OPTIMAL STATISTICAL DESIGNS OF MULTIVARIATE EWMA AND MULTIVARIATE CUSUM CHARTS BASED ON AVERAGE RUN LENGTH AND MEDIAN RUN LENGTH

by

LEE MING HA

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

November 2006

ACKNOWLEDGEMENTS

First and foremost, I would like to take this opportunity to thank my supervisor, Dr. Michael Khoo Boon Chong for his valuable assistance, patience and comments throughout the completion of this thesis. I would also like to take this opportunity to express my thanks to the Dean of the School of Mathematical Sciences, Universiti Sains Malaysia, Associate Professor Dr. Ahmad Izani Md. Ismail, his deputies Mr. Safian Uda and Dr. Adli Mustafa, lecturers and staffs of the department for their advice and support as well as those who have contributed to the implementation and completion of this study.

I am very grateful to the USM librarians for all the assistance that they have rendered me in finding the required materials for this study. With their help, I have managed to use the library's resources and facilities effectively which enabled me to search for information related to this study electronically via the online database. I also wish to express my appreciation and gratitude to the staffs of the Institute of Postgraduate Studies for their help and guidance throughout my study in USM.

I am indebted to my family for providing me with the opportunity to further my study. Special thanks to my father Mr. Lee Hock Yeo who has always respected and supported me in whatever decisions that I made. I would like to thank my cousin Ms. Lee See Ha who is also studying in USM for her support during my study.

I also wish to extend my thanks to all of my friends, especially Ms. Ang Siew Wei, Ms. Yap Ping Wei and Ms. Chua Cui King for their helpful suggestion and encouragement. But above all, I would like to thank God for all of his help.

TABLE OF CONTENTS

CHAPTER 3 : MARKOV CHAIN APPROACH FOR MULTIVARIATE EWMA AND MULTIVARIATE CUSUM CHARTS BASED ON AVERAGE RUN LENGTH

CHAPTER 4 : A PROPOSED MARKOV CHAIN APPROACH FOR MULTIVARIATE EWMA AND MULTIVARIATE CUSUM CHARTS BASED ON MEDIAN RUN LENGTH

CHAPTER 5 : PROGRAM DESCRIPTION AND OPERATION

CHAPTER 6 : PROPOSED OPTIMAL STATISTICAL DESIGNS OF MULTIVARIATE EWMA AND MULTIVARIATE CUSUM CHARTS

CHAPTER 7 : A PROPOSED GRAPHICAL METHOD FOR OPTIMAL DESIGNS OF MULTIVARIATE EWMA AND MULTIVARIATE CUSUM CHARTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF PUBLICATIONS

1. Lee, M.H. and Khoo, M.B.C. (2006). Optimal Statistical Design of a Multivariate EWMA Chart based on ARL and MRL. *Communications in Statistics – Simulation and Computation* 35(3), 831–847.

2. Lee, M.H. and Khoo, M.B.C. (2006). Optimal Statistical Design of a Multivariate CUSUM Chart based on ARL and MRL. *International Journal of Reliability, Quality and Safety Engineering* 13(5).

REKABENTUK BERSTATISTIK OPTIMA BAGI CARTA-CARTA EWMA MULTIVARIAT DAN CUSUM MULTIVARIAT BERDASARKAN PURATA PANJANG LARIAN DAN MEDIAN PANJANG LARIAN

ABSTRAK

Carta kawalan multivariat ialah alat yang berkuasa dalam kawalan proses yang melibatkan kawalan serentak beberapa cirian kualiti yang berkorelasi. Carta-carta multivariat hasil tambah longgokan (MCUSUM) dan multivariat purata bergerak berpemberat eksponen (MEWMA) sentiasa dicadangkan dalam kawalan proses apabila pengesanan cepat anjakan tetap yang kecil atau sederhana dalam vektor min adalah diingini. Tesis ini memperkembangkan suatu kaedah grafik untuk menentukan pilihan optima parameter carta MEWMA and MCUSUM bagi pengesanan cepat suatu saiz anjakan yang diingini. Carta-carta kawalan ini adalah direka bentuk secara berstatistik dan kaedah yang digunakan adalah berdasarkan pendekatan am yang hanya tersedia ada untuk carta-carta univariat EWMA dan CUSUM. Prestasi carta-carta kawalan adalah diukur dengan purata panjang larian (ARL) yang diperoleh dengan menggunakan pendekatan rantai Markov atau kaedah simulasi Reka bentuk berdasarkan median panjang larian (MRL) juga diberikan memandangkan MRL adalah lebih bermakna sebagai ukuran pemusatan berkenaan dengan taburan panjang larian yang sangat terpencong. Reka bentuk berdasarkan MRL dijalankan dengan menggunakan pendekatan rantai Markov atau kaedah simulasi. Selain memberikan pendekatan grafik untuk mempermudahkan prosedur sedia ada dalam reka bentuk carta-carta optima MEWMA dan MCUSUM yang berdasarkan terutamanya pada ARL, tesis ini memperkenalkan strategi reka bentuk optima dua carta tersebut dengan menggunakan MRL, yang hanya tersedia ada untuk carta-carta univariat EWMA dan CUSUM sahaja. Setiap skema reka bentuk adalah berdasarkan pada aspek taburan panjang larian yang diperoleh dengan menggunakan cerapan-cerapan tak bersandar daripada taburan normal multivariat. Bilangan cirian kualiti yang dipertimbangkan ialah $p = 2, 3, 4, 5, 8$ dan 10. Prosedur untuk reka bentuk optima carta-carta MEWMA dan

CHAPTER 1
INTRODUCTION

1.1 Control Charts

One of the most powerful tools that has been used extensively in quality improvement work is a control chart. The general idea of a control chart was sketched out in a memorandum that Walter Shewhart of Bell Laboratory wrote on May 16, 1924 (Montgomery, 2005). The control chart found widespread use during World War II and has been employed, with various enhancements and modifications, ever since. The construction of a control chart is based on statistical principles. Specifically, the charts are based upon some of the statistical distributions. When used in conjunction with a manufacturing process (or a non-manufacturing process), a control chart can indicate when a process is out-of-control. Ideally, we would want to detect such a situation as soon as possible after its occurrence. Conversely, we would like to have as few false alarms as possible.

