NTNU

Norwegian University of
Science and Technology

A New Static Estimator Based on Self-Optimizing Theory

Chriss Grimholt

Chemical Engineering and Biotechnology
Submission date: June 2011
Supervisor: Sigurd Skogestad, IKP
Co-supervisor: Maryam Ghardan, IKP

Norwegian University of Science and Technology
Department of Chemical Engineering

Abstract

This thesis compares the performance of the new static model based estimator proposed by Skogestad et al. (2011) with least squares (LS), principal component regression (PCR), and partial least squares (PLS) estimators on a linear, binary, and multicomponent distillation model. The performance is classified into two categories: "open-loop" performance (estimator used for monitoring) and "closed-loop" performance (estimator used for control). The new estimator is derived from a regression point of view, and it is shown that this estimator is optimal for "closed-loop" estimation. Skogestad et al. (2011) also presented a method called loss regression for applying the new estimator on data. This thesis shows that this estimator is sensitive to noise and collinearity, and a new improved method called the truncated "closed-loop" method (truncated CLM) is proposed. It is found that the new estimator and the truncated CLM have better "closed-loop" performance, but worse "open-loop" performance than LS, PCR and PLS.

Preface

This thesis was written as a final part of my masters degree from the Norwegian University of Science and Technology (NTNU) in 2011.

I would like to thank my supervisor professor Sigurd Skogestad for his support, guidance, and encouragement throughout the project.

I would also like to thank my co-supervisor Maryam Ghadrdan for the interest she has shown in my work, and all the help she has given me.

Lastly, I would like to thank my family for their never-ending love, support, and encouragement, and my friends for making five years of study pass in mere moments.

Declaration of Compliance

I hereby declare that this is an independent work in compliance with the exam regulations of the Norwegian University of Science and Technology.

Date and signature: _____

Contents

Chapter 1

Introduction

The focus of this work was to compare the new static estimator proposed by Skogestad et al. (2011) with conventional static estimators like least squares (LS), principal component regression (PCR), and partial least squares (PLS) on standard challenge problems and as a composition estimator for distillation. The distillation problem gave much insight into the estimator properties, making the standard challenge problems redundant. It was decided to only show the results for the distillation problem in the report, and to summarize the standard challenge problems in the appendix.

1.1 Estimators

In a chemical plant, there are usually a large number of hardware sensors which are used for monitoring and controlling given processes. In some processes there are process variables (e.g., composition) that are too difficult or expensive to measure. Estimators, also called soft sensors, work by predicting desired variables using existing measurements (usually variables that are easy to measure like temperature). Thus, estimators can be a good alternative to hardware sensors for these difficult process variables.

Estimators can usually be divided into four main groups based on whether they are a static or dynamic estimator, or a linear or nonlinear estimator. Because the new estimator is a static linear estimator, only this type of estimators are treated in this thesis.

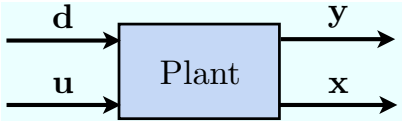


Figure 1.1: Generalization of a chemical plant.

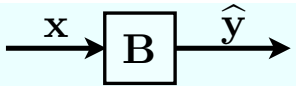


Figure 1.2: Block diagram of an estimator $\mathbf{B}$ which uses measurements $\mathbf{x}$ to predict the primary variable $\mathbf{y}$.

Figure 1.1, show a typical generalization of a chemical plant. There are two input variables $\mathbf{u}$ and $\mathbf{d}$. $\mathbf{u}$ are degrees of freedom that can be used for controlling the plant, and $\mathbf{d}$ are disturbances that affects the plant operation. The system has two outputs, the primary variable $\mathbf{y}$ and the measurements $\mathbf{x}$.

The goal of the estimator is to precisely estimate the primary variable $\mathbf{y}$, called the prediction $\hat{\mathbf{y}}$, from the measurements $\mathbf{x}$. Figure 1.2 shows a block diagram of an estimator $\mathbf{B}$.

The estimator is usually determined in two ways: (1) from data with regression techniques like least squares (LS), principal component regression (PCR), or partial least squares (PLS), or (2) from detailed models most commonly derived from first principles.

Model based estimators are primarily used for planning and designing of processing plants and are based on ideal conditions. This makes them difficult to adapt for real world processes as they are sensitive to noise and modeling errors. On the other hand, data based estimators rely on recorded data from the actual process, and are thus able to describe the real process conditions (Kadlec et al., 2008). In this thesis, data will be generated from the process models, and added random noise to mimic real data. The classical methods for finding data based estimators are discussed further in the subsequent chapter.

Estimators can be used in two different ways: (1) for monitoring or (2) "controlling" primary variables (actually it is the predicted primary variables which are controlled). Estimators used for monitoring primary variables are termed "open-loop" estimators (the predictions are not controlled), and estimators used for controlling primary variables are termed "closed-loop" estimators (predictions are controlled).

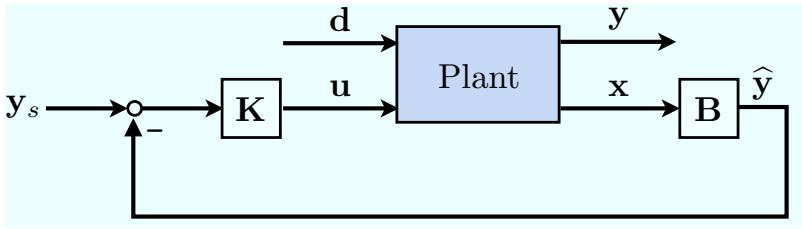


Figure 1.3: Block diagram of a "closed-loop" estimator

To clarify, all references to "open-loop" and "closed-loop" (in quotation marks) refer to whether or not the predictions are controlled. On the other hand, the terms open-loop and closed-loop (not in quotation marks) refers only to whether the primary variables (sometimes also secondary variables) are controlled.

A figure of a "closed-loop" estimator is shown in Figure 1.3. The controller $\mathbf{K}$ adjusts $\mathbf{u}$ such that the difference between the setpoints $\mathbf{y}_s$ and the predictions $\hat{\mathbf{y}}$ are zero. In contrast, For a system with primary variables $\mathbf{y}$ in closed-loop, it is the difference between the primary variables $\mathbf{y}$ and its setpoints $\mathbf{y}_s$ which goes toward zero. In the chapter about the model based estimators, we will derive optimal estimators for both "open-loop" and "closed-loop" operations.

1.2 The Self-Optimizing Background of the New Estimator

When designing a plant-wide control structure, one usually finds that there are degrees of freedom (valves) that are unconstrained degrees of freedom, which need to be specified during operation. The idea behind self-optimizing control is to use these extra degrees of freedom to control process variables (measurements) that, for when kept constant, keeps the plant close-to its optimal operation when the plant is subjected to disturbances. Optimal operation can refer to either economical or environmental considerations. That is, maximizing profit or minimizing production of harmful waste products. There are two common ways of finding the self-optimizing variables based on individual measurements: the direct loss evaluation (Skogestad, 2000) and the maximum gain rule (Halvorsen et al., 2003). Direct loss evaluation is a brute force approach and is computationally demanding. The maximum gain rule is

less accurate, but requires far less computations. A good introduction to the topic of self-optimizing control is written by Skogestad (2004).

An alternative to controlling individual process variables, is to control measurement combinations. The goal then becomes to find the optimal combination of measurements which gives the best self-optimizing properties. Halvorsen et al. (2003) found the optimal measurement combination by a quadratic optimization problem called the exact local method, and an analytical solution to the optimization problem was presented by Alstad et al. (2009). In addition, one has the very simple nullspace method (Alstad and Skogestad, 2007) but it cannot handle measurement noise. The new static estimator presented by Skogestad et al. (2011) is a continuation of the work done by Hori et al. (2005) and Alstad et al. (2009), and uses the exact local method to find an optimal combination of measurements by using minimum prediction error as the optimization objective.

1.3 Motivation

In distillation, control of product composition is of great importance, but is usually complicated by problems associated with on-line measurements. Most composition analyzers, like gas chromatographs, have high investment and maintenance cost. Also, these type of analyzers have large time delays associated with the measurements, which gives severe limitations on the control performance.

Temperature measurements are inexpensive, fast, and reliable. The control of a single tray temperature is a common method for indirectly controlling product composition in the industry. This is possible because of the close physical relationship between temperatures and composition in distillation. However, temperature measurements are not a precise indication of product composition. Mejdell and Skogestad (1991) mentioned several sources of inaccuracies of single temperature control, where improper tray mixing, pressure variations and random noise are some.

As an alternative, multiple temperature measurements can be used to predict the product composition. Weber and Brosilow (1972) proposed a static (Brosilow) estimator using linear model of the primary variables and measurements, and reported that composition control using this estimator was far

superior to the composition control achieved by fixing a single stage temperature.

Using measurement data, a static estimator can be easily found from conventional multivariate calibration methods such as LS, PCR and PLS. Mejdell and Skogestad (1991) reported that PCR and PLS had good performance for both binary and multicomponent distillation. Mejdell and Skogestad (1993) reported that the proposed Brosilow estimator had poor performance for their linear distillation example, and found that this was mainly caused by the estimators use of information about the degrees of freedom $\mathbf{u}$. They found also that the static PCR estimator had very good "closed-loop" performance, rivaling that of the dynamic Kalman filter.

The new static estimator based on self-optimizing theory uses linear models (like the Brosilow estimator), but uses only information about measurements $\mathbf{x}$, and not the degrees of freedom $\mathbf{u}$. Because of the self-optimizing nature of the estimator, it is thought that it will have good prediction ability. Skogestad et al. (2011) also presented a data based extension to the new estimator called loss regression, making it possible to extract the new estimator also from measurement data.

1.4 Project Scope

In this thesis, the new estimator based on self-optimizing theory will be derived from a regression point of view. We will show, for linear systems, that this estimator is optimal for "closed-loop" estimation (systems where the prediction is controlled), and the estimator will in this thesis be termed the optimal "closed-loop" estimator.

The loss regression will in this thesis be referred to as the "closed-loop" method (CLM). We found that this method was sensitive to collinearity and noise, and a new improved method called truncated CLM is presented.

In addition, three optimal "open-loop" model based estimators (estimation only used for monitoring purposes) were developed for comparison and insight.

Finally, the "open-loop" and "closed-loop" performance of the optimal "closed-loop" estimator, the CLM estimator, and the truncated CLM estimator will be compared with LS, PCR, and PLS estimators on three distillation examples

(linear, nonlinear binary, and nonlinear multicomponent). It will be shown that the performance of the optimal "closed-loop" estimator and the truncated CLM estimator has better "closed-loop" performance, but have worse "open-loop" performance than other estimators.

Chapter 2

Data Based Estimators

2.1 Introduction To Data Based Estimators

One way of finding a static estimator $\mathbf{B}$ is with multivariate calibration methods like least squares (LS), principal component regression (PCR) and partial least squares (PLS). Given a system with one dependent variable y and one independent variable x, where x have a direct relationship with y given by

$$y = f(x)$$

The object of multivariate calibration is to obtain the best relationship between x and y. This relationship is called a model, or by the term *estimator* which will be used in this thesis. LS, PCR, and PLS are linear methods for determining the estimator, and finds the best linear relationship between x and y,

$$y = bx + b_0 + e$$

where the estimator is given by b and b_0, and e is the residual (also called the model error). To simplify calculations, all data in this thesis are given in deviation variables, and the bias term b_0 will be zero. The relationship can then be expressed,

$$\Delta y = b \cdot \Delta x + e$$

Because all variables can be assumed to be in deviation form, the deviation notation will be dropped, giving

$$y = bx + e$$

In this example there is only one dependent variable and one independent variable, but an estimator can as easily be found from multiple dependent and independent variables, as will be shown in the next section.

2.1.1 Data Structure

Previously we mentioned a system with one independent variable y and one dependent variable x, given by

$$y = bx + e$$

If the relationship will be determined from observations, we would need to obtain several measurements of x and y called *samples*, to make the estimator as accurate as possible. Given that we have m samples and one independent $x_{1,i}$ and dependent $y_{1,i}$ variable where $i = 1, \ldots, m$, the relationship can be expressed

$$\begin{bmatrix} y_{1,1} & \cdots & y_{1,m} \end{bmatrix} = b \begin{bmatrix} x_{1,1} & \cdots & x_{1,m} \end{bmatrix} + \begin{bmatrix} e_{1,1} & \cdots & e_{1,m} \end{bmatrix}$$

This relationship can be extended further to include multiple independent variables. Given that we have n independent variables $x_{j,i}$ $(j = 1, \ldots, n)$ the independent variables can be written as a vector $\mathbf{x}_i$, and the estimator b_j written as $\mathbf{b}^T$. All vectors in this thesis are column vectors, and consequently row vectors will be designated as transposed vectors. The expression becomes,

$$\begin{bmatrix} y_{1,1} & \cdots & y_{1,m} \end{bmatrix} = \mathbf{b}^T \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_m \end{bmatrix} + \begin{bmatrix} e_{1,1} & \cdots & e_{1,m} \end{bmatrix}$$

Analog to this, the relationship can easily be extended to include multiple dependent variables. Given that we have p dependent variables y_h $(h = 1, \ldots, p)$, the dependent variable can be expressed as the vector $\mathbf{y}_i$ and will form the columns of the matrix $\mathbf{Y}$. The corresponding parameters $\mathbf{b}_h^T$ will form the rows of the matrix $\mathbf{B}$, and the residual $e_{h,i}$ can be written as the vector $\mathbf{e}_i$ and will form the columns of the matrix $\mathbf{E}$, giving

$$\overbrace{\begin{bmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_m \end{bmatrix}}^{\mathbf{Y}} = \overbrace{\begin{bmatrix} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_p^T \end{bmatrix}}^{\mathbf{B}} \overbrace{\begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_m \end{bmatrix}}^{\mathbf{X}} + \overbrace{\begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_m \end{bmatrix}}^{\mathbf{E}}$$

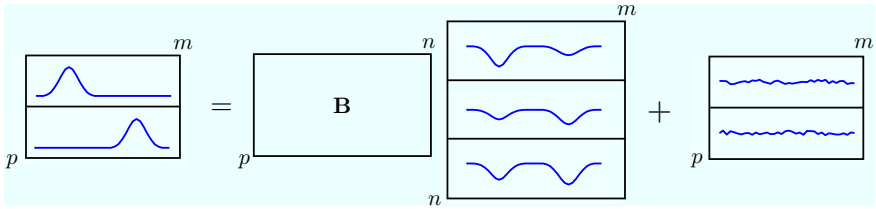


Figure 2.1: Graphical representation of the matrices

resulting in the final matrix from[1]

$$\mathbf{Y} = \mathbf{BX} + \mathbf{E}$$

A graphical representation of the relationship is shown in Figure 2.1. The dependent variables are the top and bottom product compositions, and the independent variables are three stage temperatures in a distillation column. The response is for a bell shaped change in top and bottom composition as a function of time. From this we can visually see how an estimator can be found from data. We can see that the first temperature measurement is most correlated with the top composition, and vice versa. Thus, we can assume that the first model parameter will be large for the top composition, and the third model parameter will be large for the bottom composition.

2.1.2 Finding the Estimator

Given that we have the independent $\mathbf{X}$ data with n independent variables and m samples, and we want to find the solution to the problem

$$\mathbf{E} = \mathbf{Y} - \mathbf{BX} = 0$$

there are three different cases which affects the solution of the problem:

$n = m$ there are as many samples as variables. Assumed that $\mathbf{X}$ is of full rank, the equation system is determined, and there is one unique solution

$$\mathbf{B} = \mathbf{YX}^{-1}$$

This situation is rarely encountered in real situations.

[1]Notice that this expression is the transpose of the of the expression conventionally used in multivariable calibration $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$, and implies that the estimator $\mathbf{B}$ found in this thesis is actually the transposed of the conventional estimator. This was done to avoid confusion in later chapters when comparing with results from control theory.

$n > m$ there are less samples than variables. The equation system is underdetermined and there are an infinite number of solution to the problem:

$$\mathbf{BX} = \mathbf{Y}$$

A null space solution $\mathbf{N}(p \times n)$ of $\mathbf{X}$

$$\mathbf{NX} = 0$$

can then be added to $\mathbf{B}$ and still be a solution to the problem.

Proof. $(\mathbf{N} + \mathbf{B})\mathbf{X} = \mathbf{NX} + \mathbf{BX} = \mathbf{BX}$ □

$n < m$ there are more samples than variables. The equation system is overdetermined and there are no exact solutions. However, we can find the solution that comes closest to solving the problem by minimizing the residual, giving the solution

$$\mathbf{BX} \approx \mathbf{Y}$$

where $\mathbf{B}$ is the optimal solution.

The most common case is the overdetermined case, and LS, PCR, and PLS are the common methods for finding the optimal solution. LS finds the estimator $\mathbf{B}_{\text{ls}}$ by finding the maximum correlation between $\mathbf{X}$ and $\mathbf{Y}$. PCR captures the maximum variance in $\mathbf{X}$, and uses this to find the estimator $\mathbf{B}_{\text{pcr}}$. PLS captures the maximum covariance between $\mathbf{X}$ and $\mathbf{Y}$ to find the estimator $\mathbf{B}_{\text{pls}}$.

2.1.3 Development of an Estimator

The development of an estimator is usually done in two steps. The first step is calibration or training, where an estimator $\mathbf{B}$ is found from a calibration set $\mathbf{X}_{\text{cal}}$ and $\mathbf{Y}_{\text{cal}}$. The second step is prediction testing or validation, where the estimator's ability to predict $\mathbf{Y}$ from a validation set $\mathbf{X}_{\text{val}}$ and $\mathbf{Y}_{\text{val}}$ (data not used in the calibration set) is tested and evaluated. This two step procedure requires that the data set of $\mathbf{X}$ and $\mathbf{Y}$ is initially divided into a calibration and validation set. The two step procedure is illustrated as follows. In the

calibration step, the estimator is found (assume that the calibration set is determined),

$$\mathbf{B} = \mathbf{Y}_{\text{cal}}\mathbf{X}_{\text{cal}}^{-1}$$

In the validation step, the prediction $\widehat{\mathbf{Y}}$ is calculated with the estimator from the validation set

$$\widehat{\mathbf{Y}} = \mathbf{B}\mathbf{X}_{\text{val}}$$

and the prediction error calculated

$$\mathbf{E} = \mathbf{Y}_{\text{val}} - \widehat{\mathbf{Y}}$$

When testing multiple estimators, estimators which gives low prediction error are kept, and estimators which gives high prediction error is discarded.

Cross-validation The above method of splitting the data set into a calibration and validation set, wastes one portion of the data on testing. For large data sets, this might not be a problem, but for small data sets this could greatly affect the estimators performance.

Cross-validation splits the data set into several segments I. Where $I-1$ segments are used for calibration and one is used for validation. These segments are rotated, and the calibration and validation process repeated such that each segment are used for validation only one time. When the process is finished, an average prediction error is calculated for the cross-validation. With this method, data is not wasted on validating the estimators.

2.2 Mathematical Tools

2.2.1 Singular Value Decomposition (SVD)

One very effective and useful mathematical tool is the singular value decomposition (SVD), and will in thesis be used to explain and derive LS, PCR, PLS, and other concepts. Any matrix can by SVD be decomposed into three matrices. The SVD of a matrix $\mathbf{X}$ $(n \times m)$ can be written

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$$

where the $\mathbf{U}$ $(n \times n)$ and the $\mathbf{V}$ $(m \times m)$ are unitary, and called the left and right singular vector matrix, respectively. The $\mathbf{\Sigma}$ $(n \times m)$ is a diagonal matrix where nonnegative entries, called singular values σ_i $(i = 1, ..., \min\{n, m\})$, are ordered by descending magnitude. The left singular vector matrix indicates the strongest and weakest output directions, and the right singular vector matrix indicates the strongest and weakest input directions. The singular values give the magnitudes of these directions. The ratio between the largest and smallest singular value is called the condition number $\mathrm{cond}(\mathbf{X}) = {}^{\sigma_1}/_{\sigma_{\min\{n,m\}}}$, and a matrix with a large condition number is called ill-conditioned. The rank of the matrix $\mathbf{X}$ $(\mathrm{rank}(\mathbf{X}))$ is equal to the number of nonzero singular values.

