DESIGN OF A COMPOSITION ESTIMATOR FOR INFERENTIAL CONTROL OF HIGH-PURITY DISTILLATION COLUMNS

Joonho Shin and Sunwon Park
Department of Chemical Engineering
KAIST
Taejon 305-701, Korea

Moonyong Lee
Department of Chemical Engineering
Yeungnam University
Kyongsan 712-749, Korea

Abstract

In distillation column control, temperature as a secondary measurement is widely used in order to infer product composition. This paper addresses the design of a static PLS estimator using multiple temperatures for estimating the product compositions of the distillation columns. Design guidelines on the development of the composition estimator using PLS regression are presented. It is shown that the loading vector space of the estimator exactly reflects the direction of temperatures due to corresponding input variables by considering an inverse estimation problem. We also discuss the effects of data pre-processing and variable transformation. The estimator based on the guidelines is robust to sensor noise and has a good predictive power.

Keywords

Composition estimator, PLS (Partial-Least-Squares) regression.

Introduction

Product quality measurement is one of the major difficulties associated with the composition control of distillation columns. In distillation columns, the product composition control is performed by using process analyzers such as Gas Chromatography(GC), but the analyzers suffer from large measurement delays, high investment/maintenance costs and low reliability. For these reasons, many workers (Joseph and Brosilow, 1978; Mejdell and Skogestad, 1991; Kresta, Marlin and MacGregor, 1994; Piovoso and Kosanovich, 1994) have studied the inferential model, which estimates the product composition of the distillation column. It was reported that the Partial-Least-Squares(PLS) provides models with good predictive power and robustness to the process noise and sensor failure (Kresta et al., 1994).

In developing the composition estimator using PLS regression, we have to consider the following issues: (1) the determination of the number of factors in PLS regression; (2) the selection of the secondary measurements to be used; (3) the selection of the most effective variable transformation or scaling. These estimator design issues are discussed by analyzing the results of simulation studies on a high-purity binary column example.

Problem Definition

The linear estimator inferring the product compositions of the distillation column may be written by

$$\hat{y} = K\theta \tag{1}$$

where $\hat{y}$ is the estimated composition and θ is the secondary measurement vector. The problems concerned in this work are the design problems stated in the introduction.

Evaluation Criteria

The Explained Prediction Variance(EPV) (Mejdell and Skogestad, 1991) in percent is used to evaluate the performance of the estimator:

$$EPV(k) = 100 \times \left(1 - \frac{MSEP(k)}{MSEP(0)}\right) \quad (2)$$

where the mean squared error of prediction $MSEP(k) = 1/N_{set} \sum_{i=1}^{N_{set}} (\hat{y}_i(k) - y_i)^2$ and N_{set} is the number of data sets(here, N_{set}=64).

The Prediction Error Sum of Squares(PRESS) (Kresta et al., 1994) is also used to evaluate the absolute performance:

$$PRESS = \sum_{i=1}^{N_{set}} (\hat{y}_{D,i} - y_{D,i})^2 \quad (3)$$

Example Column

The steady state simulation of the binary column of normal-hexane and cyclo-hexane with 40 theoretical stages is performed using the rigorous steady state simulator, AspenPlus. The feed stream enters the column at stage 20 as saturated liquid. The nominal operation conditions and the variation of the inputs of the binary column are given in Table 1.

Table 1. Simulation Conditions for the Binary Distillation Column.

	Base case conditions	Variation in steady state reference set
F	1000.0 kmol/hr	Constant
T_F	346.57 K	Constant
z_f	0.5	0.25~0.75
D	500.0 kmol/hr	Vary
Q_B	0.130736×10^8 cal/sec	Vary
P	1 atm	Constant
y_D	0.99	0.97~0.997
x_B	0.01	0.003~0.03

Variable Transformation

Logarithmic transformation of the product compositions was used by many investigators (Joseph and Brosilow, 1978; Mejdell and Skogestad, 1991; Kresta et al., 1994). Several transformation methods were tested by Mejdell and Skogestad (1991). They presented that the logarithmic transformation of both the composition and the temperatures improved the estimation performance. In this work, the same transformation methods are adopted and evaluated to select the most effective variable transformation method.

Multivariate Statistical Methods

Various multivariate statistical methods such as Multiple Linear Regression(MLR), Principal Component Regression(PCR) and PLS are used to build the linear static estimator inferring the product composition in this work. For the convenience of the discussion, only the results of PLS are presented in this paper.

