Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems
The International Federation of Automatic Control
December 18-20, 2013. Mumbai, India

Using Process Data for Finding Self-optimizing Controlled Variables ⋆

Johannes Jäschke and Sigurd Skogestad

Norwegian University of Science and Technology, NTNU, Trondheim, Norway. (e-mail:jaschke@ntnu.no, skoge@ntnu.no)

Abstract: In the process industry it is often not known how well a process is operated, and without a good model it is difficult to tell if operation can be further improved. We present a data-based method for finding a combination of measurements which can be used for obtaining an estimate of how well the process is operated, and which can be used in feedback as a controlled variable. To find the variable combination, we use past measurement data and fit a quadratic cost function to the data. Using the parameters of this cost function, we then calculate a linear combination of measurements, which when held constant, gives near-optimal operation. Unlike previously published methods for finding self-optimizing controlled variables, this method relies only on past plant measurements and a few plant experiments to obtain the process gain. It does not require a model which is optimized off-line to find the controlled variable.

Keywords: Process Optimization, Control, Partial least squares, Empirical modelling, Self-optimizing control

1. INTRODUCTION

Rising competition in a global market, environmental challenges and governmental regulations make it increasingly necessary to operate chemical plants close to optimality. At the same time one is often faced with not knowing exactly how well or poorly the plant is operated in practice, and which options that may exist to systematically improve operation. If a suitable first principle plant model is available, it may be used online for monitoring the performance, or for real-time optimization (RTO). In RTO a mathematical optimization problem is solved to find the optimal operating parameters for the process [Marlin and Hrymak, 1997]. Here the plant measurements are primarily used to update the model parameters in the online optimization problem.

Alternatively, the model may be used offline to develop a suitable control strategy, which results in an acceptable loss, a "self-optimizing" control structure. According to Skogestad [2000], "Self-optimizing control is when we can achieve an acceptable loss with constant setpoint values for the controlled variables (without the need to re-optimize when disturbances occur)". Another closely related concept is NCO tracking [Srinivasan et al., 2003], where the necessary conditions for optimality (NCO) are selected as the controlled variables. Usually the controlled variables in these offline approaches are found by optimizing a suitable process model.

Often a good first principle model is not available because it is prohibitively expensive to develop and maintain a model which accurately reflects the process. One class of approaches which does not require a model, is extremum seeking control [Krstic and Wang, 2000], where the process is persistently excited in order to obtain gradient information. However, for many practical process applications, persistent excitations are undesirable. An attractive alternative is to use empirical data-based models, such as regression models and partial least squares models. In virtually all chemical plants, data is collected almost continuously, and using this data to model and subsequently optimize the process seems very attractive. It can lead to significant operational savings, while requiring a relatively limited effort to develop and maintain the models. Not surprisingly, there is a large body of general literature on empirical and data based modelling, see e.g. Box and Draper [1987], Esbensen [2004], and it is widely used in industry.

In the context of process control, data-based approaches have often been used for soft-sensing applications (e.g. Lin et al. [2007]), where available measurements are used to estimate an unmeasured variable. In view of online process optimization, there have been many suggestions over the years, including evolutionary operation (EVOP) [Box, 1957], dating back to the 1950s, and in more recent years McGregor and coworkers [Yacoub and MacGregor, 2004].

In the context of using offline process optimization to find simple operational policies, there has been much less activity. Data-based methods have been developed and used by Jäschke and Skogestad [2011b] and Skogestad et al. [2011], where it is assumed that *optimal* data is available, and it is used to find optimal controlled variables. Ye et al. [2013] have also used regression methods to find controlled variables. However, they rely on a process model to generate the data, and they assume that disturbance measurements are available for the regression.

