Loss Method: A Static Estimator Applied for Product Composition Estimation From Distillation Column Temperature Profile

M. Ghadrdan\textsuperscript{1}  C. Grimholt\textsuperscript{1}  S. Skogestad\textsuperscript{1}  I.J. Halvorsen\textsuperscript{2}

\textsuperscript{1}Norwegian University of Science & Technology, Department of Chemical Engineering, 7491, Trondheim, Norway

\textsuperscript{2}SINTEF ICT, Applied Cybernetics, 7465 Trondheim, Norway

AIChE Meeting, 2011
Motivation

- Some process variables cannot be measured frequently
  - Example: Composition measurement using online analyzers (like Gas Chromatograph)
    - Large measurement delays
    - High investment/maintenance costs
    - Low reliability
Motivation

- Some process variables cannot be measured frequently
  - Example: Composition measurement using online analyzers (like Gas Chromatograph)
    - Large measurement delays
    - High investment/maintenance costs
    - Low reliability

- Sensors:
  - Temperature
  - Pressure
  - Flow rate
  - etc.
Motivation

- Some process variables can not be measured frequently
  - Example: Composition measurement using online analyzers (like Gas Chromatograph)
    - Large measurement delays
    - High investment/maintenance costs
    - Low reliability

- Sensors:
  - Temperature
  - Pressure
  - Flow rate
  - etc.

An estimator attempts to approximate the unknown parameters using the measurements
Outline

1 Introduction
   - Estimation
   - Partial Least Squares

2 Loss Method
   - Optimal estimators for different scenarios
   - Necessary data for the task of estimation

3 Examples
Estimators

Different categories: Static / Dynamic, Data-based / Model-based, Open-loop / Close-loop

- Static Estimators
- Dynamic Estimators

---

Estimators

Different categories: Static / Dynamic, Data-based / Model-based, Open-loop / Close-loop

- **Static Estimators**
  - Model-based
    - Example: Brasilow estimator\(^1\)
    - Our method is in this category
  - Data-based
    - Example: Partial Least Square (PLS)

- **Dynamic Estimators**

---
\(^1\) R. Weber, C. Brosilow, The Use of Secondary Measurements to Improve Control, AIChE J., 18, 3, p. 614-623
Estimators

Different categories: Static / Dynamic, Data-based / Model-based, Open-loop / Close-loop

- **Static Estimators**
  - Model-based
    - Example: Brasilow estimator\(^1\)
    - Our method is in this category
  - Data-based
    - Example: Partial Least Square (PLS)

- **Dynamic Estimators**
  - Model-based
    - Example: Kalman filter
  - Data-based
    - Time variant reliability analysis of existing structures using data

\(^1\) R. Weber, C. Brosilow, The Use of Secondary Measurements to Improve Control, AIChE J., 18, 3, p. 614-623
Partial Least Squares

- PC regression = weights are calculated from the covariance matrix of the predictors
- PLS = weights reflect the covariance structure between predictors and response – mostly requires a complicated iterative algorithm

Nipals and SIMPLS algorithms probably most common
- The goal is to maximize the correlation between the response(s) and component scores
- PLS can extends to multiple outcomes and allows for dimension reduction
- No collinearity – Independence of observations not required
PLS

\[ \hat{Y} = BX \]

- PLS: is not optimal for any particular problem
- Loss method: optimal for certain well-defined problems
OBJECTIVE

The main objective is to find a linear combination of measurements such that keeping these constant indirectly leads to nearly accurate estimation with a small loss $L$ in spite of unknown disturbances, $d$, and measurement noise, $\eta$.

$$\min_H \| e \|_2 = \| y - \hat{y} \|_2$$
"Open-loop" (for the purpose of Monitoring):

1. No control \((u\) is a free variable)
2. Primary variables \(y\) are controlled \((u\) is used to keep \(y = y_s\)).
3. Secondary variables \(z\) are controlled \((u\) is used to keep \(z = z_s\)).

