Surrogate model generation using self-optimizing variables

Julian Straus, Sigurd Skogestad

Department of Chemical Engineering, Norwegian Univ. of Science & Technology (NTNU), NO-7491 Trondheim, Norway

Abstract

This paper presents the application of self-optimizing concepts for more efficient generation of steady-state surrogate models. Surrogate model generation generally has problems with a large number of independent variables resulting in a large sampling space. If the surrogate model is to be used for optimization, utilizing self-optimizing variables allows to map a close-to-optimal response surface, which reduces the model complexity. In particular, the mapped surface becomes much "flatter", allowing for a simpler representation, for example, a linear map or neglecting the dependency of certain variables completely. The proposed method is studied using an ammonia reactor which for some disturbances shows limit-cycle behaviour and/or reactor extinction. Using self-optimizing variables, it is possible to reduce the number of manipulated variables by three and map a response surface close to the optimal response surface. With the original variables, the response surface would include also regions in which the reactor is extinct.

Keywords: Self-optimizing control, Surrogate model, Sampling domain definition, B-splines, Optimization of integrated processes, Steady-state optimization

1. Introduction

This paper focuses on the generation of surrogate models for steady-state process optimization. Surrogate models, also known as reduced-order models or response surfaces, are an emerging field in many applications (Forrester et al., 2008). They can be seen as mapping of a nonlinear model and are in this respect similar to lookup tables. Since the publication of Sacks et al. (1989), the investigation of surrogate models increased. Their advantage is a simple mathematical description of a complicated system. The resulting simplified model can subsequently be utilized in *e.g.* optimization routines (Forrester and Keane, 2009) or in a multi-scale modelling framework like MoDeNa (Karolius et al., 2016).

Chemical engineering processes are frequently modelled using flowsheeting software. With sequential-modular simulation packages, like Aspen Plus®, Aspen Hysys®, SimSci PRO/II, or UniSim Design Suite, numerical problems may arise especially when we have large recycles. Furthermore, certain unit operations like reactors or columns may be computationally expensive to solve. Therefore, Caballero and Grossmann (2008) developed an "algorithm for the use of surrogate models in modular flowsheet optimization". In this approach, the surrogate model substitutes "noisy" and/or computational expensive models. By a "noisy" model we mean that the output from the model may vary because of numerical issues, for example, dependencies in the initial values. Their surrogate model was given by Kriging models (Krige, 1951) and can comprise several unit operations. This approach was also applied to a sour-water stripping plant (Quirante and Caballero, 2016).

As an alternative to Kriging models and with the aim of using the surrogate model in optimization routines, Cozad et al. (2014) developed the ALAMO methodology. In this methodology, the surrogate model is based on a selection of basis functions and the model quality is improved through error maximization sampling. The advantage of the ALAMO approach is the simplicity of the basis functions and the easy availability of derivative information of the surrogate model. However, it is necessary to fit a surrogate model after each sampling iteration of the algorithm.

Surrogate models can be useful in the optimization of integrated flowsheets (Straus and Skogestad, 2016). Sequential modular simulators often have problems with convergence due to recycles, whereas equation-oriented solvers simulators are difficult to initialize. Surrogate models may be introduced for individual units or combination of units and recycle streams may be incorporated, resulting in reduced computational cost for solving the overall flowsheet. The generation of surrogate models is more difficult, however, if the number of independent variables (n_u) is high. This so-called "curse of dimensionality" says that the number of sampling points n_p grows dramatically with the number of independent variables n_u (Forrester et al., 2008). This growth is especially pronounced for nonlinear models where the interaction between the independent variables affect the dependent variables. There are two ways to solve

☆J.S. gratefully acknowledges the financial support from YARA International ASA.

this problem; reduce the number of independent variables n_u or reduce the resulting interaction between the independent variables to obtain a simpler response surface.

One way to reduce n_u is given by a factorial sampling plan as developed by Morris (1991). This approach utilizes preliminary simulations, in which it is decided which of the independent variables have an effect on the dependent variables and whether this effect is linear. The insignificant variables may be omitted in the latter design of experiments, whereas linear variables require much less sampling points. However, problems may still arise if the model is very nonlinear or interactions exist in-between the independent variables.

Another alternative to reduce n_u is the application of PLS regression (Straus and Skogestad, 2017b). Through PLS regression on the sampled space, new latent variables $\mathbf{u}'$ with $n_{u'} < n_u$ are defined. The independent variables for the surrogate model fitting are subsequently the latent variables.

A better alternative, investigated in this paper, may be to use the concepts of self-optimizing control (SOC) (Skogestad, 2000) to identify new independent variables (Straus and Skogestad, 2016). Self-optimizing control is a philosophy from control theory. The aim of SOC is to select controlled variables, which, if kept constant when disturbances occur, give a small economic loss. It allows to focus on the region we are actually interested in, and hence, reduce the complexity of the response surface and reduce the sampling domain. Furthermore, it allows a reduction in n_u in certain special cases.

This paper is structured as follows; Section 2 summarizes the application of surrogate models in the context of optimization based on the work of Straus and Skogestad (2016). Section 3 discusses how self-optimizing variables can be applied in the generation of surrogate models. Section 4 first introduces the ammonia case study and then shows results from the application of self-optimizing control in surrogate model generation. Section 5 discusses the applicability of the proposed procedure and addresses limitations and problems of utilizing self-optimizing variables. Appendix A explains the concepts of self-optimizing control as used within this paper.

2. Optimization using local surrogate models

Consider a large-scale steady-state process to be optimized, given by

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & J(\mathbf{x},\mathbf{d},\mathbf{u}) \\ \text{s.t.} \quad & \mathbf{0} = \mathbf{g}(\mathbf{x},\mathbf{d},\mathbf{u}) \\ & \mathbf{0} \geq \mathbf{h}(\mathbf{x},\mathbf{d},\mathbf{u}) \end{aligned} \tag{1}$$

where J is a scalar cost function, usually an economic cost, $\mathbf{d} \in \mathbb{R}^{n_d}$ denotes the disturbances, for example feed variables and model parameters, $\mathbf{u} \in \mathbb{R}^{n_u}$ are the independent decision variables, and $\mathbf{x} \in \mathbb{R}^{n_x}$ are the internal model state variables. The equality constraints $\mathbf{g} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ are typically given by the equations in the flowsheeting software. Operational inequality constraints $\mathbf{h} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_h}$ can be imposed on the states $\mathbf{x}$ or inputs $\mathbf{u}$. As the optimization of a large-scale process is generally difficult, the process is split into several submodels given by $\mathbf{g}_i : \mathbb{R}^{n_{x_i}} \times \mathbb{R}^{n_{d_i}} \times \mathbb{R}^{n_{u_i}} \to \mathbb{R}^{n_{x_i}}$ and $\mathbf{h}_i : \mathbb{R}^{n_{x_i}} \times \mathbb{R}^{n_{d_i}} \times \mathbb{R}^{n_{u_i}} \to \mathbb{R}^{n_{h_i}}$. This is exemplified for the distinctive submodel i and the remaining submodels k in Fig. 1. Each submodel may have individual manipulated variables $\mathbf{u}_i \in \mathbf{u}$ and disturbances $\mathbf{d}_i \in \mathbf{d}$. It is possible that a disturbance or manipulated variable appears in several submodels. In addition, each submodel has inlet $\mathbf{z}_{k,i} \in \mathbb{R}^{n_{z_{k,i}}}$ and outlet connection variables $\mathbf{z}_{i,k} \in \mathbb{R}^{n_{z_{i,k}}}$. Note that the variables $\mathbf{z}_{i,k}$ are states or outputs for the submodel i they come from, $\mathbf{y}_{i,k} = \mathbf{z}_{i,k}$, whereas they are disturbances for the submodel k they enter. The connection variables and the disturbance variables can be combined into an augmented "disturbance" vector