While the bulk of the literature on control charts deal with a single measurement on the process, methods are available which may be employed when two or more characteristics are measured at the same time on a process. Shewhart also recognized this problem but a great deal of what will be discussed is due to the work by Harold Hotelling in the 1930s and 1940s. These techniques include his T^2-procedure and its extensions to multivariate generalizations of control charts for means and standard deviations or ranges (Jackson, 1985).

Recent papers dealing with multivariate control procedures include the works of Villalobos et al. (2005), Yang and Rahim (2005), Zhou et al. (2005), Bodecchi et al. (2005), Champ et al. (2005), Marengo et al. (2006), Aparisi et al. (2006); and Testik and Runger (2006).

1.2 Basic Control Chart Principles

A control chart is a graphical display of a quality characteristic that has been measured or computed from a sample versus the sample number or time. The chart contains a centre line that represents the average value of the quality characteristic corresponding to the in-control state. Two other horizontal lines are the upper control limit and the lower control limit. These control limits are chosen so that if the process is in-control, nearly all of the sample points will fall between them. As long as the points plot within the control limits, the process is assumed to be in control and no action is necessary. However, a point that plots outside of the control limits is interpreted as evidence that the process is out-of-control, hence investigation and corrective action are required to find and eliminate the assignable cause or causes responsible for this behavior. It is customary to connect the sample points on the control chart with straight-line segments, so that it is easier to visualize how the sequence of points has evolved over time.

There is a close connection between control charts and hypothesis testing. The control chart is a test of the hypothesis that the process is in a state of statistical control. A point plotting within the control limits is equivalent to failing to reject the hypothesis of statistical control, and a point plotting outside the control limits is equivalent to rejecting the hypothesis of statistical control.

An important factor in control chart usage is the design of the control chart. By this we mean the selection of sample size, control limits, and frequency of sampling. In most quality control problems, it is customary to design the control chart using primarily statistical considerations. The use of statistical criteria such as these along with industrial experience has led to general guidelines and procedures for designing control charts. Recently, however, control chart design from an economic point of view has begun, considering explicitly the cost of sampling, losses from allowing defective product to be produced, and the costs of investigating out-of-control signals that are actually false alarms (Montgomery, 2005).

1.3 Multivariate Quality Control Charts

There are many situations in which the simultaneous monitoring or control of two or more related variables is necessary. Process-monitoring problems in which several related variables are of interest are sometimes called multivariate quality control problems. The original work in multivariate quality control was done by Hotelling (1947), who applied his procedures to bombsight data during World War II. This subject is particularly important today, as automatic inspection procedure makes it relatively easy to measure many parameters on each unit of product manufactured. Many chemical process plants and semiconductor manufacturers routinely maintain manufacturing databases with process and quality data on hundreds of variables. Often the total size of these databases is measured in millions of individual records. Monitoring or analysis of these data with univariate SPC procedure is often ineffective. The use of multivariate methods has increased greatly in recent years for this reason (Montgomery, 2005).

The most familiar multivariate process monitoring and control procedure is the Hotelling's T^2 control chart for monitoring the mean vector of a process. It is a direct analog of the univariate Shewhart $\overline{X}$ chart. The Hotelling's T^2 chart is a multivariate Shewhart type control chart that only takes into account the present information of the process, so consequently it is relatively insensitive to small and moderate shifts in the process mean vector. To provide more sensitivity to small and moderate shifts, multivariate exponentially weighted moving average (MEWMA) and multivariate cumulative sum (MCUSUM) control charts are developed. Their advantage is that they take into account the present and past information of the process. Therefore, they are followed by selecting the combination of chart parameters, i.e., the smoothing constant, r for the MEWMA chart (or the reference value, k for the MCUSUM chart) and its corresponding control limit, H. The combination of the chart parameters obtained is optimal in the sense that for a fixed in-control ARL or MRL, it produces the lowest out-of-control ARL or MRL for the specified magnitude of a shift in the mean vector for a quick detection. Graphs of the optimal chart parameters of the MEWMA and MCUSUM charts for various in-control ARLs or MRLs based on different magnitude of shifts in the mean vector are given. Examples illustrating the application of these graphs are also provided.

OPTIMAL STATISTICAL DESIGNS OF MULTIVARIATE EWMA AND MULTIVARIATE CUSUM CHARTS BASED ON AVERAGE RUN LENGTH AND MEDIAN RUN LENGTH

ABSTRACT

A multivariate control chart is a powerful tool in process control involving a simultaneous monitoring of several correlated quality characteristics. The multivariate cumulative sum (MCUSUM) and multivariate exponentially weighted moving average (MEWMA) charts are often recommended in process monitoring when a quick detection of small or moderate sustained shifts in the mean vector is desired. This thesis develops a graphical method to determine the optimal choices of the parameters of the MEWMA and MCUSUM charts for a quick detection of a desired size of a shift. These control charts are statistically designed and the method used follows the general approach that is currently available only for univariate EWMA and CUSUM charts. The performances of the control charts are measured by the average run length (ARL) that is derived using a Markov chain approach or a simulation method. The design based on the median run length (MRL) is also given as the MRL is a more meaningful measure of centrality with respect to the highly skewed run length distribution. The design based on MRL is made using the Markov chain approach or the simulation method. Besides providing a graphical approach to simplify the existing procedure in the design of optimal MEWMA and MCUSUM charts which are based mainly on the ARL, this thesis introduces the optimal design strategy of the two charts using the MRL, which are currently available for the univariate EWMA and CUSUM charts only. Each of the design schemes is based on the aspects of the run length distribution derived using independent observations from the multivariate normal distribution. The number of quality characteristics considered are p = 2, 3, 4, 5, 8 and 10. The procedure for the optimal design of the MEWMA and MCUSUM charts consists of specifying the desired in-control ARL or MRL and the magnitude of the shift (i.e., the square root of the noncentrality parameter, δ) in the process mean vector to be detected quickly. This is

