INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

University of Alberta

Optimal Sensor Selection and Parameter Estimation for Real-Time Optimization

by

Lisa Marie Fraleigh

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Process Control

Department of Chemical and Materials Engineering

Edmonton, Alberta

Spring 1999

University of Alberta

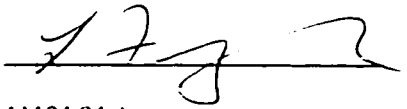
Library Release Form

Name of Author: Fraleigh, Lisa Marie

Title of Thesis: Optimal Sensor Selection and Parameter Estimation for Real-Time Optimization

Degree: Master of Science

Year this Degree Granted: 1999



11104 84 Avenue
Suite 401
Edmonton, Alberta
T6G 2R4

December 14, 1998

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Optimal Sensor Selection and Parameter Estimation for Real-Time Optimization submitted by Lisa Marie Fraleigh in partial fulfillment of the requirements for the degree of Master of Science in Process Control.

J.F. Forbes

M. Guay

T. Chen

B. Huang

J. Kresta

November 30, 1998

Abstract

A Real-Time Optimizer is an on-line steady-state model-based optimizer that aims to improve the process profitability by adjusting operations in response to process changes. Optimizer effectiveness depends on such factors as parameter estimation technique and measurement selection for parameter updating. This thesis evaluates parameter estimation techniques in terms of parameter quality, and develops Sensor System Design Cost as a tool for sensor selection.

Back-substitution (the solution of p equations in p unknowns) remains a common parameter estimation technique because it is simple to implement. This thesis shows that a least squares technique can substantially reduce parameter uncertainty, while being more robust to measurement error and sensor failure.

Available methods for sensor selection focus on parameter variance. This thesis analyzes sensor selection in terms of the cost of setpoint variance and bias. The significance of the Sensor System Design Cost is that the expected performance of various sensor systems can be compared in terms of dollars.

Acknowledgements

I would like to thank my supervisors, Fraser Forbes and Martin Guay, for their guidance and patience during this work. I would also like to thank NSERC for their financial support. Special thanks to my family and friends for their encouragement throughout this project.

Table of Contents

List of Tables

List of Figures

Nomenclature

$\mathbf{A}$	coefficient matrix for a linear model
A_i	frequency factors in Arrhenius rate Equation (2.30), hr^{-1}
$\mathbf{a}$	vector of independent and dependent variables of Problem (C.6)
B_i	activation energies in Arrhenius rate Equation (2.30), R
$\mathbf{b}$	bias vector
C	design cost
c	random variables defined in Equation (2.3)
$\mathbf{d}$	vector of dependent variables in Problem (C.6)
$\mathbf{e}$	vector of model equation residuals
$\mathbf{F}_i$	flow rate of stream i, lb/hr
$\mathbf{F}$	vector of functions defined in Equation (C.13)
$\mathbf{f}$	vector of model equations
f	probability density function
$\mathbf{g}$	vector of inequality constraints, defined for Problem (C.6)
$\mathbf{g}_1$	function relating coefficient matrix $\mathbf{A}$ to measurements $\mathbf{z}$
$\mathbf{g}_2$	function relating bias vector $\mathbf{b}$ to measurements $\mathbf{z}$
$\mathbf{h}$	vector of inequality constraints, defined for Problem (3.12)
h_j	magnitude of $\Delta\beta_j$, $j = 1..p$, defined Equation (2.29)
k_i	Arrhenius rate of reaction i, hr^{-1}
L	Lagrange function
m	number of measured variables
n	number of independent model equations
P	profit function
p	number of unknown parameters
$\mathbf{Q}_{\mathbf{x}^*}$	covariance matrix of setpoints
$\mathbf{Q}_\mathbf{z}$	covariance matrix of measurements

$\mathbf{Q}_\beta$ covariance matrix of parameters

$\mathbf{Q}_{\hat{\beta}}$ covariance matrix of parameters, estimated at $\hat{\beta}$

$\mathbf{Q}_\epsilon$ covariance matrix of measurement errors

Q heat duty, Btu/hr

$\mathbf{R}$ Cholesky factor defined in Equation (3.14)

R reflux ratio

$\mathbf{S}$ the space of possible optimal setpoints

s number of setpoints

T_R reactor temperature in case studies, R

U heat transfer coefficient

$\mathbf{u}$ vector of KKT multipliers

$\mathbf{v}$ vector of KKT multipliers

v reactor volume in case studies, lb

$\mathbf{W}$ weighting matrix for weighted least squares estimation

$\mathbf{w}$ vector of ones, defined in Equation (3.21)

X_i weight fraction of component i

$\mathbf{x}$ vector of independent variables in Equation (2.9)

$\mathbf{x}^*$ vector of setpoint values

$\tilde{\mathbf{x}}^*$ vector of nominal setpoint values, found using noise-free measurements from a given sensor system

$\mathbf{x}_m^*$ vector of setpoints found using the most rigorous model and a complete set of perfect measurements

x mole fraction of component P not in the azeotropic mixture, Equation (4.3)

$\mathbf{y}$ vector of response variables in Equation (2.9)

$\mathbf{z}$ vector of measurements

$\mathbf{z}_o$ vector of the expected values of the measurements, defined for Equation (3.3)

$\mathbf{z}^*$ true values of measured variables

$\mathbf{z}_c^*$ complete set of perfect measurements

α	relative volatility
$\boldsymbol{\beta}$	vector of true parameter values
$\hat{\boldsymbol{\beta}}$	vector of parameter estimates
$\tilde{\boldsymbol{\beta}}$	vector of parameter estimates, calculated using noise-free measurements
$\boldsymbol{\delta}$	integration variable defined on Ψ, Equation (3.32)
$\boldsymbol{\delta}_G^{(k)}$	Gauss-Newton increment at step k
$\boldsymbol{\varepsilon}$	vector of measurement noise
$\boldsymbol{\eta}$	vector of transformed setpoints, by Equation (3.13)
$\bar{\boldsymbol{\eta}}$	vector of nominal transformed setpoints, found using noise-free measurements from a given sensor system
$\boldsymbol{\eta}_m$	vector of transformed setpoints found using the most rigorous model and a complete set of perfect measurements
$\kappa_{\hat{\beta}}$	condition number of parameter covariance matrix $\mathbf{Q}_{\hat{\beta}}$
$\boldsymbol{\lambda}$	vector of Lagrange multipliers, Problem (A.8)
λ_i	eigenvalue of parameter covariance matrix, $i = 1..p$
Ξ	class of experimental designs, defined for Problem (3.8)
ξ	experimental design belonging to Ξ, defined for Problem (3.8)
σ^2	variance of measurement noise
$\varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right)$	frequency function describing the occurrence rate of a particular plant optimum
$\boldsymbol{\tau}_i$	eigenvector of parameter covariance matrix, $i = 1..p$
ϕ	objective function of Problem (C.11)
φ	objective function of Problem (A.8)
Ψ	the space of possible predicted optimum setpoints for the given measurement set and external variable values
ψ	objective function of Problem (C.6)
$\varkappa$	the set of alternative sensor systems

Operators

∇^2	Hessian
$\|\mathbf{R}\|$	determinant of matrix $\mathbf{R}$
$\mathbf{R}^T$	transpose of matrix $\mathbf{R}$
$\|\mathbf{x}\|$	L_2 norm of $\mathbf{x}$

Chapter 1

Introduction

Economic optimization of process operations is important as industries strive to achieve or maintain their competitive advantage by producing a product at minimum cost [Cutler and Perry, 1983]. This thesis will be concerned with steady-state, model-based On-line Optimization, or Real-Time Optimization (RTO). On-line process optimization is used to continuously monitor process performance and make appropriate operating adjustments to improve plant profitability.

As shown in Figure 1-1, RTO is the link between process planning and scheduling, and process control. The purpose of an RTO system is to maintain an economically optimal operations policy for processes with time-varying behaviour [Forbes, 1994]. Process behaviour can vary for a number of reasons. For example, changes in ambient conditions may affect cooling water temperatures, air cooler efficiencies and heat loss from equipment. In addition, catalysts decay, heat exchangers foul, distributors plug, and feedstocks may vary. There are economic benefits to be had when the process is operated to account for such process changes [Cutler and Perry, 1983]. The degree to which these benefits are realized will depend on the design of the RTO system [de Hennin *et al.*, 1994; Forbes, 1994]. RTO design decisions include model selection, data validation techniques, parameter estimation method, measurement selection, optimization algorithm, and so forth. In this thesis, a statistical approach is used to evaluate current parameter

estimation techniques, and to develop a new tool for making sensor selection decisions.

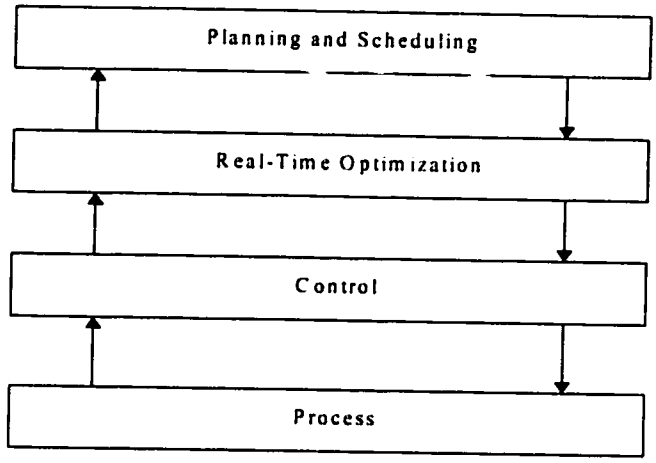


Figure 1-1: Plant decision-making hierarchy.

This chapter presents an overview of process optimization, and RTO in particular. It also serves to review the current state of RTO design technology, as a motivation for the work included in this thesis. A description of the thesis contents and conventions is given in the final sections.

1.1 Approaches to Process Optimization

There are two general categories for steady-state process optimization: direct search and model-based methods [Garcia and Morari, 1981]. Direct search methods involve actual plant experimentation to examine the process response surface directly, so that a direction for improved performance can be identified. An example of direct search method is Evolutionary Operations (EVOP) [Box and Draper, 1969]. In this method, experimentation is performed by repeatedly making a number of very small deviations from current process operation, and the resulting effect on process performance is determined. Once a direction for improved plant performance is identified, the plant operations are moved

to a new point in that direction. The experimentation is repeated at the new point, and iteratively the process operations are improved until an optimum is found. More recent direct search methods involve plant experimentation to identify a dynamic model of the process (whose structure is pre-determined), and then use this dynamic model to estimate steady-state behaviour for the evaluation of process changes that yield improved economic performance [Bamberger and Iserman, 1978; Garcia and Morari, 1981; MacFarlane and Bacon, 1989].

The benefits of direct search techniques are that the methods are simple and well understood, and can be effective when very little is known about the process. These methods can be difficult to justify in large plants with many degrees of freedom available for optimization. For example, in EVOP the number of plant experiments that must be performed increases geometrically with the number of variables to be optimized [Box and Draper, 1969]. Further, steady-state must be achieved for each experiment, which may be extremely time-consuming in processes with large settling times. Meanwhile, operating conditions are constantly changing so that the true optimum itself is moving. It therefore seems possible that an optimum may never be found.

The main drawback of direct search methods, and the reason substantial plant experimentation is required, is that no *a priori* process information is used. In theory, if a process model is used for optimization, plant experimentation could be reduced such that only a few key model parameters be determined. The more accurately the model represents the data, the less experimentation required. The model can then be used to predict plant behaviour over a wide range of operations, and to estimate the location of the plant optimum. Once operations are moved to the approximate optimum, experimentation is repeated. As model fidelity improves, the number of iterations required to reach the true plant optimum decreases.

Model-based optimizers may use either steady-state or dynamic models. Since this thesis is concerned with optimal steady-state operations, the following discussion will be limited to steady-state models. A Real-Time Optimizer (RTO) or On-line Optimizer is

a steady-state model-based optimizer.

1.2 Overview of Real-Time Optimization

A general schematic for an RTO system is shown in Figure 1-2. The component subsystems are Measurement, Data Validation, Model Updating, Optimization, Command Conditioning (also called Results Analysis), and Control. A description of these subsystems is given below.

Measurement. In order to track process changes, measurements must be taken. These measurements will be used to update the process model.

Data Validation. This subsystem includes steady-state detection, gross error detection, and data reconciliation. Since steady-state models are used, measurement data should represent steady-state conditions. To detect steady-state, automated examination of operation statistics over a recent time interval is generally performed [White, 1997]. The steady-state measurements may contain errors such as sensor malfunction or leaks (gross errors), as well as random process noise. These errors can cause inaccuracies in the model-based predictions of the plant optimum, because measurement error propagates to parameter error. This contributes to plant-model mismatch and results in erroneous setpoints being calculated by the optimizer. Therefore, the gross errors must be detected and removed or corrected. Gross error detection techniques use statistical tests to compare measurement deviations with random behaviour [Crowe *et al.*, 1983 ; Crowe, 1986; Crowe, 1988; Mah and Tamhane, 1982; Rosenberg *et al.*, 1987; Tong and Crowe, 1995]. Data reconciliation is the process of using statistical principles to allocate the remaining random errors across the measured variables in an attempt to obtain the best possible estimate of actual plant operation [Crowe *et al.*, 1983; Crowe, 1986; Crowe, 1996]. Generally, gross error detection and data reconciliation are performed sequentially, but Tjoa and Biegler [1991] have developed strategies for solving the two problems simultaneously. Albuquerque and Biegler [1996] extended the simultaneous strategy to dynamic systems.

Model Updating. By using measurements to estimate current model parameters (such as heat transfer coefficients or reaction rate constants), real-time model updating permits realistic models to be used by the optimizer [Cutler and Perry, 1983].

Optimization. Steady-state model-based optimization is performed on the updated process model, to determine the feasible operating point that generates the maximum profit.

Command Conditioning (Results Analysis). The optimizer will produce setpoint estimates that correspond to the model-based optimal plant operation. However, there will be uncertainty associated with these setpoints due to various sources of error throughout the RTO loop. Command conditioning examines the optimizer output to ensure that it represents meaningful changes in plant operations and expected improvement in profitability.

Control. To capture the expected profit, the optimizer needs a control system that can reliably move the plant to the optimum [Scott, 1996] .

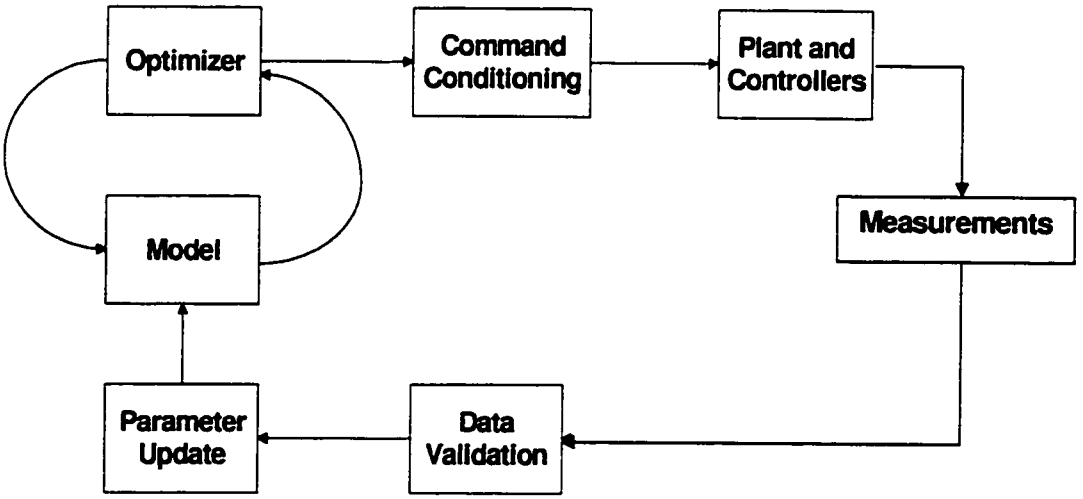


Figure 1-2: Real-Time Optimization loop.

1.3 RTO Design

As shown in Figure 1-2, RTO is a closed-loop system, and therefore the reliability of its results will depend on the performance of each of the subsystems. Cutler and Perry [1983] have expressed the probability of success of optimization in terms of the accuracy of the model parameters as well as the models themselves, the accuracy of the economic information, and the performance of the process control system. Anything that improves the accuracy of these factors will enhance the expected improvement in process profitability through on-line optimization. On the other hand, sufficient inaccuracies in even one of these areas can lead to results that can decrease the profitability of plant operation [Cutler and Perry, 1983].

A number of researchers have done work related to the improvement of RTO system design. Of central importance to model-based optimization is the quality of the process model. In testing model quality, Durbeck [1965] and Biegler *et al.* [1985] have shown that a model is adequate for optimization if the derivatives of the objective function with respect to the optimization variables match those of the process. Forbes [1994] has presented various measures for assessing model fidelity, while taking the closed RTO loop into consideration.

The problem of selecting which model parameters to update on-line has also been addressed. Krishnan *et al.* [1992] suggest that those parameters that have the greatest effect on the objective function or alter the active constraint set should be updated on-line. Once the parameters have been chosen, other important design decisions include the parameter estimation technique, and the selection of which measurements to use in the parameter update. These items are the focus of this thesis, and are discussed in detail in the following sections.

1.3.1 Parameter Estimation

Although it is generally understood that the parameter estimation technique used in the model update will have a significant effect on the parameter estimates, little work has been done to assess these effects in an RTO framework. Parameter estimation is a particularly difficult problem under current RTO practices for two important reasons. First, the estimation is usually performed using a single set of corrupted steady-state data at the current operating point. Second, since computational efficiency is an important consideration, the size of the RTO problem can limit the possible parameter estimation technique alternatives. For a complete ethylene plant the model may contain 200,000 equations and variables [White, 1997].

There are many parameter estimation schemes available, such as back-substitution, least squares, or error-in-variables estimation. A number of researchers have combined data reconciliation and parameter estimation into one problem [Kim *et al.*, 1991; Liebman *et al.*, 1992; Pages *et al.*, 1994; Tjoa and Biegler, 1992]. The most widely used parameter estimation technique for RTO is back-substitution, in which p equations are solved for p unknown parameters. This method, though simple to implement, is the least robust method in terms of the transmission of measurement error to the parameters [Dahlquist and Bjorck, 1974]. Any method that is more robust to random measurement error will give more reliable parameter estimates and hence a more reliable optimizer.

When selecting a parameter estimation scheme, observability issues are also important and have received some attention. Krishnan [1990] defined parameter observability for steady-state systems as whether or not model parameters can be uniquely determined through a steady-state process model. Singh [1997] identified the importance of including the estimation scheme within the framework of observability, since the structure of the estimator may not allow for the unique mapping of measurements to parameters. Further, Singh [1997] shows the importance of considering the degree of parameter observability when selecting an estimation scheme. In addition to estimation technique, the measurements used to update the parameters can have a significant effect on their

quality as well.

1.3.2 Measurement Selection

As stated above, it is important that the parameters be observable through the process model, estimation scheme, and process measurements. Krishnan *et al.* [1992] have presented a method for measurement selection based on parameter observability and accessibility, and the statistical properties of the parameter confidence region. Measurements are accessible to the parameters if they contain some information about the parameters. That is, the measurements change when the parameters change [Krishnan *et al.*, 1992]. When a measurement is deemed inaccessible, it is excluded from consideration for parameter estimation. Once inaccessible measurements are removed from the set of possible measurements, the measurements for model updating are chosen based on their contributions to parameter uncertainty. The method uses the singular values of the parameter covariance matrix (which are indicative of the variance associated with the parameter estimates), and chooses measurements that minimize the overall parameter uncertainty.

Although parameter variance is an important factor when selecting measurements for on-line parameter updating, Krishnan *et al.* [1992] have only considered the model updating portion of the RTO loop. Figure 1-2 shows the closed-loop relationship between the measurements, the parameters, and the setpoints. Clearly the choice of measurements will impact not just the model parameters, but the estimated setpoints and overall RTO performance. Rather than choose measurements based on parameter quality, it would seem beneficial to incorporate the knowledge of the optimization algorithm and the profit function to choose measurements based on setpoint quality and expected process profitability. Setpoint quality can be defined in terms of variance and offset from the expected plant optimum. In upfront design work, a measurement set can then be chosen for maximum economic benefit. In an operational RTO system, it would be useful to identify where an additional sensor should be installed for the maximum improvement in prof-

itability. This thesis develops such a measurement selection criterion in terms of both bias and variance measures in setpoint uncertainty, as well as expected RTO profitability.

1.4 Thesis Scope

This thesis focuses on the selection of parameter estimation technique for model updating, and the selection of measurements to be used for parameter estimation. As was discussed in the previous sections, there has been little work done in the evaluation of parameter estimation techniques for RTO. Since the RTO parameter estimation problem is poorly conditioned to begin with, it is important to find a robust yet computationally practical method to ensure that the best possible parameter estimates are being sent to the optimization portion of the RTO loop. Measurement selection plays an important part in the quality of the parameter estimates as well. Although Krishnan *et al.* [1992] have established a method for measurement selection based on parameter variance measures, the method suffers because it considers the parameter updator in isolation, separate from the overall RTO loop. Since the goal of RTO is increased profitability, it seems that the practical approach to measurement selection is to find the sensor system that is expected to give the maximum RTO profitability.

Chapter 2 investigates various parameter estimation techniques, in terms of the quality of the parameter estimates and the robustness of the methods to measurement error and sensor malfunction. It also examines the computational requirements of the different methods. Once the method has been chosen, another important factor is the measurements selected for the parameter estimation. Chapter 3 takes two approaches to measurement selection, which result in similar sensor selection criteria. First, D- and T-optimal design of experiments criteria [Kiefer and Wolfowitz, 1959; Atkinson and Federov, 1975] are used to develop criteria that will identify sensor systems to meet specific parameter quality objectives, and weighted D- and T-criteria are introduced to incorporate expected profitability. Secondly, a Design Cost [Forbes and Marlin, 1996]

approach is applied to the measurement selection problem, to meet both setpoint quality objectives and include expected profitability. The resulting Sensor System Design Cost criterion is a combination of the D- and T-optimality criteria, which may be used to make sensor selection decisions for RTO. Chapter 4 presents a case-study that demonstrates how the new sensor selection tool may be used to solve a practical problem.

Throughout this thesis, the following assumptions have been made:

1. All process measurements are taken at steady state.

2. All the gross errors have been removed from the data; only normally distributed random noise is associated with the measurements.

3. The control system is able to implement the setpoints computed by the optimizer, in order to guarantee feasible operation regardless of optimizer output. The control structure is not considered in the process model.

1.5 Thesis Conventions

This section covers terms and conventions used in this thesis.

The **process model** refers to the system of equations describing the material and energy balances, the physical phenomena, as well as operating constraints.

Process variables are divided in a number of ways. First, they may be **manipulated** or **dependent**. The manipulated variables are the ones that can be independently adjusted to optimize plant performance, while the dependent variables are uniquely determined once the manipulated variable values are set. With respect to measurements, the process variables can be **measurable** or **unmeasurable**. Measurable variables are those that *can* be measured: flows, temperatures, etc., for which a sensor may be installed. The measurable variables are those involved in sensor selection decisions, and may be **measured** or **unmeasured**. A **complete measurement set** refers to the case where all measurable variables are actually measured. Unmeasurable variables are those that

cannot be measured, either because the sensor does not exist, or it cannot be installed for some physical reason. These variables are not involved in sensor selection decisions.

The term **nominal** refers to a zero-noise situation. The nominal values of the measurements are thus the true values of those measured variables. The nominal parameter values are those obtained using the nominal measurements. The nominal calculated setpoints are those found by the optimizer when the nominal parameter values are used in the model. Unless otherwise specified, the reader is to assume the measurements are corrupted by random, normally distributed noise. A single set of measurements, or parameter estimates, or setpoints, is one sample from the distribution of possible measurements, parameter estimates, or setpoints, respectively, at the current state of operations.

Throughout this thesis, all terms are explained on first usage. The nomenclature section lists all the symbols used.

Chapter 2

Parameter Estimation for RTO

The parameter estimation portion of the RTO loop was introduced in the previous chapter. For a given measurement set $\mathbf{z}$, and the model equations $\mathbf{f}$, the parameters $\boldsymbol{\beta}$ are found using a given estimation scheme. Using the parameter estimates, the optimizer attempts to find a more profitable operating region. The accuracy of the parameter estimates is important since parameter error may cause the optimizer to send the plant to a less profitable operating region [Cutler and Perry, 1983]. Therefore, the estimation scheme should be chosen such that the error in the parameter estimates is minimized.

The purpose of this chapter is to investigate three parameter estimation techniques: back-substitution, least squares estimation, and the error-in-variables method, and to show that the common practice of back-substitution is not as robust as the alternative methods, with respect to the transmission of measurement error and random process noise. The chapter begins with a discussion of parameter observability, including the importance of examining the degree of observability. The three parameter estimation techniques investigated in this thesis are discussed in Sections 2.2 and 2.3. The chapter concludes with a case study that illustrates the advantages and disadvantages of the parameter estimation techniques.

2.1 Parameter Observability

For parameter estimation, observability refers to whether or not model parameters can be uniquely determined using a set of measurements [Sorenson, 1980]. Consider the simple example:

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{b} \tag{2.1}$$

where $\mathbf{A}_{n\times p}$ and $\mathbf{b}_{p\times 1}$ are functions of the measurements. A unique set of parameters can be determined if and only if $\mathbf{A}$ is rank p, where p is the number of parameters. Equivalently, a unique set of parameters can be found if and only if $\mathbf{A}^T\mathbf{A}$ is positive definite [Sorenson, 1980].

