Statistical Process Adjustment of Multivariate Processes with Minimum Control Efforts

Li Wang[1], Kaibo Wang[2]

[1]: Department of Industrial and System Engineering, University of South California, Los Angeles, CA 90089

[2]: Department of Industrial Engineering, Tsinghua University, Beijing 100084, China. Email: kbwang@tsinghua.edu.cn

Abstract – **In controlling a multiple-input-multiple-output (MIMO) process, usually all control variables have to be adjusted at each step, which may incur high adjustment cost. This paper proposes a Lasso adjustment algorithm, which minimizes the number of variables to be adjusted at each step. Simulation results show that the proposed algorithm can maintain acceptable output deviations while reduce the number of variables need to be adjusted significant.**

Keywords – **Statistical Process Adjustment, Statistical Process Control, Variable Selection**

I. INTRODUCTION

To meet challenges posed by ever growing quality requirements in manufacturing processes, especially in semiconductor industry, the Statistical Process Adjustment (SPA) concept has been proposed by researchers [1]. SPA combines conventional statistical process control (SPC), time series analysis and control theory and is proven with capability to reduce process variation [1, 2]. The current literatures about SPA, however, mainly focus on off-target cost, ignoring other important aspects of the manufacturing cost structure. Another limitation is that most literatures in this area focus on single-input-single-output (SIMO) problems or multiple-input-single-output (MISO) problems, while multiple-input-multiple-output(MIMO) problems are more common in use [3]. In the following sections of the introduction, after a brief introduction of SPA, the adjustment cost will be discussed, and we'll layout our multiple processes control scheme.

Traditionally, there are two approaches developed rather independently in the area of quality assurance: SPC is mainly developed in the part industry while EPC in the process industry. Although they often perform quite well, however, both approaches have serious limitations. SPC's ability in finding out potential assignable causes does not help to eliminate such causes once they are found, and its ability to deal with auto-correlated data is often less than satisfactory. In the other hand, EPC can adjust the machine variables to control the process variables as a way to reduce the variation in the final product variables, but it is often criticized as a bandage cover the wound not cure it as it does not try to find out the shock but only compensate its consequences [3]. As a result, there are some researchers proposed to combine EPC and SPC with improved performance in reduced output variation [1-3] and such studies can be categorized as SPA.

Many statisticians and engineers in the SPC area believe that adjustments are unnecessary in practice due to Deming's findings that several adjustment schemes just make larger variation than not adjustment at all in his famous funnel experiment [1]. However, Deming's conclusions are only applicable to the condition he originally set, which includes an originally on-the-target process. In other circumstances, the conclusion may not apply. For example, in some applications, there are set-up adjustment problems and the process will not come to targeted value until some adjustments are made [1, 3]. There is no simple criterion as whether the average of the process is moving to determine when one needs to utilize SPA. Such decisions can only be made based on process models and cost structures [1].

As an example of SPA, we introduce a lapping process in silicon wafer manufacturing. This manufacturing process utilizes a special wafer lapping machine. Its main purpose is to remove materials from wafer through relatively rotating upper and lower disks with slurry. The procedure can be described as: before one run begins, the operator sets controllable factors of the process, then the lapping machine begins with a low pressure lapping process and then high pressure lapping process. After one lapping run, the operator will determine the next run's parameters based on the removed volume of this run.

There is much attention paid to costs involved in manufacturing process control. Bather [4] had categorized two kinds of cost named running cost and overhaul cost. The former one is a quadratic cost associated with deviation from target, and the latter is a fixed cost associated with the adjustments made. Sampling cost is included in the cost structure by Butte and Tang [3] . In the structure's three components, off-target cost received most attention and its properties and related controller are extensively investigated [5-9]. The other two parts of the cost structure is much less studied as they are often overlooked as insignificant comparing to off-target cost. This argument is true when adjustments are only turning knobs or other trivial activities, but in many other instances, to make the necessary adjustments may cause a stop of production due to equipment re-configure or setup. In such cases, to minimize the number of adjustments is essential for the whole process's profitability (In this paper, we'll not discuss the sampling cost as the sensor

978-1-4244-8503-1/10/$26.00 ©2010 IEEE

technology develops, sampling cost is becoming more insignificant or case dependent; the consideration of sampling cost does not change the derivation and conclusions in this paper).