The contribution of this paper is to show how non-optimal open-loop plant measurement data can be used to 1)

⋆ This work was supported by the Norwegian Research Council

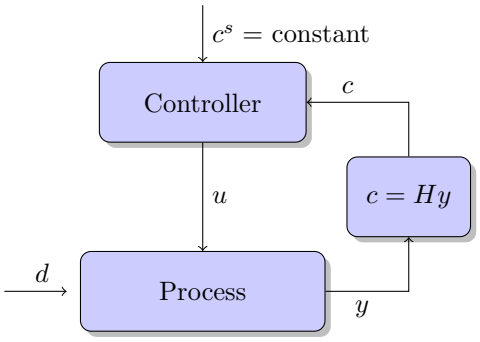

Fig. 1. Idea of self-optimizing control: By controlling $c = Hy$ at a constant setpoint, the process is kept close to optimal in presence of varying disturbances d.

obtain a quadratic model of the cost function, and 2) find simple self-optimizing controlled variables, which when controlled at constant setpoints, keep the process close to the optimum.

The paper is structured as follows. In the next section, we present some related background from self-optimizing control. Section 3 describes how these results can be used to obtain self-optimizing controlled variables from operating data. In Section 4 we apply our approach to a case study of a CSTR and present some simulation results. Finally, in Section 5 we discuss the results and draw conclusions.

2. SELF-OPTIMIZING CONTROL

In this section, we briefly give some relevant background on self-optimizing control. The general idea is to find variables which, when controlled at a constant setpoint, result in near-optimal operation with acceptable loss [Skogestad, 2000]. A possible implementation scheme of a self-optimizing control structure is given in Fig. 1. The idea is that near-optimal operation is achieved by controlling the controlled variables $c = Hy$ at constant setpoints.

We assume that the problem of optimally operating a process at steady state (or a sequence of steady states as d varies) can be formulated as a mathematical optimization problem,

$$\min_{u,x} \bar{J}(u,x,d) \quad \text{s.t.} \quad \begin{cases} h(u,x,d) = 0 \\ g(u,x,d) \leq 0 \end{cases} \tag{1}$$

where the variables u, x, and d denote the degrees of freedom, the state variables, and the disturbances, respectively. The scalar function $\bar{J}$ denotes the cost function, $h : \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \mapsto \mathbb{R}^{n_x}$ the model equations, and $g : \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \mapsto \mathbb{R}^{n_g}$ denotes the operational and safety constraints.

In addition, we assume that we have a plant measurement model

$$y = f^y(u,x,d), \tag{2}$$

where y is the n_y-dimensional vector of measurements, and f^y is the function mapping the variables u, x and d to onto the measurement space.

The following policy for implementing optimal operation is suggested [Skogestad, 2000]:

(a) Control the active constraints at their optimal values.

(b) Control "self-optimizing" variables for the remaining unconstrained degrees of freedom.

When the active constraints are controlled, they may formally be eliminated together with the states x from the optimization problem (1). This enables us to rewrite the problem for part (b) as an unconstrained lower-dimensional optimization problem,

$$\min_u J(u,d). \tag{3}$$

Around the nominal operating point $[u^{*T}, d^{*T}]$, we approximate the cost function using a second-order Taylor expansion, where $\Delta u = u - u^*$ and $\Delta d = d - d^*$:

$$\begin{aligned} J \approx J^* + \begin{bmatrix} J_u^* & J_d^* \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} \\ + \frac{1}{2} \begin{bmatrix} \Delta u^T & \Delta d^T \end{bmatrix} \begin{bmatrix} J_{uu}^* & J_{ud}^* \\ J_{du}^* & J_{dd}^* \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} \end{aligned} \tag{4}$$

where $J_u^* = \frac{\partial J}{\partial u}\big|_*$, $J_d^* = \frac{\partial J}{\partial d}\big|_*$, and $J_{uu}^* = \frac{\partial^2 J}{\partial u^2}\Big|_*$, $J_{ud}^* = J_{du}^{*\,T} = \frac{\partial J^2}{\partial u \partial d}\Big|_*$ and $J_{dd}^* = \frac{\partial^2 J}{\partial d^2}\Big|_*$ are the first and second derivatives, evaluated at the nominal point.