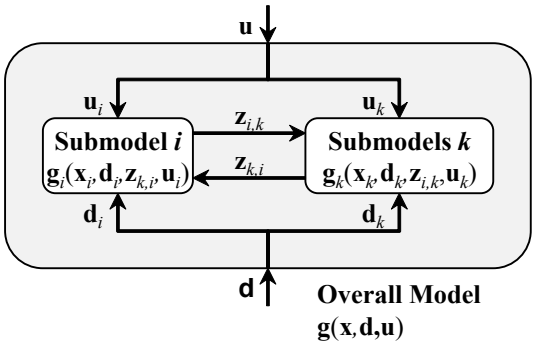

Fig. 1: Example of a submodel within an overall model.

$$\tilde{\mathbf{d}}_i = \begin{bmatrix} \mathbf{d}_i \\ \mathbf{z}_{k,i} \end{bmatrix} \tag{2}$$

Each submodel $\mathbf{g}_i$ may be reformulated as a surrogate model given by

$$\mathbf{g}'_{i,k} : \{\tilde{\mathbf{d}}_i, \mathbf{u}_i\} \mapsto \mathbf{y}_{i,k} \tag{3}$$

The total number of independent variables for each submodel is $n_i^{tot} = n_{u_i} + n_{\tilde{d}_i}$. Note that this is an input-output model with no explicit internal states. The reformulated optimization problem in terms of surrogate models then becomes

$$\begin{aligned} \min_{\tilde{\mathbf{d}},\mathbf{u}} \quad & J\left(\tilde{\mathbf{d}},\mathbf{u}\right) \\ \text{s.t.} \quad & \mathbf{0} = \mathbf{y}_{i,k} - \mathbf{g}'_{i,k}\left(\tilde{\mathbf{d}}_i,\mathbf{u}_i\right) && i \in 1,\ldots,n,\ \forall k \neq i \\ & \mathbf{0} = \mathbf{z}_{i,k} - \mathbf{y}_{i,k} && i \in 1,\ldots,n,\ \forall k \neq i \\ & \mathbf{0} \geq \mathbf{h}_i\left(\tilde{\mathbf{d}}_i,\mathbf{u}_i\right) && i \in 1,\ldots,n \end{aligned} \tag{4}$$

The sampling domain for surrogate model generation is given by bounds on the independent variables for each submodel

$$\tilde{\mathbf{d}}_{i,min} \leq \tilde{\mathbf{d}}_i \leq \tilde{\mathbf{d}}_{i,max} \tag{5}$$

$$\mathbf{u}_{i,min} \leq \mathbf{u}_i \leq \mathbf{u}_{i,max} \tag{6}$$

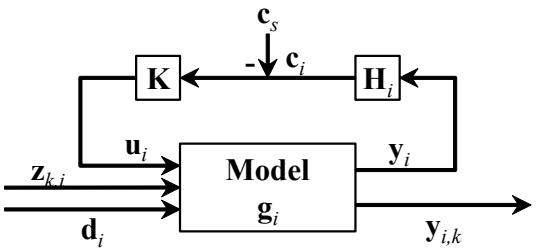

Fig. 2: Block diagram illustrating the change of independent variables.

The sampling may be performed using for example Latin hypercube sampling or regular grid sampling. Depending on the number of independent variables n_i^{tot} and the nonlinearity of the model, this can require a large sampling space. Hence, a reduction in the complexity of the surrogate model may be necessary.

3. Surrogate model generation using self-optimizing variables

Consider a detailed model representation

$$\mathbf{g}_i\left(\mathbf{x}_i, \tilde{\mathbf{d}}_i, \mathbf{u}_i\right) = \mathbf{0} \tag{7}$$

of submodel i and let $\mathbf{y}_{i,k} = \mathbf{f}_{i,k}\left(\mathbf{x}_i, \tilde{\mathbf{d}}_i, \mathbf{u}_i\right)$ represent the variables we are interested in knowing. To avoid solving the detailed model (7) every time, for example during optimization, we want to derive a surrogate model (3). To be able to introduce self-optimizing variables $\mathbf{c}_i$ to replace the original independent variables $\mathbf{u}_i$, we assume that we can define a a local cost function $J_i\left(\mathbf{x}_i, \tilde{\mathbf{d}}_i, \mathbf{u}_i\right)$. This cost function should reflect the overall cost J in (1) because we are not interested in arbitrary variations in $\mathbf{u}_i$, but in changes along the optimal surface. That is, we want to find instead a surrogate model

$$\mathbf{g}'_{i,k,SOC} : \{\tilde{\mathbf{d}}_i, \mathbf{c}_i\} \mapsto \mathbf{y}_{i,k} \tag{8}$$

where we remain close to the optimal surface for expected variations in $\tilde{\mathbf{d}}_i$ when the new variables $\mathbf{c}_i$ are kept constant (or strictly speaking, their setpoints $\mathbf{c}_{i,s}$).

This variable change is illustrated in Figure 2 using a block diagram. The controller $\mathbf{K}$ has integral action so that we have perfect control at steady state ($\mathbf{c}_i = \mathbf{c}_{i,s}$). Note that Figure 2 is just for illustrating how we can change the independent variables from $\mathbf{u}_i$ to $\mathbf{c}_i$, and there are no dynamics present in the surrogate model. The objective is that with this change in independent variables, the surrogate models become much simpler, for example linear, and in some cases we may even eliminate variables.

3.1. Motivating example

As a motivating illustration of the concept, consider the Rosenbrock function (Rosenbrock, 1960):

$$J_i\left(u_i\right) = \left(1 - d_i\right)^2 + 100\left(u_i - d_i^2\right)^2 \tag{9}$$

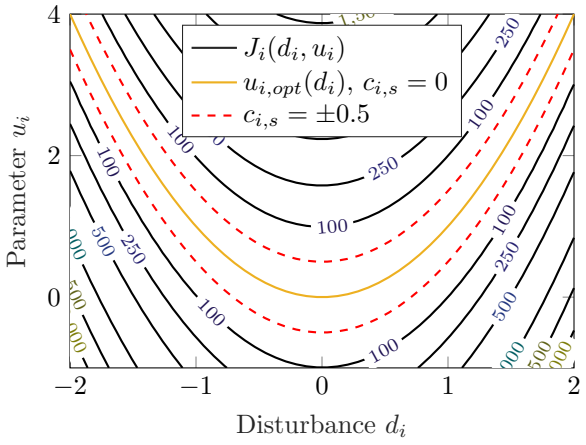

Fig. 3: Example for mapping the optimal response surface using the Rosenbrock function as case study.

The cost function is in this case also the outlet dependent variable y_i. The contour map for J_i as a function of u_i and d_i is shown in Figure 3. For a given disturbance value d_i, it would not make sense to map the whole region for u_i as it includes regions with a high value of the cost function. Instead, it is preferable to map only the region around the optimal input $u_{i,opt}\left(d_i\right)$ as given by the yellow line. Note, by introducing

$$c_i = u_i - d_i^2 \tag{10}$$

and setting $c_{i,s} = 0$, the cost function is minimized independently of the value of d_i as we indirectly get $u_i = u_{i,opt}\left(d_i\right) = d_i^2$. This allows to map along the optimal response surface as shown by the dashed lines. These bounds correspond to $c_i = \pm 0.5$. The close-to-optimal response surface has a simpler structure compared to the complete response surface. Compared to the optimal response surface approach, it is possible to vary the setpoint of the new variable, $c_{i,s}$ as well. The surrogate models according to Eqs. (3) and (8) are then given by

$$\mathbf{g}'_i : \{d_i, u_i\} \mapsto J_i \tag{11}$$

$$\mathbf{g}'_{i,SOC} : \{d_i, c_i\} \mapsto J_i \tag{12}$$

in which $\mathbf{g}'_i$ would include interaction terms between u_i and d_i and fourth-order terms in d_i. On the other hand, $\mathbf{g}'_{i,SOC}$ does not include interaction terms and at most second-order terms are needed in the model. Correspondingly, less points have to be sampled due to the simpler structure of the model. Note that this illustrative example does not correspond to the linear self-optimizing variables $\mathbf{c}_i$ as used in this paper. Unfortunately, it is in general difficult to obtain nonlinear self-optimizing variables. Hence, linear combinations of measurements are used as self-optimizing variables as outlined in the next section.