For example, in above mentioned lapping process of silicon wafer manufacturing, we can adjust inputs as slurry temperature, slurry density, lapping time, pressure and speed. It incurs little cost to adjust lapping time, pressure or speed as it can be done by setting the machine, but to adjust slurry temperature or density will be costly and time consuming, thus significantly increase the equipment's idling cost. Another cost associated with productivity reduction is that professional operators are needed to perform and confirm each parameter adjustment. Besides lost in productivity, too frequent adjustment may bring uncertain influence upon other quality characteristics.

Generally, there are two ways of incorporating adjustment cost into the SPA scheme [1].

- To associate a fixed-cost to each adjustment without considering its magnitude. This approach is closely related to the dead-band adjustment policies;
- To constrain the variance of the input variables, in other words, to constrain the overall adjustment magnitude. This approach is equivalent to assuming quadratic cost in control factors [10].

There are some works trying to combine adjustment cost into a comprehensive scheme. Crowder [11] developed an SPC model to minimized expected overall cost, but his result relied on the assumption of a symmetrical quadratic off-target cost and fixed adjustment cost, which are not often the case in the application. The determination of appropriate factors in the cost models is also difficult in application.

Patterson et al. [12] introduced an adjustment variable selection approach to minimize the number of variables to adjust and thus reduce the number of the unnecessary adjustments. Although their method can reduce the number of variables to be adjusted significantly, the variable choosing process is sometimes arbitrary. Another pitfall of the scheme is that the number of variables to be adjusted need to be specified beforehand. If one wants to change the number, one should perform analysis again which is somewhat cumbersome.

To mitigate problems faced in previous works, we propose a variable selection approach. The basic idea is to use coefficient shrinkage method, for example the Lasso method [13], to automatically include a penalty cost of each adjustment in the cost structure , then based on this and the level of variation reduction one desired, one can continuously control a parameter to balance the two costs involved.

This approach has following advantages:

- With the flexibility to dynamically adapting to the change of cost structure
- Need no specific rates a cost model will require which are often difficult even impossible to get
- Need no complex decisions by operators and thus both reduce his/her workload and the expectation loss due to the wrong choice made

In the following sections, we first introduce our process model, then develop a new Lasso adjustment algorithm. After that, we investigate the algorithms' performance using Monte Carlo simulations.

II. LASSO PROCESS ADJUSTMENT ALGORITHM

As stated before, this paper will focus on MIMO statistical adjustment problems. In this area, Tseng et al. [14] discussed the stable region of MEWMA algorithm and parameter optimization with white and IMA disturbance. Del Castillo [15] applied the double EWMA controller to the MIMO case. These works, however, did not consider the number of adjustment which will be dealt with in this paper.

To establish the process model, we assume that the relationship between the input and output variable is linear; the observation and adjustment opportunity occur at discrete equal-distance time point;

Based on above assumptions, we build our process model at $t = 1, 2, ...$ as

$$\mathbf{Y_t} = \mathbf{B}\mathbf{X_{t-1}} + \mathbf{A} + \mathbf{E_t} \tag{1}$$

where

$\mathbf{X}_{t-1}$: $m \times 1$ input vector

$\mathbf{Y}_t$: $n \times 1$ output vector

$\mathbf{E}_t$: $n \times 1$ white noise part of disturbance vector, in which each element $\overset{i.i.d}{\sim} N(0,1^2)$ and mutually independent of each other.