The left and right singular vector matrix has also the following useful properties:

$$\mathbf{U}^T = \mathbf{U}^{-1}$$
$$\mathbf{V}^T = \mathbf{V}^{-1}$$

2.2.2 Pseudoinverse

It is only possible to take the inverse of a matrix that is determined and of full rank. For other cases, the matrix becomes singular and the matrix has no inverse. The solution to the equation $\mathbf{Y} = \mathbf{BX}$ can then be found by the pseudoinverse.

Given that $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$ is the SVD of matrix $\mathbf{X}$, the pseudoinverse $\mathbf{X}^\dagger$ can then be expressed as

$$\mathbf{X}^\dagger = \mathbf{V}\Sigma^\dagger\mathbf{U}^T$$

where $\Sigma^\dagger$ is a diagonal matrix of the same dimensions as Σ, and its nonzero diagonal entries are the inverse of the nonzero singular values σ_k $(k = 1, ..., \mathrm{rank}\,(\mathbf{X}))$. That is, a matrix of rank r will have

$$\Sigma = \mathrm{diag}\,\{\sigma_1 > \sigma_2 > \ldots > \sigma_r > 0\}$$

which have the corresponding

$$\Sigma^\dagger = \mathrm{diag}\left\{\tfrac{1}{\sigma_1} < \tfrac{1}{\sigma_2} < \ldots < \tfrac{1}{\sigma_r} > 0\right\}$$

The pseudoinverse of the full rank matrix $\mathbf{X}$ $(n \times m)$ has the following properties:

$n = m$ The equation system is determined and $\mathbf{X}^\dagger = \mathbf{X}^{-1}$

$n > m$ The equation system is underdetermined and pseudoinverse becomes the left pseudoinverse[2], $\mathbf{X}^\dagger = \mathbf{X}_L^\dagger = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$. Notice that this will not give a real inverse $\mathbf{X}\mathbf{X}_L^\dagger = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T \neq \mathbf{I}$. However, known that there are an infinite number of solutions to the problem, $\mathbf{Y}\mathbf{X}_L^\dagger$ will give a unique solution. That is, of all possible solutions for the underdetermined equation system, $\mathbf{Y}\mathbf{X}_L^\dagger$ will give the shortest solution $\left(\min\left\|\mathbf{B}\right\|_F\right)$

$n < m$ The equation system is overdetermined and the pseudoinverse becomes the right pseudoinverse[3], $\mathbf{X}^\dagger = \mathbf{X}_R^\dagger = \mathbf{X}^T\left(\mathbf{X}\mathbf{X}^T\right)^{-1}$.

2.3 Least Squares (LS)

The basic principle behind the least squares problem, is to find an estimator $\mathbf{B}_{\text{ls}}$ which minimizes the sum of squared residuals Σe^2. Given the linear relationship

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{E}$$

the least square optimization problem can be stated as

$$\mathbf{B}_{\text{ls}} = \arg\min_{\mathbf{B}} \|\mathbf{Y} - \mathbf{B}\mathbf{X}\|_F^2$$

where $\|\cdot\|_F$ is the Frobenius norm, and the analytical solution for $\mathbf{B}_{\text{ls}}$ is [4]

$$\mathbf{B}_{\text{ls}} = \mathbf{Y}\mathbf{X}^T\left(\mathbf{X}\mathbf{X}^T\right)^{-1}$$

Proof. The least squares objective function, with the given residual matrix $\mathbf{E}(p\times m)$, can be stated as $S(\mathbf{B}) = \sum_{k=1}^{p}\sum_{j=1}^{m} e_{k,j}^2 = \|\mathbf{E}(\mathbf{B})\|_F^2 = \text{tr}\left(\mathbf{E}\mathbf{E}^H\right)$, where $\text{tr}\left(\mathbf{A}\right) = \sum_i a_{ii}$. We are at the optimum solution when $\partial S(\mathbf{B})/\partial\mathbf{B} = 0$, which gives the problem $\frac{\partial}{\partial\mathbf{B}}\text{tr}\left(\mathbf{E}\mathbf{E}^H\right) = 0$. Assumed that all entries in the matrix $\mathbf{E}$ are real, the conjugate transpose coincides with the transpose resulting in $\mathbf{E}^H = \mathbf{E}^T$. The optimal estimator $\mathbf{B}_{\text{ls}}$ is $\frac{\partial}{\partial\mathbf{B}}\text{tr}\left(\mathbf{E}\mathbf{E}^T\right) = 2\left(\mathbf{Y}-\mathbf{B}\mathbf{X}\right)\mathbf{X}^T = 0$. Assuming that $\mathbf{X}\mathbf{X}^T$ are of full rank (nonsingular) gives the solution $\mathbf{B}_{\text{ls}} = \mathbf{Y}\mathbf{X}^T\left(\mathbf{X}\mathbf{X}^T\right)^{-1}$. □

[2] This is called the left pseudoinverse because $\mathbf{X}_L^\dagger\mathbf{X} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X} = \mathbf{I}$.

[3] This is called the right pseudoinverse because $\mathbf{X}\mathbf{X}_R^\dagger = \mathbf{X}\mathbf{X}^T(\mathbf{X}\mathbf{X}^T)^{-1} = \mathbf{I}$. Remember that this is for the transposed system.

[4] The conventional problem is $\mathbf{B}_{\text{ls}}^T = \arg\min_{\mathbf{B}} \left\|\mathbf{Y}^T - \mathbf{X}^T\mathbf{B}^T\right\|_F^2$,which has the analytical solution $\mathbf{B}_{\text{ls}}^T = \left(\mathbf{X}\mathbf{X}^T\right)^{-1}\mathbf{X}\mathbf{Y}^T$

2.3.1 Least Squares with Pseudoinverse

Because of the inverse of $\mathbf{X}\mathbf{X}^T$ in the expression for the LS estimator $\mathbf{B}_{\text{ls}}$, we can only find estimators for equation systems that are determined ($m = n$) or overdetermined ($m > n$). In the cases where the equation systems are underdetermined ($m < n$) or $\mathbf{X}$ is rank deficient (rank lower than the highest possible rank), the matrix $\mathbf{X}\mathbf{X}^T$ is singular and has no inverse, and the expression for $\mathbf{B}_{\text{ls}}$ breaks down.

To cope with this, a more general expression for the $\mathbf{B}_{\text{ls}}$, can be found using the pseudoinverse of $\mathbf{X}$, and the analytical solution to the LS problem becomes

$$\mathbf{B}_{\text{ls}} = \mathbf{Y}\mathbf{X}^{\dagger}$$

2.3.2 Collinearity

Collinearity is defined as approximate linear dependence, and occur when the independent variables are highly correlated. Collinearity in the matrix $\mathbf{X}$ when performing LS can result in poor estimation of the $\mathbf{B}_{\text{ls}}$ parameters. This is easily shown by the singular values of the matrix. As mentioned before, the singular values indicate the magnitudes of the input and output directions. This can also be interpreted as amount of information about $\mathbf{X}$ in the direction. Direction with large singular values contain much information about $\mathbf{X}$. Directions that are highly correlated, contain little new information and will have singular values close to zero. Perfectly correlated directions contains no new information and will have a zero singular value.

When the pseudoinverse is taken of $\mathbf{X}$, all nonzero singular values are inverted. Therefore, the strongest directions becomes the weakest, and the weakest directions becomes the strongest. Thus, for an ill-conditioned matrix, directions containing irrelevant information have a big impact on the LS estimator, and small errors in the calibration data can give very different estimators $\mathbf{B}_{\text{ls}}$.

2.4 Principal Component Regression (PCR)

As mentioned above, if there is collinearity in the data when performing LS, the estimator will be very sensitive to errors in the data. The idea behind PCR

is to only use directions in the data with relevant information and exclude directions with little and no information, and hence remove the problem with collinearity.

The PCR starts with the principal component analysis (for further reading see Shlens (2005)), where truncated SVD is used to remove directions with little information. Given the data matrix $\mathbf{X}$ $(n \times m)$, the SVD is

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

with the matrices $\mathbf{U}$ $(n \times n)$, $\mathbf{\Sigma}$ $(n \times m)$ and $\mathbf{V}$ $(m \times m)$. The matrices are truncated to rank l, where l is the number of principal components, and gives

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{U}}_l \widetilde{\mathbf{\Sigma}}_l \widetilde{\mathbf{V}}_l^T$$

where $\widetilde{\mathbf{U}}_l$ $(m \times l)$, $\widetilde{\mathbf{\Sigma}}_l$ $(l \times l)$ and $\widetilde{\mathbf{V}}_l$ $(n \times l)$ are the truncated matrices. The number of principal components is usually determined by cross-validation. For the linear relationship

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{E}$$

when using only l principal components, the optimal estimator is

$$\mathbf{B}_{\text{pcr}} = \mathbf{Y}\widetilde{\mathbf{V}}_l \widetilde{\mathbf{\Sigma}}_l^{-1} \widetilde{\mathbf{U}}_l^T = \mathbf{Y}\mathbf{X}_l^\dagger$$

where $\mathbf{X}_l^\dagger$ is the inverse of the truncated SVD.

Proof. See Schreyer et al. (2002) □

2.5 Partial Least Squares (PLS)

The basic PLS regression was developed by H. Wold and have found wide usage within the field of chemometrics. The main idea behind PLS is to find directions in $\mathbf{X}$ which have the greatest covariance with $\mathbf{Y}$ and ensuring that these are treated first. This is advantageous to PCR which only prioritize directions in $\mathbf{X}$ that have the greatest variance.

PLS is an iterative process, and there are several algorithms for finding the estimator. One of the more common methods are the SIMPLS algorithm, which is the standard PLS tool in Matlab. A simplified summary of SIMPLS is shown in Algorithm 2.1. For further reading and the full Algorithm, see de Jong (1993). For an alternative algorithm see Höskuldsson (1988), and for a non-iterative procedure see Di Ruscio (2000).

Algorithm 2.1 Simplified SIMPLS algorithm for PLS regression

1. mean center the $\mathbf{X}$ and $\mathbf{Y}$ data, giving $\mathbf{X}_0$ and $\mathbf{Y}_0$.
2. compute the cross-product $\mathbf{S} = \mathbf{X}_0\mathbf{Y}_0^T$
3. **for** number of components $i = 1, \ldots, l$
   if $i = 1$: compute SVD of $\mathbf{S}$
   if $i > 1$: compute SVD of $\mathbf{S} - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{S}$
   get weights $\mathbf{w}_i$ = first left singular vector
   compute scores $\mathbf{t}_i = \mathbf{X}_0^T\mathbf{w}_i$
   compute loadings $\mathbf{p}_i = \mathbf{X}_0\mathbf{t}_i/(\mathbf{t}_i^T\mathbf{t}_i)$
   store vectors $\mathbf{w}_i$, $\mathbf{t}_i$ and $\mathbf{p}_i$ into $\mathbf{W}, \mathbf{T}$, and$\mathbf{P}$ respectively.
   end
4. compute regression coefficients $\mathbf{B}_{\text{pls}} = \mathbf{Y}_0(\mathbf{T}\mathbf{W}^T)$

Chapter 3

Optimal Model Based Estimators

Another way to develop estimators are by using detailed models of a given system. In this chapter we will derive three optimal "open-loop" estimators and one optimal "closed-loop" estimator for linear systems. With optimal is implied estimators which gives the smallest prediction error $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ for a given set of conditions. In the next chapter, the usage of the optimal "closed-loop" estimator will be extended to also apply for data. Note that the model based estimators are denoted $\mathbf{H}$ to distinguish them from the data based estimators.

With the term "open-loop" estimator, it is implied that the predicted primary variables $\hat{\mathbf{y}}$ are used for monitoring purposes, and not for control. It should noted that this is not the same as implying that primary variables (also secondary variables) are uncontrolled. They can in fact be controlled by other means than the predictions.

We have thought of three main types of "open-loop" control scenarios where $\hat{\mathbf{y}}$ are used for monitoring purposes:

S1: Predicting primary variables from a system with no control ($\mathbf{u}$ is a free variable) .

S2: Predicting primary variables from a system where primary variables $\mathbf{y}$ are controlled ($\mathbf{u}$ is used for keeping $\mathbf{y} = \mathbf{y}_s$).

S3: Predicting primary variables from a system where secondary variables $\mathbf{z}$ are controlled ($\mathbf{u}$ is used for keeping $\mathbf{z} = \mathbf{z}_s$).

With a "closed-loop" estimator it is implied that the $\hat{\mathbf{y}}$ will be used for control purposes. That is,

S4: Predicting primary variables from a system where the predictions $\hat{\mathbf{y}}$ are controlled ($\mathbf{u}$ is used to keep $\hat{\mathbf{y}} = \mathbf{y}_s$).

The block diagram of the four scenarios are shown in Figure 3.1. The estimators have been developed from the following linear system:

Linear System

For the four cases, we will consider the prediction error (residual) $\mathbf{e}$ defined as

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} \tag{3.1}$$

when using a linear estimator

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{x}_m \tag{3.2}$$

assuming linear models for measurements $\mathbf{x}$, primary variables $\mathbf{y}$, and secondary variables $\mathbf{z}$

$$\mathbf{x} = \mathbf{G}_x\mathbf{u} + \mathbf{G}_x^d\mathbf{d} \tag{3.3}$$

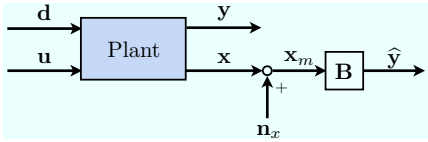
$$\mathbf{y} = \mathbf{G}_y\mathbf{u} + \mathbf{G}_y^d\mathbf{d} \tag{3.4}$$

$$\mathbf{z} = \mathbf{G}_z\mathbf{u} + \mathbf{G}_z^d\mathbf{d} \tag{3.5}$$

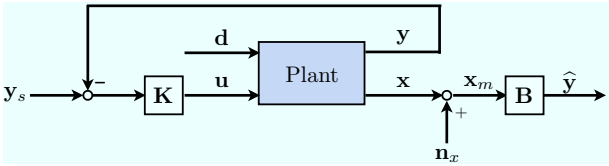
where the actual measurements $\mathbf{x}_m$, containing measurement noise $\mathbf{n}_x$, is

$$\mathbf{x}_m = \mathbf{x} + \mathbf{n}_x \tag{3.6}$$

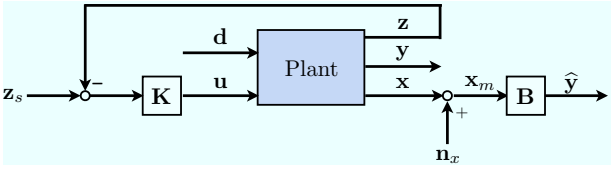
It is also assumed that $\dim(\mathbf{y}) = \dim(\mathbf{z}) = \dim(\mathbf{u})$.



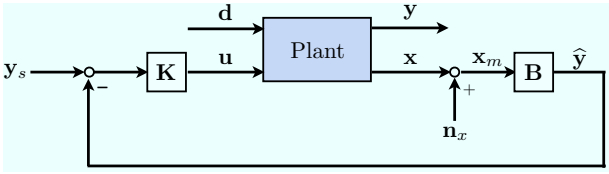
(a) S1: No control and $\mathbf{u}$ is a free variable. $\mathbf{B}$ is an "open-loop" estimator.



(b) S2: Control of primary variable $\mathbf{y}$. The controller $\mathbf{K}$ adjusts $\mathbf{u}$ such that $\mathbf{y} = \mathbf{y}_s$. $\mathbf{B}$ is an "open-loop" estimator.



(c) S3: Control of secondary variable $\mathbf{z}$. The controller $\mathbf{K}$ adjusts $\mathbf{u}$ such that $\mathbf{z} = \mathbf{z}_s$. $\mathbf{B}$ is an "open-loop" estimator.



(d) S4: Control of the predicted primary variable $\hat{\mathbf{y}}$. The controller $\mathbf{K}$ adjusts $\mathbf{u}$ such that $\hat{\mathbf{y}} = \mathbf{y}_s$. $\mathbf{B}$ is an "closed-loop" estimator.

Figure 3.1: Block diagrams of the four control scenarios.

3.1 Optimal "Open-Loop" Estimators

3.1.1 Open-Loop Operation ($\mathbf{u}$ free variable)

To find the optimal estimator for open-loop operation, prediction error has to be expressed as a function of the system and the estimator.

Lemma 1. Prediction error for a given "open-loop" estimator $\mathbf{H}$ ($\mathbf{u}$ is a free variable). *For a given linear estimator $\mathbf{H}$, the prediction error $\mathbf{e}$, when applied to the system defined above, and considering the degrees of freedom $\mathbf{u}$ as free variables, can for a given input $\mathbf{u}$, disturbance $\mathbf{d}$ and noise $\mathbf{n}_y$, be expressed as*

$$\mathbf{e}(\mathbf{H}) = \begin{bmatrix} (\mathbf{G}_y - \mathbf{H}\mathbf{G}_x) & \left(\mathbf{G}_y^d - \mathbf{H}\mathbf{G}_x^d\right) & -\mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{d} \\ \mathbf{n}_x \end{bmatrix} \tag{3.7}$$

Proof. An expression for $\hat{\mathbf{y}}$ as an explicit function of $\mathbf{u}$, $\mathbf{d}$ and $\mathbf{n}_x$ is obtained by combining (3.2), (3.6) and (3.3).

$$\hat{\mathbf{y}} = \mathbf{H}\left(\mathbf{G}_x\mathbf{u} + \mathbf{G}_x^d\mathbf{d} + \mathbf{n}_x\right)$$

Using the definition of prediction error with the expression for $\hat{\mathbf{y}}$ and (3.4) gives

$$\mathbf{e}(\mathbf{H}) = (\mathbf{G}_y - \mathbf{H}\mathbf{G}_x)\,\mathbf{u} + \left(\mathbf{G}_y^d - \mathbf{H}\mathbf{G}_x^d\right)\mathbf{d} - \mathbf{H}\mathbf{n}_x$$

which is the same as (3.7) □

Given that the operation will have different variations in the input variables $\mathbf{u}$, $\mathbf{d}$ and $\mathbf{n}_x$, it must be factored in to find the best estimator. For the "open-loop" estimators, we decided to minimize the expected prediction error. That is, input variables are expected to have a normal distribution, and we want to find the estimator that gives the best prediction for the expected variation. We could also use some bounds on the input variables and optimize for worst case scenarios. This will be done for the "closed-loop" estimators, but it was not taken into account for the "open-loop" estimators.

Lemma 2. Expected prediction error for a given "open-loop" estimator $\mathbf{H}$. *Let the disturbance and noise be normalized on the form*

$$\mathbf{u} = \mathbf{W}_u \mathbf{u}'$$
$$\mathbf{d} = \mathbf{W}_d \mathbf{d}'$$
$$\mathbf{n}_x = \mathbf{W}_{n_x} \mathbf{n}_x'$$

so that the prediction error from Lemma 1 can be expressed

$$\mathbf{e}(\mathbf{H}) = \underbrace{\begin{bmatrix} (\mathbf{G}_y - \mathbf{H}\mathbf{G}_x)\mathbf{W}_u & \left(\mathbf{G}_y^d - \mathbf{H}\mathbf{G}_x^d\right)\mathbf{W}_d & -\mathbf{H}\mathbf{W}_{n_x} \end{bmatrix}}_{\mathbf{M}_{\mathrm{ol}}(\mathbf{H})} \begin{bmatrix} \mathbf{u}' \\ \mathbf{d}' \\ \mathbf{n}_x' \end{bmatrix}$$

For the expected prediction error, assume normal distribution for the degrees of freedom, disturbances and noise, and let the normalized variables be scaled such that

$$u' \sim \mathcal{N}(0,1) \quad d' \sim \mathcal{N}(0,1) \quad n_y' \sim \mathcal{N}(0,1)$$

where u', d' and n_y' are the elements of the normalized vectors $\mathbf{u}'$, $\mathbf{d}'$ and $\mathbf{n}_y'$ respectively, and $\mathcal{N}(0,1)$ denotes a normal distribution with a zero mean and unit standard deviation. The diagonal scaling matrices $\mathbf{W}_u$, $\mathbf{W}_d$ and $\mathbf{W}_{n_x}$ contain the standard deviations of the elements in $\mathbf{u}$, $\mathbf{d}$ and $\mathbf{n}_x$, respectively. The expected prediction error (Kariwala et al., 2008) then becomes

$$\|\mathbf{e}(\mathbf{H})\|_{2,\exp} = \frac{1}{2}\|\mathbf{M}_{\mathrm{ol}}(\mathbf{H})\|_F^2$$

In Theorem 1, we find the optimal estimator for the given expression.