Observability was first introduced by Kalman [1960] to describe state-output relationships in linear dynamic systems. It was used to determine whether the states of a system could be uniquely determined through a given set of measurements. Stanley and Mah [1981] define state observability for nonlinear steady-state systems as whether or not changes in the states can be determined through a set of measurements and a steady-state process model. Krishnan [1990] expanded this definition of observability to apply to parameter observability, as well as state observability. According to her definition, a system is observable if all of the states and model parameters can be observed through a set of process measurements and the process model.

For practical purposes, the model parameters must be observable through the estimator, as well as the measurements and process model. There are a number of estimation procedures that may be used to determine the unknown quantities, and a poorly designed procedure may not allow for the unique mapping of measurements to unknowns. Singh [1997] extended the definition of parameter observability to include the effects of the estimation scheme, by using the statistical properties of the parameter covariance matrix $\mathbf{Q}_{\hat{\beta}}$. She stated that:

> For any updating procedure where parameter estimates ($\hat{\boldsymbol{\beta}}$) (assumed to be

locally normally distributed) are to be determined using measurements ($\mathbf{z}$), the parameters are observable if all the eigenvalues of $\mathbf{Q}_{\hat{\beta}}^{-1}$ are strictly positive (positive and nonzero), where $\mathbf{Q}_{\hat{\beta}}$ is the covariance matrix of $\boldsymbol{\beta}$ estimated at $\hat{\boldsymbol{\beta}}$.

This definition provides a point-wise binary test for whether the parameters are observable through the measurements and the estimation procedure. Singh [1997] also recognized a need for considering the degree of observability in parameter estimation. She considers the degree of observability to be the quality of the estimates that can be obtained from a given set of measurements through the updator. Here, quality refers to both size and shape of the parameter confidence region.

The size and shape of the parameter confidence region are related to the eigenvalues of the parameter covariance matrix [Sorenson, 1980]. Assume that the p parameter estimates can be locally approximated by a multivariate normal distribution $\hat{\boldsymbol{\beta}} \sim \mathbf{N}_p(\boldsymbol{\beta}, \mathbf{Q}_{\boldsymbol{\beta}})$, where $\boldsymbol{\beta}$ are the true parameter values and $\mathbf{Q}_{\boldsymbol{\beta}}$ is the positive definite parameter covariance matrix. The probability density function for $\hat{\boldsymbol{\beta}}$ is given by:

$$f\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{(2\pi)^{\frac{p}{2}} \left|\mathbf{Q}_{\boldsymbol{\beta}}\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^T \mathbf{Q}_{\boldsymbol{\beta}}^{-1}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right] \tag{2.2}$$

This distribution is characterized by the quadratic form that appears in the exponent:

$$c^2 = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^T \mathbf{Q}_{\boldsymbol{\beta}}^{-1}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \tag{2.3}$$

For a given value of c^2, Equation (2.3) defines the surface on which the probability density function has a constant value. This surface is a p-dimensional ellipsoid that characterizes the joint confidence region of the parameter estimates. The semi-axes of the ellipsoid have magnitudes $c^2\lambda_i$, and have directions defined by $\boldsymbol{\tau}_i$, where λ_i and $\boldsymbol{\tau}_i$ are the eigenvalues and corresponding eigenvectors of $\mathbf{Q}_{\boldsymbol{\beta}}$ [Sorenson, 1980]. As an eigenvalue of $\mathbf{Q}_{\boldsymbol{\beta}}$ increases, the confidence region elongates in the direction of the corresponding eigenvector. In the limit, when an eigenvalue approaches infinity, the confidence region becomes unbounded,

rendering the parameters unobservable. Figure 2-1 shows the confidence region for a two-dimensional case.

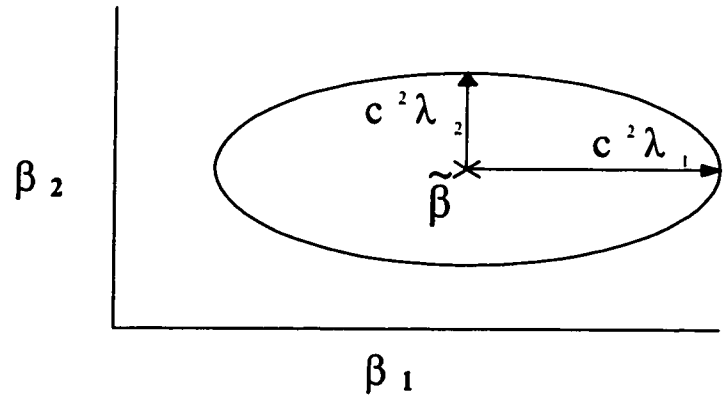


Figure 2-1: Parameter confidence region.

A common measure of confidence region volume is the determinant of the covariance matrix [Box and Lucas, 1959; Box and Hunter, 1965], since the determinant is the product of the eigenvalues. Large eigenvalues indicate a large confidence region, and give a large determinant. Shape is also important, and the condition number of $\mathbf{Q}_\beta$ may be used as an indicator of shape. The condition number is defined to be the ratio of the largest to smallest eigenvalue of the covariance matrix:

$$\kappa_{\tilde{\beta}} = \frac{\lambda_{\max}(\mathbf{Q}_\beta)}{\lambda_{\min}(\mathbf{Q}_\beta)} \tag{2.4}$$

A perfectly spherical confidence region would mean that the eigenvalues of $\mathbf{Q}_\beta$ are all equal and therefore the condition number would be unity. As the confidence region elongates, the difference between eigenvalues increases, causing the condition number to increase. Using these quality measures, Singh [1997] gives the following guidelines for choosing a parameter estimation scheme:

1. For 'strong' observability, small $|\mathbf{Q}_{\tilde{\beta}}|$ and $\kappa_{\tilde{\beta}}$ are required, where $\mathbf{Q}_{\tilde{\beta}}$ is the parameter covariance matrix estimated at $\tilde{\beta}$.

2. Among regions of similar volumes, the one yielding the smallest value for $\kappa_{\hat{\beta}}$ would be preferred, while among regions with similar conditioning the one encompassing the smallest volume (for a given confidence level) should be chosen.

2.2 Parameter Estimation Techniques

Given a set of measurements and a process model, one would like to estimate the model parameters using a scheme that minimizes the transmission of measurement noise through to the parameter estimates. The concept of parameter observability, and the degree of observability in particular, may be used to choose among parameter estimation techniques for RTO. According to Singh's [1997] guidelines, a good parameter estimation scheme is one that gives a small and/or spherical parameter confidence region. There are a number of methods used for estimating the unknown quantities in process models, and this section will present three methods: back-substitution, least squares, and error-in-variables estimation.

2.2.1 Back-Substitution

Back-substitution is the preferred practice in industrial RTO applications, because of its simplicity [Tsang, 1998]. To illustrate this method, consider the set of n independent equations:

$$\mathbf{f}(\boldsymbol{\beta}, \mathbf{z}) = \mathbf{0} \tag{2.5}$$

where $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown parameters to be determined, and $\mathbf{z} \in \mathbb{R}^m$ is a vector of measured process variables. To back-calculate the parameters, one requires p independent equations such that [Krishnamurthy and Sen, 1986]:

$$rank\left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right] = p \tag{2.6}$$

For the case where $n = p$, all the equations are used, and in the case where $n > p$, a subset of the available equations must be chosen.

When the equations are affine in the parameters, the system of independent equations may be written as:

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{b} \tag{2.7}$$

where $\mathbf{A} = \mathbf{g}_1(\mathbf{z})$ is a $p \times p$ invertible matrix, and $\mathbf{b} = \mathbf{g}_2(\mathbf{z})$ is a p-dimensional vector. The unique solution for the p parameters is simply [Householder, 1964]:

$$\hat{\boldsymbol{\beta}} = \mathbf{A}^{-1}\mathbf{b} \tag{2.8}$$

Generally, the equations will be nonlinear. There are methods generally available for solving nonlinear equations, which are iterative and have efficient local convergence [Krishnamurthy and Sen, 1986]. Global convergence cannot be guaranteed due to nonlinearity of the response surface. A common method for solving p nonlinear equations in p unknowns is Newton's method, which converges on the solution by linearizing the model equations about the current parameter estimate [Krishnamurthy and Sen, 1986].

2.2.2 Least Squares Estimation

In order to reduce the influence of random errors on the parameter estimates, it is useful to have a greater number of equations than unknowns, rendering the system over-determined [Dahlquist and Bjorck, 1974]. The unknown parameters are found by attempting to satisfy as many equations as possible, if only approximately.

The least squares model is usually formulated as [Bates and Watts, 1988] :

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\beta}, \mathbf{x}) + \boldsymbol{\epsilon} \tag{2.9}$$

where $\mathbf{y}$ are the measured *response* variables and $\mathbf{x}$ are the *independent* variables. The

term $\mathbf{f}(\boldsymbol{\beta}, \mathbf{x})$ is called the *expectation function*, and $\boldsymbol{\varepsilon}$ represents the measurement noise. The least squares estimate of the parameters, $\hat{\boldsymbol{\beta}}$, is the one that minimizes the sum of squares of the residuals:

$$\min_{\hat{\boldsymbol{\beta}}} \mathbf{e}^T \mathbf{e} \tag{2.10}$$

subject to:

$$\mathbf{y} = \mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{x}) + \mathbf{e}$$

where $\mathbf{e}$ represents the residual of the model. The least squares estimates are only appropriate when the model and the assumptions on the noise are valid. The assumptions involved in least squares estimation are [Box *et al.*, 1978]:

(i) The expectation function $\mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{x})$ provides an adequate fit of the data.

(ii) The model takes the form of Equation (2.9).

(iii) The values of the independent variables are perfectly known.

(iv) The measurement noise $\boldsymbol{\varepsilon}$ is independent of the expectation function.

(v) ε_i is distributed $N(0, \sigma_\varepsilon^2)$ $\forall$ i.

(vi) $cov(\varepsilon_i, \varepsilon_j) = 0$ $\forall$ $i \neq j$.

A possible least squares formulation for RTO follows the formulation of Box [1970], which uses the following model:

$$\mathbf{f}(\boldsymbol{\beta}, \mathbf{z}) = \boldsymbol{\varepsilon} \tag{2.11}$$

This formulation of the problem is used for two reasons. First, in RTO, the so-called response variables may not be measured directly, but are computed from other variables

that are actually measured. Secondly, the models represent relationships between a number of measured quantities that cannot be rigidly classified as dependent or independent variables: there is error associated with every variable.

The least squares solution is the value for $\hat{\boldsymbol{\beta}}$ such that $\mathbf{e}^T\mathbf{e}$ is a minimum, *i.e.*:

$$\min_{\hat{\boldsymbol{\beta}}} \mathbf{e}^T\mathbf{e} \tag{2.12}$$

subject to:

$$\mathbf{e} = \mathbf{f}(\hat{\boldsymbol{\beta}}, \mathbf{z})$$

When the equations are linear in the parameters, Equation (2.11) can be written as:

$$\mathbf{A}\boldsymbol{\beta} - \mathbf{b} = \boldsymbol{\epsilon} \tag{2.13}$$

where $\mathbf{A}$ is a $n \times p$ matrix, and $\mathbf{b}$ a n-dimensional vector, with $\mathbf{A}$ and $\mathbf{b}$ both functions of the measurements $\mathbf{z}$. Problem (2.12) is then formulated as:

$$\min_{\hat{\boldsymbol{\beta}}} \mathbf{e}^T\mathbf{e} \tag{2.14}$$

subject to:

$$\mathbf{e} = \mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{b}$$

The least squares solution is [Sorenson, 1980]:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} \tag{2.15}$$

which may be compared with the estimates found by back-substitution in Equation (2.8). Generally, rigorous process models are nonlinear in the parameters. Finding the least squares parameter estimates is difficult and requires iterative methods. A common method is the Gauss-Newton method, which iteratively converges on the solution by

linearizing the model about the current parameter estimate $\hat{\boldsymbol{\beta}}^{(k)}$, and moving to a new estimate $\hat{\boldsymbol{\beta}}^{(k+1)}$ that minimizes the sum of squares of the approximated residuals [Bates and Watts, 1988]:

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} + \boldsymbol{\delta}_G^{(k)} \tag{2.16}$$

where

$$\boldsymbol{\delta}_G^{(k)} = -\left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right]^{-1}_{\hat{\boldsymbol{\beta}}^{(k)}} \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T_{\hat{\boldsymbol{\beta}}^{(k)}} \mathbf{f}(\hat{\boldsymbol{\beta}}^{(k)}, \mathbf{z}) \tag{2.17}$$

is the Gauss-Newton increment.

Another iterative method that should be mentioned is the Newton-Raphson method [Sorenson, 1980]. This method is similar to the Gauss-Newton method except it uses a quadratic approximation to the objective function rather than first order approximation.

The simple least squares solution is appropriate when it can be assumed that all residual terms have the same variance. However, according to the model of Equation (2.11), the assumptions on the residual terms are no longer valid. Each residual term will depend on the errors associated with each measurement appearing in the particular equation. Therefore it cannot be assumed that these terms are independent, or that their variances are identical. In this situation, a weighted least squares approach should be used [Box *et al.*, 1978]. In the linear case, the expression for the parameters is [Sorenson, 1980]:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{A}^T\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{W}\mathbf{b} \tag{2.18}$$

where $\mathbf{W}$ is the weighting matrix. In order to obtain the best, linear, unbiased estimators (BLUE), the weighting matrix is chosen to be the inverse of the covariance matrix of the

equation residuals [Sorenson, 1980]:

$$\mathbf{W} = \mathbf{Q}_{\boldsymbol{\varepsilon}}^{-1} \tag{2.19}$$

The weights are inversely proportional to the variance terms, so a high variance means the corresponding relationship will be given less weight than one with a smaller variance. That is, the more likely an equation residual is near zero, the more influence it has on the solution.

2.2.3 Error-in-Variables

In RTO, the model parameters are estimated while all variables are subject to error. Therefore, the best treatment of the problem is a method that formally recognizes no distinction between dependent and independent variables, but treats all variables equally. This method is known as the error-in-variables method (EVM) [Britt and Luecke, 1973]. EVM provides both parameter estimates and reconciled data estimates that are consistent with the model [Kim *et al.*, 1990]. Deming [1943] introduced the original formulation of the problem of parameter estimation when there is error in all measured variables. Only approximate solutions could be made at that time. Although other researchers presented exact methods for fitting straight lines or polynomials [York, 1966; Williams, 1968; O'Neil *et al.*, 1969; Southwell, 1969], a general EVM algorithm was not presented until 1973, by Britt and Luecke. Further work has been done by Reilly and Patino-Leal [1981], Schwetlick and Tiller [1985], Valko and Vajda [1987] , and Kim *et al.* [1990]. A brief outline of Britt and Luecke's [1973] EVM algorithm is given below, further details can be found in Appendix A.

Britt and Luecke [1973] formulated the problem by considering the n-equation model:

$$\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta}) = \mathbf{0} \tag{2.20}$$

where $\mathbf{z}^*$ and $\boldsymbol{\beta}$ are the true values of the m measured variables and p parameters re-

spectively. The measurements $\mathbf{z}$ are corrupted by noise:

$$\mathbf{z} = \mathbf{z}^* + \boldsymbol{\varepsilon} \tag{2.21}$$

where $\boldsymbol{\varepsilon}$ represents the measurement error. In this formulation, both $\boldsymbol{\beta}$ and $\mathbf{z}^*$ are unknown quantities since $\mathbf{z}^*$ represents the true values of the measured variables. Therefore, in order to find a maximum likelihood estimate of $\boldsymbol{\beta}$, it is also necessary to simultaneously estimate $\mathbf{z}^*$. The following assumptions are made [Britt and Luecke, 1973]:

(i) $\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta})$ is twice continuously differentiable with respect to each argument.

(ii) The Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}$ has rank p, and the Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \mathbf{z}}$ has rank n.

(iii) The error $\boldsymbol{\varepsilon}$ is normally distributed having zero mean and a known positive definite covariance matrix $\mathbf{Q}_\varepsilon$.

The joint probability density function for the measurements is:

$$g(\mathbf{z}) = (2\pi)^{-m/2} |\mathbf{Q}_\varepsilon|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{z}^*)^T \mathbf{Q}_\varepsilon^{-1} (\mathbf{z} - \mathbf{z}^*)\right] \tag{2.22}$$

The likelihood function is found using the joint probability density function, by considering the measurements $\mathbf{z}$ as known, and the unknowns $(\mathbf{z}^*, \boldsymbol{\beta})$ as the variables:

$$l(\mathbf{z}^*, \boldsymbol{\beta}) = (2\pi)^{-m/2} |\mathbf{Q}_\varepsilon|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{z}^*)^T \mathbf{Q}_\varepsilon^{-1} (\mathbf{z} - \mathbf{z}^*)\right] \tag{2.23}$$

The unknowns $(\mathbf{z}^*, \boldsymbol{\beta})$ must obey the following constraint:

$$\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta}) = \mathbf{0} \tag{2.24}$$

Any solution of Equation (2.24) that maximizes the likelihood function with respect to all other solutions, is a maximum likelihood estimate of $(\mathbf{z}^*, \boldsymbol{\beta})$. Therefore the EVM

problem can be expressed as:

$$\min_{\mathbf{z}^*,\boldsymbol{\beta}} \frac{1}{2} (\mathbf{z} - \mathbf{z}^*)^T \mathbf{Q}_\epsilon^{-1} (\mathbf{z} - \mathbf{z}^*) \tag{2.25}$$

subject to:

$$\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta}) = \mathbf{0}$$

Using a Lagrange multiplier approach to solve the problem (see Appendix A for details), a system of $m + p + n$ nonlinear equations must be solved to obtain reconciled measurements and parameter estimates. Generally the number of measurements alone make the problem impractical, if not impossible to solve. Britt and Luecke [1973] therefore used a first order Taylor approximation of the constraint Equation (2.24) about the most recent estimate of $(\mathbf{z}^*, \boldsymbol{\beta})$, in an iterative scheme. The details of the development can be found in Appendix A. The final result is an iterative solution using the following equations:

$$\boldsymbol{\beta} - \boldsymbol{\beta}^{(k)} = -\left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T \left(\frac{\partial \mathbf{f}}{\partial \mathbf{z}} \mathbf{Q}_\epsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T\right)^{-1} \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right]^{-1} \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T \left(\frac{\partial \mathbf{f}}{\partial \mathbf{z}} \mathbf{Q}_\epsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T\right)^{-1} \cdot \left[\mathbf{f}\left(\mathbf{z}^{(k)}, \boldsymbol{\beta}^{(k)}\right) + \frac{\partial \mathbf{f}}{\partial \mathbf{z}} \left(\mathbf{z} - \mathbf{z}^{*(k)}\right)\right] \tag{2.26}$$

$$\mathbf{z}^* - \mathbf{z}^{*(k)} = \mathbf{z} - \mathbf{z}^{*(k)} - \mathbf{Q}_\epsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T \left(\frac{\partial \mathbf{f}}{\partial \mathbf{z}} \mathbf{Q}_\epsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T\right)^{-1} \cdot \left[\mathbf{f}\left(\mathbf{z}^{*(k)}, \boldsymbol{\beta}^{(k)}\right) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}} \left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(k)}\right) + \frac{\partial \mathbf{f}}{\partial \mathbf{z}} \left(\mathbf{z} - \mathbf{z}^{*(k)}\right)\right] \tag{2.27}$$

The method is:

1. Compute $(\mathbf{z}^*, \boldsymbol{\beta})$ using Equations (2.26) and (2.27).
2. Set $\left(\mathbf{z}^{*(k+1)}, \boldsymbol{\beta}^{(k+1)}\right) = (\mathbf{z}^*, \boldsymbol{\beta})$, relinearize the objective function about this new point, and find a new estimate $\left(\mathbf{z}^{*(k+2)}, \boldsymbol{\beta}^{(k+2)}\right)$.

3. Continue iterations until tolerances on $\left\|\beta^{(k+1)} - \beta^{(k)}\right\|$ and $\left\|\mathbf{z}^{*(k+1)} - \mathbf{z}^{*(k)}\right\|$ are met.

2.3 Robustness and Computational Requirements

For RTO, one desires a parameter estimation scheme that will give a small, spherical parameter confidence region, as well as minimum offset between the true parameter values and their expected values. Another important consideration is the computational requirements of the scheme, because some schemes may be too expensive to be practical, especially given the size of the RTO problem. The following discussion highlights some of the issues involved in each parameter estimation technique, and provides some modifications that have been made to the methods.

The first issue to be addressed is parameter error, which may result from both the model used for parameter estimation, and the numerical solution technique. Dahlquist and Bjorck [1974] have stated that the influence of random errors on the parameter estimates is reduced by having a greater number of equations than unknowns. The fewer equations, the larger the effect of random errors. Therefore, the effect of random measurement error will have less influence on the parameter estimates when a least squares or EVM technique is used instead of back-substitution (*i.e.*, when all n equations are used rather than just p equations).

Another benefit to having more equations than unknowns is the robustness to sensor malfunction. With back-substitution, if a sensor fails, the system of p equations has $p+1$ unknowns, and there is no longer a unique solution. With conventional least squares and EVM, the missing measurement becomes an additional unknown, and the parameters and missing measurement can be estimated simultaneously.

The second issue to be addressed is computational requirements. In general, RTO systems require the solution of very large parameter estimation problems. A complete ethylene plant model may contain 200,000 equations and variables with hundreds of

parameters to be estimated [White, 1997]. The computational efficiency of the parameter estimation technique is therefore extremely important. In EVM, the simultaneous estimation of parameters and true values for measured variables can yield very large problems that cannot be efficiently solved using current computing technology. Tjoa and Biegler [1992] indicate that the computational effort required to solve for both true measurement values and parameters in a nonlinear model often increases with the square or even with the cube of the number of equations. A number of modifications to EVM have been proposed. One approach is to nest the measurement correction step within the parameter estimation step [Reilly and Patino-Leal, 1981], which reduces the size of the estimation problem. Another approach is to separate the steps into a two-stage calculation [Schwetlick and Tiller, 1985; Kim *et al.*, 1990], again reducing computational requirements but there is no convergence guarantee. Therefore, it seems that EVM remains an impractical technique for use in typical RTO schemes. For this reason, it will not be considered further in this thesis. It may, however, become a useful technique once computing technology advances.

Back-substitution and least squares techniques appear to be the only practical alternatives for RTO. With respect to computational efficiency, it is clear that back-substitution is the least expensive to implement, since the number of equations involved is always equal to the number of parameters, and the solution is direct rather than an optimization. Least squares estimation, while giving more reliable parameter estimates, requires more computation than back-substitution. In the linear case, back-substitution requires p^3 (for the matrix inversion) plus p^2 (for the multiplication $\mathbf{A}^{-1}\mathbf{b}$) operations. The simple least squares method requires an additional $\frac{1}{2}np(p+3)$ operations in order to compute $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}^T\mathbf{b}$ [Dahlquist and Bjorck, 1969]. For the weighted least squares method, the computation of the weighting matrix will generally be too difficult to be practical.

When the model is nonlinear, much of the computation is involved with evaluating the Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \beta}$. A modification to Newton's method, to reduce this computation, is

to only update the Jacobian occasionally [Dahlquist and Bjorck, 1974], using:

$$\beta^{(k+1)} = \beta^{(k)} - \left[\frac{\partial \mathbf{f}}{\partial \beta}\right]^{-1}_{\beta=\beta^{(q)}} \mathbf{f}^{(k)} \tag{2.28}$$

where $k = q, ..., q+r$. Another modification for both back-substitution and least squares techniques is to use difference approximations to the derivatives rather than actual derivatives in the Jacobian [Dahlquist and Bjorck, 1974]:

$$\frac{\partial f_i(\beta)}{\partial \beta_j} \approx \frac{f_i(\beta + h_j \mathbf{e}_j) - f_i(\beta)}{h_j} \tag{2.29}$$

where $\mathbf{e}_j$ is the jth coordinate vector and $h_j \neq 0$, $i, j = 1, 2, ...p$. Substituting the approximate derivatives in for $\frac{\partial \mathbf{f}}{\partial \beta}$, the result is a p-dimensional discretized Newton's method. This method requires $\mathbf{f}(\beta)$ to be evaluated at $(p+1)$ points, so the amount of computation is similar to that of Newton's method if the evaluation of $f_i(\beta)$ takes as much work as that of $\frac{\partial f_i}{\partial \beta_j}$ [Dahlquist and Bjorck, 1974]. The trade-off in using difference approximations is poor accuracy, and the use of numerical derivatives usually increases the computing time required for convergence [Bates and Watts, 1988].

In summary, EVM seems to be an impractical and expensive parameter estimation technique for RTO, given the current state of computing technology. Among the practical alternatives, back-substitution appears to be the poorer choice with respect to parameter error and system reliability when sensors are subject to failure. In the next section, parameter estimation by back-substitution is compared to estimation by least squares in a series of case studies.

2.4 Williams-Otto Reactor Case Study

In this series of case studies, parameter estimation by back-substitution is compared to estimation by least squares under various situations. The first study compares the two methods during normal process operation, the second incorporates a bias into one of the

measurements, and the third considers the case where a sensor fails completely. It is shown that least squares is a superior method to back-substitution in all cases. In comparison to back-substitution, the least squares approach reduces the size of the parameter confidence region, and is more robust to measurement noise and bias. Furthermore, the method does not fail when a sensor fails.

