

Algorithm of Multivariate Process Quality Control Based on Joint Control Domain Conversion

NIU Qing

Institute of Industrial Design, Lanzhou Jiaotong University, Lanzhou, China, 730070

Key words: Quality control; Joint control domain; Principal components analysis; Control charts

Abstract: Aiming at the problem of multivariate process quality control, an algorithm based on joint control domain transformation is proposed. By using the principal components analysis method, original quality vector is converted to a new one whose components are unrelated, thus the joint control domain of new quality vector is a hyper ellipsoid whose axes are parallel with the corresponding axes in multidimensional space, and is contained in the hyper cube formed by the intersection of every components' control limits, so the process quality control system is simplified and the number of control charts is reduced. Theoretical analysis and experiments prove that the proposed algorithm based on joint control domain transformation could monitor the shift of process quality effectively, and the efficiency of process quality control system is improved.

Introduction

In modern global market, products' quality is one of the key factors which greatly influence the competitiveness of enterprises. Theory of statistical process control (SPC) and its tool, i.e., control chart proposed by Shewhart solved the problem of univariate process quality control. But in modern manufacturing, a process usually contains more than one quality component. Many researches and applications reveal that error conclusion will be caused if the multivariate process control problem is simplified as a series of univariate ones, because in this way, the relationship of quality components is neglected and the control information of whole process is incomplete [1-4]. In order to monitor the relationship of quality components, T^2 control chart was proposed by Hotelling originally, which is regarded as the basis theory of multivariate statistical process control (MSPC).

For a p -dimensional process quality, the control system based on MSPC contains p Shewhart control charts to monitor the quality components, and a T^2 control chart to monitor the relationship of quality components, thus the scale of control system is large. In order to reduce the number of control charts, theories of neural network is introduced to the field of multivariate process quality control [5-7]. However, in these intelligent methods, a large number of empirical data is needed to train the network; furthermore, the structure of network is mainly designed for specific applications and thus less of generality.

On the basis of above researches, an algorithm of multivariate process quality control based on joint domain transformation is proposed in this paper. By using principle component analysis method, the original quality vector is converted to a new one whose components are unrelated, so the joint control domain of new quality vector is contained in the intersection of every component control limits in multidimensional space, so the process quality control system is simplified and the efficiency can be improved greatly.

Structure of multivariate process quality control system

Suppose $\mathbf{X}=(x_1, x_2, \dots, x_p)^T$ is a p -dimensional process quality and follows the normal distribution $N(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$. Let $\mathbf{X}' = \mathbf{X} - \boldsymbol{\mu}_X$, in general, \mathbf{X}' is also denoted as \mathbf{X} , and the distribution function of \mathbf{X} is

$$f(\mathbf{X}) = (2p)^{-\frac{p}{2}} |\boldsymbol{\Sigma}_X|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{X}^T \boldsymbol{\Sigma}_X^{-1} \mathbf{X}\right) \quad (1)$$

Assume α_X is the false probability, then in p -dimensional space, region D_X exists to meet the following equation

$$\int_{D_X} (2p)^{-\frac{p}{2}} |\Sigma_X|^{-\frac{1}{2}} \exp(-\frac{1}{2} \mathbf{X}^T \Sigma_X^{-1} \mathbf{X}) d\mathbf{X} = 1 - \alpha_X \quad (2)$$

Where D_X is the joint control domain of \mathbf{X} . By the symmetry of normal distribution, we know that $f(\mathbf{X})$ is a constant on the boundary of D_X . That is, for a specific false probability α_X , there is a corresponding constant k_X to meet the following equation

$$f(\mathbf{X}) = k_X \quad \mathbf{X} \in \partial D_X \quad (3)$$

Where ∂D_X is the boundary of D_X . From (1) and (3), there is the following equation

$$\mathbf{X}^T \Sigma_X^{-1} \mathbf{X} = -2 \ln[k_X (2p)^{\frac{p}{2}} |\Sigma_X|^{-\frac{1}{2}}] \quad (4)$$

Considering the variation of process quality, Σ_X and Σ_X^{-1} are both real symmetric positive definite matrix, $\mathbf{X}^T \Sigma_X^{-1} \mathbf{X} > 0$. Let $-2 \ln[k_X (2p)^{\frac{p}{2}} |\Sigma_X|^{-\frac{1}{2}}] = R_X^2$, equation (4) can be written as

$$\mathbf{X}^T \Sigma_X^{-1} \mathbf{X} = R_X^2 \quad (5)$$

Equation (5) means that the joint control domain of \mathbf{X} is a hyper ellipsoid centered on the origin of coordinates. In general, any axis of this hyper ellipsoid is not parallel with the corresponding coordinate axis, and the angle is determined by the covariance σ_{ij} . When point $\mathbf{X}=(x_1, x_2, \dots, x_p)^T$ is in the hyper ellipsoid, it means that the relationship of components is normal.