3.2. Prcoedure for selecting self-optimizing variables

Let the "measurements"

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{x}_i \\ \tilde{\mathbf{d}}_i \\ \mathbf{u}_i \end{bmatrix} \tag{13}$$

represent all the system variables and assume that we use self-optimizing control ideas to find a linear measurement combination

$$\mathbf{c}_i = \mathbf{H}_i \mathbf{y}_i \tag{14}$$

where $\mathbf{y}_i \in \mathbb{R}^{n_{y_i}}$ are the selected measurements to replace the independent variables $\mathbf{u}_i$ and $\mathbf{c}_i \in \mathbb{R}^{n_{u_i}}$ are the local self-optimizing variables

To this effect, we define a local cost function $J_i(\mathbf{x}_i, \mathbf{d}_i, \mathbf{u}_i)$ and consider the following local optimization problem

$$\begin{aligned} \min_{\mathbf{x}_i, \mathbf{u}_i} \quad & J_i(\mathbf{x}_i, \mathbf{d}_i, \mathbf{u}_i) \\ \text{s.t.} \quad & \mathbf{0} = \mathbf{g}_i(\mathbf{x}_i, \mathbf{d}_i, \mathbf{u}_i) \\ & \mathbf{0} \geq \mathbf{h}_i(\mathbf{x}_i, \mathbf{d}_i, \mathbf{u}_i) \end{aligned} \tag{15}$$

As shown in Appendix A, we need to define the expected disturbance set through the weight $\mathbf{W_d}$ and the expected "noise" (caused by numerical errors) in the measurements $\mathbf{y}_i$ through the weight $\mathbf{W_{n^y}}$. The self-optimizing variables are then obtained as the set $\mathbf{c}_i = \mathbf{H}_i \mathbf{y}_i$ which minimizes $\left| J_i\left(\mathbf{c}_i, \tilde{\mathbf{d}}_i\right) - J_{i,opt}\left(\tilde{\mathbf{d}}_i\right) \right|$. If we neglect "noise", then we may use the nullspace method (Alstad and Skogestad, 2007), but in this paper we include "noise" and use the exact local method, (Alstad et al., 2009) as described in more detail in Appendix A.

Note that we include the inlet connection variables $\mathbf{z}_{k,i}$ as disturbances in the calculation of $\mathbf{H}_i$, that is

$$\tilde{\mathbf{d}}_i = \begin{bmatrix} \mathbf{d}_i \\ \mathbf{z}_{k,i} \end{bmatrix} \tag{16}$$

and $n_{\tilde{d}_i} = n_{d_i} + n_{z_{k,i}}$. The calculation of the SOC selection matrix $\mathbf{H}_i$ according to optimization problem (A.9) requires the solution to 1 (or $n_{\tilde{d}_i} + 1$ if the optimal sensitivity matrix is calculated using finite differences) nonlinear problem(s) and n_{u_i} mixed integer quadratic problems. Furthermore, n_i^{tot} nonlinear systems of equations have to be solved to obtain the gain matrix $\mathbf{G^y}$ and the disturbance gain matrix $\mathbf{G_d^y}$. The sampling for the calculation of the surrogate model then consists of solving n_p nonlinear systems of equations.

Based on the sampling domain in (5), we suggest using as scaling matrices

$$\mathbf{W_d} = \operatorname{diag}\left(\max\left(\tilde{\mathbf{d}}_i - \tilde{\mathbf{d}}_{i,min}, \tilde{\mathbf{d}}_{i,max} - \tilde{\mathbf{d}}_i\right)\right) \tag{17}$$

as this results in minimizing the loss within the surrogate model domain. The "measurement noise" scaling matrix should be set to the expected numerical noise in $\mathbf{y}_i$. If this noise is small compared to the disturbance scaling matrix, we can set the measurement noise scaling matrix to

$$\mathbf{W_{n^y}} = w_{n^y} \operatorname{diag}(\mathbf{1}) \tag{18}$$

where w_{n^y} is small and $\mathbf{1}$ is a vector of ones with length n_y. However, two necessities arise for the parameter w_{n^y}

1. w_{n^y} is large enough so that $\mathbf{Y}\mathbf{Y}^\mathsf{T}$ in (A.8) is nonsingular;
2. w_{n^y} should be small compared to the entries of $\mathbf{W_d}$ to reduce the effect of measurement noise in the calculation of the selection matrix $\mathbf{H}$.

It is often preferable to use a block diagonal selection matrix $\mathbf{H}_i$. The advantage of a block diagonal matrix is to reduce the computational load of adjusting the setpoints iteratively in the flowsheeting software. This corresponds to Problem 3 described by Yelchuru and Skogestad (2012) and cannot be solved using the MIQP approach of (Yelchuru and Skogestad, 2012) as it violates the convex formulation theorem. However, it is possible to calculate a "local" selection matrix $\mathbf{H}_{i,l}$ for each input $u_{i,l}$ using only measurements in the vicinity of $u_{i,l}$, see Appendix A.2 for details. The resulting block diagonal matrix is due to neglecting interactions not optimal, but is sufficient for the subsequent application. A discussion of the scaling matrices and the use of a structured selection matrix $\mathbf{H}_i$ is provided in Section 5.2.

In summary, the procedure for utilizing self-optimizing variables in the context of surrogate model generation can be summarized as follows:

1. Set up a nonlinear problem (15) for submodel i and identify the connection variables $\mathbf{z}_{k,i}$ and $\mathbf{y}_{i,k}$.
2. Construct the augmented disturbance vector

   $$\tilde{\mathbf{d}}_i = \begin{bmatrix} \mathbf{d}_i \\ \mathbf{z}_{k,i} \end{bmatrix}$$

   and the measurement vector

   $$\mathbf{y}_i = \begin{bmatrix} \mathbf{x}_i \\ \tilde{\mathbf{d}}_i \\ \mathbf{u}_i \end{bmatrix}$$

   and define the sampling domain in (5) and (6).
3. Define the scaling matrices $\mathbf{W_d}$ and $\mathbf{W_{n^y}}$, for example using (17) and (18).
4. Solve the nonlinear problem (15) for the nominal input variables and calculate the sensitivity matrix $\mathbf{F}$ either using Eq. (A.5) or through finite differences. This requires the solution of 1 or $1+n_{\tilde{d}_i}$ optimization problems, depending on the availability of analytic expressions for $\mathbf{G^y}$, $\mathbf{G_d^y}$, $\mathbf{J_{uu}}$, and $\mathbf{J_{ud}}$.
5. Define the local measurements around the manipulated variables used for the calculation of the self-optimizing variables based on the total measurement $\mathbf{y}_i$.

