Annual Reviews in Control, Vol. 43, 2017, pp. 199-223
DOI: 10.1016/j.arcontrol.2017.03.001.

Self-Optimizing Control – A Survey

Johannes Jäschke[a,*], Yi Cao[b], Vinay Kariwala[c]

[a]*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*
[b]*School of Water, Energy and Environment, Cranfield University, Cranfield, Bedford MK 43 0AL, United Kingdom*
[c]*ABB Global Industries & Services Limited, Bangalore, India*

Abstract

Self-optimizing control is a strategy for selecting controlled variables. It is distinguished by the fact that an economic objective function is adopted as a selection criterion. The aim is to systematically select the controlled variables such that by controlling them at constant setpoints, the impact of uncertain and varying disturbances on the economic optimality is minimized. If such a selection leads to an acceptable economic loss compared to perfectly optimal operation then the chosen control structure is referred to as "self-optimizing". In this comprehensive survey on methods for finding self-optimizing controlled variables, we summarize the progress made during the last one and a half decades. In particular, we present brute-force methods, local methods based on linearization, data and regression based methods, and methods for finding nonlinear controlled variables for polynomial systems. We also discuss important related topics such as handling active constraints. Finally, we point out open problems and directions for future research.

Keywords: Self-optimizing control, Control structure selection, Controlled variables, Plant-wide control

Contents

1 Introduction **2**
1.1 What is self-optimizing control? . . . 2
1.2 The purpose of this review . . . 3
1.3 Defining optimal operation . . . 3
1.4 Implementation of optimal operation . . . 4

2 Historical notes **6**

3 Brute force methods for self-optimizing CV selection **8**

4 Local methods for steady-state self-optimizing CVs **9**
4.1 Local approximation of the cost function . . . 10
4.2 Local approximation of the plant and linear measurement combination . . . 11
4.3 Loss for a given control structure (Exact Local Method) . . . 12
4.4 Finding optimal measurement combinations . . . 15

*Corresponding author
Email addresses: jaschke@ntnu.no (Johannes Jäschke), y.cao@cranfield.ac.uk (Yi Cao), vinay.kariwala@in.abb.com (Vinay Kariwala)

Published by Elsevier. This is the Author Accepted Manuscript issued with:
Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version is available online at 10.1016/j.arcontrol.2017.03.001.
Please refer to any applicable publisher terms of use.

4.4.1 Null-space method . . . 16
4.4.2 Extended Null-space method . . . 17
4.4.3 Minimum Loss method (Explicit solution) . . . 18
4.5 Incorporating measured disturbances . . . 19
4.6 Selecting subsets of measurements . . . 20
4.6.1 Tailor-made branch and bound methods . . . 20
4.6.2 MIQP formulation . . . 22
4.7 Structured H . . . 24
4.8 Conclusion . . . 26

5 Constraint handling **26**
5.1 Parametric programming approach . . . 26
5.2 Integrated approach . . . 28
5.3 Cascade control approach . . . 30
5.4 Constraint matching . . . 30
5.5 Conclusion . . . 31

6 Self-optimizing CVs for larger operating windows **31**
6.1 Polynomial zero loss-method . . . 31
6.2 Regression approach . . . 33
6.3 Controlled variable adaptation . . . 34
6.4 Global approximation of controlled variables . . . 34
6.5 Conclusion . . . 35

7 Model-free approaches for self-optimizing control **35**
7.1 Historical data approach (using non-optimal data) . . . 36
7.2 Optimal data approach . . . 37
7.3 Conclusions . . . 38

8 Towards self-optimizing control for dynamic problems **38**

9 Applications of self-optimizing control **39**
9.1 Overview of self-optimizing control applications reported in literature . . . 39
9.2 Related applications . . . 41

10 Discussion on the relationship of self-optimizing control to other approaches **41**

11 Current status and future work **42**
11.1 Summary of the status of self-optimizing control . . . 42
11.2 Future work – open issues . . . 43

Appendix A Approximated loss expression & Minimum Singular Value Rule (Maximum gain rule) **45**

1. Introduction

1.1. What is self-optimizing control?

The purpose of a process plant is to generate profit. Beside plant design choices like size and type of equipment, plant operation has a major influence on the overall economic performance. The profitability of plant operation is strongly influenced by the design of the control structure. In the control structure design phase engineers make fundamental decisions about which variables to manipulate, to measure and to control (Skogestad, 2004a). Especially when the operating conditions vary, a judicious selection of controlled variables (CVs) can lead to large operational savings and increased competitiveness. In the context of control structure design, Skogestad (2000) was the first to formulate the concept of a self-optimizing control structure. It is is characterized by the choice of self-optimizing CVs:

> A set of controlled variables is called *self-optimizing* if, when it is kept at constant setpoints, the process is operated with an acceptable loss with respect to the chosen objective function (also when disturbances occur).

It is important to note that "self-optimizing" is not a property of the controller itself, as it is in e.g. adaptive control (Åström and Wittenmark, 2008). Rather, the term *self-optimizing control* has been used for describing a strategy for desinging the control structure, where the aim is to achieve close to optimal operation by (constant) setpoint control (Skogestad, 2000, 2004a; Alstad et al., 2009). In this paper will also use the term self-optimizing control in this sense.

The successful application of self-optimizing control requires tools and methods for selecting good CVs, and this is the topic of this review paper. The main difference between self-optimizing control and other methods for designing control structures, that typically consider controllability and control performance as a selection criterion, e.g. van de Wal and de Jager (2001), is that in self-optimizing control the selection is done to systematically minimize the loss of optimality with respect to a given economic cost function. Typically, this cost function is directly linked to the economic cost of plant operation, but also other objectives, such as energy efficiency, or also indirect control type objectives are possible (Skogestad and Postlethwaite, 2005). Thus, the selection procedure is driven by a clearly defined cost function, which is minimized during plant operation by simply controlling the self-optimizing CVs at their setpoints.

Unlike in real-time optimization approaches (Marlin and Hrymak, 1997; Grötschel et al., 2001), where a cost function is repeatedly optimized online to update the setpoints of the CVs, in self-optimizing control a model is used off-line to study the structure of the optimal solution. This insight is then translated into the design of a simple control structure, that keeps the process close to the optimum despite varying disturbances. Of course, this may lead to a loss due to simplification, but in many cases the benefits of a simple scheme outweigh the increased "optimality" of complex schemes, because of the high costs of implementing and maintaining these.

Self-optimizing CVs have been used in industry for a long time. Well-known examples, where self-optimizing control is inherently realized, include ratio control with a constant ratio setpoint, or controlling constrained variables at their constrained values, e.g. keeping a pressure variable at the maximal allowable value. The aim of the research field of self-optimizing control is, however, to provide a mathematical framework and systematic methods for finding CVs that give good economic performance.

1.2. The purpose of this review

After more than one and a half decades with ongoing research on self-optimizing control methods, we feel that it is time to summarize the main results and give a self-contained overview of the state-of-the-art and open issues in the development of methods for finding self-optimizing CVs. In large part, this survey paper is written as a tutorial, where the basic concepts are presented with examples. We hope that both experienced researchers and newcomers to the field will find it a useful resource that stimulates further applications and research.

1.3. Defining optimal operation

The goal for designing a control structure is nicely captured in the statement by Morari et al. (1980), who mentioned that "our main objective is to translate the economic objective into process control objectives". Thus, process control is not an end in itself, but is always in the context

of achieving best performance in terms of economics for a given set of operating conditions and constraints. Mathematically, this can be stated as an optimization problem.

Most continuous processes are operated at a steady-state (or close to it) for most of the time, which means that the disturbances stay constant long enough to make the economic effect of the transients negligible[1]. Therefore, we formulate the problem of optimal operation as a steady state optimization problem:

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

Here $x \in \mathbb{R}^{n_x}$ denotes the state variables, $d \in \mathbb{R}^{n}$ denotes the disturbances, and $\bar{u} \in \mathbb{R}^{n_{\bar{u}}}$ the steady state degrees of freedom[2] that affect the steady state operational cost $\bar{J} : \mathbb{R}^{n_{\bar{u}}} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \mapsto \mathbb{R}$. Further, the function $f : \mathbb{R}^{n_{\bar{u}}} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \mapsto \mathbb{R}^{n_f}$ denotes the model equations, and $g : \mathbb{R}^{n_{\bar{u}}} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \mapsto \mathbb{R}^{n_g}$ the operational constraints. We denote the optimal objective value of Problem (1) as $\bar{J}^*(d)$. In this paper we assume that the optimization problems are sufficiently smooth, and have a unique (local) minimum. This assumption generally excludes problems with logic and integer decision variables, such as the schedule for shutting a pump on and off at given times.

Under operation, the cost $\bar{J}(\bar{u}, x, d)$ should be minimized while satisfying the plant constraints. If all the states x, disturbances d, were perfectly known, one could attempt to solve Problem (1) and apply the optimal inputs $\bar{u}^*$ to the plant. Under ideal conditions this would result in optimal operation with the associated cost $\bar{J}^*(d)$. However, in practice such a strategy is not implementable because the plant is never truly at steady state, and because perfect knowledge of the model states and disturbances is not available. Instead, the knowledge about the plant conditions is typically available from measurements, and we assume to have a model for the plant measurements

$$y_0 = m(\bar{u}, x, d), \tag{2}$$

where the function $m : \mathbb{R}^{n_{\bar{u}}} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \to \mathbb{R}^{n_y}$ describes the relationship between the variables $\bar{u}, x, d$ and the model outputs $y_0 \in \mathbb{R}^{n_y}$. However, the signals y that are measured in the real plant are corrupted by measurement noise $n^y \in \mathbb{R}^{n_y}$, such that

$$y = y_0 + n^y. \tag{3}$$

1.4. Implementation of optimal operation

Using the measurements y, the task of the control structure and the controllers is to implement the optimal solution of Problem (1) into the real plant. A good control structure will ensure that the plant runs close to the economically optimal point, also when the operating conditions and disturbances change.

The control system of a chemical plant is typically decomposed and organized in a hierarchical manner, where different control layers operate on different time-scales (Skogestad, 2000). An example for such a hierarchical control structure is given in Figure 1. On top of the hierarchy

[1] In the case where transient behavior significantly contributes to the operating cost, optimal operation is formulated as a dynamic optimization problem, see Section 8.

[2] For example, degrees of freedom that do not affect the steady-state are the levels in the condenser and reboiler of a distillation column.