MCUSUM terdiri daripada penetapan ARL atau MRL dalam kawalan yang diingini dan magnitud anjakan dalam vektor min proses (iaitu, punca kuasa dua parameter tak memusat, δ) yang akan dikesan dengan cepat. Ini diikuti dengan pemilihan kombinasi parameter-parameter carta, iaitu pemalar pelicinan, r untuk carta MEWMA (atau nilai rujukan, k untuk carta MCUSUM) dan had kawalannya yang sepadan, H. Kombinasi parameter-parameter carta yang diperoleh adalah optima dengan kenyataan bahawa bagi ARL atau MRL dalam kawalan yang ditetapkan, ia menghasilkan ARL atau MRL luar kawalan yang terendah untuk magnitud anjakan dalam vektor min yang ditetapkan untuk pengesanan cepat. Graf-graf parameter carta optima bagi carta MEWMA dan MCUSUM untuk pelbagai ARL atau MRL dalam kawalan berdasarkan magnitud anjakan dalam vektor min yang berlainan adalah diberikan. Contoh-contoh yang menunjukkan aplikasi graf-graf ini turut diberikan.

more powerful to detect small shifts than the Hotelling's T^2 chart. The MEWMA statistics is a more straightforward generalization of the corresponding univariate procedure than the MCUSUM statistics. Furthermore, the design of the MEWMA control chart is simpler and it can also be used as a process forecasting tool (Chua and Montgomery, 1992).

One of the design criteria for both univariate and multivariate control charts is statistical design. Statistical design procedures refer to choices of optimal chart parameters that ensure the control chart performance meets certain statistical criteria. These criteria are often based on aspects of the run length distribution of the chart, such as the average run length (ARL) or the median run length (MRL). It is recommended that a multivariate control chart is designed to have a specified ARL (or MRL) value at shift $\delta = 0$ and a minimum ARL (or MRL) value at shift $\delta = \delta_1$, where δ_1 is the smallest magnitude of a shift in the process mean vector considered important enough to be detected quickly. Note that δ is referred to as the square root of the noncentrality parameter. The values of the chart parameters are optimal as these values appear to minimize the ARL (or MRL) at shift $\delta = \delta_1$ for a given ARL (or MRL) at shift $\delta = 0$.

Multivariate control charting procedures can be computationally intensive. The availability of computer software will be a major determining factor in the future use of such charts.

1.4 Objectives of the Thesis

In this thesis, we propose the statistical designs of the MEWMA and MCUSUM charts. The objectives of the thesis are as follows:

(i) to develop the optimal designs of the MEWMA and MCUSUM charts. Besides simplifying the existing procedures in designing an optimal MEWMA or MCUSUM chart, based mainly on the ARL, this thesis proposes the design strategies of the charts using the MRL. Thus, ARL is used as a primary criterion for evaluating the performances of these control charts, whereas the use of MRL as a secondary criterion for designing the MEWMA and MCUSUM charts, which is an important contribution of this thesis is also recommended. Optimal designs of the MEWMA and MCUSUM charts allow a more complete study of the performances of these control charts. The optimal designs based on MRL are currently available in the literature for the univariate EWMA and CUSUM charts only. This study extends the optimal designs to the multivariate cases.

(ii) to propose a four step procedure in obtaining the chart parameters of optimal MEWMA and MCUSUM charts based on ARL and MRL. The four step procedure is currently available for the univariate EWMA and CUSUM control charts only in the literature. This procedure is extended to the multivariate EWMA and CUSUM control charts in this thesis. Although optimal statistical designs of the MEWMA and MCUSUM charts based on ARL are given in Prabhu and Runger (1997) and Crosier (1988) respectively, this thesis provides a more complete step-by-step approach (i.e., a four step procedure) for the optimal designs of the charts compared to the existing methods which are insufficient and incomplete.

(iii) to obtain an optimal set of values for the chart parameters associated with the MEWMA or MCUSUM procedure: the control limit H and the smoothing constant r for the MEWMA chart or the control limit H and the reference value k for the MCUSUM chart.

(iv) to present user-friendly programs that allow practitioners to determine the optimal chart parameters of the MEWMA and MCUSUM control charts in all possible situations. The programs are implemented using the Markov chain approach or the simulation method, and the programs calculate the ARL, MRL, control limits and percentiles of the run length distribution for the MEWMA and MCUSUM charts.

(v) to provide a graphical approach by means of graphs that guide practitioners in the optimal designs of the MEWMA and MCUSUM charts for detecting shifts of a desired magnitude in the mean vector. These graphs can give immediate approximation of the optimal chart parameters of the control charts for a given in-control ARL or MRL.

1.5 Methodologies and Organization of the Thesis

In this thesis, the one-dimensional Markov chain approach described by Runger and Prabhu (1996) is applied for the in-control case of the optimal designs of the MEWMA and MCUSUM charts based on ARL. The two-dimensional Markov chain proposed by Runger and Prabhu (1996) is used to approximate the ARL for the out-of-control case of the MEWMA chart. As for the MCUSUM chart, the simulation method is used for the out-of-control case to study the performance of the chart based on the ARL.

Using the theory of probability distribution of the run length given by Brook and Evans (1972), the in-control MRL values of the MEWMA and MCUSUM charts are computed using the one-dimensional Markov chain approach. This theory is also used to obtain the out-of-control MRL values of the MEWMA chart using the two-dimensional Markov chain approach. For the out-of-control case of the MCUSUM chart based on MRL, the simulation method is implemented.