2.4.1 Process Description

The Williams-Otto reactor [Williams and Otto, 1960] is modelled as a continuously stirred tank reactor, with temperature T_R and volume $v = 4640$ lb, in which three simultaneous reactions involving six chemical species are taking place:

$$\begin{aligned} A + B &\rightarrow C \\ B + C &\rightarrow P + E \\ C + P &\rightarrow G \end{aligned}$$

The reactions rate coefficients have Arrhenius temperature dependencies:

$$k_i = A_i \exp(-B_i/T_R) \tag{2.30}$$

where $i = 1, 2, 3$. The details of these reactions are given in Table 2.1. There are two entering flows, F_A and F_B, of pure reactants A and B respectively. The exit flow F_R contains all six species. There are ten measurable variables (flows, temperature, exit species concentrations), and the parameters updated on-line are the frequency factors A_i in Equation (2.30). The activation energies B_i are assumed to be known. The steady-state

Reaction	Frequency Factor (hr^{-1})	Activation Energy (R)	Basis
A + B $\rightarrow$ C	5.9755×10^9	12×10^3	lb of A
B + C $\rightarrow$ P + E	2.5962×10^{12}	15×10^3	lb of B
C + P $\rightarrow$ G	9.6283×10^{15}	20×10^3	lb of C

Table 2.1: Reaction data

model consists of seven independent mass balance equations - an overall flow balance plus six species balances:

$$F_R - F_A - F_B = 0 \tag{2.31}$$

$$F_A - vk_1 X_A X_B - F_R X_A = 0 \tag{2.32}$$

$$F_B - vk_1 X_A X_B - vk_2 X_B X_C - F_R X_B = 0 \tag{2.33}$$

$$2vk_2 X_B X_C - F_R X_E = 0 \tag{2.34}$$

$$vk_2 X_B X_C - \frac{1}{2} vk_3 X_C X_P - F_R X_P = 0 \tag{2.35}$$

$$2vk_1 X_A X_B - 2vk_2 X_B X_C - vk_3 X_C X_P - F_R X_C = 0 \tag{2.36}$$

$$1.5vk_3 X_C X_P - F_R X_G = 0 \tag{2.37}$$

2.4.2 Method

The first step in each case is to generate process data about a nominal operating point by adding random, normally distributed noise. The chosen nominal operating point and measurement covariance matrix can be found in Appendix B. One thousand sample data sets are used in each study.

Parameter estimates are obtained for each set of data. In order to estimate the three unknown parameters by back-substitution, three equations must be chosen from the seven available equations. Since Equation (2.31) does not contain any parameters it is not used. To minimize the transmission of measurement error to the parameters, a well-conditioned set of equations is desired. For this reason, Equations (2.32), (2.34), and (2.37) are selected. These equations form a linear, diagonal system in the parameters, meaning the equations are orthogonal and well-conditioned. To estimate the parameters by least squares, Equations (2.32) through (2.37) are used. Again, the flow balance is omitted because there are no parameters in that equation.

After completing the parameter estimation step, the results from back-substitution and least squares estimation are compared. Two measures are used to compare the

parameter estimates from the two methods. First, to measure the transmission of noise from the measurements to the parameters, the size of the parameter confidence regions are compared. A common method for quantifying the volume of a confidence region is simply the determinant of the covariance matrix of the data $|\mathbf{Q}_{\tilde{\beta}}|$ [Wald, 1943; Box and Lucas, 1959; Kiefer and Wolfowitz, 1959] . To measure the transmission of measurement bias to parameter bias, the sum of squared deviations between the true parameter values and the parameter estimates based on noise-free biased data, $\left\|\beta^* - \tilde{\beta}\right\|^2$, are compared.

In the first of the three studies, all process measurements are available and unbiased, to simulate normal process operation. For the second study, a 10% bias in the measurement of F_R is introduced, so that its nominal value reads 57,500 lb/hr rather than its true value of 52,500 lb/hr. The third study considers the case in which the F_R measurement fails completely.

2.4.3 Results

The problem was scaled such that the true values of the unknown parameters are:

$$\begin{aligned} A_1 &= 10 \\ A_2 &= 15 \\ A_3 &= 20 \end{aligned}$$

This allows for a more reliable solution, and effectively reduces the scale of the covariance matrix, its determinant, and the bias measure given below. The results are as follows:

Case	Method	$\lvert\mathbf{Q}_{\hat{\beta}}\rvert \times 10^{-5}$	$\left\lVert\beta^* - \bar{\beta}\right\rVert^2$
1	Back-substitution	2066	0.0037802
	Least Squares	25.01	0.0036608
2	Back-substitution	2997	5.5808
	Least Squares	25.45	0.55815
3	Back-substitution	fail	fail
	Least Squares	2.337	0.0036853

Table 2.2: Results Summary

2.4.4 Discussion

The results of this study show that least squares estimation outperforms back-substitution in all cases. In each case, the least squares method results in a substantially smaller confidence region for the parameters, giving the parameters a higher degree of observability as defined in Section 2.2. When a bias is introduced in the measurement of F_R (Case 2), the least squares method outperforms back-substitution by reducing the transmission of the bias to the parameters. Thus it is shown that, in this example, the least squares method not only reduces the effects of random process noise, it also reduces the impact of measurement bias caused by malfunction or miscalibration. When a sensor fails completely, as in Case 3, the back-substitution method also fails. The reason is the back-substitution model has three equations but four unknown quantities, and no unique solution exists. Least squares, on the other hand, has six equations in four unknowns, and a locally optimal set of parameter estimates can be found. It is possible, however, to use heuristics in a back-substitution scheme to select an alternate set of equations on-line.

It was pointed out in Chapter 1 that RTO is a closed-loop application, and the performance of the system will be adversely affected by the propagation of various noise sources through the loop. The motivation behind suggesting the use of a least squares

parameter estimation technique over the preferred method of back-substitution is to reduce the transmission of error through the model updating portion of the RTO loop. In this case study it is demonstrated that error transmission may be reduced by using a least squares technique instead of back-substitution, thus confining the parameter estimates to a smaller region and reducing the effects of measurement bias. One might therefore expect that these robust parameter estimates will lead to robust setpoints calculated during the optimization portion of the RTO loop. The first logical extension of this study, then, is to examine the transmission of noise from measurements through the RTO loop beyond parameter estimates, to calculated setpoints, since the setpoints will ultimately determine the plant profitability. Recognizing that the goal of RTO is increased profit, a second extension is to consider the relationships between profit and setpoint variance. A third direction of study may consider the selection of measurements used for the parameter update. It was shown in Case 3 that the determinant measure actually decreased an entire order of magnitude from the Case 1 results, for least squares estimation, while the bias measure increased only marginally. It is therefore questionable whether F_R should be measured at all. While adopting a least squares parameter estimation technique, these issues will be addressed in Chapter 3.

Chapter 3

Optimal Sensor Selection for Setpoint Determination

Since RTO is a closed-loop control system, error propagation is an important consideration in its design. Chapter 2 examined parameter estimation techniques, and found that least squares estimation is more effective than the preferred method of back-substitution with respect to filtering out the effects of process noise and measurement error. The case study of Section 2.4 indicated that with a least squares estimation, it may not be necessary to measure every possible variable in the process model.

This chapter continues the analysis of error propagation through to the setpoints determined by the optimizer. Given a least squares parameter estimation scheme, the goal is to find a sensor system that minimizes bias and variance in the calculated setpoints. The chapter begins with an overview of D- and T-optimal experimental design criteria and how they may be applied to sensor system selection in RTO. Recognizing the need to combine the two criteria, the second part of the chapter develops Sensor System Design Cost, which effectively balances D- and T-optimality criteria and allows the designer to compare various sensor systems with respect to uncaptured profit due to setpoint bias and variance. The final section of this chapter revisits the Williams-Otto reactor [1960], as a demonstration of how the Sensor System Design Cost may be applied in RTO design.

3.1 Optimal Design of Experiments and RTO

Since the parameter update portion of the RTO loop is a model-fitting procedure, DOE criteria may be used to determine which measurements should be taken to meet the chosen objectives. In this particular case, the desired objectives are to minimize bias and variance in the calculated setpoints, which result from the parameters estimated using process measurements.

In regression analysis, statistical models are developed to represent the observed relationship between experimental conditions and measured variables. In a typical situation, a practitioner may be faced with a number of candidate model representations. Furthermore, depending on the amount of noise and the availability of informative data, it may be difficult to estimate model parameters with certainty. In both cases, further experimentation must be performed.

Optimal DOE criteria have been developed for the analysis of regression models. D-optimal experimental design [Kiefer and Wolfowitz, 1959] has been developed to determine the experimental conditions that will most likely minimize the uncertainty in parameter estimates. T-optimal design [Atkinson and Federov, 1975] is used to decide which measurements to take such that one can discriminate between rival models, and is based on prediction error. In this section, the basic theory of D- and T-optimal design is presented, and the application of DOE theory to RTO is developed.

3.1.1 D-Optimal Design

The basis for D-optimal experimental design was first introduced by Wald [1943], and is a design used to minimize the uncertainty in the parameter estimates found when fitting a model to data. Consider the linear statistical model:

$$\mathbf{A}\boldsymbol{\beta} - \mathbf{b} = \boldsymbol{\varepsilon} \tag{3.1}$$

where the noise terms $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. For a linear regression model, the parameter estimates $\hat{\boldsymbol{\beta}}$ have a joint confidence region whose volume is proportional to $\left|\mathbf{A}^T\mathbf{A}\right|^{-1/2}$ [Box and Lucas, 1959]. The entries in $\mathbf{A}$ depend on the experimental design conditions: temperature, pressure, and so forth. A D-optimal design, $\mathbf{A}$, is a design that minimizes the volume of the parameter inference region. Equivalently, the design is chosen to maximize the D-optimal design criterion, $\left|\mathbf{A}^T\mathbf{A}\right|$, also called the determinant criterion. In a statistical sense, a D-optimal design is said to maximize the determinant of the information or, equivalently, minimize the generalized variance of the parameter estimates.

Box and Lucas [1959] extended the application of the determinant criterion to nonlinear models:

$$\mathbf{f}(\boldsymbol{\beta}, \mathbf{z}) = \boldsymbol{\varepsilon} \tag{3.2}$$

where $\boldsymbol{\beta}$ are the model parameters and $\mathbf{z}$ are the measured variables. Assuming the existence of a representative set of prior estimates of the parameters, $\hat{\boldsymbol{\beta}}$, one can write a first order Taylor series expansion of the model about $\hat{\boldsymbol{\beta}}$ and $E[\mathbf{z}] = \mathbf{z}_o$:

$$\boldsymbol{\varepsilon} = \mathbf{f}(\boldsymbol{\beta}, \mathbf{z}) \approx \left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right]_{\hat{\boldsymbol{\beta}}, \mathbf{z}_o} \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{z}}\right]_{\hat{\boldsymbol{\beta}}, \mathbf{z}_o} (\mathbf{z} - \mathbf{z}_o) \tag{3.3}$$

Equation (3.3) is equivalent to Equation (3.1) if:

$$\mathbf{A} = \left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right]_{\hat{\boldsymbol{\beta}}, \mathbf{z}_o} \tag{3.4}$$

$$\mathbf{b} = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{z}}\right]_{\hat{\boldsymbol{\beta}}, \mathbf{z}_o} (\mathbf{z} - \mathbf{z}_o) - \left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right]_{\hat{\boldsymbol{\beta}}, \mathbf{z}_o} \hat{\boldsymbol{\beta}} \tag{3.5}$$

Box and Lucas [1959] have demonstrated that a locally D-optimum design can be obtained by maximizing:

$$\left| \left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right]^T_{\hat{\boldsymbol{\beta}}, \mathbf{z}_o} \left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right]_{\hat{\boldsymbol{\beta}}, \mathbf{z}_o} \right| \tag{3.6}$$

The difficulty with this particular design is that D-optimality is lost as soon as a new design forces parameter estimates to move away from their nominal point. In practice, this phenomenon requires a sequential approach where locally D-optimal designs are iteratively generated and implemented [Box and Hunter, 1965].

3.1.2 T-Optimal Design

T-optimal experimental designs were introduced by Atkinson and Federov [1975], to provide designs that yield maximum discrimination between rival models. Consider the following problem, which in general need not be linear:

$$\mathbf{f}_t(\mathbf{z}, \boldsymbol{\beta}) = \boldsymbol{\varepsilon} \tag{3.7}$$

where $\mathbf{z}$ are the measured variables and $\boldsymbol{\beta}$ are unknown parameters. The noise terms $\boldsymbol{\varepsilon} \sim \mathbf{N}(0, \sigma^2\mathbf{I})$ are assumed to be independently and identically distributed. The function $\mathbf{f}_t(\mathbf{z}, \boldsymbol{\beta})$ is one of two known functions $\mathbf{f}_1(\mathbf{z}_1, \boldsymbol{\beta}_1)$ and $\mathbf{f}_2(\mathbf{z}_2, \boldsymbol{\beta}_2)$ where the unknown parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2 \in \mathbb{R}^{p_1}$ and $\mathbb{R}^{p_2}$ respectively. The purpose of the experiment is to determine which of the two models is true. Assuming that the first model is true, $\mathbf{f}_t(\mathbf{z}, \boldsymbol{\beta}) = \mathbf{f}_1(\mathbf{z}_1, \boldsymbol{\beta}_1)$, the T-optimal experiment is designed to maximize the sum of squares lack of fit of the second model [Atkinson and Federov, 1975]:

$$\sup_{\xi \in \Xi} \Delta(\xi) \tag{3.8}$$

where:

$$\Delta(\xi) = \inf_{\hat{\boldsymbol{\beta}}_2} \left\| \mathbf{f}_1(\mathbf{z}_1, \boldsymbol{\beta}_1) - \mathbf{f}_2\left(\mathbf{z}_2, \hat{\boldsymbol{\beta}}_2\right) \right\|_2^2 \tag{3.9}$$

In these expressions, ξ is the experimental design belonging to a class of designs Ξ.

For linear models $n\Delta(\xi)/\sigma^2$ is the noncentrality parameter of the χ^2 distribution of the residual sum of squares for the rival model. The T-optimum design provides the most

powerful F-test for lack of fit of the second model when the first is true. If the models are nonlinear in the parameters, the exact F-test is replaced by an asymptotic one [Atkinson and Federov, 1975].

3.1.3 DOE for RTO

Both D- and T-optimal designs may be applied to the measurement selection problem for setpoint determination in RTO. In this section, D- and T-optimal designs are extended to sensor selection for RTO. Further, weighted D- and T-optimal criteria are developed to include the effects of expected RTO profitability. The section concludes with a discussion of the calculation involved and the methods available for performing these calculations.

In conventional D-optimal design, the experimenter can choose the variables to measure, their settings, and the number of replicates of each variable, in order to minimize the uncertainty in the parameters in a given model. In the current RTO design, however, the settings for the variables cannot be chosen, and replicate measurements are not taken. Therefore, in this thesis, the sole decision to be made is the best choice of model variables to measure. The goal is to choose a measurement set, among the possible alternatives, which minimizes the variance associated with the calculated setpoints (confidence region volume).

This problem can be easily re-stated as a D-optimal design problem that aims to minimize the volume of the setpoint confidence region. The D-optimal design criterion requires the selection of a set of measurements that minimize the determinant of the setpoint covariance matrix, $|\mathbf{Q}_{\mathbf{x}^\bullet}|$. In order to use the proposed D-optimal criterion for sensor selection, the setpoint covariance matrix must be approximated. A method for approximating $\mathbf{Q}_{\mathbf{x}^\bullet}$ can be found in Appendix C.

In RTO, it is desirable for the process models to be rigorous, first-principles models. These more rigorous models permit more accurate coverage of a much wider operating region than is possible with empirical models [White, 1997]. Rigorous models contain a large number of process variables, some of which can be measured. There are two

obvious reasons why certain variables may not be measured. First, it may be physically impossible to install the required sensor. Second, the number of sensors that can be purchased may be limited by a budget. For these reasons, RTO designers may require an alternate, simpler model, so that the desired parameters can be estimated using the available measurements. The T-optimal sensor system for RTO is the one that corresponds to the model giving a minimum sum of squared deviations between the nominal setpoints $\tilde{\mathbf{x}}^*$ found with this sensor system, and the setpoints $\mathbf{x}_m^*$ determined using the most rigorous model and a complete and perfectly known set of measurements. The problem is stated as:

$$\min_{\varkappa} \|\mathbf{x}_m^* - \tilde{\mathbf{x}}_{\varkappa}^*\|_2^2 \tag{3.10}$$

where $\varkappa$ is the set of potential sensor systems.

Weighting

The D- or T-optimal criteria ensure that the volume of the confidence region of the setpoints, or the bias in the setpoints, is minimized. However, the topology of the profit surface may cause setpoint deviations in one direction to be much more costly than deviations in another direction. Since the primary concern of RTO is economics and not necessarily setpoint deviation, it would be desirable to weight the two criteria such that variance or bias in the more costly directions is penalized more than variance or bias in the less costly directions. In this way, a weighted D-optimal or T-optimal criterion would aim to minimize lost dollars caused by setpoint variance or bias, rather than minimize setpoint variance or bias directly.

This weighting may be accomplished through the use of the reduced Hessian of the profit function to be maximized, $\nabla_r^2 P$. The reduced Hessian is the symmetric matrix of second derivatives of the profit function $P(\mathbf{x}^*)$, in the space of the setpoints used for

optimization [Edgar and Himmelblau, 1988]. In a two-dimensional case:

$$\nabla_r^2 P = \begin{bmatrix} \frac{\partial^2 P}{\partial x_1^2} & \frac{\partial^2 P}{\partial x_1 \partial x_2} \\ \frac{\partial^2 P}{\partial x_2 \partial x_1} & \frac{\partial^2 P}{\partial x_2^2} \end{bmatrix} \tag{3.11}$$

Generally, the RTO problem is formulated as:

$$\max_{\mathbf{x}^*} P(\mathbf{x}^*) \tag{3.12}$$

subject to:

$$\mathbf{f}(\boldsymbol{\beta}, \mathbf{z}) = \mathbf{0}$$

$$\mathbf{h}(\boldsymbol{\beta}, \mathbf{z}) \leq \mathbf{0}$$

Since this is a maximization problem the reduced Hessian will be negative definite at the optimum, and therefore the negative reduced Hessian is positive definite.

The goal of the weighting procedure is to transform the setpoint variables to a coordinate system based on dollars, so that in the new coordinates, variance or bias in any direction has the same cost associated with it. Assuming that, in some small neighbourhood about the nominal setpoint values, the setpoints follow a multivariate normal distribution, the following linear transformation can be made:

$$\boldsymbol{\eta} = \mathbf{R}\mathbf{x}^* \tag{3.13}$$

where:

$$-\nabla_r^2 P = \mathbf{R}^T \mathbf{R} \tag{3.14}$$

is the Cholesky decomposition of the positive definite reduced Hessian, $-\nabla_r^2 P$. Since $\mathbf{x}^*$ is assumed to be $N(\mathbf{x}_m^*, \mathbf{Q}_{\mathbf{x}^*})$ in some small neighbourhood about $\mathbf{x}_m^*$, the linear

transformation of Equation (3.13) means that [Chatfield and Collins, 1980]:

$$\boldsymbol{\eta} \sim \mathbf{N}\left(\mathbf{R}\mathbf{x}_m^*, \mathbf{R}\mathbf{Q}_{\mathbf{x}^*}\mathbf{R}^T\right) \tag{3.15}$$

The D-optimal criterion can then be applied to the covariance matrix of $\boldsymbol{\eta}$:

$$\min \left|\mathbf{R}\mathbf{Q}_{\mathbf{x}^*}\mathbf{R}^T\right| \tag{3.16}$$

Equation (3.16) is the weighted D-optimal criterion. The effect of the weighting can be seen in the following simple example.

Example 1 *Suppose a setpoint covariance matrix:*

$$\mathbf{Q}_{\mathbf{x}^*} = \begin{bmatrix} 15 & 0 \\ 0 & 4 \end{bmatrix}$$

Let the negative reduced Hessian of the profit function be:

$$-\nabla_r^2 P = \begin{bmatrix} 2 & 0 \\ 0 & 64 \end{bmatrix} = \begin{bmatrix} 1.414 & 0 \\ 0 & 8 \end{bmatrix}^T \begin{bmatrix} 1.414 & 0 \\ 0 & 8 \end{bmatrix}$$

In the unweighted problem the x_1^* variance contributes the most to the determinant, as shown in the $\mathbf{Q}_{\mathbf{x}^*}$ matrix. Any minimization of the determinant would primarily aim to reduce the variance associated with this variable. The negative Hessian of the profit surface, however, indicates that variance in the x_2^* variable causes a much larger loss of profit than variance in the x_1^* variable. It would be beneficial to weight the optimization problem to recognize this. Using Equation (3.16), the weighted covariance matrix becomes:

$$\mathbf{R}\mathbf{Q}_{\mathbf{x}^*}\mathbf{R}^T = \begin{bmatrix} 1.414 & 0 \\ 0 & 8 \end{bmatrix} \begin{bmatrix} 15 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1.414 & 0 \\ 0 & 8 \end{bmatrix}^T = \begin{bmatrix} 30 & 0 \\ 0 & 256 \end{bmatrix}$$

With this transformation, variance in the x_2^* variable contributes the most to the deter-

minant, and hence an optimization would aim to reduce the variance in x_2^*, rather than x_1^*, since variance in x_2^* is more costly.

A weighted T-optimal criterion can be developed in a similar fashion. Using the transformation of Equation (3.13), the weighted bias in the new variable $\boldsymbol{\eta}$ is written as:

$$(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_m) = \mathbf{R}\,(\tilde{\mathbf{x}}^* - \mathbf{x}_m^*) \tag{3.17}$$

The weighted sum of squared bias is:

$$\begin{aligned} \|\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}_m\|_2^2 &= \left|(\tilde{\mathbf{x}}^* - \mathbf{x}_m^*)^T\,\mathbf{R}^T\mathbf{R}\,(\tilde{\mathbf{x}}^* - \mathbf{x}_m^*)\right| \\ &= \left|(\tilde{\mathbf{x}}^* - \mathbf{x}_m^*)^T\left(-\nabla^2 P\,|_{\mathbf{x}_m^*}\right)(\tilde{\mathbf{x}}^* - \mathbf{x}_m^*)\right| \end{aligned} \tag{3.18}$$

and the weighted T-optimal criterion is:

$$\min_{x} \left|(\tilde{\mathbf{x}}^* - \mathbf{x}_m^*)^T\left(-\nabla^2 P\,|_{\mathbf{x}_m^*}\right)(\tilde{\mathbf{x}}^* - \mathbf{x}_m^*)\right| \tag{3.19}$$

3.1.4 Discussion

D- and T-optimality criteria have been developed to apply to the sensor selection issue within the RTO system, with respect to setpoint determination. The D-optimal criterion can be used to find the sensor system that will minimize the uncertainty in the calculated setpoints, while the T-optimal criterion may be used to find the system that gives the least amount of bias between the nominal calculated setpoints (using the chosen sensor system), and those that would be calculated on a complete and perfectly known measurement set. Further, both criteria may be modified to exploit the geometry of the problem using the reduced Hessian of the profit function. These optimality criteria, however, will not necessarily give the same results. Since both objectives are important, a compromise must be made between the D- and T-optimality criteria. The evaluation of Sensor Network Design Cost will provide this function.

3.2 Sensor System Design Cost

The Sensor System Design Cost (SSDC) is defined to be the expected lost profit due to deviations between calculated setpoints $\mathbf{x}_m^*$ and $\mathbf{x}^*$:

$$C = E\left[P(\mathbf{x}_m^*) - P(\mathbf{x}^*)\right] \tag{3.20}$$

where E is the expectation operator, and $P(\mathbf{x}^*)$ represents the plant profit at setpoints $\mathbf{x}^*$. The setpoints $\mathbf{x}_m^*$ are those that would be obtained through a rigorous process model and a complete, perfectly known measurement set $\mathbf{z}_c^*$. The setpoints $\mathbf{x}^*$ are determined through a particular model that is based upon a particular sensor system with measurements $\mathbf{z}$. There will be two types of lost profit. The first is due to variance, which exists because of random process noise. The second type of lost profit is bias cost, which exists because of deviations between the predictions based on the rigorous model and those based on the alternate model. Figure 3-1 [Forbes, 1994] depicts the bias and variance cost associated with the setpoint predictions. In this section, the SSDC criterion is developed based on statistical principles. Further, the SSDC criterion combines the D- and T-optimal criteria into one diagnostic tool for sensor selection.

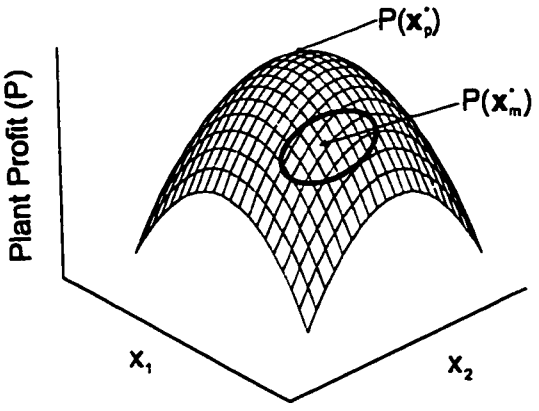


Figure 3-1: Bias and variance cost associated with setpoint predictions.

The development of the SSDC follows the Design Cost work of Forbes and Marlin [1996]. The primary difference in this work is the assumption that the rigorous process model closely approximates the true plant, and hence identifies a reasonable approximation of the true plant optimal operation. Therefore, this optimum is taken to be the benchmark to which all other predicted setpoints, possibly from simpler RTO models, are compared. Any deviation in setpoints, and hence loss of profit, from the benchmark optimal operation is a result of model simplification due to the elimination of measurements, and random process noise, as well as plant/model mismatch. The detailed development may be found in Appendix D. The final SSDC equation is:

$$C = -\frac{1}{2}\left[\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right) + \mathbf{w}^T\left[\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right]\mathbf{w}\right] \tag{3.21}$$

where $\tilde{\mathbf{x}}^*$ are the nominal setpoints found with a particular sensor system, $\nabla_r^2 P$ is the reduced Hessian of the profit function, $\mathbf{Q}_{\mathbf{x}^*}$ is the setpoint covariance matrix, and $\mathbf{w} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$.