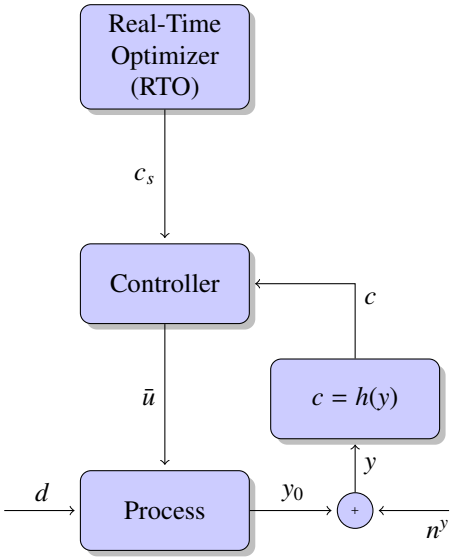


Figure 1: Hierarchical control structure. The setpoint c_s is calculated in the RTO, and passed down to a controller. The controller adjusts the inputs $\bar{u}$ such that the CV $c = h(y)$ tracks the value of the setpoint c_s closely.

is the real-time optimizer (RTO), which usually operates on a time scale of several hours and computes the setpoints c_s for the controller below which operates on a time scale of seconds and minutes. In many cases, the real-time "optimization" is done by plant operators, who adjust the setpoints of the controllers according to their experience and best practices. However, with availability of cheap computing power, the optimization of the setpoints is increasingly performed by a computer.

The controller then adapts the inputs dynamically to keep the CVs, which are functions of the measurements,

$$c = h(y), \tag{4}$$

at the setpoints c_s that are given from the RTO. The choice of the control structure is manifested in the choice of the variable c. Generally $c = h(y)$ may be any function of measurements, but it is often chosen to be linear, such that $c = Hy$ where H is a constant matrix of suitable dimensions.

Skogestad and Postlethwaite (2005) state some requirements for a good CVs:

1. The CV should be easy to control, that is, the inputs u should have a significant effect (gain) on c.
2. The *optimal value* of c should be insensitive to disturbances.
3. The CV should be insensitive to noise.
4. In case of several CVs, the variables should not be closely correlated.

A good choice of c that satisfies the requirements above will not require the RTO to update the setpoints every time the operating conditions and disturbances change. Simply controlling $c = h(y)$ at its setpoint will indirectly lead to the corresponding optimal (or near optimal) inputs u, and the control structure is referred to as a self-optimizing control structure.

Instead of focusing on finding the *optimal inputs* to the plant when disturbances vary, as done in open-loop and controller design inspired approaches such as economic model predictive control, in self-optimizing control we find or design the *optimal outputs* $c = h(y)$ of the plant. The controllers will then generate the corresponding (near-optimal) plant inputs u. Although setpoint changes may still be necessary for some major disturbances, a good control structure will require much fewer setpoint changes, and often result in good performance without an RTO layer.

To quantify performance, we define the loss L_c associated with a particular control structure as the difference between the cost resulting from that control structure (represented by the chosen $c = h(y)$) and the cost resulting from truly optimal operation (Halvorsen et al., 2003),

$$L_c = \bar{J}(\bar{u}, x, d) - \bar{J}^*(d). \tag{5}$$

The loss will be used to compare control structures, and to evaluate if a certain control structure is self-optimizing, which accomplished when "... we can achieve an acceptable loss with constant setpoint values for the controlled variables (without the need to reoptimize when disturbances occur)" (Skogestad, 2000).

2. Historical notes

With growing complexity of process plants many authors have found it increasingly necessary to find systematic ways for designing control structures that optimize process performance. Foss (1973) pointed out that one of the most prevalent problems in chemical engineering is: " Which variables should be measured, which inputs should be manipulated, and what links should be made between these two sets?". This important problem encountered by many engineers is still a research topic, and is addressed in the field of control structure selection.

A systematic approach to this question was enabled by the framework of hierarchical control systems, for which much of the theoretical foundations were laid by Mesarović et al. (1970), and which gained more recognition in the process control community by the work of Findeisen et al. (1980). The concept of a hierarchical decomposition of the control structure provided the ground for the idea of self-optimizing control. Morari et al. (1980) formulated the core idea when writing about finding good CVs, that "we want to find a function c of the process variables which when held constant, leads automatically to the optimal adjustments of the manipulated variables, and with it, the operating conditions". They also introduced the idea of the loss as a criterion to select the best feedback structure.

Closely related to the ideas presented by Morari et al. (1980) is the paper by Shinnar (1981), with the main difference being that in the latter, the main objective is tracking some variables while, Morari et al. (1980) consider a more general cost function to minimize. However, the idea is the same: To reach the objective indirectly by controlling certain measured process variables. Arbel et al. (1996) follow the ideas and introduce the concept of "dominant variables". These are variables which dominate the process behavior and are considered good CV candidates, which when kept constant, give good overall process performance. Although some examples for selecting dominant variables are given in the paper, a clear definition of the dominant variables is missing

Luyben (1988) introduces the "eigenstructure" of a process to define a control structure with self-regulating/optimizing properties. In other publications, Luyben (1975); Yi and Luyben (1995) studied the selection of CVs, which are similar to self-optimizing variables, with the

difference being that they propose to use CVs which make $\partial u/\partial d$ small, while the criterion in self-optimizing control is to find variables for which the loss is insensitive to the disturbance, i.e. $\partial L/\partial d$ small (or zero).

Rijnsdorp (1991) included a procedure for plantwide control structure selection, and proposes to use on-line algorithms to adjust the degrees of freedom optimally for the plant. Narraway et al. (1991) and Narraway and Perkins (1993) explicitly include the economics into the control structure selection, and discuss the effect of disturbances, but they do not present a general procedure for selecting CVs. The paper by Zheng et al. (1999) approaches the problem in a very similar fashion to self-optimizing control, and apply it to a reactor with separator. The work uses an economic objective function as a criterion, but does not include the effect of implementation error and measurement noise.

More recently, Engell and coworkers, (Engell et al., 2005; Engell, 2007; Pham and Engell, 2009) developed an approach for control structure selection that is based on the economic cost. Beside the effect of disturbances, they also consider measurement noise. However, instead of considering the loss during operation, they consider the effect of the control on the cost directly, and the setpoint for CVs is found by optimizing over a set of discrete representative disturbances. Heath et al. (2000) present an economics-driven approach, where a mixed integer linear problem is solved to obtain the CVs.In their case, they assume that all degrees of freedom are used to satisfy constraints.

Enagandula and Riggs (2006) propose a method for selecting the control structure by predicting the variability of the products. Rangaiah and coworkers developed an approach integrating heuristics and simulations for designing the control structure of chemical plants (Vasudevan and Rangaiah, 2011). A good review of current approaches for designing a control structure was compiled by Rangaiah and Kariwala (2012).

Bonvin and coworkers (Srinivasan et al., 2003a,b; François et al., 2005) propose to use the necessary conditions for optimality (NCO) as CV, and the goal is to make the system track the NCO. The NCO are the ideal self-optimizing CVs, except that they are generally expressed in terms of unmeasured variables, and are thus difficult to implement. In order to track the NCO, their current values must be re-constructed, either using a model, or by plant experiments.

There is also a large body of literature that considers the problem of selecting the control structure without considering economics at all. That is, only control properties like stability and interaction measures are taken into account. Examples for these approaches include Banerjee and Arkun (1995), who consider the problem from a purely control/stability perspective, as well as e.g. Samuelsson et al. (2005); Salgado and Conley (2004); Shaker and Komareji (2012). However, these approaches are out of the scope of this paper and will not discussed further at this point.

An approach that has recently gained popularity is to include the economic optimization directly into the controllers, such that the controller uses estimates of the disturbances d and states x and solves a dynamic optimization problem to calculate the optimal inputs $\bar{u}$ for the plant. This approach is commonly referred to as *Economic Model Predictive Control* (MPC) (Diehl et al., 2011; Ellis et al., 2014; Jäschke et al., 2014). Here typically a dynamic optimization problem is discretized on a finite time horizon, and repeatedly solved at given sampling times to compute the economically optimal inputs to the plant. The input for the first sampling time is injected into the plant, and the procedure is repeated at the next sampling time when new measurements are available.

Although conceptually simple, this approach is currently not used very much in practice because it requires a model that is accurate on several time-scales, from long-term economics to

short-term regulatory control and plant stabilization. Moreover, in many cases it is also difficult to obtain accurate estimates of the states (Kolås et al., 2008). For an overview over the current state of the art, we refer the reader to a recently published special issue of the Journal of Process Control that focuses on Economic Model Predictive Control edited by Christofides and El-Farra (2014). This issue also contains a review article by Ellis et al. (2014).

The class of extremum seeking methods (Ariyur and Krstic, 2003) is in this respect similar to economic MPC. In extremum seeking methods, a controller is designed that maximizes the value of a specified measurement. Although extremum seeking methods do not rely on a process model, the objective is similar to the Economic MPC approach, where control and optimization is performed simultaneously.

Finally, the concept of *self-optimizing control* was introduced in the end of the 1990s (Skogestad et al., 1998), as a strategy to achieve near-optimal operation by selecting CVs that are simply kept at their constant setpoint values. It gained wide recognition through the seminal paper by Skogestad (2000). This review paper covers the developments from 2000 onwards.

3. Brute force methods for self-optimizing CV selection

The earliest methods for finding self-optimizing CVs use a brute-force approach (Skogestad, 2000; Larsson et al., 2001; Govatsmark, 2003). The idea is to simply evaluate the performance of all possible candidate CV sets for all possible values of disturbances and measurement noise.

For example, consider a plant with $n_y = 3$ measurements, and $n_{\bar{u}} = 2$ degrees of freedom. For all possible disturbance and noise values, we calculate the loss resulting from keeping the CV candidates $c_{(1)} = [y_1\, y_2]^T$, $c_{(2)} = [y_1\, y_3]^T$, and $c_{(3)} = [y_2\, y_3]^T$ at constant setpoints. Then we select the CV candidate that gives the lowest loss. If the loss is acceptable, then self-optimizing control is achieved.

The evaluation may be based on either the worst-case loss, or the average loss associated with a particular choice of CVs. The worst-case loss is calculated as

$$L_{c,wc} = \max_{d\in\mathcal{D},\, n^y\in\mathcal{N}} L_c, \tag{6}$$

and the average loss is defined as

$$L_{c,av} = \mathop{\mathbb{E}}_{d\in\mathcal{D},n^y\in\mathcal{N}} [L_c]\,. \tag{7}$$

Here $\mathbb{E}[\cdot]$ denotes the expectation operator and the sets $\mathcal{D} \subset \mathbb{R}^{n_d}$ and $\mathcal{N} \subset \mathbb{R}^{n_y}$ contain all allowable disturbance and noise values. The variable L_c denotes the loss associated with controlling the variable c at a constant setpoint for a given value of disturbance and measurement noise.