The control limits of x_i are two hyper planes in p -dimensional space which are perpendicular to the corresponding coordinate axis. The intersection of all components' control limits is a hyper cube centered on coordinate origin. When point $\mathbf{X}=(x_1, x_2, \dots, x_p)^T$ is in this hyper cube, it means every components are under control.

The multivariate process quality is normal means all components and their relationship are both under control, that is, point $\mathbf{X}=(x_1, x_2, \dots, x_p)^T$ must be in the intersection of hyper cube and hyper ellipsoid in p -dimensional space. In general, there are non-overlapping regions between the joint control domain (hyper ellipsoid) and the intersection of components' control limits (hyper cube). For example, in bivariate process quality $\mathbf{X}=(x_1, x_2)^T$, the joint control domain is a ellipse, the intersection of two components' control limits is a rectangle, shown in figure 1, and the shaded area are the non-overlapping regions. So in multivariate process quality system, the Shewhart control charts to monitor every components and T^2 control chart to monitor the relationship of all components should be established, the number of control charts in this system is $p+1$.

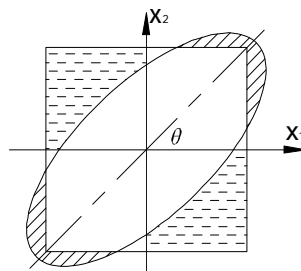


Fig.1 Joint control domain and component control limits of binary process quality

Simplification of process quality control system

In above multivariate process quality control system, the number of control charts is $p+1$, thus the whole control procedure is complicated, so it is necessary to find a new way to simplify the

control system and improve the control efficiency.

Basic idea. In order to reduce the number of control charts in above control system, take the bivariate process quality as an example. In figure 1, if the ellipse is rotated properly to make its axes coincide to the corresponding coordinate axes, then the ellipse will be contained in the rectangle, shown in figure 2. This means that if the relationship of x_1 and x_2 is normal, then x_1 and x_2 are both normal, the process quality is under control. Now the process quality can be monitored by only one T^2 control chart.

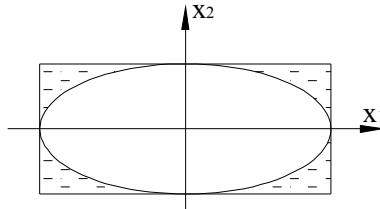


Fig.2 Joint control domain and component control limits after rotation

If above idea is generalized to p -dimensional space, the conclusion is similar. So the core of control system simplification is converting the original process quality \mathbf{X} to a new one, denoted as \mathbf{Y} , where \mathbf{Y} 's joint control domain is a hyper ellipsoid whose axes are coincide to the corresponding coordinate axes, then by monitoring the relationship of \mathbf{Y} 's components, the whole process quality can be monitored. The new process quality \mathbf{Y} should meet the following conditions:

- (1) The sum of components information in \mathbf{X} should equal to the one in \mathbf{Y} ;
- (2) \mathbf{Y} follows the normal distribution, .i.e., $\mathbf{Y} \sim N(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$;
- (3) $\boldsymbol{\Sigma}_Y$ is a diagonal matrix, .i.e., the covariance of y_i and y_j ($i \neq j$) is 0.

The first condition ensures that there is no information loss in conversion; the second one ensures that the joint control domain of \mathbf{Y} is a hyper ellipsoid in p -dimensional space, .i.e., by the similar derivation from equation (1) to (5), the joint control domain of \mathbf{Y} can be written as

$$\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y} = R_Y^2 \quad (6)$$

Where $R_Y^2 = -2 \ln[k_Y (2\pi)^{p/2} / |\boldsymbol{\Sigma}_Y|^{1/2}]$, k_Y is a constant determined by the false probability α_Y , and α_Y can be calculated by α_X .

The last condition ensures that the axes of the hyper ellipsoid which is represented as the joint control domain of \mathbf{Y} is coincide to the corresponding coordinate axes.