$\mathbf{B}$: $n \times m$ coefficient matrix

$\mathbf{A}$: $n \times 1$ intercept

$\mathbf{T}$: $n \times 1$ target vector

This is a general model using m inputs to adjust n outputs.

Lasso (Least Absolute Shrinkage and Selection Operator) algorithm [13] was proposed by Robert Tibshirani in 1996. It minimizes the residual sum of squares as the traditional least square estimation method, while adds a constraint that the sum of coefficients' absolute values must less or equal to a non-negative constant. The reason we choose Lasso as the base of our adjustment algorithm can be stated as follows:

One only has to control one parameter to dynamically balance off-target and adjustment cost, namely, the constant in the constraint or its equivalence. For example, if the constant term decreases, the allowed adjustment amplitude becomes smaller, which further results in less adjustment but more centered process output.

Lasso's geometry makes it is natural to reduce the number of adjustment needed. As compared to ridge regression, Lasso's absolute value constraint creates corners that are more likely to produce zero adjustment in some inputs.

Lasso has efficient implementation, namely, Lars (Least Angel Regression) algorithm [16], to accommodate the need for real time adjustment.

With the process model (1), to minimize the sum of output residual squares at time $t\,(t = 2,3,...)$, the adjustment problem's object function can be formulated as to find input $\mathbf{X}_{t-1}$ to:

$$
\begin{aligned}
&\text{Minimize}(\mathbf{T}-\mathbf{Y_t})^T(\mathbf{T}-\mathbf{Y_t}) \\
&= (\mathbf{T}-\mathbf{Y_t}+\mathbf{Y_{t-1}}-\mathbf{Y_{t-1}})^T(\mathbf{T}-\mathbf{Y_t}+\mathbf{Y_{t-1}}-\mathbf{Y_{t-1}}) \\
&= [(\mathbf{T}-\mathbf{Y_{t-1}})-(\mathbf{Y_t}-\mathbf{Y_{t-1}})]^T[(\mathbf{T}-\mathbf{Y_{t-1}})-(\mathbf{Y_t}-\mathbf{Y_{t-1}})] \\
&= [\mathbf{T}'-(\mathbf{Y_t}-\mathbf{Y_{t-1}})]^T[\mathbf{T}'-(\mathbf{Y_t}-\mathbf{Y_{t-1}})](\mathbf{T}'=\mathbf{T}-\mathbf{Y}_{t-1}) \\
&= [\mathbf{T}'-\mathbf{B}(\mathbf{X}_{t-1}-\mathbf{X}_{t-2})]^T[\mathbf{T}'-\mathbf{B}(\mathbf{X}_{t-1}-\mathbf{X}_{t-2})] \\
&(\mathbf{Z}_t=\mathbf{X}_{t-1}-\mathbf{X}_{t-2}) \\
&= (\mathbf{T}'-\mathbf{B}\mathbf{Z}_t)^T(\mathbf{T}'-\mathbf{B}\mathbf{Z}_t) \\
&\text{subject to} \sum_{i=1}^{m} |z_{i,t}| \le \lambda, \text{ where } \lambda \ge 0
\end{aligned}
$$

The constraint means the sum of each input's adjustment must less than or equal to a non-negative constant. We can treat $\mathbf{Z}_t$ as the needed coefficient matrix, $\mathbf{B}$ as predictor matrix and $\mathbf{T}'$ as response in coefficient estimation problem which the original Lasso method intend to solve, and with $\mathbf{Z}_t$ we can find the new input levels $\mathbf{X}_{t-1}$ based on the Lasso algorithm.

III. Performance and Stability Analysis

After the development of adjustment algorithms, we will investigate their performance in minimizing the mean square error (MSE) from target values as well as in reduction of adjustment efforts, then we'll analyze algorithms' stability against parameter estimation.

As it is unpractical to apply different adjustment algorithms in actual manufacturing processes due to time and cost constraints as well as a process's unrepeatable nature, we apply Monte Carlo simulations based on the above mentioned process model.