Since the sets of all possible disturbances and noise realizations, $\mathcal{D}$ and $\mathcal{N}$, generally have an infinite number of points, we evaluate different samples (scenarios) $i = 1 \ldots M$, which may be, for example, uniformly distributed in $\mathcal{D}$ and $\mathcal{N}$. The values of d and n^y for the i-th scenario are denoted by $d^{(i)}, n^{y(i)}$, respectively, and the corresponding optimal cost $\bar{J}^*(d^{(i)})$ is obtained by solving (1) parametrized by $d^{(i)}$.

Given a candidate CV set $c_{(j)}$ and its setpoint $c_{s(j)}$, for each scenario $i = 1 \ldots M$ the cost $\bar{J}^{(i)}_{c_{(j)}}$

obtained by controlling $c_{(j)}$ to its setpoint is found as the solution of the following problem:

$$\begin{aligned} \min_{\bar{u}} \; & \bar{J}(\bar{u}^{(i)}, x^{(i)}, d^{(i)}) \\ & \text{s.t.} \\ & f(\bar{u}^{(i)}, x^{(i)}, d^{(i)}) = 0 \\ & g(\bar{u}^{(i)}, x^{(i)}, d^{(i)}) \leq 0 \\ & c_{(j)}(\bar{u}^{(i)}, x^{(i)}, d^{(i)}, n^{y(i)}) - c_{s(j)} = 0 \end{aligned} \tag{8}$$

Typically, the setpoint $c_{s,(j)}$ is chosen as the nominally optimal value that is obtained when optimizing the system with $d = d_{nom}$ and $n^y = 0$. In the cases where this choice may be infeasible, Govatsmark (2003) presents an approach to obtain a robust and feasible setpoint. All candidate sets $c_{(j)}$, $j = 1 \ldots n_j$ are then ranked using one of the loss expressions (6) or (7). For example, the average loss

$$L_{c_{(j)},av} = \mathop{\mathbb{E}}_{i=1..M} \left[\bar{J}^{(i)}_{c_{(j)}} - \bar{J}^*(d^i) \right]. \tag{9}$$

A similar approach may also be applied for finding combinations of measurements as CVs, $c = Hy$ (Umar et al., 2012). This, however, leads to large-scale nonlinear bilevel optimization problems that can be very difficult to solve.

The brute force approaches requires the solution of many optimization problems. Problem (8) needs to be solved for each CV candidate $j = 1 \ldots n_j$ for all $i = 1 \ldots M$ scenarios. When selecting $n_{\bar{u}}$ single measurements out of a total n_y measurements as CVs, there are

$$C^{n_{\bar{u}}}_{n_y} = \left(\begin{array}{c} n_y \\ n_{\bar{u}} \end{array} \right) = \frac{n_y!}{(n_y - n_{\bar{u}})! n_{\bar{u}}!} \tag{10}$$

different control structures that may be chosen. Thus, the number of possible control structures grows rapidly with the number of measurements n_y. For real plants where the number of candidate CVs can become very large, this approach becomes intractable.

To reduce the number of optimization problems, the noise may be neglected by setting $n^y = 0$. However, if there are many disturbances and CV candidates, the approach will still be intractable. Moreover neglecting measurement noise may render a CV candidate infeasible when noise is present in the real plant (Govatsmark, 2003).

Generally the optimization problems (1) and (8) solved at each scenario $i = 1 \ldots M$ are large-dimensional and non-convex. Thus, the very large number of difficult optimization problems may render the approach practically infeasible, unless some heuristics are devised to reduce the number of CV candidates.

4. Local methods for steady-state self-optimizing CVs

To limit the number of CVs that are considered, and to exclude poor CV choices early in the control structure design, local methods have been developed. The motivation for local methods is that a candidate CV must perform well locally around the nominal operating point where the process is expected to operate most of the time, otherwise it may be excluded immediately. Only CVs that are found to perform well close to the nominal conditions are then further examined and tested over the whole operating region.

4.1. Local approximation of the cost function

Starting from the steady-state problem (1), the model equations $f(\bar{u}, x, d) = 0$ are used to formally eliminate the states x from the optimization problem. This yields a simplified reduced problem

$$\begin{aligned} &\min_{\bar{u}} \bar{J}^r(\bar{u}, d) \\ &\text{s.t.} \\ &g(\bar{u}, d) \leq 0. \end{aligned} \tag{11}$$

Assuming further that the set of active constraints does not change under operation[3], the first step is to control the active constraints at their optimal values. Then the active constraints may be formally eliminated from (11) together with their corresponding degrees of freedom. This results in the unconstrained problem

$$\min_{u} J(u, d), \tag{12}$$

where $J(u, d)$ denotes the cost function of the unconstrained problem, and u denotes the remaining unconstrained degrees of freedom that are left after satisfying all the active constraints. These unconstrained degrees of freedom u will be used to control self-optimizing CVs. We denote the optimal cost function value of (12) as $J^*(d)$, and we note that as long as the disturbance d does not cause the set of active constraints to change, we have $J^*(d) = \bar{J}^*(d)$.

To find good self-optimizing CV candidates, the nonlinear optimization problem (12) is approximated locally by a quadratic function around the nominally optimal operating point. Introducing deviation variables $\Delta u = u - u_{nom}$ and $\Delta d = d - d_{nom}$ and where u_{nom} is the nominally optimal input corresponding to the nominal disturbance d_{nom}, a Taylor expansion around the nominally optimal point gives

$$\begin{aligned} J(\Delta u, \Delta d) \approx J_{nom} &+ \begin{bmatrix} J_u & J_d \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} \\ &+ \frac{1}{2} \begin{bmatrix} \Delta u^T & \Delta d^T \end{bmatrix} \begin{bmatrix} J_{uu} & J_{ud} \\ J_{du} & J_{dd} \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} \end{aligned} \tag{13}$$

Here $J_u = \frac{\partial J}{\partial u}$ and $J_d = \frac{\partial J}{\partial d}$ denote the derivatives of the cost function with respect to u and d, respectively, and $J_{uu} = \frac{\partial^2 J}{\partial u^2}$, $J_{ud} = J_{du}^T = \frac{\partial^2 J}{\partial u \partial d}$ and $J_{dd} = \frac{\partial^2 J}{\partial d^2}$ denote the second derivatives of the cost function (12), evaluated at $u = u_{nom}$ and $d = d_{nom}$.

Since we are approximating the cost around the nominally *optimal* point, we have that $J_u = 0$. Differentiating (13) with respect to Δu and equating the expression to zero yields the optimality condition that must hold in order to remain locally optimal.

$$\frac{\mathrm{d}J}{\mathrm{d}u} \approx \underbrace{J_u}_{=0} + J_{uu}\Delta u + J_{ud}\Delta d = \begin{bmatrix} J_{uu} & J_{ud} \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} = 0 \tag{14}$$

Note that J_u, J_{uu}, J_{ud} are evaluated at the *nominal optimal* point, while $\frac{\mathrm{d}J}{\mathrm{d}u}$ is the gradient value at a point nearby given by Δu and Δd. Solving for the *optimal input* $\Delta u^*(d)$, yields

$$\Delta u^*(d) = -J_{uu}^{-1} J_{ud} \Delta d. \tag{15}$$

[3]This assumption can be relaxed by using the methods described in Section 5.

Using the quadratic cost function (13) and the optimal input (15), it can be shown that the loss from optimality can be expressed as (Alstad, 2005)

$$L = J(u,d) - J^*(d) = \frac{1}{2}\left(\Delta u - \Delta u^*(d)\right) J_{uu} \left(\Delta u - \Delta u^*(d)\right), \tag{16}$$

or alternatively

$$L = \frac{1}{2}\|z\|_2^2 \tag{17}$$

where $\|\cdot\|_2$ denotes the two-norm, and with z defined as

$$z = J_{uu}^{1/2}(\Delta u - \Delta u^*(d)). \tag{18}$$

This loss expression is the basis for all local methods presented in this section.

Moreover, by inserting (15) into (18), we obtain

$$z = J_{uu}^{-1/2}\begin{bmatrix} J_{uu} & J_{ud} \end{bmatrix}\begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}. \tag{19}$$

Comparing (19) with the gradient expression in (14) we observe that the local loss from (17) can be written as the squared weighted 2-norm of the gradient value:

$$\begin{aligned} L &= \frac{1}{2}\left\| J_{uu}^{-1/2}\begin{bmatrix} J_{uu} & J_{ud} \end{bmatrix}\begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}\right\|_2^2 \\ &= \frac{1}{2}\left\|\left(J_{uu}^{-1/2}\right)\frac{\mathrm{d}J}{\mathrm{d}u}\right\|_2^2. \end{aligned} \tag{20}$$

Here the gradient expression $\frac{\mathrm{d}J}{\mathrm{d}u}$ is the actual plant gradient under operation (not at the nominal point). We see that when the gradient is zero (a necessary condition for optimality), then also the local loss becomes zero.

4.2. Local approximation of the plant and linear measurement combination

The measurement model (3) is linearized around the nominal point,

$$\Delta y = G^y \Delta u + G_d^y \Delta d + n^y, \tag{21}$$

where $\Delta y = y - y_{nom}$, and $G^y = \frac{\partial y}{\partial u}$, and $G_d^y = \frac{\partial y}{\partial d}$ are Jacobian matrices of appropriate sizes, that are evaluated around the nominally optimal point, and which represent the gain from the inputs and disturbances, respectively, to the outputs. Choosing the CVs to be a linear combination of measurements y gives

$$\Delta c = H\Delta y, \tag{22}$$

where we call $H \in \mathbb{R}^{n_u \times n_y}$ the measurement selection or combination matrix. Inserting (21) into (22) yields

$$\Delta c = HG^y \Delta u + HG_d^y \Delta d + Hn^y. \tag{23}$$

Generally the elements of H may take arbitrary values, as long as $\mathrm{rank}(HG^y) = n_u$. This is required to obtain a linearly independent set of CVs that fully specify the system. If a set of single measurements is selected as CVs, then each row of H will contain exactly one "1", and have zero-entries otherwise. In this case H has the property

$$HH^T = I_{n_u}, \tag{24}$$

where I_{n_u} denotes the Identity matrix of dimension $n_u \times n_u$.