Under *optimal nominal operation*, we have that $J_u^* = 0$. Under the same assumptions used to obtain Eq. (4), the gradient can be approximated around the optimal nominal point ($J_u^* = 0$) as

$$J_u = \underbrace{J_u^*}_{=0} + \begin{bmatrix} J_{uu}^* & J_{ud}^* \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}. \tag{5}$$

For optimal operation, the first-order optimality conditions require that the gradient is zero, i.e

$$J_u = \begin{bmatrix} J_{uu}^* & J_{ud}^* \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} = 0. \tag{6}$$

If we could measure or evaluate the gradient, it would be the ideal self-optimizing controlled variable. Unfortunately, this is not the case in practice. Although the cost is measured, we cannot simply fit a response surface to it, because it depends on the disturbance variables d, which can be neither measured nor manipulated. Instead, we propose to approximate the gradient in terms of measurements y only, and use this as a self-optimizing variable [Jäschke and Skogestad, 2011a].

To express the gradient (6) in terms of the plant measurements, we linearize the measurement model (2) around the nominal operating point, and upon eliminating the state variables, we obtain

$$y = G^y \Delta u + G_d^y \Delta d = \tilde{G}^y \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}. \tag{7}$$

If there is a sufficient number of measurements available [1], i.e. $n_y \geq n_u + n_d$, we can use the measurement model (7) to eliminate the unknowns u and d from the gradient, and thus obtain a controlled variable which is equivalent to the gradient. Inverting (7), and inserting into (5) yields

$$J_u = \begin{bmatrix} J_{uu}^* & J_{ud}^* \end{bmatrix} \left[\tilde{G}^y\right]^\dagger \Delta y, \tag{8}$$

where $(\cdot)^\dagger$ denotes the pseudo-inverse. Defining

[1] The degrees of freedom u are generally also included in the measurement vector y.

$$H = [\, J^*_{uu} \;\; J^*_{ud} \,] \left[\tilde{G}^y\right]^\dagger \tag{9}$$

we have that the desired self-optimizing controlled variable is

$$\Delta c = H \Delta y. \tag{10}$$

Controlling $\Delta c = c - c^*$ to zero gives optimal operation. This is equivalent to the "null-space method" [Alstad and Skogestad, 2007] [2].

The derivation of H does not take measurement noise into account and assumes that there are sufficiently many independent measurements such that $\tilde{G}^y$ can be inverted. If the measurement noise is large, or if there are too few measurements, it will not be possible to make the gradient close to zero, and there will be an additional loss. These cases are treated in Alstad et al. [2009]. In this paper, we assume that the measurement noise is negligible, and that there are enough independent measurements such that $\tilde{G}^y$ can be inverted.

3. OBTAINING H FROM OPERATIONAL DATA

In the previous section, we have shown how a locally optimal controlled variable combination $\Delta c = H\Delta y$, with H given in (9), can be obtained from a linear process model (7) and a quadratic approximation of the cost function (4). However, this assumes that we know J_{uu}, J_{ud} and $\tilde{G}^y$, which may be difficult to obtain in practice. Here, we show how the H-matrix in (9) can be obtained from historical process operation data. The idea is to express the cost function in terms of the measurements y, and to use available measurement data to estimate the cost parameters.

3.1 Fundamental relationships

To obtain the desired controlled variable combination, we first express the quadratic cost function (4) in terms of the measurements. Solving (7) for $[\, \Delta u \;\; \Delta d \,]^T$ and inserting into the cost function (4) gives

$$\begin{aligned} J &= J^* + \underbrace{[\, J^*_u \;\; J^*_d \,] \left[\tilde{G}^y\right]^\dagger}_{J^*_y} \Delta y \\ &\quad + \frac{1}{2}\Delta y^T \underbrace{\left[\tilde{G}^y\right]^{\dagger^T} \begin{bmatrix} J^*_{uu} & J^*_{ud} \\ J^*_{du} & J^*_{dd} \end{bmatrix} \left[\tilde{G}^y\right]^\dagger}_{J^*_{yy}} \Delta y \\ &= J^* + J^*_y \Delta y + \frac{1}{2} y^T J^*_{yy} \Delta y. \end{aligned} \tag{11}$$

Inspecting the term J^*_{yy} closer, we see that H from (9) is contained in it:

$$\begin{aligned} J^*_{yy} &= \left[\tilde{G}^{y\dagger}\right]^T \begin{bmatrix} [\, J^*_{uu} \;\; J^*_{ud} \,] \tilde{G}^{y\dagger} \\ [\, J^*_{du} \;\; J^*_{dd} \,] \tilde{G}^{y\dagger} \end{bmatrix} \\ &= \left[\tilde{G}^{y\dagger}\right]^T \begin{bmatrix} H \\ [\, J^*_{du} \;\; J^*_{dd} \,] \tilde{G}^{y\dagger} \end{bmatrix} \end{aligned} \tag{12}$$