"Close-loop" (for the purpose of Control)
Assumption: Linear models for the primary variables $y$, measurements $x$, and secondary variables $z$

\[ y = G_y u + G_y^d d \quad x = G_x u + G_x^d d \quad z = G_z u + G_z^d d \]

\[ G_y = \left( \frac{\partial y}{\partial u} \right)_d, \quad G_y^d = \left( \frac{\partial y}{\partial d} \right)_u \]
\[ G_x = \left( \frac{\partial x}{\partial u} \right)_d, \quad G_x^d = \left( \frac{\partial x}{\partial d} \right)_u \]
\[ G_z = \left( \frac{\partial z}{\partial u} \right)_d, \quad G_z^d = \left( \frac{\partial z}{\partial d} \right)_u \]

The actual measurements $x_m$, containing measurement noise $n^x$ is

\[ x_m = x + n^x \]

The linear estimator is of the form

\[ \hat{y} = H x_m \]
Loss Method

- Linear model from inputs to primary variables $G_y$
- Linear model from disturbances to primary variables $G_{d_y}$
- Linear model from inputs to measurements $G_x$
- Linear model from disturbances to measurements $G_{d_x}$
- Linear model from inputs to primary variables $G_y$

$y = G_y u + G_{d_y} d + G_x n_x + H x_{sw} + \hat{y}$

Prediction error
Loss Method

Linear model from inputs to primary variables
\( G_y \)

Linear model from disturbances to primary variables
\( G_{dy} \)

Linear model from inputs to measurements
\( G_x \)

Linear model from disturbances to measurements
\( G_{dx} \)

Prediction error

\( y \)

\( u \)

\( + \)

\( - \)

\( d \)

\( n_x \)

\( x_m \)

\( \hat{y} \)
**Optimal estimators for different scenarios (Loss Method)**

"Open-loop" 1

\[ H_1 = Y_1 X_1^\dagger \]

\[ Y_1 = \begin{bmatrix} G_y W_u & G_d^d W_d & 0 \end{bmatrix} \]

\[ X_1 = \begin{bmatrix} G_x W_u & G_d^d W_d & W_{n^x} \end{bmatrix} \]

"Open-loop" 2

\[ H_2 = Y_2 X_2^\dagger \]

\[ Y_2 = \begin{bmatrix} W_y & 0 & 0 \end{bmatrix} \]

\[ X_2 = \begin{bmatrix} G_x^c W_{y^s} & FW_d & W_{n^x} \end{bmatrix} \]

"Open-loop" 3

\[ H_3 = Y_3 X_3^\dagger \]

\[ Y_3 = \begin{bmatrix} G_y^c W_{z^s} & F_y^c W_d & 0 \end{bmatrix} \]

\[ X_3 = \begin{bmatrix} G_x^c W_{z^s} & F_x^c W_d & W_{n^x} \end{bmatrix} \]

"Closed-loop"

\[ \min_H \| H \begin{bmatrix} FW_d & W_{n^x} \end{bmatrix} \|_F \]

s.t. \( H G_x = G_y \)

* All subject to the constraint of independent variables values
Optimal "open-loop" estimator, when $y=ys$ (Loss Method)

\[
H_2 = Y_2 X_2^+ \\
Y_2 = \begin{bmatrix} W_y s & 0 & 0 \end{bmatrix} \\
X_2 = \begin{bmatrix} G_y^c W_y s & FW_d & W_n^x \end{bmatrix}
\]

Initial Equations

\[
y = G_y u + G_y^d d \\
x = G_x u + G_x^d d \\
x_m = x + n^x \\
\hat{y} = H x_m
\]

\[
u = G_y^{-1} y_s - G_y^{-1} G_y^d d \\
\hat{y} = H \left[ G_x G_y^{-1} y_s + \left( G_x^d - G_x G_y^{-1} G_y^d \right) d + n^x \right] \\
e = \begin{bmatrix} (I - HG_x^c) W_y s & -HFW_d & -HW_n^x \end{bmatrix} \begin{bmatrix} y_s' \\
d' \\
n^x' \end{bmatrix}
\]

\[
\|e(H)\|_2 = \frac{1}{2} \|M_{ol}(H)\|_F^2
\]
Optimal "open-loop" estimator, when $y = y_s$ (Loss Method)

