equiv \tilde{B}_0 (\tilde{B}_0^T \tilde{B}_0)^{-1} \tilde{B}_0^T$ 。

將  $\tilde{A}_0$  及  $\tilde{B}_0$  表示成  $\tilde{A}_0 = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{\tilde{p}})$  及  $\tilde{B}_0 = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{\tilde{q}})$ ，其中  $\mathbf{a}_i$  及  $\mathbf{b}_i$  分別為  $\tilde{A}_0$  及  $\tilde{B}_0$  的第  $i$  個行向量。因

$$E[(X - \mu) P_{\tilde{B}_0} (X - \mu)^T] = \sum_{j=1}^{\tilde{q}} E[(X - \mu) (\mathbf{b}_j \mathbf{b}_j^T) (X - \mu)^T] = \sum_{j=1}^{\tilde{q}} (\mathbf{b}_j \otimes I_p)^T \Sigma (\mathbf{b}_j \otimes I_p) \text{ 及}$$

$$E[(X - \mu)^T P_{\tilde{A}_0} (X - \mu)] = \sum_{i=1}^{\tilde{p}} E[(X - \mu)^T (\mathbf{a}_i \mathbf{a}_i^T) (X - \mu)] = \sum_{i=1}^{\tilde{p}} (I_q \otimes \mathbf{a}_i)^T \Sigma (I_q \otimes \mathbf{a}_i) \text{ ,}$$

其中  $\otimes$  為克羅內克積 (Kronecker product)。因此  $\tilde{A}_0$  包括  $\sum_{j=1}^{\tilde{q}} (\mathbf{b}_j \otimes I_p)^T \Sigma (\mathbf{b}_j \otimes I_p)$  的前  $\tilde{p}$  大的特徵值所對應的特徵向量； $\tilde{B}_0$  包括  $\sum_{i=1}^{\tilde{p}} (I_q \otimes \mathbf{a}_i)^T \Sigma (I_q \otimes \mathbf{a}_i)$  的前  $\tilde{q}$  大的

特徵值所對應的特徵向量。附錄中有我們對上面兩式詳細的推導過程。

另外，由定理二性質 (b') 進一步可推得

$$\max_{\substack{A \in O_{p \times \tilde{p}} \\ B \in O_{q \times \tilde{q}}} E \|A^T (X - \mu) B\|_F^2 = \max_{\substack{A \in O_{p \times \tilde{p}} \\ B \in O_{q \times \tilde{q}}} \text{trace}\{(B \otimes A)^T \Sigma (B \otimes A)\} \text{ .} \quad (3)$$

以上是 Hung et al. (2012) 針對定理二中的性質 (b') 及 (c') 所做之進一步推導。文中並提出下列之性質一，但並未對定理二中之性質 (a') ~ (c') 提出證明，其證明請見附錄。

**性質一 (Hung et al., 2012)** 對任一大小為  $pq \times pq$  之半正定矩陣  $\Sigma$ ，式(3)右式



之最大解會存在。

在 Hung et al. (2012) 文章中提到，由性質一可知  $\max_{\substack{A \in O_{p \times \tilde{p}} \\ B \in O_{q \times \tilde{q}}} E \|A^T (X - \mu) B\|_F^2$  的

解會存在，進而得到式(2)的解亦會存在。至於求算  $\tilde{A}_0$  和  $\tilde{B}_0$  的演算法，Hung et al. (2012) 並未提到，我們仿照 GLRAM 演算法提出母體觀點演算法來求算  $\tilde{A}_0$  和  $\tilde{B}_0$  並敘述在第三章中。

### 2.3.3 維度之選擇

Hung et al. (2012) 利用 PCA 的觀念，提出藉由解釋變異來決定  $\tilde{p}$  及  $\tilde{q}$ 。首先定義累積變異 (cumulative variance)，其用來量度張量資料投影到 MPCA 子空間上所能解釋的變異。

**定義一 (Hung et al., 2012)** 令  $\{\tilde{A}_0, \tilde{B}_0\}$  為式(3)的一組解，

$\Phi(\tilde{p}, \tilde{q}) = E \| \tilde{A}_0^T (X - \mu) \tilde{B}_0 \|_F^2$  為  $X$  在秩(rank)  $-(\tilde{p}, \tilde{q})$  內的累積變異，及

$\rho(\tilde{p}, \tilde{q}) = \Phi(\tilde{p}, \tilde{q}) / \Phi(p, q)$  為  $X$  的總變異在秩  $-(\tilde{p}, \tilde{q})$  內被解釋的比例，其中

$\Phi(p, q) = E \| X - \mu \|_F^2$  為  $X$  的總變異。對於樣本亦有類似的定義：

$$\hat{\Phi}(\tilde{p}, \tilde{q}) = \frac{1}{n} \sum_{i=1}^n \| \hat{A}_0^T (X_i - \bar{X}) \hat{B}_0 \|_F^2, \quad \hat{\Phi}(p, q) = \frac{1}{n} \sum_{i=1}^n \| X_i - \bar{X} \|_F^2 \text{ 及}$$

$$\hat{\rho}(\tilde{p}, \tilde{q}) = \hat{\Phi}(\tilde{p}, \tilde{q}) / \hat{\Phi}(p, q)。$$

令  $\lambda_1 \geq \dots \geq \lambda_{\tilde{p}} \geq 0$  及  $\xi_1 \geq \dots \geq \xi_{\tilde{q}} \geq 0$  分別為  $E[(X - \mu) P_{\tilde{B}_0} (X - \mu)^T]$  的前  $\tilde{p}$  大及  $E[(X - \mu)^T P_{\tilde{A}_0} (X - \mu)]$  的前  $\tilde{q}$  大的特徵值，那麼可得  $\Phi(\tilde{p}, \tilde{q}) = \sum_{i=1}^{\tilde{p}} \lambda_i = \sum_{j=1}^{\tilde{q}} \xi_j$ 。同

樣地令  $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_{\tilde{p}} \geq 0$  及  $\hat{\xi}_1 \geq \dots \geq \hat{\xi}_{\tilde{q}} \geq 0$  分別為  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) P_{\hat{B}_0} (X_i - \bar{X})^T$  的前  $\tilde{p}$

大的及  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^T P_{\hat{A}_0} (X_i - \bar{X})$  的前  $\tilde{q}$  大的特徵值，亦可得

$\hat{\Phi}(\tilde{p}, \tilde{q}) = \sum_{i=1}^{\tilde{p}} \hat{\lambda}_i = \sum_{j=1}^{\tilde{q}} \hat{\xi}_j$ ，推導過程見附錄。另外，因  $\Phi(\tilde{p}, \tilde{q}) \leq \Phi(p, q)$ ，進而得

$\rho(\tilde{p}, \tilde{q}) \leq 1$ ；而對於樣本觀點亦有類似情形。因此可由  $\rho(\tilde{p}, \tilde{q})$  達到預定的解釋比例  $\rho_0 \in (0, 1)$  來決定  $\tilde{p}$  及  $\tilde{q}$ ，而此部分在 Hung et al. (2012) 的文章中則以假設檢定來呈現。



## 第三章 監控方法

本研究將一維剖面之監控方法推廣到二維剖面上，並考慮二維剖面間具有個體差異性，藉由 PCA 及 MPCA 提出第二階段的監控策略。假設二維剖面資料模型是  $Y_{ij} = Z(s_i, t_j) + \varepsilon_{ij}$ ,  $i=1, \dots, p, j=1, \dots, q$ ， $p$  和  $q$  為正整數且  $\varepsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ 。我們可知二維剖面的資料型態是矩陣且矩陣元素為  $Z(s_i, t_j) + \varepsilon_{ij}$ 。

### 3.1 PCA 方法

先將新進的二維剖面資料矩陣  $Y_{p \times q}$  中心化，這邊的中心化是指  $Y$  減掉藉由經驗或第一階段而得的平均矩陣。再藉由 thin-plate splines 方法將其平滑化以去除噪音之干擾，所得的矩陣以符號  $X_{p \times q}$  表之。假設在利用 thin-plate splines 平滑化時每個剖面所使用之平滑參數都一樣，那麼每個第二階段之  $X$  仍彼此獨立且具相同分配。在第二階段製程監控的研究中，假設在管制內的  $vec(X)$  其結合分配服從  $N_{pq}(\mathbf{0}, \Sigma_X)$ 。首先，我們先對  $\Sigma_X$  做特徵根分解得到  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{pq} \geq 0$  及所對應的特徵向量  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{pq}$ 。可知第  $r$  個主成分就是第  $r$  大的特徵根  $\lambda_r$  所對應的特徵向量  $\mathbf{v}_r$ 。接下來，我們需決定要保留多少個主成分，以符號  $k$  代表此量。最常見的方法是利用由這些主成分所解釋的變異在原本資料總變異上所佔的比例（亦即  $\sum_{i=1}^k \lambda_i / \sum_{i=1}^{pq} \lambda_i$ ）是否達到預定的合適值來決定  $k$ ，其中  $k \leq rank(\Sigma_X)$ 。將每個新進的  $vec(X)$  對  $k$  個主成分  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  個別投影得到  $s_1, s_2, \dots, s_k$ ，亦即  $s_r = \mathbf{v}_r' vec(X)$ ,  $r=1, \dots, k$ 。因  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  彼此間獨立且對於在管制內的  $vec(X)$  而言其結合分配服從  $N_{pq}(\mathbf{0}, \Sigma_X)$ ，因此  $s_r$  彼此間獨立且服從  $N(0, \lambda_r)$ ,  $r=1, \dots, k$ 。那麼我們可以考慮藉由監控統計量

$$T_{pca}^2 = \sum_{i=1}^k \frac{S_i^2}{\lambda_i} \quad (4)$$

來建構 $T^2$ 管制圖。當製程在管制內時，此統計量服從自由度為 $k$ 之卡方分配。

接下來我們將利用 MPCA 建構出第二階段的監控方法。在這之前，我們仿 GLRAM 演算法提出母體觀點演算法來求得 $\tilde{A}_0$ 及 $\tilde{B}_0$ 。

### 3.2 母體觀點演算法

**母體觀點演算法** 給定一個隨機初始值 $A^{(0)} \in O_{p \times \tilde{p}}$ ，對於 $N = 0, 1, 2, \dots$

步驟一： $B^{(N+1)}$  為  $\max_{B \in O_{q \times \tilde{q}}} E \|A^{(N)T} (X - \mu) B\|_F^2$  的解。

步驟二： $A^{(N+1)}$  為  $\max_{A \in O_{p \times \tilde{p}}} E \|A^T (X - \mu) B^{(N+1)}\|_F^2$  的解。

步驟三：重複步驟一及步驟二，直到 $E \|A^{(N)T} (X - \mu) B^{(N)}\|_F^2$  和

$E \|A^{(N+1)T} (X - \mu) B^{(N+1)}\|_F^2$  之間沒有顯著差異，即可令 $\tilde{A}_0 = A^{(N+1)}$ 及 $\tilde{B}_0 = B^{(N+1)}$ 。

類似於樣本觀點，上述的母體觀點演算法確保當 $N$ 遞增時

$E \|A^{(N)T} (X - \mu) B^{(N)}\|_F^2$  會單調遞增（指的是不嚴格遞增），又由附錄內定理二性

質(b')的證明中之式(A1)可知 $E \|X - \mu\|_F^2$  為 $E \|A^{(N)T} (X - \mu) B^{(N)}\|_F^2$  的上界，故此演

算法會收斂。在隨機初始值的選擇上可藉由 $E[(X - \mu)(X - \mu)^T]$  的前 $\tilde{p}$ 大的特徵

值所對應的特徵向量來當作隨機初始值 $A^{(0)}$ 。因此藉由上述的母體觀點演算法及

定理二即可得到 $\{\tilde{A}_0, \tilde{B}_0\}$ 及 $\tilde{U}_0$ 。

### 3.3 MPCA 方法

以下我們分成製程參數已知和未知兩種情況來討論。

**第一種情況：參數已知**

先將新進的二維剖面資料矩陣  $Y_{p \times q}$  中心化，如有必要再藉由 thin-plate splines 方法將其平滑化以去除噪音之干擾，所得矩陣以符號  $X_{p \times q}$  表之。在第二階段製程監控的研究中，假設在管制內的  $X$  其  $E(X) = \mathbf{0}_{p \times q}$  及  $\text{cov}(\text{vec}(X)) = \Sigma_X$ ，其中  $\Sigma_X$  為已知。在預先給定的  $\tilde{p}$  及  $\tilde{q}$  下，將  $\Sigma_X$  套用母體觀點演算法得到  $\tilde{A}_{0, \tilde{p} \times \tilde{p}}$  及  $\tilde{B}_{0, \tilde{q} \times \tilde{q}}$ 。由定理二性質 (a') 可知，每個矩陣  $X$  皆可得到相對應之

$$U_{X, \tilde{p} \times \tilde{q}} = \tilde{A}_{0, \tilde{p} \times \tilde{p}}^T X_{p \times q} \tilde{B}_{0, \tilde{q} \times \tilde{q}}。$$

**性質二** 假設  $X$  在管制內，且  $\text{vec}(X) \sim N_{pq}(\mathbf{0}_{pq \times 1}, \Sigma_X)$ ，則

$$\text{vec}(U_X) \sim N_{\tilde{p}\tilde{q}}(\mathbf{0}_{\tilde{p}\tilde{q} \times 1}, T_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}})，其中 T = \text{cov}(\text{vec}(U_X))。$$