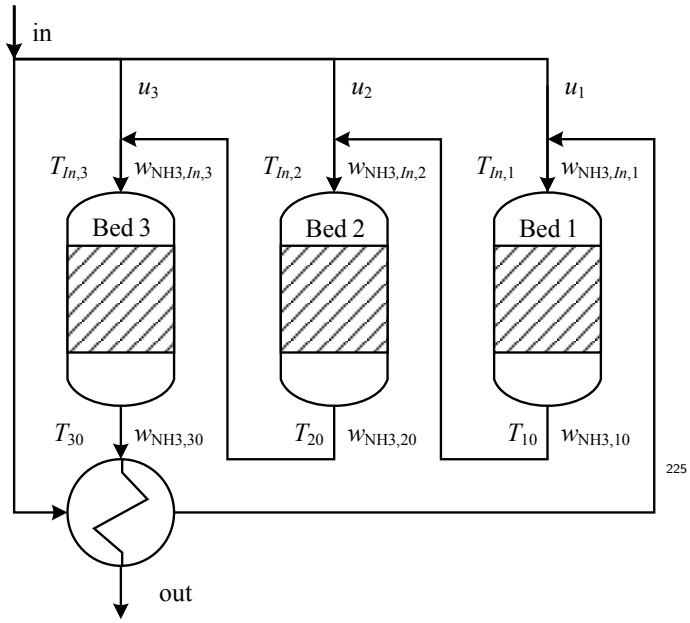

Fig. 4: Heat-integrated 3 bed reactor system of the ammonia synthesis gas loop.

6. Define the maximum number of measurements used for each manipulated variable u_l.
7. Calculate the optimal selection matrices $\mathbf{H}_{i,l}$ for the different manipulated variables $u_{i,l}$ using the MIQP (A.9).
8. Add the linear equality constraints given by Eq. (14) to the model $\mathbf{g}_i$ and sample n_p points.
9. Construct the surrogate models $\mathbf{g}'_{i,k,SOC}$ for the dependent variables $\mathbf{y}_{i,k}$ as given in Eq. (8).

4. Case study - ammonia reactor

The case study is a heat-integrated ammonia reactor, shown in Figure 4. This reactor was previously used for stability analysis (Morud and Skogestad, 1998) and application of economic nonlinear model predictive control (Straus and Skogestad, 2017a). A detailed model description can be found in Morud and Skogestad (1998). The subscript i is dropped in the following for the sake of simplicity. At the steady-state optimal operation point, small disturbances lead to limit-cycle behaviour and/or reactor extinction (Straus and Skogestad, 2017a). Varying the manipulated variables $\mathbf{u}$ individually results in creating a response surface that includes undesirable operating regions with reactor extinction and limit cycle behaviour. Hence, the response surface is complicated and it is necessary to sample a lot of points to achieve a good surrogate model.

4.1. Model description and modification

The aim of the reactor is to maximize the conversion *per pass*, which can be expressed in this example as the extent of reaction ξ [kg/s],

$$\xi = \dot{m}_{in}\left(w_{\mathrm{NH_3},30} - w_{\mathrm{NH_3},in}\right) \tag{19}$$

Table 1: Bounds and units for the connection variables.

	$\dot{m}_{in}$ [kg/s]	p_{in} [bar]	T_{in} [°C]	$w_{\mathrm{NH_3},in}$ [wt.%]	$R_{\mathrm{H_2/N_2},in}$ [-]
Lower Bound	59.5	185	235	7	2.8
Nominal Value	70.0	200	250	8	3.0
Upper Bound	80.5	215	265	9	3.2

where $\dot{m}$ [kg/s] is the mass flow and $w_{\mathrm{NH_3},i}$ the ammonia mass fraction. Correspondingly, the cost function for the optimization problem (15), which is posed as minimization problem, is given by

$$J = -\xi \tag{20}$$

The equality constraints are given by the ammonia mass balance and the energy balance described by Straus and Skogestad (2017a) for each CSTR j in the CSTR cascade used to represent each reactor bed. The number of CSTRs in each bed is $n = 10$.

In order to increase the applicability of the resulting surrogate model, the hydrogen to nitrogen molar ratio is not considered to be fixed as in (Straus and Skogestad, 2017a) and (Morud and Skogestad, 1998). Instead, the molar ratio of hydrogen to nitrogen,

$$R_{\mathrm{H_2/N_2},j} = \frac{\dot{n}_{\mathrm{H_2},j}}{\dot{n}_{\mathrm{N_2},j}} \tag{21}$$

in each reaction section j is introduced as an algebraic state. This results in 30 additional algebraic constraints

$$0 = R_{\mathrm{H_2/N_2},j} - \frac{\dot{n}_{\mathrm{H_2},j-1} + r_{\mathrm{H_2},j} m_{cat,j}/M_{\mathrm{H_2}}}{\dot{n}_{\mathrm{N_2},j-1} + r_{\mathrm{N_2},j} m_{cat,j}/M_{\mathrm{N_2}}} \tag{22}$$

in which M_i is the respective molar mass and $r_{i,j}$ the reaction rate in [kg i/kg$_{cat}$ h].

For this system, the independent variables are given by the three split ratios

$$\mathbf{u} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^\mathsf{T} \tag{23}$$

which may be viewed as the real manipulated variables. The five disturbances are the inlet conditions to the system

$$\tilde{\mathbf{d}} = \begin{bmatrix} \dot{m}_{in} & p_{in} & T_{in} & w_{\mathrm{NH_3},in} & R_{\mathrm{H_2/N_2},in} \end{bmatrix}^\mathsf{T} \tag{24}$$

These are the connection variables $\mathbf{z}_{k,i}$. If desired, it is as well possible to include parameter changes as disturbances, for example in the reaction rate constant, but this was not done in this case study. The bounds on the disturbance variables are given in Table 1. Two output variables have to be fitted in the surrogate model. These are the (mass) extent of reaction ξ and the outlet temperature T_{out}. The outlet ratio $R_{\mathrm{H_2/N_2},out}$ can be calculated through the respective outlet molar flows $\dot{n}_{i,out}$ which in turn are calculated from exact mass balances using ξ. This furthermore guarantees mass conservation in the resulting surrogate model. To summarize,

$$\mathbf{y}_{i,k} = \begin{bmatrix} \xi \\ T_{out} \end{bmatrix} \tag{25}$$

The system was modelled using CasADi (Andersson, 2013) and optimized using IPOPT (Wächter and Biegler, 2006).

4.2. Application of SOC

As $n_u = 3$, three SOC variables $\mathbf{c} = \mathbf{H}\mathbf{y}$ have to be obtained. We want to use local variables for each reactor bed to simplify the calculations when using a flowsheet simulator. Hence, the MIQP approach in (A.9) as proposed by Yelchuru and Skogestad (2012) is applied individually for each bed resulting in a block diagonal matrix given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & 0 & 0 \\ 0 & \mathbf{H}_2 & 0 \\ 0 & 0 & \mathbf{H}_3 \end{bmatrix} \tag{26}$$

In order to have a small number of measurements, we consider for each bed $n_y = 1$ and $n_y = 2$. This is compared to a more intuitive control structure, where the the inlet temperatures (*In*) as well as the inlet and outlet temperatures of the respective beds are used (*In,Out*). The scaling matrix $\mathbf{W_d}$ according to Eq. (17) and Table 1 is

$$\mathbf{W_d} = \mathrm{diag}\left(\begin{bmatrix} 10.5 & 15 & 15 & 1 & 0.2 \end{bmatrix}\right) \tag{27}$$

whereas the parameter w_{n^y} in the calculation of $\mathbf{W_{n^y}}$ is selected as $w_{n^y} = 10^{-3}$.