$$H_2 = Y_2 X_2^+$$

$$Y_2 = \begin{bmatrix} W_{y_s} & 0 & 0 \end{bmatrix}$$

$$X_2 = \begin{bmatrix} G_x^l W_{y_s} & FW_d & W_{n^x} \end{bmatrix}$$

Initial Equations

$$y = G_y u + G_y^d d$$

$$x = G_x u + G_x^d d$$

$$x_m = x + n^x$$

$$\hat{y} = H x_m$$

$$u = G_y^{-1} y_s - G_y^{-1} G_y^d d$$

$$\hat{y} = H \left[ G_x G_y^{-1} y_s + \left( G_x^d - G_x G_y^{-1} G_y^d \right) d + n^x \right]$$

$$e = \left[ (I - H G_x^c) W_{y_s} - H F W_d - H W_{n^x} \right] \begin{bmatrix} y'_s \\ d' \\ n^{x'} \end{bmatrix}$$

$$\|e(H)\|_2 = \frac{1}{2} \|M_{ol}(H)\|_F^2$$
Optimal "open-loop" estimator, when \( y = y_s \) (Loss Method)

\[
\tilde{d} = \begin{bmatrix} u' \\ d' \\ n^x' \end{bmatrix} \sim \mathcal{N} \left( 0, \mathbf{I}_{n_u+n_d+n_x} \right)
\]

\[
\| \mathbf{e}(H) \|_{2,\text{exp}} = \frac{1}{2} \mathbb{E} \left[ \text{tr} \left( \mathbf{M} \tilde{d} \tilde{d}^T \mathbf{M}^T \right) \right] = \frac{1}{2} \text{tr} \left( \mathbf{M}^T \mathbf{M} \mathbb{E} \left[ \tilde{d} \tilde{d}^T \right] \right)
\]

\[
\mathbb{E} \left[ \tilde{d} \tilde{d}^T \right] = \text{Cov} \left( \tilde{d}, \tilde{d} \right) + \mu \mu^T
\]

\[
\min_{\mathbf{H}} \| \begin{bmatrix} \mathbf{W}_{y_s} & 0 & 0 \end{bmatrix} - \mathbf{H} \begin{bmatrix} \mathbf{G}_{x}^{cl} \mathbf{W}_{y_s} & \mathbf{F} \mathbf{W}_d & \mathbf{W}_{n^x} \end{bmatrix} \| \text{=} \min_{\mathbf{H}} \| \mathbf{Y}_2 - \mathbf{H} \mathbf{X}_2 \|
\]
Optimal "close-loop" estimator (Loss Method)

\[
\begin{align*}
\text{Initial Equations} \\
y &= G_y u + G_y^d d \\
x &= G_x u + G_x^d d \\
x_m &= x + n^x \\
\hat{y} &= Hx_m \\
u &= -(HG_x)^{-1} H \left( G_x^d d + n^x \right) + (HG_x)^{-1} y_s \\
y &= -G_y (HG_x)^{-1} H \left( G_x^d d + G_x G_y^{-1} G_y^d d \right) n^x + G_y (HG_x)^{-1} y_s \\
e &= y - \hat{y} = y - y_s = -G_y (HG_x)^{-1} H (F_d + n^x) + \left[ G_y (HG_x)^{-1} - I \right] y_s \\
\min_H \| H \begin{bmatrix} F & W_n^x \end{bmatrix} \|_F \\
\text{s.t. } HG_x = G_y
\end{align*}
\]
Optimal "close-loop" estimator (Loss Method)

Initial Equations

\[
\begin{align*}
  y &= G_y u + G_y^d d \\
  x &= G_x u + G_x^d d \\
  x_m &= x + n^x \\
  \hat{y} &= H x_m \\

  u &= - (HG_x)^{-1} H \left( G_x^d d + n^x \right) + (HG_x)^{-1} y_s \\
  y &= -G_y (HG_x)^{-1} H \left( \underbrace{G_x^d - G_x G_y^{-1} G_y^d}_{F} \right) d + n^x + G_y (HG_x)^{-1} y_s \\
  e &= y - \hat{y} = y - y_s = -G_y (HG_x)^{-1} H (F d + n^x) + \left[ G_y (HG_x)^{-1} - I \right] y_s
\end{align*}
\]

\[
\begin{align*}
  \min_{H} \| H \left[ \begin{array}{cc}
    F W_d & W n^x
  \end{array} \right] \|_F \\
  \text{s.t. } HG_x = G_y
\end{align*}
\]
Optimal "close-loop" estimator (contd.)