性質二的證明請見附錄。令  $\text{rank}(T_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}}) = k_1$ ，亦即矩陣  $T$  的秩。將  $T$  正交對角化使得  $T = GDG^T$ ，其中  $G$  是單範正交矩陣 (orthonormal matrix) 及  $D$  是對角矩陣其對角元素是  $T$  的特徵值。令  $\lambda_1^* \geq \dots \geq \lambda_{k_1}^* \geq \dots \geq \lambda_{\tilde{p}\tilde{q}}^* \geq 0$  為  $T$  的特徵值。因

$$\text{vec}(U_X) \sim N_{\tilde{p}\tilde{q}}(\mathbf{0}_{\tilde{p}\tilde{q} \times 1}, T_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}}) \text{ 且 } D_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}} = G_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}}^T T_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}} G_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}}，則$$

$$G^T \text{vec}(U_X) \sim N_{\tilde{p}\tilde{q}}(\mathbf{0}_{\tilde{p}\tilde{q} \times 1}, D_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}})。令 G^T \text{vec}(U_X) = (s_1^*, \dots, s_{k_1}^*, \dots, s_{\tilde{p}\tilde{q}}^*)^T，明顯地，$$

$s_i^* \stackrel{id}{\sim} N(0, \lambda_i^*)$ ,  $i = 1, \dots, \tilde{p}\tilde{q}$ 。特別注意第  $k_1$  個以後的特徵值是等於 0，因此下面的統計量只考慮前  $k_1$  個。類似於 PCA 我們可以藉由監控統計量

$$T_{mpca}^2 = \sum_{i=1}^{k_1} \frac{s_i^{*2}}{\lambda_i^*} \quad (5)$$

來建構  $T^2$  管制圖。當製程在管制內時，此統計量服從自由度為  $k_1$  之卡方分配。

### 第二種情況：參數未知

假設  $\text{vec}(X)$  的共變異矩陣  $\Sigma_X$  未知，則可藉由第一階段製程管制獲得一組管制內且已中心化及平滑化的二維剖面資料，並藉由這組資料獲得樣本共變異矩陣

$\hat{\Sigma}_X$ 。在預先給定的  $\tilde{p}$  及  $\tilde{q}$  下，將這組管制內的資料藉由 GLRAM 演算法獲得  $\hat{A}_{0,p \times \tilde{p}}$  及  $\hat{B}_{0,q \times \tilde{q}}$  分別用來估計  $\tilde{A}_0$  及  $\tilde{B}_0$ 。接下來，套用第一種情況所敘述的方法來監控新進的二維剖面資料。

當 PCA 及 MPCA 的解釋比例皆為 100% 時，則 PCA 和 MPCA 兩個監控統計量會相等。定理三描述 PCA 和 MPCA 在何種情形下解釋比例皆為 100%。

**定理三** 假設對於特定的  $p_0$  及  $q_0$ ， $X = A_0 U_X B_0^T$ ，並假設式(4)中主成分個數  $k = \text{rank}(\Sigma_X)$ 。那麼當  $\tilde{p} = p_0$  和  $\tilde{q} = q_0$  時，則式(5)會等於式(4)。

定理三之證明請見附錄。在母體觀點下，定理三的  $\tilde{p} = p_0$  和  $\tilde{q} = q_0$  表示  $\tilde{A}_0 = A_0$  及  $\tilde{B}_0 = B_0$ 。明顯地，當  $p_0 = p$  且  $q_0 = q$  時，對任何例子而言，定理三中的  $A_0$  及  $B_0$  一定存在。不過，當  $p_0 < p$  且  $q_0 < q$  時，在某些情況下，定理三中的  $A_0$  及  $B_0$  仍會存在。當定理三成立時，對於 PCA 而言， $k = \text{rank}(\Sigma_X)$  表示第  $k$  個以後的特徵值皆為 0，因此前  $k$  個特徵值的總和會等於資料總變異，亦即 PCA 的解釋比例為 100%。對於 MPCA 而言，重建結果等於原本，表示  $\tilde{A}_0$  及  $\tilde{B}_0$  皆分別抓到行及列方向上的所有資訊，亦即 MPCA 的解釋比例為 100% (亦可由 2.3.3 中的  $\rho(\tilde{p}, \tilde{q})$  公式及附錄定理二性質 (b') 證明中之式(A1)獲得)。在第四章的範例二我們將呈現 PCA 和 MPCA 有相同的平均連串長度 (Average Run Length，簡稱 ARL) 曲線來反應定理三之情形。

### 3.4 ARL

我們藉由 ARL 來評估與比較 PCA 及 MPCA 兩方法的偵測能力。

#### 3.4.1 PCA 方法的 ARL 算法

新進的二維剖面資料矩陣  $Y_{p \times q}$  經由中心化及去除噪音之干擾後得到矩陣

$X_{p \times q}$ ，假設  $\text{vec}(X) \sim N_{pq}(\mathbf{0}_{pq \times 1} + \boldsymbol{\delta}_{pq \times 1}, \Sigma_X_{pq \times pq})$ 。那麼式(4)的統計量  $T_{pca}^2$  會服從自由度為  $k$  及非中心參數為  $\lambda_c = (\Gamma^{*T} \boldsymbol{\delta})^T D_p^{*-1} (\Gamma^{*T} \boldsymbol{\delta})$  之非中心卡方分配 (noncentral chi-square distribution)，其中  $\Gamma^* = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$  及  $D_p^*$  為對角矩陣其元素為  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ ，則  $T^2$  管制圖的偵測力可藉由  $P(T_{pca}^2 > \chi_{k,\alpha}^2) = P(\chi_k^2(\lambda_c) > \chi_{k,\alpha}^2)$  獲得，其中  $\chi_{k,\alpha}^2$  是中心卡方分配 (chi-square distribution)  $\chi_k^2$  的  $(1-\alpha) \times 100\%$  百分位數。

### 3.4.2 MPCA 方法的 ARL 算法

算法和 3.4.1 一樣，只不過  $\lambda_c = (((\tilde{B}_0 \otimes \tilde{A}_0) G^*)^T \boldsymbol{\delta})^T D^{*-1} (((\tilde{B}_0 \otimes \tilde{A}_0) G^*)^T \boldsymbol{\delta})$ ，其中  $D^*$  為對角矩陣其元素為  $\lambda_1^* \geq \dots \geq \lambda_{k_1}^*$  及  $G^*$  為由  $G$  的前  $k_1$  個行向量依序排置而成之矩陣。

## 第四章 模擬與比較研究

在本章中，我們藉由範例一及範例二來比較 PCA 和 MPCA 兩方法之間的偵測能力及執行時間，並呈現在某種特殊情況下 PCA 和 MPCA 兩方法之間的關係。值得一提的是我們所提出的方法皆假設  $\text{vec}(X)$  服從多元常態分配，而我們知道如果一個隨機過程 (stochastic process) 是一個高斯過程 (Gaussian process)，則任意有限個時間點的結合分配會服從多元常態分配。故而底下的範例其模型皆採用高斯過程。

### 4.1 範例一

此例我們考慮布朗運動 (Brownian motion) 及它的變化型式來模擬二維剖面資料，底下對布朗運動及它的變化型式做簡單介紹。

(a) **布朗運動過程 (Brownian motion process)**：一個隨機過程  $[B(t), t \geq 0]$  是一個布朗運動如果 (i)  $B(0) = 0$ ；(ii)  $[B(t), t \geq 0]$  具有穩定獨立遞增 (stationary independent increments) 之性質；(iii) 對於  $t > 0$  而言， $B(t) \sim N(0, \sigma^2 t)$ ，其中  $\sigma$  是大於 0 的常數。布朗運動，又稱為維納過程 (Wiener process)，是一個高斯過程，具有  $E(B(t)) = 0$  及  $\text{cov}(B(t_1), B(t_2)) = \sigma^2 t_1, t_1 \leq t_2$ 。

(b) **布朗橋 (Brownian bridge)**：如果  $[B(t), t \geq 0]$  是一個布朗運動，則

$[B(t), 0 \leq t \leq 1 | B(1) = 0]$  稱為布朗橋。它亦是一個高斯過程，具有

$E(B(t) | B(1) = 0) = 0, t < 1$  及  $\text{cov}(B(t_1), B(t_2) | B(1) = 0) = \sigma^2 t_1(1 - t_2), t_1 \leq t_2 \leq 1$ 。

(c) **積分布朗運動 (integrated Brownian Motion)**：如果  $[B(t), t \geq 0]$  是一個

布朗運動，則定義  $IB(t) = \int_0^t B(s) ds$ ，稱  $\{IB(t), t \geq 0\}$  為積分布朗運動。它亦是

一個高斯過程，具有  $E(IB(t)) = 0$  及  $\text{cov}(IB(t_1), IB(t_2)) = \sigma^2 t_1^2 \left( \frac{t_2}{2} - \frac{t_1}{6} \right), t_1 \leq t_2$ 。

假設此例的模型為  $Y_{ij} = Z(s_i, t_j) + \varepsilon_{ij}, i = 1, \dots, 30, j = 1, \dots, 29$ ，其中



$Z(s_i, t_j) = X_1(s_i) + X_2(t_j), \forall i, j$ ，且  $X_1(s)$  及  $X_2(t)$  皆屬於上述(a),(b),(c)三種高斯過程的其中一種，照理應有八種情況，底下只呈現(i)布朗橋+積分布朗運動和(ii)布朗運動+積分布朗運動兩種情況。值得一提的是，我們只知道  $X_1(s)$  及  $X_2(t)$  個別的平均向量和共變異矩陣，因此對它們做特徵根分解分別得到特徵值

$\lambda_1^{(1)} \geq \dots \geq \lambda_{30}^{(1)} \geq 0$  和  $\lambda_1^{(2)} \geq \dots \geq \lambda_{29}^{(2)} \geq 0$  及個別所對應的特徵向量  $\mathbf{v}_1^{(1)}, \dots, \mathbf{v}_{30}^{(1)}$  和

$\mathbf{v}_1^{(2)}, \dots, \mathbf{v}_{29}^{(2)}$ 。讓  $K_1$  和  $K_2$  分別代表  $X_1$  和  $X_2$  的特徵值非零的主成分個數。則可藉由

$$(X_1(s_1), \dots, X_1(s_{30}))^T = \sum_{i=1}^{K_1} c_i^{(1)} \mathbf{v}_i^{(1)} \quad \text{及} \quad (X_2(t_1), \dots, X_2(t_{29}))^T = \sum_{i=1}^{K_2} c_i^{(2)} \mathbf{v}_i^{(2)}$$

分別建構  $X_1(s_i), i=1, \dots, 30$  和  $X_2(t_j), j=1, \dots, 29$ ，進而建構  $Z(s_i, t_j), \forall i, j$ ，其中

$c_i^{(1)} \sim N(0, \lambda_i^{(1)}), i=1, \dots, K_1$  及  $c_i^{(2)} \sim N(0, \lambda_i^{(2)}), i=1, \dots, K_2$ 。

我們將考慮下列兩種偏移方式：

(a)  $(X_1(s_1), \dots, X_1(s_{30}))^T = \sum_{i=1}^{K_1} c_i^{(1)} \mathbf{v}_i^{(1)}$  且  $c_i^{(1)} \sim N(0 + \delta \sqrt{\lambda_i^{(1)}}, \lambda_i^{(1)}), i=1, \dots, K_1$  及

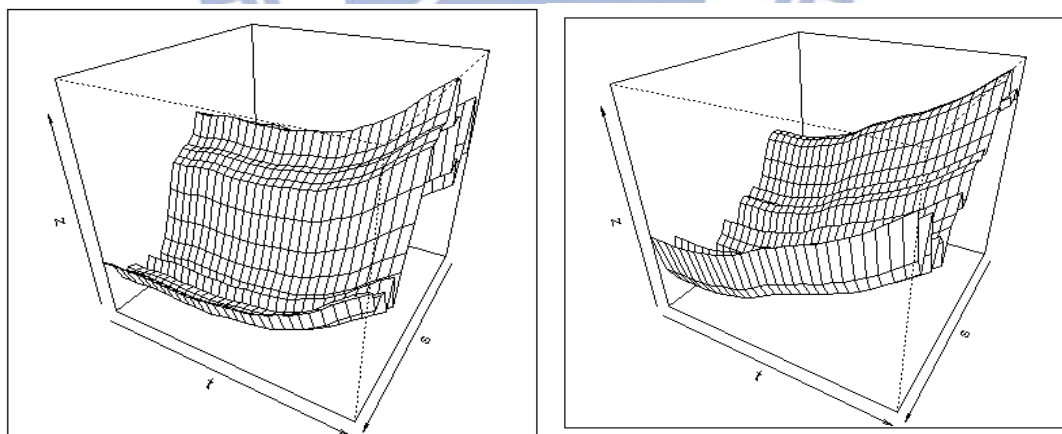
$(X_2(t_1), \dots, X_2(t_{29}))^T = \sum_{i=1}^{K_2} c_i^{(2)} \mathbf{v}_i^{(2)}$  且  $c_i^{(2)} \sim N(0 + \delta \sqrt{\lambda_i^{(2)}}, \lambda_i^{(2)}), i=1, \dots, K_2$ 。