In Chapter 2, the MEWMA and MCUSUM control chart procedures are discussed. The properties and performance evaluation of these control charts are also explained in this chapter. The Markov chain representation of the multivariate control procedures for designing control charts based on the ARL and MRL is given in Chapters 3 and 4 respectively. Chapter 5 explains the Statistical Analysis System (SAS) programs that are used to calculate the ARL, MRL and other percentiles of the run length distribution of the control charts. These programs are given in Appendix C. Chapter 6 defines the concept of the proposed optimal statistical design of the MEWMA and MCUSUM control charts. Chapter 7 presents the proposed graphical method for the optimal design of the control charts. The graphs of optimal parameters of the control charts are provided in Appendix D. Examples for the application of these graphs are also given in Chapter 7. Finally, the conclusion of this thesis and suggestions for further research are presented in Chapter 8.

CHAPTER 2
SOME PRELIMINARIES AND REVIEW OF MULTIVARIATE EWMA AND MULTIVARIATE CUSUM CHARTS

2.1 The Multivariate Normal Distribution

The normal distribution is generally used to describe the behavior of a continuous quality characteristic in univariate statistical quality control. The univariate normal probability density function (Montgomery, 2005) is

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad \text{for } -\infty < x < \infty. \tag{2.1}$$

Apart from the minus sign, the term in the exponent of the normal distribution can be written as

$$(x-\mu)(\sigma^2)^{-1}(x-\mu) \tag{2.2}$$

where the mean of the normal distribution is μ and its variance is σ^2.

This quantity measures the squared standardized distance from the random variable, X to the mean μ, where the term "standardized" means that the distance is expressed in standard deviation units (Montgomery, 2005).

A similar approach can be used in the multivariate case. Assume that there are p variables, $X_1, X_2,..., X_p$ in a p-component vector, $\mathbf{X}' = (X_1, X_2,..., X_p)$. Let the mean vector of the $\mathbf{X}$'s be $\boldsymbol{\mu}' = (\mu_1, \mu_2,..., \mu_p)$, and the $p \times p$ covariance matrix, Σ contains the variances and covariances of the random variables in $\mathbf{X}$, where the main diagonal elements of Σ are the variances of X_j, for $j = 1, 2,..., p$, and the off-diagonal elements are the covariances. The squared standardized (generalized) distance from $\mathbf{X}$ to $\boldsymbol{\mu}$ is

$$(\mathbf{X}-\boldsymbol{\mu})' \Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu}). \tag{2.3}$$

The multivariate normal density function is obtained by replacing the squared standardized distance in Equation (2.2) by the multivariate generalized distance in Equation (2.3) and changing the constant term $1/\sqrt{2\pi\sigma^2}$ to a more general form that makes the area under the probability density function unity regardless of the value of p. Thus, the multivariate normal probability density function (Montgomery, 2005) is

$$f(\mathbf{x})=\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{2.4}$$

where $\mathbf{x}'=(x_1, x_2,\ldots, x_p)$.

We will now give a brief description on the sample mean vector and covariance matrix of a random sample from a multivariate normal distribution. Assume that we have a random sample of size n, $\mathbf{X}_1, \mathbf{X}_2,\ldots, \mathbf{X}_n$, from a multivariate normal distribution, where the i^{th} sample vector, $\mathbf{X}_i$ contains observations on each of the p variables, $X_{i1}, X_{i2},\ldots, X_{ip}$, $i=1,2,\ldots,n$. It follows that the sample mean vector (Montgomery, 2005) is

$$\overline{\mathbf{X}}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i \tag{2.5}$$

and the sample covariance matrix is

$$\mathbf{S}=\frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i-\overline{\mathbf{X}})(\mathbf{X}_i-\overline{\mathbf{X}})'. \tag{2.6}$$

The sample variances on the main diagonal of the matrix S are computed as

$$S_j^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ij}-\overline{X}_j\right)^2, \tag{2.7}$$

for $j = 1, 2,\ldots, p$, while the sample covariances as

$$S_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)\left(X_{ik} - \bar{X}_k\right), \qquad (2.8)$$

for $j = 1, 2, \ldots, p$, $k = 1, 2, \ldots, p$ and $j \neq k$. Note that the sample mean of variable j is computed as follows:

$$\bar{X}_j = \frac{\sum_{i=1}^{n} X_{ij}}{n}. \qquad (2.9)$$

It can be shown that the sample mean vector and sample covariance matrix are unbiased estimators of the corresponding population quantities, i.e., $E(\overline{\mathbf{X}}) = \boldsymbol{\mu}$ and $E(\mathbf{S}) = \Sigma$ (Montgomery, 2005).

The sample covariance matrix in correlation form is made up of elements R_{jk} representing the pairwise correlation coefficient between quality characteristics X_j and X_k in vector $\mathbf{X}$; that is, the element in the j^{th} row and the k^{th} column of the sample covariance matrix in correlation form is given by (Tracy et al., 1992)