The candidate measurements for the MIQP approach are

$$\mathbf{y}_1 = \begin{bmatrix} T_{In,1} \\ \mathbf{T}_{1:10} \end{bmatrix} \qquad \text{for Bed 1, } u_1 \tag{28}$$

$$\mathbf{y}_2 = \begin{bmatrix} T_{In,2} \\ \mathbf{T}_{11:20} \end{bmatrix} \qquad \text{for Bed 2, } u_2 \tag{29}$$

$$\mathbf{y}_3 = \begin{bmatrix} T_{In,3} \\ \mathbf{T}_{21:30} \end{bmatrix} \qquad \text{for Bed 3, } u_3 \tag{30}$$

Therefore, each of the three selection matrices considers 11 measurements. The mass fraction measurements $\mathbf{w}_{\mathrm{NH_3}}$ would not be viable measurements for control purposes. However, in the case of surrogate model generation, they can still be used. We found that including the mass fractions in the measurements did not change the selected subset of measurements. Hence, the mass fractions are excluded in the candidate measurements. The optimization problem (A.9) was solved using $m = 100$ in the big-m approach of Eq. (A.11). The solution to the problem with one measurement ($n_y = 1$, $MIQP_1$) for each bed gives as chosen measurements:

$$\text{Bed 1: } T_9 \qquad \text{Bed 2: } T_{18} \qquad \text{Bed 3: } T_{25} \tag{31}$$

The solution to the optimization problem (A.9) with $n_y = 2$ is given in Table 2 ($MIQP_2$). Similar to the results reported by Yelchuru and Skogestad (2012), the chosen measurements change depending on the chosen number of measurements n_y. That means that a measurement which is optimal with only one measurement is not necessarily included with two measurements.

Table 2: Optimal selection matrix for a fixed selection (*In,Out*) as well as the optimal measurement subset for each input and the corresponding optimal selection matrix $\mathbf{H}_i$ with $n_y = 2$ ($MIQP_2$).

		Chosen Variables	Selection Matrix $\mathbf{H}_i$
In,Out	Bed 1	$T_{In,1}$, T_{10}	$\begin{bmatrix} 0.067 & -1.000 \end{bmatrix}$
	Bed 2	$T_{In,2}$, T_{20}	$\begin{bmatrix} 0.098 & 1.000 \end{bmatrix}$
	Bed 3	$T_{In,3}$, T_{30}	$\begin{bmatrix} 1.000 & 0.721 \end{bmatrix}$
$MIQP_2$	Bed 1	T_4, T_6	$\begin{bmatrix} 0.952 & -1.000 \end{bmatrix}$
	Bed 2	$T_{In,2}$, T_{11}	$\begin{bmatrix} 0.982 & -1.000 \end{bmatrix}$
	Bed 3	T_{28}, T_{30}	$\begin{bmatrix} 1.000 & -0.994 \end{bmatrix}$

Table 3: Estimation error ϵ with fixing the three SOC variables using different selection matrices $\mathbf{H}$.

$\mathbf{H}$ definition	Extent of Reaction ξ		Outlet Temperature T_{out}	
	$\max\lvert\epsilon\rvert$	$\overline{\lvert\epsilon\rvert}$	$\max\lvert\epsilon\rvert$	$\overline{\lvert\epsilon\rvert}$
In	4.353 %	0.742 %	8.00 K	1.457 K
In,Out	0.540 %	0.092 %	1.02 K	0.184 K
$MIQP_1$	0.211 %	0.027 %	0.41 K	0.055 K
$MIQP_2$	0.022 %	0.003 %	0.04 K	0.005 K

4.3. Fitting of the surrogate model

The surrogate models are cubic B-splines fitted through the application of the SPLINTER library (Grimstad et al., 2015), which requires a regular grid in the independent variables $\mathbf{c}$ (which here replace $\mathbf{u}$) and $\tilde{\mathbf{d}}$. In this case study, the overall cost function is minimized by optimizing locally the degrees of freedom ($\mathbf{c} = \mathbf{H}\mathbf{y}$), that is, the local cost J_i is equal to the global cost J, so it is not necessary to include the setpoints $\mathbf{c}_s$ as degrees of freedom as it normally would be. The regular grid is given by four points for each of the varied variable $\tilde{\mathbf{d}}$, $\dot{m}_{in}$, p_{in}, T_{in}, $w_{\mathrm{NH_3},in}$, and $R_{\mathrm{H_2/N_2},in}$. This results in $n_p = 4^5 = 1024$ sampling points. The advantage of using B-splines of order two or higher is that it gives continuity of the first derivative of the surrogate model. This gives advantages for the subsequent optimization. If self-optimizing variables are not used, we would need to consider all variables ($\tilde{\mathbf{d}}$ and $\mathbf{u}$) simultaneously giving $4^8 = 65536$ sampling points. Alternatively, other surrogate model structures like Kriging or the ALAMO approach (Cozad et al., 2014) could be used.

4.4. Evaluation of the surrogate model performance

The resulting surrogate models for the outlet temperature T_{out} and ξ were evaluated using 5000 randomly sampled validation points. These validation points are the optimal response surface for this model. This implies that the surrogate model may theoretically give perfect fit for the self-optimizing variables response surface. However, this is only of minor interest as the aim of the surrogate model is to utilize it in further optimization.

In order to compare the different methods, the maximum absolute error $\max\lvert\epsilon\rvert$ and the mean absolute error $\overline{\lvert\epsilon\rvert}$ are calculated with respect to the optimal response surface. The results of the four different combination matrices can be found in Table 3. We can see that arbitrarily chosen measurements (*In* and *In,Out*) do not necessarily

result in a good surrogate model fit. Using only the three inlet temperatures (In) results in a training space with infeasible points. For example, two of the split ratios are negative for

$$\tilde{\mathbf{d}} = \begin{bmatrix} 80.5 & 185 & 235 & 9 & 2.8 \end{bmatrix}^{\mathsf{T}} \tag{32}$$

Adding the outlet temperature of each bed to the selected measurements (In,Out) reduces the error by one order of magnitude. Furthermore, all points in the training space are feasible. Selecting only one optimal measurement in each bed ($MIQP_1$, see (31)) reduces the error by more than a factor two compared to using two "arbitrary" measurements (In,Out) in each bed. Finally, increasing the number of measurements in the MIQP approach from $n_y = 1$ ($MIQP_1$) to $n_y = 2$ ($MIQP_2$) in each bed gives a further decrease by one order of magnitude. This shows that it is important to select the best measurements. Otherwise, the resulting surface is more complicated and it may be even necessary to reduce the sampling space to avoid sampling infeasible points.

Importantly, the resulting response surface is simple. To show this, the surrogate models were validated using a response surface created through incorporation of constraints (14). The comparison to this validation space results in a maximum absolute error for the surrogate model using the inlet temperatures (In) of 0.37 % in ξ and 0.7 K in T_{out}. Compared to the optimal response surface, this error is one order of magnitude smaller. In comparison, with the MIQP approach with $n_y = 2$ ($MIQP_2$), the maximum absolute error is given by 0.001 % in ξ and 0.001 K in T_{out}. This indicates, that the resulting response surface is indeed simpler and it is possible to reduce the number of sampling points. Using instead quadratic B-splines with three points for each variable $\tilde{\mathbf{d}}$, $n_p = 5^3 = 243$ points have to be sampled. The maximum absolute error is then given by 0.01 % in ξ and 0.02 K in T_{out} which is still below the error of using the inlet temperatures with 4 points for each variable.

Sampling the space without variable transformation, *i.e.* selecting $\mathbf{c} = \mathbf{u}$, is not advisable for this case study. First, as already mentioned, it would require the sampling of much more sampling points. In addition, the resulting surface is more complicated. To illustrate this, consider the case when all disturbance variables are at their lower bound (Table 1)

$$\tilde{\mathbf{d}} = \begin{bmatrix} 59.5 & 185 & 235 & 7 & 2.8 \end{bmatrix}^{\mathsf{T}} \tag{33}$$

and the manipulated variables are fixed at their nominal optimum ($\mathbf{u} = \mathbf{u}_{opt}\left(\tilde{\mathbf{d}}_{nom}\right)$). In this situation, the reactor is extinct. Hence, using the split ratios $\mathbf{u}$ as independent variables would require the mapping of regions in which the reactor is extinct as well as crossing the limit-cycle region (Morud and Skogestad, 1998). This region is exemplified in Figure 5 where the inlet pressure is at its lower bound and the other disturbances at their nominal value.