PLS Regression

This method is a variation of the PCR which recently has become popular among analytical chemists (Geladi and Kowalski, 1986). The latent variables are determined in order to have the largest covariance with the dependent variables. The PLS estimator (Mejdell and Skogestad, 1991) based on k factors is given by

$$\mathbf{K_{PLS}} = \mathbf{Q}\left(\mathbf{P}^{\mathrm{T}}\mathbf{W}\right)^{-1}\mathbf{W}^{\mathrm{T}} \quad (4)$$

where $\mathbf{Q}^{p\times k}$ is the loading matrix for the dependent variables (e.g. the product compositions), $\mathbf{P}^{q\times k}$ is the loading matrix for the independent variables (e.g. temperatures and flow rates), and $\mathbf{W}^{q\times k}$ is the matrix formed in the PLS algorithm.

Results of a Case Study

As many researchers (Mejdell and Skogestad, 1991; Kresta et al., 1994) reported, the performance of the inferential model using a single temperature measurement is very sensitive to noise (e.g. EPV decreases by 32 % when the noise 0.1 $^{\circ}C$ is added to the temperature measurement). In order to reduce the sensitivity of the inferential models using a single temperature, we should use the inferential model based on multiple temperature measurements. The design guidelines on the development of the composition estimator will be presented by analyzing the binary column case study in this section. The results presented in this work are based on all tray temperature measurements.

Determination of the Number of Factors in PLS Estimator

There is an optimal model dimension which minimizes the prediction error (Kresta et al., 1994). Therefore, when we build an inferential model, the optimal model dimension (here the number of factors in PLS estimator) should be considered.

Unique temperature profiles are obtained by specifying different values of feed composition z_f, distillate composition y_D and bottom product composition x_B (or z_f, reflux ratio R, and reboiler duty Q_B) at constant pressure. That means the number of degrees of freedom is three for the general binary columns assuming that the unmeasured disturbance is only the feed composition: the binary column system is fully defined by specifying three variables such as z_f, y_D, and x_B (or z_f, R, and Q_B). Therefore the number of directions of the temperature profile affected by the independent variables are clearly three for the distillation column of this case study. To verify this, let's consider the following inverse estimation problem:

$$\hat{\theta} = \mathbf{A}\left[z_f \; y_D \; x_B\right]^{\mathrm{T}} \quad (5)$$

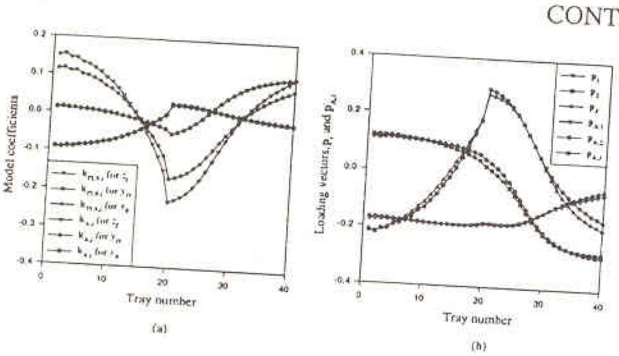

Figure 1. Comparison of (a) $\mathbf{K_{PLS}}$ *and* $\mathbf{K_A}$ *and (b) loading vectors of* $\mathbf{A}$ *and* $\mathbf{K_{PLS}}$.

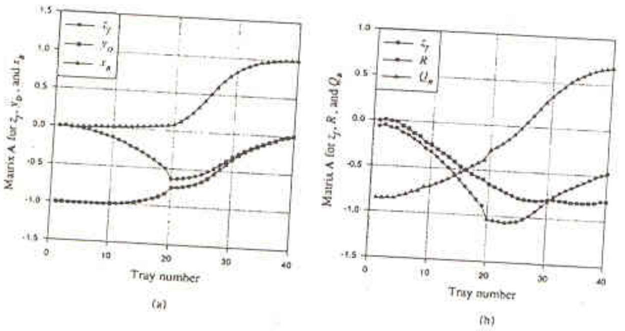

Figure 2. Plot of $\mathbf{A}$ *for input variables (a)* z_f, y_D, *and* x_B *and (b)* z_f, R, *and* Q_B.

One can estimate the temperature profile by the above equation and the inverse problem of it is the estimation of the compositions :

$$[\hat{z}_f \; \hat{y}_D \; \hat{x}_B]^{\mathrm{T}} = \mathbf{K_A} \, \theta \tag{6}$$

From Eqs. 5 and 6, one can easily find

$$\mathbf{K_A} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}} \tag{7}$$

Matrix $\mathbf{A}$ is calculated using z_f, y_D and x_B with unit variance scaling, and the EPV for temperature is 93.87. The EPV value shows that the estimation of the temperature is good using z_f ,y_D, and x_B. Fig. 1(a) compares matrix $\mathbf{K_A}$ with matrix $\mathbf{K_{PLS}}$ of which the number of factors k is three.