The goal of sensor selection is to minimize the design cost with respect to the available sensor system alternatives $\varkappa$:

$$\min_{\varkappa} -\frac{1}{2}\left[\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right) + \mathbf{w}^T\left[\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right]\mathbf{w}\right] \tag{3.22}$$

which may be rewritten as:

$$\min_{\varkappa} \left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)^T\left(-\nabla_r^2 P\,|_{\mathbf{x}_m^*}\right)\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right) + \mathbf{w}^T\left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right]\mathbf{w} \tag{3.23}$$

Since $-\nabla_r^2 P$ and $\mathbf{Q}_{\mathbf{x}^*}$ are positive definite by definition, Equation (3.23) is equivalent to:

$$\min_{\varkappa} \left|\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)^T\left(-\nabla_r^2 P\,|_{\mathbf{x}_m^*}\right)\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)\right| + \left|\mathbf{w}^T\left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right]\mathbf{w}\right| \tag{3.24}$$

Problem (3.24) includes the Hessian of the model-based profit function. Although it may be possible to evaluate the Hessian, the profit surface curvature of the model may

not accurately represent the true plant profit behaviour. In this case, Problem (3.24) can be modified so that the Hessian is removed and an upper bound on the SSDC is minimized instead. Using the Cauchy-Schwarz Inequality [Ortega, 1987], upper bounds may be placed on each of the two terms in Problem (3.24):

$$\left|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right)(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\right| \leq \left\|-\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right\|_2 \|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\|_2^2 \tag{3.25}$$

and

$$\left|\mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^\bullet}\right] \mathbf{w}\right| \leq s \left\|-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^\bullet}\right\|_2 \leq s \left\|\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right\|_2 \|\mathbf{Q}_{\mathbf{x}^\bullet}\|_2 \tag{3.26}$$

where s is the number of setpoints. The upper bound on the SSDC is therefore:

$$C \leq \left\|\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right\|_2 \left(\|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\|_2^2 + s \|\mathbf{Q}_{\mathbf{x}^\bullet}\|_2\right) \tag{3.27}$$

The minimization of Equation (3.24) is now expressed in terms of upper bounds:

$$\min_{\mathbf{x}} \left\|-\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right\|_2 \left(\|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\|_2^2 + s \|\mathbf{Q}_{\mathbf{x}^\bullet}\|_2\right) \tag{3.28}$$

Since $\left\|-\nabla_r^2 P\right\|_2$ evaluated at $\mathbf{x}_m^*$ is a constant, positive, scalar quantity, Equation (3.28) reduces to:

$$\min_{\mathbf{x}} \|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\|_2^2 + s \|\mathbf{Q}_{\mathbf{x}^\bullet}\|_2 \tag{3.29}$$

Problem (3.29) is the unweighted point-wise criterion for sensor system selection. It can be seen immediately that the first term in Problem (3.29) is the T-optimal criterion of Problem (3.10). The second term in Problem (3.29) contains the largest eigenvalue of $\mathbf{Q}_{\mathbf{x}^\bullet}$, which is proportional to the length of the largest semi-axis of the setpoint confidence region. It is also the largest contributor to the D-optimal criterion. Since all of the eigenvalues of $\mathbf{Q}_{\mathbf{x}^\bullet}$ are less than or equal to $\|\mathbf{Q}_{\mathbf{x}^\bullet}\|_2$, Problem (3.29) aims to minimize

the sum of the T-optimal criterion and the worst case D-optimal criterion.

The SSDC Problems (3.24) and (3.29) represent sensor selection at a given operating point. The model-based plant optimum $\mathbf{x}_m^*$ will change however, depending on the values of the measured variables $\mathbf{z}_c^*$. So, the model optimum is drawn from a space $\mathbf{S}$ of all possible optima. There is a frequency function $\varsigma(\mathbf{z}_c^*, \boldsymbol{\upsilon})$ associated with the optima in $\mathbf{S}$, which describes the occurrence rate of a particular plant optimum. The total uncaptured model-based profit for a particular sensor system, for all possible disturbances, is:

$$C_T = \int \left[E\left[P\left(\mathbf{x}_m^*\right)\right] - \int P(\boldsymbol{\delta})\, f\left(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) d\Psi \right] \varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right) d\mathbf{S} \tag{3.30}$$

where P and ς are both Lebesgue integrable on $\mathbf{S}$, and $f\left(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right)$ is the probability density function associated with the predicted setpoints $\mathbf{x}^*$. The maximum theoretically attainable model-based plant profit is:

$$P_T = \int P\left(\mathbf{x}_m^*\right) \varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right) d\mathbf{S} \tag{3.31}$$

So Equation (3.30) can be re-written as:

$$C_T = P_T - \int \int P(\boldsymbol{\delta})\, f\left(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) \varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right) d\Psi d\mathbf{S} \tag{3.32}$$

The goal is to minimize the total loss in profit by selecting the appropriate sensor system from the set of choices $\varkappa$. This is equivalent to:

$$\min_{\varkappa} - \int \int P(\boldsymbol{\delta})\, f\left(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) \varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right) d\Psi d\mathbf{S} \tag{3.33}$$

It is unlikely that $\varsigma(\mathbf{z}_c^*, \boldsymbol{\upsilon})$ would be known, thus the minimization of Problem (3.33) will have to be approximated by a sum over an expected set of values for $\mathbf{z}_c^*$.

If changes in the measurements are considered to occur as steps, Problem (3.33)

becomes:

$$\min_{\varkappa} \sum_{\mathbf{z}_c^*} \left[- \int P(\boldsymbol{\delta})\, f(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*})\, d\Psi \right] \varsigma(\mathbf{z}_c^*) \tag{3.34}$$

The objective in Problem (3.34) is to minimize the weighted sum of the solutions to the point-wise SSDC Problem (3.29), for the set of possible measurement values. Then using the developments for the point-wise design cost problem, the total weighted SSDC problem is:

$$\min_{\varkappa} \sum_{\mathbf{z}_c^*} \left[\left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right| + \left| \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \right] \varsigma(\mathbf{z}_c^*) \tag{3.35}$$

The unweighted SSDC is:

$$\min_{\varkappa} \sum_{\mathbf{z}_c^*} \left[\| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \|_2^2 + s \, \| \mathbf{Q}_{\mathbf{x}^*} \|_2 \right] \varsigma(\mathbf{z}_c^*) \tag{3.36}$$

The solutions to Problems (3.35) and (3.36) provide the minimum weighted and unweighted SSDC for the expected set of measurable variable values. The RTO designer can then compare sensor system alternatives for the complete range of operations, rather than one point at a time.

3.2.1 Discussion

Sensor System Design Cost was developed to find a criterion that could be used to select a sensor system among alternatives, with the goal of simultaneously minimizing both variance and bias costs. The result is a minimization problem that combines both D-optimal and T-optimal experimental design criteria, as developed in Section 3.2 for setpoint determination. Consider the point-wise Problem (3.24):

$$\min_{\varkappa} \left\{ \left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right| + \left| \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \right\}$$

The first term in this expression is the weighted T-optimal criterion from Problem (3.19). The upper limit on the second term is simply a multiple of the upper limit of the weighted D-optimal criterion. Recall the weighted D-optimal criterion:

$$\min \left| \mathbf{R}\mathbf{Q}_{\mathbf{x}^*}\mathbf{R}^T \right| \tag{3.37}$$

where:

$$-\nabla_r^2 P = \mathbf{R}^T\mathbf{R} \tag{3.38}$$

is the Cholesky decomposition of the negative reduced Hessian of the profit function. The upper bound on the determinant of $\mathbf{R}\mathbf{Q}_{\mathbf{x}^*}\mathbf{R}^T$ is:

$$\left| \mathbf{R}\mathbf{Q}_{\mathbf{x}^*}\mathbf{R}^T \right| \leq \|\mathbf{R}\|_2 \|\mathbf{Q}_{\mathbf{x}^*}\|_2 \left\|\mathbf{R}^T\right\|_2 = \left\| -\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right\|_2 \|\mathbf{Q}_{\mathbf{x}^*}\|_2 \tag{3.39}$$

while the upper bound on the second term of Problem (3.24) is:

$$\left| \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \leq s \left\| -\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right\|_2 \|\mathbf{Q}_{\mathbf{x}^*}\|_2 \tag{3.40}$$

Therefore the weighted SSDC aims to minimize the sum of the weighted T-optimal criterion and the worst case weighted D-optimal criterion. That is, it aims to minimize the lost dollars caused by both setpoint variance and bias.

If the profit surface curvature of the model does not accurately represent the true plant profit behaviour, the reduced Hessian can be removed from the problem as in Problem (3.29):

$$\min_{\mathbf{x}} \left\| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right\|_2^2 + s \|\mathbf{Q}_{\mathbf{x}^*}\|_2$$

Again the first term is the T-optimal criterion, while the second term is a multiple of the largest eigenvalue of $\mathbf{Q}_{\mathbf{x}^*}$. Hence the SSDC aims to minimize bias and worst case

variance in the setpoints.

There are two important issues to be recognized if the SSDC criterion is to be used for sensor selection. First, in order to use the weighted SSDC and minimize lost profit due to variance and bias, the reduced Hessian of the profit function must adequately describe the curvature of the plant profit surface. Secondly, the SSDC criterion includes an upper bound on the determinant criterion, rather than the determinant criterion directly. Therefore, it aims to minimize the worst possible volume of the confidence region rather than the actual volume.

The reduced Hessian of the profit function describes the local curvature of the profit surface, with respect to the setpoint variables. If the true plant profit curvature is not accurately represented by the Hessian, the weighted SSDC results may be inaccurate. An alternate approach is to remove the Hessian from the problem, and use the unweighted SSDC. Note that when using the unweighted SSDC criterion, the goal is to minimize variance and bias in the setpoints directly, rather than lost dollars due to variance and bias.

The SSDC includes the largest eigenvalue of the setpoint covariance matrix, which is representative of the largest possible determinant rather than the actual determinant. Therefore the optimization aims to minimize the largest eigenvalue, ignoring the remaining eigenvalues. The true volume of the confidence region is thus ignored. By minimizing the largest eigenvalue, it is possible that the entire confidence region is becoming smaller. But it is also possible that while the longest axis is becoming shorter, the shorter axes are becoming longer. So, by decreasing the largest eigenvalue, the volume of the confidence region is not necessarily decreasing. It would be beneficial to rework the SSDC so that the actual determinant of the setpoint covariance matrix can be used.

3.3 Sensor Selection Procedure for RTO

The developments of this chapter lead to a structured method for choosing sensors for RTO. This method can be broken down into the following steps:

1. Determine the process model, adjustable parameters, and the possible sensor system alternatives. The parameters to be updated on-line must be observable through the sensor systems, process model and chosen updating scheme.

2. Compare the performance of each alternative sensor system by calculating the SSDC for each. This will involve:

   (a) approximation (or calculation) of the following sensitivity matrices: $\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$, $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}}$, and $\frac{\partial \mathbf{x}}{\partial \boldsymbol{\beta}}$.

   (b) knowledge of measurement covariance matrix $\mathbf{Q}_{\mathbf{z}}$.

   (c) approximation (or calculation) of the reduced Hessian of the profit function, $\nabla_r^2 P$.

   (d) evaluation of the expected setpoint bias for each sensor system, using the process model and the nominal values of the process measurements.

The case study of the next section demonstrates the use of the SSDC method for sensor selection in the Williams-Otto reactor [Williams and Otto, 1960].

3.4 Williams-Otto Reactor Case Study

This case study revisits the Williams-Otto Reactor [Williams and Otto, 1960] that was studied in Section 2.4. Chapter 2 concluded with the suggestion that it may not be necessary to measure every model variable in order to achieve good RTO results. In particular, it was shown in Section 2.4 that for a least squares parameter estimation, the measurement of the variable F_R adversely affected the quality of the parameter estimates.

In this chapter, the Sensor System Design Cost criterion was developed for the purposes of evaluating various sensor systems with respect to the quality of the setpoints each system generates. Further, the concept of weighting the criterion was introduced, so that each sensor system may be evaluated in terms of uncaptured profit caused by implementing each system. The purpose of this study is to find the optimal sensor system with respect to calculated setpoints, based on the weighted and unweighted SSDC criteria.

3.4.1 Process Description

The process and system model are the same as described in Section 2.4. The available sensors are:

1. Flow of feed A to the reactor.
2. Flow of feed B to the reactor.
3. Flow of reactor exit stream.
4. Reactor temperature.
5. Weight fraction of component G in the reactor exit stream.
6. Weight fractions of components A, B, E, P and C in the reactor exit stream, obtained simultaneously.

Measurement noise is given in Appendix B. The profit function to be maximized is:

$$P = (0.30X_P + 0.0068X_E)\,F_R - 0.03F_B - 0.02F_A \tag{3.41}$$

This function assumes that products P and E may be sold at \$0.30/lb and \$0.0068/lb respectively, and the cost of feeds A and B are \$0.02/lb and \$0.03/lb respectively [Williams and Otto, 1960].

3.4.2 Method

The first task is to find the sensor system alternatives. First consider the parameter estimation portion of the RTO loop. The model consists of seven equations, containing ten measurable variables and three parameters to be estimated (A_1, A_2, and A_3). Assuming linear independence of the model equations, at least six of the ten measurable variables must actually be measured in order to perform a parameter update by back-substitution or least squares. Now consider the optimization portion of the RTO loop. Again, there are seven model equations, and ten process variables. As in Williams and Otto [1960], F_A is fixed. Therefore, for the system to be completely determined, two setpoint variables must be chosen from the remaining nine. That is, once the values for these two setpoints are chosen, the remaining process variables in the steady-state model can be uniquely determined. A common choice for setpoint variables are feed flow (F_B) and reactor temperature (T_R), since they are more easily measured and controlled than component weight fractions. It will be assumed that the setpoint variables must be measured in order for them to be controlled.

To summarize, the sensor system choices must contain at least six variables, including F_B and T_R. Further, the measurements of X_A, X_B, X_E, X_P, and X_C occur simultaneously in one sensor. There are eight possible sensor systems, given in Table 3.1.

Sensor System, $\varkappa$	Measured Variables, $\mathbf{z}$
1	F_A F_B F_R T_R X_G X_A X_B X_C X_E X_P
2	F_B F_R T_R X_G X_A X_B X_C X_E X_P
3	F_A F_B T_R X_G X_A X_B X_C X_E X_P
4	F_A F_B F_R T_R X_A X_B X_C X_E X_P
5	F_B T_R X_G X_A X_B X_C X_E X_P
6	F_B F_R T_R X_A X_B X_C X_E X_P
7	F_A F_B T_R X_A X_B X_C X_E X_P
8	F_B T_R X_A X_B X_C X_E X_P

Table 3.1: Sensor System Alternatives

The unweighted point-wise design cost problem is:

$$\min_{\mathbf{x}} \left\{ \left\| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right\|_2^2 + s \left\| \mathbf{Q}_{\mathbf{x}^*} \right\|_2 \right\} \tag{3.42}$$

and the weighted point-wise design cost problem is:

$$\min_{\mathbf{x}} \left\{ \left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right| + \left| \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \right\} \tag{3.43}$$

The design cost will be calculated for the sensor system design alternatives, and the system that gives the lowest design cost will be chosen as the optimal design. In order to calculate the SSDC, $\mathbf{x}_m^*$, $\tilde{\mathbf{x}}^*$, $\mathbf{Q}_{\mathbf{x}^*}$, and $-\nabla_r^2 P$ must be calculated. The following methods were used:

1. To calculate $\mathbf{x}_m^*$, the known true values of the parameters were used to find the optimal setpoints by minimizing Equation (3.41), subject to the model equations. These setpoints were used to find the nominal values of the measurable variables at the optimum, $\mathbf{z}_c^*$.

2. To calculate $\tilde{\mathbf{x}}^*$, the noise-free measurements $\mathbf{z}^*$ were used. A least squares parameter estimation was performed, using all model equations, and only those measurements corresponding to the particular sensor system under consideration. These parameter were used to find the optimal setpoints $\tilde{\mathbf{x}}^*$ by minimizing Equation (3.41) subject to the model equations.

3. To linearly approximate the closed-loop $\mathbf{Q}_{\mathbf{x}^*}$, Equation (C.1) of Appendix C was used [Forbes, 1994]:

$$\mathbf{Q}_{\mathbf{x}^*} = \sum_{i=0}^{\infty} \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right)^i_{\mathbf{x}^*, \boldsymbol{\beta}^*, \mathbf{z}^*} \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \right)_{\boldsymbol{\beta}^*, \mathbf{z}^*} \mathbf{Q}_{\mathbf{z}} \cdot \left[\left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right)^i_{\mathbf{x}^*, \boldsymbol{\beta}^*, \mathbf{z}^*} \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \right)_{\boldsymbol{\beta}^*, \mathbf{z}^*} \right]^T \tag{3.44}$$

In this equation, $\mathbf{Q}_{\mathbf{x}^*}$ is expressed as an infinite weighted sum in the powers of the products of RTO subsystem sensitivities, as well as the sensitivities of process variables to setpoint changes. Although this is an infinite sequence, it was possible to approximate $\mathbf{Q}_{\mathbf{x}^*}$ with just one term, each element having an accuracy of five significant digits. Since the model equations are linear in the parameters, and the problem is small, all the required derivatives were computed analytically using the methods of Appendix C.

4. To calculate $-\nabla_r^2 P$, the analytical method of Ganesh and Biegler [1987] was used.

The above calculations give an approximate design cost. Because of nonlinearity in the process, particularly the exponential temperature dependence in the reaction equations, there will be mismatch between the approximate $\mathbf{Q}_{\mathbf{x}^*}$ and the $\mathbf{Q}_{\mathbf{x}^*}$ found through process simulation. Therefore, for comparison, the SSDC for the available sensor system choices was also found through simulation. For each simulation, normally distributed random noise was added to the measurements, according to the variances and covariances given in Appendix B. One thousand sets of setpoints were generated, closed-loop, from which the covariance matrix was calculated.

3.4.3 Results

The known true values of the (scaled) parameters are:

$$\boldsymbol{\beta} = \begin{bmatrix} 10 & 15 & 20 \end{bmatrix}^T$$

where $\boldsymbol{\beta} = \begin{bmatrix} A_1 & A_2 & A_3 \end{bmatrix}^T$. The optimal setpoints found using Equation (3.41):

$$\mathbf{x}_m^* = \begin{bmatrix} 0.37987 & 0.65314 \end{bmatrix}^T$$

where $\mathbf{x} = \begin{bmatrix} F_B & T_R \end{bmatrix}^T$. Note that all flows are scaled by a factor of 10^{-5} and the temperature by 10^{-3}. The nominal values for the measurable variables, which result

from the implementation of these optimal setpoints are:

$$\mathbf{z}_c^* = [\ 0.145 \quad 0.37987 \quad 0.52487 \quad 0.65314 \quad 0.10754$$
$$0.087449 \quad 0.38963 \quad 0.015306 \quad 0.29062 \quad 0.10946\ ]^T$$

where $\mathbf{z} = \left[\ F_A \quad F_B \quad F_R \quad T_R \quad X_G \quad X_A \quad X_B \quad X_C \quad X_E \quad X_P\ \right]^T$.

At the optimum, the negative reduced Hessian was evaluated as:

$$-\nabla_r^2 P = \begin{bmatrix} 0.14684 & -0.45627 \\ -0.45627 & 3.3913 \end{bmatrix}$$

In this expression the profit is scaled by 10^{-5}, to correspond with the scaled flows. The profit at the process optimum (in actual units) is \$397.68/hr.

SSDC Estimation

Table 3.2 summarizes the nominal setpoint values, maximum eigenvalue for the expected setpoint covariance matrix, and the sensor design costs for each sensor system. The complete $\mathbf{Q}_{\mathbf{x}^*}$ matrices can be found in Appendix B, along with the derivative matrices used in the computation of $\mathbf{Q}_{\mathbf{x}^*}$, evaluated at the optimum.

Sensor System	$\tilde{F}_B^*$ 10^5 lb/hr	$\tilde{T}_R^*$ 10^3 R	$\|\mathbf{Q}_{\mathbf{x}^*}\|_2 \times 10^4$	SSDC $\times 10^4$	Weighted SSDC \$/hr
1	0.37987	0.65314	2.9111	5.8223	10.23
2	0.37987	0.65314	2.9212	5.8424	9.61
3	0.37987	0.65314	2.9171	5.8342	10.16
4	0.37987	0.65314	7.9867	15.973	26.84
5	0.37987	0.65314	2.9075	5.8152	9.40
6	0.37987	0.65314	6.6912	13.382	20.39
7	0.37987	0.65314	6.7997	13.599	22.66
8	0.37987	0.65314	6.6107	13.221	20.52

Table 3.2: SSDC Results by Estimation

SSDC by Simulation

Table 3.3 shows the results found by simulation. Again, the complete $\mathbf{Q}_{\mathbf{x}^*}$ matrices can be found in Appendix B.

Sensor System	$\tilde{F}_B^*$ 10^5 lb/hr	$\tilde{T}_R^*$ 10^3 R	$\|\mathbf{Q}_{\mathbf{x}^*}\|_2 \times 10^4$	SSDC$\times 10^4$	Weighted SSDC \$/hr
1	0.37987	0.65314	2.8629	5.7259	9.60
2	0.37987	0.65314	2.8071	5.6142	8.98
3	0.37987	0.65314	2.8125	5.6249	9.53
4	0.37987	0.65314	7.7213	15.443	25.95
5	0.37987	0.65314	2.9158	5.8317	9.14
6	0.37987	0.65314	6.6815	13.363	20.26
7	0.37987	0.65314	6.5160	13.032	22.20
8	0.37987	0.65314	6.3911	12.782	19.86

Table 3.3: SSDC Results by Simulation

3.4.4 Discussion

The SSDC analysis shows the following:

1. The bias cost in all cases is zero.
2. Sensor system #1, which contains all possible measurements, is not the optimal choice.
3. In both the weighted and unweighted cases, sensor system #5 (which excludes both F_A and F_R) is optimal when $\mathbf{Q}_{\mathbf{x}^*}$ is linearly approximated, while sensor system #2 (which excludes F_A) is optimal when $\mathbf{Q}_{\mathbf{x}^*}$ is found by simulation.
4. The linearly approximated results capture the trends observed in the simulation results.

The first observation can be explained by the fact that the linear least squares parameter estimates are unbiased [Bates and Watts, 1988]. Therefore, with noise-free measurements each sensor system allows the true parameter values to be calculated, which

in turn allows the profit optimizer to find the true optimal setpoints. Bias cost might appear if the model was nonlinear in the parameters [Box, 1971]. For the current study, the differences in the design costs among the various sensor systems can be attributed to variance alone. The weighted SSDC can therefore be considered the *expected lost profit due to setpoint variance for each sensor system*. This variance is caused by the transmission of measurement error through the RTO loop.

The second item is quite significant, since it shows that it is not necessary to measure every variable to get similar RTO results. In all cases, the results from sensor systems #1, #2, #3 and #5 are very similar. This is contrary to the general feeling that measuring every possible variable is a good idea. In fact, the weighted results from both the linear approximation and simulation studies show that the process is more profitable when F_A and/or F_R is *not* measured.

The closeness in SSDC values for sensor systems #1, #2, #3 and #5 may not allow one to distinguish which is the most profitable. It is clear, however, that the values for these systems are much better than those values for the remaining four sensor systems. One can conclude that sensor systems #4, #6, #7 and #8 should not be implemented, given the other alternatives.

Weighting of the SSDC had an effect on the optimal sensor systems found in this study, and the difference lies in the economic considerations. The reduced Hessian shows that there is more cost associated with variance in T_R than in F_B, and this influences the weighted SSDC. Any sensor system that reduces variance in T_R will have a lower weighted SSDC. For example, the unweighted SSDC (by simulation) shows that sensor system #5 is less desirable than systems #1 and #3. But the covariance matrix associated with #5 has a smaller variance in T_R (see Appendix B), and therefore the weighted SSDC is smaller.

The results of this study have some powerful implications. Consider the following problem. The reactor currently has sensor system #7 (which excludes F_R and X_G) installed for RTO. Management has decided to invest in another sensor, and the RTO

designer must decide which additional sensor to purchase: F_R or X_G. The simulation results show that the expected lost profit with system #7 is \$22.20/hr. If the F_R sensor is then added, the expected lost profit becomes \$25.95/hr, an additional loss of \$3.75/hr. Alternately, if the X_G sensor is purchased, the expected lost profit is only \$9.53/hr, and there is a \$12.67/hr return on the investment.

3.4.5 Conclusions

The SSDC criterion is an alternative approach to sensor selection that examines the RTO loop as a whole rather than looking at the parameter updator in isolation. More importantly, as shown in this case study, the SSDC criterion can be a powerful tool for sensor selection, because it expresses the sensor selection problem in terms of the expected lost profit resulting from various sensor system alternatives. Since the goal of RTO is profitability, it seems appropriate to include a profitability measure in RTO system design.

The accuracy of the SSDC results will be a function of model accuracy and the nonlinearity present in the model. Since the entire SSDC analysis is model-based, it is crucial that an appropriate plant model be used so that actual results will be similar to the expected results. In addition, the profit function and the economic data should also be accurate. An accurate process model will rarely be linear. Therefore, when possible, an SSDC analysis by simulation will be preferable over an analysis that estimates $\mathbf{Q}_{\mathbf{x}^*}$ as a function of the first order system derivatives: $\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}}$, $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}}$ and $\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$.

Although this case study demonstrated the application of the SSDC criterion for sensor system design, it is a rather small process in terms of RTO. Further, it was assumed the reactor model was perfectly known. The next chapter will examine the use of the SSDC criterion for sensor selection in a larger, integrated plant, while incorporating both structural and parametric mismatch between the RTO model and the true plant.