$$r_{jk} = \frac{S_{jk}}{S_j \cdot S_k}. \qquad (2.10)$$

2.2 The Multivariate EWMA Chart

The MEWMA chart is first studied by Lowry et al. (1992). Yumin (1996) investigates the MEWMA chart with the generalized smoothing parameter matrix. He introduces an orthogonal transformation such that the principal components of the original variables are independent of one another. Sullivan and Woodall (1996) use the MEWMA and multivariate CUSUM charts for a preliminary analysis of multivariate observations. They propose that the MEWMA should be applied to the data in a reverse time order as well as in a time order, to avoid asymmetrical performance in the detection of shifts. Margavio and Conerly (1995) consider two alternatives to the MEWMA chart. One of these alternatives is an arithmetic moving average control chart which is the arithmetic average of the sample means for the last k periods. The other alternative is a truncated version of the EWMA which truncates the EWMA after a fairly short period of time so that more emphasis is placed on the most current observation. Simulated ARL results indicate that for some situations these alternative charts outperform the MEWMA chart. Kramer and Schmid (1997) propose a MEWMA control chart which is a generalization of the control scheme of Lowry et al. (1992) for multivariate time independent observations. Runger et al. (1999) show how the shift detection capability of the MEWMA control chart can be significantly improved by transforming the original process variables to a lower-dimensional subspace through the use of a U-transformation. Stoumbos and Sullivan (2002) investigate the effects of non-normality on the statistical performance of the MEWMA control chart. They show that with individual observations, and therefore, by extension, with subgroups of any size, the MEWMA chart can be designed to be robust to non-normality and very effective at detecting process shifts of any size and direction, even for highly skewed and extremely heavy-tailed multivariate distributions. Reynolds and Kim (2005) investigate MEWMA charts based on sequential sampling and show that the MEWMA chart based on sequential sampling is much more efficient in detecting changes in the process mean vector than standard control charts based on non-sequential sampling. Pan (2005) proposes a MEWMA scheme that is an alternative to the traditional MEWMA and the distribution of the chart statistic is derived from the Box quadratic form. Kim and Reynolds (2005) show how the MEWMA chart is applied to monitor the process mean vector when the sample sizes for the p variables are not all equal. Yeh et al. (2005) propose a MEWMA chart that effectively monitors changes in the population variance-covariance matrix of a multivariate normal process when individual observations are collected.

Suppose that the $p \times 1$ random vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots$, each representing the p quality characteristics to be monitored simultaneously, are observed over time. These vectors may represent individual observations or sample mean vectors. It is assumed that $\mathbf{X}_i$, $i = 1, 2, \ldots$, are independent multivariate normal random vectors with an in-control mean vector, $\boldsymbol{\mu}_0$. For simplicity, it is assumed that each of the random vectors has a known covariance matrix, Σ (Lowry et al., 1992). In the multivariate case, Lowry et al. (1992) extend the univariate exponentially weighted moving average (EWMA) to vectors of EWMA's which can be written as

$$\mathbf{Z}_t = \mathbf{R}\mathbf{X}_t + (\mathbf{I} - \mathbf{R})\mathbf{Z}_{t-1}, \tag{2.11}$$

for $t = 1, 2, \ldots$, where $\mathbf{Z}_0 = \mathbf{0}$ and $\mathbf{R} = \text{diag}(r_1, r_2, \ldots, r_p)$, for $0 < r_j \leq 1, j = 1, 2, \ldots, p$. The MEWMA chart gives an out-of-control signal as soon as

$$T_t^2 = \mathbf{Z}_t' \Sigma_{Z_t}^{-1} \mathbf{Z}_t > H \tag{2.12}$$

where $H > 0$ is chosen to achieve a specified in-control ARL and $\Sigma_{Z_t} = \frac{r}{2-r}\left[1 - (1-r)^{2t}\right]\Sigma$ is the covariance matrix of $\mathbf{Z}_t$. If there is no prior reason to weigh past observations differently for the p quality characteristics being monitored, then $r_1 = r_2 = \cdots = r_p = r$. It will be discussed in Section 2.4 that the ARL performance of the MEWMA chart depends only on the square root of the noncentrality parameter, $\delta = \left[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\right]^{\frac{1}{2}}$, where $\boldsymbol{\mu}_1$ is the out-of-control mean vector.

If $r_1 = r_2 = \cdots = r_p = r$, then the MEWMA vector of Equation (2.11) is defined as

$$\mathbf{Z}_t = r\mathbf{X}_t + (1-r)\mathbf{Z}_{t-1}, \tag{2.13}$$

for $t = 1, 2, \ldots$. As MacGregor and Harris (1990) point out for the univariate case, using the exact variance of the EWMA statistic leads to a natural fast initial response for the EWMA chart. Thus, initial out-of-control conditions are detected quicker. This is also true for the MEWMA chart. Because, however, it may be more likely that the process will stay in-control for a while and then shift out-of-control, the asymptotic (as $t \to \infty$) covariance matrix, that is,

$$\Sigma_{Z_t} = \frac{r}{2-r}\Sigma \tag{2.14}$$

can be used to calculate the MEWMA statistic in Equation (2.12) (Lowry et al., 1992).

2.3 The Multivariate CUSUM Chart

Several versions of multivariate CUSUM charts have appeared in the literature. Woodall and Ncube (1985) consider the simultaneous use of several univariate cumulative sum (CUSUM) charts to be a single multivariate CUSUM for monitoring a multivariate normal process of p quality characteristics. The multivariate CUSUM chart is out-of-control whenever any of the univariate CUSUM chart is out-of-control. Healy (1987) applies the CUSUM procedure to the multivariate normal distribution for detecting a shift in the mean vector or for detecting a shift in the covariance matrix. Crosier (1988) proposes two new multivariate CUSUM charts. The first chart is referred to as COT, which reduces each observation to a scalar (Hotelling's T^2 statistic) and forms a CUSUM of the T^2 statistics. The second chart is referred to as the multivariate CUSUM, which forms a CUSUM vector directly from the observations. Note that the second chart is the multivariate version of the two-sided CUSUM chart introduced by Crosier (1986). Pignatiello and Runger (1990) also propose two new multivariate CUSUM charts. The first multivariate CUSUM chart (multivariate CUSUM #1) accumulates the observations before producing the quadratic forms of the mean vector while the second multivariate CUSUM chart (multivariate CUSUM #2) calculates a quadratic form for each observation and then accumulates the quadratic forms of the mean vector. The multivariate CUSUM #2 has a better ARL performance than the multivariate CUSUM #1. Hawkins (1991) proposes a CUSUM chart for regression adjusted variables. This chart is based on the vector $\mathbf{Z}$ of scaled residuals from the regression of each variable on all other variables. Wierda (1994) performs a rough ranking based on restricted simulation studies. He shows that the mean estimating multivariate CUSUM #1 and the CUSUM based on regression adjusted variables seem to be promising, based on both the ARL considerations and the ability to interpret an out-of-control signal. Ngai and Zhang (2001) develop a natural multivariate extension of the CUSUM chart via projection pursuit. This chart is more effective in avoiding inertia problem and coping with delayed shifts. Qiu and Hawkins (2001) suggest a rank-based multivariate CUSUM procedure which is based on the cross-sectional antiranks of the measurements. This procedure is distribution free in the sense that all its properties depend on the distribution of the antirank vector only. Qiu and Hawkins (2003) also propose a nonparametric multivariate CUSUM chart which is based on the order information among the measurement components, and on the order information between the measurement components and their in-control means. Runger and Testik (2004) provide a comparison of the advantages and disadvantages of several multivariate CUSUM charts. They also give the performance evaluation and interrelationship of the charts.