(b) 平均矩陣平移一個常數  $c$ ，亦即垂直平移。

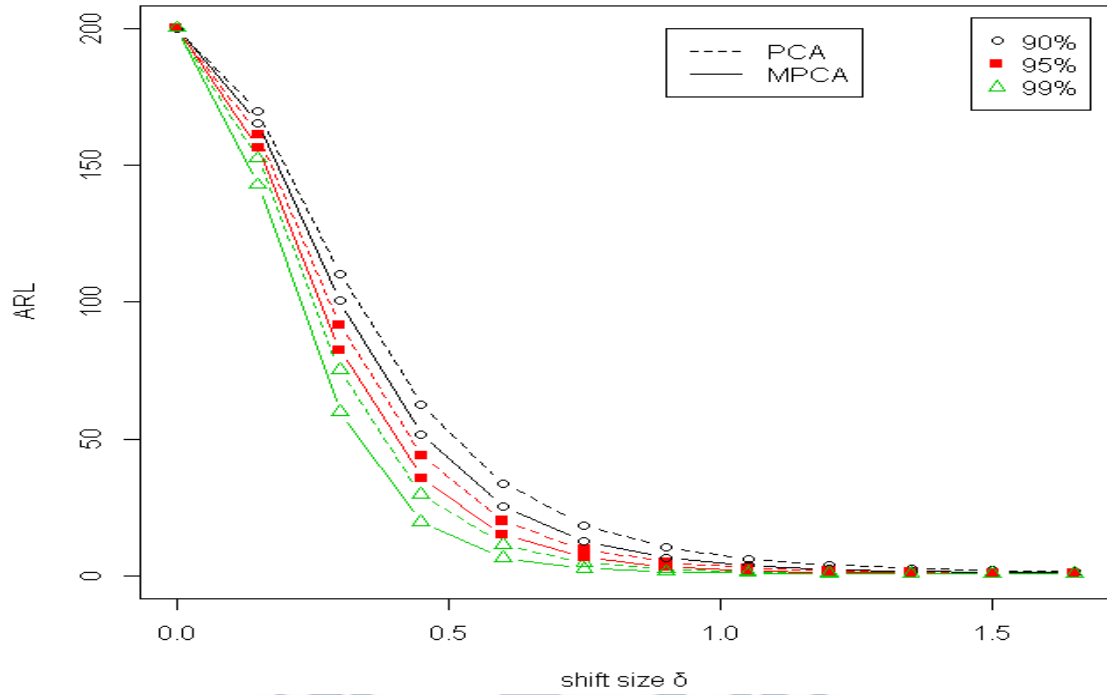
套用第三章的方法，先將  $Y$  中心化後再去除噪音之干擾所得的矩陣  $X$  其元素應相當靠近  $Z(s_i, t_j), \forall i, j$ ，再藉由 3.1 及 3.3 小節的方法來監控  $X$ 。我們用 ARL 曲線來比較 PCA 和 MPCA 兩方法；但因二者無法有完全相同的解釋比例，我們只能擇相差不大的情形來比較，例如在 4.1.1 小節中所選擇三種水準之一的 90% 其 PCA 的解釋比例為 90.86% 而 MPCA 為 89.92%；對 MPCA，我們所選擇的解釋比例皆略低於 PCA。另外，我們想了解不同的解釋比例是否會影響偵測力，於是我們將這三種水準放在同一張圖上。以符號  $ARL_0$  來表示在管制內的 ARL，此處  $ARL_0=200$ ，其所對應的假警報機率為 0.005。

#### 4.1.1 $X_1(s) + X_2(t)$ 為布朗橋+積分布朗運動之情況

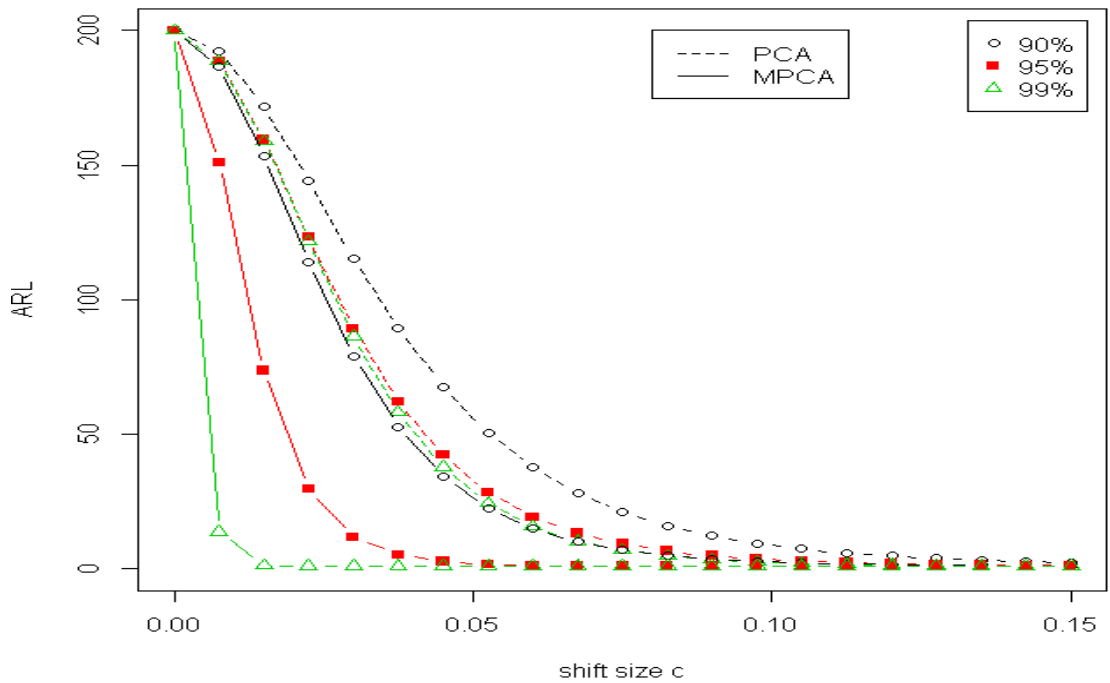
圖二呈現藉由此例之模型且  $X_1(s)$  為布朗橋及  $X_2(t)$  為積分布朗運動所模擬出來的兩個管制內之二維剖面。對於此例我們選擇的解釋比例為 90%、95%及 99%三種水準，在此設  $c = 0(0.0075)0.15$ 、 $\delta = 0(0.15)1.65$ ，且  $s$  及  $t$  的值皆為 0 和 1 之間的等分值。圖三及圖四以 ARL 曲線呈現 PCA 和 MPCCA 兩方法分別對於平均矩陣在(a)和(b)兩方式偏移下的偵測能力。我們首先觀察到 PCA 及 MPCCA 對大部分的解釋比例而言皆有解釋比例越高所對應的 ARL 會越低的情形。圖三是平均矩陣以(a)方式偏移，觀察到對於每個解釋比例而言，MPCCA 的偵測力均比 PCA 佳；而綜觀三個水準來看，則是一對接在一對後面交錯呈現。因此對於平均矩陣以(a)方式偏移時，MPCCA 表現比 PCA 來得好。圖四是平均矩陣以(b)方式垂直平移，MPCCA 表現比(a)方式更佳，由圖四觀察到 MPCCA 幾乎所有的 ARL 曲線皆在 PCA 的 ARL 曲線下方。



圖二、範例一之假設模型在情況(i)布朗橋+積分布朗運動下所模擬出來兩個管制內的二個維剖面。



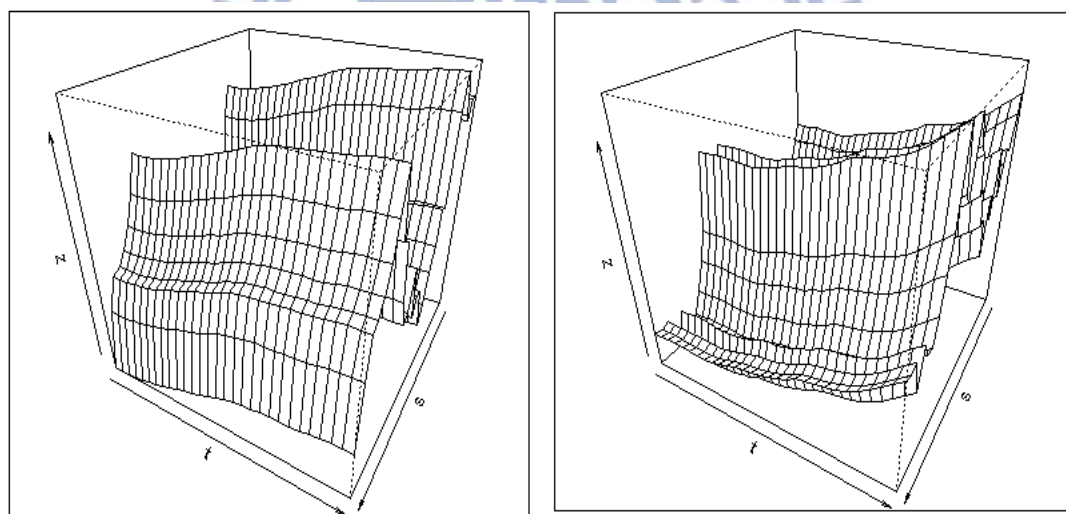
圖三、範例一之假設模型在情況(i)下以(a)方式偏移時，PCA 及 MPCA 兩方法的 ARL 曲線。



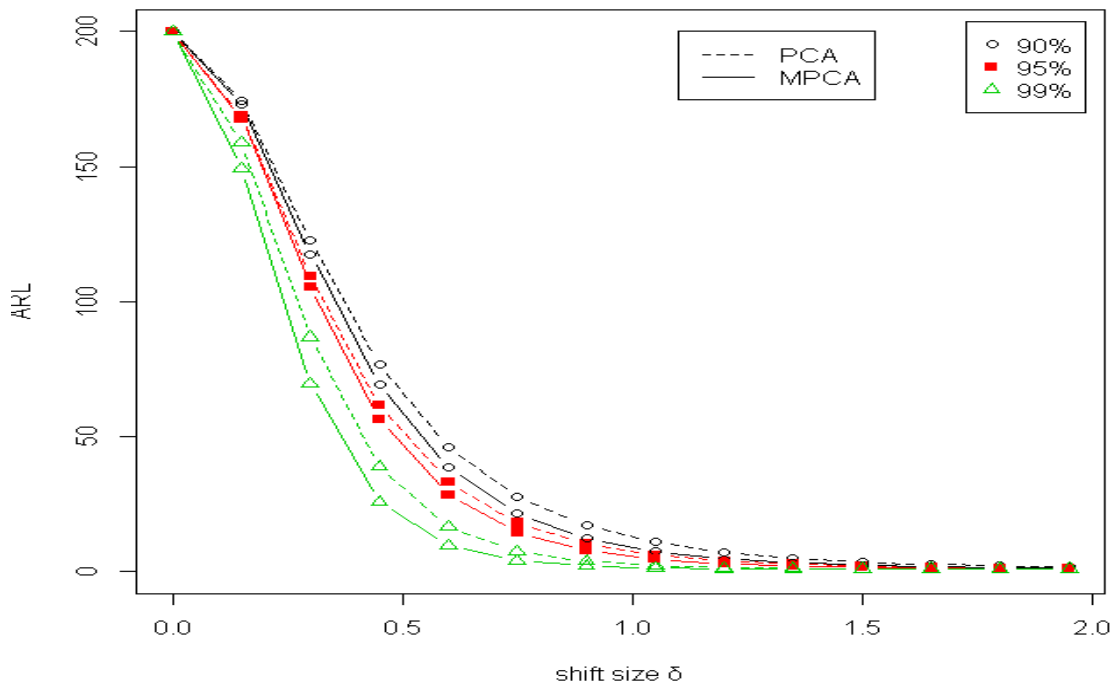
圖四、範例一之假設模型在情況(i)下以(b)方式偏移時，PCA 及 MPCA 兩方法的 ARL 曲線。

#### 4.1.2 $X_1(s)+Y_1(t)$ 為布朗運動+積分布朗運動之情況

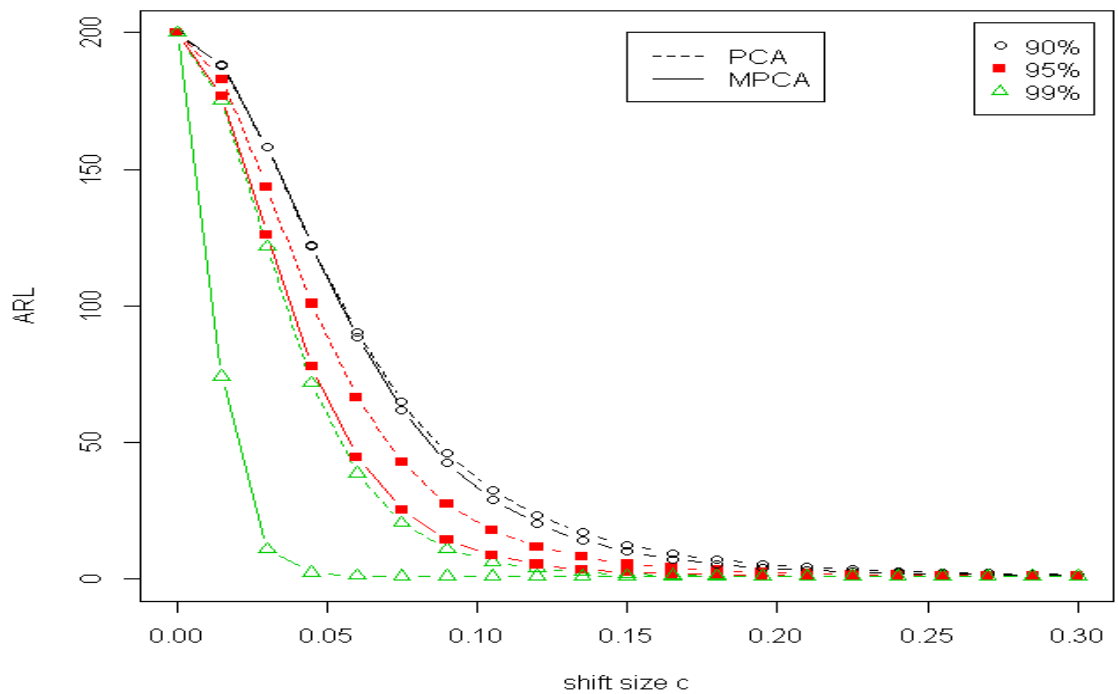
針對此情形，我們所選擇的解釋比例及  $s$  和  $t$  之設法仍相同，但  $c = 0(0.015)0.3$  及  $\delta = 0(0.15)1.95$ 。圖五、圖六及圖七依序呈現此例模型在  $X_1(s)$  為布朗運動且  $X_2(t)$  為積分布朗運動情況下所模擬出來兩個管制內的二維剖面及 PCA 和 MPCSA 兩方法對於平均矩陣以(a)及(b)兩方式偏移時的 ARL 曲線。由圖六觀察到，PCA 及 MPCSA 兩方法以(a)方式偏移所獲得的 ARL 曲線之間的關係和圖三很類似。從圖七觀察到對每個解釋比例而言，MPCSA 的偵測能力仍皆比 PCA 來得好。不過在此例 MPCSA 和 PCA 之 ARL 曲線就交錯出現，並非如圖四所呈現 MPCSA 之 ARL 曲線幾乎皆在 PCA 的下方。