Theorem 1. Optimal "open-loop" estimator $\mathbf{H}$ for open-loop operation. *The optimal "open-loop" estimator $\mathbf{H}$ for open-loop operation where the degrees of freedom $\mathbf{u}$ are considered free variables, when applied to the system defined above and considering the expected prediction error (see Lemma 1 and 2), is*

$$\mathbf{H}_1 = \mathbf{Y}_1 \mathbf{X}_1^\dagger$$

where $\mathbf{X}^\dagger$ is the pseudoinverse of $\mathbf{X}$, and

$$\mathbf{Y}_1 = \begin{bmatrix} \mathbf{G}_y\mathbf{W}_u & \mathbf{G}_y^d\mathbf{W}_d & 0 \end{bmatrix}$$

$$\mathbf{X}_1 = \begin{bmatrix} \mathbf{G}_x\mathbf{W}_u & \mathbf{G}_x^d\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix}$$

Proof. In Lemma 2 we showed that minimizing $\|\mathbf{e}(\mathbf{H})\|_2$ is equivalent to minimizing $\frac{1}{2}\|\mathbf{M}(\mathbf{H})\|_F^2$ for the expected prediction error. Because the scaling factor $1/2$ and the squaring of the norm will not affect the optimal solution of $\mathbf{H}$, they will be omitted from the optimization problem. By expanding the optimization problem $\min_{\mathbf{H}} \|\mathbf{M}(\mathbf{H})\|$ to

$$\min_{\mathbf{H}} \Big\| \underbrace{\begin{bmatrix} \mathbf{G}_y\mathbf{W}_u & \mathbf{G}_y^d\mathbf{W}_d & 0 \end{bmatrix}}_{\mathbf{Y}_1} - \mathbf{H} \underbrace{\begin{bmatrix} \mathbf{G}_x\mathbf{W}_u & \mathbf{G}_x^d\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix}}_{\mathbf{X}_1} \Big\| = \min_{\mathbf{H}} \|\mathbf{Y}_1 - \mathbf{H}\mathbf{X}_1\|$$

we recognize that this is the least squares problem with the known optimal solution

$$\mathbf{H}_1 = \mathbf{Y}_1\mathbf{X}_1^\dagger$$

□

3.1.2 Closed-Loop Primary Variables ($\mathbf{y}$)

In Theorem 1, we find the optimal estimator when primary variables $\mathbf{y}$ are controlled. The derivation is analog to the derivation of the previous estimator.

Theorem 2. Optimal "open-loop" estimator for closed-loop operation (controlled $\mathbf{y}$). *The optimal "open-loop" estimator $\mathbf{H}$ for closed-loop operation where the degrees of freedom $\mathbf{u}$ are adjusted such that the primary variables $\mathbf{y}$ are kept at the setpoints $\mathbf{y}_s$*

$$\mathbf{y} = \mathbf{y}_s$$

when applied to the system defined above and considering the expected prediction error, is

$$\mathbf{H}_2 = \mathbf{Y}_2\mathbf{X}_2^\dagger$$

and

$$\mathbf{Y}_2 = \begin{bmatrix} \mathbf{W}_{y_s} & 0 & 0 \end{bmatrix}$$

$$\mathbf{X}_2 = \begin{bmatrix} \mathbf{G}_x^{cl}\mathbf{W}_{y_s} & \mathbf{F}\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix}$$

where $\mathbf{G}_x^{cl} = \mathbf{G}_x\mathbf{G}_y^{-1}$ *and* $\mathbf{F} = \mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{G}_y^d$

Proof. Considering that $\mathbf{u}$ is used for keeping $\mathbf{y} = \mathbf{y}_s$. Solving (3.4) with respects to $\mathbf{u}$ when $\mathbf{y} = \mathbf{y}_s$ gives

$$\mathbf{u} = \mathbf{G}_y^{-1}\mathbf{y}_s - \mathbf{G}_y^{-1}\mathbf{G}_y^d\mathbf{d}$$

An expression for $\hat{\mathbf{y}}$ as an explicit function of $\mathbf{y}_s$, $\mathbf{d}$ and $\mathbf{n}_x$ is obtained by combining (3.2), (3.6), (3.3) and the expression for $\mathbf{u}$.

$$\hat{\mathbf{y}} = \mathbf{H}\left[\mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{y}_s + \left(\mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{G}_y^d\right)\mathbf{d} + \mathbf{n}_x\right]$$

Here $\left(\mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{G}_y^d\right)$ is recognized as the optimal sensitivity $\mathbf{F}$ (see section A.1 for derivation), and $\mathbf{G}_x\mathbf{G}_y^{-1}$ as the closed-loop gain $\mathbf{G}_x^{cl}$. The expression becomes

$$\hat{\mathbf{y}} = \mathbf{H}\left[\mathbf{G}_x^{cl}\mathbf{y}_s + \mathbf{F}\mathbf{d} + \mathbf{n}_x\right]$$

Using the definition of prediction error with the expression for $\hat{\mathbf{y}}$ and the assumption $\mathbf{y} = \mathbf{y}_s$ gives

$$\mathbf{e}(\mathbf{H}) = \begin{bmatrix} \left(\mathbf{I} - \mathbf{H}\mathbf{G}_x^{cl}\right) & (-\mathbf{H}\mathbf{F}) & -\mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{y}_s \\ \mathbf{d} \\ \mathbf{n}_x \end{bmatrix}$$

Proceeding analogous to Lemma 2 and Theorem 1, will result in the given proposition. □

3.1.3 Closed-Loop Secondary Variables ($\mathbf{z}$)

In Theorem 1, we find the optimal estimator when secondary variables $\mathbf{z}$ are controlled. This is probably the most common way of controlling distillation columns, where a single stage temperature is controlled instead of composition.

Theorem 3. Optimal "open-loop" estimator for closed-loop operation (controlled $\mathbf{z}$). *The optimal "open-loop" estimator $\mathbf{H}$ for closed-loop operation where the degrees of freedom $\mathbf{u}$ are adjusted such that the secondary variables $\mathbf{z}$ are kept at the setpoints $\mathbf{z}_s$*

$$\mathbf{z} = \mathbf{z}_s$$

when applied to the system defined above, assuming a linear model for the secondary variables

$$\mathbf{z} = \mathbf{G}_z\mathbf{u} + \mathbf{G}_z^d\mathbf{d} \tag{3.8}$$

and considering the expected prediction error, is

$$\mathbf{H}_3 = \mathbf{Y}_3\mathbf{X}_3^{\dagger}$$

and

$$\mathbf{Y}_3 = \begin{bmatrix} \mathbf{G}_y^{cl}\mathbf{W}_{z_s} & \mathbf{F}'_y\mathbf{W}_d & 0 \end{bmatrix}$$
$$\mathbf{X}_3 = \begin{bmatrix} \mathbf{G}_x^{cl}\mathbf{W}_{z_s} & \mathbf{F}'_x\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix}$$

where $\mathbf{G}_y^{cl} = \mathbf{G}_y\mathbf{G}_z^{-1}$, $\mathbf{G}_x^{cl} = \mathbf{G}_x\mathbf{G}_z^{-1}$, $\mathbf{F}'_y = \mathbf{G}_y^d - \mathbf{G}_y\mathbf{G}_z^{-1}\mathbf{G}_z^d$ *and* $\mathbf{F}'_x = \mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_z^{-1}\mathbf{G}_z^d$

Proof. Considering that $\mathbf{u}$ is used for keeping $\mathbf{z} = \mathbf{z}_s$. Solving (3.8) with respects to $\mathbf{u}$ when $\mathbf{z} = \mathbf{z}_s$ gives

$$\mathbf{u} = \mathbf{G}_z^{-1}\mathbf{z}_s - \mathbf{G}_z^{-1}\mathbf{G}_z^d\mathbf{d}$$

An expression for $\mathbf{y}$ as an explicit function of $\mathbf{z}_s$, $\mathbf{d}$ and $\mathbf{n}_x$ is obtained by combining (3.4) and the expression for $\mathbf{u}$

$$\mathbf{y} = \mathbf{G}_y\mathbf{G}_z^{-1}\mathbf{z}_s + \left(\mathbf{G}_y^d - \mathbf{G}_y\mathbf{G}_z^{-1}\mathbf{G}_z^d\right)\mathbf{d}$$

Recognizing the optimal sensitivity $\mathbf{F}'_y$ and the closed-loop gain $\mathbf{G}_y^{cl}$ for the primary variable (when $\mathbf{y}_2$ in closed-loop), gives

$$\mathbf{y} = \mathbf{G}_y^{cl}\mathbf{z}_s + \mathbf{F}'_y\mathbf{d}$$

An expression for $\hat{\mathbf{y}}$ as an explicit function of $\mathbf{z}_s$, $\mathbf{d}$ and $\mathbf{n}_x$ is obtained by combining (3.2), (3.6), (3.3) and the expression for $\mathbf{u}$

$$\hat{\mathbf{y}} = \mathbf{H}\left[\mathbf{G}_x\mathbf{G}_z^{-1}\mathbf{z}_s + \left(\mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_z^{-1}\mathbf{G}_z^d\right)\mathbf{d} + \mathbf{n}_x\right]$$

Recognizing the optimal sensitivity $\mathbf{F}'_x$ and the closed-loop gain $\mathbf{G}_x^{cl}$ (when $\mathbf{y}_2$ in closed-loop), gives

$$\hat{\mathbf{y}} = \mathbf{H}\left(\mathbf{G}_x^{cl}\mathbf{z}_s + \mathbf{F}'_x\mathbf{d} + \mathbf{n}_x\right)$$

Using the definition of prediction error with the expression for $\hat{\mathbf{y}}$ and $\mathbf{y}$ gives

$$\mathbf{e}(\mathbf{H}) = \begin{bmatrix} \left(\mathbf{G}_y^{cl} - \mathbf{H}\mathbf{G}_x^{cl}\right) & \left(\mathbf{F}'_y - \mathbf{H}\mathbf{F}'_x\right) & -\mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{z}_s \\ \mathbf{d} \\ \mathbf{n}_x \end{bmatrix} \tag{3.9}$$

Proceeding analogous to Lemma 2 and Theorem 1, will result in the given proposition. □

3.2 Optimal "Closed-Loop" Estimator

In Lemma 3 we find an expression for the prediction error under the assumption that the prediction is used for controlling the primary variables.

Lemma 3. Prediction error for a given "closed-loop" estimator $\mathbf{H}$. *For a given linear estimator $\mathbf{H}$, the prediction error $\mathbf{e}$, when applied to the system defined above, and considering the "closed-loop" case where the degrees of freedom $\mathbf{u}$ is adjusted such that the predicted variables $\hat{\mathbf{y}}$ are kept at the setpoints $\mathbf{y}_s$*

$$\hat{\mathbf{y}} = \mathbf{y}_s$$

can for given disturbances $\mathbf{d}$, noise $\mathbf{n}_y$ and setpoints $\mathbf{y}_s$, be expressed as

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = -\mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{H} \begin{bmatrix} \mathbf{F} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ \mathbf{n}_x \end{bmatrix} + \left[\mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} - \mathbf{I}\right] \mathbf{y}_s \tag{3.10}$$

Proof. An expression for $\hat{\mathbf{y}}$ as an explicit function of $\mathbf{u}$, $\mathbf{d}$, and $\mathbf{n}_x$ is obtained by combining (3.2), (3.6) and (3.3)

$$\hat{\mathbf{y}} = \mathbf{H}\left(\mathbf{G}_x\mathbf{u} + \mathbf{G}_x^d\mathbf{d} + \mathbf{n}_x\right)$$

We assume that the predictions $\hat{\mathbf{y}}$ are held at the setpoints $\mathbf{y}_s$ by manipulating the free variable $\mathbf{u}$. Solving $\hat{\mathbf{y}}$ with respect to $\mathbf{u}$ when $\hat{\mathbf{y}} = \mathbf{y}_s$, gives

$$\mathbf{u} = -\left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{H}\left(\mathbf{G}_x^d\mathbf{d} + \mathbf{n}_x\right) + \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{y}_s$$

and inserting the expression $\mathbf{u}$ for into (3.4) gives $\mathbf{y}$ as an explicit function of $\mathbf{d}$, $\mathbf{n}_x$ and $\mathbf{y}_s$

$$\begin{aligned} \mathbf{y} &= -\mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{H}\left(\mathbf{G}_x^d\mathbf{d} + \mathbf{n}_x\right) + \mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{y}_s + \mathbf{G}_y^d\mathbf{d} \\ &= -\mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{H}\left[\left(\mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{G}_y^d\right)\mathbf{d} + \mathbf{n}_x\right] + \mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{y}_s \end{aligned}$$

Here $\left(\mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{G}_y^d\right)$ is recognized as the optimal sensitivity $\mathbf{F}$ (see section Appendix A.1 for derivation), and the expression becomes

$$\mathbf{y} = -\mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{H}\left(\mathbf{F}\mathbf{d} + \mathbf{n}_x\right) + \mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{y}_s$$

the prediction error is then

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{y}_s = -\mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{H}\left(\mathbf{F}\mathbf{d} + \mathbf{n}_x\right) + \mathbf{G}_y \left(\mathbf{H}\mathbf{G}_x\right)^{-1} \mathbf{y}_s - \mathbf{y}_s$$

which is the same as (3.10) □

Lemma 4. Simplifying the minimization problem. *As shown in Lemma 3, the prediction error can be written as the sum of two terms*

$$\mathbf{e}(\mathbf{H}) = \underbrace{-\mathbf{G}_y(\mathbf{H}\mathbf{G}_x)^{-1}\mathbf{H}\begin{bmatrix}\mathbf{F} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{d}\\ \mathbf{n}_x\end{bmatrix}}_{\mathbf{e}_1(\mathbf{H})} + \underbrace{\left[\mathbf{G}_y(\mathbf{H}\mathbf{G}_x)^{-1} - \mathbf{I}\right]\mathbf{y}_s}_{\mathbf{e}_2(\mathbf{H})}$$

Consider minimizing the prediction error $\|\mathbf{e}_1 + \mathbf{e}_2\|_2$ *with the estimator* $\mathbf{H}$ *as a degree of freedom. In the first term* ($\mathbf{e}_1$) *we have extra degrees of freedom in* $\mathbf{H}$*, which always can be used to set the second term to zero* ($\mathbf{e}_2 = 0$) *by adding the constraint* $\mathbf{H}\mathbf{G}_x = \mathbf{G}_y$*.*

This means that the problem of finding the optimal $\mathbf{H}$ *that minimize the prediction error for a given* $\mathbf{d}$*,* $\mathbf{n}_y$ *and* $\mathbf{y}_{1s}$ *is reduced to minimizing* $\|\mathbf{e}_1(\mathbf{H})\|_2$ *where*

$$\mathbf{e}_1(\mathbf{H}) = -\mathbf{H}\begin{bmatrix}\mathbf{F} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{d}\\ \mathbf{n}_y\end{bmatrix}$$

subjected to the constraint

$$\text{s.t.} \quad \mathbf{H}\mathbf{G}_x = \mathbf{G}_y$$

Proof. There are extra degrees of freedom ($\mathbf{D}$) related to the first term $\mathbf{e}_1(\mathbf{H})$. Specifically $\mathbf{e}_1(\mathbf{H}) = \mathbf{e}_1(\mathbf{D}\mathbf{H})$ where $\mathbf{D}$ is any square nonsingular matrix. This follows because

$$(\mathbf{D}\mathbf{H}\mathbf{G}_x)^{-1}\mathbf{D}\mathbf{H} = (\mathbf{H}\mathbf{G}_x)^{-1}\mathbf{D}^{-1}\mathbf{D}\mathbf{H} = (\mathbf{H}\mathbf{G}_x)^{-1}\mathbf{H}$$

$\mathbf{D}$ can be chosen freely without affecting $\mathbf{e}_1(\mathbf{H})$, so we may choose it such that the last term is zero, $\mathbf{e}_2(\mathbf{H}) = 0$, corresponding to having $\mathbf{H}\mathbf{G}_x = \mathbf{G}_y$. □

It might seem reasonable that the optimum solution for $\mathbf{H}$ would be if $\mathbf{e}_1 = -\mathbf{e}_2$. This would be possible if the specific disturbance and noise where known in advanced. In Lemma 5, the prediction error will be expressed for a set of disturbances and noise, where the specific disturbances and noise can be positive or negative, and are unknown in advance. Thus the terms cannot be used to counteract each other.

The above Lemma shows that the prediction error for a given $\mathbf{d}$, $\mathbf{n}_x$ and $\mathbf{y}_s$ is not affected by the setpoint $\mathbf{y}_s$ as long as the constraint $\mathbf{H}\mathbf{G}_x = \mathbf{G}_y$ is fulfilled. Because of this, and to simplify further derivations, only $\mathbf{e}_1$ will be considered in the Lemma below, where expressions for the expected and worst case prediction error for a given *set* of $\mathbf{d}$ and $\mathbf{n}_x$ are proposed.

Lemma 5. Expected and worst case prediction error for a given "closed-loop" estimator $\mathbf{H}$. *Let the disturbance and noise be normalized on the form*

$$\mathbf{d} = \mathbf{W}_d \mathbf{d}'$$
$$\mathbf{n}_x = \mathbf{W}_{\mathbf{n}_x} \mathbf{n}'_x$$

so that the first term in the prediction error from Lemma 4 can be expressed

$$\mathbf{e}_1(\mathbf{H}) = -\underbrace{\mathbf{G}_y (\mathbf{H}\mathbf{G}_x)^{-1} \mathbf{H} \begin{bmatrix} \mathbf{F}\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix}}_{\mathbf{M}(\mathbf{H})} \begin{bmatrix} \mathbf{d}' \\ \mathbf{n}'_x \end{bmatrix}$$

1. For the expected prediction error, assume normal distribution for disturbances and noise, and let the normalized variables be scaled such that
$$d' \sim \mathcal{N}(0,1) \quad n'_y \sim \mathcal{N}(0,1)$$
where d' and n'_y are the elements of the normalized vectors $\mathbf{d}'$ and $\mathbf{n}'_y$ respectively, and $\mathcal{N}(0,1)$ denotes a normal distribution with zero mean and unit standard deviation. The diagonal scaling matrices $\mathbf{W}_d$ and $\mathbf{W}_{n_x}$ contain the standard deviations of the elements in $\mathbf{d}$ and $\mathbf{n}_y$, respectively. The expected prediction error (Kariwala et al., 2008) then becomes
$$\|\mathbf{e}_1\|_{2,\exp} = \frac{1}{2} \|\mathbf{M}(\mathbf{H})\|_F^2$$
2. For the worst case prediction error, let the normalized variables be scaled such that the combined norm is smaller or equal to one
$$\left\| \begin{bmatrix} \mathbf{d}' \\ \mathbf{n}'_y \end{bmatrix} \right\|_2 \leq 1$$
Then the worst case prediction error (Halvorsen et al., 2003) becomes
$$\|\mathbf{e}_1\|_{2,\mathrm{wc}} = \frac{1}{2} \bar{\sigma}(\mathbf{M}(\mathbf{H}))^2$$
where $\bar{\sigma}(\mathbf{M}(\mathbf{H}))$ is the larges singular value of $\mathbf{M}(\mathbf{H})$

Proof. From the definition of the norm, the negative sign in the prediction error $\mathbf{e}$ will not affect the solution and can be omitted. The expected and worst case prediction error follows from the definition of the norm, and are proven by Kariwala et al. (2008) and Halvorsen et al. (2003), respectively. □

Theorem 4. Optimal "closed-loop" estimator $\mathbf{H}$. *The optimal "closed-loop" estimator $\mathbf{H}$ for both the expected and worst case prediction error (see Lemma 5), can be found by solving the optimization problem*

$$\begin{aligned} \mathbf{H}_4 = \arg\min_{\mathbf{H}} & \left\| \mathbf{H} \begin{bmatrix} \mathbf{F}\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix} \right\|_F \\ \text{s.t.} \quad & \mathbf{H}\mathbf{G}_x = \mathbf{G}_y \end{aligned} \tag{3.11}$$

Proof. In Lemma 4 we showed that by introducing the constraint $\mathbf{H}\mathbf{G}_x = \mathbf{G}_y$, we only need to consider minimizing the term $\|\mathbf{e}_1(\mathbf{H})\|_2$. In Lemma 5 we showed that minimizing $\|\mathbf{e}_1(\mathbf{H})\|_2$ is equivalent to minimizing $\|\mathbf{M}(\mathbf{H})\|_F$ and $\bar{\sigma}(\mathbf{M}(\mathbf{H}))$ for the expected and worst case prediction error, respectively. It turns out that for this particular problem, the optimal $\mathbf{H}$ that minimizes $\|\mathbf{M}(\mathbf{H})\|_F$ also minimizes $\bar{\sigma}(\mathbf{M}(\mathbf{H}))$. This is not obvious, but is proven by Kariwala et al. (2008). Thus, we only need to consider minimizing $\|\mathbf{M}(\mathbf{H})\|_F$, which by introducing the constraint $\mathbf{H}\mathbf{G}_x = \mathbf{G}_y$ (using the extra degrees of freedom in $\mathbf{H}$), becomes (3.11). □

When scaled for normal distributions, the weights $\mathbf{W}_d$ and $\mathbf{W}_{n_x}$ will contain the standard deviation of the elements in $\mathbf{d}$ and $\mathbf{n}_x$. This scaling might be smaller than for the 2-norm bound case, where we might have chosen to scale with two standard deviations to ensure that the 2-bound constraint is fulfilled. However, if the relative sizes between the weights are the same for both cases, the absolute sizes of the weights will not affect the optimal $\mathbf{H}$.