4.3. Loss for a given control structure (Exact Local Method)

To evaluate the loss corresponding to a given control structure we substitute the input that is generated by controlling $\Delta c = H\Delta y$ to zero, together with the expression for the optimal input (15) into the loss expression (16). For a given disturbance Δd and given control structure represented by H, the input Δu generated by the controllers can be found by solving (23) for Δu,

$$\begin{aligned}\Delta u &= (HG^y)^{-1}\,(\underbrace{\Delta c}_{=0} - HG^y_d\Delta d - Hn^y)\\ &= -(HG^y)^{-1}\,H\left(G^y_d\Delta d + n^y\right).\end{aligned} \tag{25}$$

Inserting the input Δu derived in (25) and the optimal input from (15) into the loss expressions (16) and (17), the loss variable z for a given disturbance and measurement combination matrix H becomes (Halvorsen et al., 2003)

$$z = -J_{uu}^{1/2}\,(HG^y)^{-1}\,H\left[\left(G^y_d - G^yJ_{uu}^{-1}J_{ud}\right)\Delta d + n^y\right]. \tag{26}$$

This may be simplified to

$$z = -J_{uu}^{1/2}\,(HG^y)^{-1}\,H\left[F\Delta d + n^y\right], \tag{27}$$

where we used the optimal sensitivity matrix

$$F = G^y_d - G^yJ_{uu}^{-1}J_{ud}. \tag{28}$$

Given a measurement combination matrix H, a disturbance Δd and a current noise value n^y, the loss is calculated as $L = \|z\|_2^2$.

The matrix F represents the sensitivities of the optimal measurement values with respect to disturbances

$$F = \frac{\partial y^*}{\partial d}. \tag{29}$$

To see this, we insert the optimal input Δu^* from (15) into the measurement equation (21) to yield the optimal measurement variation Δy^* for a change Δd,

$$\begin{aligned}\Delta y^* = y^* - y_{nom} &= -G^yJ_{uu}^{-1}J_{ud}\Delta d + G^y_d\Delta d + n^y\\ &= \left(-G^yJ_{uu}^{-1}J_{ud} + G^y_d\right)\Delta d + n^y\\ &= F\Delta d + n^y.\end{aligned} \tag{30}$$

Noting that $\Delta d = d - d_{nom}$, by differentiating (30) with respect to d, we obtain $F = \frac{\partial y^*}{\partial d}$.

The optimal sensitivity matrix F can be calculated using either (28), re-optimization and finite differences, or nonlinear programming sensitivity (Fiacco, 1983; Pirnay et al., 2012) based on the inverse function theorem.

To evaluate the performance of a given set of CVs $\Delta c = H\Delta y$ for a range of disturbances, we define diagonal scaling matrices of appropriate sizes for the disturbances and noise, W_d and W_n, respectively, such that

$$\Delta d = W_d d'; \quad n^y = W_n n'. \tag{31}$$

Here d' and n' denote the scaled disturbance and measurement noise, respectively. This scaling of the variables allows us to make statements about the loss for a given set of disturbances and noise with magnitudes defined in W_d and W_n.

Using the scaling matrices, the loss Eq. (17) for a specific value of d' and n' is

$$L = \frac{1}{2}\left\| J_{uu}^{1/2}(HG^y)^{-1}HY\begin{bmatrix} d' \\ n' \end{bmatrix}\right\|_2^2 \tag{32}$$

where

$$Y = [FW_d \quad W_n]. \tag{33}$$

To simplify notation, we introduce the loss matrix M as

$$M = J_{uu}^{1/2}(HG^y)^{-1}HY, \tag{34}$$

so (32) can be written compactly as

$$L = \frac{1}{2}\left\| M\begin{bmatrix} d' \\ n' \end{bmatrix}\right\|_2^2 \tag{35}$$

Remark 1 (Important observation). *The loss matrix $M = J_{uu}^{1/2}(HG^y)^{-1}HY$, has an interesting property that will be used later for finding combinations of measurements that minimize the loss: The value of M does not change when H is pre-multiplied by any invertible matrix Q. To see this, consider the matrix M and assume that the CV is $\Delta c = \hat{H}\Delta y$, where $\hat{H} = QH$. Then we have*

$$\begin{aligned} M &= J_{uu}^{1/2}(\hat{H}G^y)^{-1}\hat{H}Y \\ &= J_{uu}^{1/2}(QHG^y)^{-1}QHY \\ &= J_{uu}^{1/2}(HG^y)^{-1}Q^{-1}QHY \\ &= J_{uu}^{1/2}(HG^y)^{-1}HY. \end{aligned} \tag{36}$$

This reflects the fact that scaling a CV will generally not affect the loss at steady state.

Depending on how the disturbances and errors are assumed to be distributed, we have different cases for the worst-case and the average errors.

- **Two-norm bounded disturbance and noise (Halvorsen et al., 2003).** When assuming that the disturbances and noise are independent and uniformly distributed over the set

  $$\mathcal{DN}_2 = \left\{(d', n') \,\middle|\, \left\|[d' \quad n']^T\right\|_2 \le 1, \right\} \tag{37}$$

  the worst-case loss is derived based on (35):

  $$L_{\text{worst}} = \max_{\left\| \begin{bmatrix} d' \\ n' \end{bmatrix} \right\|_2 \le 1} \frac{1}{2}\left\| M\begin{bmatrix} d' \\ n' \end{bmatrix}\right\|_2^2 = \frac{1}{2}\|M\|_2^2 = \frac{1}{2}\bar{\sigma}^2(M), \tag{38}$$

  where $\bar{\sigma}(\cdot)$ denotes the largest singular value.

- **Infinity-norm bounded disturbance and noise (Kariwala et al., 2008).** If the noise and the disturbance variables n' and d' are assumed independent and uniformly distributed in the set

  $$\mathcal{DN}_\infty = \left\{(d', n') \,\middle|\, \left\|[d' \quad n']^T\right\|_\infty \le 1, \right\}, \tag{39}$$

the average (expected value of the loss) is

$$L_{\text{av}} = \mathop{\mathbb{E}}_{n',d' \in \mathcal{DN}_\infty} \left[\frac{1}{2} \left\| M \begin{bmatrix} d' \\ n' \end{bmatrix} \right\|_2^2 \right] = \frac{1}{6} \|M\|_F^2 \,. \tag{40}$$

where $\mathbb{E}[\cdot]$ again denotes the expectation operator.

- **Normally distributed disturbance and noise (Kariwala et al., 2008).** Assuming that d' and n' are normally distributed with zero mean and unit variance,

$$\mathcal{DN}_{\mathcal{N}} = \{d' \sim \mathcal{N}(0, I), n' \sim \mathcal{N}(0, I)\}\,, \tag{41}$$

the worst-case loss becomes unbounded $L_{\text{wc}} = \infty$, as n' and d' may become arbitrarily large. The average loss, however is

$$L_{\text{av},\mathcal{N}} = \mathop{\mathbb{E}}_{n',d' \in \mathcal{DN}_{\mathcal{N}}} \left[\frac{1}{2} M \begin{bmatrix} d' \\ n' \end{bmatrix} \right] = \frac{1}{2} \|M\|_F^2 \,. \tag{42}$$

We observe that the assumptions on the disturbance and noise distribution affect the worst-case and average loss in form of a scaling factor only. Thus, any of the loss expressions given above may be used to rank the candidate CVs.

Remark 2. *Note that other loss expressions were derived, too, such as the average loss expression for the case with 2-norm bounded, uniform distributed noise and disturbances (Kariwala et al., 2008). However, those results do not make practical sense, as the problem was formulated in such a way that the loss value is scaled with the number of measurements n_y. This gives misleading results, because the loss value may be changed arbitrarily, e.g. by adding measurements (increasing n_y), while adding corresponding columns of zeros in H.*

Example 1. *Consider the following process (Halvorsen et al., 2003), with $n_y = 4, n_u = 1, n_d = 1$. The measurement gains from the input u and the disturbance d are*

$$G^y = \begin{bmatrix} 0.1 \\ 20 \\ 10 \\ 1 \end{bmatrix}, \quad G^y_d = \begin{bmatrix} -0.1 \\ 0 \\ -5 \\ 0 \end{bmatrix}, \tag{43}$$

and the cost to minimize is

$$J = (u - d)^2, \tag{44}$$

For this process, we have that $J_{uu} = 2, J_{ud} = -2$, and the optimal sensitivity matrix becomes

$$F = \frac{\partial y^*}{\partial d} = G^y_d - G^y J_{uu}^{-1} J_{ud} = \begin{bmatrix} 0 \\ 20 \\ 5 \\ 1 \end{bmatrix}. \tag{45}$$

Assuming further that the noise and disturbance magnitudes are given as $W_d = 1$ and $W_n = I_4$, and that $\left\| [d'\ n']^T \right\|_2 \leq 1$, we evaluate the worst-case loss obtained from controlling single

measurements:

$$\begin{aligned}
\Delta c_1 &= H_1 \Delta y = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \Delta y = \Delta y_1 \\
\Delta c_2 &= H_2 \Delta y = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \Delta y = \Delta y_2 \\
\Delta c_3 &= H_3 \Delta y = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \Delta y = \Delta y_3 \\
\Delta c_4 &= H_4 \Delta y = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \Delta y = \Delta y_4
\end{aligned} \tag{46}$$

The worst-case loss obtained with a constant setpoint policy ($\Delta c = 0$) is evaluated using (38) *as $L_{wc}(c_1) = 100$, $L_{wc}(c_2) = 1.0025$, $L_{wc}(c_3) = 0.26$, $L_{wc}(c_4) = 2$. Therefore, controlling c_3 is the best choice, as it yields the smallest worst-case loss for the given set of noise and disturbances.*

We further observe that the loss varies by orders of magnitude, depending on the selected control structure. This shows that significant economic benefits can be achieved by simply designing a good control structure for a process.

Remark 3. *Alstad (2005) showed that the loss for a given control structure is locally independent of the chosen setpoint. In particular, as long as the quadratic approximation of the cost and the linearized model do not change, the best control structure remains the best, even when the setpoint is not optimal.*

Remark 4 (Maximum scaled gain rule). *Before the exact local results presented above were found, Skogestad and Postlethwaite (1996) described an approximative method, called the "Maximum Scaled Gain Rule" (also known as "minimum singular value rule") for ranking CV candidates. Since the exact methods presented in this section are just as easy to apply, we do not present the Maximum Scaled Gain rule here. However, for completeness sake we describe it in Appendix A.*

4.4. Finding optimal measurement combinations

Using the results above, it is easily possible to evaluate the average and worst-case loss corresponding to a given control structure represented by H. This gives rise to the question of how to find a control structure which results in a minimal loss. Loosely speaking this can be formulated as an optimization problem to minimize the loss for a set of implementable control structures. For example, for finding a measurement combination $\Delta c = H\Delta y$ that minimizes the average loss for normal distributed disturbances and noise we could write

$$\begin{aligned}
\min_H L = \frac{1}{2} \left\| J_{uu}^{1/2} (HG^y)^{-1} HY \right\|_F^2 \\
\text{s.t. } HG^y \text{ invertible .}
\end{aligned} \tag{47}$$

In principle, any of the local loss expressions (38), (40), (42) given in the previous section may be used in the cost function of (47) to find combination matrices. In general, these loss expressions lead to non-convex optimization problems, which may have multiple optima, and are difficult to solve. However, in the case where H may take arbitrary values, or when the structure of H is preserved upon left-multiplication with any non-singular matrix, the non-convex problem (47) may be simplified to a convex problem, such that the optimal measurement combination H is easy to find. These cases will be discussed later in this section.