Although our study is based on above mentioned lapping process example, to protect corporative confidentiality, the parameters in the following simulation model are already treated, then the model parameters can be summarized as follows:

- The process has six inputs and three outputs.
- The original input $\mathbf{X}_0 = (0.74, 0.94, 0.16, 0.49, 0.06, 0.26)^T$.
- The intercept $\mathbf{A} = (0.87, 0.02, 0.42)^T$.
- The coefficient matrix $$\mathbf{B} = \begin{pmatrix} 5.64 & 4.68 & 4.24 & 4.93 & 5.62 & 5.89 \\ 5.21 & 4.47 & 4.84 & 5.21 & 5.78 & 5.36 \\ 5.97 & 4.58 & 5.17 & 5.39 & 5.79 & 4.9 \end{pmatrix}.$$
- The target for output is $\mathbf{T} = (14.41, 13.14, 14.23)^T$, which means that the output is on target originally.

To evaluate the performance of the Lasso adjustment algorithm, the MEWMA algorithm [14] is used as a benchmark. The EWMA controller is widely used in statistical adjustment areas. It is known that the EWMA controller is optimal to minimize MSE with properly tuned smoothing parameter if the disturbance series follows an IMA(1,1) [6]. Therefore, in the following simulation, we assume the process in Model (1) suffers not only initial bias but also an IMA(1,1) disturbance. Its MA coefficient is denoted by θ.

Each process is 200 steps long and 100 replicas are used to calculate the average values of interest.

Mean Square Error Performance

It is worth noting that there is a coefficient in the application of Lasso adjustment algorithm: a constant which represents the absolute value constraining the sum of adjustments. In our simulation, we'll use a fraction number to regularize this parameter, which ranges from 0 to 1. It refers to the ratio of L1 norm of the adjustment vector to the full linear square solution. In general, there is more room to adjust if the number is larger.

First one should investigate the MSE performance of Lasso adjustment algorithm. If should be noted that if there are no adjustment at all, the MSE will be 71.37.

In Fig. 1, different curve denotes different choice of the fraction number. From Fig. 1, one will notice that when the fraction is large, the performance of Lasso adjustment will only slightly worsen as θ increases..