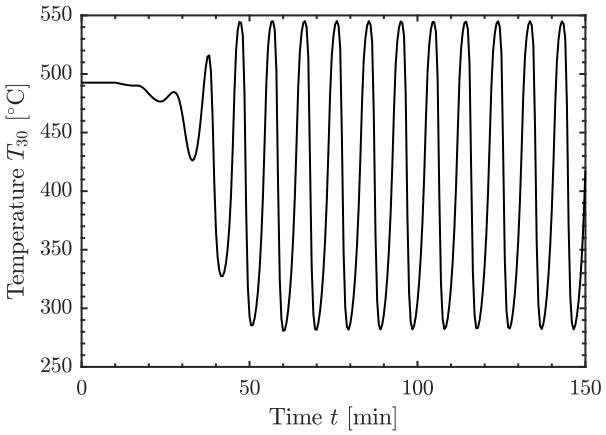

Fig. 5: Outlet temperature of Bed 3 with a pressure drop of $\Delta p_{in} = -15$ bar at $t = 10$ min with a constant input $\mathbf{u}$ at the optimal point.

We can see that the system displays limit-cycle behaviour and it is not possible to define a steady-state value for this operating point. However, these regions are not important for the subsequent optimization, and hence, should not be sampled. These regions are avoided through the application of self-optimizing variables.

5. Discussion

5.1. Advantages of the proposed method

The proposed utilization of self-optimizing variables to map the optimal response surface is a promising new method in the generation of surrogate models. The main advantages are given by

1. A response surface which is close to the optimal response surface but does not require the solution of a large number of nonlinear problems;
2. Potentially a reduced number of sampling points compared to sampling with the original independent variables.

This allows us to sample only regions we are interested in, and to neglect regions that are not encountered in practice. In the case study, it is in fact *not* possible to use the original inputs $\mathbf{u}_i$ (split ratios) as independent variables. Thus, a variable transformation would be required independently of the application of self-optimizing control. For example, one could use the variable transformation utilizing the existing control structure as proposed by Straus and Skogestad (2016). If it is necessary to have surrogate models for other states than the dependent variables $\mathbf{y}_{i,k}$, it is possible to calculate them as well, *e.g.* for the actual split ratios $\mathbf{u}_i$ or for additional potential measurements. In certain cases, as for our case study, it is as well feasible to reduce the number of independent variables.

An alternative to utilizing self-optimizing variables is to directly sample the optimal response surface given by

$$\mathbf{g}'_{i,opt} : \{\tilde{\mathbf{d}}_i\} \mapsto \mathbf{y}_{i,k} \tag{34}$$

This approach is however computationally expensive. It would require the solution to n_p nonlinear problems whereas in the application of the proposed method, only $n_{\tilde{d}}+1$ have to be solved in the calculation of the optimal sensitivity matrix $\mathbf{F}$. In addition, this approach does not allow for having the set points $\mathbf{c}_s$ as degree of freedom for solving the overall optimization problem as it is allowed with surrogate model (8).

Certain limitations of the proposed method can be identified and need to be addressed.

5.2. Practical use of the proposed method

A first important point is the selection of the disturbance and measurement scaling matrices, $\mathbf{W_d}$ and $\mathbf{W_{n^y}}$. These matrices will influence the performance of the resulting surrogate model. The 2-norm is used for scaling of the disturbances and measurement noises as we can see in Eq. (A.7). This implies that all disturbances and measurement noises may not be at their upper or lower limit simultaneously. In the case of control, this seems reasonable and a detailed discussion for using the 2-norm is given by Halvorsen et al. (2003). However, this is not the case, if we want to use the self-optimizing variables in the calculation of surrogate models. We actually want to sample these so-called corner points to avoid extrapolation. The best would be to use another norm, for example the 1-norm, but it can be partly circumvented by multiplying $\mathbf{W_d}$ by $\sqrt{n_{\tilde{d}}}$.

A second important point relates to the selection of the "measurements" $\mathbf{y}$ as they influence the loss when disturbances are present. In the case of surrogate model generation, the measurements do not need to be actual measurement as it is in the control application of SOC. Hence, it is possible to extend the measurements to states that are generally not considered as they are hard to measure, *e.g.* concentrations. As a result, the number of possible measurements $n_{y^{tot}}$ can be high. This requires the application of the MIQP approach as developed by Yelchuru and Skogestad (2012) and described in Appendix A.2. Unfortunately, the MIQP approach for measurement selection does not handle structural zeros in the selection matrix $\mathbf{H}$. The reason is that the the convex reformulation for obtaining the optimization problem (A.3) does not hold in this case as a pre-multiplication of $\mathbf{H}$ with a non-singular matrix $\mathbf{Q}$ will not preserve the structure of $\mathbf{H}$. Consequently, it is necessary to minimize the nonlinear loss expression

$$L = \frac{1}{2}\left\|\mathbf{J}_{\mathbf{uu}}^{1/2}\left(\mathbf{H}\mathbf{G}^{\mathbf{y}}\right)^{-1}\mathbf{H}\mathbf{Y}\right\|_F^2 \tag{35}$$

This optimization could be performed using a global nonlinear mixed integer optimization solver like BARON (Tawarmalani and Sahinidis, 2005) or ANTIGONE (Misener and Floudas, 2014). Unfortunately, there are no simple methods for solving this problem in a convincing way as highlighted by Jäschke et al. (2017). The development of an approach to include structural zeros is not the scope of this paper and will therefore not be discussed further. When we used alternatively a full selection matrix $\mathbf{H}$, we found that complicated adjustments in the flowsheet solver occurred. The proposed "local" approach and the resulting block diagonal selection matrix does not guarantee the optimal measurement combination in the combined measurement matrix and may lead to cases, where problems may arise. However, it is not possible to generalize when problems may occur and when not.

5.3. Number of independent variables

It is in general not possible to say, when the application of self-optimizing variables allows a reduction in the number of independent variables. There are however certain conditions, which have to be fulfilled as it is the case in the case study. One prerequisite is that the cost function J_i corresponds to the overall cost function J. In addition, it is necessary that the cost function is flat with respect to the self-optimizing variables as it was already stated by Skogestad (2000). The simpler response surface as aim of the introduction of self-optimizing variables is still likely to hold, independently of whether it is possible to reduce the number of independent variables. Especially if there are many disturbances (or connection variables), this may give a much simpler surrogate model which requires fewer sampling points to get a desired accuracy, as in the case study.

5.4. Application in flowsheeting software

The application of self-optimizing variables $\mathbf{c}_i$ in flowsheeting software can be difficult. It requires the use of additional equality constraints which can cause problems because it may require many iterations in sequential-modular simulators. As a result, the computational expense is increased. Hence, we proposed to use a structured selection matrix $\mathbf{H}_i$ with measurements in the vicinity of the respective manipulated variables. The variables $\mathbf{c}_i$ are then more decoupled and it is not necessary to converge the complete flowsheet for each adjustment. This problem is less pronounced in equation-oriented simulators. There, the application of surrogate model-based optimization results in smaller models, and hence, a simpler initialization of the models. The application of self-optimizing variables will then only increase the number of equality constraints.