Conversion of joint control domain. The information of a component is usually represented by its variance [8], so the sum of variance after conversion should equal to the one before conversion. A normal vector follows normal distribution after linear conversion. Principal components analysis method is a linear conversion in which the sum of components' variance is constant, and in the new quality vector, the covariance of different components is 0. This method meet all the above three conditions.

Let $\mathbf{X} = (x_1, x_2, \dots, x_p)^T$ is the original process quality, $\mathbf{Y} = (y_1, y_2, \dots, y_p)^T$ is the principal components vector, the conversion equation is

$$\mathbf{Y} = \mathbf{P}^T \mathbf{X} \quad (7)$$

Where $\mathbf{P} = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p)$, \mathbf{e}_i is the eigenvector related to λ_i which is the i th eigenvalue of $\boldsymbol{\Sigma}_X$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. After linear conversion, $\mathbf{Y} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_Y)$, where

$$\boldsymbol{\Sigma}_Y = \mathbf{P}^T \boldsymbol{\Sigma}_X \mathbf{P} = \text{diag}(I_1, I_2, \dots, I_p) \quad (8)$$

From equation (6) and (8), we know the joint control domain of \mathbf{Y} is

$$\sum_{i=1}^p \frac{y_i^2}{I_i} = R_Y^2 \quad (9)$$

Equation (9) means that the joint control domain of \mathbf{Y} is a hyper ellipsoid whose axes are coincide to the corresponding coordinate axes and is contained in the hyper cube formed by the intersection of every component's control limits. So in the new control system, the whole process quality can be monitored by a T^2 control chart which is used to monitor the relationship of new quality components. The establishment of control system and its rules is shown as follow

- (1) Calculate the eigenvalues and eigenvectors of Σ_X , and establish the conversion matrix P ;
- (2) Subtract μ_X form \mathbf{X} , then calculate the new process quality vector \mathbf{Y} by equation (7);
- (3) Calculate $T_Y^2 = \sum_{i=1}^p \frac{y_i^2}{I_i}$. If $T_Y^2 \leq R_Y^2$, the process is normal, otherwise, the process quality is out of control.

Key parameters. In application, it needs to confirm the key parameters, including \mathbf{Y} 's false probability α_Y and joint control domain limit R_Y^2 . From the properties of linear conversion and normal distribution, it's obvious that \mathbf{X} and \mathbf{Y} have the same false probability, .i.e.

$$\mathbf{a}_Y = \mathbf{a}_X \quad (10)$$

Form equation (8), we know

$$\Sigma_Y^{-1} = P^{-1} \Sigma_X^{-1} (P^T)^{-1} \quad (11)$$

Then from equation (6), (7), (11), and consider that P is an orthogonal matrix, it can be concluded that

$$R_Y^2 = \mathbf{Y}^T \Sigma_Y^{-1} \mathbf{Y} = \mathbf{X}^T \Sigma_X^{-1} \mathbf{X} = R_X^2 \quad (12)$$

Equation (10) and (12) means that in principal analysis method, the false probability and control limit of joint control domain of process quality are constant.

Form the knowledge of multivariate statistical analysis, when μ_X and Σ_X are known, $R_X^2 \sim \chi^2(p)$, where p is the freedom of χ^2 distribution. Then for any given false probability α_X , R_X^2 can be calculated by the following equation

$$R_X^2 = c_{\alpha_X}^2(p) \quad (13)$$

Sample analysis

A process contains 3 components, denoted as $\mathbf{X}=(x_1, x_2, x_3)^T$. 20 data are gathered to calculate the unbiased estimation of \mathbf{X} 's mean vector and covariance matrix as below

$$\hat{\mathbf{m}}_X = (22.8710, 32.8815, 55.1795)^T$$

$$\hat{\Sigma}_X = \begin{bmatrix} 34.6316 & 30.3684 & -1.0000 \\ 30.3684 & 30.8158 & -2.5526 \\ -1.0000 & -2.5526 & 6.8158 \end{bmatrix} \times 10^{-5}$$

By calculating the eigenvalues and eigenvectors of $\hat{\Sigma}_X$, the conversion matrix P is

$$P = \begin{bmatrix} -0.7277 & 0.1970 & -0.6570 \\ -0.6845 & -0.1473 & 0.7139 \\ 0.0438 & 0.9693 & 0.2421 \end{bmatrix}$$