Chapter 4

Case Study: Williams-Otto Plant

Chapter 3 introduced the SSDC criterion as a tool for a structured approach to sensor system selection in RTO. The Williams-Otto [1960] reactor case study demonstrated how the SSDC criterion may be used to identify good and bad sensor systems in terms of expected profit. The case study of Chapter 3 did not present a very realistic RTO situation, since there was a single unit that was perfectly modelled. The purpose of this chapter is to demonstrate the use of the SSDC criterion for RTO sensor selection in a larger, integrated plant. In this study, plant-model mismatch is incorporated by using one set of equations to represent the plant, and using a different set of equations for the model that forms the basis for the RTO system.

4.1 Williams-Otto Plant

The Williams-Otto [1960] plant was chosen for this case study. Originally posed as a test problem for comparing various control strategies, it has since been studied by many researchers (Krishnan *et al.* [1992], MacFarlane and Bacon [1989], and Roberts [1979]). Figure (4-1) shows the flow diagram of the plant. The major pieces of equipment are a reactor, a heat exchanger, a decanter, and a distillation column. The stream naming convention conforms to that used in Williams and Otto [1960].

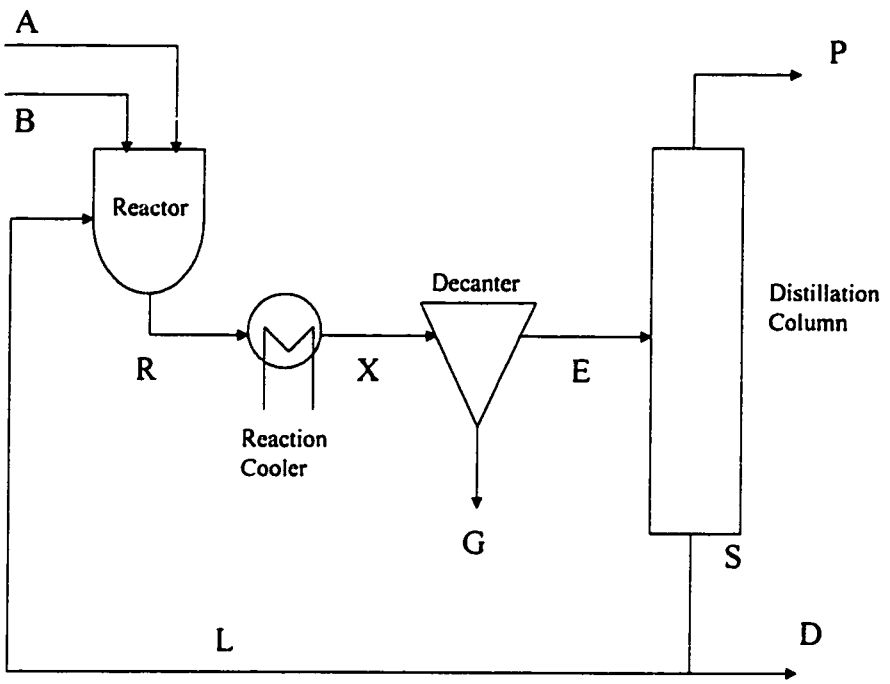


Figure 4-1: Williams-Otto Plant

Three streams enter the reactor: pure reactants A and B, plus the recycle from the distillation column bottoms. The reactor is a continuously stirred tank reactor with constant hold-up by weight. There are three exothermic reactions taking place in the reactor, and their reaction rate coefficients have an Arrhenius temperature dependence:

$$k_i = A_i \exp(-B_i/T_R) \tag{4.1}$$

where A_i and B_i are the frequency factors and activation energies respectively, $i = 1, 2, 3$. The details of the reactions are given in Table 4.1. The reactor temperature is maintained by heating and cooling tubes within it. The heating and cooling media are steam and cooling water respectively. The molecular weights of components A, B and P is 100, for components C and E it is 200, and for component G it is 300.

Reaction	Frequency Factor, hr^{-1}	Activation Energy, R	Heat of Reaction, Btu	Basis
$A + B \rightarrow C$	5.9755×10^{9}	12,000	-125	lb of A
$B + C \rightarrow P + E$	2.5962×10^{12}	15,000	-50	lb of B
$C + P \rightarrow G$	9.6283×10^{15}	20,000	-143	lb of C

Table 4.i: Reaction Data

The heat exchanger cools the reactor effluent in order to stop the reaction and to bring the effluent temperature below the point where the undesirable product G becomes insoluble (100°F). A decanter is used to remove this product G from the distillation column feed. The desired product P is concentrated by the distillation column in the overhead stream. The bottom stream is split into a recycle stream that returns to the reactor, and a purge stream that can be burned for fuel. Table 4.2 gives the physical specifications for the equipment.

Equipment	Specifications
Reactor	Capacity=4640 lb
Reactor Cooling Coils	Effective cooling area=100 ft^2 U= 0.01947 $F^{0.8}$ Btu/h/ft^2/R
Reaction Cooler	Area=569 ft^2 U=0.005856 $F^{0.8}$ Btu/h/ft^2/R
Distillation Column	Ideal stages=21 Relative Volatility=2.8
Reboiler	Area=2770 ft^2 U=0.004063 $F^{0.8}$ Btu/h/ft^2/R
Condenser	Area=4940 ft^2 (max) U=0.0054316 $F^{0.8}$ Btu/h/ft^2/R

Table 4.2: Williams-Otto plant equipment data

The Real-Time Optimizer will be used to maximize the percentage return on invest-

ment, given by Williams and Otto [1960]:

$$\% \; return = \frac{1}{0.278} \times \left\{ \begin{array}{c} -168F_A - 252F_B + 50.03712F_D - 84F_G + 2207.52F_P - 1.27173F_R - \\ 0.0336538(F_{DW} + F_{RW} + F_{XW}) - 8.4Q_{reboiler} - 2.76 \end{array} \right\} \quad (4.2)$$

where F_{DW}, F_{RW}, F_{XW} and F_{DS} represent the flows of water to the distillation column condenser, the reactor, the reaction cooler, and the flow of steam to the distillation column reboiler, respectively. The heat capacity of water is C_{Pw}, and $Q_{reboiler}$ represents the reboiler duty. Equation (4.2) includes the costs of feed streams and the prices of products, plus the charges associated with the disposal of waste product G. It also includes fixed charges (depreciation, labour, etc.), utility charges, charges associated with sales, administration, research and engineering, and the total capital investment involved.

In this study, the plant is represented with one set of equations, while the model is a *different* set of equations. The *plant* is modelled according to the original paper, but for the following modifications [Forbes, 1994]:

1. steady-state simulations were performed,
2. the heat exchangers were modelled with a log-mean temperature driving force,
3. all heat transfer coefficients were made flow dependent with flow exponents set to 0.8 [Holman, 1972],
4. the heat transfer coefficient for the reaction cooler was set at 0.42414 Btu/h/ft^2/R to match the nominal operating conditions given by Williams and Otto [1960],
5. the distillation column was modelled using tray-by-tray equilibrium relationships [Luyben, 1973] and assuming constant molal overflow,
6. the separation in the distillation column was represented as pseudo-binary with constant relative volatility,

7. the relative volatility was set at 2.8 in order to match the nominal operating conditions given by Williams and Otto [1960],

8. the distillation column has a partial condenser, with the heat transfer rate controlled by both liquid level and coolant flow,

9. the minimum acceptable concentration of product component P in the distillation column overhead stream is 95 wt%.

The plant operating constraints consist of:

1. a maximum production rate for the distillation column overhead stream of 4,763 lb/hr,

2. a minimum concentration of product component P in the distillation column overhead stream of 95 wt%,

3. a maximum available distillation column condenser heat transfer area of 4940 ft^2,

4. no by-product component G present in the distillation column feed stream and any subsequent process streams.

The feed rate of reactant A to the reactor is fixed by upstream processes. The manipulated variables available in the plant are the feed rate of reactant B to the reactor, recycle flow rate from the distillation column bottoms, reactor operating temperature, flow rate of by-product G from the decanter bottoms, flow rate of the column overhead product stream, condenser heat transfer area, and cooling water and steam flow rates in the heat exchange equipment.

Given the nominal value of feed stream F_A, there are five degrees of freedom available for plant optimization, once all of the operating constraints, thermodynamic relationships, mass and energy balances are satisfied. For this study, the five independent manipulated variables selected for optimization were the flow of reactant B to the reactor,

the recycle flow rate from the distillation column to the reactor, the reactor operating temperature, the flow of product from the top of the distillation column, and the concentration of component P in the product stream.

Through the solution of the optimization problem using the equations used to represent the plant, the optimum operation exists at the operating point shown in Table 4.3. At this operating point, the optimum flow of reactor cooling water is zero, reducing the

F_B	29461.1 lb/hr
F_L	63357.7 lb/hr
T_R	635.759 R
F_P	4763.0 lb/hr
$X_{p,SP}$	0.95
Return	56.663%

Table 4.3: Plant Optimum

degrees of freedom available for optimization. Further, the table shows that at this operating point, the overhead product constraints are active. Therefore, this study focuses on the setpoint variables F_B and T_R, leaving F_L as a manipulated variable for the control of reactor temperature.

4.2 Sensor System Design Problem

Suppose the Williams-Otto plant is in operation, and a RTO system is being designed to improve the profitability of the plant. For this purpose a model has been built, and the key parameters to update on-line have been identified. The currently available sensors are insufficient to perform the desired parameter update, and therefore an additional sensor must be installed. The sensor system design problem is to determine where the sensor should be located. In the following sections the details of the plant model, adjustable parameters, and sensor system alternatives are outlined.

4.2.1 Model

Section 4.1 outlined the set of equations used to describe the plant. From this point on, the term *plant* refers to that set of equations, which would normally be unknown. The term *model* will refer to the set of equations to be described in this section. This model will be used in the RTO system, and will incorporate mismatch with the plant.

Since component C is a reaction intermediate, it is possible to represent the three original reaction equations by a simpler two-reaction model:

$$A + 2B \rightarrow P + E$$
$$A + B + P \rightarrow G$$

These reactions have Arrhenius temperature dependencies. Since these equations do not include intermediate product C, all sensor selection work in the SSDC analysis does not consider this component.

Plant-model mismatch also exists in the heat exchanger models. All heat transfer coefficients were made flow dependent with flow exponents set to 0.7 rather than 0.8 as in the true plant. Additional mismatch is incorporated by representing the distillation column by Smoker's equation [Jafarey *et al.*, 1979] rather than a complete tray-by-tray model:

$$N = \frac{\ln\left[\frac{x_D(1-x_W)}{(1-x_D)x_W}\right]}{\ln\left[\frac{\alpha}{\left(1+\frac{1}{Rx_F}\right)^{1/2}}\right]} \tag{4.3}$$

where x_D, x_W, and x_F are the mole fractions of component P not in the azeotropic mixture for the distillate, bottoms, and feed streams respectively, R is the reflux ratio and α is the relative volatility.

4.2.2 Adjustable Parameters and Sensor System Alternatives

During RTO system design, it has been determined that the following parameters β will be updated on-line:

A_1	Arrhenius pre-exponential factor, reaction 1
A_2	Arrhenius pre-exponential factor, reaction 2
U_{rc}	Heat transfer coefficient, reaction cooler
α	Relative volatility of P to mixture of other components

The current process measurements are the flows and temperatures of all process and utility streams, and the composition of product stream P. These sensors are insufficient to perform the desired parameter update because, about the reactor in particular, there are more unknown variables than available equations. Therefore, in order to perform RTO, a gas chromatograph (GC) must be installed. There are two possible locations for the additional GC: on the recycle stream L (sensor system alternative #1), or on the reactor exit stream R (sensor system alternative #2). The SSDC criterion will be used to decide which sensor should be installed.

The problem is complicated by the fact that the RTO system does not yet exist, and therefore there is no way to validate the results of the SSDC analysis for the closed loop system. Further, because of the plant/model mismatch, there is no way to know what operating point the RTO system will converge to. For this reason, the problem is posed as a one-step-ahead problem stated as: *Given the process measurements at the current operating point, find the sensor system with the minimum SSDC for the next RTO step.*

4.3 Method

Recall the unweighted and weighted SSDC criteria from Chapter 3:

$$\min_{x} \|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\|_2^2 + s \|\mathbf{Q}_{\mathbf{x}^*}\|_2$$

and

$$\min_{x} \left|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\right| + \left|\mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right] \mathbf{w}\right|$$

In order to evaluate the SSDC criteria, the following steps were taken for each of the two sensor system alternatives:

1. Find the setpoints $\mathbf{x}_m^*$ that would be calculated if every measurable variable was perfectly measured (all flows, temperatures and compositions). This serves as the benchmark that would be achieved if one could perfectly measure all process variables. It represents the limiting best case RTO result given the plant model, to which the possible sensor systems may be compared.

2. Find the setpoints $\tilde{\mathbf{x}}^*$ that are calculated using perfect measurements from each sensor system alternative. These setpoints may have some deviation from $\mathbf{x}_m^*$, a result of not having perfect measurements of all process variables.

3. Approximate the one-step-ahead parameter and setpoint covariance matrices, based on each sensor system alternative.

4. Approximate the reduced Hessian of the profit function at $\mathbf{x}_m^*$. The Hessian describes the curvature of the profit surface as the setpoints deviate from $\mathbf{x}_m^*$.

5. Evaluate the Sensor System Design Costs.

The first two steps are straightforward, since they involve passing noise-free measurements to the parameter updator, and passing the resulting parameter values to the

optimizer to calculate setpoints. It must be emphasized that even with perfect measurements, each sensor system may calculate different parameter estimates and hence different optimal setpoints. The reasons for this are plant/model mismatch, the nonlinear nature of the process model, and the behaviour of the parameter estimation and optimization algorithms. Flowsheet (4-2) shows the sequence of calculations, where the subscripts represent the sensor system alternatives.

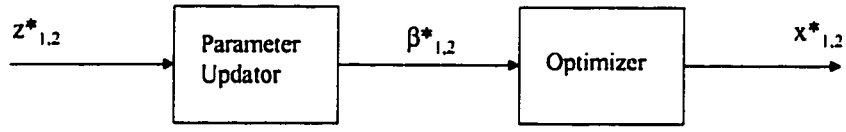


Figure 4-2: Calculation sequence for nominal setpoints.

The open-loop parameter covariance matrices are approximated as:

$$\mathbf{Q}_\beta = \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \mathbf{Q}_z \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}}^T \tag{4.4}$$

Since this is an open-loop analysis, an open-loop approximation of the setpoint covariance matrix for each sensor system is made:

$$\mathbf{Q}_{\mathbf{x}^*} = \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \right)_{\boldsymbol{\beta}^*, \mathbf{z}^*} \mathbf{Q}_z \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \right)^T_{\boldsymbol{\beta}^*, \mathbf{z}^*} \tag{4.5}$$

This is the first term of the closed-loop approximation used in Chapter 3 (see Appendix C).

In order to evaluate Equations (4.4) and (4.5), two sensitivity matrices must be approximated for each of the two sensor system alternatives. The RTO model was used to numerically estimate the sensitivity matrices $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}}$ and $\frac{\partial \mathbf{x}}{\partial \boldsymbol{\beta}}$ for each case, using central difference approximations. For each matrix $\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}}$, small perturbations were made to the

available measurements about their values $\mathbf{z}^*$ at the assumed plant optimum. For each matrix $\frac{\partial \mathbf{x}}{\partial \beta}$, small perturbations were made to the nominal parameter values $\boldsymbol{\beta}^*$ found at the assumed plant optimum. Again note that each sensor system will not necessarily produce the same nominal parameter values. The measurement covariance matrices $\mathbf{Q}_z$ for each sensor are available in Appendix E.

To approximate the reduced Hessian of the profit surface, second order difference approximations were used. The equations are:

$$\frac{\partial^2 P}{\partial \left(x^*_{m,1}\right)^2} \approx \left[\frac{P(x^*_{m,1} + \Delta x^*_{m,1}) - 2P\left(x^*_{m,1}\right) + P\left(x^*_{m,1} - \Delta x^*_{m,1}\right)}{\left(\Delta x^*_{m,1}\right)^2} \right]_{x^*_{m,2}} \tag{4.6}$$

$$\frac{\partial^2 P}{\partial \left(x^*_{m,2}\right)^2} \approx \left[\frac{P(x^*_{m,2} + \Delta x^*_{m,2}) - 2P\left(x^*_{m,2}\right) + P\left(x^*_{m,2} - \Delta x^*_{m,2}\right)}{\left(\Delta x^*_{m,2}\right)^2} \right]_{x^*_{m,1}} \tag{4.7}$$

$$\begin{aligned} \frac{\partial^2 P}{\partial x^*_{m,1} \partial x^*_{m,2}} \approx & \frac{1}{4\left(\Delta x^*_{m,1}\right)\left(\Delta x^*_{m,2}\right)} \\ & \times \left[\begin{array}{c} P(x^*_{m,1} + \Delta x^*_{m,1}, x^*_{m,2} + \Delta x^*_{m,2}) - P\left(x^*_{m,1} + \Delta x^*_{m,1}, x^*_{m,2} - \Delta x^*_{m,2}\right) - \\ P\left(x^*_{m,1} - \Delta x^*_{m,1}, x^*_{m,2} + \Delta x^*_{m,2}\right) + P\left(x^*_{m,1} - \Delta x^*_{m,1}, x^*_{m,2} - \Delta x^*_{m,2}\right) \end{array} \right] \end{aligned} \tag{4.8}$$

where $\Delta x^*_{m,1}$ and $\Delta x^*_{m,2}$ are small perturbations about $x^*_{m,1}$ and $x^*_{m,2}$ respectively.

The SSDC analysis is based on linear approximations of the setpoint covariance matrix. In order to test the validity of the results, an open-loop simulation is performed using each of the two sensor system alternatives. The method is:

1. Generate measurements that have random, normally distributed noise associated with them, using the measurement covariance matrix.

2. Use these measurements in the parameter updator to produce a distribution of parameter estimates, $\mathbf{Q}_\beta$.

3. Use the distribution of parameter estimates in the optimizer to produce a distribution of setpoints, $\mathbf{Q}_{\mathbf{x}^*}$. This setpoint covariance matrix will be based on the nonlinear nature of the model rather than a linear approximation.

4. Evaluate the SSDC based on this $\mathbf{Q}_{\mathbf{x}^*}$.

Figure (4-3) shows a schematic for the open-loop simulation.

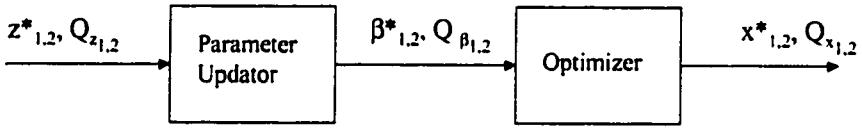


Figure 4-3: Covariance calculations.

4.4 Results

Table 4.4 shows the true optimal plant setpoints, the setpoints that would be found if all process variables were perfectly measured, and the setpoints found using perfect measurements from the two sensor system alternatives.

Variable	$\mathbf{x}^*$	$\mathbf{x}^*_m$	$\bar{\mathbf{x}}^*_1$	$\bar{\mathbf{x}}^*_2$
F_B, lb/hr	29461	27871	27782	27576
T_R, R	635.76	628.709	629.93	628.94

Table 4.4: Optimal Setpoints.

The sensitivity data can be found in Appendix E. Using these sensitivities and the measurement covariance matrices, the setpoint covariance matrices were approximated

as:

$$\mathbf{Q}_{\mathbf{x}^*,1} = \begin{bmatrix} 2.5002 \times 10^6 & 1.7615 \times 10^3 \\ 1.7615 \times 10^3 & 4.4527 \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x}^*,2} = \begin{bmatrix} 1.9215 \times 10^6 & 1.3674 \times 10^3 \\ 1.3674 \times 10^3 & 1.8657 \end{bmatrix}$$

where the rows and columns correspond to F_B and T_R. The reduced Hessian at $\mathbf{x}_m^*$ was approximated as:

$$\nabla_r^2 P = \begin{bmatrix} -5.8560 \times 10^{-5} & -0.0037680 \\ -0.0037680 & -0.36080 \end{bmatrix}$$

Table 4.5 shows the calculated SSDC criteria for each sensor system alternative.

Sensor System	SSDC	Weighted SSDC, $/hr
1	5.0082×10^6	$161.48
2	3.9301×10^6	$128.08

Table 4.5: SSDC results by linear approximation.

By simulation, the open-loop setpoint covariance matrices are:

$$\mathbf{Q}_{\mathbf{x}^*,1} = \begin{bmatrix} 3.3404 \times 10^6 & 2.8510 \times 10^3 \\ 2.8510 \times 10^3 & 6.0068 \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x}^*,2} = \begin{bmatrix} 2.9707 \times 10^6 & 2.6970 \times 10^3 \\ 2.6970 \times 10^3 & 3.6059 \end{bmatrix}$$

Using these matrices, the data in Table 4.4, and the reduced Hessian, the simulated SSDC results are:

Sensor System	SSDC, unscaled	Weighted SSDC, $/hr
1	6.6886×10^6	$219.45
2	5.9492×10^6	$195.77

Table 4.6: SSDC results by simulation.

4.5 Discussion

In this section the following issues were addressed: plant/model setpoint mismatch, setpoint bias cost, variance cost, SSDC results, and sensor cost. Table 4.4 shows there is some deviation between the plant optimum, and the optimum found by the optimizer when all process variables are perfectly known. This may suggest that the model is inadequate for RTO [Forbes, 1994]. In fact, it is possible for the model to produce the true plant optimal setpoints by adjusting the parameter values. The problem is that at the plant optimum, these parameter values are not calculated by the parameter updator given the model and the measured variables. Therefore, setpoint mismatch results in the optimizer. The mismatch is 5.4% for F_B, and 1.1% for T_R.

The best possible situation is to have perfect knowledge of all the process variables. The RTO results from the sensor system alternatives are therefore compared to this benchmark. Table 4.4 shows that there is some bias between the nominal setpoint values found when all the measurable variables are perfectly known, $\mathbf{x}_m^*$, and those found using each of the two sensor system alternatives, $\bar{\mathbf{x}}_{1,2}^*$. These deviations are small, and therefore the bias costs will be relatively small when compared to variance cost.

The variance cost is a result of measurement variance, which is translated to parameter variance and setpoint variance through the parameter updator and optimizer. The setpoint covariance matrices show that the second sensor system alternative (with the GC on stream R) gives a smaller setpoint confidence region than the first alternative. This may be explained through the parameter covariance matrices (refer to Appendix E for the matrices). The approximated parameter covariance matrices show that the

second alternative gives a tighter confidence on parameters A_1 and A_2. Since the set-point variables exist about the reactor, one would imagine that higher confidence in these two parameters would translate to a smaller confidence region in the setpoint variables. An additional point of interest is that some of the signs associated with the parameter covariances are different depending on the sensor system. For example, $\mathbf{Q}_{\beta,1}$ shows a positive covariance between A_1 and A_2 while $\mathbf{Q}_{\beta,2}$ shows this relationship to be negative.

The reduced Hessian shows that the profit is most sensitive to T_R. This makes sense since the temperature of the reactor has the largest effect on the product distribution. Therefore, one would expect that the sensor system that has the smallest variance in T_R would be the better choice in terms of SSDC: sensor system #2. Table 4.5 indicates that the weighted SSDC is \$161.48/hr for sensor system 1, and \$128.08/hr for sensor system 2. That is, compared to the best possible case in which all process variables are perfectly known, sensor system #1 gives an expected loss of \$161.48/hr, while sensor system #2 gives an expected loss of \$128.08/hr. Therefore the analysis indicates that the GC on stream R should be chosen over the GC on stream L. Even when the reduced Hessian is not used, the unweighted SSDC gives the same conclusion. The open-loop simulation results also indicate that sensor system #2 should be chosen over sensor system #1, although the magnitudes of the SSDC criteria are larger than in the linearly approximated analysis. Table 4.7 shows the ratios between the predicted and simulated SSDC results, both weighted and unweighted. Table 4.8 shows the ratios of the system #1 results to the system #2 results for each test.

	Unweighted SSDC	Weighted SSDC
Pred./Sim., System 1	0.749	0.736
Pred./Sim., System 2	0.661	0.654

Table 4.7: Ratio of predicted SSDC to simulated SSDC.

The cost of the sensors has not been addressed in this case study. Sensor cost is a very important issue, and may alter the conclusions from the SSDC analysis. This case study has assumed equal sensor costs. Suppose that the cost of the GC on stream L

	Unweighted SSDC	Weighted SSDC
System #1/System #2, Predictions	1.2743	1.2608
System #1/System #2, Simulation	1.1243	1.1210

Table 4.8: Ratio of system 1 results to system 2 results.

was much cheaper to purchase, install and maintain. In this case there would be some trade-off between the SSDC and sensor costs, which may be addressed separately, or can in fact be included in the return on investment function.

For this case study, the SSDC analysis shows that the GC should be placed on stream R rather than stream L. This placement allows for more accurate estimation of two key reaction rate parameters and hence more accurate setpoint estimation when compared to the best case benchmark. The simulation results lead to the same conclusion.

Chapter 5

Summary and Conclusions

In the competitive marketplace, many industries are striving to be the low-cost producer by operating their processes in an economically optimal fashion. Among the on-line optimization alternatives, model-based optimization has enjoyed its popularity because of the reduction in plant experimentation and the increased speed of convergence. This thesis is concerned with such on-line model-based optimizers, commonly called Real-Time Optimizers or On-line Optimizers. Using a statistical approach, this thesis evaluated current parameter estimation techniques in terms of parameter quality, and developed a new tool for making sensor placement decisions by considering the RTO loop as a whole.