The optimization problem in Theorem 4 is expressed with open-loop gains ($\mathbf{G}_x$ and $\mathbf{G}_y$), but can also be expressed with closed-loop gains by just substituting the open-loop gains for the closed-loop gains. This can easily be shown by multiplying the constraint $\mathbf{H}\mathbf{G}_x = \mathbf{G}_y$ with $\mathbf{G}_y^{-1}$ on both sides of the equality

$$\text{s.t.} \qquad \mathbf{H} \underbrace{\mathbf{G}_x \mathbf{G}_y^{-1}}_{\mathbf{G}_x^{cl}} = \underbrace{\mathbf{G}_y \mathbf{G}_y^{-1}}_{\mathbf{G}_y^{cl}}$$

It can also be shown by using the system defined above when considering a cascade like arrangement where the exact values of $\mathbf{y}$ is known. In the inner-loop, $\mathbf{u}$ is used for keeping $\mathbf{y}$ at the setpoints $\mathbf{y}'_s$ given by the outer-loop. In the outer-loop, $\mathbf{y}'_s$ is used for keeping the estimates $\hat{\mathbf{y}}$ (expressed as a explicit function of $\mathbf{y}'_s$, $\mathbf{d}$ and $\mathbf{n}_x$) at the setpoints $\mathbf{y}_s$. Assuming perfect control of the inner- and outer-loop such that $\mathbf{y} = \mathbf{y}'_s$ and $\hat{\mathbf{y}} = \mathbf{y}_s$, expressing the prediction

error as a function of $\mathbf{y}_s$, $\mathbf{d}$ and $\mathbf{n}_x$, and following Lemma 4, Lemma 5 and Theorem 4 will give the following optimization problem

$$\begin{aligned} \min_{\mathbf{H}} & \left\| \mathbf{H} \begin{bmatrix} \mathbf{F}\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix} \right\|_F \\ \text{s.t.} & \quad \mathbf{H}\mathbf{G}_x^{cl} = \mathbf{G}_y^{cl} = \mathbf{I} \end{aligned}$$

This is also the case when using secondary variables. For detailed derivation see Appendix A.2.1 and A.2.2.

The optimization problem of finding the optimal estimate $\mathbf{H}$ can be solved by using conventional optimization techniques, but as we show in Theorem 3.2.1, the optimal estimate $\mathbf{H}$ has, under certain assumptions, an analytical solution.

3.2.1 Analytical Solution

Writing the optimization problem (3.11) from Theorem 4 as

$$\min_{\mathbf{H}} \left\| \mathbf{H} \underbrace{\begin{bmatrix} \mathbf{F}\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix}}_{\widetilde{\mathbf{F}}} \right\|_F$$

$$\text{s.t.} \quad \mathbf{H}\mathbf{G}_x = \mathbf{G}_y$$

Under the assumption that $\left(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T\right)$ is of full rank, the optimal "closed-loop" estimator $\mathbf{H}$ have the following analytical solution (Alstad et al., 2009)

$$\mathbf{H}_4^T = \left(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T\right)^{-1}\mathbf{G}_x\left(\mathbf{G}_x^T\left(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T\right)^{-1}\mathbf{G}_x\right)^{-1}\mathbf{G}_y^T \tag{3.12}$$

Chapter 4

Optimal "Closed-Loop" Estimator Used on Data

In the previous Chapter, we derived a "closed-loop" estimator based on explicit models of a system. Now, we want to extend the usage of this estimator to also apply for systems where all available information is given by data. That is, situations where all available information about the primary variables $\mathbf{y}$ and the measurements $\mathbf{x}$ are given by the data matrices $\mathbf{Y}$ and $\mathbf{X}$, respectively. To use the presented "closed-loop" estimator, we would need to separate the data into data blocks that contain information equivalent to the information given by $\mathbf{G}_y$, $\mathbf{G}_x$ and $\widetilde{\mathbf{F}}$. The data block equivalent to $\widetilde{\mathbf{F}}$ must contain information about $\mathbf{y}$ for "optimal" variation in $\mathbf{u}$. Here "optimal" refers to variations in $\mathbf{u}$ that keeps $\mathbf{y}$ constant regardless of disturbances $\mathbf{d}$. The data blocks equivalent to $\mathbf{G}_y$ and $\mathbf{G}_x$ must contain information about $\mathbf{y}$ and $\mathbf{x}$ for "non-optimal" variations in $\mathbf{u}$. That is, variation in $\mathbf{u}$ that does not keep $\mathbf{y}$ constant.

We then present the following two step procedure for using the optimal "closed-loop" estimator on data:

1. Separate data into "optimal" and "non-optimal" data, using techniques described below.

2. Use the "non-optimal" data as $\mathbf{G}_y$ and $\mathbf{G}_x$, and the "non-optimal" data as $\widetilde{\mathbf{F}}$ in the expression for the optimal "closed-loop" estimator.

Separation into data block can be achieved in two ways, by selecting an experimental design that yields the blocks directly, or by transforming the data such that the blocks of interest can be identified.

4.0.2 Experimental Design

Assume that the data contained in both $\mathbf{Y}$ and $\mathbf{X}$ are obtained from two different sources of data ("non-optimal" and "optimal" data), and the data constructed as follows

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} \mathbf{Y}_{\mathrm{non-opt}} & \mathbf{Y}_{\mathrm{opt}} \end{bmatrix} \\ \mathbf{X} &= \begin{bmatrix} \mathbf{X}_{\mathrm{non-opt}} & \mathbf{X}_{\mathrm{opt}} \end{bmatrix} \end{aligned}$$

The "non-optimal" data $\mathbf{Y}_{\mathrm{non-opt}}$ and $\mathbf{X}_{\mathrm{non-opt}}$ (for $\mathbf{y}$ and $\mathbf{x}$ respectively) are found by varying $\mathbf{u}$ and keeping $\mathbf{d}$ constant. Note that the data must be expressed as deviation variables. By scaling this data with the perturbations in $\mathbf{u}$ we obtain $\mathbf{G}_y$ and $\mathbf{G}_x$.

It is strictly not necessary to scale the "non-optimal" data with the perturbation of $\mathbf{u}$. This comes from the constraint $\mathbf{H}\mathbf{G}_x = \mathbf{G}_y$ used in the optimization problem, where the scaling used for obtaining $\mathbf{G}_y$ and $\mathbf{G}_x$ are the same for both. Thus we can equivalently use the constraint $\mathbf{H}\mathbf{X}_{\mathrm{non-opt}} = \mathbf{Y}_{\mathrm{non-opt}}$.

The "optimal" data $\mathbf{Y}_{\mathrm{opt}}$ and $\mathbf{X}_{\mathrm{opt}}$ are found by keeping $\mathbf{y}$ constant for various disturbances $\mathbf{d}$ (including noise). Because the data is expressed as deviation variables, the optimal data $\mathbf{Y}_{\mathrm{opt}}$ becomes a zero matrix. The variations in $\mathbf{d}$ and consequently the variation in $\mathbf{X}_{\mathrm{opt}}$ should be representative for expected operation. This will then directly give $\widetilde{\mathbf{F}}$ with representative weights.

We can now find the optimal "closed-loop" estimator $\mathbf{H}$ by solving (3.11) using the following equivalent data blocks $\mathbf{G}_y = \mathbf{Y}_{\mathrm{non-opt}}$, $\mathbf{G}_x = \mathbf{X}_{\mathrm{non-opt}}$ and $\widetilde{\mathbf{F}} = \mathbf{X}_{\mathrm{opt}}$.

4.0.3 "Closed-Loop" Method (CLM)

Assume that the data matrices $\mathbf{Y}$ and $\mathbf{X}$ do not have separate "optimal" and "non-optimal" data blocks, and that $\mathbf{x}$ and $\mathbf{y}$ are subjected to perturbations in

$\mathbf{u}$ and disturbances $\mathbf{d}$ simultaneously. The data can then be divided into "optimal" and "non-optimal" blocks by the following two-step procedure (Skogestad et al., 2011).

1. Perform a singular value decomposition (SVD) on the data matrix $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

2. Multiply the data matrices $\mathbf{Y}$ and $\mathbf{X}$ with the unitary matrix $\mathbf{V}$

$$\begin{aligned} \mathbf{YV} &= \begin{bmatrix} \mathbf{Y}_{\text{non-opt}} & 0 \end{bmatrix} \\ \mathbf{XV} &= \begin{bmatrix} \mathbf{X}_{\text{non-opt}} & \mathbf{X}_{\text{opt}} \end{bmatrix} \end{aligned}$$

We can then use the same approach as above, where we find the optimal "closed-loop" estimator $\mathbf{H}$ by solving (3.11) and using the following equivalent data blocks $\mathbf{G}_y = \mathbf{Y}_{\text{non-opt}}$, $\mathbf{G}_x = \mathbf{X}_{\text{non-opt}}$ and $\widetilde{\mathbf{F}} = \mathbf{X}_{\text{opt}}$. This method is termed closed-loop method (CLM), and its estimator is denoted $\mathbf{B}_{\text{clm}}$.

Proof. Given that $\mathbf{V}$ is a real unitary matrix from the SVD of the $p \times m$ data matrix $\mathbf{Y}$, the magnitude of the prediction error $\left\|\mathbf{e}\right\|_F = \left\|\mathbf{Y} - \mathbf{HX}\right\|_F$ when multiplied by $\mathbf{V}$ will not be affected, and $\left\|\mathbf{YV} - \mathbf{HXV}\right\|_F = \left\|\mathbf{Y} - \mathbf{HX}\right\|_F$. Thus we can use $\mathbf{V}$ to divide the data up without changing the results. Multiplying $\mathbf{Y}$ with $\mathbf{V}$, and remembering that $\mathbf{V}^T = \mathbf{V}^{-1}$, gives

$$\mathbf{YV} = \mathbf{U}\mathbf{\Sigma}$$

where $\mathbf{\Sigma}$ contains a diagonal matrix Σ_1 of real non-zero singular values σ_i arranged in descending order

$$\mathbf{\Sigma} = \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix}$$

and

$$\Sigma_1 = \operatorname{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_k\}; \quad k = \min(p, m)$$

Writing the unitary matrix $\mathbf{U}$ in block form as $\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix}$ gives

$$\mathbf{YV} = \mathbf{US} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{U}_1\Sigma_1 & 0 \end{bmatrix}$$

Thus by multiplying $\mathbf{Y}$ and $\mathbf{X}$ with the unitary matrix $\mathbf{V}$, the data is transformed into the same shape as the "non-optimal – optimal" data arrangement (zeros in the right part of $\mathbf{Y}$), without affecting the optimal $\mathbf{H}$. □

4.0.4 Truncated "Closed-Loop" Method (Truncated CLM)

CLM suffers from the same weakness as LS, giving poor results for ill-conditioned matrices and underdetermined systems. As an improvement, we propose to perform a principal component analysis prior to the CLM on the $\mathbf{X}$ data. Thus, removing the weaker directions containing collinearity and noise. Also, because of the mathematical constraint imposed by $(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T)^{-1}$ in the analytical expression for the optimal "closed-loop" estimator, the expression breaks down if the system becomes underdetermined. Therefore, it is proposed to use the truncated pseudoinverse on the $\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T$ in the analytical expression. The procedure then becomes as follows,

Before the CLM, perform a principal component analysis on the $\mathbf{X}$ data. Given the data $\mathbf{X}$, with the following SVD

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

Truncate the decomposition to l number of principal components, giving

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{U}}_l\widetilde{\mathbf{\Sigma}}_l\widetilde{\mathbf{V}}_l^T$$

Continue with the CLM described above, but using $\widetilde{\mathbf{X}}$ instead of $\mathbf{X}$. Given the SVD $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, the data transformation becomes

$$\begin{aligned}\mathbf{Y}\mathbf{V} &= \begin{bmatrix} \mathbf{Y}_{\text{non-opt}} & 0 \end{bmatrix} \\ \widetilde{\mathbf{X}}\mathbf{V} &= \begin{bmatrix} \widetilde{\mathbf{X}}_{\text{non-opt}} & \widetilde{\mathbf{X}}_{\text{opt}} \end{bmatrix}\end{aligned}$$

Using the following equivalent data blocks $\mathbf{G}_y = \mathbf{Y}_{\text{non-opt}}$, $\mathbf{G}_x = \widetilde{\mathbf{X}}_{\text{non-opt}}$ and $\widetilde{\mathbf{F}} = \widetilde{\mathbf{X}}_{\text{opt}}$, the truncated CLM estimator can then be fund by the following expression

$$(\mathbf{B}_{\text{clm}}^{\dagger})^T = \big(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T\big)_l^{\dagger}\mathbf{G}_x\big(\mathbf{G}_x^T\big(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T\big)_l^{\dagger}\mathbf{G}_x\big)^{-1}\mathbf{G}_y^T \tag{4.1}$$

where l is the number of principal components used in the initial step.

Chapter 5

Example System: Distillation Column

The distillation model used in this thesis is the multicomponent distillation model made by Antonio Arauju, which is a modification of the model made by Stathis Skouras. The model is based on the "column A" model described in Skogestad and Postlethwaite (2010). The model is readily available on Skogestad's homepage.

Please notice that in this chapter, x, y, and z stands for liquid, vapor and feed composition, *and not* measurements, primary and secondary variables.

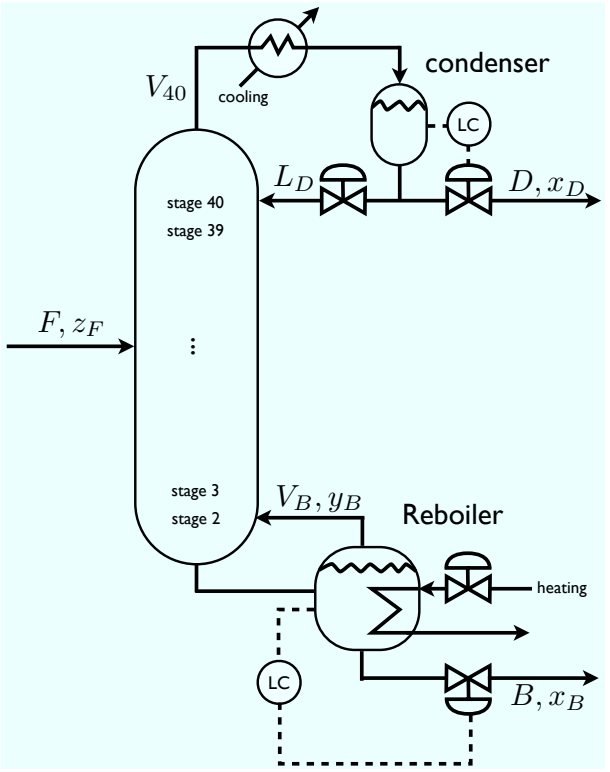
The column has 40 theoretical stages (including the reboiler) and a total condenser, for a total of 41 stages. The feed is introduced on stage 20 as saturated liquid. The column is shown in Figure 5.1 with the used notation, and mass and component balances for the column is shown in Table 5.1.

We assumed constant molar flows, and no vapor holdup. Because the feed was saturated, the vapor flow throughout the column becomes equal to the vapor boilup in the reboiler.

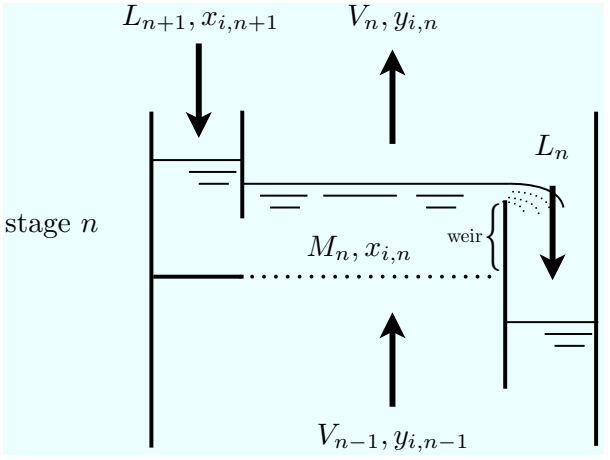
$$V_n = V_B$$

Further we assumed equilibrium on all stages, constant pressure and constant relative volatility. The vapor composition for component i on stage n, then becomes

$$y_{i,n} = \frac{\alpha_i x_{i,n}}{1 + \Sigma_i^{N_C - 1} (\alpha_i - 1) x_{i,n}}$$



(a) Distillation setup with stage numbering



(b) General stage with used notation

Figure 5.1

Table 5.1: Mass and component balances for example column

Reboiler ($n = 1 = B$)

$$\frac{dM_B}{dt} = L_2 - V_B - B$$
$$\frac{d\left(x_{i,B}M_B\right)}{dt} = x_{i,2}L_2 - y_{i,B}V_B - x_{i,B}B$$

Feed stage ($n = 20 = F$)

$$\frac{dM_F}{dt} = F + V_{19} + L_{21} - V_F - L_F$$
$$\frac{d\left(x_{i,F}M_F\right)}{dt} = z_{F,i}F + y_{i,19}V_{19} + x_{i,21}L_{21} - y_{i,F}V_F - x_{i,F}L_F$$

Condenser ($n = 41 = D$)

$$\frac{dM_D}{dt} = V_{40} - L_D - D$$
$$\frac{d\left(x_{i,D}M_D\right)}{dt} = y_{i,40}V_{40} - x_{i,D}L_D - x_{i,D}D$$

Generic stage for enriching and stripping section ($n = 2, \ldots, 19, 21, \ldots, 40$)

$$\frac{dM_n}{dt} = V_{n-1} + L_{n+1} - V_n - L_n$$
$$\frac{d(x_{i,n}M_n)}{dt} = y_{i,n-1}V_{n-1} + x_{i,n+1}L_{n+1} - y_{i,n}V_n - x_{i,n}L_n$$

where α_i is the relative volatility for component i, and N_C is the total number of components.