The prediction error $e$

$$e = y - \hat{y} = y - y_s = -G_y (H G_x)^{-1} H (F d + n^x) + \left[ G_y (H G_x)^{-1} - I \right] y_s$$

Introducing the normalized variables:

$$e = -G_y (H G_x)^{-1} H \left[ \begin{array}{cc} F W_d & W_n^x \end{array} \right] \left[ \begin{array}{c} d' \\ n^{x'} \end{array} \right] + \left[ G_y (H G_x)^{-1} - I \right] y_s$$

Degree of Freedom

$$e_1 (H) = e_1 (DH)$$
If \( \tilde{F} = \begin{bmatrix} FW_d & W_n x \end{bmatrix} \) is full rank, which is always the case if we include independent measurement noise, then \(^2\)

\[
H = D \left( \left( X_{opt} X_{opt}^T \right)^{-1} G_x \right)^T
\]

where

\[
D = G_y \left( G_x^T \left( X_{opt} X_{opt}^T \right)^{-1} G_x \right)^{-1}
\]

\(^2\) Alstad et al. (2009), Optimal measurement combinations as controlled variables, J. Proc. Control, 19 (1), 138-148
Necessary data for the task of estimation (Model-based)

\[ \begin{bmatrix} Y \\ X \end{bmatrix}_{all} = \begin{bmatrix} G_Y & 0 \\ G_X & X_{opt} \end{bmatrix} \]

where \( X_{opt} = \begin{bmatrix} F \ W_d \\ W_{nx} \end{bmatrix} \)

\[ \begin{bmatrix} Y = Y_{non-opt} & 0 \\ X = X_{non-opt} & X_{opt} \end{bmatrix} \]
Necessary data for the task of estimation (Data-based)

Theorem

Closed Loop Regressor (CLR) \(^a\). The data matrices can be transformed to the “optimal – non-optimal” structure by

1. Performing a singular value decomposition on the data matrix \(Y\)
2. Multiplying the data matrices \(X\) and \(Y\) with the unitary matrix \(V\)

\[
\begin{align*}
YV &= \begin{bmatrix} Y_{\text{non-opt}} & 0 \end{bmatrix} \\
XV &= \begin{bmatrix} X_{\text{non-opt}} & X_{\text{opt}} \end{bmatrix}
\end{align*}
\]

\(^a\)Skogestad et al (2011). Selected Topics on Constrained and Nonlinear Control Workbook
Necessary data for the task of estimation (Data-based)

Theorem

Closed Loop Regressor (CLR)\(^a\). The data matrices can be transformed to the “optimal – non-optimal” structure by

1. Performing a singular value decomposition on the data matrix \(Y\)
2. Multiplying the data matrices \(X\) and \(Y\) with the unitary matrix \(V\)

\[
\begin{align*}
YV &= \begin{bmatrix} Y_{\text{non-opt}} & 0 \end{bmatrix} \\
XV &= \begin{bmatrix} X_{\text{non-opt}} & X_{\text{opt}} \end{bmatrix}
\end{align*}
\]

\(^a\)Skogestad et al (2011). Selected Topics on Constrained and Nonlinear Control Workbook

Proof.

Since \(V\) is unitary, so \(\|YV - HXV\|_F = \|Y - HX\|_F\)

Writing the unitary matrix \(U\) in block form as \(U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}\), we will have

\[
\begin{align*}
YV &= US = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \\
      &= \begin{bmatrix} U_1 \Sigma_1 & 0 \end{bmatrix}
\end{align*}
\]
Example 1: Results

- Binary Distillation (Col. A), 41 trays, 8 measurements
- Secondary variables: Reflux, temperature in 25th tray

The mean prediction error of the model-based estimators applied to four operation scenarios

<table>
<thead>
<tr>
<th>Calibration Data</th>
<th>Validation Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
</tr>
<tr>
<td>S1</td>
<td>0.0085</td>
</tr>
<tr>
<td>S2</td>
<td>0.0591</td>
</tr>
<tr>
<td>S3</td>
<td>0.0599</td>
</tr>
<tr>
<td>S4</td>
<td>0.0099</td>
</tr>
</tbody>
</table>
Example 1: Results