Several authors have developed multivariate extensions of the univariate CUSUM. Crosier (1988) proposes two multivariate CUSUM charts. The one with the better ARL performance and will be considered in this thesis is denoted by MCUSUM and is based on the following statistics:

$$C_t = \left\{(\mathbf{S}_{t-1} + \mathbf{X}_t - \mathbf{a})'\Sigma^{-1}(\mathbf{S}_{t-1} + \mathbf{X}_t - \mathbf{a})\right\}^{1/2} \tag{2.15}$$

and

$$\mathbf{S}_t = \mathbf{0} \qquad\qquad \text{if } C_t \le k,$$

or

$$\mathbf{S}_t = (\mathbf{S}_{t-1} + \mathbf{X}_t - \mathbf{a})\left(1 - \frac{k}{C_t}\right) \qquad \text{if } C_t > k,$$

for $t = 1, 2, \ldots$. Here, $\mathbf{S}_0 = \mathbf{0}$ and $k > 0$ is the reference value of the scheme with Σ representing the covariance matrix of $\mathbf{X}_t$. In this thesis, it is assumed without loss of generality that $\mathbf{a} = \mathbf{0}$, where $\mathbf{a}$ is the aim point or target value for the mean vector. An out-of-control signal is generated (Crosier, 1988) when

$$Y_t = \left[\mathbf{S}_t' \Sigma^{-1} \mathbf{S}_t\right]^{1/2} > H \tag{2.16}$$

where $H > 0$ is the control limit of the scheme.

The MCUSUM procedure is often based on the assumption that the observations $\mathbf{X}_t$ belong to an independently and identically distributed process, from a multivariate normal distribution. The MCUSUM scheme enables faster detection of small shifts in the mean vector than the Hotelling's T^2 chart.

The MCUSUM chart is a directionally invariant scheme (Crosier, 1988), that is, the ARL performance of the chart is determined by the distance of the off-target mean vector, $\boldsymbol{\mu}_1$ from the on-target mean vector, $\boldsymbol{\mu}_0$. The distance is defined as the square root of the noncentrality parameter, $\delta = \left[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\right]^{1/2}$. When the process is in-control, $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 = \mathbf{0}$ and $\delta = 0$, whereas when the process is out-of-control, $\boldsymbol{\mu}_1$ is different from $\boldsymbol{\mu}_0$, i.e., $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0 \neq \mathbf{0}$ and $\delta > 0$.

2.4 The Directional Invariance Property of Multivariate EWMA and Multivariate CUSUM Charts

If a process is in statistical control, the $p \times 1$ random vector $\mathbf{X}_t$, representing the p quality characteristics to be monitored simultaneously at time t has a mean vector, $\boldsymbol{\mu}_0$

and a covariance matrix, Σ. One possible effect of an assignable cause is that it will lead to a process having an out-of-control mean vector, μ_1 but with the same covariance matrix Σ, i.e., a shift of the process parameters from (μ_0, Σ) to (μ_1, Σ). This results in a shift in the process mean vector (Palm, 1990).

Multivariate charts can be divided into two categories, i.e., the directionally invariant schemes and the direction specific schemes. The ARL performance of the directionally invariant control charts is determined solely by the distance of the off-target mean vector, μ_1 from the on-target mean vector, μ_0 and not by the particular direction of that departure from the mean. Distance is defined as the square root of the noncentrality parameter given as $\delta = \left[(\mu_1 - \mu_0)' \Sigma^{-1}(\mu_1 - \mu_0)\right]^{1/2}$, which describes the size of a shift (Lowry and Montgomery, 1995). Typically, most symmetric two-sided univariate control charts are directionally invariant, whereas multiple univariate schemes used for a multivariate process are not directionally invariant. These types of multivariate charts are generally used for the detection of shifts in the process parameters along their respective axes, and must be aimed in a particular direction. Another way of using these charts is to aim them in the direction of the principal components (Lowry and Montgomery, 1995).

It is well known that the χ^2 chart is directionally invariant. For example, suppose that there are $p = 2$ dimensions, the target mean vector is $(0, 0)'$ and the covariance matrix is the 2 by 2 identity matrix. Then, because of directional invariance, the χ^2 chart has the property that the ARL is the same for any vector that is the same distance from the target. Thus, for example, shifts from the target value for the mean vector to $(1,0)'$, $(0,-1)'$, $(0.5\sqrt{2}, 0.5\sqrt{2})'$ and $(-0.5\sqrt{2}, -0.5\sqrt{2})'$ all have the same ARL (Pignatiello and Runger, 1990).

It has been noted by Pignatiello and Runger (1990) that many multivariate procedures, such as the χ^2 chart, Hotelling's T^2 chart, and most of the multivariate CUSUM charts, are directionally invariant. Lowry and Montgomery (1995) and other researchers such as Linna et al. (2001) also note the directional invariance property of many of these multivariate control charting methods which include the MEWMA chart.