圖五、範例一之假設模型在情況(ii)布朗運動+積分布朗運動下所模擬出來兩個管制內的二維剖面。



圖六、範例一之假設模型在情況(ii)下以(a)方式偏移時，PCA 及 MPCA 兩方法的 ARL 曲線。



圖七、範例一之假設模型在情況(ii)下以(b)方式偏移時，PCA 及 MPCA 兩方法的 ARL 曲線。

## 4.2 範例二

我們採用和範例一相同做法，也利用  $X_1(s_i) + X_2(t_j)$  來建構二維剖面資料

$Z(s_i, t_j), \forall i, j$ 。假設模型為  $Y_{ij} = Z(s_i, t_j) + \varepsilon_{ij}, i = 1, \dots, 20, j = 1, \dots, 20$ ，其中

$Z(s_i, t_j) = X_1(s_i) + X_2(t_j), \forall i, j$  而  $X_1(s) = Z_1 \cos \omega_1 s + Z_2 \sin \omega_1 s$  且假設

$Z_1, Z_2 \sim N(0, \sigma_1^2)$ ，同樣地  $X_2(t) = Z_3 \cos \omega_2 t + Z_4 \sin \omega_2 t$  且  $Z_3, Z_4 \sim N(0, \sigma_2^2)$ 。在此

假設  $Z_1$ 、 $Z_2$ 、 $Z_3$  和  $Z_4$  彼此間獨立。那麼我們可以很容易獲得  $\mu_Z(s, t) = 0$  及

$\text{cov}(Z(s_1, t_1), Z(s_2, t_2)) = \sigma_1^2 \cos \omega_1 (s_2 - s_1) + \sigma_2^2 \cos \omega_2 (t_2 - t_1)$ 。對於參數設定，令

$\omega_1 = 0.2, \omega_2 = 0.5$ 、 $\sigma_1 = 2, \sigma_2 = 3$  以及  $s = 0(2\pi / (2 \times 19))\pi$  和  $t = 0(2\pi / (3 \times 19))\frac{2}{3}\pi$ 。圖

八呈現藉由此例的模型所模擬出來的兩個管制內之二維剖面。接下來套用 3.1 及 3.3 小節的方法來監控這樣的二維剖面。在 MPCA 方法中我們是採用母體觀點以

及製程參數已知的情況。對於此例的偏移是藉由  $Z_1, Z_2 \sim N(0 + \delta\sigma_1, \sigma_1^2)$  及

$Z_3, Z_4 \sim N(0 + \delta\sigma_2, \sigma_2^2)$  來設定，其中  $\delta = 0(0.25)3$ 。明顯地我們可發現到，像這樣的二維剖面，其  $\text{rank}(\Sigma_X) = 4$ 。對於 PCA，讓  $k = \text{rank}(\Sigma_X) = 4$ ，而對於 MPCA，

讓  $\tilde{p} = 3$  及  $\tilde{q} = 3$ ，則發現  $X$  的重建剛好等於原本的  $X$ ，並發現  $k_1 = 4 = k$ ，進而

發現 PCA 和 MPCA 的 ARL 相同並呈現在圖九，且在此情況 PCA 和 MPCA 的解釋比例皆為 100%。因此，此情況驗證了定理三。圖十所呈現之 PCA 取  $k = 3$  (解

釋比例約為 85%) 而 MPCA 取  $\tilde{p} = 3$  及  $\tilde{q} = 2$  (解釋比例約為 70%)。在  $X_1$  和  $X_2$

的平均向量皆有偏移及  $X_1$  的平均向量有偏移而  $X_2$  的平均向量沒有偏移兩種情

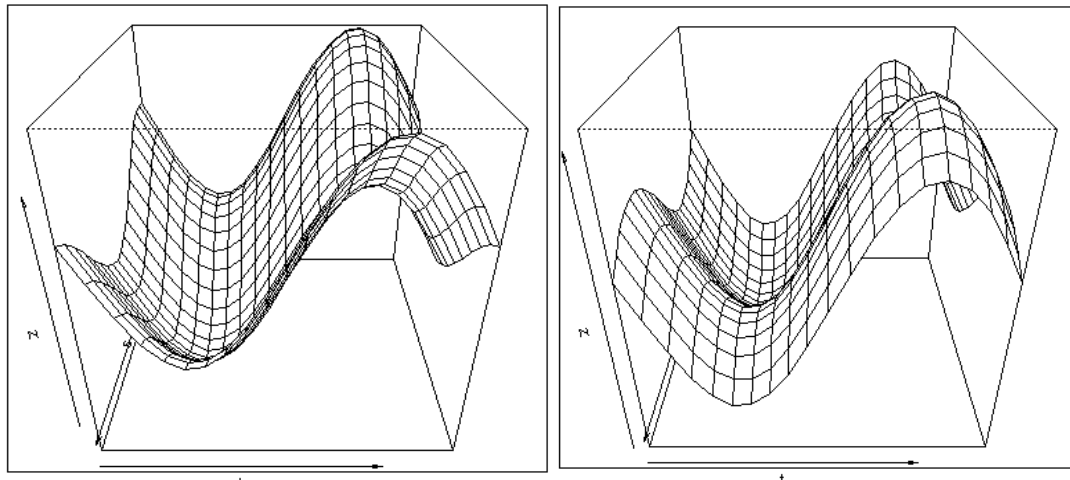
形下，計算 PCA 及 MPCA 的 ARL 曲線。由圖十可發現，當只有  $X_1$  的平均向量

偏移時，MPCA 的偵測力明顯較 PCA 好。此可能是因為  $X$  在  $\tilde{p} = 3$  及  $\tilde{q} = 3$  時可

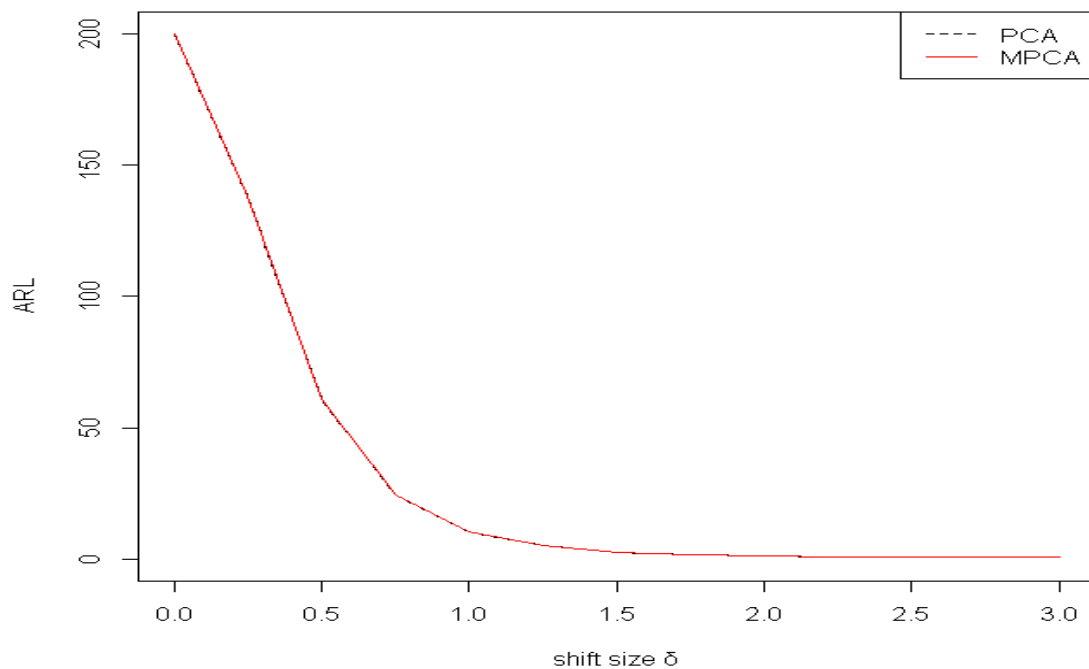
完全百分百地被重建，意味著  $\tilde{p} = 3$  就能夠反應行方向上所有的資訊；而現在

的情形是只有  $X_1$  的平均向量改變，意味著改變全在行方向上，因此 MPCA 能很快

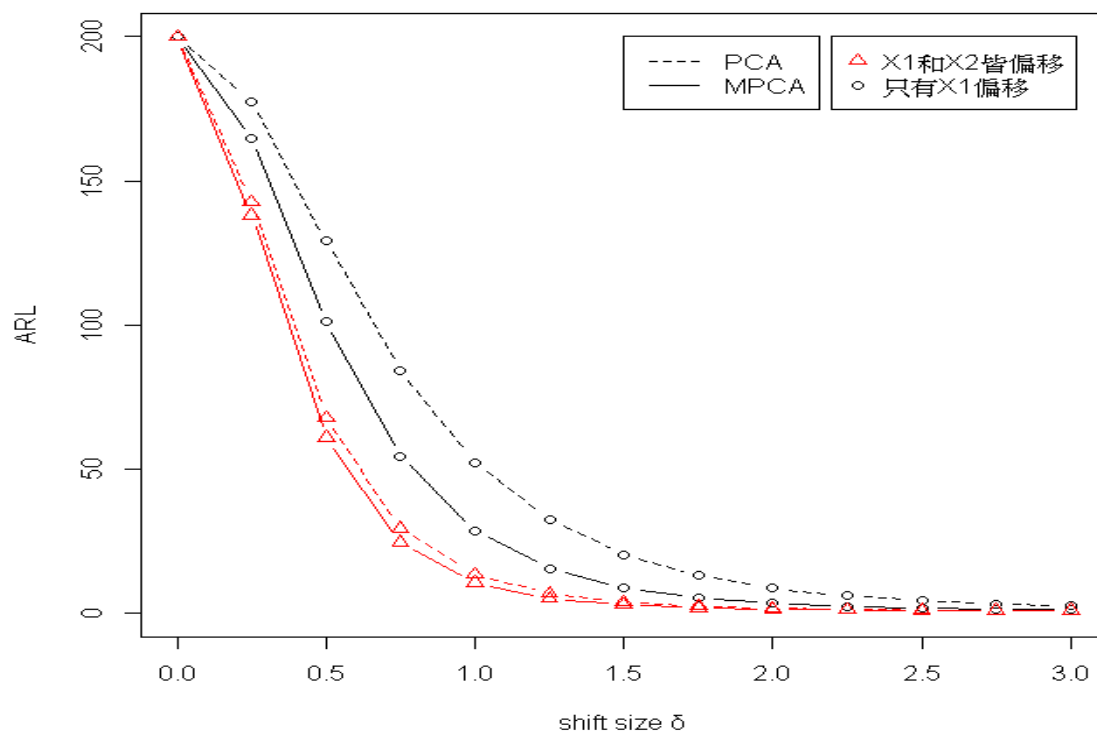
反應此變化。對於  $X_1$  和  $X_2$  的平均向量皆有偏移的情形，MPCA 的偵測力雖略比 PCA 來得好但相較於只有  $X_1$  平均向量有偏移的情形，二者相差有限。我們猜測此乃因 PCA 的解釋比例已達 85% 表示已具有大多的資訊，而 MPCA 的解釋比例只有 70%，且由上述可知行方向上的資訊已夠，進而表示列方向上的資訊還不足。



圖八、由 4.2 小節之假設模型所模擬出來兩個管制內的二維剖面。



圖九、PCA 在  $k=4$  及 MPCA 在  $\tilde{p}=3$  及  $\tilde{q}=3$  時個別的 ARL 曲線，兩者完全重合。



圖十、PCA (解釋比例約為 85%) 和 MPCA (解釋比例約為 70%) 在  $X_1$  和  $X_2$  的平均向量皆有偏移及只有  $X_1$  的平均向量有偏移兩種情形的 ARL 曲線。

### 4.3 PCA 與 MPCA 的執行時間

在進行製程監控前，對 PCA 而言，需先對  $\Sigma_x$  做特徵根分解，以得到它的特徵值及特徵向量；對於 MPCA 而言，在母體觀點下需先獲得矩陣  $A$ 、 $B$ 、 $T$  及  $G$ 。底下的執行時間是使用 R 程式在硬體設備 Intel Q8300 處理器 2.5 GHz 且高速快取記憶體 (Cache RAM) 容量為 2G 之下所需之 CPU 執行時間。當  $p$  及  $q$  增大時，PCA 在進行特徵根分解時需花較多時間，且當  $p = q = 80$  時，會有記憶體 (memory space) 容量不足而沒辦法進行分解的情形。至於純粹的第二階段線上監控，PCA 及 MPCA 的計算時間則差不多。在不同的維度  $p$  和  $q$  下，表一及表二分別呈現 PCA 及 MPCA 在範例一及範例二的執行時間，此執行時間 (單位：秒)



包括在監控前需獲得參數的時間。由表一及表二我們觀察到，PCA 在此二範例的執行時間類似；反觀 MPCSA 在兩範例的執行時間有明顯差異。我們發現此乃因範例一之疊代演算收斂較久並且利用  $2D^2$  PCA 求算初始值時也耗費較多時間，因為此範例需利用到判斷式故需多花一些時間。總而言之，對於執行時間而言，除了維度  $p$  和  $q$  很小的幾個情形外，MPCSA 在計算速度上優於 PCA；而對於記憶量而言，PCA 在做特徵分解時所需的記憶量遠大於 MPCSA 所需。