We can expect that the SVD(Singular Value Decomposition) of $\mathbf{A}$ may give us the same eigenvectors as the loading vectors of Θ which is the matrix of the temperatures for calibration. The SVD of matrix $\mathbf{A}$ is performed and the largest three eigenvectors ($\mathbf{p_{A,1}}$, $\mathbf{p_{A,2}}$, and $\mathbf{p_{A,3}}$) are plotted with loading vectors($\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$) from the PLS in Fig. 1(b). It is clearly shown that the directions of the temperature profile have a one-to-one correspondency with the input variables z_f, y_D, and x_B.

The plot of matrix $\mathbf{A}$ in Fig. 2(a) shows the direct relation between the temperatures and z_f, y_D, and x_B. One may try to use R and Q_B instead of y_D and x_B. But EPV using z_f, R, and Q_B is only 50.4 and the performance of temperature estimation is not good. Therefore the directions of the temperature profile cannot be explained in terms of z_f, R, and Q_B in the linear form. The plot of the matrix $\mathbf{A}$ for input variables z_f, R, and Q_B is shown in Fig. 2(b).

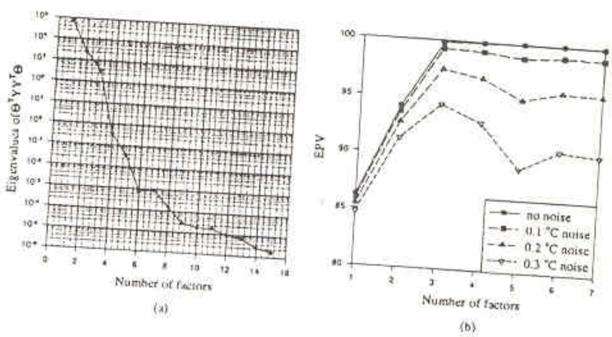

Figure 3. (a) Plot of eigenvalues of $\Theta^{\mathrm{T}}\mathbf{YY}^{\mathrm{T}}\Theta$ *and (b) Effect of measurement noise on EPV.*

Additionally the eigenvalues of $\Theta^{\mathrm{T}}\mathbf{YY}^{\mathrm{T}}\Theta$ for PLS are plotted in Fig. 3(a) where $\mathbf{Y}$ is the matrix of product composition y_D for calibration runs. The magnitudes of eigenvalues drastically decrease as the number of factors increases. One can easily find that the condition number of $\Theta^{\mathrm{T}}\mathbf{YY}^{\mathrm{T}}\Theta$ is directly related to the sensitivity of the PLS estimator with corresponding factors. One should select the first three factors as the condition number increases by the order of 10^3 if the number of factors increases from 3 to 4.

We checked the sensitivity of the composition estimator by the cross-validation procedure (Mejdell and Skogestad, 1991). As shown in Fig. 3(b), the EPV increases monotonically as the number of factors increases. However, when the inferential model with more than 4 factors is applied to noise corrupted data, the EPV decreases. As we expect, the EPV without noise is saturated when the number of factors are three and the estimator with k=3 is robust to the noise. We can conclude that the optimal number of factors in the sense of the robustness and the prediction ability is three for our binary column and the factors describe the effects of z_f, y_D and x_B. This results can be extended for multicomponent distillation columns. The recommended number of factors is equal to the number of independent variables (e.g. z_f, y_D, x_B, and column pressure P) which have influence on the temperature profile. This guideline for determination of the number of factors is verified by other examples of the multi-components columns. The number of factors for BTX column (Kresta et al., 1994) and for a 4 component column is 4 and 5 respectively assuming that the unmeasured disturbance is only the feed composition.