Since $n_y \geq n_u + n_d$, we have that $\tilde{G^y}^T \tilde{G^y}^{\dagger^T} = I$, so the upper n_u rows of $\tilde{G^y}^T J^*_{yy}$ are exactly the H-matrix given in equation (9),

$$\tilde{G^y}^T J^*_{yy} = \begin{bmatrix} H \\ [\, J^*_{du} \;\; J^*_{dd} \,] \tilde{G}^{y\dagger} \end{bmatrix}. \tag{13}$$

For control purposes we are primarily interested in H, so we do not need all elements in $\tilde{G^y}$, but only the first part, G^y. Thus, we obtain H by premultiplying J^*_{yy} with $\left[\, G^y \;\; 0_{n_y \times n_d} \,\right]^T$, which yields

$$\left[\, G^y \;\; 0_{n_y \times n_d} \,\right]^T J^*_{yy} = \begin{bmatrix} H \\ 0_{n_d \times n_y} \end{bmatrix}. \tag{14}$$

In summary, given J^*_{yy} and the gain matrix G^y, we can easily calculate the optimal measurement combination H, and use it to control the process.

Remark 1. To find controlled variables without a rigorous model, it is an advantage that only $G^y = \frac{\partial y}{\partial u}$ is required (instead of the full matrix $\tilde{G}^y = [G^y \, G^y_d]$), because G^y can be easily found using a few plant experiments. On the other hand, obtaining $G^y_d = \frac{\partial y}{\partial d}$ from plant experiments is difficult, because it is not possible to manipulate the disturbance d.

3.2 Obtaining G^y

One approach to obtain the measurement gain matrix $G^y = \frac{\partial y}{\partial u} = [g^{(1)}, \ldots, g^{(n_u)}]$, is to perform step changes in the inputs u_i and record the changes in the outputs y. The rows $i = 1..n_u$ of the gain matrix can then be calculated by

$$g^{(i)} = \left(y^{\text{pert}} - y^*\right) / \left(u_i^{pert} - u_i^*\right),$$

where the subscript i of the input u denotes the i-th input, and the superscript pert denotes the perturbed value. For better accuracy, one may perform several plant experiments of this kind and use the average value of the gain.

3.3 Obtaining J_{yy}

Gathering the data. Before we proceed to gather data to find J^*_{yy}, we make some assumptions:

(1) The data is collected while the process is operating in open loop.
(2) The number of independent measurements is greater or equal to the number of independent inputs and disturbances [3], $n_y \geq n_u + n_d$.
(3) Active constraints are controlled and not changing [4].
(4) Important disturbance changes are present in the data [5].

[2] Note that the self-optimizing control paradigm is fundamentally different from the alternative "neighbouring extremal control" approach (see e.g. Gros et al. [2009]), because the goal in self-optimizing control is to find a variable *offline*, which is held constant online. The term "neighbouring extremal", however, typically refers to schemes, where the gradient is estimated *online* in the controller. Nevertheless, at the basis of both approaches lies the first-order optimality condition $J_u = 0$.

[3] This can be tricky, because one might not know what unmeasured disturbances may affect the plant.

[4] The active constraints can often be identified by physical insight, without need of a model

[5] If the data is taken from a sufficiently long period, it is reasonable to assume that all relevant disturbances are present in the data.

(5) The data is collected when the plant is at steady state.
(6) The process data is sampled in a region close to the optimum, where the cost can be approximated by a quadratic cost function.

We collect all the "raw" measurement data in the matrix Y_{raw},

$$Y_{raw} = \left[\, y^{(1)} \;\ldots\; y^{(i)} \;\ldots\; y^{(N)} \,\right],$$

where the superscript (i) denotes the sample number.

Preparing the data. Before using using the data further, it should be centered by subtracting the mean, and scaled such that the variance of the measurements is equal. Note that the data now is in form of deviation variables.