表一、PCA 及 MPCSA 在範例一的 CPU 執行時間 (單位：秒)

	$p, q = 20$	$p, q = 30$	$p, q = 40$	$p, q = 50$	$p, q = 60$	$p, q = 70$	$p, q = 80$
PCA	0.36	2.03	10.02	35.92	105.93	275.69	----
MPCSA	2.14	4.02	10.99	27.94	67.05	209.53	404.54

表二、PCA 及 MPCSA 在範例二的 CPU 執行時間 (單位：秒)

	$p, q = 20$	$p, q = 30$	$p, q = 40$	$p, q = 50$	$p, q = 60$	$p, q = 70$	$p, q = 80$
PCA	0.42	2.06	12.58	35.95	104.3	258.13	----
MPCSA	1.30	1.87	3.30	6.83	13.80	37.67	73.11

#### 4.4 晶圓應用範例

藉由模擬方式來模擬晶圓氧化厚度，首先利用球方程式生成一個  $30 \times 29$  的矩陣來模擬晶圓氧化目標厚度，以符號  $T_0$  表之，再藉由範例一之模型來納入晶圓與晶圓之間的變異，最後加上量測誤差  $\varepsilon$  而得  $Y_{30 \times 29}$ ，其模型為  $Y_{ij} = T_{0ij} + Z(s_i, t_j) + \varepsilon_{ij}$ ,  $i = 1, \dots, 30, j = 1, \dots, 29$ 。因晶圓為圓形，我們將矩陣  $Y$  所對應之晶圓缺值部分的數值改為 0。圖十一及圖十二分別呈現此模型在 4.1.1 和 4.1.2 小節所述之兩種情況下並藉由將部分值改為 0 的方法所模擬出來兩個管制內的二維剖面。對於製程

監控，我們考慮兩種情況(I)藉由上述模型模擬出方形資料 $Y$ 及(II)將(I)所得的方形資料 $Y$ 內所對應之晶圓缺值部分的數值改為0，此種情況較貼近實際的晶圓資料。對於情況(II)而言，因有些部分的數值是0，所以無法經由推導求得在母體觀點中所需用到的 $\sum_x$ ；因此，我們採用樣本觀點進行製程監控之模擬。

**模擬 $T_0$ 及 $Y$ 的作法：**

假設在管制內的晶圓氧化厚度為在4~6奈米內，且假設晶圓氧化目標厚度是在4.5~5.2奈米內。令下面式(6)的 $R=171.9571$ 、 $s$ 和 $t$ 的值皆為0和1之間的等分值及由式(6)所得到的值中最小的值為 $T_{\min}$ ，

$$-\sqrt{R^2 - (29(s_i - 0.5))^2 - (28(t_j - 0.5))^2} \quad (6)$$

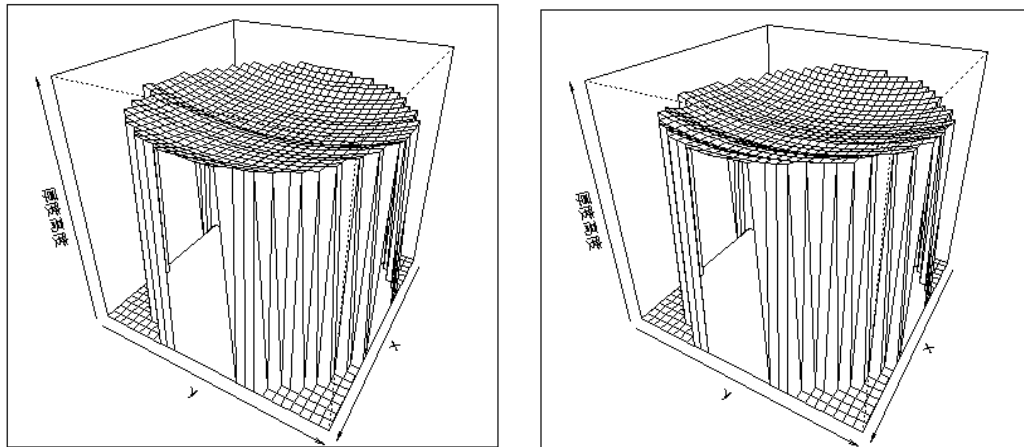
其中 $i=1, \dots, 30, j=1, \dots, 29$ 。藉由將式(6)所有值減掉 $T_{\min}$ 再加上4.5後所得到的值來當作 $T_{0,ij}, \forall i, j$ 。利用4.1.1小節的情況來模擬 $Z(s_i, t_j), \forall i, j$ ，則藉由模型可得 $Y$ 。

**在情況(I)及情況(II)下監控 $Y$ 的作法：**

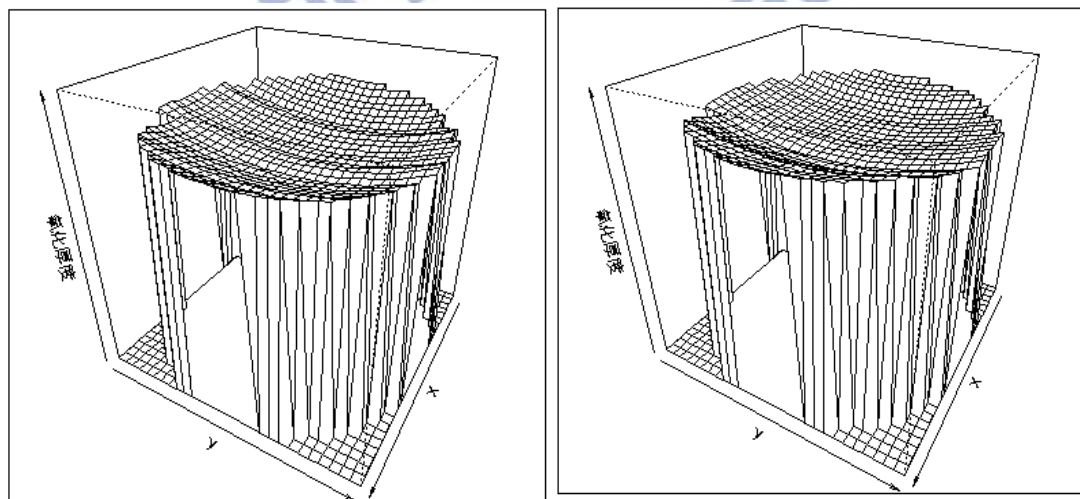
在情況(I)下，藉由上述生成一組300筆 $Y$ ；將這組資料一一中心化及平滑化而得300筆在管制內的 $X$ ；而情況(II)則將在情況(I)下所得的300筆管制內的 $X$ 中之晶圓缺值部分的數值改為0來當作管制內的一組晶圓資料。對於PCA和MPCA兩方法的解釋比例無論情況(I)或(II)皆設定約為96%。在PCA方面，情況(I)和(II)分別藉由各自的300筆管制內資料來求得 $\hat{\Sigma}_X^{(I)}$ 和 $\hat{\Gamma}^{(I)}$ 及 $\hat{\Sigma}_X^{(II)}$ 和 $\hat{\Gamma}^{(II)}$ 。在MPCA方面，為了使解釋比例均在96%附近，對於情況(I)採用 $\tilde{p}=11$ 和 $\tilde{q}=10$ 而情況(II)採用 $\tilde{p}=12$ 和 $\tilde{q}=12$ ，並分別將各自的300筆管制內資料套用樣本觀點的GLRAM演算法來求得 $\hat{A}_0^{(I)}$ 、 $\hat{B}_0^{(I)}$ 、 $\hat{T}_0^{(I)}$ 和 $\hat{G}_0^{(I)}$ 及 $\hat{A}_0^{(II)}$ 、 $\hat{B}_0^{(II)}$ 、 $\hat{T}_0^{(II)}$ 和 $\hat{G}_0^{(II)}$ 。

接下來先針對情況(I)模擬三組資料，分別為管制內及在4.1小節提及的兩種偏移方式，每組資料皆有50筆，其中兩偏移方式所需的參數設為 $\delta=1.5$ 和 $c=0.0375$ 。情況(II)是將情況(I)的三組方形資料中之晶圓缺值部分的數值改為0

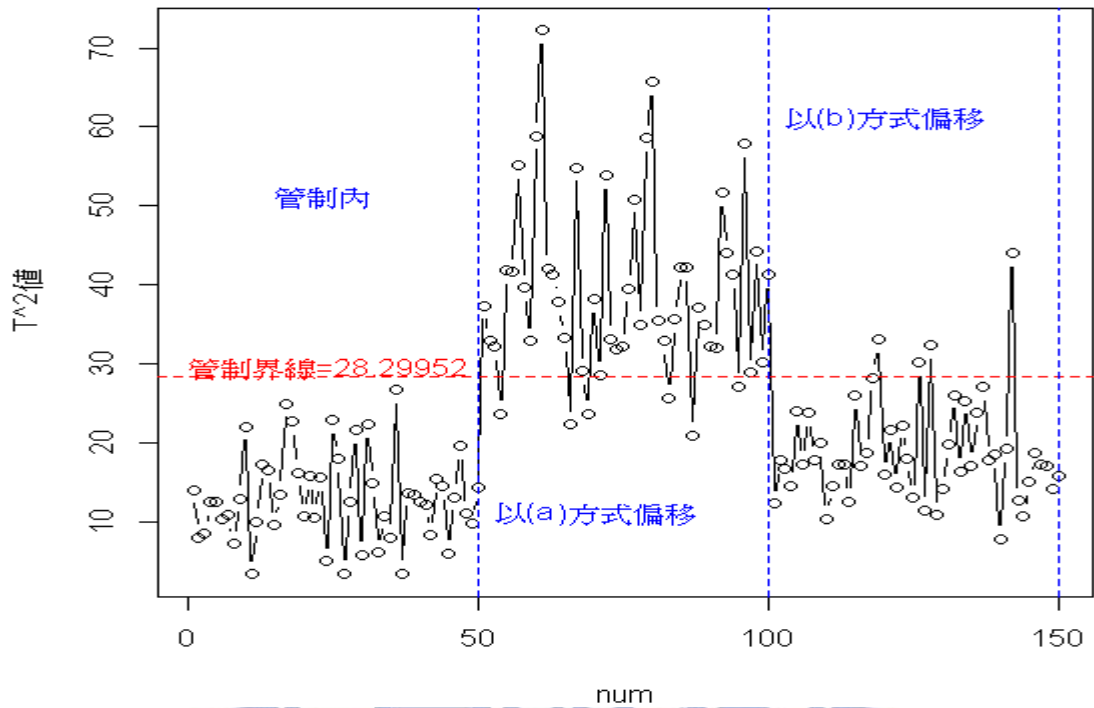
而得的 150 筆資料。圖十三和圖十四及圖十五和圖十六分別為在情況(I)及情況(II)下藉由 PCA 和 MPCA 方法來監控兩情況各自的 150 筆資料的管制圖。由圖十三和圖十五其最右方發現，無論情況(I)或(II)，PCA 對平均矩陣以垂直方式偏移之偵測力很差。由圖十四和圖十六發現雖然情況(II)沒符合監控方法的前提假設，但對於以(a)或(b)方式偏移的偵測力和情況(I)相似。



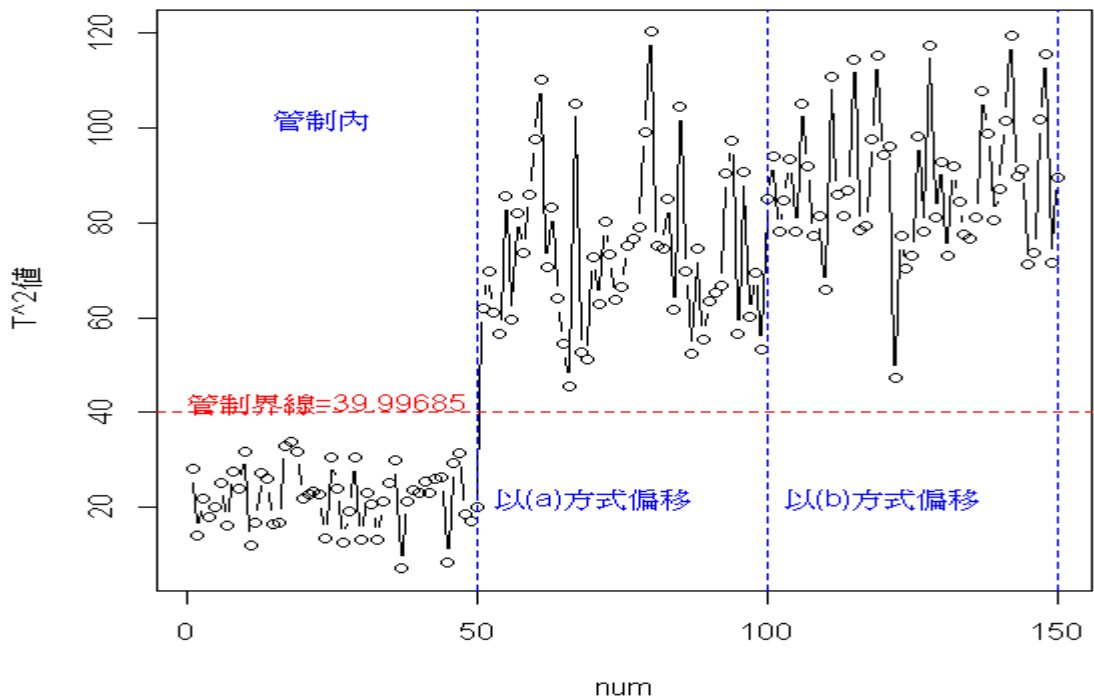
圖十一、晶圓氧化的目標厚度在 4.1.1 小節之情形下所模擬出來的兩個管制內的二維剖面。