4.4.1. Null-space method

In the Null-space method (Alstad and Skogestad, 2007) for finding the optimal measurement combination H, we assume that the measurement noise can be neglected ($W_n = 0$), so we only need to compensate for disturbances d. Moreover, we assume that we have $n_y \geq n_u + n_d$ independent measurements. Under these assumptions, it is possible to find a matrix H that gives zero loss $L_{wc} = L_{av} = 0$, by simply selecting H in the left null-space of F, such that

$$HF = 0. \tag{48}$$

To show this, consider the worst-case loss $L_{wc} = \frac{1}{2}\bar{\sigma}\left(J_{uu}^{1/2}(HG^y)^{-1}HY\right)^2$. Without noise $W_n = 0$, such that $Y = [FW_d \quad 0]$, and the resulting worst-case loss is simplified to $L_{wc} = \frac{1}{2}\bar{\sigma}\left(J_{uu}^{1/2}(HG^y)^{-1}HFW_d\right)^2$. Selecting H such that $HF = 0$ will then result in zero worst-case and average loss $L_{wc} = L_{av} = 0$ (as long as HG^y is non-singular).

Remark 5. *Alternatively we may arrive at this result by requiring the optimal setpoint change to be zero. From* (22)*,* (29) *and* (31) *the optimal change in the CVs can be expressed as*

$$\Delta c_{\text{opt}} = H\Delta y_{\text{opt}} = HFW_d d' = 0. \tag{49}$$

Selecting H in the left null-space of F will make the optimal variation $\Delta c_{\text{opt}} = 0$. This means that keeping the setpoint constant $\Delta c = 0$ will result in the optimal change in the measurements, which in turn implies that the inputs assume their optimal values.

Remark 6. *There exists an interesting relationship between the null-space method and the gradient (Jäschke and Skogestad, 2011a). Using $J_u = [J_{uu} \quad J_{ud}]\begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}$ from* (14)*, Setting $n^y = 0$ and defining $\tilde{G}^y = [G^y \quad G_d^y]$, we can write* (21) *as*

$$\Delta y = G^y\Delta u + G_d^y\Delta d = \tilde{G}^y\begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}. \tag{50}$$

Since we have $n_y \geq n_u + n_d$, we can invert the measurement relationship (50) *and inserting in the gradient* (14) *gives*

$$\Delta c = \Delta J_u = [J_{uu} \quad J_{ud}]\left[\tilde{G}^y\right]^{\dagger}\Delta y. \tag{51}$$

Therefore, controlling Δc to zero forces the gradient to become zero, $J_u = 0$. It can be verified that

$$H = [J_{uu} \quad J_{ud}]\left[\tilde{G}^y\right]^{\dagger} \tag{52}$$

is in the left null-space of F, which from (28) *can be written as*

$$F = \tilde{G}^y\begin{bmatrix} -J_{uu}^{-1}J_{ud} \\ I \end{bmatrix}, \tag{53}$$

where I denotes the identity matrix.

Considering the optimal sensitivity matrix F in (53) more closely, we can see that the null-space of F can be decomposed into two parts. The first component is given by $H = \left[J_{uu} \quad J_{ud}\right]\left[\tilde{G}^y\right]^{\dagger}$ and the second component is given by the left null-space of $\tilde{G}^y$. The directions corresponding to

this second part with $\tilde{H}\tilde{G}^y = 0$ correspond to invariant output directions, that cannot be controlled and that are always zero.[4] When applying the null-space method, it must therefore be verified that the selected measurement combination is *not* in the null-space of $\tilde{G}^y$.

The null-space method is not optimal in a realistic setting, because it does not take measurement noise into account. Furthermore, the requirement of at least as many measurements as the sum of number of inputs and disturbances leads to complex control structures involving many variables. However, the derivation of the null-space method in Remark 6 can be used as a starting point for approaching a more difficult problem, where measurement polynomials are used as CVs, see Section 6.1.

Example 2 (Null-space method). *Consider* $F = \begin{bmatrix} 0 & 20 & 5 & 1 \end{bmatrix}^T$ *from* (45) *in Example 1. A measurement selection matrix in the left null-space of* F *is* $H_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$*, corresponding to selecting the CV* $c_1 = y_1$*. However the left null-space of* F *is spanned by three basis vectors, so we may also chose different measurement combinations, such as given by* $H_2 = \begin{bmatrix} 0 & 1 & -4 & 0 \end{bmatrix}$*, and* $H_3 = \begin{bmatrix} 0 & 1 & 0 & -20 \end{bmatrix}$.

Without measurement noise ($W_n = 0$*) controlling* c_1 *and* c_2 *gives* $L_{wc}(c_1) = L_{wc}(c_2) = 0$*, yielding optimal operation in spite of disturbances. However, the measurement combination* c_3 *lies in the left null-space of* F*, but also in the left null-space of* $\tilde{G}^y$*, and hence* G^y*. This "controlled variable" is therefore not controllable because the gain* HG^y *from inputs to CVs is zero. Consequently, the loss* $L_{wc} = \frac{1}{2}\bar{\sigma}(J_{uu}^{1/2}(HG^y)^{-1}HY)$ *may assume arbitrarily large values.*

Next we show that if measurement noise is present ($W_n \neq 0$*) the loss can depend significantly on the choice of the basis vectors from the left null-space. Given measurement noise and disturbances of magnitude* $W_n = I_4$ *and* $W_d = 1$*, we evaluate the worst-case loss for the different alternatives to be*

$$\begin{aligned} L_{wc}(c_1) &= 100 \\ L_{wc}(c_2) &= 0.04250 \\ L_{wc}(c_3) &= \infty. \end{aligned} \tag{54}$$

4.4.2. Extended Null-space method

As shown in the example above, the choice of vectors in the left null-space of F does have a significant influence on the performance in presence of noise. If $n_u + n_d > n_y$, the H-matrix can be chosen such that beside perfectly rejecting disturbances, the effect of noise is minimized. That is, after all disturbance effects are rejected, the remaining measurements are used to minimize the effect of noise. Next, we give a new derivation of the results in Alstad et al. (2009). First use the scaling matrix W_n to define scaled measurements y', such that

$$y' = W_n^{-1}\Delta y. \tag{55}$$

In presence of measurement noise n^y, the expression for the scaled measurement is calculated using (21) and $\tilde{G}^y = [G^y \quad G_d^u]$ as

$$y' = W_n^{-1}\Delta y = W_n^{-1}\tilde{G}^y \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} + W_n^{-1}n^y. \tag{56}$$

[4] When $\tilde{H}$ is selected such that $\tilde{H}\tilde{G}^y = 0$, then also $\tilde{H}G^y = 0$, and we have selected a measurement combination that is uncontrollable.

Rearranging gives

$$W_n^{-1} n^y = -W_n^{-1} \Delta y + W_n^{-1} \tilde{G}^y \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} \tag{57}$$

The least squares solution $[\Delta u\, \Delta d]^T$ for (57), which minimizes the sum of squares of the weighted noise $n^{yT} W_n^{-1} W_n^{-1} n^y$ is given by

$$\begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} = (W_n^{-1} \tilde{G}^y)^\dagger W_n^{-1} \Delta y. \tag{58}$$

Inserting this least squares solution into the expression for the gradient (14) we obtain

$$c = \begin{bmatrix} J_{uu} & J_{ud} \end{bmatrix} (W_n^{-1} \tilde{G}^y)^\dagger W_n^{-1} \Delta y, \tag{59}$$

and the corresponding H-matrix is

$$H = \begin{bmatrix} J_{uu} & J_{ud} \end{bmatrix} (W_n^{-1} \tilde{G}^y)^\dagger W_n^{-1}. \tag{60}$$

Except for a constant scaling factor that does not affect the loss, this is the same expression as originally derived in Alstad et al. (2009).

Note that the matrix H from (60) is in the left null-space of F. To see this, we write F as in (53), and obtain

$$\begin{aligned} HF &= \begin{bmatrix} J_{uu} & J_{ud} \end{bmatrix} (W_n^{-1} \tilde{G}^y)^\dagger W_n^{-1} \tilde{G}^y \begin{bmatrix} -J_{uu}^{-1} J_{ud} \\ I \end{bmatrix} \\ &= -J_{ud} + J_{ud} = 0. \end{aligned} \tag{61}$$

By selecting H as in (60), the basis vectors in the left null-space of F are chosen to minimizes the effect of noise.

Example 3 (Extended Null-space method). *Inserting the values for Example 1 into* (60) *yields*

$$H_4 = \begin{bmatrix} 0.0085 & -0.0997 & 0.3998 & -0.0050 \end{bmatrix}, \tag{62}$$

and controlling $c_4 = H_4 y$ at a constant value leads to a loss of $L_{wc}(c_4) = 0.04247$. This loss is significantly lower than $L_{wc}(c_1) = 100$ that is obtained by selecting $H_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$. Note that both, H_1 and H_4 are in the left null-space of F.

4.4.3. Minimum Loss method (Explicit solution)

In practice the assumption of exact measurements is not valid, and the optimal measurement combination matrix H must take the measurement noise n^y into account. Although this is done in the Extended Null-Space Method described above, there are two drawbacks in this approach. (1) it only works when we have more measurements than the sum of number of inputs and disturbances, so the number of required measurements can become very large, and (2) the Extended Null-Space Method will generally not give an optimal trade-off between noise and disturbance rejection. For example, in some cases it may be beneficial to not reject a disturbance completely in order to compensate for the even more detrimental effect of noise.

In the literature there are several derivations to the solution of finding a locally optimal measurement combination matrix H, that finds the optimal trade-off between rejecting the disturbances and measurement noise. Kariwala (2007) and Kariwala et al. (2008) present methods

which are based on determining eigenvalues of a matrix, while Heldt (2010) gives a related method based on a generalized singular value decomposition. Alstad et al. (2009) reformulate the problem of minimizing the nonlinear loss expression to obtain a convex quadratic optimization program. All these approaches lead to the same solution, so for brevity we only present the last approach by Alstad et al.