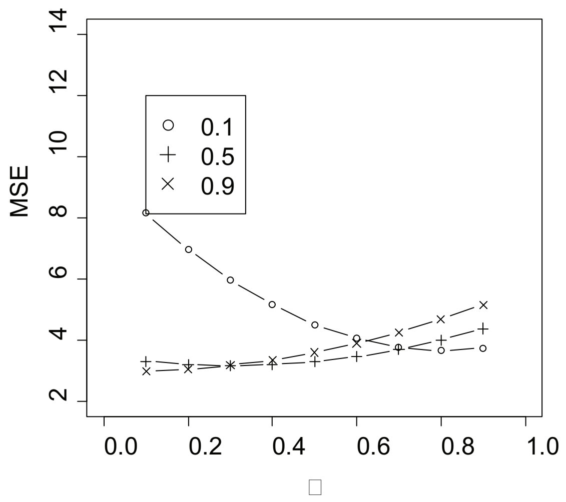

Figure 1. Lasso Adjustment MSE

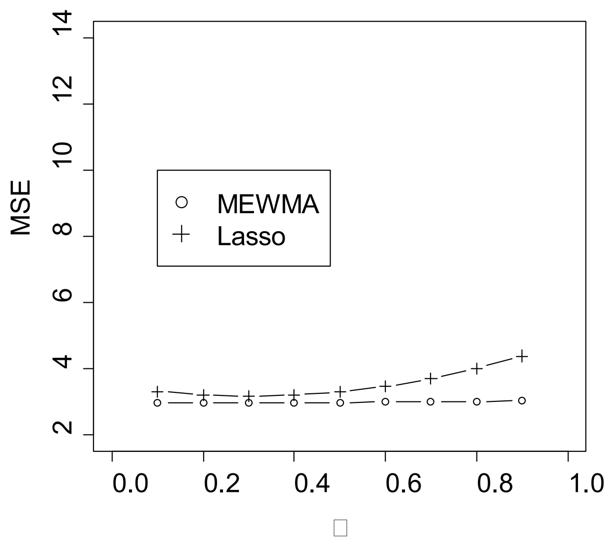

Fig. 2 MSE Comparison between MEWMA and Lasso

After experimenting with fraction number ranging from 0.1 to 1, we find the MSE performance is best when fraction number is 0.5 for the Lasso adjustment algorithm. Comparing the cases with the MEWMA algorithm (smoothing coefficient is obtained as 1□□) in Fig. 2, one will find that 1) when □ is small, all algorithms performs well; 2) The Lasso adjustment algorithm performs nearly as stable as MEWMA algorithm.

Adjustment Effort Comparison

While retaining good MSE performance, our new adjustment algorithms' main advantage is the significant reduction of adjustment effort required.

To illustrate this point, we further study Fig. 3 and Fig. 4 (To avoid jam in the figure, we only keep first three inputs). Instead of zigzag adjustment series of MEWMA, the Lasso adjustment algorithm offers less adjustment which can be seen as flat segments of the adjustment series. More interesting, using the Lasso adjustment algorithm, one needs not to adjust the second input at all, which is a very good characteristic that has the potential to greatly facilitate the adjustment if this input is difficult to control.

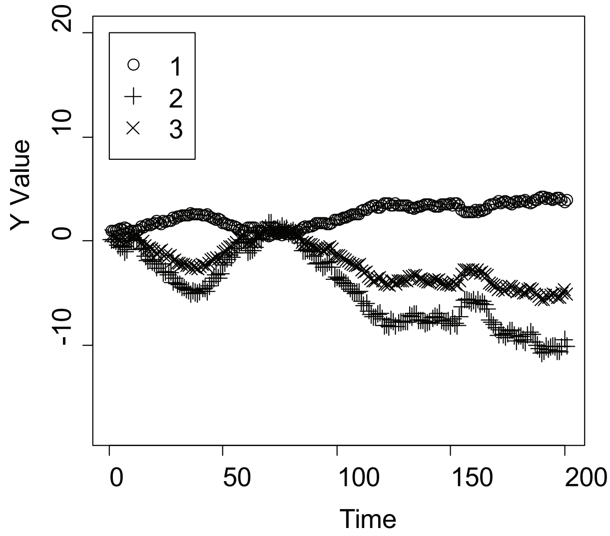

Fig. 3 MEWMA Adjust Trajectory

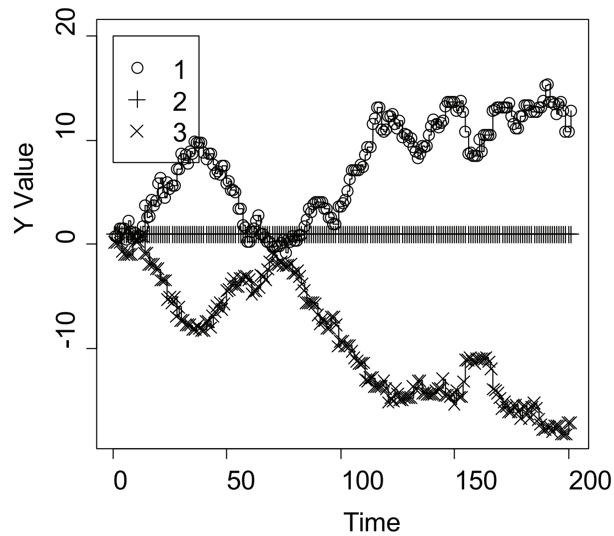

Fig. 4 Lasso Adjust Trajectory

After this simple comparison, we now investigate the adjustment effort performance of the Lasso adjustment algorithm against θ and the fraction number. The initial simulation shows, however, that θ has little influence on this performance criteria. Therefore, we will only study this performance index under different fraction numbers while keep $\theta = 0.5$.