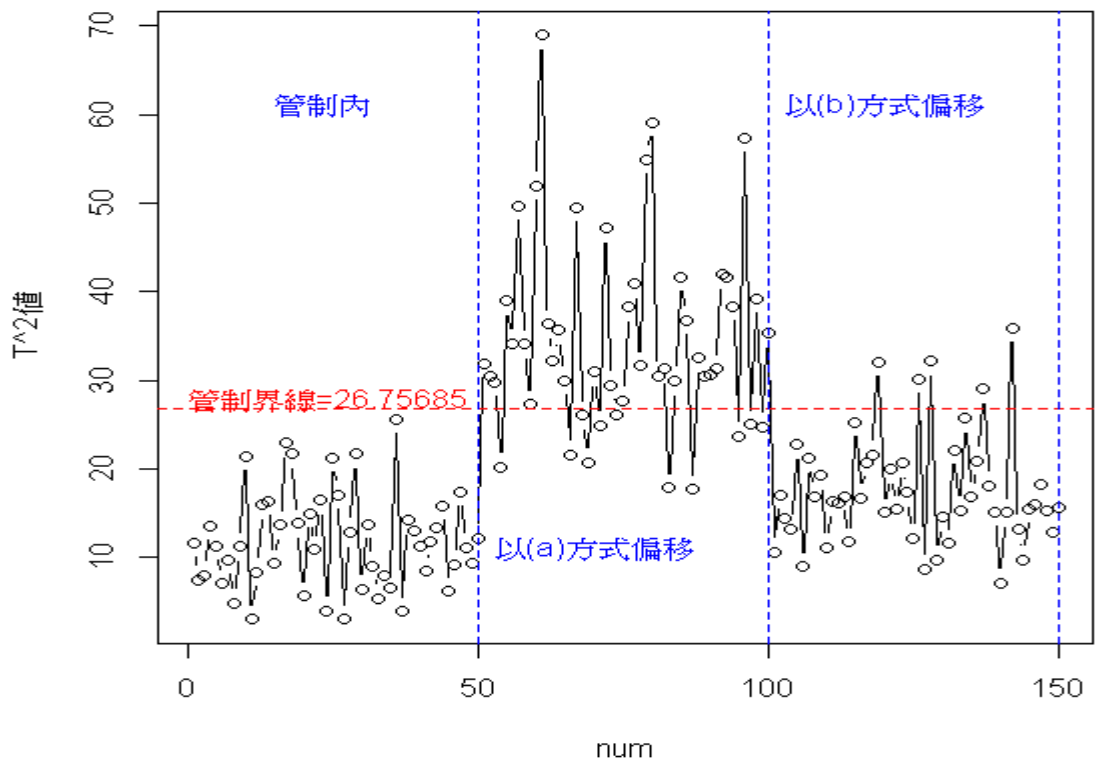
圖十二、晶圓氧化的目標厚度在 4.1.2 小節之情形下所模擬出來的兩個管制內的二維剖面。



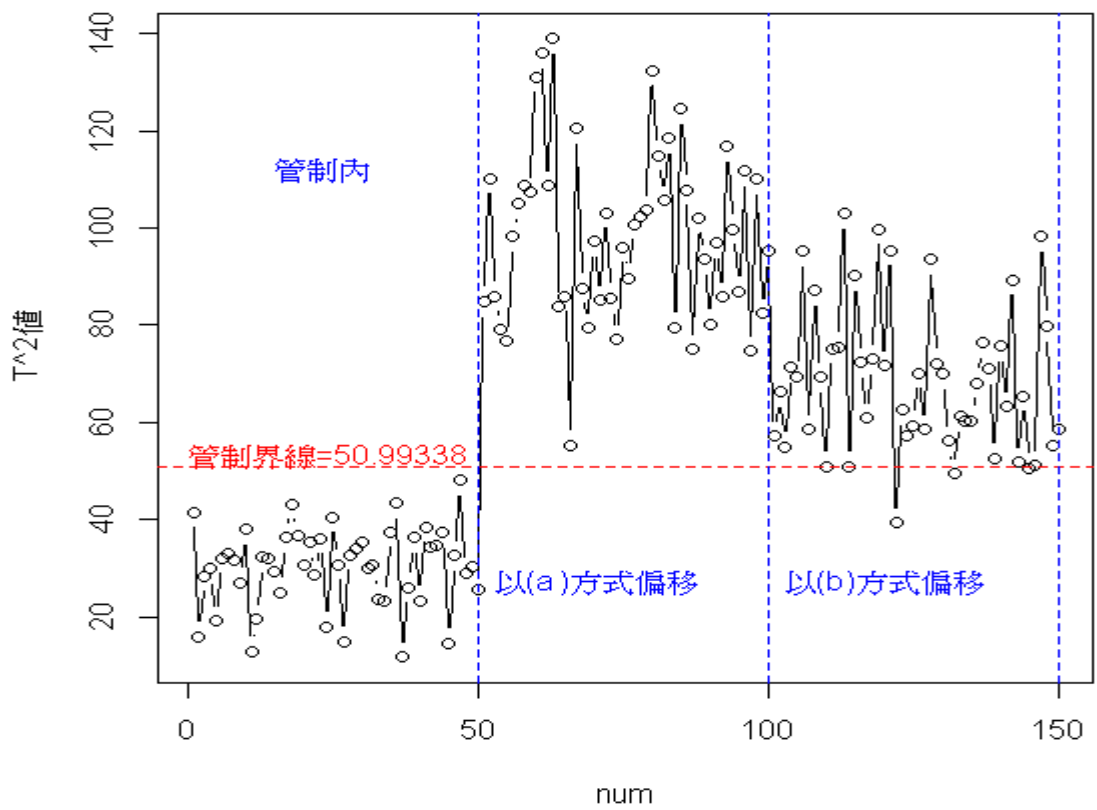
圖十三、在情況(I)下之 150 筆測試資料用 PCA 方法監控的管制圖。



圖十四、在情況(I)下之 150 筆測試資料用 MPCA 方法監控的管制圖。



圖十五、在情況(II)下之 150 筆測試資料用 PCA 方法監控的管制圖。



圖十六、在情況(II)下之 150 筆測試資料用 MPCA 方法監控的管制圖。

## 第五章 結論和未來展望

### 5.1 結論

在先前的文獻中，較少關於二維剖面監控之研究，Gardner et al. (1997) 對於空間特徵提出如何建構這樣的曲面，並在不考慮晶圓與晶圓之間具有個體差異性下，利用四個衡量標準來量測新進資料和基準曲面之間的差距，進而偵測研磨過程中哪邊出狀況。本篇文章主要探究在剖面與剖面之間存在個體差異性下二維剖面之監控，並運用 PCA 及 MPCA 提出第二階段的監控策略。對於偵測能力，我們利用平均連串長度 ARL 來評估 MPCA 及 PCA 兩方法的績效，大體上 MPCA 方法較 PCA 方法來得好。MPCA 的兩個基底矩陣分別抓取行方向及列方向上的資訊，因此當新進的二維剖面資料若只有在行或列方向上有變化且 MPCA 的基底矩陣能抓取所有行方向或列方向上的資訊時，則能較快偵測出平均矩陣已發生變化。而在建構正交基底時，PCA 所需的參數會較 MPCA 來得多，此極可能就是 MPCA 優於 PCA 的緣故。對於執行時間方面，當  $p$  及  $q$  皆很大時，將二維剖面拉成一維向量其共變異矩陣之維度會很大，使得 R 程式在對其做特徵根分解時需要較多計算時間；反之 MPCA 需特徵分解的共變異矩陣的維度已變小，則所需的計算時間會較少。因此相較二者我們比較建議使用 MPCA 來執行第二階段的製程監控。

### 5.2 未來展望

本篇文章只針對第二階段提出二維剖面監控策略，並沒有探討第一階段二維剖面的監控，未來可探究第一階段的監控。在第四章中，我們藉由  $X_1(s) + X_2(t)$  來建構  $Z(s,t)$  這或許太簡單化，未來可考慮較實際或較複雜的  $Z(s,t)$ 。

## 參考文獻

- [1] De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000a), “A Multilinear Singular Value Decomposition,” *SIAM Journal on Matrix Analysis and Applications*, 21, pp. 1253–1278.
- [2] De Lathauwer, L., De Moor, B., and Vandewalle, J. (2000b), “On the Best Rank-1 and Rank-( $R_1, R_2, \dots, R_N$ ) Approximation of Higher-Order Tensors,” *SIAM Journal on Matrix Analysis and Applications*, 21, pp. 1324–1342.
- [3] Ding, Y., Zeng, L., and Zhou, S. (2006), “Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes,” *Journal of Quality Technology*, 38, pp. 199–216.
- [4] Eyvazian, M., Noorossan, R., Saghaei, A., and Amiri, A. (2011), “Phase II Monitoring of Multivariate Multiple Linear Regression Profiles,” *Quality and Reliability Engineering International*, 27, pp. 281–296.
- [5] Gardner, M. M., Lu, J. C., Gyurcsik, R. S., Wortman, J. J., Hornung, B. E., Heinisch, H. H., Rying, E. A., Rao, S., Davis, J. C., and Mozumder, P. K. (1997), “Equipment Fault Detection Using Spatial Signatures,” *IEEE Transactions on Components Packaging and Manufacturing Technology*, 20, pp. 295–303.
- [6] Hung, H., Wu, P. S., Tu, I. P., and Huang, S. Y. (2012), “On Multilinear Principal Component Analysis of Order-Two Tensors,” *Biometrika*, pp. 1–15.
- [7] Jensen, W. A. and Birch, J. B. (2009), “Profile Monitoring via Nonlinear Mixed Models,” *Journal of Quality Technology*, 41, pp. 18–34.
- [8] Jensen, W. A., Birch, J. B., and Woodall, W. H. (2008), “Monitoring Correlation Within Linear Profiles Using Mixed Models,” *Journal of Quality Technology*, 40, pp. 167–183.
- [9] Kang, L. and Albin, S. L. (2000), “On-Line Monitoring When the Process Yields a Linear Profile,” *Journal of Quality Technology*, 32, pp. 418–426.
- [10] Kazemzadeh, R. B., Noorossana, R., and Amiri, A. (2008), “Phase I Monitoring of Polynomial Profiles,” *Communications in Statistics: Theory and Methods*, 37, pp. 1671–1686.
- [11] Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003), “On the Monitoring of Linear Profiles,” *Journal of Quality Technology*, 35, pp. 317–328.
- [12] Kolda, T. G. and Bader, B. W. (2009), “Tensor Decompositions and Applications,” *SIAM Review*, 51(3), pp. 455–500.
- [13] Li, B., Kim, M. K., and Altman, N. (2010), “On Dimension Folding of Matrix- or Array-valued Statistical Objects,” *Annals of Statistics*, 38, pp. 1094–1121.

- [14] Lu, H., Plataniotis, K. N., and Venetsanopoulos, A. N. (2008), “MPCA: Multilinear Principal Component Analysis of Tensor Objects,” *IEEE Transactions on Neural Networks*, 19, pp. 18–39.
- [15] Mahmoud, M. A., Parker, P. A., Woodall, W. H., and Hawkins, D. M. (2007), “A Change Point Method for Linear Profile Data,” *Quality and Reliability Engineering International*, 23, pp. 247–268.
- [16] Qiu, P., Zou, C., and Wang, Z. (2010), “Nonparametric Profile Monitoring by Mixed Effects Modeling,” *Technometrics*, 52:3, pp. 265–277.
- [17] Saghaei, A., Mehrjoo, M., and Amiri, A. (2009), “A CUSUM-Based Method for Monitoring Simple Linear Profiles,” *The International Journal of Advanced Manufacturing Technology*, 45, pp. 1252–1260.
- [18] Shiau, J. - J. H. and Weng, Z. - P. (2004), “Profile Monitoring by Nonparametric Regression,” Technical Report, Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.
- [19] Shiau, J. - J. H., Huang, H. L., Lin, S.-H., and Tsai, M. - Y. (2009), “Monitoring Nonlinear Profiles with Random Effects by Nonparametric Regression,” *Communications in Statistics - Theory and Methods*, 38, pp. 1664–1679.
- [20] Williams, J. D., Woodall, W. H., and Birch, J. B. (2007), “Statistical Monitoring of Nonlinear Product and Process Quality Profiles,” *Quality and Reliability Engineering International*, 23, pp. 925–941.
- [21] Yang, J., Zhang, D., Frangi, A. F., and Yang, J. Y. (2004), “Two-dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, pp. 131–137.
- [22] Ye, J. (2005), “Generalized Low Rank Approximations of Matrices,” *Machine Learning*, 61, pp. 167–191.
- [23] Zhang, D. and Zhou, Z. H. (2005), “(2D) 2PCA: Two-directional Two-dimensional PCA for Efficient Face representation and recognition,” *Neurocomputing*, 69, pp. 224–231.
- [24] Zou, C., Ning, X., and Tsung, F. (2012), “LASSO-Based Multivariate Linear Profile Monitoring,” *Annals of Operations Research*, 192, pp. 3–19.
- [25] Zou, C., Tsung, F., and Wang, Z. (2007), “Monitoring General Linear Profiles Using Multivariate EWMA Schemes,” *Technometrics*, 49, pp. 395–408.
- [26] Zou, C., Zhang, Y., and Wang, Z. (2006), “Control Chart Based on Change-Point Model for Monitoring Linear Profiles,” *IIE Transactions*, 38, pp. 1093–1103.