Liquid dynamics where implemented with the Franci's weir formula

$$L_n = K \cdot (M_{\text{ow}})^{3/2}$$

where K is a flow constant and M_{ow} is the holdup that is over the weirs capacity. M_{ow} can also be expressed as

$$M_{\text{ow}} = M_n - M_{\text{uw}}$$

where M_n is the holdup on stage n, and M_{uw} is the holdup limit (holdup under weir) for the weir. See Table 5.2 for specifications.

Stage temperatures where estimated from the linear relationship

$$T_n = \sum_{i=1}^{N_c} x_{i,n}T_i^b$$

Table 5.2: Constants for liquid dynamics

symbol	description	constant
K_{of}	weir constant over feed	21.65032
K_{uf}	weir constant below feed	29.65032
M_{uw}	holdup under weir	0.25

where $x_{i,n}$ is the mole fraction for component i on stage n, and T_i^b is the components boiling temperature. This might seem as a crude estimation, but is sufficient for our needs.

5.1 Column Control

The liquid levels in the condenser and reboiler are controlled by the top and bottom product streams, respectively. The column has two remaining degrees of freedom, namely the reflux L_D and vapor boilup V_B. These are used in the subsequent chapters for controlling either product compositions directly by measuring product compositions, indirectly by the predicted product compositions, or by using specific stage temperatures.

Chapter 6

Linear System

Model based estimators have been developed for four different control scenarios, that is:

S1: no control, and $\mathbf{u}$ is a free variable. ("open-loop" estimator)

S2: control of *primary variables* $\mathbf{y}$ where $\mathbf{u}$ is used for keeping $\mathbf{y} = \mathbf{y}_s$. ("open-loop" estimator)

S3: control of *secondary variables* $\mathbf{z}$ where $\mathbf{u}$ is used for keeping $\mathbf{z} = \mathbf{z}_s$. ("open-loop" estimator)

S4: control of the *predicted primary variables* $\hat{\mathbf{y}}$ where $\mathbf{u}$ is used for keeping $\hat{\mathbf{y}} = \mathbf{y}_s$. ("closed-loop" estimator)

6.1 Problem Definition

Estimators performance for the four scenarios will be tested when using a linear system, and the following problems will be treated in the first part of this chapter:

1. Which of the optimal "open-loop" and "closed-loop" estimators performs best on the different scenarios.

2. Which calibration data gives the best estimators. This especially apply to scenario 4, for which we cannot generate calibration data. This is a consequence of the implicit nature of scenario 4. That is, the estimator $\mathbf{B}$ is a part of $\mathbf{u}$, and thus we cannot generate calibration data before the estimators is known.

3. Which regression method gives the best estimators.

In the second part, a further investigation into "closed-loop" performance is conducted where the following problem will be treated:

1. For different number of available stage temperature measurements, which method gives the best performance.

6.1.1 Linear Example Column

The system used in this chapter is the linear approximation to "column A" example with binary mixture from Chapter 5. The degree of freedom $\mathbf{u}$ is the internal streams L_D and V_B, and the disturbance $\mathbf{d}$ are changes in feed composition. The primary variable (top and bottom composition) is given by

$$\mathbf{y} = \begin{bmatrix} -1.297 & 1.278 \\ -1.503 & 1.522 \end{bmatrix} \mathbf{u} + \begin{bmatrix} -0.847 \\ -0.921 \end{bmatrix} \mathbf{d}$$

For the first part, only the eight stage temperatures T_n ($n = 5, 10, 15, 20, 25, 30, 35, 40$) were used, and expressed by

$$\mathbf{x}_m = \begin{bmatrix} 67.6 & -67.4 \\ 188.6 & -189.4 \\ 303.1 & -204.2 \\ 217.0 & -216.3 \\ 293.1 & -291.2 \\ 242.2 & -239.8 \\ 104.6 & -102.3 \\ 28.8 & -28.4 \end{bmatrix} \mathbf{u} + \begin{bmatrix} 44.1 \\ 119.7 \\ 192.9 \\ 154.4 \\ 198.1 \\ 165.6 \\ 76.4 \\ 45.7 \end{bmatrix} \mathbf{d} + \mathbf{n}_x$$

In the second part, all stage temperatures were used, and the linear model is shown in Appendix B. The secondary variables, L_D and stage temperature

Table 6.1: Summary of the four scenarios

scenario	*expression for* $\mathbf{u}$	*variable distribution*
S1 ($\mathbf{u}$)	$\mathbf{u}$	$\mathbf{u} \sim \mathcal{N}(0, 0.005^2 I_2)$
S2 ($\mathbf{y} = \mathbf{y}_s$)	$\mathbf{G}_y^{-1}\left(\mathbf{y}_s - \mathbf{G}_y^d \mathbf{d}\right)$	$\mathbf{y}_s \sim \mathcal{N}(0, 0.005^2 I_2)$
S3 ($\mathbf{z} = \mathbf{z}_s$)	$\mathbf{G}_z^{-1}\left(\mathbf{z}_s - \mathbf{G}_z^d \mathbf{d}\right)$	$\mathbf{z}_s \sim \mathcal{N}(0, [0.05^2\ 0.5^2] I_2)$
S4 ($\hat{\mathbf{y}} = \mathbf{y}_s$)	$(\mathbf{H}\mathbf{G}_x)^{-1}\left[\mathbf{H}\left(\mathbf{G}_x^d \mathbf{d} + \mathbf{n}_x\right) + \mathbf{y}_s\right]$	$\mathbf{y}_s \sim \mathcal{N}(0, 0.005^2 I_2)$

all scenarios $\mathbf{d} \sim \mathcal{N}(0, 0.05^2 I_2)$ and $\mathbf{n}_y \sim \mathcal{N}(0, 0.2^2 I_8)$

T_{25}, is given by

$$\mathbf{z} = \begin{bmatrix} 1 & 0 \\ 293.1 & -291.2 \end{bmatrix} \mathbf{u} + \begin{bmatrix} 0 \\ 298.0 \end{bmatrix} \mathbf{d}$$

Nominal operation of the system was decided to have a variation in the product specification (primary variable) equal to $\mathrm{std} \approx 0.005$. The linear system was arranged into the four different control scenarios described previously, and are summarized in Table 6.1.

The variation in input variables for scenario 2, 3 and 4 were chosen such that the resulting standard deviation in the primary variable $\mathbf{y}$ was similar for the three systems ($\mathrm{std} \approx 0.005$).

Because scenario 1 has no control, the primary variable is very sensitive to disturbances which gives rise to large standard deviations in $\mathbf{y}$ ($\mathrm{std} \approx 0.08$), and the product specification cannot be met. The standard deviation in $\mathbf{u}$ ($\mathrm{std} \approx 0.08$) was selected to give a small standard deviation in $\mathbf{y}$, but also sufficiently large to give noticeable contribution to the variation ($\mathrm{std} \approx 0.10$).

6.1.2 Testing Procedure

Table 6.2 gives an overview of the testing procedure. For data based estimators, this procedure had three dimensions (calibration, regression method and validation). Estimators were trained on each of the three scenarios with each of the four regression methods. In all, 12 estimators were trained, which in turn were validated on the four scenarios. For the model based estimators,

Table 6.2: Overview of the testing procedure

	dimension 1	*dimension 2*		*dimension 3*
	calibration	method	estimator	validation
model		optimal S1	$\mathbf{H}_1$	S1
		optimal S2	$\mathbf{H}_2$	S2
		optimal S3	$\mathbf{H}_3$	S3
		optimal S4	$\mathbf{H}_4$	S4
data	S1	LS	$\mathbf{B}_{\mathrm{ls},i}$	S1
	S2	PCR	$\mathbf{B}_{\mathrm{pcr},i}$	S2
	S3	PLS	$\mathbf{B}_{\mathrm{pls},i}$	S3
		CLM	$\mathbf{B}_{\mathrm{clr},i}$	S4

explanation: S1($\mathbf{u}$), S2($\mathbf{y} = \mathbf{y}_s$), S3($\mathbf{z} = \mathbf{z}_s$) and S4($\hat{\mathbf{y}} = \mathbf{y}_s$)

the testing procedure had only two dimensions. The estimators were found from the model using the analytical expressions given in Chapter 3 with scaling equal to the distribution of the input variable, and validated against the four scenarios.

It was found through trial and error, that 4 principal components gave the best performance for PCR, and 3 components gave the best performance for PLS and truncated CLM.

6.1.3 Calibration and Validation Data

For the data based estimators, calibration data was generated by drawing 32 random values for $\mathbf{u}$,$\mathbf{d}$, $\mathbf{n}_x$, $\mathbf{y}_s$, and $\mathbf{z}_s$ with the distributions given in Table 6.1, and calculating the corresponding output variables $\mathbf{x}_m$ and $\mathbf{y}$ for the respective scenarios (except scenario 4). This gave one set of calibration data with 32 samples: $\mathbf{X}$ (8×32) and $\mathbf{Y}$ (2×32).

Validation data, was created in the same way as the calibration data. Because of the implicit nature of scenario 4, validation data was generated after the estimators were found and used in "closed-loop".

6.1.4 Evaluation Criteria

The square norm of the prediction error was used as a measure of the estimators performance, that is

Table 6.3: The mean prediction error of the model based estimators when applied to the four operation scenarios.

	validation data			
	S1	S2	S3	S4
$\mathbf{H}_1$	0.0085	0.2749	0.0215	0.0506
$\mathbf{H}_2$	0.0591	0.0093	0.0104	0.0104
$\mathbf{H}_3$	0.0599	0.0166	0.0098	0.0132
$\mathbf{H}_4$	0.0099	0.0099	0.0099	0.0099

explanation: S1($\mathbf{u}$), S2($\mathbf{y} = \mathbf{y}_s$), S3($\mathbf{z} = \mathbf{z}_s$) and S4($\hat{\mathbf{y}} = \mathbf{y}_s$)

$$e = \left\| \mathbf{Y} - \widehat{\mathbf{Y}} \right\|_F$$

were $\mathbf{Y}$ is the primary variable from the validation data, and $\hat{\mathbf{Y}}$ is the predicted primary variable. To average out the random nature of the system, the testing procedure was repeated 150 times, each time with a new set of calibration and validation data. At first, the mean prediction error of the 150 times was used as a measurement of the estimators prediction performance, but because of random noise and variation of input variables, some data sets had severe deteriorating effect on the performance. This affected the mean prediction error by giving an unstable and distorted picture of the performance. By taking the median of the prediction error for the 150 times, the sets with severe deterioration in performance are excluded as only the middle prediction error is used. This gives thus a more stable and more "accurate" picture of the estimators performance.

6.2 Results

Model Based Estimators Resulting median prediction errors for the model based estimators are given in Table 6.3. As expected, the optimal estimators were best when applied to its intended scenario, and can be seen by a minimum prediction error along the diagonal.

The optimal "closed-loop" estimator $\mathbf{H}_4$ had a constant prediction error when applied to all four scenarios, and is probably due to the constraint $\mathbf{G}_y = \mathbf{H}\mathbf{G}_x$. The three optimal "open-loop" estimators $\mathbf{H}_{1-3}$ does not have this constraint

and $\mathbf{G}_y \neq \mathbf{H}\mathbf{G}_x$. The prediction ability will therefor change from one scenario to another. It should be mentioned that the three "open-loop" estimators can come very close to having $\mathbf{G}_y = \mathbf{H}\mathbf{G}_x$, but this depends heavily on the scaling (which in turn depends on the operation of the system). If expected disturbances are large compared to expected changes in $\mathbf{u}$ (as in this case), then the estimators will prioritize the prediction of disturbances by approaching $\mathbf{G}_y^d = \mathbf{H}\mathbf{G}_x^d$.

When using "closed-loop" estimators (scenario 4), the optimal "closed-loop" estimator gave the best performance. This was expected, as the optimal "closed-loop" estimator derived to be optimal for linear systems with the given distribution. The optimal "open-loop" estimator $\mathbf{H}_2$ gave also good performance when used in "closed-loop", and it is expected that data based estimators trained on scenario 2 will give good performance as "closed-loop" estimators. Scenario 2 is usually not feasible or is irrelevant in real life. This because we would normally use estimators to predict primary variables that we cannot measure (measurements that are too slow, too expensive or not possible). If the primary variables could be measured, we would use this measurement instead of the predicted variable. However, scenario 2 is still interesting as this is the most common scenario for finding estimators from computer simulations when working with complex models.

Scenario 3 is the most used arrangement in the industry, where a secondary variable is controlled instead of the more desired primary variable. The optimal "open-loop" estimator for scenario 3 gave fair performance when used in "closed-loop" (though this is dependent on the operation of the system).

Data Based Estimators The performance of the data based estimators are shown in Table 6.4. The data based estimators performed best for the scenarios they where trained on, and estimators trained on scenario 2 gave the best "closed-loop" performance. This is reasonable because scenario 2 and 4 have similar control structures, and the two scenarios will have similar internal variation in the column. The directions in the calibration data generated from scenario 2 will then be similar to validation data generated for scenario 4 (given a good estimator).

Figure 6.1 shows the largest left singular vector for the calibration data for the first three scenarios, and for the validation data for scenario 4 (given all column temperatures). As expected scenario 2 and 4 were almost identical, while scenario 3 was similar for the lower part of the column. Scenario 1

Table 6.4: Performance of Data Based Estimator

			validation data			
			S1	S2	S3	S4
calibration data	S1	$\mathbf{B}_{\mathrm{ls},1}$	0.0100	0.2788	0.0223	0.0523
	S2	$\mathbf{B}_{\mathrm{ls},2}$	0.0788	0.0108	0.0116	0.0119
	S3	$\mathbf{B}_{\mathrm{ls},3}$	0.1053	0.0346	0.0113	0.0265
	S1	$\mathbf{B}_{\mathrm{pcr},1}$	0.0092	0.3339	0.0257	0.0534
	S2	$\mathbf{B}_{\mathrm{pcr},2}$	0.0812	0.0101	0.0112	0.0112
	S3	$\mathbf{B}_{\mathrm{pcr},3}$	0.1064	0.0315	0.0105	0.0241
	S1	$\mathbf{B}_{\mathrm{pls},1}$	0.0092	0.2935	0.0232	0.0552
	S2	$\mathbf{B}_{\mathrm{pls},2}$	0.0813	0.0101	0.0111	0.0112
	S3	$\mathbf{B}_{\mathrm{pls},3}$	0.1099	0.0333	0.0105	0.0239
	S1	$\mathbf{B}_{\mathrm{clr},1}$	0.0103	0.2900	0.0233	0.0516
	S2	$\mathbf{B}_{\mathrm{clr},2}$	0.0641	0.0113	0.0115	0.0115
	S3	$\mathbf{B}_{\mathrm{clr},3}$	0.0872	0.0368	0.0118	0.0266
	S1	$\mathbf{B}^{\dagger}_{\mathrm{clr},1}$	0.0094	0.3863	0.0291	0.0530
	S2	$\mathbf{B}^{\dagger}_{\mathrm{clr},2}$	0.0552	0.0105	0.0107	0.0107
	S3	$\mathbf{B}^{\dagger}_{\mathrm{clr},3}$	0.0800	0.0359	0.0107	0.0244

explanation: S1($\mathbf{u}$), S2($\mathbf{y} = \mathbf{y}_s$), S3($\mathbf{z} = \mathbf{z}_s$) and S4($\hat{\mathbf{y}} = \mathbf{y}_s$)

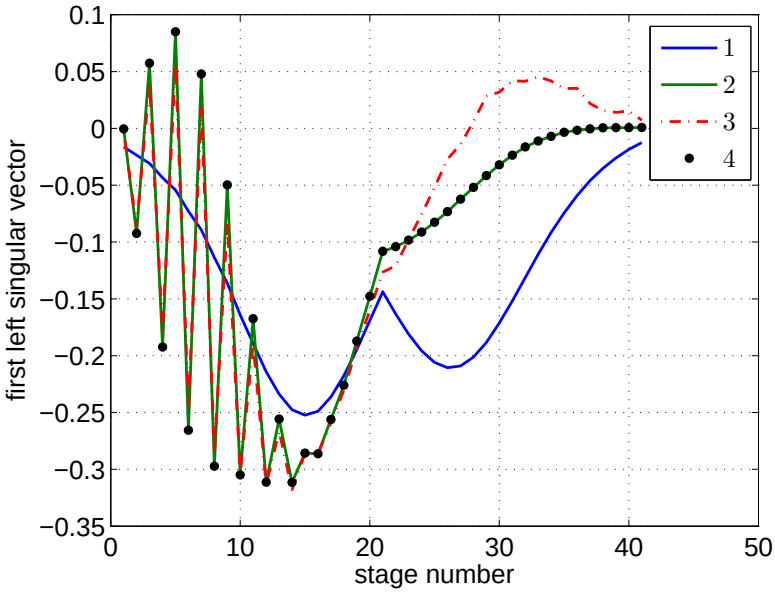


Figure 6.1: The first vector of the left singular matrix of $\mathbf{X}$ calibration data for Scenario 1–3 , and $\mathbf{X}$ validation data for Scenario 4.

differed most from the others, and explains why estimators trained on scenario 1 performed poorly on other control scenarios.

For estimators trained on scenario 1–3, there is little consistency in the relative performance between LS, PCR, PLS and CLM, and there is no clear overall advantageous method. For estimators trained and validated on the same scenario, that is the diagonal elements, PCR and PLS gave slightly better results. CLM had lower performance than PCR and PLS, and is assumed to be caused by assimilation of noise. Truncated CLM filters out this noise resulting in the improved performance. When trained on scenario 2, the truncated CLM gave the best "closed-loop" estimator (scenario 4).

When comparing the model based estimator in Table 6.3 with their data based counterparts in Table 6.4, we see that the model based estimators gave better results. This is reasonable because, as proven in Chapter 3, for a linear system with an estimator on the form $\mathbf{y} = \mathbf{H}\mathbf{x}_m$, where the variation in the variables follows a normal distribution, the model based estimators are the optimal estimators for their respective system. Thus for our linear system, the data based method can at best give the same performance as their model based

counterparts. Therefore, given enough variation in the calibration data, it is expected that estimators trained on scenario 1 ($\mathbf{B}_1$) will approach the optimal "open-loop" estimator $\mathbf{H}_1$. Similarly it is expected that $\mathbf{B}_2 \rightarrow \mathbf{H}_2$ and $\mathbf{B}_3 \rightarrow \mathbf{H}_3$.

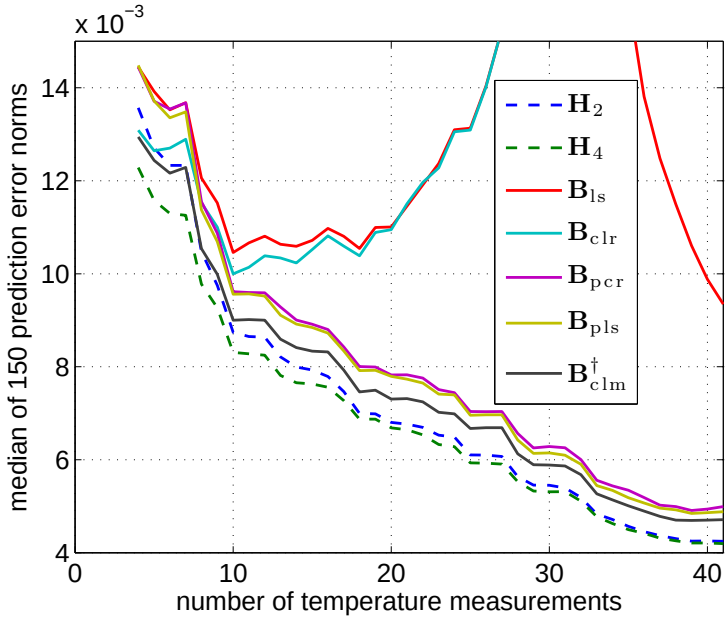
Because the CLM uses the analytical expression for the optimal "closed-loop" estimator, it is expected that $\mathbf{B}_{\text{clr}} \rightarrow \mathbf{H}_4$. As seen in Table 6.4, the truncated CLM trained on scenario 2, $\mathbf{B}^{\dagger}_{\text{clr},2}$, comes close to this. Notice also that the performance is nearly constant for the scenarios 2–4, and the performance for scenario 1 is greatly improved compared to the other methods. The untruncated CLM gave similar results when trained on scenario 2, but with slightly reduced performance.