It is important to emphasize that there are two implicit assumptions in the run length comparisons based on the noncentrality parameter. First, it is assumed that any shift from the in-control mean vector, regardless of the size of the shift, is to be detected as quickly as possible. Second, it is assumed that a shift from $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ to $\boldsymbol{\mu} = \boldsymbol{\mu}_1$ is to be detected as quickly as a shift from $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ to $\boldsymbol{\mu} = \boldsymbol{\mu}_2$ if $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_0)' \Sigma^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_0)$, that is, the run length is a function of the square root of the noncentrality parameter, δ (Lowry et al., 1992).

The performances of the multivariate control charts, i.e., the MEWMA and MCUSUM charts considered in this study are also directionally invariant. This indicates that the properties of these control charts depend on $\boldsymbol{\mu}_1$ only through the value of the square root of the noncentrality parameter, δ. This simplifies the evaluation of the properties of the control charts.

2.5 Measures of Performance Evaluation of Control Charts

2.5.1 Average Run Length (ARL)

The performance of a control chart can be evaluated in terms of its sensitivity to detect shifts in the parameter that is monitored. This sensitivity can be measured by the number of subgroups taken until the chart signals a shift. For each subgroup a point is plotted on the control chart. The number of plotted points until an out-of-control signal occurs is a discrete random variable, usually called the run length and the ARL is the expected value of this random variable. A sequence of ARL values, for a given control scheme and a set of process average shifts, is called the scheme's ARL profile. ARL profiles are useful when alternative quality control schemes are evaluated and compared (Klein, 1997).

2.5.2 Median Run Length (MRL)

Palm (1990) points out that the run length distributions are usually highly skewed, hence the ARL should not be used as a sole measure of a chart's performance and that medians may be more useful than averages as measures of centrality. Hence, the MRL can be used in conjunction with the ARL and percentiles of run length distribution, since it is a better measure of central tendency for a skewed run length distribution (Dyer et al., 2003). The MRL is estimated to be the 50^{th} percentage point of the probability distribution of the run length.

2.5.3 Percentiles of Run Length Distribution

The sensitivity of a control chart is often summarized using the mean of the run length distribution, i.e., the ARL. Designs and analyses of the control schemes are generally based on the ARL considerations. However, Palm (1990) does not believe that this single parameter contains enough information on the actual run length distribution to make it particularly useful in practical applications and that more than

ARL may be needed. A practitioner will be more interested in percentiles or percentage points of the run length distribution because they give more details regarding the expected behavior of the run lengths. The percentile of a run length distribution is the cumulative proportion or percent of signals given by the number of plotted statistics following the shift and it can be referred to as the cumulative distribution function of the run length. It should be noted that the percentiles of the run length completely characterizes the run length distribution, while the ARL does not.

Although it may be of interest to obtain a median run length profile for a given scheme, such a criterion does not directly address another practical concern, namely the occurrence of too many early false, out-of-control signals, represented by the low percentiles of the run length distribution (Klein, 1997). Crowder (1987a) points out that in many cases, a practitioner may be concerned with the probability of early false out-of-control signals for a given control scheme. Setting the in-control ARL at a desired level may not ensure that the probability of an early false signal is acceptable. He recommends that once ARL considerations have led to a particular control scheme, an analysis of the probability of such early false signals should be done. Both Palm (1990) and Crowder (1987a) have noted that the in-control (i.e., zero process mean shift) run length distributions for the control schemes also deserved consideration. The in-control run length distribution is the distribution of the number of plotted statistics until a false out-of-control signal is given by the control scheme. Thus, false out-of-control signals occur when the process is in-control. The early false out-of-control signals are signals associated with the in-control run lengths which are less than the ARL. The probabilities of the occurrences of these early signals are reflected in the lower percentiles of the in-control run length distribution. The 5^{th}, 10^{th} and 25^{th} percentile values of the in-control run length distribution are usually used to represent the early false out-of-control signals. Using these percentiles as a secondary criterion, and assuming that two schemes have essentially the same ARL profiles, the scheme with the higher run lengths at each of these percentiles is chosen. The reason for this choice is that the scheme chosen will, in the long run, give rise to less frequent, short, false out-of-control signals. For example, suppose that scheme A has run lengths of 5, 52 and 152 at the 5^{th}, 10^{th} and 25^{th} percentiles of the run length distribution, while scheme B has run lengths of 23, 64 and 155 at these three percentiles. Then, if these two schemes have the same ARL profiles, scheme B would be preferred.

As a conclusion, parameters of the run length distribution are generally accepted as appropriate measures of a chart's performance and they are used in practice to select an appropriate control chart with the ARL being the most commonly used measure. Supplementing the ARL with the percentiles of the run length distribution gives a more complete picture of a chart's performance (Jones et al., 2004). Thus, the performance criteria of a chart designed to detect a specified shift in the process mean consists of measures based on the ARL, MRL and percentiles of the run length distribution. These measures allow a thorough study of the performance of a control chart.

CHAPTER 3
MARKOV CHAIN APPROACH FOR MULTIVARIATE EWMA AND MULTIVARIATE CUSUM CHARTS BASED ON AVERAGE RUN LENGTH

3.1 Introduction

There have been considerable number of approaches used to study the distributions and expectation of run lengths for quality control schemes, and one of the methods is the Markov chain approach. The Markov chain approach is introduced by Brook and Evans (1972) for evaluating the performance of the one-sided CUSUM chart. This approach is used by many authors to derive the ARL or the entire run length distribution for various control chart schemes to study the run length characteristics for the quality control schemes. Since each author focuses on one or more charts, different Markov chain applications are tailored to each case. For instance, the state space of the Markov chain has been formulated differently by different authors, and in several cases it is not specified at all (Fu et al., 2003). For example, Lucas and Saccucci (1990) consider the Markov chain approximation to examine the performance of a two-sided EWMA chart and provide design recommendations for the chart. Crowder (1987a) derives an integral equation for the EWMA chart and a computer program is presented by Crowder (1987b) to calculate the ARL of the EWMA chart using integral equation. Champ and Rigdon (1991) show that if the product midpoint rule is used to approximate the integral equation, the integral equation approach and the Markov chain approach yield the same approximation for the ARL. Jiang et al. (2000) applies a two-dimensional Markov chain model to approximate the run length of an autoregressive moving average chart. Calzada and Scariano (2003) study the integral equation and Markov chain approaches for computing the ARL of a two-sided EWMA chart. Fu et al. (2003) introduce a general unified framework on the Markov chain embedding technique which is based either on a simple boundary crossing rule, or on a compound rule. Recently, the application of the Markov chain approach is used by Costa and Magalhaes (2005) for developing the cost function of a two-stage $\bar{X}$