Effects of data pre-processing and variable transformation

The results with unit-variance scaling are better than those with no scaling. Unit variance scaling makes the temperature measurements, which are collinear with T_1, have the same information as T_1(see Fig. 2(a).). The unit variance scaling of the temperature measurements provides us the useful information about T_1 or T_N highly related with y_D or x_B (note that $y_D = f(T_1)$ for the binary column). But the estimator with unit variance scaling is sensitive to noise because the weight on the temperature measurement, which has small variance (e.g. T_1), is large. The performances of the inferential models based on the scaled variables are almost the same except for the case of the logarithmic transformation on the composition only. For k=3, EPV=98.048 and PRESS=0.575×10^{-4} without variable transformation and scaling. When the transformation $\ln(y_D/(1-y_D))$ on y_D is applied, the relationship between $\ln(y_D/(1-y_D))$ and T_i is linear around the middle of the column. But the performance of the transformation is not good because the linear relationship between $\ln(y_D/(1-y_D))$ and T_i around the feed tray becomes distorted when the feed disturbances such as z_f is introduced. The logarithmic transformation on the composition is not desirable when the change of z_f is frequent. The performance of the logarithmic transformation is not good (EPV=92.698, PRESS=0.531×10^{-3} with k=3) because the change in z_f is very large for our column. The performance of the estimator is improved by using the logarithmic transformation both on the composition and temperature by $\ln(y_D/(1-y_D))$ and $\ln((T_i-T_L^b)/(T_H^b-T_i))$ (Mejdell and Skogestad, 1991). The EPV with no noise is close to 100 % after only 3 factors due to the linearizing effect of the transformation (EPV=99.943, PRESS=0.958×10^{-6} with k=3). But the transformation is somewhat sensitive to noise because $(T_i-T_L^b)/(T_H^b-T_i)$ term becomes zero or infinite at the end of the column. Estimation without variable transformation is best from the view points of prediction accuracy and robustness for our binary column.

Use of Measured Inputs

The relationship between the auxiliary measured inputs(e.g. R and Q_B) and the tray temperatures cannot be described in linear forms. Therefore the estimation using the auxiliary measured inputs is not desirable. Actually the estimation performance decreases when we use R and Q_B in addition to all tray temperatures. The results using all tray temperatures, R, and Q_B with three factors were EPV=99.119 and PRESS=0.263×10^{-4} with unit variance scaling while EPV=99.805 and PRESS=0.588×10^{-5} were obtained when only the temperatures were used. Similar results were also reported by previous researchers (Mejdell and Skogestad, 1993; Piovoso and Kosanovich, 1994).

The projection estimator (Joseph and Brosilow, 1978) has been reported to be sensitive to noise and modeling error (Mejdell and Skogestad, 1993). The R and Q_B are used to infer the composition in the projection estimator. It is clear that the performance of the estimator using the measured inputs is not good due to the modeling errors. Additionally, the projection estimator cannot handle the collinear measurements properly and provide the advantage of the multiple temperature measurements (e.g. robust to the process noise).

Conclusions

The guidelines on the design of composition estimator via PLS have been presented: The recommended number of factors is equal to the number of independent variables(e.g. z_f, y_D, x_B, and column pressure P) which have influence on the temperature profile. It has been shown that the loading vector space of the estimator exactly reflects the direction of temperatures due to corresponding input variables. The relationship between the auxiliary measured inputs (R and Q_B) and the product composition cannot be described in the linear form. Thus, the estimation using the measured inputs is not desirable. The performance of the logarithmic transformation on the composition is not good if the change of z_f is frequent. The transformation $\ln(y_D/(1-y_D))$ and $\ln((T_i-T_L^b)/(T_H^b-T_i))$ is effective but somewhat sensitive to measurement noise. When all tray temperature measurements are used, unit variance scaling enhances the estimation performance by making the collinear measurements with T_1 have the same information as T_1, but makes the estimators sensitive to noise.

Acknowledgment

Financial support from the Korea Science and Engineering Foundation through the Automation Research Center at Postech is gratefully acknowledged. This paper was also supported in part by Korea Research Foundation.

References

Geladi, P. and B.R. Kowalski. (1986). Partial least-squares regression: A tutorial, *Analytica Chimica Acta*, **185**(3), 1-17.

Joseph, B. and C.B. Brosilow. (1978). Inferential control of processes : Part I. Steady state analysis and design, *AIChE*. **24**(3), 485-492.

Kresta, J.V., T.E. Marlin and J.F. MacGregor. (1994). Development of inferential process models using PLS, *Computers Chem. Eng*., **18**(7), 597-611.

Mejdell, T. and S. Skogestad. (1991). Estimation of distillation compositions from multiple temperature measurements using partial-least-squares regression, *Ind. Eng. Chem. Res*., **30**, 2543-2555.

Mejdell, T. and S. Skogestad. (1993). Output estimation using multiple secondary measurements: high-purity distillation, *AIChE J*., October, **39**(10), 1641-1653.

Piovoso, M.J. and K.A. Kosanovich. (1994). Application of multivariate statistical methods to process monitoring and controller design, *Int. J. Control*, **59**(3), 743-765.