Median prediction error for 150 data set with 200 samples.
Example 2: Multi-component distillation

\[ u = y = \begin{bmatrix} x_{C_3} \text{ in } D \\ x_{C_2} \text{ in } B \end{bmatrix} \]

\[ G_y = I \]

\[ G_x^d = F \]

\[ G_y^d = 0 \]
Example 2: Results

\[
H = \begin{bmatrix}
0.0004 & 0.0014 \\
0.0081 & -0.0045 \\
-0.005 & 0.0074 \\
-0.0047 & 0.0006 \\
0.0062 & -0.0104 \\
-0.003 & 0.0126 \\
-0.0013 & 0.0051 \\
0.0024 & -0.0162 \\
-0.0028 & 0.0042
\end{bmatrix}
\]

\[
B = \begin{bmatrix}
0.0002 & 0.0013 \\
0.0087 & -0.0041 \\
-0.006 & 0.0068 \\
-0.0051 & 0.0003 \\
0.0077 & -0.0096 \\
-0.0034 & 0.0124 \\
-0.0016 & 0.0049 \\
0.0026 & -0.016 \\
-0.0031 & 0.004
\end{bmatrix}
\]
Example 2: Results

(a) +5% disturbance in feed flow

(b) -1% disturbance in Feed composition $z_{1F}$
Example 3: Kaibel distillation column

The model used for this study is simulated in UNISIM. The feed stream is an equimolal mixture of Methanol, Ethanol, 1-Propanol, 1-butanol and saturated liquid. All the optimal operating points for different sets of the disturbances are found by applying an optimisation solver in MATLAB with the full non-linear model in UNISIM.

The right figure in Figure 2 shows the composition profiles in different sections of the dividing-wall column. As it is obvious, the most difficult separation is taking place in the prefractionator and the other sections are performing close to binary separation with small light or heavy impurity. Because of this, the focus of our study is on the prefractionator part.

Partial Least Square (PLS) Method

In chemometrics, partial least squares (PLS) regression has become an established tool for modelling linear relations between multivariate measurements (Martens and Næs 1989). This biased regression method is used to compress the predictor data matrix

\[
x = \begin{bmatrix} x_1, & \cdots, & x_p \end{bmatrix}
\]

that contains the values of \( p \) predictors for \( n \) samples, into a set of \( A \) latent variable or factor scores \( t = \begin{bmatrix} t_1, & \cdots, & t_A \end{bmatrix} \), where \( A \leq p \).

The factors \( t = \begin{bmatrix} t_1, & \cdots, & t_A \end{bmatrix} \) are determined sequentially using the nonlinear iterative partial least squares (NIPALS) algorithm [2]. The orthogonal factor scores are used to fit a set of \( n \) observations to \( m \) dependent variables \( y = \begin{bmatrix} y_1, & \cdots, & y_m \end{bmatrix} \). The main attraction of the method is that it finds a parsimonious model even when the predictors are highly collinear or linearly dependent. So, the final fitting

DoF

\[
u = \begin{bmatrix} R_L & R_V & L & V & S_1 & S_2 \end{bmatrix}
\]

Extra Degrees of Freedom:

- Vapor Split \( (R_V) \)
- Liquid Split \( (R_L) \)

Disturbances:

- Feed flowrate, composition and quality
- Column Pressure
- Setpoints for splits
Example 3: Results

Possible Improvement for Loss method: Structured $H^3$

---

Yelchuru et al., MIQP formulation for Controlled Variable Selection in Self Optimizing Control
Conclusion

- Loss method is more systematic method to design soft-sensor compared to PLS
- For the example we showed, PLS and Loss method show almost the same result although two different approaches are used
Comment on PLS

- Shrinkage properties\(^4\)

\[
MSE = E(b - \beta)' S (b - \beta) = \sum_i \lambda_i (Ea_i - \alpha_i)^2 + \sum_i \lambda_i \text{Var}(a_i)
\]

\[
a_i = f(\lambda_i) a_i^0
\]

\(f(\lambda_i) = 0\ or\ 1\) for OLS, PCR, Ridge

Butler et al.: PLS is not a shrinkage method. PLSR nearly always can be improved