We start by recalling that pre-multiplying H by any invertible matrix Q will not affect the value of the loss matrix $M = J_{uu}^{1/2}(HG^y)^{-1}HY$, see (36). Thus, we may select Q to make $HG^y = J_{uu}^{1/2}$. This cancels the nonlinearity in M, and a matrix H that minimizes the average loss can be found by solving the convex optimization problem (Alstad et al., 2009)

$$\begin{aligned} &\min_H \|HY\|_F \\ &\text{s.t.} HG^y = J_{uu}^{1/2}. \end{aligned} \tag{63}$$

An explicit solution to this problem was found by Alstad et al. (2009), and later simplified by Yelchuru and Skogestad (2012) to

$$H = (G^y)^T \left(YY^T\right)^{-1}. \tag{64}$$

Equation (64) provides the locally best measurement combination for a given set of measurements, provided YY^T is invertible, which always is the case when all measurements are affected by noise. It was found by Kariwala et al. (2008) that the H-matrix that minimizes the average loss, also minimizes the worst-case loss. Note that the full loss expressions for the worst-case loss (38) or the average loss (40), (42) must be used for evaluating the value of loss that is obtained with H from (64).

Example 4 (Exact local method). *Using* (64) *for calculating the measurement combination matrix for the system in Example 1 with the noise and disturbances magnitudes $W_n = I_4$ and $W_d = 1$, gives*

$$H_5 = \begin{bmatrix} 0.1000 & -1.1241 & 4.7190 & -0.0562 \end{bmatrix}, \tag{65}$$

and the corresponding loss is $L_{wc}(c_5) = 0.0405$.

This is the measurement combination which gives the best performance locally. We observe that the loss is slightly lower than in the case of the Extended Null-space method. This is because an optimal trade-off between disturbance rejection and noise rejection is found. Here again, we see that the measurements with the largest weights are y_2 and y_3.

4.5. Incorporating measured disturbances

To improve the self-optimizing properties of the control structure in presence of known (measured) disturbances, e.g. prices or feed quality, a straightforward approach proposed by Jäschke and Skogestad (2011) is to include the measured disturbances into the measurement vector, and use e.g. (64) to calculate the optimal measurement combination. The augmented measurement vector y^{aug} becomes

$$y^{aug} = \begin{bmatrix} y \\ d^m \end{bmatrix}, \tag{66}$$

where d^m denotes the measured disturbances. The corresponding optimal CV can then be found by any of the methods described above, and written as

$$\begin{aligned} \Delta c^{aug} &= H^{aug}\Delta y^{aug} = \begin{bmatrix} H & H^d \end{bmatrix}\begin{bmatrix} \Delta y \\ \Delta d^m, \end{bmatrix} \\ &= H\Delta y + H^d\Delta d^m. \end{aligned} \tag{67}$$

Instead of keeping this CV at a constant setpoint, $\Delta c^{aug} = H\Delta y + H^d\Delta d^m = 0$, one may use the measured disturbance to change the setpoints of the CVs, and instead control

$$\Delta c = H\Delta y, \tag{68}$$

to the setpoint

$$\Delta c_s = -H^d\Delta d^m. \tag{69}$$

The concept is illustrated in the block diagram in Figure 2, and has also been applied in Umar et al. (2014).

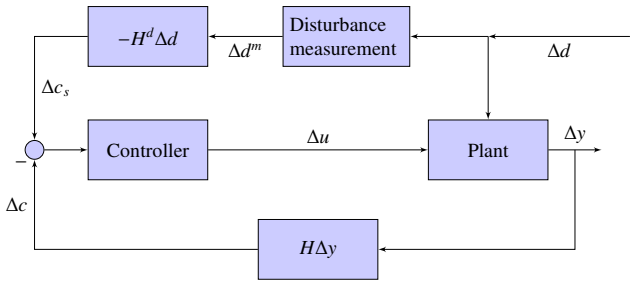


Figure 2: Optimal setpoint adaption to measured disturbances

4.6. Selecting subsets of measurements

In most practical cases, using a combination of all available measurements in the H matrix is neither desired nor required. Usually, controlling a combination of a subset of the available measurements results in a performance which is similar to when using all available measurements (Kariwala, 2007; Kariwala et al., 2008). The result is a simpler control structure at a usually insignificant increase in loss.

Selecting the best subset of n variables from n_y variables leads to a combinatorial optimization problem, where the loss L has to be evaluated individually for each CV. The number of possible combinations can become very large, and selecting n variable combinations as CVs from a set of n_y measurements leads to $C^n_{n_y} = \begin{pmatrix} n_y \\ n \end{pmatrix} = \frac{n_y!}{(n_y-n)!n!}$ possible structures. In the literature, the problem of selecting the best subset of measurements has been approached in two ways: The first one is to develop tailor-made branch and bound algorithms (Cao and Kariwala, 2008; Kariwala and Cao, 2009, 2010) , and the second approach formulates the selection problem as a mixed integer quadratic optimization problem (MIQP), and uses standard MIQP solvers to obtain the best measurement set (Yelchuru and Skogestad, 2012).

4.6.1. Tailor-made branch and bound methods

These methods exploit monotonicity of the optimal loss value as a function of the number of variables. This means that when removing one measurement from the CV, the optimal loss cannot decrease. Here we present the general idea for minimizing the average loss for normally distributed disturbances and noise. For a more detailed description and treatment of the worst-case, we refer to Cao and Kariwala (2008), and Kariwala and Cao (2009, 2010).

Denote by X the index set describing the selected measurements, and by $|X|$ its cardinality. Further, let G_X^y denote the gain matrix corresponding to the selected measurements, i.e. the rows of G^y corresponding to the indices in X. Analogously Y_X denotes the matrix consisting of the rows in Y corresponding to the index set X, and $H_X^o \in \mathbb{R}^{n_u \times |X|}$ denote the *optimal* H-matrix corresponding to the chosen subset of measurements. The matrix H_X^o corresponds to the optimal solution of problem (63), when using G_X^y and Y_X

Then, for given noise and disturbances that are normally distributed, the *optimal* average loss is a function of the index set X:

$$L_{av}(X) = \frac{1}{2} \left\| J_{uu}^{1/2} \left((H_X^o G_X^y) \right)^{-1} H_X^o Y_X \right\|_2^2 \tag{70}$$

This optimal average loss is monotonously decreasing with increasing number of measurements, because adding a measurement to the combination can only improve the optimal performance.

Using the monotonicity properties, an efficient algorithm can be developed for selecting the best subset of measurements. Now denote by X_{all} the index set for all available measurements, and let X_n be any set corresponding to a selection of n measurements. The optimal index set for the best CV containing n variables is

$$X_n^* = \arg \min_{X_n \subset X_{all}} L_{av}(X_n) \tag{71}$$

The monotonicity property implies that if one index set X_n is contained in another index set X_m ($X_n \subset X_m$) then the optimal loss of the superset must be less or equal to the optimal loss of the subset:

$$L(X_m) \leq L(X_n). \tag{72}$$

Assuming that B is a known upper bound to the minimum loss for the case when n variables are selected, such that

$$B \geq L(X_n), \tag{73}$$

then this upper bound can be used to exclude (prune) sets of measurements. In the case where $X_n \subset X_m$, whenever we find

$$L(X_m) > B \tag{74}$$

such that

$$L(X_m) > L(X_n), \tag{75}$$

all n-index subsets included in X_m can be excluded from evaluation, because they will not lead to a lower optimal loss. Equation (74) is called "pruning condition". This pruning condition can be used in an algorithm for systematically eliminating suboptimal measurement subsets from consideration.

We briefly describe such an algorithm for the case where we select 2 out of 6 measurements, for a more general description, we refer to Cao and Kariwala (2008). If we were to evaluate all possibilities for selecting 2 out of 6 measurements, we would need to evaluate $C_6^2 = 15$ combinations of measurements. Using the pruning condition (74), we can organize the search for the best combination in a tree structure, see Figure 3. Each node corresponds to a measurement that is eliminated from further consideration, because including it will not reduce the loss along a path.

At the (initial) root node in Figure 3, the indices corresponding to all six measurements are included, and the loss is evaluated. This gives the best current upper bound. Next, a branch

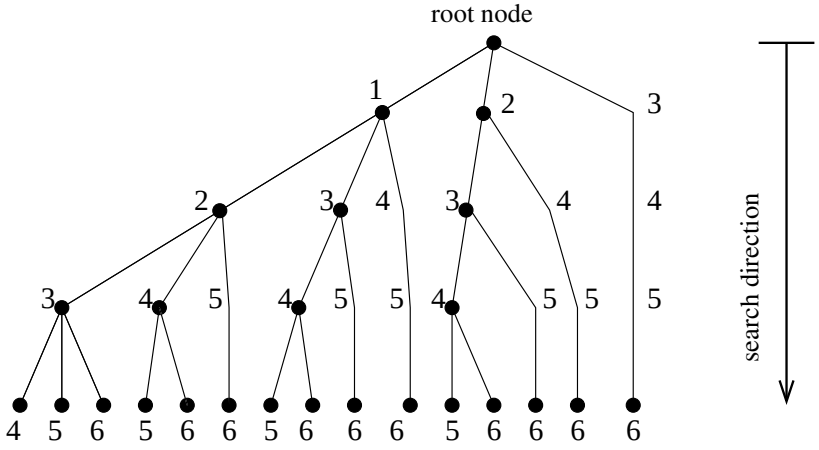


Figure 3: Branch and Bound solution tree for selecting 2 out of 6 measurements (Adapted from Cao and Kariwala (2008)). The number beside each node denotes the measurement which is discarded for that particular node and all nodes below.

is selected, and we move down the branch to the next node on that branch and exclude the corresponding measurement. If the loss of this node is higher than the loss of any other node on the same level, this node and all its sub-nodes may be pruned, because the loss on the lower nodes (where more measurements are excluded) must be higher or equal to the loss of the current node. Then one may proceed along the same branch, or alternatively select another branch either to obtain a new lower bound, or to prune the branch. This procedure is repeated until the optimal measurement set is identified at the leaf of the final remaining branch.

Example 5 (Branch and Bound). *Consider the case in Figure 3, where we want to select the 2 best measurements out of 6 candidates. On the first level, we evaluate the losses for the cases where we exclude measurement 1, 2, and 3. Assume that the loss corresponding to the node where y_1 is excluded is $L(X = \{2, 3, 4, 5, 6\}) = 100$\$, and the loss with y_2 excluded $L(X = \{1, 3, 4, 5, 6\}) = 59$\$, and y_3 excluded $L(X = \{1, 2, 4, 5, 6\}) = 25$\$. In this case we can immediately disregard (prune) the branches corresponding to the two left nodes (where y_1 and y_2, respectively, are excluded). Considering the remaining branch, it is seen that only two measurements remain at its leaf, $X = \{1, 2\}$. So selecting a combination of y_1 and y_2 will give the best performance. Thus we have efficiently found the optimal solution while evaluating only 3 of all 15 possible combinations.*

More sophisticated bidirectional branch and bound methods have been developed, which prune both from the top and the bottom of the tree, thus becoming even more efficient (Kariwala and Cao, 2009, 2010). Free software implementations for these branch and bound methods can be downloaded from the internet[5].