It is generally understood that the choice of numerical technique used in the parameter estimator will affect the quality of the parameter estimates, but industry continues to back-substitute rather than take a least squares approach. Although back-substitution is a simple method to implement, the trade-off in terms of RTO performance can be substantial. Chapter 2 discussed various estimation techniques, and showed through a case study that adopting a simple least squares estimation scheme over a back-substitution scheme can dramatically reduce the size of the parameter confidence region. Further, in the study, the least squares method was shown to provide increased robustness to measurement error and sensor failure. The error-in-variables method would further improve the quality of the parameter estimates, but current computing technology makes this

method impractical for RTO.

Chapter 3 investigated the issue of sensor selection. The available literature provides some design methods for choosing sensors [Krishnan *et al.*, 1992], but that research considers the problem in terms of parameter confidence region. That is, it treats the updator as an isolated system rather than examining the effects of sensor selection on the RTO system as a whole. Since the goal of RTO is increased profitability, this thesis analyzed sensor system design in terms of profit by examining variance and bias in the calculated setpoints, and the plant profit surface. Chapter 3 began with a discussion of optimal design of experiments criteria, and developed two independent criteria that can be used for sensor selection in terms of variance cost and bias cost. These criteria were the weighted D-optimal and T-optimal design criteria, respectively. Recognizing the need to combine the two criteria in a single design tool, a measure called the Sensor System Design Cost (SSDC) was developed. The SSDC criterion effectively balances the D- and T-optimality criteria and allows the RTO designer to compare various sensor systems with respect to uncaptured profit due to setpoint bias and variance. The most practical benefit to analyzing sensor system alternatives, in terms of profit, is that it provides a tangible measure of expected RTO performance. It would be much easier to justify the purchase of an additional sensor to management when additional profit can be associated with it, rather than merely associating some improved confidence in a particular model parameter.

The case study of Chapter 4 demonstrated how the SSDC criterion may be used to choose among sensor system alternatives for RTO. It should be noted that the analysis need not be restricted to choosing among sensors to be installed. In the case where there are a multitude of measurements available in the plant, the SSDC may instead be used to decide which *subset* of the available measurements to use for RTO calculations. For example, if certain sensors are inherently noisy it may be beneficial to exclude those measurements for improved parameter and setpoint accuracy.

Although this alternative approach to sensor selection has its merits, it also has its

shortcomings. The primary issue is that the analysis is based on linear approximations of the plant model at a particular operating point. There will always be error caused by plant/model mismatch. Further error is introduced because the nonlinearities in the model will limit the region in which linear approximations are valid. There are no guarantees that any conclusions drawn at one operating point will hold at some nearby operating point. However, using the SSDC may be justified since it is a knowledge-based approach to choosing sensors. Without the use of a design tool based on some process knowledge and expected RTO results, the RTO designer has no information for choosing among sensor system alternatives.

An additional point that was discussed in Chapter 3 but should be restated, is that the SSDC criterion aims to minimize the largest eigenvalue of the setpoint covariance matrix. Effectively, this minimizes the longest axis of the setpoint confidence region. This does not guarantee that the total volume of the confidence region is minimized, because as the longest axis becomes smaller, other axes may become larger. It would be beneficial to develop an alternative SSDC measure that explicitly involves the minimization of the determinant of the covariance matrix.

Bibliography

[1] J. S. Albuquerque and L. T. Biegler. Data reconciliation and gross error detection for dynamic systems. *AIChE J.*, 42(10):2841–2856, 1996.

[2] A. C. Atkinson and V. V. Federov. The design of experiments for discriminating between two rival models. *Biometrika*, 62(1):57–70, 1975.

[3] W. Bamberger and R. Iserman. Adaptive on-line steady-state optimization of slow dynamic processes. *Automatica*, 14:223–230, 1978.

[4] D. M. Bates and D. G. Watts. *Nonlinear Regression Analysis and its Applications.* John Wiley Sons, New York, 1988.

[5] L. T. Biegler, I. E. Grossman, and A. W. Westerberg. A note on approximation techniques used for process optimization. *Comp. Chem. Eng.*, 9(2):201–206, 1985.

[6] G. E. P. Box and N. R. Draper. *Evolutionary Operations.* John Wiley Sons, New York, 1969.

[7] G. E. P. Box and W. G. Hunter. Sequential design of experiments for nonlinear models. *IBM Symposium in Statistics*, pages 113–137, 1965.

[8] G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters.* John Wiley Sons, New York, 1978.

[9] G. E. P. Box and H. L. Lucas. Design of experiments in nonlinear situations. *Biometrika*, 46:77–90, 1959.

[10] M. J. Box. Improved parameter estimation. *Technometrics*, 12(2):219–229, 1970.

[11] M. J. Box. Bias in nonlinear estimation. *J. Royal Stat. Soc. Ser. B*, 33:171–190, 1971.

[12] H. I. Britt and R. H. Luecke. The estimation of parameters in nonlinear implicit models. *Technometrics*, 15(2):233–247, 1973.

[13] C. Chatfield and A. J. Collins. *Introduction to Multivariate Analysis*. Chapman and Hall, New York, 1980.

[14] C. M. Crowe. Reconciliation of process flow rates by matrix projection. II: The nonlinear case. *AIChE J.*, 32(4):616–623, 1986.

[15] C. M. Crowe. Recursive identification of gross errors in linear data reconciliation. *AIChE J.*, 34(4):541–550, 1988.

[16] C. M. Crowe. Formulation of linear data reconciliation using information theory. *Chem. Eng. Science*, 51(12):3359–3366, 1996.

[17] C. M. Crowe, Y. A. Garcia Campos, and A. N. Hrymak. Reconciliation of process flow rates by matrix projection. I: The linear case. *AIChE J.*, 29(6):881–888, 1983.

[18] C. R. Cutler and R. T. Perry. Real time optimization with multivariable control is required to maximize profits. *Comp. Chem. Eng.*, 7(5):663–667, 1983.

[19] G. Dahlquist and A. Bjorck. *Numerical Methods*. Prentice-Hall, N.J., 1974.

[20] S. de Hennin, J. D. Perkins, and G. W. Barton. Structural decisions in on-line optimization. In *Proceedings of the International Conference on Process Systems Engineering PSE '94*, pages 297–302, 1994.

[21] W. E. Deming. *Statistical Adjustment of Data*. Wiley, New York, 1943.

[22] R. C. Durbeck. *Principles for Simplification of Optimizing Control Models.* PhD thesis, Case Institute of Technology, 1965.

[23] T. F. Edgar and D. M. Himmelblau. *Optimization of Chemical Processes.* McGraw-Hill, New York, 1988.

[24] J. F. Forbes. *Model Structure and Adjustable Parameter Selection for Operations Optimization.* PhD thesis, McMaster University, Hamilton, 1994.

[25] J. F. Forbes and T. E. Marlin. Design cost: a systematic approach to technology selection for model-based real-time optimization. *Comp. Chem. Eng.*, 20(6/7):717-734, 1996.

[26] N. Ganesh and L. T. Biegler. A reduced hessian strategy for sensitivity analysis of optimal flowsheets. *AIChE J.*, 33(2):282–296, 1987.

[27] C. E. Garcia and M. Morari. Optimal operation of integrated processing systems. *AIChE J.*, 27(6):960–968, 1981.

[28] P. Gillett. *Calculus and Analytical Geometry.* D.C. Heath and Company, Toronto, 1984.

[29] J. P. Holman. *Heat Transfer.* McGraw-Hill, New York, 1972.

[30] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, 1991.

[31] A. S. Householder. *The Theory of Matrices in Numerical Analysis.* Blaisdell Pub. Co., New York, 1964.

[32] A. Jafarey, J. M. Douglas, and T. J. McAvoy. Short-cut techniques for distillation column design and control: 1. column design. *Ind. Eng. Chem. Proc. Des. Dev.*, 18(2):197–202, 1979.

[33] R. E. Kalman. Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5:102–119, 1960.

[34] J. Kiefer and J. Wolfowitz. Optimum designs in regression problems. *Ann. Math. Stat.*, 30, 1959.

[35] I. W. Kim, M. J. Liebman, and T. F. Edgar. Robust error-in-variables estimation using nonlinear programming techniques. *AIChE J.*, 36(7):985–993, 1990.

[36] I. W. Kim, M. J. Liebman, and T. F. Edgar. A sequential error-in-variables method for nonlinear dynamic systems. *Comp. Chem. Eng.*, 15(9):663–670, 1991.

[37] E. V. Krishnamurthy and S. K. Ken. *Numerical Algorithms: Computations in Science and Engineering.* Affiliated East-West Press Private Limited, New Delhi, 1986.

[38] S. Krishnan. PhD thesis, University of Sydney, Australia, 1990.

[39] S. Krishnan, G. W. Barton, and J. D. Perkins. Robust parameter estimation in on-line optimization - part i: methodology and simulated case study. *Comp. Chem. Eng.*, 16(6):545–562, 1992.

[40] M. J. Liebman, T. F. Edgar, and L. S. Lasdon. Efficient data reconciliation and estimation using nonlinear programming techniques. *Comp. Chem. Eng.*, 16(10/11):963–986, 1992.

[41] W. L. Luyben. *Process Modelling, Simulation and Control for Chemical Engineers.* McGraw-Hill, New York, 1973.

[42] R. C. MacFarlane and D. W. Bacon. Empirical strategies for open-loop on-line optimization. *Can. J. Chem. Eng.*, 67:665–677, August 1989.

[43] R. S. H. Mah and A. C. Tamhane. Detection of gross errors in process data. *AIChE J.*, 28(5):828–830, 1982.

[44] M. O'Neil, I. G. Sinclair, and F. J. Smith. Polynomial curve fitting when abscissas and ordinates are both subject to error. *Computer J.*, 12:52, 1969.

[45] J. M. Ortega. *Matrix Theory: A Second Course.* Plenum Press, New York, 1987.

[46] A. Pages, H. Pingaud, M. Meyer, and X. Joulia. A strategy for simultaneous data reconciliation and parameter estimation on process flowsheets. *Comp. Chem. Eng.*, 18:S223–S227, 1994.

[47] P. M. Reilly and H. Patino-Leal. A Bayesian study of the error-in-variables model. *Technometrics*, 23(3):221–231, 1981.

[48] P. D. Roberts. An algorithm for steady-state optimization and parameter estimation. *Int. J. Sys. Sci.*, 10(7):719–734, 1979.

[49] J. Rosenberg, R. S. H. Mah, and C. Iordache. Evaluation of schemes for detecting and identifying gross errors in process data. *Ind. Eng. Chem. Res.*, 26:555, 1987.

[50] H. Schwetlick and V. Tiller. Numerical methods for estimating parameters in non-linear models with error in the variables. *Technometrics*, 27(1):17, 1985.

[51] D. Scott. Achieving profit and efficiency with on-line optimization. *Measurement and Control*, 29(5):139–142, 1996.

[52] A. Singh. Modeling and model updating in the real-time optimization of gasoline blending. Master's thesis, University of Toronto, Canada, 1997.

[53] H. W. Sorenson. *Parameter Estimation.* Marcel Dekker Inc., New York, 1980.

[54] W. H. Southwell. Fitting experimental data. *J. Comput. Phys.*, 4:465, 1969.

[55] G. M. Stanley and R. S. H. Mah. Observability and redundancy in process data estimation. *Chem. Eng. Sci.*, 36:259–272, 1981.

[56] I. B. Tjoa and L. T. Biegler. Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Comp. Chem. Eng.*, 15(10):679–690, 1991.

[57] I. B. Tjoa and L. T. Biegler. Reduced successive quadratic programming strategy for error-in-variables estimation. *Comp. Chem. Eng.*, 16(6):523–533, 1992.

[58] H. Tong and C. M. Crowe. Detection of gross errors in data reconciliation by principal component analysis. *AIChE J.*, 41(7):1712–1722, 1995.

[59] E. Tsang, 1998. Personal conversation.

[60] P. Valko and S. Vajda. An extended Marquardt-type procedure for fitting error-in-variables models. *Comp. Chem. Eng.*, 11(1):37, 1987.

[61] A. Wald. On the efficient design of statistical investigations. *Ann. Math. Statist.*, 14:134–140, 1943.

[62] D. C. White. Online optimization: what, where and estimation roi. *Hydrocarbon Processing*, pages 43–51, June 1997.

[63] J. H. Williams. Least squares fitting of a straight line. *Can. J. Physics*, 46:1845, 1968.

[64] T. J. Williams and R. E. Otto. A generalized chemical processing model for the investigation of computer control. *AIEE Transactions*, pages 458–473, November 1960.

[65] P. York. Least squares fitting of a straight line. *Can. J. Physics*, 44:1709, 1966.

Appendix A

Error-In-Variables Method

Chapter 2 briefly outlined a number of parameter estimation techniques, including the error-in-variables method. This appendix provides further details about this method.

In RTO, the model parameters are estimated while all variables are subject to error. Therefore, the best treatment of the parameter estimation problem is a method that formally recognizes no distinction between dependent and independent variables, but treats all variables equally. This method is known as the error-in-variables method (EVM) [Britt and Luecke, 1973]. EVM provides both parameter estimates and reconciled data estimates that are consistent with the model [Kim *et al.*, 1990]. Britt and Luecke [1973] formulated the estimation problem by considering the n-equation model:

$$\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta}) = \mathbf{0} \tag{A.1}$$

where $\mathbf{z}^*$ and $\boldsymbol{\beta}$ are the true values of the m measured variables and p parameters respectively. The measurements are corrupted by noise:

$$\mathbf{z} = \mathbf{z}^* + \boldsymbol{\varepsilon} \tag{A.2}$$

where $\boldsymbol{\varepsilon}$ represents the measurement error. In this formulation, both $\boldsymbol{\beta}$ and $\mathbf{z}^*$ are unknown quantities since $\mathbf{z}^*$ represents the true values of the measured variables. Therefore,

in order to find a maximum likelihood estimate of $\boldsymbol{\beta}$, it is also necessary to simultaneously estimate $\mathbf{z}^*$. The following assumptions are made [Britt and Luecke, 1973]:

(i) $\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta})$ is twice continuously differentiable with respect to each pair of arguments.

(ii) The Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}$ has rank p, and the Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \mathbf{z}}$ has rank n.

(iii) The error $\boldsymbol{\varepsilon}$ is normally distributed having zero mean and a known positive definite covariance matrix $\mathbf{Q}_\varepsilon$.

The joint probability density function for the measurements is:

$$g(\mathbf{z}_m) = (2\pi)^{-m/2} |\mathbf{Q}_\varepsilon|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{z}^*)^T \mathbf{Q}_\varepsilon^{-1} (\mathbf{z} - \mathbf{z}^*)\right] \tag{A.3}$$

The likelihood function is found using the joint probability density function, by considering the measurements $\mathbf{z}$ as known, and the unknowns $(\mathbf{z}^*, \boldsymbol{\beta})$ as the variables:

$$l(\mathbf{z}^*, \boldsymbol{\beta}) = (2\pi)^{-m/2} |\mathbf{Q}_\varepsilon|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{z} - \mathbf{z}^*)^T \mathbf{Q}_\varepsilon^{-1} (\mathbf{z} - \mathbf{z}^*)\right] \tag{A.4}$$

The unknowns $(\mathbf{z}^*, \boldsymbol{\beta})$ must obey the following constraint:

$$\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta}) = \mathbf{0} \tag{A.5}$$

Any solution of Equation (A.5) that maximizes the likelihood function with respect to all other solutions, is a maximum likelihood estimate of $(\mathbf{z}^*, \boldsymbol{\beta})$. Therefore the EVM problem can be expressed as:

$$\min_{\mathbf{z}^*, \boldsymbol{\beta}} \frac{1}{2}(\mathbf{z} - \mathbf{z}^*)^T \mathbf{Q}_\varepsilon^{-1} (\mathbf{z} - \mathbf{z}^*) \tag{A.6}$$

subject to: (A.7)

$$\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta}) = \mathbf{0}$$

To solve the problem, Lagrange multipliers are used to establish the objective function:

$$\varphi(\mathbf{z}^*, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{2}(\mathbf{z} - \mathbf{z}^*)^T \mathbf{Q}_\epsilon^{-1}(\mathbf{z} - \mathbf{z}^*) + \boldsymbol{\lambda}^T \mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta}) \tag{A.8}$$

The necessary conditions for a stationary point of Equation (A.8) are that the first derivatives of φ with respect to each of $(\mathbf{z}^*, \boldsymbol{\beta}, \boldsymbol{\lambda})$ are all equal to zero, which leads to the following system of equations:

$$\begin{aligned} -\mathbf{Q}_\epsilon^{-1}(\mathbf{z} - \mathbf{z}^*) + \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T \boldsymbol{\lambda} &= \mathbf{0} \\ \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T \boldsymbol{\lambda} &= \mathbf{0} \\ \mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta}) &= \mathbf{0} \end{aligned} \tag{A.9}$$

Unfortunately, this system of $m + p + n$ nonlinear equations is quite large for solution by usual methods. Generally the number of measurements alone make this problem impractical, if not impossible, to solve. Britt and Luecke [1973] therefore used a first order Taylor approximation of the constraint equation about the most recent estimate of $(\mathbf{z}^*, \boldsymbol{\beta})$, in an iterative scheme. The objective function becomes:

$$\varphi(\mathbf{z}^*, \boldsymbol{\beta}, \boldsymbol{\lambda}) = \frac{1}{2}(\mathbf{z} - \mathbf{z}^*)^T \mathbf{Q}_\epsilon^{-1}(\mathbf{z} - \mathbf{z}^*) + \boldsymbol{\lambda}^T \left[\mathbf{f}(\mathbf{z}_k^*, \boldsymbol{\beta}_k) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta} - \boldsymbol{\beta}_k) + \frac{\partial \mathbf{f}}{\partial \mathbf{z}}(\mathbf{z}^* - \mathbf{z}_k^*) \right] \tag{A.10}$$

where k is the iteration number, and the Jacobian matrices are evaluated at $(\mathbf{z}_k^*, \boldsymbol{\beta}_k)$.

The necessary conditions for a stationary point are:

$$\begin{aligned} -\mathbf{Q}_\varepsilon^{-1}(\mathbf{z}-\mathbf{z}^*) + \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T \boldsymbol{\lambda} &= \mathbf{0} \\ \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T \boldsymbol{\lambda} &= \mathbf{0} \\ \mathbf{f}(\mathbf{z}_k^*, \boldsymbol{\beta}_k) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}-\boldsymbol{\beta}_k) + \frac{\partial \mathbf{f}}{\partial \mathbf{z}}(\mathbf{z}^*-\mathbf{z}_k^*) &= \mathbf{0} \end{aligned} \tag{A.11}$$

from which Britt and Luecke [1973] derived the following equations:

$$\begin{aligned} \boldsymbol{\beta}-\boldsymbol{\beta}^{(k)} = -\left[\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T \left(\frac{\partial \mathbf{f}}{\partial \mathbf{z}}\mathbf{Q}_\varepsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T\right)^{-1} \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\right]^{-1} \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T \left(\frac{\partial \mathbf{f}}{\partial \mathbf{z}}\mathbf{Q}_\varepsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T\right)^{-1} \\ \bullet \left[\mathbf{f}\left(\mathbf{z}^{*(k)}, \boldsymbol{\beta}^{(k)}\right) + \frac{\partial \mathbf{f}}{\partial \mathbf{z}}\left(\mathbf{z}-\mathbf{z}^{*(k)}\right)\right] \end{aligned} \tag{A.12}$$

$$\begin{aligned} \mathbf{z}^*-\mathbf{z}^{*(k)} = \mathbf{z}-\mathbf{z}^{*(k)} - \mathbf{Q}_\varepsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T \left(\frac{\partial \mathbf{f}}{\partial \mathbf{z}}\mathbf{Q}_\varepsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T\right)^{-1} \\ \bullet \left[\mathbf{f}\left(\mathbf{z}^{*(k)}, \boldsymbol{\beta}^{(k)}\right) + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^{(k)}\right) + \frac{\partial \mathbf{f}}{\partial \mathbf{z}}\left(\mathbf{z}-\mathbf{z}^{*(k)}\right)\right] \end{aligned} \tag{A.13}$$

The method is as follows:

1. Compute $(\mathbf{z}^*, \boldsymbol{\beta})$ using Equations (A.12) and (A.13). This will be a stationary point of Equation (A.10), which is just an approximation of the true objective function.
2. Set $\left(\mathbf{z}^{*(k+1)}, \boldsymbol{\beta}^{(k+1)}\right) = (\mathbf{z}^*, \boldsymbol{\beta})$, relinearize the objective function about this new point, and find a new estimate $\left(\mathbf{z}^{*(k+2)}, \boldsymbol{\beta}^{(k+2)}\right)$.
3. Continue iterations until tolerances on $\left\|\boldsymbol{\beta}^{(k+1)} - \boldsymbol{\beta}^{(k)}\right\|$ and $\left\|\mathbf{z}^{*(k+1)} - \mathbf{z}^{*(k)}\right\|$ are met.

This algorithm is a simultaneous $m \times p$ search, and involves the inversion of an $n \times n$ matrix as well as a $p \times p$ matrix. Britt and Luecke [1973] point out that the

matrix $\frac{\partial \mathbf{f}}{\partial \mathbf{z}} \mathbf{Q}_\epsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T$ will usually be sparse, and therefore easy to invert. Note that assumption *(iii)* on $\mathbf{f}(\mathbf{z}^*, \boldsymbol{\beta})$ and the positive definiteness of $\mathbf{Q}_\epsilon$ ensures that $\frac{\partial \mathbf{f}}{\partial \mathbf{z}} \mathbf{Q}_\epsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T$ and $\frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}^T \left(\frac{\partial \mathbf{f}}{\partial \mathbf{z}} \mathbf{Q}_\epsilon \frac{\partial \mathbf{f}}{\partial \mathbf{z}}^T \right)^{-1} \frac{\partial \mathbf{f}}{\partial \boldsymbol{\beta}}$ are nonsingular. Also, if the procedure converges, condition *(i)* ensures that the final result is a stationary point of Equation (A.8). However, as in any nonlinear problem, it is possible that the result is a stationary point other than the absolute minimum. It is also possible that the minimum occurs at more than one point.

In EVM, the simultaneous estimation of parameters and true values for measured variables can yield very large problems that cannot be efficiently solved using current computing technology. A number of modifications to EVM have been proposed. One approach is to nest the measurement correction step within the parameter estimation step [Reilly and Patino-Leal, 1981], which reduces the size of the estimation problem. Another approach is to separate the two steps into a two-stage calculation [Schwetlick and Tiller, 1985; Kim *et al.*, 1990], again reducing computational requirements but there is no convergence guarantee.

Appendix B

Case Study Data

B.1 Williams-Otto Reactor Study, Section 2.4.

The nominal operating point used in the case study of Section 2.4 is given in Table B.1. Random, normally distributed measurement noise was added to component measure-

F_A	14,500 lb/hr
F_B	38,000 lb/hr
F_R	52,500 lb/hr
T_R	653.2 R
X_G	0.1075
X_A	0.0874
X_B	0.3896
X_C	0.0153
X_E	0.2907
X_P	0.1095

Table B.1: Nominal operating point, Section 2.4 study.

ments according to the chosen covariance matrix:

$$\mathbf{Q} = \begin{bmatrix} 0.0001 & -0.00005 & 0 & 0 & 0 & 0 \\ -0.00005 & 0.0001 & 0 & 0 & -0.00003 & 0 \\ 0 & 0 & 0.0001 & -0.00005 & -0.00001 & 0 \\ 0 & 0 & -0.00005 & 0.0001 & -5.48 \times 10^{-9} & 0 \\ 0 & -0.00003 & -0.00001 & -5.48 \times 10^{-9} & 0.0001 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0001 \end{bmatrix}$$

where the columns and rows correspond to $\begin{bmatrix} X_A & X_B & X_C & X_E & X_P & X_G \end{bmatrix}$. The variances on flows and temperature were chosen so that $3\sigma = 5\%$ error on flows, and $3\sigma = 0.5R$ for temperature.

B.2 Williams-Otto Reactor Study, Section 3.4

The variance and covariance information is the same as given above.