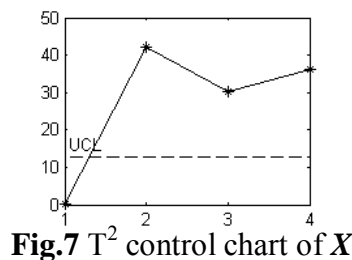
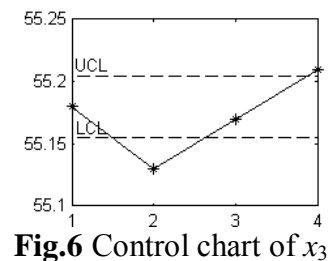
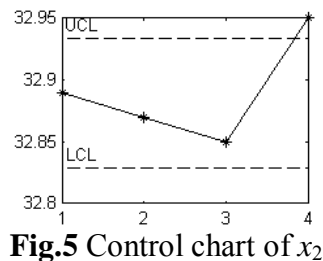
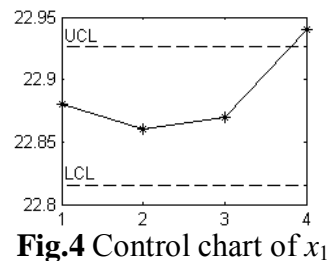
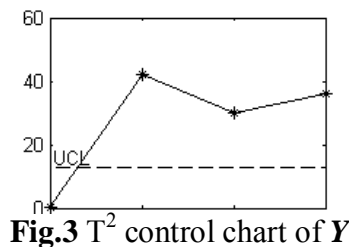
In table 1, column 2~4 are 4 quality data. Subtract \hat{m}_x from every row and the convert the result by equation (7), shown as column 5~7.

Tab.1 The instance data and transformation results

No.	x_1	x_2	x_3	y_1	y_2	y_3
1	22.88	32.89	55.18	-0.0123	0.0010	0.0003
2	22.86	32.87	55.13	0.0137	-0.0485	-0.0130
3	22.87	32.85	55.17	0.0219	-0.0048	-0.0241
4	22.93	32.94	55.20	-0.0958	0.0331	0.0110

Suppose false probability $\alpha_Y=0.005$, by χ^2 distribution table we know $R_Y^2=12.838$. Plot the T^2 control chart of Y by equation (9), shown as figure 3. Its clear that the last 3 quality data are out of control.

In order to validate the above conclusion, establish Shewhart control charts of every component in X , and X 's T^2 control chart, shown as figure 4~7. Form these figures we know that in the 2nd point, the relationship of X and x_3 are abnormal, so the process quality is out of control; in the 3rd point, the cause of process quality out of control is the relationship of X is abnormal; in the last one, all components and their relationship are abnormal, these cause the process quality out of control. The conclusion is same as the one based on joint control domain conversion method.



Conclusion

Linear transformation based on the principal components analysis method could convert the joint

control domain of multivariate process quality to a hyper ellipsoid whose axes coincide to the corresponding coordinate axes in p -dimensional space, and the converted hyper ellipsoid is contained in the hyper cube formed by the intersection of every components' control limits, so the number of control chars can be reduced from $p+1$ to 1, the structure of control system is simplified. Theoretical analysis and experiments prove that the proposed algorithm could monitor the abnormal shift of process quality effectively, and the efficiency of process quality control system is improved highly.

References

- [1] Zhang Gongxu, Diagnosis Theory of Two Kinds of Quality and Its Application, Science press, Beijing, 2001
- [2] Wang Suli, Ma Yizhong, Study and Software Design for the Technology of Multivariate Quality Control, J. Computer Engineering and Applications. 41 (2005) 186-188
- [3] Fang Jitao, Yang Musheng, The Quality Monitoring System of Manufacturing Process Based on The Shewhart control Chart, J. Journal of Shandong University of Technology. 20 (2006) 62-65
- [4] Tian Haijing, Pan Erdun, Multivariate Statistical Process Control for Autocorrelated Process, J. Journal of Shanghai Jiaotong University, 42 (2008) 496-499
- [5] R.S.Guh, Simultaneous Process Mean and Variance Monitoring Using Artificial Neural Network. Computers and Industrial Engineering, 58 (2010) 739-753
- [6] A.Shaban, M.Shalaby, E.Abdelhafiez, Automated Identification of Basic Control Charts Patterns Using Neural Networks, J. Journal of Software Engineering and Applications, 3(2010) 208-220
- [7] Jia Feng, Jiang Pinyu, Liu Daoyu, Error Propagation Control Method for Multistage Batches Maching Processes of Blades, J. Computer Intergraded Manufacturing System, 18(2012) 76-88
- [8] Xiang Dongjin, Practical Multivariate Statistical Analysis, China university of geosciences press, Wuhan, 2005