4.6.2. MIQP formulation

Alternatively, Yelchuru and Skogestad (2012) propose to find the best measurement subset by modifying the quadratic problem (63) to a mixed integer quadratic problem (MIQP), which

[5]http://www.mathworks.com/matlabcentral/fileexchange/?term=authorid:22524

can be solved by standard MIQP solvers. The MIQP may be formulated by rewriting

$$H = \begin{bmatrix} h_{1,1} & \cdots & h_{1,n_y} \\ \vdots & \ddots & \vdots \\ h_{n_u,1} & \cdots & h_{n_u,n_y} \end{bmatrix} \tag{76}$$

as a vector by stacking the rows of H in a column vector

$$h_\delta = \begin{bmatrix} h_{1,1} \cdots h_{1,n_y}, & h_{2,1} \cdots h_{2,n_y}, & \cdots & h_{n_u,1} \cdots h_{n_u,n_y} \end{bmatrix}^T \tag{77}$$

with dimension $\mathbb{R}^{n_u n_y \times 1}$. By restructuring the other matrices in a similar fashion, an equivalent vectorized version of problem (63) can be stated as:

$$\begin{aligned} &\min_{h_\delta} h_\delta^T Y_\delta h_\delta \\ &\text{s.t. } G_\delta^{yT} h_\delta = j_\delta \end{aligned} \tag{78}$$

where G_δ^y, j_δ, and Y_δ are the corresponding re-structured versions of G^y, J_{uu} and Y, respectively.

To modify problem (78) for selecting the best subset of measurements, we use the property that controlling a subset of measurements is equivalent to setting all columns associated with excluded measurements to zero. Therefore the problem of selecting only a subset of all available measurements can be formulated as a mixed integer quadratic program, where a constraint is included that enforces the usage of a given number of measurements, while letting the optimizer choose which ones to include in the set. In Yelchuru and Skogestad (2012) these are implemented as "big-M" constraints. Here, a vector of binary variables is defined as

$$\sigma = [\sigma_1 \quad \sigma_2 \quad \cdots \quad \sigma_{n_y}], \quad \sigma_j = \{0, 1\}, \tag{79}$$

where $\sigma_j = 1$ corresponds to a measurement that is included in the measurement combination (nonzero weight in the corresponding elements in H), while $\sigma_j = 0$ corresponds to variables that are not included in the measurement combination, and have a zero element in H. The constraints on the binary variables can be written as

$$P\sigma = s, \tag{80}$$

where $P = \underline{1}_{1\times n_y}^T$ is a n_y dimensional vector of ones, and s is the number of measurements that we want to include in the measurement combination. Then the problem of selecting the optimal measurement subset can be written as

$$\begin{aligned} \min_{h_\delta,\sigma} \quad & h_\delta^T Y_\delta h_\delta \\ \text{s.t. } & G_\delta^{yT} h_\delta = j_\delta \\ & P\sigma = s \\ & -M\sigma_j \le h_{i,j} \le M\sigma_j \quad j = 1 \ldots n_y,\ i = 1 \ldots n_u \\ & \sigma \in \{0, 1\}. \end{aligned} \tag{81}$$

Here, $M \in \mathbb{R}_+^{n_u}$ is a vector of positive constants which are used in the big-M constraints for ensuring that whenever σ_j is zero, the corresponding elements in the H matrix are zero, too.

The values in the M vector are upper bounds on the elements in H. Selecting appropriate values for the elements in M is not straightforward, because a too large value causes the computational load to become very high, while a too small value may result in a falsely active constraint, and a suboptimal solution.

Solving (81), results in a H matrix, which minimizes the average and worst-case local loss, (38) and (42), respectively. This is guaranteed by the constraint $G_\delta^y h_\delta = j_\delta$, which ensures that $J_{uu}^{1/2}(HG^y)^{-1} = I$. The main advantage of casting the problem as MIQP is that it allows for usage of standard MIQP solvers, such as e.g. CPLEX (International Business Machines, 2014) or Gurobi (Gurobi Optimization, 2015).

Remark 7. *From a mathematical point of view it is not required to write the MIQP* (81) *in terms of vectors, as shown above. It may as well have been stated in terms of matrices. However, for numerical software it is often convenient to have the optimization problem in the form given above.*

Remark 8. *The best subset of measurements may not be the same for worst-case and average loss minimization. While the tailor-made branch and bound approaches can find the best measurement subset for worst-case and average loss minimization, the MIQP method will only find the measurement set that minimizes the average loss. However, for most cases it is recommended to minimize the average loss, as the worst-case may not occur very often in practice.*

Example 6 (Selecting subset of measurements). *We consider the problem from Example 1 with $W_d = 1$ and $W_n = I_4$, and find the best measurement combinations for the case where we include 1,2,3,4 measurements in H. Figure 4 shows the value of the optimal loss that can be achieved when including the different numbers of measurements into the CV. We observe that the loss is reduced significantly when increasing the number of measurements from 1 to 2, while adding more than 2 measurements does not have a significant impact on the loss. In this case controlling a measurement combination of more than 2 measurements would basically only add capital cost without improving operation.*

This behavior, that the loss decreases very much initially and then flattens out, is found frequently in chemical processes, where there are typically many measurements that could potentially be included in H.

4.7. Structured H

The problem described in the previous section, where we search for the best measurement subsets to include in H can be considered a special case of a more general problem, where we impose a certain structure on the measurement combination matrix H. For example, it may be undesirable to combine measurements that are located far away from each other, in order to avoid combining variables with very different dynamics and long time-delay. Also, combining e.g. similar measurements (e.g. 3 pressure measurements) may give CVs that have an intuitive physical meaning, and thus increased acceptance among operators.

In such cases a structure is imposed on H. For example, when looking for the best single measurements to control, we require that there is only one non-zero element in each row of H, similarly, if we want to find a subset containing 5 measurements, this corresponds to the requirement of having 5 non-zero columns in H.

The special type of structural constraints for picking optimal subsets of measurements (Section 4.6) have in common that the solution (when the optimal measurements to be included are

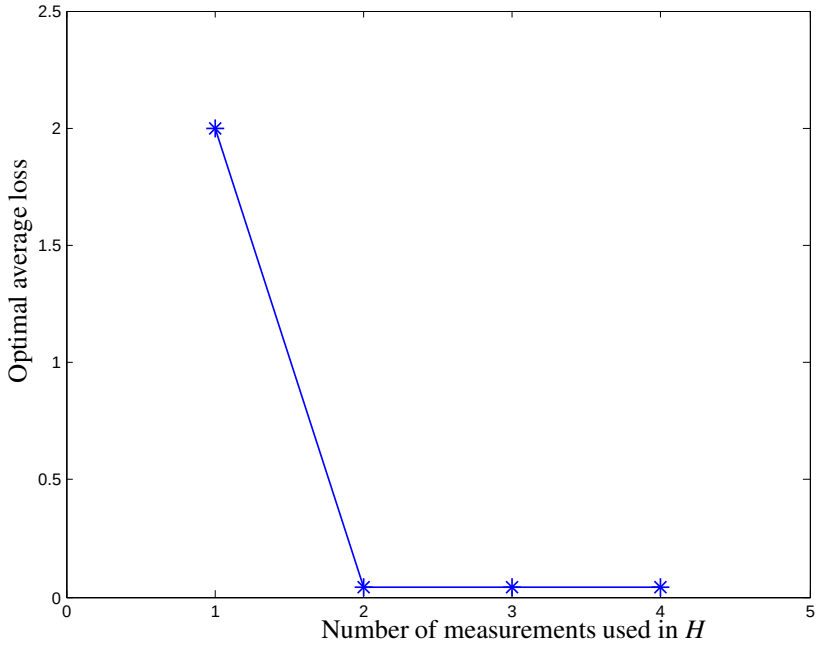


Figure 4: Optimal average loss value that is obtained for the best subsets of measurements

known) can be re-formulated in terms of (63). Equivalently, the optimal H (with the imposed structure) does not change structure when pre-multiplied by any invertible matrix Q. In particular, any zero column in H will remain zero upon multiplication with Q. Therefore solving the MIQP with convex cost, (81) gives the H that also minimizes the exact non-convex loss expression $L = \frac{1}{2}\left\|J_{uu}^{1/2}(HG^y)^{-1}HY\right\|_F^2$. Similarly, the monotonicity property which states that the optimal loss cannot decrease when adding more measurements is preserved in that case.

However, when more general structural constraints are required, such that H has e.g. a block-diagonal structure, the problem becomes more complicated. In this case, pre-multiplying H with any non-singular matrix Q does not preserve the structure of H, so minimizing the convex re-formulation will not result in a solution that minimizes the loss expression $L = \frac{1}{2}\left\|J_{uu}^{1/2}(HG^y)^{-1}HY\right\|_F^2$. Also note that for such general constraints the monotonicity property that is used in the branch and bound algorithm, may not hold any longer. For a collection of structural constraints that can be handled with currently available methods, we refer to Heldt (2010), and Yelchuru and Skogestad (2012).

Currently there are no simple and tractable methods for handling cases with structural constraints that are not preserved upon pre-multiplication with a non-singular matrix, and that do not satisfy the monotonicity property required for the branch and bound algorithm. These cases result in mixed integer optimization problems with non-convex subproblems for every possible variable combination.

4.8. Conclusion

The local methods presented in this section are all based on a linearization around the nominally optimal operating point. They are subject to the following main limitations:

1. The results are only guaranteed to be valid in a vicinity of the nominal point
2. The disturbances must not change the set of active constrains
3. Only structures that are preserved upon pre-multiplication by any non-singular matrix can be imposed on H.

The methods only give candidate sets of CVs, which have to be evaluated and tested over the whole region using the nonlinear model. However, the advantage of these local methods is that they can be used for efficient screening of candidate sets of CVs, because CVs that give a large loss close to the expected (nominal) operating point can be excluded immediately from further consideration. Moreover the local methods can handle measurement noise in a systematic manner.

The problem of selecting a subset of measurements to include in the CVs, leads to a difficult combinatorial optimization problem. Simple structural constraints (e.g. selecting the optimal subset of measurements), can be addressed using off-the shelf MIQP solvers and tailor-made branch and bound methods. However, more complicated structures that are not preserved upon pre-multiplication with a non-singular matrix remain an open problem.