B.2.1 Approximation of $\mathbf{Q_x}$

Sensitivity Matrices

The sensitivity of the setpoints to the parameters is given by:

$$\frac{d\mathbf{x}}{d\boldsymbol{\beta}} = \begin{bmatrix} 0.0050246 & 0.0057398 & -0.0047359 \\ -0.0014235 & 0.0018218 & -0.0016640 \end{bmatrix}$$

The sensitivity matrices $\frac{d\beta}{dz}$ were dependent on the measurement set. In the following set of matrices, the subscript refers to the associated sensor system. For sensor systems #1

through #8, the following matrices are obtained:

$$\frac{d\boldsymbol{\beta}^T}{d\mathbf{z}_1} = \begin{bmatrix} 23.123 & 10.334 & 57.773 \\ 28.418 & 33.757 & 46.591 \\ -7.8923 & 1.3418 & -11.445 \\ -281.06 & -526.67 & -934.12 \\ 8.6149 & -3.2539 & 149.64 \\ -126.51 & -5.4652 & -30.554 \\ -40.585 & -56.151 & -24.540 \\ 13.782 & -990.19 & -1328.9 \\ 5.5580 & 24.585 & -11.737 \\ -0.039561 & 13.357 & -237.64 \end{bmatrix}$$

$$\frac{d\boldsymbol{\beta}^T}{d\mathbf{z}_2} = \begin{bmatrix} 19.025 & 29.509 & 23.086 \\ 5.2819 & 7.2197 & 21.397 \\ -281.32 & -527.47 & -937.62 \\ 9.7242 & -2.7512 & 152.02 \\ -121.21 & -3.0605 & -17.095 \\ -42.502 & -57.046 & -29.195 \\ 14.633 & -990.86 & -1329.7 \\ 6.2737 & 24.863 & -9.9141 \\ -0.10357 & 13.348 & -238.33 \end{bmatrix}$$

$$\frac{d\boldsymbol{\beta}^T}{d\mathbf{z}_3} = \begin{bmatrix} 16.652 & 11.413 & 48.250 \\ 19.967 & 35.128 & 34.233 \\ -281.32 & -527.47 & -937.62 \\ 9.0802 & -3.3281 & 149.92 \\ -126.46 & -5.4222 & -30.211 \\ -39.505 & -56.367 & -22.839 \\ 13.508 & -991.70 & -1333.1 \\ 6.4853 & 24.384 & -10.357 \\ 0.21679 & 13.301 & -237.85 \end{bmatrix}$$

$$\frac{d\boldsymbol{\beta}^T}{d\mathbf{z}_4} = \begin{bmatrix} 36.610 & 5.2211 & 291.62 \\ 39.289 & 29.586 & 235.18 \\ -19.498 & 5.7214 & -212.67 \\ -281.31 & -527.48 & -937.43 \\ -133.58 & -2.7420 & -153.16 \\ -46.290 & -54.029 & -123.45 \\ 6.5878 & -989.20 & -1444.3 \\ 2.8121 & 25.577 & -59.251 \\ -13.118 & 18.268 & -464.08 \end{bmatrix}$$

$$\frac{d\boldsymbol{\beta}^T}{d\mathbf{z}_5} = \begin{bmatrix} 26.323 & 39.484 & 52.650 \\ -281.32 & -527.47 & -937.63 \\ 9.5254 & -3.0229 & 151.21 \\ -119.40 & -0.58283 & -9.7514 \\ -44.520 & -59.804 & -37.370 \\ 14.531 & -991.00 & -1330.1 \\ 5.3583 & 23.612 & -13.622 \\ -0.49496 & 12.813 & -239.91 \end{bmatrix}$$

$$\frac{d\boldsymbol{\beta}^T}{d\mathbf{z}_6} = \begin{bmatrix} 25.635 & 27.639 & 126.41 \\ 0.49793 & 8.5732 & -53.389 \\ -281.31 & -527.48 & -937.44 \\ -126.10 & -1.6758 & -93.604 \\ -50.861 & -54.681 & -159.86 \\ 8.0478 & -989.00 & -1432.6 \\ 3.4353 & 25.666 & -54.287 \\ -16.026 & 17.853 & -487.24 \end{bmatrix}$$

$$\frac{d\boldsymbol{\beta}^T}{d\mathbf{z}_7} = \begin{bmatrix} 21.235 & 9.7328 & 123.92 \\ 18.217 & 35.769 & 5.3451 \\ -281.31 & -527.48 & -937.42 \\ -134.54 & -2.4616 & -163.58 \\ -44.255 & -54.626 & -101.26 \\ 7.1328 & -989.36 & -1438.3 \\ 4.9096 & 24.962 & -36.372 \\ -14.198 & 18.585 & -475.86 \end{bmatrix}$$

and

$$\frac{d\boldsymbol{\beta}^T}{d\mathbf{z}}_8 = \begin{bmatrix} 26.322 & 39.484 & 52.645 \\ -281.31 & -527.48 & -937.44 \\ -125.92 & 1.4873 & -113.30 \\ -51.038 & -57.736 & -140.84 \\ 8.0509 & -988.94 & -1433.0 \\ 3.3529 & 24.248 & -45.457 \\ -16.032 & 17.744 & -486.56 \end{bmatrix}$$

The sensitivities of the measurements to the parameters is given by the following matrix:

$$\frac{d\mathbf{z}}{d\mathbf{x}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ -0.42239 & 2.3665 \\ -0.20771 & -1.3575 \\ 1.0094 & -2.1139 \\ -0.027751 & -0.37527 \\ -0.32828 & 1.5127 \\ -0.023283 & -0.032475 \end{bmatrix}$$

where:

$$\mathbf{x} = \begin{bmatrix} F_B & T_R \end{bmatrix}^T$$

$$\boldsymbol{\beta} = \begin{bmatrix} A_1 & A_2 & A_3 \end{bmatrix}^T$$

and for $\frac{d\mathbf{z}}{d\mathbf{x}}$:

$$\mathbf{z} = \begin{bmatrix} F_A & F_B & F_R & T_R & X_G & X_A & X_B & X_C & X_E & X_P \end{bmatrix}^T$$

In the matrices $\frac{d\beta}{d\mathbf{z}_i}$, the measurement set $\mathbf{z}$ corresponds with sensor set i. For the different sensor systems, only the necessary rows of $\frac{d\mathbf{z}}{d\mathbf{x}}$ are included in the estimation of $\mathbf{Q}_{\mathbf{x}^*}$.

Setpoint Covariance Matrices $\mathbf{Q}_{\mathbf{x}^*}$

The following matrices are the power series approximations of the setpoint covariance matrices for the alternate sensor systems. They were obtained using Equation (C.1) of Appendix C. Each covariance value is accurate to 5 significant digits using just one term of the power series. The number in the subscript corresponds to the sensor system number. Note that these matrices are expressed such that $x_1 = F_B$ is scaled by 10^{-5}, and $x_2 = T_R$ is scaled by 10^{-3}. For sensor systems #1 through #8, the approximate setpoint covariance matrices are:

$$\mathbf{Q}_{\mathbf{x},1} = \begin{bmatrix} 0.00026481 & 8.1187 \times 10^{-5} \\ 8.1187 \times 10^{-5} & 4.0537 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},2} = \begin{bmatrix} 0.00026617 & 8.1106 \times 10^{-5} \\ 8.1106 \times 10^{-5} & 3.8644 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},3} = \begin{bmatrix} 0.00026529 & 8.1484 \times 10^{-5} \\ 8.1484 \times 10^{-5} & 4.0396 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},4} = \begin{bmatrix} 0.00069255 & 0.0002681 \\ 0.0002681 & 0.00012129 \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},5} = \begin{bmatrix} 0.00026529 & 8.0258 \times 10^{-5} \\ 8.0258 \times 10^{-5} & 3.7837 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},6} = \begin{bmatrix} 0.00059098 & 0.00021241 \\ 0.00021241 & 9.1701 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},7} = \begin{bmatrix} 0.00059371 & 0.00022344 \\ 0.00022344 & 0.00010123 \end{bmatrix}$$

and

$$\mathbf{Q}_{\mathbf{x},8} = \begin{bmatrix} 0.00058220 & 0.00021180 \\ 0.00021180 & 9.2299 \times 10^{-5} \end{bmatrix}$$

B.2.2 Setpoint Covariance Matrices $\mathbf{Q}_{\mathbf{x}^\star}$ by Simulation

The following results were obtained through a closed-loop RTO simulation, as described in Section 3.4.

$$\mathbf{Q}_{\mathbf{x},1} = \begin{bmatrix} 0.00026304 & 7.6080 \times 10^{-5} \\ 7.6080 \times 10^{-5} & 3.7391 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},2} = \begin{bmatrix} 0.00025833 & 7.4123 \times 10^{-5} \\ 7.4123 \times 10^{-5} & 3.5234 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},3} = \begin{bmatrix} 0.00025778 & 7.5663 \times 10^{-5} \\ 7.5663 \times 10^{-5} & 3.7296 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},4} = \begin{bmatrix} 0.00067533 & 0.00025218 \\ 0.00025218 & 0.00011514 \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},5} = \begin{bmatrix} 0.00026743 & 7.8492 \times 10^{-5} \\ 7.8492 \times 10^{-5} & 3.6502 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},6} = \begin{bmatrix} 0.00059339 & 0.00020791 \\ 0.00020791 & 8.9982 \times 10^{-5} \end{bmatrix}$$

$$\mathbf{Q}_{\mathbf{x},7} = \begin{bmatrix} 0.00057431 & 0.00020716 \\ 0.00020716 & 9.6350 \times 10^{-5} \end{bmatrix}$$

and

$$\mathbf{Q}_{\mathbf{x},8} = \begin{bmatrix} 0.00056613 & 0.00020055 \\ 0.00020055 & 8.8024 \times 10^{-5} \end{bmatrix}$$

In all cases, the covariance matrices found by the power series approximation are very similar to those found by simulation.

Appendix C

Setpoint Covariance Approximation

C.1 Setpoint Covariance Approximation

Since the closed-loop setpoint covariance matrix will not be known for the various sensor system alternatives, it must be approximated. Assuming that setpoint variance is primarily the result of measurement noise propagating through the RTO loop, it is possible to approximate the covariance matrix using an equation given by Forbes [1994]:

$$\mathbf{Q}_{\mathbf{x}^*} = \sum_{i=0}^{\infty} \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right)^i_{\mathbf{x}^*,\boldsymbol{\beta}^*,\mathbf{z}^*} \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \right)_{\boldsymbol{\beta}^*,\mathbf{z}^*} \mathbf{Q}_{\mathbf{z}} \cdot \left[\left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right)^i_{\mathbf{x}^*,\boldsymbol{\beta}^*,\mathbf{z}^*} \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \right)_{\boldsymbol{\beta}^*,\mathbf{z}^*} \right]^T \quad \text{(C.1)}$$

This equation is developed by considering the RTO system as a series of nonlinear maps, and representing them by linearizations for small deviations from the plant optimum [Forbes, 1994]. In this expression, the process noise is assumed to be a random sample from a Gaussian distribution, having zero mean and covariance matrix $\mathbf{Q}_{\mathbf{z}}$. This assumption is appropriate when process disturbances are stationary and approximately normally distributed. In Equation (C.1), $\mathbf{Q}_{\mathbf{x}^*}$ is expressed as an infinite weighted sum in the powers of the products of RTO subsystem sensitivities, as well as the sensitivities of

process variables to setpoint changes. Although this is an infinite sequence, it is usually possible to approximate $\mathbf{Q}_{\mathbf{x}^*}$ with a small number of terms, provided that:

$$\left\| \frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right\|_2 \ll 1 \tag{C.2}$$

As this norm approaches unity, a large number of terms would have to be included in the estimation of $\mathbf{Q}_{\mathbf{x}^*}$. The norm must be less than unity for Equation (C.1) to be true.

In order to evaluate Equation (C.1), the sensitivity matrices must be determined. Methods for obtaining these matrices analytically are outlined in the next sections. Usually, however, they are difficult to compute analytically and must be approximated using numerical techniques.

When it is impossible to obtain the derivatives of the process variables with respect to the setpoint variables, $\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$, a one step-ahead approximation of the setpoint covariance matrix can be obtained through an open-loop approximation [Forbes, 1994]:

$$\mathbf{Q}_{\mathbf{x}^*} \approx \frac{\partial \mathbf{x}^*}{\partial \mathbf{z}} \mathbf{Q}_{\mathbf{z}} \frac{\partial \mathbf{x}^{*T}}{\partial \mathbf{z}} \tag{C.3}$$

Here, $\mathbf{Q}_{\mathbf{z}}$ is the known measurement covariance matrix, and $\mathbf{z}$ are the measured variables. In the open-loop case, measurement variance only propagates through the parameter estimator and profit optimization, so $\frac{\partial \mathbf{x}^*}{\partial \mathbf{z}}$ can be decomposed as:

$$\frac{\partial \mathbf{x}^*}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \tag{C.4}$$

and Equation (C.3) can be rewritten as:

$$\mathbf{Q}_{\mathbf{x}^*} \approx \frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \mathbf{Q}_{\mathbf{z}} \left(\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} \right)^T \tag{C.5}$$

This expression for $\mathbf{Q}_{\mathbf{x}^*}$ is the first term in the power series of Equation (C.1). If higher order terms are significant, Equation (C.5) will underestimate the variance contribution

to the SSDC criterion.

C.2 Sensitivity Calculations

The sensitivities of the setpoints to the model parameters, $\frac{\partial \mathbf{x}^*}{\partial \boldsymbol{\beta}}$, may be found using the method of Ganesh and Biegler [1987]. Consider the general optimization problem:

$$\begin{aligned} &\min_{\mathbf{x}^*} \psi(\mathbf{a}, \boldsymbol{\beta}) \\ &\text{subject to:} \\ &\quad \mathbf{f}(\mathbf{a}, \boldsymbol{\beta}) = 0 \\ &\quad \mathbf{g}(\mathbf{a}, \boldsymbol{\beta}) \leq 0 \end{aligned} \tag{C.6}$$

where ψ is the objective function for the setpoint calculation problem, $\mathbf{a} = \{\mathbf{x}^*, \mathbf{d}\}$ where $\mathbf{x}^*$ are the setpoints to be calculated (decision variables), $\mathbf{d}$ are tear variables (dependent), and $\boldsymbol{\beta}$ are the parameters supplied by the model updator. The model equations are given by $\mathbf{f}$ and the inequality constraints by $\mathbf{g}$. At the optimal solution the Karush-Kuhn-Tucker (KKT) conditions are satisfied:

$$\begin{aligned} \nabla\psi(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) + \nabla\mathbf{g}(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ)\,\mathbf{u}^\circ + \nabla\mathbf{f}(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ)\,\mathbf{v}^\circ &= 0 \\ \mathbf{u}^{\circ T}\mathbf{g}(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) &= 0 \\ \mathbf{u}^\circ &\geq 0 \\ \mathbf{g}(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) &\leq 0 \\ \mathbf{f}(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) &= 0 \end{aligned} \tag{C.7}$$

where $\mathbf{a}^\circ$ is the base case optimal solution and $\mathbf{u}^\circ$ and $\mathbf{v}^\circ$ are the KKT multipliers at the base case optimum. That is, for the supplied parameters $\boldsymbol{\beta}^\circ$, $\mathbf{a}^\circ$ is a local minimum of the optimization problem with KKT multipliers $\mathbf{u}^\circ$ and $\mathbf{v}^\circ$. The first order sensitivity of the optimal setpoints with respect to the parameters is desired. Assume the following:

(*i*) The functions defining the optimization problem are at least twice continuously differentiable in $\mathbf{a}$ and at least once in $\boldsymbol{\beta}$ for a neighbourhood of $(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ)$.

(ii) The constraint gradients are linearly independent at $(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ)$ and strict complementary slackness holds for the problem at $(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ)$ with unique KKT multipliers $\mathbf{u}^\circ$ and $\mathbf{v}^\circ$.

(iii) The second-order sufficiency conditions are met.

The KKT conditions at the optimum $\mathbf{a}^\circ$ state:

$$\begin{aligned} \nabla_{\mathbf{a}} L(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) &= 0 \\ \mathbf{g}_A(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) &= 0 \\ \mathbf{f}(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) &= 0 \end{aligned} \tag{C.8}$$

where L is the Lagrange function and $\mathbf{g}_A$ are the active inequality constraints. To satisfy these conditions for a perturbation $\Delta\boldsymbol{\beta}$ in the parameters about $\boldsymbol{\beta}^\circ$, the first order corrections are found by noting that:

$$\begin{aligned} d[\nabla_{\mathbf{a}} L(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ)] &= \nabla^2_{\mathbf{a}} L^\circ d\mathbf{a} + \nabla_{\mathbf{a}} \mathbf{g}_A^\circ d\mathbf{u} + \nabla_{\mathbf{a}} \mathbf{f}^\circ d\mathbf{v} + \nabla_{\boldsymbol{\beta}\mathbf{a}} L^{T\circ} d\boldsymbol{\beta} = \mathbf{0} \\ d\mathbf{g}_A(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) &= \nabla_{\mathbf{a}} \mathbf{g}_A^{T\circ} d\mathbf{a} + \nabla_{\boldsymbol{\beta}} \mathbf{g}_A^{T\circ} d\boldsymbol{\beta} = \mathbf{0} \\ d\mathbf{f}(\mathbf{a}^\circ, \boldsymbol{\beta}^\circ) &= \nabla_{\mathbf{a}} \mathbf{f}^{T\circ} d\mathbf{a} + \nabla_{\boldsymbol{\beta}} \mathbf{f}^{T\circ} d\boldsymbol{\beta} = \mathbf{0} \end{aligned} \tag{C.9}$$

Rearranging these expressions and partitioning $\mathbf{a}$ gives the following system of linear equations:

$$\begin{bmatrix} \nabla_{\boldsymbol{\beta}\mathbf{x}^*} L^{T\circ} \\ \nabla_{\boldsymbol{\beta}\mathbf{d}} L^{T\circ} \\ \nabla_{\boldsymbol{\beta}} \mathbf{g}_A^{T\circ} \\ \nabla_{\boldsymbol{\beta}} \mathbf{f}^{T\circ} \end{bmatrix} = - \begin{bmatrix} \nabla^2_{\mathbf{x}^*} L^\circ & \nabla_{\mathbf{x}^*\mathbf{d}} L^\circ & \nabla_{\mathbf{x}^*} \mathbf{g}_A^\circ & \nabla_{\mathbf{x}^*} \mathbf{f}^\circ \\ \nabla_{\mathbf{d}\mathbf{x}^*} L^\circ & \nabla^2_{\mathbf{d}} L^\circ & \nabla_{\mathbf{d}} \mathbf{g}_A^\circ & \nabla_{\mathbf{d}} \mathbf{f}^\circ \\ \nabla_{\mathbf{x}^*} \mathbf{g}_A^{T\circ} & \nabla_{\mathbf{d}} \mathbf{g}_A^{T\circ} & 0 & 0 \\ \nabla_{\mathbf{x}^*} \mathbf{f}^{T\circ} & \nabla_{\mathbf{d}} \mathbf{f}^{T\circ} & 0 & 0 \end{bmatrix} \begin{bmatrix} \nabla_{\boldsymbol{\beta}} \mathbf{x}^{*T\circ} \\ \nabla_{\boldsymbol{\beta}} \mathbf{d}^{T\circ} \\ \nabla_{\boldsymbol{\beta}} \mathbf{u}^{T\circ} \\ \nabla_{\boldsymbol{\beta}} \mathbf{v}^{T\circ} \end{bmatrix} \tag{C.10}$$

The derivative matrix on the right hand side of Equation (C.10) contains the desired matrix $\nabla_{\beta}\mathbf{x}^{*T\circ}$ (or $\frac{\partial \mathbf{x}^{*}{}^{T}}{\partial \beta}$). For small problems, it may be possible to obtain the elements of this system of equations analytically. For large problems, however, numerical approximations would have to be used.

It may be possible to calculate the sensitivities of the model parameters to the process measurements $\frac{\partial \beta}{\partial \mathbf{z}}$, when the problem is small. Consider a least squares parameter estimation scheme. The objective function is:

$$\begin{aligned} \phi &= \mathbf{e}^{T}\mathbf{e} \\ &= \mathbf{f}^{T}(\beta, \mathbf{z})\,\mathbf{f}(\beta, \mathbf{z}) \end{aligned} \tag{C.11}$$

where β are the p parameters to be determined, and $\mathbf{z}$ are the process measurements. At the optimum:

$$\frac{\partial \phi}{\partial \beta} = 2\mathbf{f}^{T}(\beta, \mathbf{z})\frac{\partial \mathbf{f}}{\partial \beta} = \mathbf{0} \tag{C.12}$$

Define:

$$\mathbf{F}(\beta, \mathbf{z}) = \left[\mathbf{f}^{T}(\beta, \mathbf{z})\frac{\partial \mathbf{f}}{\partial \beta}\right]^{T} = \mathbf{0} \tag{C.13}$$

Using the Implicit Function Theorem [Gillett, 1984]:

$$\frac{\partial \mathbf{F}}{\partial \beta}d\beta + \frac{\partial \mathbf{F}}{\partial \mathbf{z}}d\mathbf{z} = \mathbf{0} \tag{C.14}$$

Recognizing that $\mathbf{F}(\beta, \mathbf{z})$ is a square $p \times p$ matrix, and with some rearrangement:

$$\frac{d\beta}{d\mathbf{z}} = -\left[\frac{\partial \mathbf{F}}{\partial \beta}\right]^{-1}\frac{\partial \mathbf{F}}{\partial \mathbf{z}} \tag{C.15}$$

Again, a numerical approximation may be more practical if the problem is large.

Without considerable plant knowledge, it is impossible to determine the process gain matrix $\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$. These sensitivities of process variables to the setpoints would have to be found through plant experimentation. For simulation purposes, a process model is assumed, and $\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$ can be computed analytically when the problem is small. Assume the system is completely determined. That is, once the setpoints are calculated, the resulting model consists of n equations in n unknown dependent process variables. Partition $\mathbf{z}$ such that:

$$\mathbf{z} = \begin{bmatrix} \mathbf{x}^* & \mathbf{d} \end{bmatrix}^T \tag{C.16}$$

where $\mathbf{x}^*$ are the setpoint variables and $\mathbf{d}$ are the dependent variables. It is clear that

$$\frac{\partial \mathbf{x}^*}{\partial \mathbf{x}^*} = \mathbf{I} \tag{C.17}$$

and $\frac{\partial \mathbf{d}}{\partial \mathbf{x}^*}$ may be calculated similarly to $\frac{d\beta}{d\mathbf{z}}$. The process model for simulation is:

$$\mathbf{f}(\boldsymbol{\beta}^*, \mathbf{d}, \mathbf{x}^*) = \mathbf{0} \tag{C.18}$$

where $\boldsymbol{\beta}^*$ are the perfectly known process parameters. Again using the Implicit Function Theorem [Gillett, 1984]:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{d}} d\mathbf{d} + \frac{\partial \mathbf{f}}{\partial \mathbf{x}^*} d\mathbf{x}^* = \mathbf{0} \tag{C.19}$$

Noting that in this case $\frac{\partial \mathbf{f}}{\partial \mathbf{d}}$ is a square, full-rank matrix:

$$\frac{d\mathbf{d}}{d\mathbf{x}^*} = -\frac{\partial \mathbf{f}}{\partial \mathbf{d}}^{-1} \frac{\partial \mathbf{f}}{\partial \mathbf{x}^*} \tag{C.20}$$

where $\frac{d\mathbf{d}}{d\mathbf{x}^*}$ represents the steady-state process gain. Therefore, the complete sensitivity matrix is:

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}^*} = \begin{bmatrix} \mathbf{I} \\ \frac{d\mathbf{d}}{d\mathbf{x}^*} \end{bmatrix} \tag{C.21}$$

C.2.1 Central Difference Approximation

A common numerical approach for approximating the sensitivity matrices is the central difference approximation [Krishnamurthy and Sen, 1986], and the method is demonstrated here for the estimation of $\frac{d\beta_i}{dz_j}$. While keeping all other variables constant, the parameter β_i can be written as some function of measurement z_j, $\beta_i(z_j)$, whose derivatives are single-valued, finite, and continuous functions. The following Taylor expansions can be written:

$$\beta_i(z_j + \Delta z_j) = \beta_i(z_j) + \Delta z_j \frac{\partial \beta_i}{\partial z_j} + \frac{\Delta z_j^2}{2!}\frac{\partial^2 \beta_i}{\partial z_j^2} + \frac{\Delta z_j^3}{3!}\frac{\partial^3 \beta_i}{\partial z_j^3} + \ldots \tag{C.22}$$

$$\beta_i(z_j - \Delta z_j) = \beta_i(z_j) - \Delta z_j \frac{\partial \beta_i}{\partial z_j} + \frac{\Delta z_j^2}{2!}\frac{\partial^2 \beta_i}{\partial z_j^2} - \frac{\Delta z_j^3}{3!}\frac{\partial^3 \beta_i}{\partial z_j^3} + \ldots \tag{C.23}$$

for some small Δz_j. Ignoring third order terms and higher, and subtracting Equation (C.23) from Equation (C.22):

$$\frac{\partial \beta_i}{\partial z_j} = \frac{\beta_i(z_j + \Delta z_j) - \beta_i(z_j - \Delta z_j)}{2\Delta z_j} \tag{C.24}$$

which has an error $O\left(\Delta z_j^2\right)$. Equation (C.24) can be used to approximate any of the required sensitivities involved in the estimation of $\mathbf{Q}_{\mathbf{x}^*}$, using process models and/or plant experimentation.

Appendix D

Design Cost Development

The developments of this appendix follow the Design Cost work of Forbes and Marlin [1996]. Forbes and Marlin [1996] introduced Design Cost as a measure of the loss of RTO performance due to imperfect optimization. The benchmark was the true plant optimum, and RTO performance was assessed in terms of uncaptured profit with respect to the benchmark. This development is similar to Forbes and Marlin's [1996] work, in that it quantifies the performance of a particular RTO system in terms of uncaptured profit. The difference lies in the benchmark. Forbes and Marlin's [1996] benchmark was the true plant optimum, which is generally unknown. Here, the benchmark is the best possible performance of the RTO system, given the model and the process variables which can be measured. RTO performance based on a particular sensor system is compared to this benchmark.

The Sensor System Design Cost (SSDC) can be expressed as the expected difference between the profit at the closed-loop RTO benchmark setpoints, $\mathbf{x}_m^*$, and the profit at the nominal setpoints $\mathbf{x}^*$, which are based on a particular sensor system:

$$C = E\left[P(\mathbf{x}_m^*) - P(\mathbf{x}^*)\right] \tag{D.1}$$

where C represents the unrealized profit caused by the use of an alternate model. E is

the expectation operator, and $P(\mathbf{x})$ is the plant profit at setpoints $\mathbf{x}$.