In order to obtain a quadratic model, we need to take also the product of measurements into account. This is done by augmenting the data by all second order terms, such that each column of the data matrix Y contains data corresponding to (where the Δ for marking deviation variables has been omitted):

$$y_{aug} = \left[\, y_1 .. y_{n_y} \;\; y_1^2 \;\; y_1 y_2 \cdots y_1 y_{n_y} \;\; y_2 y_2 \;\; y_2 y_3 \ldots y_{n-1} y_n \;\; y_n^2 \,\right]^T$$

In addition, we assume that the cost function can be measured at each sample time, and we collect all cost data into a separate [6] vector:

$$J_m = \left[\, J^{(1)} \;\ldots\; J^{(i)} \;\ldots\; J^{(N)} \,\right]^T$$

Partial least squares regression If the measurements y were independent variables, we could simply use regression to fit the measurements y to the quadratic cost function in order to find J^*, J_y^* and J_{yy}^*. Unfortunately, the measurements will generally *not* be independent, and simply fitting a cost function to y-data will result in an ill-posed optimization problem, and give very poor results. To obtain a sufficiently good estimate of the cost function parameters, we use a "partial least squares" (PLS) method [Esbensen, 2004]. This method enables us to handle collinearity and linear dependence in the data well, and can be used to find a model which describes the cost function well.

The basic idea of PLS is to find a linear transformation which explains the variation in the prediction variables (in our system: Y) as well as the variation in the response variables (in our case J_m). The PLS algorithm projects the Y and J_m data onto a lower dimensional space, which still captures all the essential correlations:

$$\begin{aligned} Y^T &= TP^T + E_1 \\ J_m &= UQ^T + E_2 \end{aligned} \tag{15}$$

The matrices T, P, U and Q are chosen such that the covariance between the data Y^T and J_m is maximized, and E_1, E_2 are the residuals. Based on this decomposition, a regression factor β is determined, which predicts J as a function of y_{aug}. After applying the PLS algorithm, which is implemented e.g. in Matlab, the prediction of the cost can be calculated as

$$J = [1 \quad y_{aug}^T]\beta, \tag{16}$$

where the vector β is obtained from the PLS algorithm. We do not present further details of PLS here, instead we refer to the literature, see e.g. Esbensen [2004], where the procedure is described in detail.

[6] Note that the measurements of the cost function must not be included in the data matrix Y, because this would cause the model to use J_m to predict the cost, and we would not obtain a quadratic model

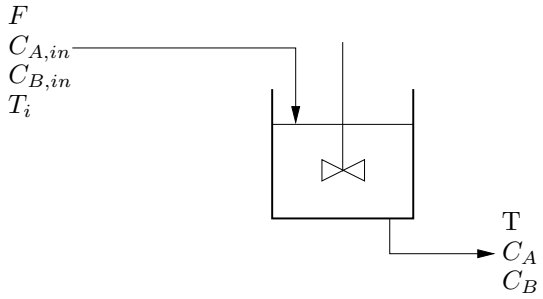

Fig. 2. Simple CSTR

Since y_{aug} contains all the products of the measurements with each other, we can simply re-arrange Eq. (16) into the form of Eq. (11). For example, in the case of 2 measurements, the augmented measurement vector is

$$y_{aug} = \left[\, y_1 \;\; y_2 \;\; y_1^2 \;\; y_1 y_2 \;\; y_2^2 \,\right]^T,$$

and Eq. (16) becomes

$$J = \beta_0 + y_1\beta_1 + y_2\beta_2 + y_1^2\beta_3 + y_1 y_2\beta_4 + y_2^2\beta_5$$

This can be re-written as the quadratic form given in Eq. (11),

$$J = \beta_0 + \left[\, \beta_1 \;\; \beta_2 \,\right] \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \frac{1}{2} \left[\, y_1 \;\; y_2 \,\right] \begin{bmatrix} 2\beta_3 & \beta_4 \\ \beta_4 & 2\beta_5 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \tag{17}$$

and comparing the coefficients in (17) with (11), we see that

$$J^* = \beta_0, \qquad J_y^* = \left[\, \beta_1 \;\; \beta_2 \,\right], \qquad J_{yy}^* = \begin{bmatrix} 2\beta_3 & \beta_4 \\ \beta_4 & 2\beta_5 \end{bmatrix}.$$

Thus we have obtained all parameters of the quadratic cost function (11), which are required for calculating H.