Based on simulated runs, we found that the maximum average number of adjustment is 3 for the Lasso algorithm. The MEWMA algorithm's average number of adjustment, however, is almost a constant 6, which is the maximum number of adjustable variables. Besides a better performance in average number of adjustment index, the Lasso based algorithms' adjustment actions tend to be concentrated on some input variables and some input variables may never need to be adjusted at all, which reduce the complexity of the adjustment scheme and thus reduce potential errors.

IV. CONCLUSION

This paper has proposed a statistical adjustment algorithm based on Lasso to control MIMO processes. The proposed algorithms can greatly reduce the number of adjustment needed while retaining MSE performance of the tradition MEWMA algorithm; the proposed method can result in significant saving in operational cost.

ACKNOWLEDGMENT

This research was supported by the National Natural Science Foundation of China (NSFC) under grant No. 70802034.

REFERENCES

[1]. Del Castillo, E., *Statistical process adjustment: a brief retrospective, current status, and some opportunities for further work.* Statistica Neerlandica, 2006. **60**(3): p. 309-326.

[2]. Montgomery, D.C., et al., *Integrating Statistical Process Control and Engineering Process Control.* Journal of Quality Technology, 1994. **26**(2): p. 79-87.

[3]. Butte, V. and L. Tang, *Engineering Process Control: A Review.* Handbook of Performability Engineering, 2008: p. 203-223.

[4]. Bather, J., *Control charts and the minimization of costs.* Journal of the Royal Statistical Society. Series B (Methodological), 1963. **25**(1): p. 49-80.

[5]. Box, G.E.P. and T. Kramer, *Statistical Process Monitoring and Feedback Adjustment - A Discussion.* Technometrics, 1992. **34**(3): p. 251-267.

[6]. Ingolfsson, A. and E. Sachs, *Stability and sensitivity of an EWMA controller.* Journal of Quality Technology, 1993. **25**(4): p. 271-287.

[7]. Luceno, A. and F.J. Gonzalez, *Effects of dynamics on the properties of feedback adjustment schemes with dead band.* Technometrics, 1999. **41**(2): p. 142-152.

[8]. O'Shaughnessy, P.T. and L.D. Haugh, *An EWMA-based bounded adjustment scheme with adaptive noise variance estimation.* Journal of Quality Technology, 2002. **34**(3): p. 327-338.

[9]. Colosimo, B.M., R. Pan, and E. Del Castillo, *Setup adjustment for discrete-part manufacturing processes with asymmetric cost functions.* International Journal of Production Research, 2005. **43**(18): p. 3837-3854.

[10]. Astrom, K., *Introduction to stochastic control theory.* 1970: Academic Pr.

[11]. Crowder, S.V., *An Spc Model for Short Production-Runs - Minimizing Expected Cost.* Technometrics, 1992. **34**(1): p. 64-73.

[12]. Patterson, O.D., et al., *Methodology for feedback variable selection for control of semiconductor manufacturing processes - Part 1: Analytical, and simulation results.* IEEE Transactions on Semiconductor Manufacturing, 2003. **16**(4): p. 575-587.

[13]. Tibshirani, R., *Regression shrinkage and selection via the Lasso.* Journal of the Royal Statistical Society Series B-Methodological, 1996. **58**(1): p. 267-288.

[14]. Tseng, S.T., R.J. Chou, and S.P. Lee, *A study on a multivariate EWMA controller.* Iie Transactions, 2002. **34**(6): p. 541-549.

[15]. Del Castillo, E. and R. Rajagopal, *A multivariate double EWMA process adjustment scheme for drifting processes.* Iie Transactions, 2002. **34**(12): p. 1055-1068.

[16]. Efron, B., et al., *Least Angle Regression.* Annals of Statistics, 2004. **32**(2): p. 407-451.