6.2.1 Performance of Estimators in "Closed-Loop"

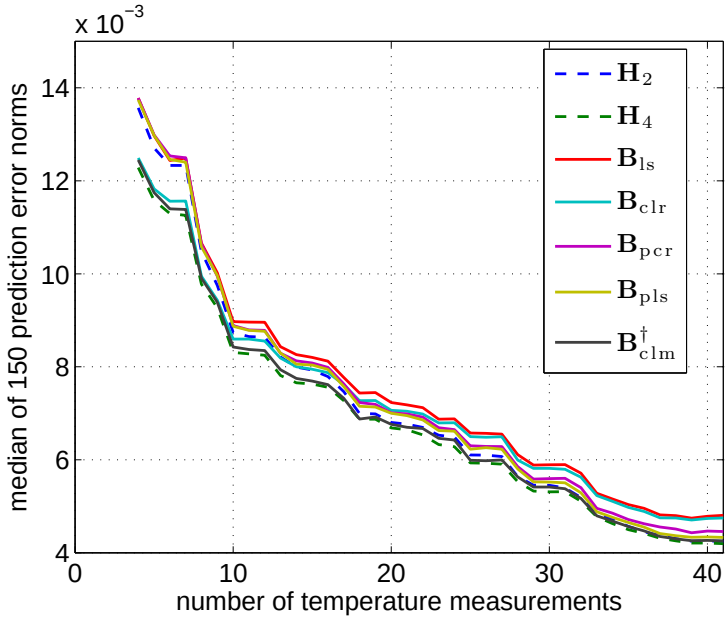
Figure 6.2a, shows estimators "closed-loop" performance for a range of used stage temperatures (number of measurements increased from 8 to 41, see Appendix B for model). All data estimators were trained on calibration sets generated from scenario 2, and model based estimators were found from the linear model. The estimators were validated on scenario 4. Calibration and validation data with each 32 samples were generated, and the test was repeated 300 times. The median prediction error for the 300 times was used as a measurement of estimator performance. The order of the temperature measurements can be seen in Table 7.2 on page 53.

$\mathbf{H}_4$ gave the best performance for all measurements. $\mathbf{H}_2$ gave poor performance for few measurements, but approached $\mathbf{H}_4$ when the number of measurements increase. PLS and PCR gave very similar performance.

Because the calibration sets had only 32 samples, the equations system went from overdetermined to underdetermined when the number of measurements increased. For the overdetermined systems (4-31 temperature measurements), LS had similar performance to PCR and PLS, but deteriorated rapidly as the system became more determined. For determined and underdetermined systems (32-41 temperature measurements) PCR and PLS gave much better performance than LS. The reason for this is the existence of collinearity and noise in the data set. When the system is overdetermined, the collinearity and noise will be filtered out as only the strongest directions in the data will be used. When the system is close to determined, LS is forced to use the weaker



(a) Median prediction error for 150 datasets with 32 samples.



(b) Median prediction error for 150 datasets with 200 samples.

Figure 6.2: "closed-loop" performance for estimators trained on scenario 2, and for the optimal "open-loop" and "closed-loop" estimator $\mathbf{H}_2$ and $\mathbf{H}_4$.

directions. Thus assimilating much/all of the collinearity and noise, resulting in a deterioration of the performance. PCR and PLS uses only the strongest directions (equal to the number of principal components), thus reducing the assimilation of collinearity and noise, giving better results.

The CLM estimator $\mathbf{B}_{\text{clm}}$, gave good results when the system was very overdetermined (4-10 temperature measurements), but there was a fast deterioration in the performance when the system became more determined and the performance converged with the LS estimator. As the case with LS estimator, the deterioration in performance for the overdetermined system is assumed to be caused by assimilation of collinearity and noise.

When the system became underdetermined, the $\mathbf{B}_{\text{clm}}$ performance becomes unstable. The unstable performance for the underdetermined system is due to the mathematical constraint imposed by $(\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T)^{-1}$ in the analytical expression (3.12) for the optimal "closed-loop" estimator. When the system is underdetermined, the inverse of $\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T$ becomes singular and the inverse does not exists. However, because of numerical error in Matlab the inverse was still found (becoming the inverse of very ill-conditioned matrix), which results in the unstable and poor performance.

The truncated CLM estimator $\mathbf{B}_{\text{clm}}^{\dagger}$ gave improved performance for all measurements, resulting in the best "closed-loop" estimator. This is accredited to filtration of waker directions in the calibration data, and to the truncated pseudoinverse of $\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^T$.

As mentioned in the previous section, the model based estimators, when applied to the exact model, represents the optimal estimator for their respective scenarios. Thus data estimators can at best become as good their model based counterparts. To show this, the number of samples in a given set were increased to 200, and the results are shown in Figure 6.2b. It can be clearly seen that as the data methods get enough data, they do approach their model based counterparts. It is also clearly seen that LS, PCR and PLS cannot perform better than the scenarios they are trained on. Given enough data, the truncated CLM estimator $\mathbf{B}_{\text{clm}}^{\dagger}$ approaches $\mathbf{H}_4$, becoming close-to optimal for "closed-loop" estimation.

6.2.2 Summary

- The optimal "open-loop" and "closed-loop" estimators gave the best performance on their intended scenario.

- "Closed-loop" estimators (that is, estimators validated on scenario 4), had the best performance when trained on Scenario 2.
- The truncated CLM estimators had the best "closed-loop" performance, while the PLS and PCR estimators had the best "open-loop" performance.
- The PLS and PCR estimators approached the optimal "closed-loop" estimator for higher number of temperature measurements (independent variables).
- The truncated CLM estimator approached the optimal "closed-loop" estimator for few temperature measurements (independent variables), given that there is sufficient variation in the calibration data.
- The truncated CLM estimator performed better than CLM estimator.

Chapter 7

Nonlinear System

7.1 Problem Definition

Given a range of possible temperature measurements, find the best estimators when used in "open-loop" (scenario 2) and "closed-loop" (scenario 4) on a nonlinear model. Only PCR, PLS, and truncated CLM estimators will be tested, as initial tests showed that the other methods gave poor performance.

Two cases will be considered: (1) binary feed mixture and (2) multicomponent feed mixture composed of one heavy and one light nonkey component in addition the key components in the binary mixture. Physical data and nominal operation of the column was set equal to the specifications used in Mejdell and Skogestad (1991), and are given in Table 7.1.

The multicomponent mixture consists of four components were the objective is, as for the binary distillation, to separate the two key components. This separation can be quantified with the pseudobinary mole fraction,

$$x' = \frac{x_L}{x_L + x_H}$$

where x' is the pseudobinary mole fraction of the light key component, and x_L and x_H are actual mole fraction for the light and heavy key components respectively. For both binary and multicomponent distillation, nominal operation of the column gives $x'_D = 0.99$ and $x'_B = 0.01$.

Table 7.1: Data for Distillation Column

binary mixture

		nominal operation ($F = 1$)				
i	compt	z_F	x_D	x_B	$\alpha_{i,i+1}$	$T^b\,[K]$
1	LK	0.500	0.990	0.010	1.5	341.9
2	HK	0.500	0.010	0.990		355.4

multicomponent mixture

		nominal operation ($F = 1$)				
i	compt	z_F	x_D	x_B	$\alpha_{i,i+1}$	$T^b\,[K]$
1	LNK	0.050	0.125	0.000	2.0	321.4
2	LK	0.350	0.866	0.006	1.5	341.9
3	HK	0.350	0.009	0.577	2.0	355.4
4	HNK	0.250	0.000	0.417		381.6

7.1.1 Calibration Set

The calibration set used in this thesis was generated with the same specifications used in Mejdell and Skogestad (1991). The calibration set consists of 32 different simulated runs where z_F, x_B and x_D were specified as shown in Table 7.2, and the temperature profiles recorded. The first run is the nominal operation point, and the next 15 runs were randomly chosen. The last 16 runs were selected by a 2-composite design in four levels. Normal distributed random noise with a standard deviation of 0.2 was added to the temperature profiles. For the multicomponent case, the pseudobinary composition were the same as in the binary case. In addition, the feed composition of the light nonkey component was randomly varied between 0.025 and 0.075, and the heavy nonkey between 0.15 and 0.35. Noise was added to the temperature profiles in the same way as the binary case.

7.1.2 Testing Procedure

Estimators were trained, and estimator performance was measured when using only four temperature measurements. This was repeated, but with one additional temperature measurement, until all column temperatures were used. The order of the temperature measurements are shown in Table 7.3. The first four temperatures were chosen to give a cross-section of the column, the rest

Table 7.2: Specifications used in simulations to obtain column temperature profiles

run	z_F	x_D	x_B
1	0.5000	0.9900	0.0100
2	0.5375	0.9913	0.0262
3	0.4250	0.9738	0.0151
4	0.5250	0.9700	0.0132
5	0.4125	0.9801	0.0058
6	0.6000	0.9849	0.0044
7	0.5125	0.9942	0.0066
8	0.5500	0.9827	0.0076
9	0.4875	0.9962	0.0189
10	0.4750	0.9956	0.0087
11	0.5625	0.9934	0.0115
12	0.4625	0.9772	0.0300
13	0.4375	0.9950	0.0038
14	0.4500	0.9924	0.0173
15	0.5750	0.9868	0.0228
16	0.5875	0.9885	0.0050
17	0.4000	0.9700	0.0300
18	0.4000	0.9700	0.0033
19	0.4000	0.9967	0.0300
20	0.4000	0.9967	0.0033
21	0.6000	0.9700	0.0300
22	0.6000	0.9700	0.0033
23	0.6000	0.9967	0.0300
24	0.6000	0.9967	0.0033
25	0.4500	0.9827	0.0173
26	0.4500	0.9827	0.0058
27	0.4500	0.9942	0.0173
28	0.4500	0.9942	0.0058
29	0.5500	0.9827	0.0173
30	0.5000	0.9827	0.0058
31	0.5500	0.9942	0.0173
32	0.5500	0.9942	0.0058

Table 7.3: Order of available stage temperatures T_n

n	25,15,5,35,20,22,17,27,34,4,14,24,3,37,19,23,6,36,16,26,2,38,21,18, 28,12,39,9,29,40,10,30,7,11,31,8,32,13,33,41 and 1

were chosen in a semi-random order. That is, random, but a certain spread in the measurements were ensured.

7.1.3 Number of Principal Components

To estimate the optimum number of principal components to use in the regression, the calibration set was cross-validated as follows. From one principal component to the maximum number of principal components: The 32 runs were divided into seven groups, were six of the groups were used for calibration and the last group used for validation. This was repeated seven times, but rotating the groups such that all groups had been used for testing once. For each cross-validation, the mean squared error of prediction (MSEP) was calculated by

$$\mathrm{MSEP}\,(l) = {}^{1}\!/_{32} \sum_{i=1}^{32} (y_i - \hat{y}_i\,(l))^2$$

where l is the number of principal components, and y_i is x_D or x_B for the ith run. Because we are estimating both top and bottom compositions, the MSEP for both compositions were summed and used as the performance for the respective number principal components. The number of principal components giving the lowest summed MSEP was used in the regression, and the specific number of principal components for each run are shown in Appendix C.

7.1.4 Evaluation

For the binary case, each estimator was tested when used in "open-loop" and "closed-loop" with dynamic simulations. Normal distributed random noise was added each 0.1 time unit to the temperature profile with standard deviation of 0.2. Disturbances in composition and changes in setpoints were changed according to the calibration data in Table 7.2. A typical response can be seen in Figure 7.1. For each run, the median prediction error was taken of the last

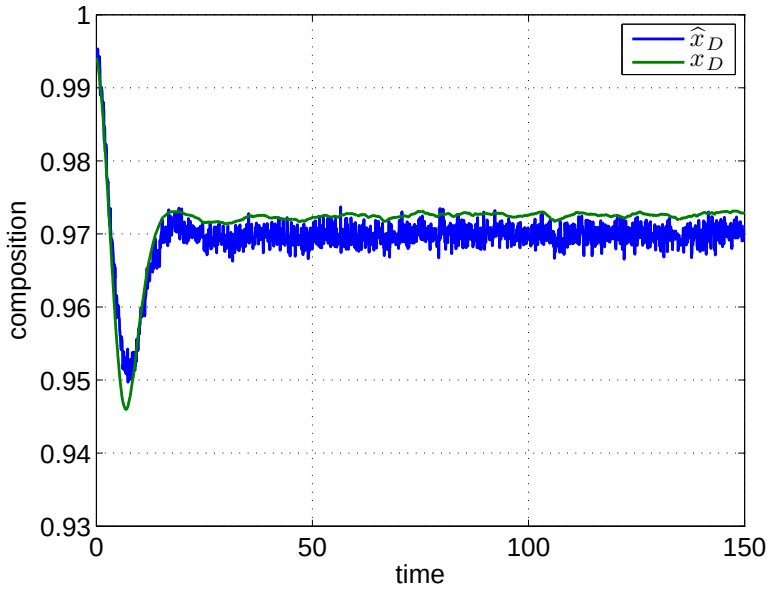


Figure 7.1: The dynamic response of the primary variable for run 18 in the calibration data for binary distillation.

50 time units and the summed MSEP was calculated for the 32 runs.

Given the time consuming nature of dynamic simulations, the multicomponent case was tested by solving the control problem as an constrained optimization problem. Further, noise was not added to the temperatures column temperatures. This is assumed to have little effect on the relative performance of each estimator. Because of steady-state offset in the estimators predictions, certain runs in Table 7.2 were not achievable (runs: 18 an20). For these runs, the pseudobinary compositions were set as close to the unachievable compositions as possible. For the 32 runs, the summed MSEP was calculated.

7.2 Results

The estimator's "open-loop" and "closed-loop" performance are shown in figure 7.2 for binary mixture, and figure 7.3 for multicomponent mixture. The results will be discussed further below.

7.2.1 Binary Mixture

Seen in Figure 7.2a, PLS gives good performance for "open-loop" prediction, and is only surpassed by PCR and CLM for higher number of temperature measurements. Noise added to calibration have corrupted the smaller PLS components, resulting in a lower number of principal components used in the regression. Figure 7.2b shows the estimators "closed-loop" performance. PLS had the worst performance, and CLM had the best.

For both "open-loop" and "closed-loop" predictions, CLM and PCR had very similar performance, but for "open-loop" cases PCR was better, and for "close-loop" cases CLM was better. Given that these two methods uses the same pre-treatment it is reasonable that they will give similar results.

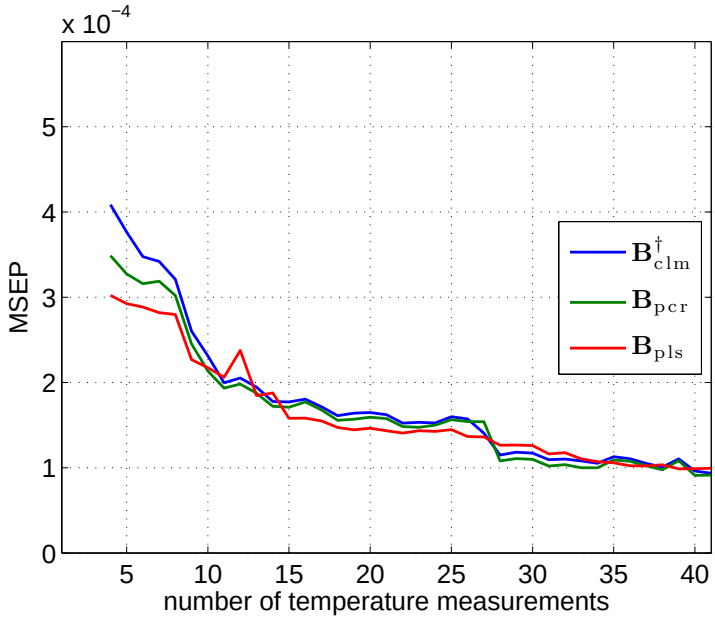
7.2.2 Multicomponent Mixture

Figure 7.3a shows the estimators "open-loop" performance, and we see the results are similar to the binary case. PLS gave the best performance for most temperatures, and the performance of PCR and CLM was similar to each other.

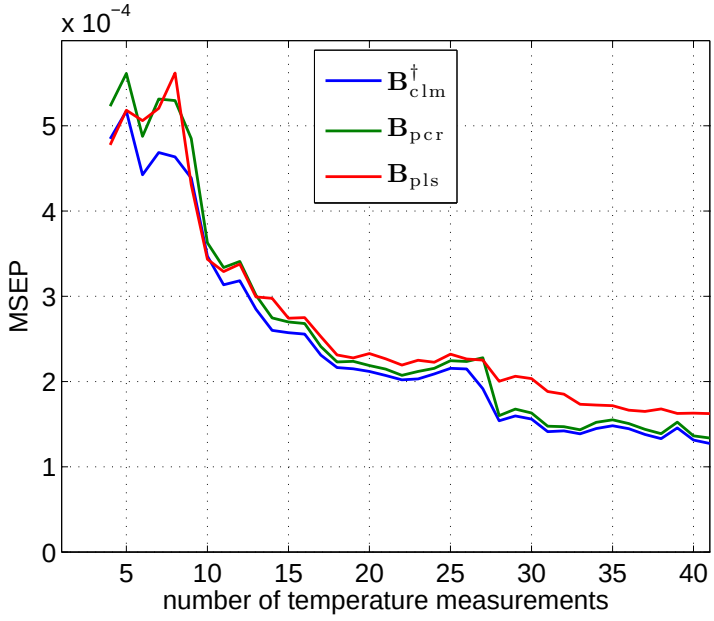
The estimators "closed-loop" performance are shown in Figure 7.3b. CLM had the best performance for all measurements. PCR had worse performance for few temperature measurements, but the performance approached CLM for higher number of measurements. PLS had the worst performance for most number of measurements.

The poor performance of PLS and PCR are mainly due to runs 12, 17, and 19 were the estimators have twice the prediction error compared to CLM. The estimators also have large prediction error for these runs on binary distillations, but on the same level as CLM. The runs have large perturbations in feed, top, and bottom compositions, and it is expected that the estimators would have larger prediction errors than other runs with smaller perturbations. For multicomponent distillation, these perturbation is amplified by the presents of nonkey components. If too few principal components are used in the regression, the estimators would have trouble predicting the primary variables when subjected to large disturbances in nonkey compositions.