4. EXOTHERMIC CSTR CASE STUDY

4.1 Process description

To demonstrate our approach, we consider a simple CSTR studied by Economou and Morari [1986], Kariwala [2007], Jäschke and Skogestad [2011a,b]. A schematic diagram of the process is given in Fig. 2, where also the main variables are introduced. The feed stream containing mainly component A enters the reactor, where an equilibrium reaction

$$A \rightleftharpoons B$$

takes place. The reactor effluent contains a mixture of A and B with the same concentrations as in the tank. The manipulated variable is the feed temperature T_i, which can be adjusted to minimize the operating costs, which are calculated as the cost for heating the feed minus the income generated by selling the product B,

$$J = -\left(p_B c_B - (p_{T_i} T_i)^2\right).$$

From the mass and energy balances we obtain the following dynamic system:

$$\frac{dC_A}{dt} = \frac{1}{\tau}\left(C_{A,in} - C_A\right) - r \tag{18}$$

$$\frac{dC_B}{dt} = \frac{1}{\tau}\left(C_{B,in} - C_B\right) + r \tag{19}$$

$$\frac{dT}{dt} = \frac{1}{\tau}\left(T_i - T\right) + 5r, \tag{20}$$

Table 1. CSTR parameters

Symbol	Parameter description	Value
p_B	Price for product c_B	2.009 \$l/mol
p_{T_i}	Price for heating feed	$1.657e^{-3}\$/K$
τ	Time constant	1 min
$C^*_{A,in}$	Nominal feed concentration A	1 mol/l
$C^*_{B,in}$	Nominal feed concentration B	0 mol/l
T^*_i	Nominal input value	424.303 K

where the variables C_A and C_B denote the concentrations of component A and B, respectively, in the reactor. The temperature in the reactor is denoted by T, and the feed temperature is denoted by T_i. The feed concentrations of the two components are denoted by the variables $C_{A,in}$ and $C_{B,in}$. Finally, the reaction rate r is calculated as

$$r = 5000e^{-\frac{10000}{1987}\frac{1}{T}}C_A - 10^6 e^{-\frac{15000}{1987}\frac{1}{T}}C_B. \quad (21)$$

The values of the price parameters, the time constant τ and the nominal process values are given in Table 1. We assume that the process has four measurements, which are

$$y = [\, C_A,\, C_B,\, T,\, T_i \,]^T, \quad (22)$$

and two unmeasured disturbances, namely the feed concentrations:

$$d = [\, C_{A,in},\, C_{B,in} \,]^T. \quad (23)$$

The exact values of the disturbances are not known under operation, but we assume that it is known that $C_{A,in} = 0.5\frac{\text{mol}}{\text{l}} - 1.5\frac{\text{mol}}{\text{l}}$, and $C_{B,in} = 0\frac{\text{mol}}{\text{l}} - 0.5\frac{\text{mol}}{\text{l}}$.

This process has an unconstrained optimum, because increasing the feed temperature will lead to increased production of C_B, which contributes to lowering the cost. At the same time a higher feed temperature contributes to increasing the cost function due to the price for heating. Therefore, the optimum occur at some optimal trade-off point. The minimally required number of measurements for this process is $n_u + n_d = 1 + 2 = 3$, so there are enough measurements to apply our approach.

4.2 Simulations and Results

To generate the measurement data we run the CSTR in open loop with random inputs T_i in the range of 420-430 K, and disturbances in the ranges given above. We added some measurement noise to the data to make the case study more realistic. The measurement noise on the concentrations is uniformly distributed with within $\pm 0.01\frac{\text{mol}}{\text{l}}$, and the noise for the temperature measurements varies uniformly between $\pm$0.5K.

A total of 1000 samples was taken. The first 500 samples were used to calibrate the model, and the rest were used to validate the model. Based on inspection of the residuals for the two data sets, it was found that a PLS model with 10 components reproduced both data sets reasonably well.