5.5. Local cost function

The application of self-optimizing variables requires the definition of a local cost J_i corresponding to the overall cost J. This is possible for the investigated case study as it is general advantageous to maximize the conversion per pass of a chemical reactor. If this is not the case, the self-optimizing variables must be obtained using the overall cost and optimization problem (1). In the investigated case study, the introduction of an additional heat-exchanger with external cooling duty would for example complicate the cost function as there is no direct cost

linked to the conversion per pass. This brings us back to the starting point of the application of surrogate model generation and is similar to the "hen and egg" problem. If it is not possible to optimize the overall flowsheet, how can we then calculate the self-optimizing variables for each submodel $\mathbf{g}_i$?

One approach to achieve this is the utilization of a simplified overall model for the calculation of the self-optimizing variables. These can then be used in the generation of the surrogate models for the submodels. In addition, this would require the incorporation of the setpoints $\mathbf{c}_s$ of the self-optimizing variables into the sampling domain as they do not correspond to the actual optimal values due to the simplified overall model. The advantage here is the simplified response surface. Another approach, which is is probably better, is to define a reasonable local cost function, for example, based on physical arguments.

5.6. Manipulated variables and disturbances affecting several submodels

As mentioned in Section 2, it possible that disturbances and manipulated variables affect several submodels. This can include, for example, a disturbance in the cooling water temperature or if several compressors are connected to the same turbine shaft. This leads to the question if we can apply the procedure in these cases as well.

On one hand, it is not a problem with a disturbance affecting several submodels. The reason is that the disturbances are included as independent variables in the surrogate model generation (see Eqs. (8) and (2)). Therefore, the connection of the submodels for optimization will provide to all surrogate models the same disturbance value.

On the other hand, manipulated variables $\mathbf{u}_i^*$, which affect several submodels, have to be handled more carefully. Due to the variable transformation, it is not possible to calculate self-optimizing variables for these manipulated variables in each submodel. Instead, the manipulated variables must be assigned to one submodel. In addition, it is required in this submodel to fit a surrogate model

$$\mathbf{g}'_{u_i} : \{\tilde{\mathbf{d}}_i, \mathbf{c}_i\} \mapsto \mathbf{u}_i^* \tag{36}$$

to calculate the value of $\mathbf{u}_i^*$. This value is then used as connection variable in the other submodels that contain $\mathbf{u}_i^*$ as independent variable.

It has to be noted that it is more common in chemical process that disturbances affect several submodels. Furthermore, the developer of the surrogate models can decide to include all unit operations with the same manipulated variables $\mathbf{u}_i^*$ into one submodel. Then, it is not necessary to fit a surrogate model (36) to the manipulated variables $\mathbf{u}_i^*$.

5.7. Alternative approaches for independent variable reduction

Self-optimizing control is a "direct" approach that applies optimization explicitly in the calculation of the optimal sensitivity matrix $\mathbf{F}$. There exist several other approaches for optimization in the context of real-time optimization with parameter uncertainty. One "indirect" approach is (adaptive) directional modifier adaptation (Costello et al., 2016; Singhal et al., 2017). This approach utilizes the conditions of optimality expressed by the derivative of the Langrangian with respect to the inputs and parameters instead of using optimization explicitly for variable reduction. Through the application of singular-value decomposition to the derivative of the Lagrangian, it is possible to identify input directions in which the cost function is sensitive with respect to the uncertain parameters. In the case of surrogate model generation, the uncertain parameters would correspond to the augmented disturbances.

Using directional derivatives is advantageous in comparison to the proposed procedure using self-optimizing variables if there are many input connection streams, and hence, the number of independent variables is very large. In this situation, it may not be possible to calculate the optimal sensitivity matrix $\mathbf{F}$ explicitly and a reduction in independent variables may be necessary.

6. Conclusion

Combining principles from control theory and surrogate modelling, a new method was developed to simplify the structure of surrogate models. The main idea is to replace the original independent variables $\mathbf{u}_i$ by a better set $\mathbf{c}_i$ using the approach of self-optimizing control, see Section 3.2 for details. This approach allows for omitting regions in which the submodel is suboptimal, for example because a reactor is extinct. In addition, it may in some cases result in fewer independent variables.

Appendix A. Previous results on self-optimizing control

Consider the following optimization problem

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \quad & J(\mathbf{x}, \mathbf{d}, \mathbf{u}) \\ \text{s.t.} \quad & \mathbf{0} = \mathbf{g}(\mathbf{x}, \mathbf{d}, \mathbf{u}) \\ & \mathbf{0} \geq \mathbf{h}(\mathbf{x}, \mathbf{d}, \mathbf{u}) \end{aligned} \tag{A.1}$$

For example, this could be the local optimization problem in (15), but with subscript i omitted. The aim of self-optimizing control is to identify controlled variables $\mathbf{c}$ which, when kept constant, result in a minimum loss in the presence of disturbances ($\mathbf{d}$). Frequently, linear combinations of measurements $\mathbf{y}$ are used

$$\mathbf{c} = \mathbf{H}\mathbf{y} \tag{A.2}$$

where $\mathbf{H} \in \mathbb{R}^{n_u \times n_y}$ is a combination matrix. The question is: how can we identify the optimal selection matrix and correspondingly the self-optimizing variables? A detailed review answering this question can be found in Jäschke et al. (2017).

Appendix A.1. Summary of self-optimizing control approaches for obtaining $\mathbf{H}$

The optimal selection matrix $\mathbf{H}$ as introduced in Eq. (A.2) that minimizes the expected value of $|J(\mathbf{c},\mathbf{d}) - J_{opt}(\mathbf{d})|$ can be calculated using the nullspace method (Alstad and Skogestad, 2007) or the exact-local method (Alstad et al., 2009). The latter is given by the solution to the following optimization problem

$$\begin{aligned} \min_{\mathbf{H}} \quad & \left\|\mathbf{HY}\right\|_F \\ \text{s.t.} \quad & \mathbf{HG^y} = \mathbf{J_{uu}^{1/2}} \end{aligned} \tag{A.3}$$

with $\mathbf{G^y} \in \mathbb{R}^{n_y \times n_u}$ representing the measurement gain matrix with respect to the input $\mathbf{u}$. $\mathbf{Y}$ is given by

$$\mathbf{Y} = \begin{bmatrix} \mathbf{FW_d} & \mathbf{W_{n^y}} \end{bmatrix} \tag{A.4}$$

The optimal sensitivity matrix $\mathbf{F} = \frac{\partial \mathbf{y}^{opt}}{\partial \mathbf{d}}$ can be calculated as

$$\mathbf{F} = -\left(\mathbf{G^y J_{uu}^{-1} J_{ud}} - \mathbf{G_d^y}\right) \tag{A.5}$$

where $\mathbf{J_{uu}} \in \mathbb{R}^{n_u \times n_u}$ is the Hessian of the cost function, and $\mathbf{J_{ud}} \in \mathbb{R}^{n_u \times n_d}$ which is the second order derivative of J with respect $\mathbf{u}$ and $\mathbf{d}$. Alternatively, if it is not possible to easily obtain the analytic matrices of the cost function $\mathbf{J_{uu}}$ and $\mathbf{J_{ud}}$, the optimal sensitivity matrix $\mathbf{F}$ can also be calculated using finite differences. This results in n_d additional optimization problems. $\mathbf{W_d}$ and $\mathbf{W_{n^y}}$ are the disturbance and measurement noise scaling matrices given by

$$\Delta\mathbf{d} = \mathbf{W_d}\mathbf{d}'; \qquad \mathbf{n^y} = \mathbf{W_{n^y}}\mathbf{n^{y'}} \tag{A.6}$$

The vectors $\mathbf{d}'$ and $\mathbf{n^{y'}}$ are assumed to satisfy

$$\left\| \begin{bmatrix} \mathbf{d}' \\ \mathbf{n^{y'}} \end{bmatrix} \right\|_2 \leq 1 \tag{A.7}$$