As all estimators used the estimated optimal number of principal components found from cross-validation, the poor performance for PLS might be caused by

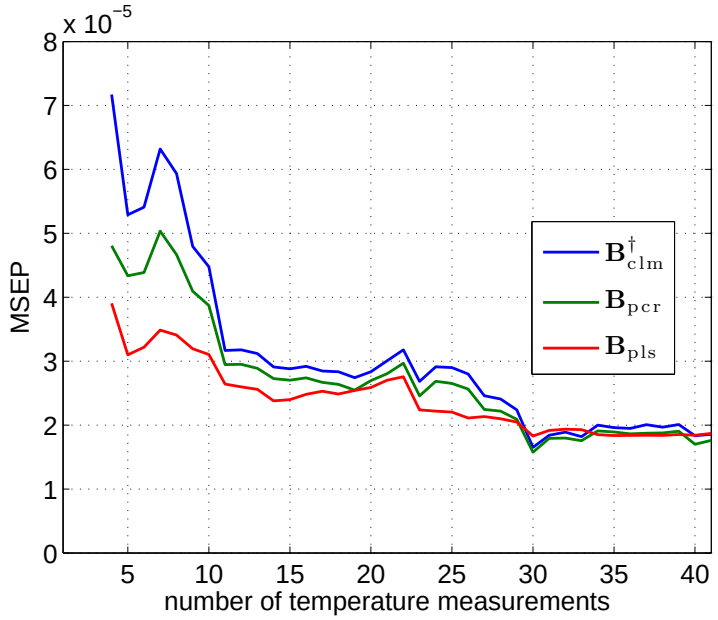


(a) "Open-loop" prediction

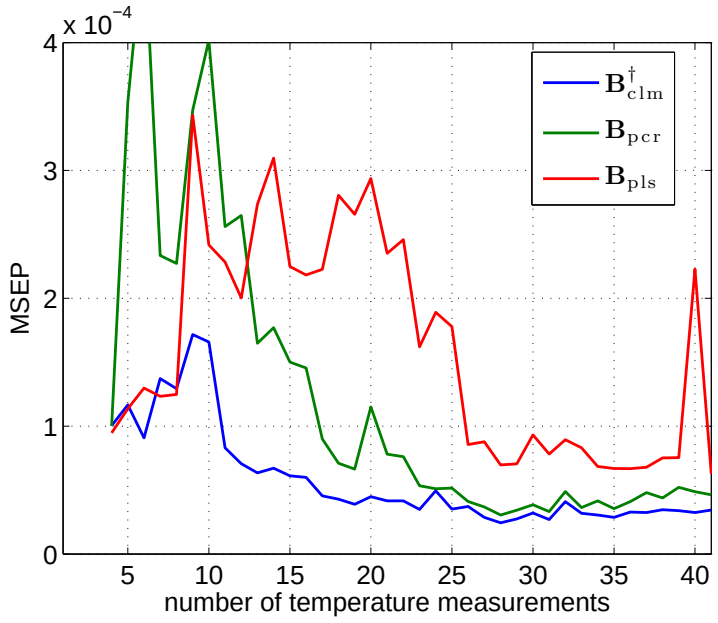


(b) "Closed-loop" prediction

Figure 7.2: Estimator performance binary distillation



(a) "Open-loop" prediction



(b) "Closed-loop" prediction

Figure 7.3: Estimator performance for multicomponent distillation

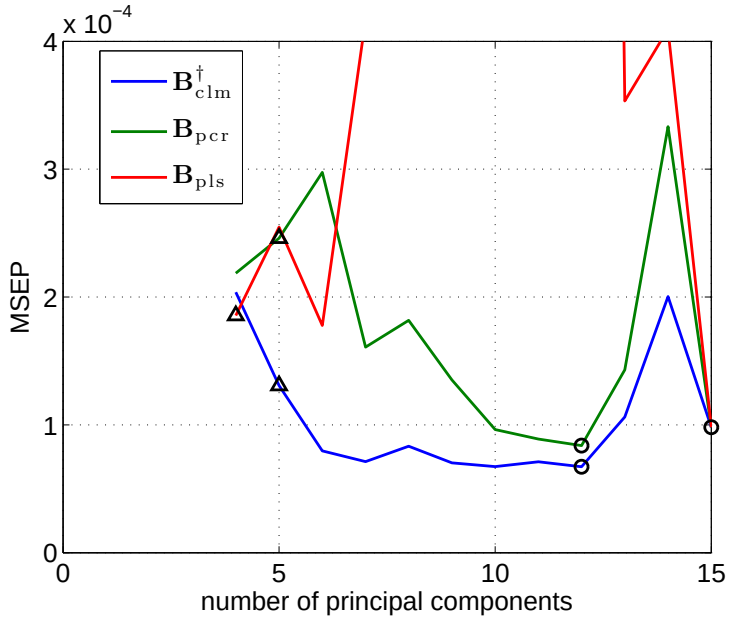
under estimating the number of components. In the cross-validation process, the estimators are trained on Scenario 2 and validated on Scenario 2. The estimated optimal number of components, is in fact for "open-loop" estimation. We see from Figure 7.3a, that cross-validation gives good "open-loop" performance for the given estimators.

For cross-validating "closed-loop" estimators, it would be more correct to validate the performance on Scenario 4. This would be possible when working on computer models, but would also be very be time consuming.

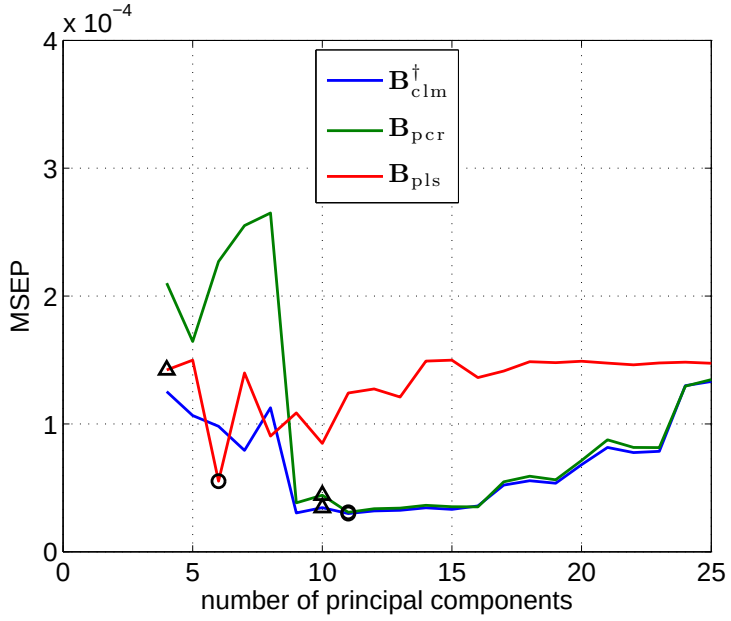
Figure 7.4 shows the estimators "closed-loop" performance for 15 and 25 temperature measurements as a function of the number of principal components. To avoid over fitting, the column temperatures were added random normal distributed noise of magnitude 0.2 to each run. As seen from the figure, the performance can be improved for the three methods. PLS have the best performance with six principal components while PCR and CLM have the best performance with about 10 components. This indicated that the cross-validation process underestimated the number of principal components needed in the PLS regression and the performance can be improved. It is also seen that even though the performance of PLS can be improved, it will not outperform the truncated CLM with the given noise and disturbances.

7.2.3 Summary

- CLM estimators had generally the best "closed-loop" performance for both binary and multicomponent distillation.
- PLS estimators had generally best "open-loop" performance for both binary and multicomponent distillation.
- Cross-validation on Scenario 2 works well for "open-loop" estimators, but works only to a lesser degree for "closed-loop" estimators.



(a) 15 temperature measurements



(b) 25 temperature measurements

Figure 7.4: Estimators "closed-loop" performance (multicomponent distillation) for given number of temperature measurements as a function of the number of principal components. (△) is the estimated optimum number of components from cross-validation. (◯) is the optimum number of component for the given noise.

Chapter 8

Discussion

8.1 Optimal "Open-Loop" Estimators

Shown previously, the three optimal "open-loop" estimators have the same solution as LS, but uses model gains and scaling matrices instead of data. The model gains with scaling can be thought of as having data with only the relevant directions. From this we can deduce that, for linear systems, an optimal estimator can be found from data with an LS method (given that the calibration data contain the same variation as expected from nominal operation). Methods like PCR and PLS, which can be thought of as LS with different preprocessing of the data before regression, will also give an optimal estimator. PCR and PLS, will have an advantage when the calibration data contains collinearity and noise.

8.2 Optimal "Closed-Loop" Estimator

Seen from the expression for the optimal "closed-loop" estimator, the optimization problem does not have the same form as LS (Though, it can be claimed that the optimization problem is a weighted LS problem). Thus from data, we would expect that we cannot find the optimal "closed-loop" estimator with conventional techniques like LS, PCR and PLS. For conditions were expected variation in setpoints are large compared to disturbances, the

optimal "open-loop" estimator with closed-loop primary variable $\mathbf{y}$, will give approximately the same estimator as the optimal "closed-loop" estimator. The optimal "open-loop" estimators for closed-loop $\mathbf{y}$ is

$$\mathbf{H}_2 = \arg\min_{\mathbf{H}} \left\| \; (\mathbf{I} - \mathbf{H}\mathbf{G}_x^{cl})\mathbf{W}_{ys} \quad (-\mathbf{H}\mathbf{F}\mathbf{W}_d) \quad (-\mathbf{H}\mathbf{W}_{n_x}) \; \right\|_F$$

Relatively large variations in setpoint gives ($\mathbf{I} \approx \mathbf{H}\mathbf{G}_x^{cl}$) and the optimization problem can be approximated to

$$\mathbf{H}_2 = \arg\min_{\mathbf{H}} \left\| \mathbf{H} \begin{bmatrix} \mathbf{F}\mathbf{W}_d & \mathbf{W}_{n_x} \end{bmatrix} \right\|_F$$

$$\text{s.t.} \quad \mathbf{H}\mathbf{G}_x^{cl} = \mathbf{I}$$

which is the same as the optimization problem for the optimal "closed-loop" estimator. This can also be shown for the two other optimal "open-loop" estimators. This is not immediately obvious, but it is easily shown when recognizing that, with the given constraint, $\mathbf{F}$ reduces to $\mathbf{G}_y^d - \mathbf{H}\mathbf{G}_x^d$. Thus, for some conditions the "open-loop" estimators are also optimal "closed-loop" estimators.

For linear systems with no constraints on the product compositions (that is, setpoint changes can be set arbitrarily large), all the "open-loop" estimators could also be an optimal "closed-loop" estimator. If there are constraints on the product compositions or the system is nonlinear, this would not be the case, and the optimal "open-loop" estimators would probably not result in an optimal "closed-loop" estimator.

8.2.1 Further Improvements of CLM

We have seen throughout this thesis, that the truncated CLM estimators have correlated performance with the PCR estimators. And as mentioned before, this is thought to be an effect of the principal component analysis, which both methods have in common. This belief is amplified further when comparing the performance of the untruncated CLM estimator with the LS estimator, and notice a similar correlation between the performance (Figure 6.2a on page 48). This leads to the conclusion that the untruncated CLM improves the "closed-loop" performance of LS, and that the truncated CLM improves the "closed-loop" performance of PCR. A natural extension to this would be to use PLS in combination with the CLM, and improve the "closed-loop" performance of PLS. We have tried several novel approaches to this by using the PLS scores and loading matrices, but non have been successful.

8.2.2 Weighting and Logarithmic Transformations

The temperatures in the column ends contain the most representative information about the product compositions, but are also affected most by nonlinearities and noise. Top and bottom temperatures vary little with disturbances and product compositions compared to other temperatures. Mejdell and Skogestad (1991) reported that weighting the variables such that the information in the column ends are not lost, gave better results. They also reported that logarithmic transformations of the temperatures gave very good results, but also that it was imperative that the transformed data was weighted to reduce the noise sensitivity. Due to limited time, weighting and logarithmic transformations of the temperatures were not tested, but it is believed that this will also improve the performance of the truncated CLM.

Chapter 9

Conclusion

9.1 The Optimal Model Based Estimators

The new static estimator proposed by Skogestad et al. (2011) was found to be optimal for "closed-loop" estimation, and was termed the optimal "closed-loop" estimator. When using the optimal "open-loop" estimators (especially $\mathbf{H}_2$) in "closed-loop" on the linear distillation model, we found that their performance approached the optimal "closed-loop" performance when using additional temperature measurements.

9.2 The "Closed-Loop" Method

CLM, which is a data based extension to the optimal "closed-loop" estimator, was found to approach the optimal "closed-loop" estimator for few temperature measurements. However, the CLM was found to be sensitive to noise and collinearity. The proposed truncated CLM uses a principal component analysis on the temperature data, filtrating out this collinearity and noise. The truncated CLM was found to have significantly improved performance on noisy data compared to CLM when dealing with closed-to determined calibration data. For the distillation examples, the truncated CLM was found to give the best "closed-loop" performance of the data methods.

CLM has, due to the mathematical constraint imposed by $(\mathbf{F}\mathbf{F}^T)^{-1}$, requiring $\mathbf{F}\mathbf{F}^T$ to be of full rank in the analytical formula (3.12), no solution for underdetermined calibration data. The truncated CLM use the truncated pseudoinverse of $\mathbf{F}\mathbf{F}^T$, circumventing the requirement of full rank. The truncated CLM was found to give good "close-loop" performance for these underdetermined systems.

9.3 PLS and PCR

PLS and PCR were both found to have lower "closed-loop" performance on the distillation examples than the truncated CLM. PCR performance was found to be often very correlated with the truncated CLM, giving in most cases a very similar performance. Of the two methods however, PCR gave the best "open-loop" performance, and truncated CLM gave the best "closed-loop" performance. PLS gave generally the best "open-loop" performance.

9.4 "Closed-Loop" Systems

One inherent problem of "close-loop" systems are their implicit nature, which inhibits us from generating calibration data. Thus, it is necessary to use calibration data from "open-loop" systems to train the estimators. We found from the linear distillation example that estimators trained on calibration data from "open-loop" operation with controlled primary variables $\mathbf{y}$ (Scenario 2) gave the best "closed-loop" performance. Controlling the primary variables $\mathbf{y}$ are usually not feasible in real distillation columns, and it is more common to control certain stage temperatures. That is, "open-loop" operation with controlled secondary variables $\mathbf{z}$ (Scenario 3). We found for our linear distillation example that estimators trained on this control scenario gave fair "closed-loop" performance.

Nomenclature

α	relative volatility
$\mathbf{B}$	data based estimator
B	bottom liquid flow rate
$\mathbf{d}$	disturbances
D	distillate flow rate
$\mathbf{d}'$	normalized disturbance
$\mathbf{e}$	prediction error
$\mathbf{F}$	optimal sensitivity
F	Feed flow rate
$\mathbf{G}_x$,$\mathbf{G}_x^d$	gain matrix from inputs (degree of fredom $\mathbf{u}$ and disturbance $\mathbf{d}$) to measurements $\mathbf{x}$
$\mathbf{G}_y$,$\mathbf{G}_y^d$	gain matrix from inputs (degree of fredom $\mathbf{u}$ and disturbance $\mathbf{d}$) to primary variable $\mathbf{y}$
$\mathbf{G}_z$,$\mathbf{G}_z^d$	gain matrix from inputs (degree of fredom $\mathbf{u}$ and disturbance $\mathbf{d}$) to secondary variable $\mathbf{z}$
$\mathbf{H}$	model based estimator
I_n	identity matrix with n diagonal entries
$\mathbf{K}$	controller
L_D	reflux flow rate

M	stage holdup
Muw	stage holdup under weir
$M_o w$	stage holdup over weir
$\mathbf{n}_x$	measurement noise
$\mathbf{n}'_x$	normalized measurement noise
$\mathbf{\Sigma}$	sigular value matrix
T_n	stage temperature for stage n
$\mathbf{U}$	left sigular vector matrix
$\mathbf{u}$	degres of freedom / manipulated variable
$\mathbf{u}'$	normalized degres of freedom
$\mathbf{V}$	right sigular vector matrix
V_B	vapor boiluop flow rate
$\mathbf{W}_d$	scaling matrix for the disturbance $\mathbf{d}$
$\mathbf{W}_u$	scaling matrix for the degree of freedom $\mathbf{u}$
$\mathbf{W}_{n_x}$	scaling matrix for the mesurement noise $\mathbf{n}_x$
$\mathbf{X}$	dependent variable data matrix
$\mathbf{x}$	measurements, independent variable
x'	pseudocomposition of light component
x_B	mole fraction of light component in bottom product
x_D	mole fraction of light component in top product
$\mathbf{Y}$	independent variable data matrix
$\mathbf{y}$	primary variable, dependent variable
$\mathbf{y}_s$	setpoint for primary variable
$\hat{\mathbf{y}}$	predicted primary variable
z_F	feed mole fraction of light component

Bibliography

Alstad, V. and S. Skogestad (2007). Null space method for selecting optimal measurement combinations as controlled variables. *Ind. Eng. Chem. Res. 46*(846-853).

Alstad, V., S. Skogestad, and E. S. Hori (2009). Optimal measurement combinations as controlled variables. *Journal of Process Control 19*, 138–148.

de Jong, S. (1993). Simpls: an alternative approach simpls: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems 18*, 251–263.

Di Ruscio, D. (2000). A weighted view on the partial least-squares algorithm. *Automatica 36*, 831–350.

Halvorsen, I. J., S. Skogestad, J. C. Morud, and V. Alstad (2003). Optimal selection of controlled variables. *Ind. Eng. Chem. Res. 42*, 3270–3284.

Hori, E. S., S. Skogestad, and V. Alstad (2005). Perfect steady-state indirect control. *Ind. Eng. Chem. Res. 44*, 863–868.

Höskuldsson, A. (1988). Pls regression methods. *J. Chemom 2*, 211–228.

Kadlec, P., B. Gabrys, and S. Strandt (2008). Data-driven soft sensors in the process industry. *Computers and Chemical Engineering*.

Kalivas, J. H. (1997). Two data sets of near infrared spectra. *Chemometrics and Intelligent LaboratorySystems 37*, 255–257.

Kariwala, V., Y. Cao, and S. Janardhanan (2008). Local self-optimizing control with average loss minimization. *Ind. Eng. Chem. Res. 47*, 1150–1158.

Martens, H., J. P. Nielsen, and S. B. Englesen (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures. *Anal. Chem 75*, 394–404.

Mejdell, T. and S. Skogestad (1991). Estimation of distillation compositions from multiple temperature measurements using partial-least-squares regression. *Ind.Eng.Chem.Res 30*, 2543–3555.

Mejdell, T. and S. Skogestad (1993). Output estimation using multiple secondary measurements: High-purity distillation. *AIChE Journal 39*(10).

Schreyer, S. K., M. Bidinosti, and P. D. Wentzell (2002). Application of maximum likelihood principal components regression to fluorescence emission spectra. *Applied Spectroscopy 56*(6), 789–796.

Shlens, J. (2005). *A Tutorial on Principal Component Analysis*. http://www.cs.cmu.edu/∼elaw/papers/pca.pdf.

Skogestad, S. (2000). Plantwide control: the search for the self-optimizing control structure. *Journal of Process Control 10*, 487–507.

Skogestad, S. (2004). Near-optimal operation by self-optimizing control: from process control to marathon running and business systems. *Computers and Chemical Engineering 29*, 127–137.

Skogestad, S. and I. Postlethwaite (2010). *Multivariable Feedback Control – Analysis and Design* (2 ed.). John Wiley & Sons, Ltd.

Skogestad, S., R. Yelchuru, and J. Jäschke (2011). Optimal use of measurements for control, optimization and estimation using loss method: Summary of existing results and some new. In M. Huba, S. Skogestad, M. Fikar, M. Hovd, T. Johansen,

and B. Rohal'-Ilkiv (Eds.), *Selected Topics on Constrained and Nonlinear Control – Workbook*. STU Bratislava – NTNU Trondheim.

Weber, R. and C. Brosilow (1972). The use of secondary measurements to improve control. *AIChE Journal 18*(3).

Appendices

Appendix A

Extra Derivations

A.1 Optimal Sensitivity $\mathbf{F}$

The optimal sensitivity $\mathbf{F}$ is the sensitivity of the optimal measurements $\mathbf{x}$ with respects to disturbances $\mathbf{d}$. For our purposes the measurements that keep the primary variables at the setpoints are the optimal. The optimal sensitivity can then be defined as

$$\mathbf{F} = \left(\frac{d\mathbf{x}_{\text{opt}}}{d\mathbf{d}}\right) = \left(\frac{d\mathbf{x}}{d\mathbf{d}}\right)_{\mathbf{y}=\mathbf{y}_s}$$

Given the linear expression for the primary variables and measurements

$$\mathbf{y} = \mathbf{G}_y\mathbf{u} + \mathbf{G}_y^d\mathbf{d}$$

$$\mathbf{x} = \mathbf{G}_x\mathbf{u} + \mathbf{G}_x^d\mathbf{d}$$

Assume that $\mathbf{u}$ is used to keep the predicted primary variable $\hat{\mathbf{y}}$ at their setpoint $\mathbf{y}_s$. The optimal value for $\mathbf{y}$, is when $\mathbf{y} = \mathbf{y}_s$ (which implies no prediction error), and gives

$$\mathbf{u}_{\text{opt}} = \mathbf{G}_y^{-1}\mathbf{y}_s - \mathbf{G}_y^{-1}\mathbf{G}_y^d\mathbf{d}$$

The optimal variation in the measurements $\mathbf{x}$ can be written as

$$\mathbf{x}_{\text{opt}} = \mathbf{G}_x\mathbf{u}_{\text{opt}} + \mathbf{G}_x^d\mathbf{d}$$

$$\mathbf{x}_{\text{opt}} = \left(\mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{G}_y^d\right)\mathbf{d} + \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{y}_s$$

Applying the definition of optimal sensitivity gives

$$\mathbf{F} = \frac{d\mathbf{x}_{\text{opt}}}{d\mathbf{d}} = \mathbf{G}_x^d - \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{G}_y^d$$

A.2 Cascade Arrangements

The following derivations will show that we can use closed-loop gains in the expression for the optimal "closed-loop" estimator. Also, it will show that when using closed-loop gains we assume a cascade control arrangement.