## 附錄

**定理二** 在預先給定的  $\tilde{p}$  及  $\tilde{q}$  下，令  $\tilde{A}_{0,p \times \tilde{p}}$ 、 $\tilde{B}_{0,q \times \tilde{q}}$  及  $\tilde{U}_{0,\tilde{p} \times \tilde{q}}$  為式(2)的解。則

$$(a') \text{Prob}(\tilde{U}_0 = \tilde{A}_0^T (X - \mu) \tilde{B}_0) = 1 ;$$

(b')  $\{\tilde{A}_0, \tilde{B}_0\}$  為  $\max_{\substack{A \in O_{p \times \tilde{p}} \\ B \in O_{q \times \tilde{q}}}} E \|A^T (X - \mu) B\|_F^2$  的解；

(c')  $\tilde{A}_0$  的行向量是由  $E[(X - \mu) P_{\tilde{B}_0} (X - \mu)^T]$  的前  $\tilde{p}$  大的特徵值所對應的特徵向

量依序排置而成， $\tilde{B}_0$  的行向量是由  $E[(X - \mu)^T P_{\tilde{A}_0} (X - \mu)]$  的前  $\tilde{q}$  大的特徵值所

對應的特徵向量依序排置而成，其中  $P_{\tilde{A}_0} \equiv \tilde{A}_0 (\tilde{A}_0^T \tilde{A}_0)^{-1} \tilde{A}_0^T$  及

$$P_{\tilde{B}_0} \equiv \tilde{B}_0 (\tilde{B}_0^T \tilde{B}_0)^{-1} \tilde{B}_0^T .$$

### 定理二之證明

對於性質 (a') ~ (c') 的證明，在不失一般性下可令  $\mu = 0$ 。

### 性質 (a') 之證明：

$$\begin{aligned} E \|X - AUB^T\|_F^2 &= E[\text{trace}((X - AUB^T)(X - AUB^T)^T)] \\ &= E[\text{trace}(XX^T - XBU^T A^T - AUB^T X^T + AUB^T BU^T A^T)] \\ &= E[\text{trace}(XX^T)] + E[\text{trace}(UU^T) - 2\text{trace}(UB^T X^T A)] \end{aligned} \quad (7)$$

令  $M = B^T X^T A$ ，則式(7)可寫成  $E[\sum_{i=1}^p \sum_{j=1}^q X_{ij}^2] + E[\sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} U_{ij}^2 - 2\sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} U_{ij} M_{ji}]$

$$= E[\sum_{i=1}^p \sum_{j=1}^q X_{ij}^2] + E[\sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} (U_{ij} - M_{ji})^2] - E[\sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} M_{ji}^2] \quad (8)$$

假設已找到使得式(8)最小的矩陣  $\tilde{A}_0$  及  $\tilde{B}_0$ ，而現在要找尋  $U$  使得式(8)最小，則

$E[\sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} X_{ij}^2]$  及  $E[\sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} M_{ji}^2]$  可視為常數，那麼可得

$$E[\sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} (\tilde{U}_{0,ij} - M_{ji})^2] = \sum_{i=1}^{\tilde{p}} \sum_{j=1}^{\tilde{q}} E(\tilde{U}_{0,ij} - M_{ji})^2 = 0, \text{ 則得 } Prob(\tilde{U}_{0ij} = M_{ji}) = 1, \forall i, j.$$

因此，可證得  $Prob(\tilde{U}_0 = \tilde{A}_0^T X \tilde{B}_0) = 1$ 。

**性質(b')之證明：**由性質(a')得知  $Prob(\tilde{U}_0 = \tilde{A}_0^T X \tilde{B}_0) = 1$ ，則

$$\begin{aligned} E\|X - \tilde{A}_0 \tilde{U}_0 \tilde{B}_0^T\|_F^2 &= E\|X - \tilde{A}_0 \tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T\|_F^2 \\ &= E[\text{trace}((X - \tilde{A}_0 \tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T)(X - \tilde{A}_0 \tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T)^T)] \\ &= E[\text{trace}(XX^T) - \text{trace}(X \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0 \tilde{A}_0^T) - \text{trace}(\tilde{A}_0 \tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T X^T) \\ &\quad + \text{trace}(\tilde{A}_0 \tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0 \tilde{A}_0^T)] \\ &= E[\text{trace}(XX^T) - \text{trace}(\tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0) - \text{trace}(\tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0) \\ &\quad + \text{trace}(\tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0)] \\ &= E[\text{trace}(XX^T) - \text{trace}(\tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0)] \\ &= E[\|X\|_F^2 - \|\tilde{A}_0^T X \tilde{B}_0\|_F^2] \\ &= E\|X\|_F^2 - E\|\tilde{A}_0^T X \tilde{B}_0\|_F^2 \end{aligned} \tag{A1}$$

因  $\tilde{A}_0$ 、 $\tilde{B}_0$  及  $\tilde{U}_0$  會使得  $E\|X - \tilde{A}_0 \tilde{U}_0 \tilde{B}_0^T\|_F^2$  最小且  $E\|X\|_F^2$  可視為常數，因此，可證得  $\tilde{A}_0$  及  $\tilde{B}_0$  會使得  $E\|\tilde{A}_0^T X \tilde{B}_0\|_F^2$  為最大。

**性質(c')之證明：**已知  $P_{\tilde{A}_0} = \tilde{A}_0(\tilde{A}_0^T \tilde{A}_0)^{-1} \tilde{A}_0^T$  又  $\tilde{A}_0^T \tilde{A}_0 = I_{\tilde{p}}$ ，則  $P_{\tilde{A}_0} = \tilde{A}_0 \tilde{A}_0^T$ 。同樣地，

已知  $P_{\tilde{B}_0} = \tilde{B}_0(\tilde{B}_0^T \tilde{B}_0)^{-1} \tilde{B}_0^T$  又  $\tilde{B}_0^T \tilde{B}_0 = I_{\tilde{q}}$ ，則  $P_{\tilde{B}_0} = \tilde{B}_0 \tilde{B}_0^T$ 。由性質(b')可知， $\tilde{A}_0$  及  $\tilde{B}_0$  會

使得  $E\|\tilde{A}_0^T X \tilde{B}_0\|_F^2$  最大，且

$$\begin{aligned}
E\left\|\tilde{A}_0^T X \tilde{B}_0\right\|_F^2 &= E[\text{trace}((\tilde{A}_0^T X \tilde{B}_0)(\tilde{A}_0^T X \tilde{B}_0)^T)] \\
&= E[\text{trace}(\tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0)] \\
&= \text{trace}(\tilde{A}_0^T E[X P_{\tilde{B}_0} X^T] \tilde{A}_0)
\end{aligned}$$

因此，給定  $\tilde{B}_0$  下，則  $\tilde{A}_0$  使得  $E\left\|\tilde{A}_0^T X \tilde{B}_0\right\|_F^2$  最大等同於使得  $\text{trace}(\tilde{A}_0^T E[X P_{\tilde{B}_0} X^T] \tilde{A}_0)$  最大，間接獲得矩陣  $\tilde{A}_0$  包括  $E[X P_{\tilde{B}_0} X^T]$  的前  $\tilde{p}$  大的特徵值所對應的特徵向量。

因任意矩陣與其轉置矩陣有相同的佛賓尼斯範數，我們同理可得在給定  $\tilde{A}_0$  下使得  $E\left\|\tilde{A}_0^T X \tilde{B}_0\right\|_F^2$  最大的  $\tilde{B}_0$  其包括  $E[X^T P_{\tilde{A}_0} X]$  前  $\tilde{q}$  大的特徵值所對應的特徵向量。

故性質 (c') 得證。

接著要證明性質二，先敘述兩個需用到的克羅內克積之性質，如下。

**克羅內克積的性質** 假設矩陣  $M_1$  維度為  $m_1 \times n_1$ 、矩陣  $M_2$  維度為  $m_2 \times n_2$ 、矩陣  $M_3$  維度為  $n_1 \times m_3$  及矩陣  $M_4$  維度為  $n_2 \times m_4$ ，則

1.  $(M_1 \otimes M_2)^T = M_1^T \otimes M_2^T$ 。
2.  $(M_1 \otimes M_2)(M_3 \otimes M_4) = M_1 M_3 \otimes M_2 M_4$ 。

**性質二** 假設  $X$  在管制內，且  $\text{vec}(X) \sim N_{pq}(\mathbf{0}_{pq \times 1}, \sum_X pq \times pq)$ ，則

$\text{vec}(U_X) \sim N_{\tilde{p}\tilde{q}}(\mathbf{0}_{\tilde{p}\tilde{q} \times 1}, T_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}})$ ，其中  $T = \text{cov}(\text{vec}(U_X))$ 。

**性質二之證明：**由定理二的性質 (a') 可得  $\text{Prob}(U_X = \tilde{A}_0^T X \tilde{B}_0) = 1$ ，亦即事件

$\{U_X \neq \tilde{A}_0^T X \tilde{B}_0\}$  的機率為零。對  $U_X = \tilde{A}_0^T X \tilde{B}_0$  兩邊取  $\text{vec}$  而得

$\text{vec}(U_X) = (\tilde{B}_0^T \otimes \tilde{A}_0^T) \text{vec}(X) = (\tilde{B}_0 \otimes \tilde{A}_0)^T \text{vec}(X)$ 。因  $\text{vec}(X) \sim N_{pq}(\mathbf{0}, \sum_X)$ ，則

$\text{vec}(U_X) \sim N_{\tilde{p}\tilde{q}}(\mathbf{0}, T)$ ，其中  $T = (\tilde{B}_0 \otimes \tilde{A}_0)^T \sum_X (\tilde{B}_0 \otimes \tilde{A}_0)$ 。

**定理三** 假設對於特定的  $p_0$  及  $q_0$ ， $X = A_0 U_X B_0^T$ ，並假設式(4)中主成分個數

$k = \text{rank}(\Sigma_X)$ 。那麼當  $\tilde{p} = p_0$  和  $\tilde{q} = q_0$  時，則式(5)會等於式(4)。

在母體觀點下，定理三的  $\tilde{p} = p_0$  和  $\tilde{q} = q_0$  表示  $\tilde{A}_0 = A_0$  及  $\tilde{B}_0 = B_0$ 。在證明定理三前先提出四個引理。

**引理一**  $(\tilde{B}_0 \otimes \tilde{A}_0)^T (\tilde{B}_0 \otimes \tilde{A}_0) = I_{\tilde{p}\tilde{q}}$ ，其中矩陣  $\tilde{B}_0 \otimes \tilde{A}_0$  維度為  $pq \times \tilde{p}\tilde{q}$ 。

**引理一之證明**：藉由克羅內克積性質及  $\tilde{A}_0^T \tilde{A}_0 = I_{\tilde{p}}$  和  $\tilde{B}_0^T \tilde{B}_0 = I_{\tilde{q}}$ ，可得

$$(\tilde{B}_0 \otimes \tilde{A}_0)^T (\tilde{B}_0 \otimes \tilde{A}_0) = (\tilde{B}_0^T \otimes \tilde{A}_0^T)(\tilde{B}_0 \otimes \tilde{A}_0) = \tilde{B}_0^T \tilde{B}_0 \otimes \tilde{A}_0^T \tilde{A}_0 = I_{\tilde{q}} \otimes I_{\tilde{p}} = I_{\tilde{p}\tilde{q}}。$$

由此可直接得到下述引理二。

**引理二**  $\text{rank}(\tilde{B}_0 \otimes \tilde{A}_0) = \tilde{p}\tilde{q}$ ，其中矩陣  $\tilde{B}_0 \otimes \tilde{A}_0$  維度為  $pq \times \tilde{p}\tilde{q}$  且  $\tilde{p}\tilde{q} \leq pq$ 。

**引理三** 在定理三的假設下， $\text{rank}(\Sigma_X) = \text{rank}(T)$ 。

**引理三之證明**：在定理三的假設下， $X = A_0 U_X B_0^T$  以及  $\tilde{A}_0 = A_0$  且  $\tilde{B}_0 = B_0$ ，故

$X = \tilde{A}_0 U_X \tilde{B}_0^T$ 。由此我們可獲得  $\text{vec}(X) = (\tilde{B}_0 \otimes \tilde{A}_0) \text{vec}(U_X)$ 。接下來兩邊取 cov，

得  $\text{cov}(\text{vec}(X)) = (\tilde{B}_0 \otimes \tilde{A}_0) \text{cov}(\text{vec}(U_X)) (\tilde{B}_0 \otimes \tilde{A}_0)^T$  亦即  $\Sigma_X = (\tilde{B}_0 \otimes \tilde{A}_0) T (\tilde{B}_0 \otimes \tilde{A}_0)^T$ 。

藉由奇異值分解定理，存在維度為  $pq \times pq$  之正交矩陣  $Q_1$ 、維度為  $\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}$  之正交

矩陣  $Q_2$  以及維度為  $pq \times \tilde{p}\tilde{q}$  的矩陣  $\Lambda_{pq \times \tilde{p}\tilde{q}} = \begin{pmatrix} \Lambda_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}} \\ \mathbf{0}_{(pq - \tilde{p}\tilde{q}) \times \tilde{p}\tilde{q}} \end{pmatrix}$ ，使得

$\tilde{B}_0 \otimes \tilde{A}_0 = Q_1 \Lambda Q_2^T$ ，其中  $\Lambda_1$  是維度為  $\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}$  的對角矩陣。由引理二可知

$\text{rank}(\tilde{B}_0 \otimes \tilde{A}_0) = \tilde{p}\tilde{q}$ ，因此  $\text{rank}(\Lambda_1) = \tilde{p}\tilde{q}$ 。則  $\Sigma_X = Q_1 \Lambda Q_2^T T Q_2 \Lambda^T Q_1^T$ 。令

$$T^* = Q_2^T T Q_2，那麼 \Sigma_X = Q_1 \Lambda T^* \Lambda^T Q_1^T = Q_1 \begin{pmatrix} \Lambda_1 \\ \mathbf{0} \end{pmatrix} T^* (\Lambda_1^T \quad \mathbf{0}) Q_1^T$$