Thus, $\mathbf{W_d}$ and $\mathbf{W_{n^y}}$ represent the magnitude of the expected variations in $\mathbf{d}$ and $\mathbf{y}$. The solution to problem (A.3) when $\mathbf{H}$ is a full matrix is (Yelchuru and Skogestad, 2012):

$$\mathbf{H}^\mathsf{T} = \left(\mathbf{YY}^\mathsf{T}\right)^{-1}\mathbf{G}^y \tag{A.8}$$

Appendix A.2. Measurement selection for $\mathbf{H}$

It is in general desirable to use few measurements $\mathbf{y}$. In order to select the an optimal subset of measurements n_y, Yelchuru and Skogestad (2012) developed a mixed integer quadratic programming approach. It requires the reformulation of the problem given in Eq. (A.3) in vectorized form:

$$\begin{aligned} \min_{\mathbf{h}_\delta \boldsymbol{\sigma}_\delta} \quad & \mathbf{h}_\delta^\mathsf{T}\mathbf{F}_\delta\mathbf{h}_\delta \\ \text{s.t.} \quad & \mathbf{G}_\delta^{y\mathsf{T}}\mathbf{h}_\delta = \mathbf{j}_\delta \\ & \sum_{k=1}^{n_{y^{tot}}} \sigma_k = n_y \end{aligned} \tag{A.9}$$

where $\sigma_k \in \{0,1\}$ with $k = 1 \ldots n_{y^{tot}}$ are binary variables to indicate, whether measurements are used in the selection matrix. The quadratic cost term is given by

$$\mathbf{F}_\delta = \mathbf{Y}_\delta\mathbf{Y}_\delta^\mathsf{T} \tag{A.10}$$

and is block diagonal. The same holds true for $\mathbf{G}_\delta^{y\mathsf{T}}$ whereas $\mathbf{h}_\delta$ and $\mathbf{j}_\delta$ are a vectorized form of $\mathbf{H}$ and $\mathbf{J_{uu}}$ respectively. Further constraints have to be imposed on $\mathbf{h}_\delta$ to guarantee that $h_{jk} = 0$ for $\sigma_k = 0$ and input u_j and measurement y_k. In this problem, the big-m approach is chosen. This results in bounds for the entries in the selection matrix $\mathbf{H}$ given by

$$-\begin{bmatrix} m \\ m \\ \vdots \\ m \end{bmatrix}\sigma_k \leq \begin{bmatrix} h_{1k} \\ h_{2k} \\ \vdots \\ h_{n_u k} \end{bmatrix} \leq \begin{bmatrix} m \\ m \\ \vdots \\ m \end{bmatrix}\sigma_k, \qquad \forall k \in 1,2,\ldots,n_{y^{tot}} \tag{A.11}$$

For a detailed description and derivation of the MIQP approach for measurement selection, the reader is referred to Yelchuru and Skogestad (2012).

References

Alstad, V., Skogestad, S., 2007. Null space method for selecting optimal measurement combinations as controlled variables. Industrial & engineering chemistry research 46 (3), 846–853.

Alstad, V., Skogestad, S., Hori, E. S., 2009. Optimal measurement combinations as controlled variables. Journal of Process Control 19 (1), 138 – 148.

Andersson, J., October 2013. A General-Purpose Software Framework for Dynamic Optimization. PhD thesis, Arenberg Doctoral School, KU Leuven, Department of Electrical Engineering (ESAT/SCD) and Optimization in Engineering Center, Kasteelpark Arenberg 10, 3001-Heverlee, Belgium.

Caballero, J. A., Grossmann, I. E., 2008. An algorithm for the use of surrogate models in modular flowsheet optimization. AIChE Journal 54 (10), 2633–2650.

Costello, S., Franois, G., Bonvin, D., 2016. A directional modifier-adaptation algorithm for real-time optimization. Journal of Process Control 39, 64 – 76.

Cozad, A., Sahinidis, N. V., Miller, D. C., 2014. Learning surrogate models for simulation-based optimization. AIChE Journal 60 (6), 2211–2227.

Forrester, A., Sobester, A., Keane, A., 2008. Engineering Design via Surrogate Modelling: A Practical Guide. Wiley.

Forrester, A. I., Keane, A. J., 2009. Recent advances in surrogate-based optimization. Progress in Aerospace Sciences 45 (1), 50 – 79.

Grimstad, B., et al., 2015. SPLINTER: a library for multivariate function approximation with splines. `http://github.com/bgrimstad/splinter`, accessed: 2017-11-26.

Halvorsen, I. J., Skogestad, S., Morud, J. C., Alstad, V., 2003. Optimal selection of controlled variables. Industrial & Engineering Chemistry Research 42 (14), 3273–3284.

Jäschke, J., Cao, Y., Kariwala, V., 2017. Self-optimizing control A survey. Annual Reviews in Control 43 (Supplement C), 199 – 223.

Karolius, S., Preisig, H. A., Rusche, H., 2016. Multi-scale modelling software framework facilitating simulation of interconnected scales using surrogate-models. In: Kravanja, Z., Bogataj, M. (Eds.), 26th European Symposium on Computer Aided Process Engineering. Vol. 38 of Computer Aided Chemical Engineering. Elsevier, pp. 463 – 468.

Krige, D. G., 1951. A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis, University of Witwatersrand, South Africa.

Misener, R., Floudas, C. A., Jul 2014. ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations. Journal of Global Optimization 59 (2), 503–526.

Morris, M. D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.

Morud, J. C., Skogestad, S., 1998. Analysis of instability in an industrial ammonia reactor. AIChE Journal 44 (4), 888–895.

Quirante, N., Caballero, J. A., 2016. Large scale optimization of a sour water stripping plant using surrogate models. Computers & Chemical Engineering 92 (Supplement C), 143 – 162.

Rosenbrock, H. H., 1960. An automatic method for finding the greatest or least value of a function. The Computer Journal 3 (3), 175–184.

Sacks, J., Welch, W. J., Mitchell, T. J., Wynn, H. P., 11 1989. Design and analysis of computer experiments. Statist. Sci. 4 (4), 409–423.

Singhal, M., Marchetti, A. G., Faulwasser, T., Bonvin, D., 2017. Improved directional derivatives for modifier-adaptation schemes. IFAC-PapersOnLine 50 (1), 5718 – 5723, 20th IFAC World Congress.

Skogestad, S., 2000. Plantwide control: the search for the self-optimizing control structure. Journal of Process Control 10 (5), 487 – 507.

Straus, J., Skogestad, S., 2016. Minimizing the complexity of surrogate models for optimization. In: Kravanja, Z., Bogataj, M. (Eds.), 26th European Symposium on Computer Aided Process Engineering. Vol. 38 of Computer Aided Chemical Engineering. Elsevier, pp. 289 – 294.

Straus, J., Skogestad, S., June 2017a. Economic NMPC for heat-integrated chemical reactors. In: 2017 21st International Conference on Process Control (PC). pp. 309–314.

Straus, J., Skogestad, S., Jan 2017b. Variable reduction for surrogate modelling. In: Proceedings of Foundations of Computer-Aided Process Operations 2017, Tucson, AZ, USA, 8-12 Jan. 2017.

Tawarmalani, M., Sahinidis, N. V., 2005. A polyhedral branch-and-cut approach to global optimization. Mathematical Programming 103, 225–249.

Wächter, A., Biegler, L. T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming 106 (1), 25–57.

Yelchuru, R., Skogestad, S., 2012. Convex formulations for optimal selection of controlled variables and measurements using mixed integer quadratic programming. Journal of Process Control 22 (6), 995 – 1007.