The gain matrix G^y was obtained by step testing as

$$G^y = [\, -0.0012,\, 0.0012,\, 1.0057,\, 1.0 \,]^T, \quad (24)$$

and the data for our cost function obtained with the PLS regression is

Table 2. Loss comparison without noise

Controlled variable	Average loss	Worst case loss
$H_{min\,loss}$	0.0132	0.0632
H_{data}	0.0260	0.1131

Table 3. Loss comparison with noise

Controlled variable	Average loss	Worst case loss
$H_{min\,loss}$	0.0133	0.0632
H_{data}	0.0267	0.1193

$$J^* = 0.0022$$
$$J^*_y = [\, -0.2070\ -5.4507\ 0.0089\ -0.0028 \,]$$
$$J^*_{yy} = \begin{bmatrix} -0.0783 & -0.2955 & -0.0083 & -0.0164 \\ -0.2955 & 0.4341 & 0.0135 & 0.0035 \\ -0.0083 & 0.0135 & 0.0017 & -0.0001 \\ -0.0164 & 0.0035 & -0.0001 & -0.0016 \end{bmatrix}. \quad (25)$$

Applying our data based procedure from Section 3, we obtain the H-matrix as

$$H_{data} = [\, 0.3205\ -0.2245\ -0.0206\ 0.0219 \,]. \quad (26)$$

We compare the performance of the process when controlling $\Delta c = H\Delta y$ with the truly optimal operation for 1000 random disturbances for the cases with and without measurement noise. For comparison, we have also simulated the CSTR with the controlled variable combination obtained from the model based "minimum loss method" Alstad et al. [2009]

$$H_{min,loss} = [\, 0.3205\ -0.2691\ -0.0001\ -0.0022 \,]. \quad (27)$$

The average loss and the maximum loss for the simulations without measurement noise are given in Table 2. We observe that the loss when using the data method is about twice as large as the loss when using the exact local method. However, the actual value of the cost function is around 0.8, so the relative loss is still quite small.

Although the method is not designed to handle noise, we have also tried the method on the case study with measurement noise included. The results for the same data points (but this time with measurement noise), are given in Table 3. Here the trend is similar to the noise-free case, but the absolute values are a little bit larger.

5. DISCUSSION AND CONCLUSION

We presented a method for finding a self-optimizing controlled variable combination. The approach does not require a model or disturbance measurements, and the controlled variable is an estimate of the gradient.

If disturbances were measurable, surface response methods [Box and Draper, 1987] could have be used for optimization. However, in practice the disturbances are generally not measurable online. An alternative approach, which estimates the gradient online, is extremum seeking [Krstic and Wang, 2000]. However, this approach requires excitation of the process, which often is undesirable.

As expected, the model based "minimum loss method" gave better results than our data-based method. However, when there is no model available this alternative does not exist, and our approach can be used to obtain at least some approximation of the optimal H. The loss values for the simulation case with noise and the case without noise

have been shown to be very similar, so we conclude that our method can handle measurement noise to some degree. Larger noise will affect our method in two ways: 1) it will deteriorate the estimation of the quadratic cost function (although this effect may partially be eliminated when using PLS), and 2) it will enter online when controlling $c = H\Delta y$ (this effect may be handled in analogy to Alstad et al. [2009]. However a detailed study of the effect of noise is outside the scope of this paper.

The combination matrix H can be used for feedback control, as we have done in our case study, but it may also be used for monitoring purposes. In this case, the process operators simply monitor magnitude of the elements in $\Delta c = H\Delta y$, and use it as an indication of optimality.

The approach in this paper is local, because it is based on linearization around the optimal point and the data model cannot be extrapolated outside the data range. However, since the data is from plant measurements, it can be verified how well the model matches the plant. If the disturbances become too large for the linearized model to be valid, it becomes necessary to use a different approach, most likely based on a first principles model. However, a model based approach will not be able to capture unmodelled disturbances either, unless is is somehow corrected by measurements.