chart. A design procedure for a dual CUSUM chart is developed by Zhao et al. (2005) and an analytical formula for the ARL calculation is obtained via the Markov chain method. Magalhaes et al. (2006) obtain the performance measures of adaptive control charts through the Markov chain approach. The steady-state ARL of a synthetic control chart is evaluated by Costa and Rahim (2006) using the Markov chain model. Shu and Jiang (2006) develop a two-dimensional Markov chain model to analyze the performance of adaptive CUSUM charts.

For the MEWMA scheme, Runger and Prabhu (1996) describe a two-dimensional Markov chain approach to determine the run length performance of a MEWMA control chart. Prabhu and Runger (1997) use the Markov chain method to provide design recommendations for the MEWMA chart that parallels to many results provided for the univariate EWMA by Lucas and Saccucci (1990). Molnau et al. (2001b) then present a computer program that calculates the ARL for the MEWMA control chart using the Markov chain approximation. Rigdon (1995a,b) considers an integral equation and a double integral equation to calculate the in-control and out-of-control ARLs for the MEWMA scheme respectively. Bodden and Rigdon (1999) provide a computer program for the in-control ARL approximation of the MEWMA chart that can be expressed as the solution of an integral equation.

For the MCUSUM scheme, Crosier (1988) studies a Markov chain approach for the approximation of the in-control ARL of a MCUSUM procedure. In this thesis, the one-dimensional Markov chain approach is used for the in-control case of the MCUSUM chart but the proposed Markov chain approach is based on the method of Runger and Prabhu (1996). The Markov chain approach for the MCUSUM chart in this thesis is based on a noncentral chi-square distribution which is different with the method of Crosier (1988), where Crosier (1988) applies a chi-square distribution with the state space of the Markov chain also formulated differently. Consequently, the proposed technique to develop a model for the in-control case of the MCUSUM chart is similar to the MEWMA chart of Runger and Prabhu (1996). As discussed in Section 4.4, the two-dimensional Markov chain approach is infeasible to compute the ARL, MRL and the percentiles of the run length distribution of the out-of-control case for the MCUSUM chart. Consequently, the out-of-control run length performance of the MCUSUM chart is evaluated by the simulation method, discussed in Section 5.3.

For ease of interpretation, the assumptions that are initially made for this thesis are (i) the target mean vector, $\boldsymbol{\mu}_0$ is a zero vector, $\mathbf{0}$ and the covariance matrix, Σ is an identity matrix, $\mathbf{I}$ (although this assumption is made, in practice any process parameters, $\boldsymbol{\mu}_0$ and Σ can be considered provided that their values are known), (ii) the sample size is one; that is to say, the data are structured only as individual observations, and (iii) the process is a multivariate normal process with independently and identically distributed observations.

3.2 The Basic Theory of Markov Chain Approach for Evaluating the Average Run Length

The basic idea for defining the states of a Markov chain is to partition the set of real numbers into a finite set of intervals and to choose a number i in each interval to approximate the true state of the process. This means that by discretizing the values of the statistic used in the Markov chain, q_t, to a finite set of possible values, the run length distribution of a control scheme can be approximated using results from the theory of the Markov chain (Champ and Ridgon, 1991).

Let a control scheme be modeled as a Markov chain, the range of the possible values of statistic q_t is discretized into the $m+2$ states with state 0: $q_t = 0$,

state i: $(i-1)g < q_t \le ig$,

state $m+1$: $q_t > \text{UCL}$

where $i = 1, 2,\ldots, m$, the thickness of each state is $g = \dfrac{2\text{UCL}}{2m+1}$, and UCL is the limit of the control region (Hawkins, 1992).

The absorbing state (i.e., state $m+1$) which extends past the UCL and corresponds to a value for which a signal is given by a control chart scheme is referred to as the out-of-control state. The remaining states are called the transient states, i.e., states $0, 1, 2,\ldots, m$ are referred to as the in-control states. The state 0 corresponds to a zero value for the statistic q_t and the state $m+1$ corresponds to a value of the statistic q_t greater than the control limit of the control chart. The transition probability matrix for a control chart scheme can be partitioned as (Brook and Evans, 1972)

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_t & (\mathbf{I}-\mathbf{P}_t)\mathbf{1} \\ \mathbf{0}' & 1 \end{pmatrix} \tag{3.1}$$

where $\mathbf{I}$ is the $(m+1)\times(m+1)$ identity matrix, $\mathbf{0}'$ is a $1\times(m+1)$ null vector and $\mathbf{1}$ is a $(m+1)\times 1$ column vector of ones. The ij^{th} entry of the transition probability matrix $p_{ij} = \Pr(q_t \text{ in state } j \mid q_{t-1} \text{ in state } i)$ represents the probability that the statistic q_t moves from state i to state j. The first $m+1$ rows and columns of $\mathbf{P}$ correspond to the in-control states. All the absorbing states can be combined into a single absorbing state, i.e., state $m+1$. Here, the absorbing state is the last row and column in $\mathbf{P}$. Hence, the sub-matrix $\mathbf{P}_t$ contains the in-control transition probabilities for moving from one transient state to another transient state. The probabilities of moving from a transient state to the absorbing state are found by subtraction since the elements in the rows of the transition probability matrix must sum to one.