The uncertainty in the predicted setpoints $\mathbf{x}^*$ can be expressed by the probability density function $f(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*})$, where $\mathbf{Q}_{\mathbf{x}^*}$ is the covariance matrix of the predicted setpoint values. If both P and f are Lebesgue integrable on the domain of interest, the expected profit of the RTO system is:

$$E[P(\mathbf{x}^*)] = \int_{\Psi} P(\boldsymbol{\delta})\, f(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*})\, d\boldsymbol{\delta} \tag{D.2}$$

where $\boldsymbol{\delta}$ is an integration variable defined on Ψ, the space of possible predicted optimum setpoints for the given measurement set and external variable values. Substituting Equation (D.2) into Equation (D.1), the model-based uncaptured profit at $\mathbf{x}_m^*$ is:

$$C = E[P(\mathbf{x}_m^*)] - \int_{\Psi} P(\boldsymbol{\delta})\, f(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*})\, d\boldsymbol{\delta} \tag{D.3}$$

Assume that there are excess degrees of freedom in the optimization problem, so that the reduced space of the profit maximization problem has at least one dimension. Further, assume that the profit surface is at least twice differentiable in the reduced space, so the reduced Hessian exists. The profit function can then be represented by the truncated Taylor series expansion:

$$P(\boldsymbol{\delta}) = P(\mathbf{x}_m^*) + \nabla_r P\,|_{\mathbf{x}_m^*}\,(\mathbf{x}_m^* - \boldsymbol{\delta}) + \frac{1}{2}(\mathbf{x}_m^* - \boldsymbol{\delta})^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,(\mathbf{x}_m^* - \boldsymbol{\delta}) + O\left(\|\mathbf{x}_m^* - \boldsymbol{\delta}\|^3\right) \tag{D.4}$$

The reduced gradient will disappear at $\mathbf{x}_m^*$, and ignoring terms third order and higher, Equation (D.4) can be re-written as:

$$P(\boldsymbol{\delta}) = P(\mathbf{x}_m^*) + \frac{1}{2}\left[(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) + (\tilde{\mathbf{x}}^* - \boldsymbol{\delta})\right]^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,\left[(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) + (\tilde{\mathbf{x}}^* - \boldsymbol{\delta})\right] \tag{D.5}$$

Rearranging:

$$P(\boldsymbol{\delta}) = P(\mathbf{x}_m^*) + \frac{1}{2}(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) + (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}) + \frac{1}{2}(\tilde{\mathbf{x}}^* - \boldsymbol{\delta})^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}) \quad \text{(D.6)}$$

Equation (D.6) can be substituted into Equation (D.3), and the result simplified through the following developments. By definition:

$$E[\tilde{\mathbf{x}}^* - \boldsymbol{\delta}] = \mathbf{0} \quad \text{(D.7)}$$

therefore:

$$\int_{\Psi} (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,(\tilde{\mathbf{x}}^* - \boldsymbol{\delta})\, f(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*})\, d\boldsymbol{\delta} = 0 \quad \text{(D.8)}$$

Now consider the second order term in $(\tilde{\mathbf{x}}^* - \boldsymbol{\delta})$:

$$\int_{\Psi} (\tilde{\mathbf{x}}^* - \boldsymbol{\delta})^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,(\tilde{\mathbf{x}}^* - \boldsymbol{\delta})\, f(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*})\, d\boldsymbol{\delta} \quad \text{(D.9)}$$

Expanding term-wise, this expression may be re-written as:

$$\int_{\Psi} (\tilde{\mathbf{x}}^* - \boldsymbol{\delta})^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,(\tilde{\mathbf{x}}^* - \boldsymbol{\delta})\, f(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*})\, d\boldsymbol{\delta} = \sum_i \sum_j \frac{d^2 P}{dx_i dx_j} \sigma_{ij}^2(\mathbf{x}^*) \quad \text{(D.10)}$$

Since $\mathbf{Q}_{\mathbf{x}^*} = [\sigma_{ij}^2]$, Equation (D.10) can be re-expressed using the Hadamard product of the reduced Hessian of the profit function and the setpoint covariance matrix:

$$\int_{\Psi} (\tilde{\mathbf{x}}^* - \boldsymbol{\delta})^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\,(\tilde{\mathbf{x}}^* - \boldsymbol{\delta})\, f(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*})\, d\boldsymbol{\delta} = \mathbf{w}^T \left[\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right] \mathbf{w} \quad \text{(D.11)}$$

where $\mathbf{w} = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^T$.

Now $P(\mathbf{x}_m^*)$ is constant at the given operating point, and by definition

$$\int_{\Psi} f\left(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) d\boldsymbol{\delta} = \mathbf{1} \tag{D.12}$$

so that:

$$\int_{\Psi} P(\mathbf{x}_m^*) f\left(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) d\boldsymbol{\delta} = P(\mathbf{x}_m^*) \tag{D.13}$$

By combining the results from Equations (D.8) and (D.11), an expression for the expected profit is obtained:

$$\int_{\Psi} P\left(\boldsymbol{\delta}\right) f\left(\tilde{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) d\boldsymbol{\delta} =$$
$$P(\mathbf{x}_m^*) + \frac{1}{2}\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right) + \frac{1}{2}\mathbf{w}^T\left[\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right]\mathbf{w} \tag{D.14}$$

Since $E\left[P(\mathbf{x}_m^*)\right] = P(\mathbf{x}_m^*)$ the SSDC is given by:

$$C = -\frac{1}{2}\left[\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right) + \mathbf{w}^T\left[\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right]\mathbf{w}\right] \tag{D.15}$$

The goal of sensor selection is to minimize the design cost with respect to the available sensor system alternatives $\varkappa$:

$$\min_{\varkappa} -\frac{1}{2}\left[\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)^T \nabla_r^2 P\,|_{\mathbf{x}_m^*}\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right) + \mathbf{w}^T\left[\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right]\mathbf{w}\right] \tag{D.16}$$

Taking the negative sign inside each expression and eliminating the factor of 1/2, Equation (D.16) can be re-written as:

$$\min_{\varkappa}\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right)^T\left(-\nabla_r^2 P\,|_{\mathbf{x}_m^*}\right)\left(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*\right) + \mathbf{w}^T\left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right]\mathbf{w} \tag{D.17}$$

For a RTO problem with a unique optimum the reduced Hessian is negative definite, so the negative of the reduced Hessian is positive definite. The covariance matrix of the

calculated setpoints is positive definite by definition. Since the Hadamard product of two positive definite matrices is itself positive definite [Horn and Johnson, 1991] the minimum possible value of Problem (D.17) is zero. Problem (D.17) can therefore be rewritten as:

$$\min_{\varkappa} \left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) + \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \tag{D.18}$$

The terms in Equation (D.18) can be separated using the triangle inequality, which states that:

$$\left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) + \sum_{i=1}^{s} \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \right| \leq$$
$$\left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right| + \left| \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \tag{D.19}$$

All of the matrices in this expression are at least positive semi-definite, so the quadratic terms are positive or zero by definition and Inequality D.19 becomes an equality. Now Problem (D.18) can be expressed as:

$$\min_{\varkappa} \left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right| + \left| \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \tag{D.20}$$

Problem (D.20) includes the Hessian of the model-based profit function. Although it may be possible to evaluate the Hessian, the profit surface curvature of the plant may not be accurately represented in the model. In this case, Problem (D.20) can be modified so that the Hessian is removed and an upper bound on the SSDC is minimized instead. Using the Cauchy-Schwarz Inequality [Ortega, 1987], upper bounds may be placed on each of the two terms in Problem (D.20):

$$\left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right| \leq \left\| -\nabla_r^2 P \mid_{\mathbf{x}_m^*} \right\|_2 \left\| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right\|_2^2 \tag{D.21}$$

and

$$\left|\mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right] \mathbf{w}\right| \leq s \left\|-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right\|_2 \leq s \left\|\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right\|_2 \left\|\mathbf{Q}_{\mathbf{x}^*}\right\|_2 \tag{D.22}$$

where s is the number of setpoints. The upper bound on the SSDC is therefore:

$$C \leq \left\|\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right\|_2 \left(\left\|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\right\|_2^2 + s \left\|\mathbf{Q}_{\mathbf{x}^*}\right\|_2\right) \tag{D.23}$$

The minimization is therefore:

$$\min_{\varkappa} \left\|-\nabla_r^2 P \mid_{\mathbf{x}_m^*}\right\|_2 \left(\left\|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\right\|_2^2 + s \left\|\mathbf{Q}_{\mathbf{x}^*}\right\|_2\right) \tag{D.24}$$

but, recognizing that the reduced Hessian is fixed, the problem can be reduced to:

$$\min_{\varkappa} \left\|(\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)\right\|_2^2 + s \left\|\mathbf{Q}_{\mathbf{x}^*}\right\|_2 \tag{D.25}$$

Problem (D.25) is the unweighted point-wise criterion for sensor system selection. It can be seen immediately that the first term in Problem (D.25) is the T-optimal criterion of Problem (3.10). The second term in Problem (D.25) contains the largest eigenvalue of $\mathbf{Q}_{\mathbf{x}^*}$, which is proportional to the length of the largest semi-axis of the setpoint confidence region. It is also the largest contributor to the D-optimal criterion. Since all of the eigenvalues of $\mathbf{Q}_{\mathbf{x}^*}$ are less than or equal to $\left\|\mathbf{Q}_{\mathbf{x}^*}\right\|_2$, Problem (D.25) aims to minimize the sum of the T-optimal criterion and the worst case D-optimal criterion.

The SSDC Problem (D.25) represents sensor selection at a given operating point. The model-based plant optimum $\mathbf{x}_m^*$ will change however, depending on the values of the measurable variables $\mathbf{z}_c^*$. So, the model optimum is drawn from a space $\mathbf{S}$ of all possible optima. There is a frequency function $\varsigma\left(\mathbf{z}_c^*, \upsilon\right)$ associated with the optima in $\mathbf{S}$, which describes the occurrence rate of a particular plant optimum. The total uncaptured

model-based profit for a particular sensor system, for all possible disturbances, is:

$$C_T = \int \left\{ E\left[P\left(\mathbf{x}_m^*\right)\right] - \int P\left(\boldsymbol{\delta}\right) f\left(\bar{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) d\Psi \right\} \varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right) d\mathbf{S} \tag{D.26}$$

where P and ς are both integrable on $\mathbf{S}$. The maximum theoretically attainable model-based plant profit is:

$$P_T = \int P\left(\mathbf{x}_m^*\right) \varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right) d\mathbf{S} \tag{D.27}$$

So Equation (D.26) can be re-written as:

$$C_T = P_T - \int \int P\left(\boldsymbol{\delta}\right) f\left(\bar{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) \varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right) d\Psi d\mathbf{S} \tag{D.28}$$

The goal is to minimize the total loss in profit by selecting the appropriate sensor system from the set of choices $\varkappa$. This is equivalent to:

$$\min_{\varkappa} - \int \int P\left(\boldsymbol{\delta}\right) f\left(\bar{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) \varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right) d\Psi d\mathbf{S} \tag{D.29}$$

It is unlikely that $\varsigma\left(\mathbf{z}_c^*, \boldsymbol{\upsilon}\right)$ would be known, thus the minimization of Problem (D.29) will have to be approximated by a sum over an expected set of values for $\mathbf{z}_c^*$.

In RTO, as the process changes from one steady state to the next, the measurements appear to occur as steps. Therefore Problem (D.29) can be expressed as:

$$\min_{\varkappa} \sum_{\mathbf{z}_c^*} \left[- \int P\left(\boldsymbol{\delta}\right) f\left(\bar{\mathbf{x}}^* - \boldsymbol{\delta}, \mathbf{Q}_{\mathbf{x}^*}\right) d\Psi \right] \varsigma\left(\mathbf{z}_c^*\right) \tag{D.30}$$

The objective in Problem (D.30) is to minimize the weighted sum of the solutions to the point-wise SSDC Problem (D.25), for the set of possible measurement values. Then using the developments for the point-wise design cost problem, the total SSDC problem

is:

$$\min_{\varkappa} \sum_{\mathbf{z}_c^*} \left[\left\| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right\|_2^2 + s \left\| \mathbf{Q}_{\mathbf{x}^*} \right\|_2 \right] \varsigma (\mathbf{z}_c^*) \tag{D.31}$$

The solution to Problem (D.31) provides the minimum SSDC for the expected set of measurable variable values. The RTO designer can then compare sensor system alternatives for the complete range of operations, rather than one point at a time.

D.1 Weighting

The profit surface will have some topology. Thus, variance and bias may be more costly in certain directions. It may be desirable to weight the SSDC minimization in order that bias in more costly directions is penalized, and variance in more costly directions is penalized. These goals are met by solving the point-wise Problem (D.20) as:

$$\min_{\varkappa} \left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \left|_{\mathbf{x}_m^*} \right. \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right| + \left| \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \tag{D.32}$$

or the problem over the complete range of operations as:

$$\min_{\varkappa} \sum_{\mathbf{z}_c^*} \left[\left| (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*)^T \left(-\nabla_r^2 P \left|_{\mathbf{x}_m^*} \right. \right) (\mathbf{x}_m^* - \tilde{\mathbf{x}}^*) \right| + \left| \mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*} \right] \mathbf{w} \right| \right] \varsigma (\mathbf{z}_c^*) \tag{D.33}$$

The first term in each problem is the weighted T-optimal criterion from Problem (3.19). The upper limit on the second term is simply a multiple of the upper limit of the weighted D-optimal criterion. Recall the weighted D-optimal criterion:

$$\min \left| \mathbf{R} \mathbf{Q}_{\mathbf{x}^*} \mathbf{R}^T \right| \tag{D.34}$$

where

$$-\nabla_r^2 P = \mathbf{R}^T \mathbf{R} \tag{D.35}$$

is the Cholesky decomposition of the negative reduced Hessian of the profit function. The upper bound on the determinant of $\mathbf{R}\mathbf{Q}_{\mathbf{x}^*}\mathbf{R}^T$ is:

$$\left|\mathbf{R}\mathbf{Q}_{\mathbf{x}^*}\mathbf{R}^T\right| \leq \|\mathbf{R}\|_2 \|\mathbf{Q}_{\mathbf{x}^*}\|_2 \left\|\mathbf{R}^T\right\|_2 = \left\|-\nabla_r^2 P\,|_{\mathbf{x}_m^*}\right\|_2 \|\mathbf{Q}_{\mathbf{x}^*}\|_2 \tag{D.36}$$

while the upper bound on the second term of Equations (D.32) and (D.33) is:

$$\left|\mathbf{w}^T \left[-\nabla_r^2 P \circ \mathbf{Q}_{\mathbf{x}^*}\right] \mathbf{w}\right| \leq s \left\|-\nabla_r^2 P\,|_{\mathbf{x}_m^*}\right\|_2 \|\mathbf{Q}_{\mathbf{x}^*}\|_2 \tag{D.37}$$

Therefore the weighted SSDC aims to minimize the sum of the weighted T-optimal criterion and the worst case weighted D-optimal criterion. That is, it aims to minimize the lost dollars caused by both setpoint variance and bias.

Appendix E

Case Study Data for Chapter 4

This appendix provides the variance/covariance data for measurement noise, as well as the sensitivity matrices and parameter covariance matrices for each sensor system alternative.

E.1 Measurement Noise Data

All flows and temperature measurements were considered to be independent. The flow variances are given in Table E.1, where flows are expressed in lb/hr and values have been scaled by 10^{-5}. Temperature variances are equal for all temperature measurements, at 2.78×10^{-8} R, where temperatures are expressed in R and values have been scaled by 10^{-3}. The temperature measurements are all process streams, plus inlet and outlet utility temperatures for the reaction cooler water, condenser water, and reboiler steam. The covariance matrix for component weight fractions in streams L and P:

$$\mathbf{Q}_L = \begin{bmatrix} 6.14 \times 10^{-5} & -1.03 \times 10^{-4} & 0 & 0 \\ -1.03 \times 10^{-4} & 3.51 \times 10^{-4} & 0 & -2.63 \times 10^{-5} \\ 0 & 0 & 3.92 \times 10^{-4} & 0 \\ 0 & -2.63 \times 10^{-5} & 0 & 4.03 \times 10^{-6} \end{bmatrix}$$

F_A	feed component A	5.84×10^{-6}
F_B	feed component B	2.41×10^{-5}
F_D	fuel product, purge flow from column bottoms	3.54×10^{-5}
F_E	top flow from decanter	3.00×10^{-4}
F_G	flow of heavy product G from decanter	3.35×10^{-7}
F_L	recycle flow to reactor	1.12×10^{-4}
F_P	light product flow, top of column	6.3×10^{-7}
F_R	reactor exit flow	3.20×10^{-4}
F_S	flow from column bottoms	2.73×10^{-4}
F_X	flow from reaction cooler	3.20×10^{-4}
FU_{XW}	utility flow, reaction cooler water	7.93×10^{-4}
FU_{DW}	utility flow, condenser cooling water	0.00762
FU_{DS}	utility flow, reboiler steam	0.0224

Table E.1: Flow variances, Williams-Otto plant study.

$$\mathbf{Q}_P = \begin{bmatrix} 1.67 \times 10^{-7} & -2.79 \times 10^{-7} & 0 & 0 \\ -2.79 \times 10^{-7} & 9.53 \times 10^{-7} & 0 & -1.14 \times 10^{-5} \\ 0 & 0 & 1.06 \times 10^{-6} & 0 \\ 0 & -1.14 \times 10^{-5} & 0 & 2.78 \times 10^{-4} \end{bmatrix}$$

where the rows and columns correspond to components A, B, E and P, respectively. The covariance matrix for stream R is:

$$\mathbf{Q}_R = \begin{bmatrix} 5.26 \times 10^{-5} & -8.81 \times 10^{-5} & 0 & 0 & 0 \\ -8.81 \times 10^{-5} & 3.01 \times 10^{-4} & 0 & 0 & -4.81 \times 10^{-5} \\ 0 & 0 & 3.35 \times 10^{-4} & 0 & 0 \\ 0 & 0 & 0 & 2.62 \times 10^{-6} & 0 \\ 0 & -4.81 \times 10^{-5} & 0 & 0 & 1.57 \times 10^{-5} \end{bmatrix}$$

where the rows and columns correspond to components A, B, E, G and P, respectively.

E.2 Sensitivity Data

This section provides the sensitivity data for the matrices $\frac{\partial \beta}{\partial z}$ and $\frac{\partial x}{\partial \beta}$. All values were found by central difference approximation about the nominal values for $\mathbf{z}$ and $\boldsymbol{\beta}$, for each of the two sensor system alternatives. The size of perturbations were about $\pm 0.3\%$ for flows, $\pm 1.5\%$ for temperatures, $\pm 0.2\%$ for compositions, $\pm 0.5\%$ for the frequency factors, $\pm 0.2\%$ for the reaction cooler heat transfer coefficient, and $\pm 0.5\%$ for the relative volatility. Exceptions are pointed out in the following subsections. It was important that the size of perturbation be large compared to the level of accuracy in the solver, but still represent a small change with respect to the perturbed variable.

E.2.1 Sensitivities of Parameters to Measurements

The following tables give the sensitivity data for matrices $\frac{d\beta}{dz}$. Since the data are different depending on the sensor system, this section is split into two parts.

Sensor system #1

Process Flows	A_1	A_2	U_{rc}	α
F_A	-126.0	-0.6881	0	-0.2890
F_B	75.41	19.87	0	-2.169
F_D	0	0	0	0
F_E	0.0009631	-0.3531	0	-0.05008
F_G	28.53	388.9	0	6.361
F_L	-116.3	26.33	0	0.8364
F_P	127.6	-139.4	0	16.72
F_R	120.0	-8.032	0.1635	0.1335
F_S	-59.35	-58.74	0	-0.4033
F_X	48.78	49.39	0	-0.1255
FU_{XW}	0	0	-0.04153	0
FU_{DW}	-0.3394	-0.3018	0	-0.1055
FU_{DS}	0.1985	0.1765	0	0.02338

Process Temperatures	A_1	A_2	U_{rc}	α
T_A	-10.93	-6.018	0	0.005031
T_B	-22.21	-12.23	0	0.009434
T_D	0	0	0	0
T_E	0	0	0	0
T_G	0	0	0	0
T_L	-47.76	-26.29	0	0.020238
T_P	-205.8	-183.1	0	-64.015
T_R	-2030	-621.5	0.2351	-0.03459
T_S	-15.48	-13.77	0	-1.824
T_X	0	0	-6.071	0
$TUin_{XW}$	0	0	-5.375	0
$TUin_{DW}$	53.81	47.86	0	16.76
$TUin_{DS}$	56.54	50.28	0	6.66
$TUout_{XW}$	0	0	11.20	0
$TUout_{DW}$	179.6	159.7	0	55.82
$TUout_{DS}$	24.17	21.49	0	2.845

Stream L Components	A_1	A_2	U_{rc}	α
X_A	-325.3	-85.36	0	1.107
X_B	-339.9	-40.84	0	1.806
X_E	108.6	-15.60	0	7.917
X_P	52.2	-155.3	0	-88.01

Stream P Components	A_1	A_2	U_{rc}	α
X_A	-32.63	-13.53	0	-2.750
X_B	-50.27	-11.36	0	-3.179
X_E	48.5	-10.28	0	-2.854
X_P	4.193	-15.69	0	-0.7795

Sensor system #2

Process Flows	A_1	A_2	U_{rc}	α
F_A	82.37	35.70	0	1.752
F_B	167.3	25.75	0	11.45
F_D	0	0	0	0
F_E	-0.05955	0.1263	0	-0.1692
F_G	0	0	0	0
F_L	5.505	25.59	0	-4.590
F_P	55.13	-31.98	0	104.4
F_R	-1.985	-13.50	0.1635	-0.5637
F_S	-1.811	-14.66	0	1.441
F_X	-0.7370	14.92	0	-5.382
FU_{XW}	0	0	-0.04153	0
FU_{DW}	0.1142	0.1985	0	-0.2972
FU_{DS}	-0.06562	-0.1141	0	0.1317

Process Temperatures	A_1	A_2	U_{rc}	α
T_A	-3.037	-1.894	0	-0.1780
T_B	-6.172	-3.848	0	-0.3623
T_D	0	0	0	0
T_E	0	0	0	0
T_G	0	0	0	0
T_L	-13.27	-8.275	0	-0.7786
T_P	69.14	120.2	0	-180.3
T_R	-2163	-675.1	0.2351	1.318
T_S	5.115	8.885	0	-10.25
T_X	0	0	-6.071	0
$TUin_{XW}$	0	0	-5.375	0
$TUin_{DW}$	-18.12	-31.49	0	47.15
$TUin_{DS}$	-18.68	-32.48	0	37.51
$TUout_{XW}$	0	0	11.20	0
$TUout_{DW}$	-60.36	-104.9	0	158.4
$TUout_{DS}$	-8.030	-13.96	0	16.10

Stream P Components	A_1	A_2	U_{rc}	α
X_A	2.781	3.000	0	-5.281
X_B	4.718	3.359	0	-7.346
X_E	-7.378	2.451	0	-6.280
X_P	2.431	2.232	0	-0.007778

Stream R Components	A_1	A_2	U_{rc}	α
X_A	-464.2	-118.5	0	2.340
X_B	-426.7	-54.64	0	-1.308
X_E	65.51	-9.495	0	32.52
X_G	-67.48	404.5	0	-37.80
X_P	33.65	-212.0	0	-214.2

E.2.2 Sensitivities of Setpoints to Parameters

This section is again split into two parts, because the nominal parameter values are different for each sensor system and the perturbations are about the nominal values.

Sensor System #1

	F_B	T_R
A_1	-0.002356	-0.0003106
A_2	0.006279	-0.0005802
U_{rc}	0	0
α	-0.005127	-0.0009550

Sensor System #2

	F_B	T_R
A_1	-0.002236	-0.0002952
A_2	0.005948	-0.0005741
U_{rc}	0	0
α	-0.004067	-0.0007841

Clearly the parameter U_{rc} has no effect on the setpoint variables. It seems this was a poor choice of a parameter to update on-line.

E.3 Parameter Covariance Matrices

The parameter covariance matrices were found by approximation using Equation (C.5), and by simulation.

E.3.1 Covariance Matrices by Approximation

Sensor System #1

$$\mathbf{Q}_{\hat{\beta}} = \begin{bmatrix} 38.06 & 1.289 & 0.006265 & -0.6127 \\ 1.289 & 2.114 & -0.004243 & -0.08297 \\ 0.006265 & -0.004243 & 1.524 \times 10^{-5} & 6.987 \times 10^{-6} \\ -0.6127 & -0.08297 & 6.987 \times 10^{-6} & 0.06583 \end{bmatrix}$$

Sensor System #2

$$\mathbf{Q}_{\hat{\beta}} = \begin{bmatrix} 34.94 & -1.242 & -0.0001180 & -3.594 \\ -1.242 & 0.8475 & -0.0007106 & -0.04100 \\ -0.0001180 & -0.0007106 & 1.524 \times 10^{-5} & -2.948 \times 10^{-5} \\ -3.594 & -0.04100 & -2.948 \times 10^{-5} & 1.078 \end{bmatrix}$$

E.3.2 Covariance Matrices by Simulation

Sensor System #1

$$\mathbf{Q}_{\hat{\beta}} = \begin{bmatrix} 39.75 & 2.025 & 0.007342 & -0.6040 \\ 2.025 & 2.204 & -0.0001080 & -0.1368 \\ 0.007342 & -0.0001080 & 1.541 \times 10^{-5} & 1.376 \times 10^{-5} \\ -0.6040 & -0.1368 & 1.376 \times 10^{-5} & 0.08013 \end{bmatrix}$$

Sensor System #2

$$\mathbf{Q}_{\hat{\beta}} = \begin{bmatrix} 32.43 & -1.825 & -0.0002672 & -0.9486 \\ -1.825 & 0.8261 & -0.0005874 & 0.03734 \\ -0.0002672 & -0.0005874 & 1.580 \times 10^{-5} & 4.302 \times 10^{-5} \\ -0.9486 & 0.03734 & 4.302 \times 10^{-5} & 0.1158 \end{bmatrix}$$

For both systems, the approximated matrices are very similar to the simulated results. There is some deviation for the relative volatility in the case of the second sensor system. The second sensor system has a smaller variance for both of the frequency factors.