$= Q_1 \begin{pmatrix} \Lambda_1 T^* \Lambda_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q_1^T$ 。因  $Q_1$  及  $Q_2$  為正交矩陣且  $\Lambda_1$  為滿秩 (full rank)，則

$$\text{rank}(\Sigma_X) = \text{rank}\left(Q_1 \begin{pmatrix} \Lambda_1 T^* \Lambda_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} Q_1^T\right) = \text{rank}\left(\begin{pmatrix} \Lambda_1 T^* \Lambda_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right)$$

$$= \text{rank}(\Lambda_1 T^* \Lambda_1^T) = \text{rank}(T^*) = \text{rank}(Q_2^T T Q_2) = \text{rank}(T)$$

因此，可證得  $\text{rank}(\Sigma_X) = \text{rank}(T)$ 。

**引理四**  $((\tilde{B}_0 \otimes \tilde{A}_0)G)^T ((\tilde{B}_0 \otimes \tilde{A}_0)G) = I_{\tilde{p}\tilde{q}}$ 。

**引理四之證明：**

$$((\tilde{B}_0 \otimes \tilde{A}_0)G)^T ((\tilde{B}_0 \otimes \tilde{A}_0)G) = G^T (\tilde{B}_0 \otimes \tilde{A}_0)^T (\tilde{B}_0 \otimes \tilde{A}_0) G = G^T I_{\tilde{p}\tilde{q}} G = G^T G = I_{\tilde{p}\tilde{q}}。$$

上述為所提出的四個引理之敘述及其證明，底下為定理三的證明。

**定理三之證明：**

先將式(4)表示成矩陣型態。因  $\Sigma_X$  為半正定矩陣，則經由特徵根分解可得

$$\Sigma_{X, pq \times pq} = \Gamma_{pq \times pq} D_{P, pq \times pq} \Gamma_{pq \times pq}^T, \text{ 其中 } \Gamma = (\mathbf{v}_1, \dots, \mathbf{v}_k, \dots, \mathbf{v}_{pq}) \text{ 為正交矩陣而 } D_P \text{ 為對}$$

角矩陣，對角元素為  $\lambda_1 \geq \dots \geq \lambda_k \geq \dots \geq \lambda_{pq}$ ，且  $\lambda_{k+1} = \dots = \lambda_{pq} = 0$ 。因此

$$D_{P, pq \times pq} = \Gamma_{pq \times pq}^T \Sigma_{X, pq \times pq} \Gamma_{pq \times pq}。 \text{ 令 } \Gamma^* = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k), \mathbf{s} = (s_1, s_2, \dots, s_k)^T \text{ 及}$$

$$D_P^* = \begin{pmatrix} \lambda_1 & & o \\ & \ddots & \\ o & & \lambda_k \end{pmatrix}, \text{ 明顯地 } \mathbf{s} = \Gamma^{*T} \text{vec}(X)。 \text{ 因此 } T_{pca}^2 = \sum_{i=1}^k \frac{s_i^2}{\lambda_i}$$

$$= (s_1, \dots, s_k) \begin{pmatrix} \frac{1}{\lambda_1} & & o \\ & \ddots & \\ o & & \frac{1}{\lambda_k} \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_k \end{pmatrix} = (\Gamma^{*T} \text{vec}(X))^T D_P^{*-1} (\Gamma^{*T} \text{vec}(X)), \text{ 而 } \Gamma^* \text{ 是由 } \Gamma \text{ 之前 } k$$

個行向量所形成；則進一步可得  $D_{P, k \times k}^* = \Gamma_{k \times pq}^{*T} \Sigma_{X, pq \times pq} \Gamma_{pq \times k}^*$ ，明顯地  $\Gamma^{*T} \Gamma^* = I_k$ 。

另一方面， $G^T \text{vec}(U_X) = G^T (\tilde{B}_0 \otimes \tilde{A}_0)^T \text{vec}(X) = ((\tilde{B}_0 \otimes \tilde{A}_0)G)^T \text{vec}(X)$ ，兩邊取

cov，而得  $\text{cov}(G^T \text{vec}(U_X)) = \text{cov}(((\tilde{B}_0 \otimes \tilde{A}_0)G)^T \text{vec}(X))$ ，亦即

$$G^T T G = ((\tilde{B}_0 \otimes \tilde{A}_0)G)^T \Sigma_X ((\tilde{B}_0 \otimes \tilde{A}_0)G)。 \text{ 又 } D = G^T T G, \text{ 則}$$

$$D_{\tilde{p}\tilde{q} \times \tilde{p}\tilde{q}} = ((\tilde{B}_0 \otimes \tilde{A}_0)G)_{\tilde{p}\tilde{q} \times pq}^T \Sigma_{X, pq \times pq} ((\tilde{B}_0 \otimes \tilde{A}_0)G)_{pq \times \tilde{p}\tilde{q}}。 \text{ 令 } G_{\tilde{p}\tilde{q} \times k_1}^* \text{ 為由矩陣 } G \text{ 的前 } k_1$$

個行向量所形成之矩陣，則矩陣  $(\tilde{B}_0 \otimes \tilde{A}_0)G^*$  為矩陣  $(\tilde{B}_0 \otimes \tilde{A}_0)G$  的前  $k_1$  行，因此由引理四可獲得  $((\tilde{B}_0 \otimes \tilde{A}_0)G^*)^T ((\tilde{B}_0 \otimes \tilde{A}_0)G^*) = I_{k_1}$ 。令

$$D_{k_1 \times k_1}^* = \begin{pmatrix} \lambda_1^* & & o \\ & \ddots & \\ o & & \lambda_{k_1}^* \end{pmatrix}, \text{ 又 } D = ((\tilde{B}_0 \otimes \tilde{A}_0)G)^T \Sigma_X ((\tilde{B}_0 \otimes \tilde{A}_0)G), \text{ 則可獲得}$$

$$D_{k_1 \times k_1}^* = ((\tilde{B}_0 \otimes \tilde{A}_0)G^*)_{k_1 \times pq}^T \Sigma_{X, pq \times pq} ((\tilde{B}_0 \otimes \tilde{A}_0)G^*)_{pq \times k_1}.$$

現在將式(5)表示成矩陣型態。在 3.3 小節中已宣告

$$G^T \text{vec}(U_X) = (s_1^*, \dots, s_{k_1}^*, \dots, s_{pq}^*)^T, \text{ 則 } G^{*T} \text{vec}(U_X) = (s_1^*, \dots, s_{k_1}^*)^T. \text{ 因此}$$

$$T_{mpca}^2 = \sum_{i=1}^{k_1} \frac{s_i^{*2}}{\lambda_i^*} = (s_1^*, \dots, s_{k_1}^*) \begin{pmatrix} \frac{1}{\lambda_1^*} & & o \\ & \ddots & \\ o & & \frac{1}{\lambda_{k_1}^*} \end{pmatrix} \begin{pmatrix} s_1^* \\ \vdots \\ s_{k_1}^* \end{pmatrix}$$

$$= (G^{*T} \text{vec}(U_X))^T D^{*-1} (G^{*T} \text{vec}(U_X)) = (((\tilde{B}_0 \otimes \tilde{A}_0)G^*)^T \text{vec}(X))^T D^{*-1} (((\tilde{B}_0 \otimes \tilde{A}_0)G^*)^T \text{vec}(X))$$

由引理三得知  $\text{rank}(\Sigma_X) = \text{rank}(T)$ ，而定理三的前提假設  $k = \text{rank}(\Sigma_X)$  及先前宣告  $\text{rank}(T) = k_1$ ，因此  $k = k_1$ 。又由上述可知  $D_P^* = \Gamma_{k \times pq}^{*T} \Sigma_{X, pq \times pq} \Gamma_{pq \times k}^*$  及  $D^* = ((\tilde{B}_0 \otimes \tilde{A}_0)G^*)_{k_1 \times pq}^T \Sigma_{X, pq \times pq} ((\tilde{B}_0 \otimes \tilde{A}_0)G^*)_{pq \times k_1}$ ，則  $D_P^* = D^*$  且

$\text{span}((\tilde{B}_0 \otimes \tilde{A}_0)G^*) = \text{span}(\Gamma^*)$ 。當所有特徵值都不相同時，則  $\lambda_i = \lambda_i^*$  且

$s_i^2 = s_i^{*2}, 1 \leq i \leq k$ ；而當某些特徵值有相同值時， $\text{vec}(X)$  投影至其所對應之特徵空間之投影量長度平方即為相對應之  $s_i^2$  之和亦為  $s_i^{*2}$  之和，故兩者相等。因此可證得式(4)等於式(5)。

另外，在 Hung et al. (2012) 文章中有一些公式並未附帶推導過程，我們有對一些公式給予詳細推導，其推導如下：

$$(i) E[(X - \mu)P_{\tilde{B}_0}(X - \mu)^T] = \sum_{j=1}^{\tilde{q}} E[(X - \mu)(\mathbf{b}_j \mathbf{b}_j^T)(X - \mu)^T] = \sum_{j=1}^{\tilde{q}} (\mathbf{b}_j \otimes I_p)^T \Sigma (\mathbf{b}_j \otimes I_p)$$

推導：

$$\text{因 } P_{\tilde{B}_0} = \sum_{j=1}^{\tilde{q}} \mathbf{b}_j \mathbf{b}_j^T, \text{ 則 } E[(X - \mu)P_{\tilde{B}_0}(X - \mu)^T] = \sum_{j=1}^{\tilde{q}} E[(X - \mu)(\mathbf{b}_j \mathbf{b}_j^T)(X - \mu)^T]$$

$$= \sum_{j=1}^{\tilde{q}} E[\text{vec}(I_p(X - \mu)\mathbf{b}_j)(\text{vec}(I_p(X - \mu)\mathbf{b}_j))^T]$$

$$= \sum_{j=1}^{\tilde{q}} E[(\mathbf{b}_j \otimes I_p)^T \text{vec}(X - \mu)(\text{vec}(X - \mu))^T (\mathbf{b}_j \otimes I_p)]$$

$$= \sum_{j=1}^{\tilde{q}} (\mathbf{b}_j \otimes I_p)^T \Sigma (\mathbf{b}_j \otimes I_p)$$

$$(ii) E[(X - \mu)^T P_{\tilde{A}_0}(X - \mu)] = \sum_{i=1}^{\tilde{p}} E[(X - \mu)^T (\mathbf{a}_i \mathbf{a}_i^T)(X - \mu)] = \sum_{i=1}^{\tilde{p}} (I_q \otimes \mathbf{a}_i)^T \Sigma (I_q \otimes \mathbf{a}_i)$$

推導：

$$\text{因 } P_{\tilde{A}_0} = \sum_{i=1}^{\tilde{p}} \mathbf{a}_i \mathbf{a}_i^T, \text{ 則 } E[(X - \mu)^T P_{\tilde{A}_0}(X - \mu)] = \sum_{i=1}^{\tilde{p}} E[(X - \mu)^T (\mathbf{a}_i \mathbf{a}_i^T)(X - \mu)]$$

$$= \sum_{i=1}^{\tilde{p}} E[(\text{vec}(\mathbf{a}_i^T (X - \mu) I_q))^T (\text{vec}(\mathbf{a}_i^T (X - \mu) I_q))^T]$$

$$= \sum_{i=1}^{\tilde{p}} E[(I_q \otimes \mathbf{a}_i)^T \text{vec}(X - \mu)(\text{vec}(X - \mu))^T (I_q \otimes \mathbf{a}_i)]$$

$$= \sum_{i=1}^{\tilde{p}} (I_q \otimes \mathbf{a}_i)^T \Sigma (I_q \otimes \mathbf{a}_i)$$

$$(iii) \Phi(\tilde{p}, \tilde{q}) = \sum_{i=1}^{\tilde{p}} \lambda_i = \sum_{j=1}^{\tilde{q}} \xi_j \text{ 及 } \hat{\Phi}(\tilde{p}, \tilde{q}) = \sum_{i=1}^{\tilde{p}} \hat{\lambda}_i = \sum_{j=1}^{\tilde{q}} \hat{\xi}_j$$

推導：

$$\text{由前面定理二性質 (c') 之證明可得 } E \left\| \tilde{A}_0^T X \tilde{B}_0 \right\|_F^2 = E[\text{trace}(\tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0)] =$$

$$\text{trace}(\tilde{A}_0^T E[XP_{\tilde{B}_0} X^T] \tilde{A}_0) \text{ 又 } E[\text{trace}(\tilde{A}_0^T X \tilde{B}_0 \tilde{B}_0^T X^T \tilde{A}_0)] = E[\text{trace}(\tilde{B}_0^T X^T \tilde{A}_0 \tilde{A}_0^T X \tilde{B}_0)]$$

，則可得知  $\text{trace}(\tilde{A}_0^T E[XP_{\tilde{B}_0} X^T] \tilde{A}_0) = \text{trace}(\tilde{B}_0^T E[X^T P_{\tilde{A}_0} X] \tilde{B}_0)$  ， 而

$$\text{trace}(\tilde{A}_0^T E[XP_{\tilde{B}_0} X^T] \tilde{A}_0) = \sum_{i=1}^{\tilde{p}} \lambda_i \text{ 和 } \text{trace}(\tilde{B}_0^T E[X^T P_{\tilde{A}_0} X] \tilde{B}_0) = \sum_{j=1}^{\tilde{q}} \xi_j \text{ ， 因此可得}$$

$$\Phi(\tilde{p}, \tilde{q}) = \sum_{i=1}^{\tilde{p}} \lambda_i = \sum_{j=1}^{\tilde{q}} \xi_j \text{ 。 同理亦可得 } \hat{\Phi}(\tilde{p}, \tilde{q}) = \sum_{i=1}^{\tilde{p}} \hat{\lambda}_i = \sum_{j=1}^{\tilde{q}} \hat{\xi}_j \text{ 。}$$