Two topics have not been discussed in detail this paper: 1.) the question of how many components to use in the PLS model, and 2) how many data points are required to obtain a good model of the cost function. For topic 1), we have divided the available data in two sets, and used one data set as a validation set. The number of components has been chosen such that the norm of the difference between the predictions and the actual value of the cost function for the validation data set was minimized. Generally, the number of components required will depend on the process (the directionality of the measurements) and disturbance directions. Concerning topic 2), the number of data points required depends on the number of disturbances and the directions in the disturbances. In principle, we need the disturbances (which are the predictors in the PLS regression) to be sufficiently excited so that we can obtain an estimate of their effect on the cost function, and that the effect of noise is averaged out. We did not attempt to find the minimum number of data points (samples). This is left for future work. Other directions for future work include testing the approach on a larger case study, and systematically studying how measurement noise affects the method.

ACKNOWLEDGEMENTS

The authors thank the reviewers for helpful comments and acknowledge the fruitful discussions with Dr. Hertzberg.

REFERENCES

Vidar Alstad and Sigurd Skogestad. Null space method for selecting optimal measurement combinations as controlled variables. *Industrial & Engineering Chemistry Research*, 46:846–853, 2007.

Vidar Alstad, Sigurd Skogestad, and Eduardo Shigueo Hori. Optimal measurement combinations as controlled variables. *Journal of Process Control*, 19(1):138–148, 2009.

George E. P. Box. Evolutionary operation: A method for increasing industrial productivity. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 6(2): 81–101, 1957.

George E. P. Box and Norman R. Draper. *Empirical Model-building and Response Surfaces*. John Wiley, New York, 1987.

Constantin G. Economou and Manfred Morari. Internal model control. 5. extension to nonlinear systems. *Industrial & Engineering Chemistry Process Design and Development*, 25:403–411, 1986.

Kim H. Esbensen. *Multivariate Data Analsys – in practice, 5th Edition*. CAMO, 2004.

Sébastien Gros, Bala Srinivasan, and Dominique Bonvin. Optimizing control based on output feedback. *Computers & Chemical Engineering*, 33(1):191 – 198, 2009.

Johannes Jäschke and Sigurd Skogestad. NCO tracking and Self-optimizing control in the context of real-time optimization. *Journal of Process Control*, 21(10):1407 – 1416, 2011a.

Johannes Jäschke and Sigurd Skogestad. Controlled variables from optimal operation data. In M.C. Georgiadis E.N. Pistikopoulos and A.C. Kokossis, editors, *21st European Symposium on Computer Aided Process Engineering*, volume 29 of *Computer Aided Chemical Engineering*, pages 753 – 757. Elsevier, 2011b.

Vinay Kariwala. Optimal measurement combination for local self-optimizing control. *Industrial & Engineering Chemistry Research*, 46(11):3629–3634, 2007.

Miroslav Krstic and Hsin-Hsiung Wang. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica*, 36(4):595 – 601, 2000.

Bao Lin, Bodil Recke, Jørgen K.H. Knudsen, and Sten Bay Jørgensen. A systematic approach for soft sensor development. *Computers & Chemical Engineering*, 31(56): 419 – 425, 2007.

Thomas E. Marlin and A.N. Hrymak. Real-time operations optimization of continuous processes. In *Proceedings of CPC V, AIChE Symposium Series vol. 93.*, pages 156–164, 1997.

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

Sigurd Skogestad, Ramprasad Yelchuru, and Johannes Jäschke. *Optimal use of Measurements for Control, Optimization and Estimation using the Loss Method: Summary of Existing results and Some New. In "Selected Topics on Constrained and Nonlinear Control"*, chapter 2, pages 53–90. STU-NTNU, 2011.

Bala Srinivasan, Dominique Bonvin, Erik Visser, and Srinivas Palanki. Dynamic optimization of batch processes: II. role of measurements in handling uncertainty. *Computers & Chemical Engineering*, 27(1):27 – 44, 2003.

Francois Yacoub and John F. MacGregor. Product optimization and control in the latent variable space of nonlinear {PLS} models. *Chemometrics and Intelligent Laboratory Systems*, 70(1):63 – 74, 2004. ISSN 0169-7439.

Lingjian Ye, Yi Cao, Yingdao Li, and Zhihuan Song. Approximating necessary conditions of optimality as controlled variables. *Industrial & Engineering Chemistry Research*, 52(2):798–808, 2013.