A.2.1 Inner Loop - Closed-Loop $\mathbf{y}$

The following expressions will be derived from the linear system on page 18. For the inner loop, assume that $\mathbf{u}$ is adjusted such that $\mathbf{y} = \mathbf{y}_s'$, and for the outer loop assume that $\mathbf{y}_s'$ is adjusted such that $\hat{\mathbf{y}} = \mathbf{y}_s$.

The expression for the primary variable and the predicted primary variable

$$\mathbf{y} = \mathbf{y}_s^{'} = \mathbf{G}_y\mathbf{u} + \mathbf{G}_y^d\mathbf{d}$$

$$\hat{\mathbf{y}} = \mathbf{y}_s = \mathbf{H}\left(\mathbf{G}_x\mathbf{u} + \mathbf{G}_x^d\mathbf{d} + \mathbf{n}_x\right)$$

Finding $\mathbf{u}$ as an explicit function of $\mathbf{y}_s'$

$$\mathbf{u} = \mathbf{G}_y^{-1}\left(\mathbf{y}_s' - \mathbf{G}_y^d\mathbf{d}\right)$$

Finding $\mathbf{y}_s'$ as a function of $\mathbf{y}_s$

$$\mathbf{y}_s = \mathbf{H}(\underbrace{\mathbf{G}_x\mathbf{G}_y^{-1}}_{\mathbf{G}_x^{cl}}\left(\mathbf{y}_s' - \mathbf{G}_y^d\mathbf{d}\right) + \mathbf{G}_x^d\mathbf{d} + \mathbf{n}_x)$$

$$\mathbf{y}_s' = \left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\left(\mathbf{y}_s - \mathbf{H}\left(\mathbf{G}_x^d\mathbf{d} + \mathbf{n}_x\right)\right) + \mathbf{G}_y^d\mathbf{d}$$

Expressing the prediction error as a function of $\mathbf{y}_s$

$$\mathbf{e} = -\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{H}\left(\left(\mathbf{G}_x^d + \mathbf{G}_x^{cl}\mathbf{G}_y^d\right)\mathbf{d} + \mathbf{n}_x\right) - \left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{y}_s - \mathbf{y}_s$$

$$\mathbf{e} = -\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{H}\begin{bmatrix}\mathbf{F} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{d}\\ \mathbf{I}\end{bmatrix} - \left[\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1} - \mathbf{I}\right]\mathbf{y}_s$$

Using Lemma 4 and Lemma 5, we find the optimal "closed-loop" estimator

$$\mathbf{H} = \arg\min_{\mathbf{H}} \left\|\mathbf{H}\begin{bmatrix}\mathbf{F}W_d & W_{n_x}\end{bmatrix}\right\|_{\mathbf{F}}$$

$$\text{s.t.} \quad \mathbf{H}\mathbf{G}_x^{cl} = \mathbf{I}$$

Where $\mathbf{G}_x^{cl} = \mathbf{G}_x\mathbf{G}_y^{-1}$, and $\mathbf{I} = \mathbf{G}_y^{cl} = \mathbf{G}_y\mathbf{G}_y^{-1}$.

A.2.2 Inner Loop - Closed-Loop z

The following expressions will be derived from the linear system on page 18. For the inner loop, assume that $\mathbf{u}$ is adjusted such that $\mathbf{z} = \mathbf{z}_s$, and for the outer loop assume that $\mathbf{z}_s$ is adjusted such that $\hat{\mathbf{y}} = \mathbf{y}_s$.

The expression for the secondary variable and the predicted primary variables

$$\mathbf{z} = \mathbf{z}_s = \mathbf{G}_z\mathbf{u} + \mathbf{G}_z^d\mathbf{d}$$

$$\hat{\mathbf{y}} = \mathbf{y}_s = \mathbf{H}\left(\mathbf{G}_x\mathbf{u} + \mathbf{G}_x\mathbf{d} + \mathbf{n}_x\right)$$

Finding $\mathbf{u}$ as an explicit function of $\mathbf{z}_s$

$$\mathbf{u} = \mathbf{G}_z^{-1}\left(\mathbf{z}_s - \mathbf{G}_z^d\mathbf{d}\right)$$

Finding $\mathbf{z}_s$ as a function of $\mathbf{y}_s$

$$\mathbf{y}_s = \mathbf{H}\big(\underbrace{\mathbf{G}_x\mathbf{G}_z^{-1}}_{\mathbf{G}_x^{cl}}\left(\mathbf{z}_s - \mathbf{G}_z^d\mathbf{d}\right) + \mathbf{G}_x\mathbf{d} + \mathbf{n}_x\big)$$

$$\mathbf{z}_s = \left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{y}_s - \left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{H}\left(\mathbf{G}_x\mathbf{d} + \mathbf{n}_x\right) + \mathbf{G}_z^d\mathbf{d}$$

Expressing the primary variable as an explicit function of $\mathbf{y}_s$ ($\mathbf{G}_y^{cl} = \mathbf{G}_y\mathbf{G}_z^{-1}$)

$$\mathbf{y} = \mathbf{G}_y^{cl}\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{y}_s - \mathbf{G}_y^{cl}\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{H}\left(\mathbf{G}_x\mathbf{d} + \mathbf{n}_x\right) + \mathbf{G}_y^d\mathbf{d}$$

$$\mathbf{y} = \mathbf{G}_y^{cl}\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{y}_s - \mathbf{G}_y^{cl}\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{H}\big(\left[\mathbf{G}_x - \mathbf{G}_x^{cl}\left(\mathbf{G}_y^{cl}\right)^{-1}\mathbf{G}_y^d\right]\mathbf{d} + \mathbf{n}_x\big)$$

Recognizing that

$$\mathbf{G}_x^{cl}\left(\mathbf{G}_y^{cl}\right)^{-1} = \mathbf{G}_x\mathbf{G}_z^{-1}\left(\mathbf{G}_y\mathbf{G}_z^{-1}\right)^{-1} = \mathbf{G}_x\mathbf{G}_z^{-1}\mathbf{G}_z\mathbf{G}_y^{-1} = \mathbf{G}_x\mathbf{G}_y^{-1}$$

and the optimal sensitivity ($\mathbf{F} = \mathbf{G}_x - \mathbf{G}_x\mathbf{G}_y^{-1}\mathbf{G}_y^d$) gives

$$\mathbf{y} = \mathbf{G}_y^{cl}\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{y}_s - \mathbf{G}_y^{cl}\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{H}\left(\mathbf{F}\mathbf{d} + \mathbf{n}_x\right)$$

Expressing the prediction error as a function of $\mathbf{y}_s$

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = -\mathbf{G}_y^{cl}\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1}\mathbf{H}\left(\mathbf{F}\mathbf{d} + \mathbf{n}_x\right) + \left[\mathbf{G}_y^{cl}\left(\mathbf{H}\mathbf{G}_x^{cl}\right)^{-1} - \mathbf{I}\right]\mathbf{y}_s$$

Using Lemma 4 and Lemma 5, we find the optimal "closed-loop" estimator

$$\mathbf{H} = \arg\min \left\|\mathbf{H}\begin{bmatrix}\mathbf{F}W_d & W_{n_x}\end{bmatrix}\right\|$$

$$\text{s.t.} \qquad \mathbf{H}\mathbf{G}_x^{cl} = \mathbf{G}_y^{cl}$$

Where $\mathbf{G}_x^{cl} = \mathbf{G}_x\mathbf{G}_z^{-1}$, and $\mathbf{G}_y^{cl} = \mathbf{G}_y\mathbf{G}_z^{-1}$.

Appendix B

Full Linear Model

The linear approximation to the binary mixture "column A" for all 41 stage temperatures are given in the following table. The model was found by inducing small perturbations, and recording the response.

$\mathbf{G}_x$		$\mathbf{G}_x^d$
u_1 (L_D)	u_2 (V_B)	d (z_F)
-14.6	14.5	-16.1
-20.8	22.0	-22.8
-28.5	27.3	-31.4
-38.3	40.7	-41.9
-50.1	48.2	-55.3
-64.8	68.0	-71.0
-81.9	80.2	-90.2
-102.0	105.3	-111.8
-124.2	123.5	-136.3
-148.1	150.9	-161.9
-171.9	172.4	-187.8
-194.2	196.5	-211.3
-212.4	213.7	-230.5
-224.7	226.8	-243.0
-229.6	231.1	-247.4
-226.3	227.9	-243.3

-215.4	216.7	-231.3
-198.4	199.4	-213.2
-177.2	177.9	-191.0
-154.1	154.5	-167.0
-130.8	130.9	-143.0
-149.1	148.9	-160.1
-166.0	165.5	-175.5
-180.2	179.3	-187.7
-190.2	189.1	-195.5
-195.0	193.7	-197.9
-194.1	192.5	-194.6
-187.4	185.7	-185.9
-175.6	174.0	-172.5
-160.0	158.4	-155.7
-142.0	140.5	-137.0
-122.9	121.6	-117.7
-104.0	102.8	-99.0
-86.2	85.2	-81.7
-70.0	69.2	-65.9
-55.8	55.0	-52.2
-43.5	42.9	-40.6
-33.2	32.7	-30.8
-24.6	24.2	-22.8
-17.5	17.3	-16.2
-11.8	11.6	-10.9

Appendix C

Number of Principal Components

The table below gives the number of principal components used for each regression in Chapter 7.

Table C.1: Number of principal component used in regression for column A.

number of	Binary Mixture			Multicomponent mixture		
measurements	CLM	PCR	PLS	CLM	PCR	PLS
4	4	4	4	4	4	4
5	4	4	4	4	4	4
6	5	5	4	4	4	4
7	5	5	4	5	4	4
8	5	5	5	4	4	4
9	6	6	6	5	5	4
10	7	7	6	5	5	4
11	7	7	6	5	5	4
12	7	7	5	5	5	4
13	7	7	5	5	5	4
14	7	7	6	5	5	4
15	8	8	4	5	5	4
16	9	9	4	5	5	4
17	9	9	4	5	7	4
18	9	9	4	6	7	4
21	12	12	4	7	7	4

19	10	10	4	7	5	4
20	11	11	4	5	7	4
22	12	12	4	7	9	4
23	9	9	4	9	9	4
24	14	14	4	5	10	4
25	14	14	4	10	9	4
26	14	14	4	5	9	4
27	12	14	4	9	10	4
28	6	6	4	10	15	4
29	6	6	4	15	15	4
30	6	6	4	15	16	4
31	7	7	4	19	16	7
32	8	8	4	16	17	7
33	8	8	4	17	16	4
34	9	9	4	17	17	4
35	12	12	4	17	16	4
36	13	13	4	16	17	4
37	13	13	4	6	6	4
38	13	13	4	16	16	4
39	14	14	4	16	6	4
40	13	9	4	6	6	5
41	9	9	4	11	6	5

Appendix D

Standard Challenge Problems

The truncated CLM, PCR, and PLS will in this chapter be tested on two standard regression problems. The two examples are "open-loop" system. That is, the estimator will only be used for monitoring purposes.

D.1 Gluten Test Example

The object is to precisely predict the gluten concentration from NIR absorbance. The data set (Martens et al., 2003) has one dependent variable y (gluten concentration), 100 independent variables $\mathbf{x}$ (NIR absorbents) with each 100 samples. The data set was split up into two parts, and cross-validated.

Figure D.1, shows the performance of the truncated CLM, PLS, and PCR as a function of the number of principal components used under calibration. PLS had the best performance when using 10 principal components, while both CLM and PLS (performing almost identical) had their best performance when using 19 components. PLS had the best performance of the three method.

D.2 Wheat Test Example

The object is to predict water and protein content from NIR absorbance. The data set (Kalivas, 1997) has two dependent variables $\mathbf{y}$ (wheat and water

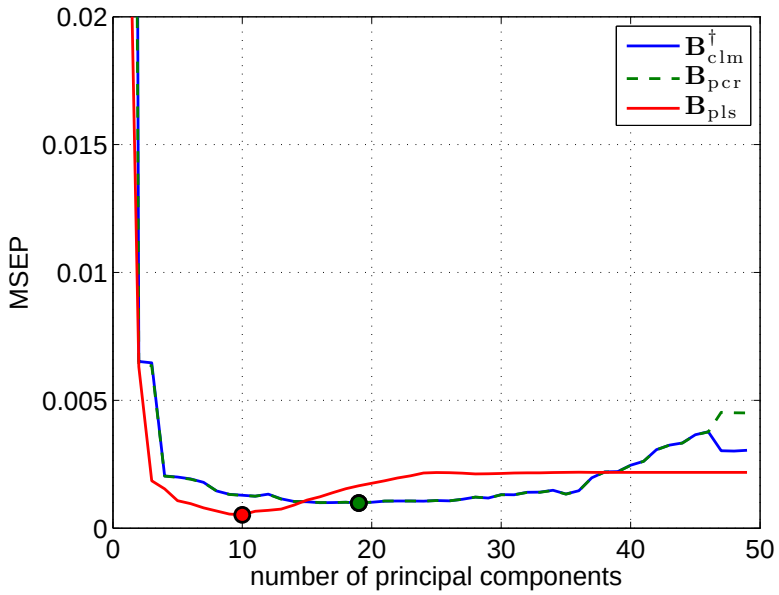


Figure D.1: Estimator performance for different number of principal components used under calibration for the gluten test example.

contests), and 701 independent variables $\mathbf{x}$ (NIR absorbent) with each 100 samples. The data set was split into two parts; even numbered samples in one part and odd numbered samples in the other, and cross-validated.

Figure D.2, shows the performance of the truncated CLM, PLS, and PCR as a function of the number of principal components used under calibration. PLS had the best performance for 28 principal components, while CLM and PCR had their best performance when using 47 components . PLS gave the overall best performance.

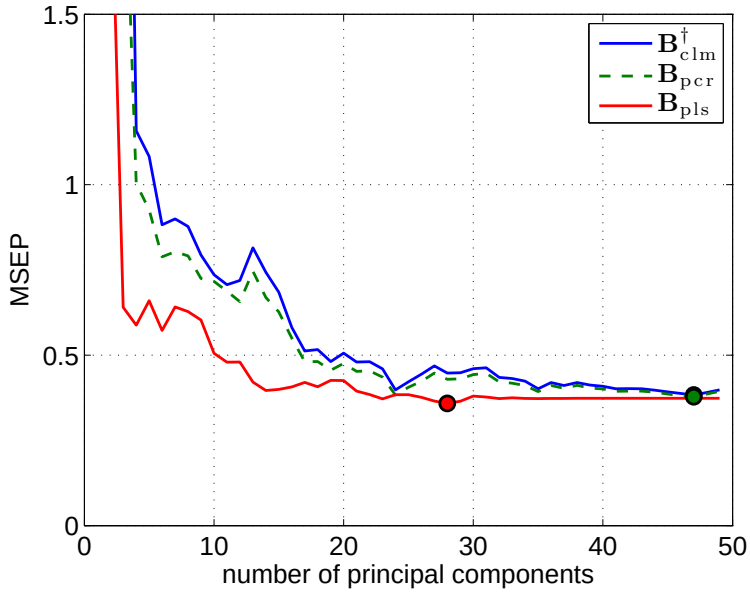


Figure D.2: Estimator performance for different number of principal components used under calibration for the wheat test example.

Appendix E

Matlab Code

E.1 Truncated CLM

```
% Function for finding the truncated "closed-loop"
% method (truncated CLM) estimator
%           written by Chriss Grimholt, april 2011
function B = tclm(Y,X,n)
% Y is the dependent variables
% X is the independent variable
% n is the number of principal components
% Principal Component Analysis
[u,s,v] = svd(X);
u = u(:,1:n); s = s(1:n,1:n); v = v(:,1:n);
X = u*s*v';
% The Truncated CLM
[u,s,v] = svd(Y); Y = Y*v; X = X*v;
p  =  size(Y,1); % number of independent variables
G1 = Y(:,1:p); Gy = X(:,1:p); Ftilde = X(:,p+1:end);
[u,s,v] = svd(Ftilde*Ftilde');
u = u(:,1:n); s = s(1:n,1:n); v = v(:,1:n);
s=s^-1; invF = v*s'*u';
% Optimal "Closed-Loop" Estimator
B = invF*Gy*pinv(Gy'*invF*Gy)*G1'; B=B';
end
```

E.2 Cross-Validation

E.2.1 Data Rotation Function

```
% splits up data into SPLIT number of blocks and assigns block P as the
% validation block. The remaning blocks are merged into the calibration
% blocks.        written by: Chriss Grimholt, 3/5-2011
function [Ycal,Xcal,Yval,Xval] = datarot(Y,X,split,p)
m = size(X,2);    % number of samples in data
set = 1:m;        % generating index for the samples
d = mod(m,split); % finds the leftover after split
n=(m-d)/split;    % finding number of samples in split
% split into split sets
setval={}; n_set = 0;
n0 = 1; % start sample
n1 = n; % number of samples in plit
while n0 <= m    % distributing data between the sets
    temp = n0 : n1;
    if d > 0
        temp = [temp, n1+1]; d = d-1; n1=n1+1;
    end
    n0 = n1+1; n1 = n0+(n-1); n_set = n_set+1; setval{n_set}=temp;
end
% validataion data
Yval = Y(:,setval{p}); Xval = X(:,setval{p}); setcal=[];
% finding which samples are not in the validation data
for i = 1:m
    if ismember(i,setval{p})
        i=i+1;
    else
    setcal = [setcal;i];
    end
end
Ycal = Y(:,setcal); Xcal = X(:,setcal); % calibration data
% special case for no split
if split == 1
Yval = Y; Xval = X; Ycal = Y; Xcal = X;
end
end
```

E.2.2 Cross-Validation Function

```
% Cross-Validation Function, B is the optimal estimator,
% nopt is the estimated optimal number of components,
% MSEP_opt is the squared error,
% Y and X is the input data, and split is the CV split.
%            written by Chriss Grimholt, april 2011
function [B, nopt, MSEP_opt] = cvtclm(Y,X,split)
% maximum number of principal components
n_max = min(size(X,2)-ceil(size(X,2)/split), size(X,1));
% Cross-Validating for n = 1 to n = n_max
for n = 1:n_max
    res = [];
    for p = 1:split
        [Ycal,Xcal,Yval,Xval] = datarot(Y,X,split,p);
        B=tclm(Ycal,Xcal,n);
        error(:,p) = sum((Yval-B*Xval).^2,2); % validating
    end
    MSEP(:,n)=sum(error,2)/size(X,2); %storing MSEP for each n
end
MSEP=sum(MSEP);
% summing up MSEP for multiple independent variables
[MSEP_opt, nopt] = min(MSEP); % findng best component
% Estimator
B=hdata(Y,X,n);
% Plot of crossvalidation
Pl = plot(1:n_max,MSEP,nopt,MSEP_opt,'or',nopt,0,'xr');
set(Pl(1),'LineStyle','-','LineWidth',1.5,'color','blue')
set(Pl(2),'LineStyle','o','LineWidth',1.5,'color','red')
set(Pl(3),'LineStyle','x','LineWidth',1.5,'color','red')
axis([0 n_max 0 max(MSEP)])
set(gca,'FontSize',14)
grid on
xlab = xlabel('Number of principal components');
ylab = ylabel('MSEP');
set(ylab,'FontSize',14,'interpreter','tex')
set(xlab,'FontSize',14,'interpreter','tex')
end
```