5. Constraint handling

The methods described so far assume that the disturbances do not change active constraints. However, in general the disturbances can cause the active set to change, and the plant may be operated in different regions that are defined by the active constraints. Recently some approaches have been proposed in the literature for overcoming this problem. In this section we present the multi-parametric programming approach (Manum and Skogestad, 2012), the integrated approach (Hu et al., 2012b), and the cascade control approach (Cao, 2004).

5.1. Parametric programming approach

In this method, every active constraint region has its separate control structure. The main idea is that controlling self-optimizing CVs is equivalent or similar to controlling the optimality conditions in the regions. On the boundary between two regions the optimality conditions of both adjacent regions are satisfied, and the CVs and constraints of both regions assume their optimal values. Thus, whenever a constraint or self-optimizing variable of a neighbouring region assumes its optimal value, we switch the control structure to the corresponding region.

This idea was used in Jäschke and Skogestad (2012a), and was further refined to require only monitoring certain carefully designed scalar descriptor functions in each region (Manum and Skogestad, 2012) Below, we give a brief outline of the ideas, and for a more detailed presentation, we refer to the paper by Manum and Skogestad (2012).

In this method it is assumed that there is no measurement noise $n^y = 0$. Starting from the reduced problem (11), the cost function $\bar{J}^r$ is approximated by a quadratic cost function, as in (13), and the constraints $g(\bar{u}, d) \leq 0$ are linearized:

$$\begin{aligned} \min_{\bar{u}} \begin{bmatrix} J^r_{\bar{u}} & J^r_d \end{bmatrix} \begin{bmatrix} \Delta\bar{u} \\ \Delta d \end{bmatrix} + \frac{1}{2}[\Delta\bar{u}^T\ \Delta d^T] \begin{bmatrix} J^r_{\bar{u}\bar{u}} & J^r_{\bar{u}d} \\ J^r_{d\bar{u}} & J^r_{dd} \end{bmatrix} \begin{bmatrix} \Delta\bar{u} \\ \Delta d \end{bmatrix} \\ \text{s.t.} \quad \Delta g = A \begin{bmatrix} \Delta\bar{u} \\ \Delta d \end{bmatrix} \leq 0. \end{aligned} \tag{82}$$

Here $\Delta g = g - g_{nom}$ is the deviation of constraint from the nominal (linearization) point, and $A = \left[\begin{array}{cc} \frac{\partial g}{\partial \bar{u}} & \frac{\partial g}{\partial d}. \end{array} \right]$ denotes the constraint Jacobian. It is assumed that the constraints Δg, can be measured and controlled using the available degrees of freedom.

For each set i of active constraints, one can find (by eliminating the active constraints for this region and using the null-space method) a CV $\Delta c^{(i)} = H^{(i)} \Delta y$, such that the optimality conditions can be written as[6]

$$\Delta c^{(i)} = H^{(i)} \Delta y = 0 \tag{83}$$

$$\Delta g^{(i)} = A^{(i)} \left[\begin{array}{c} \Delta u \\ \Delta d \end{array} \right] = 0. \tag{84}$$

Here $A^{(i)}$ denotes the rows of A that correspond to the active set in region i. Combining the vectors Δy and Δg into an augmented measurement vector

$$\Delta y^{aug} = [\Delta y, \ \Delta g]^T, \tag{85}$$

we can write the optimality conditions as

$$\Delta c^{(i),aug} = H^{(i),aug} \Delta y^{aug} = 0, \tag{86}$$

where

$$H^{(i),aug} = \left[\begin{array}{cc} H^{(i)} & 0 \\ 0 & \alpha^{(i)} \end{array} \right]. \tag{87}$$

Here $\alpha^{(i)}$ is a diagonal matrix with the $\alpha^{(i)}_{j,j} = 1$ if Δg_j is an active constraint, and zero otherwise. In every region, optimal operation can be achieved by controlling $\Delta c^{(i),aug} = H^{(i),aug} \Delta y^{aug}$ to zero, that is, by controlling the corresponding CVs and constraints at their optimal values.

Starting by controlling $\Delta c^{(i),aug} = 0$ in region i, a naive approach for detecting active set changes is to simply monitor $\Delta c^{(k),aug}$ in all the neighboring regions $k \neq i$. When a disturbance moves the process from region i to region k, we have $\Delta c^{(i),aug} = \Delta c^{(k),aug} = 0$ at the boundary (i.e. the optimality conditions of region i and k are both satisfied), and the control structure must be changed to region k. This requires to monitor $n_{\bar{u}}(n_{neighbours,i})$ scalar variable values in each region i, where $n_{\bar{u}}$ is the total number of degrees of freedom including active constraints, and $n_{neighbours,i}$ is the number of regions neighbouring region i. This number may become quite large for real systems.

To reduce the number of variables to be monitored Manum and Skogestad (2012) proposed to construct a scalar descriptor function (Baotić et al., 2008) in each region. This function is monitored in each neighboring region, and active set changes are identified by comparing its value in the current region with the values of the descriptor functions of the neighbouring regions. This reduces the number of variables to be monitored from $n_{\bar{u}}(n_{neighbors})$ to $n_{neighbors} + 1$, which can be a significant reduction.

The scalar valued descriptor function $f_i(y)$ for each region i is constructed by pre-multiplying the CVs by a non-zero constant vector $w^{(i)} \in \mathbb{R}^{n_u}$, such that

$$f_i(y) = {w^{(i)}}^T H^{(i),aug} \Delta y^{aug}. \tag{88}$$

[6]Note that controlling Δc to zero corresponds to controlling the reduced gradient for that region to zero.

Baotić et al. (2008) show that any vector $w^{(i)}$ may be used for constructing a scalar piecewise affine descriptor function, as long as it is not in the left null-space of $(H^{(i),aug} - H^{(j),aug})$, where i and j denote two neighboring regions. However, from a numerical point of view, one would desire vectors w which are robust to small errors, and an algorithm for systematically constructing such a vector is presented in Manum and Skogestad (2012); Baotić et al. (2008).

Whenever the difference between the value of the descriptor function of the current region and a neighbouring region changes sign, an active set change occurs, and the control structure is switched to the corresponding neighbouring region. An example for such a descriptor function is given in Fig. 5. Assume that the process is initially operated in Region 3, where $f_3(y)$ takes values

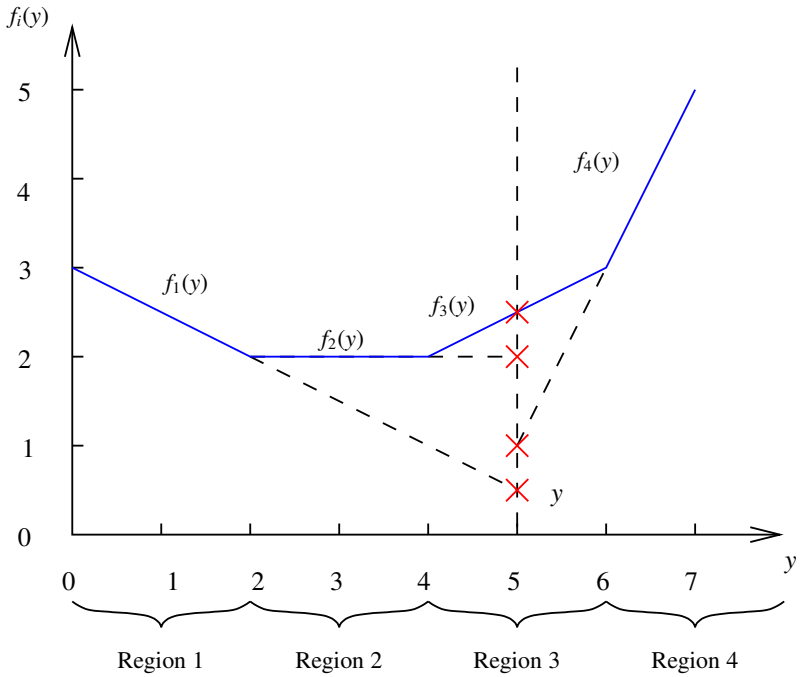


Figure 5: Illustration for changing regions (Figure adapted from Baotić et al. (2008))

within the interval $[2, 3]$. It can be seen that the function values of the neighbouring functions $f_2(y) = 2$ and $f_4(y) = 2y - 9$ evaluated in Region 3 are below the function value of $f_3(y)$. Thus we have $\text{sign}(f_3 - f_4) = 1$, and $\text{sign}(f_3 - f_2) = 1$. When moving to Region 2, the sign of $f_3(y) - f_2(y)$ changes to -1, while the sign of $f_4(y) - f_2(y)$ remains unchanged at 1. Thus, by monitoring the sign of the differences in the descriptor functions, the new region to switch to can be detected.

5.2. Integrated approach

The parametric programming approach described above can result in complicated control systems with many different CVs (one for each region) and corresponding switching laws. Especially when the number of regions is large, the complexity of this approach can render it practically infeasible. In contrast, the integrated approach aims at finding a single control structure,

which makes sure that all variables remain in the pre-specified bounds, regardless of the value of the disturbance or the noise. Although there will be a loss due to not satisfying active constraints, the idea is that the simplicity of the control structure outweighs this disadvantage.

In the local integrated approach proposed by Hu et al. (2012b) it is assumed that at the nominal (linearization) point no constraints are active, or that the disturbances do not cause any nominally active constraint to become inactive. A new constraint variable $z \in \mathbb{R}^{n_g}$ is introduced that corresponds to the value of the inequality constraint in (1), so that

$$z = g_z(z, d) \leq 0, \tag{89}$$

where z may also contain inputs u and states x. Linearizing $g_z(z, d)$ around the nominal operating point yields

$$\Delta z = G^z \Delta u + G_d^z \Delta d \leq 0. \tag{90}$$

From equations (21)-(23), keeping $\Delta c = H\Delta y$ at a constant setpoint ($\Delta c = 0$) we obtain upon solving for u

$$\Delta u = -(HG^y)^{-1} H [G_d^y W_d \quad W_n] \begin{bmatrix} d' \\ n' \end{bmatrix}. \tag{91}$$

Inserting (91) into (90) yields

$$\Delta z = \left(-G^z (HG^y)^{-1} H \left[G_d^y W_d \;\; W_n\right] + G_d^z \left[W_d \, 0\right]\right) \begin{bmatrix} d' \\ n' \end{bmatrix} \leq 0. \tag{92}$$

We observe that here again, the term $(HG^y)^{-1}H$ appears in (92). Thus, following the argument in Section 4.3, pre-multiplying H with any invertible matrix Q will not change the value of Δz. Using this information, Q may be selected to make $(HG^y)^{-1} = I$. Then we define

$$B = -G^z H \left[G_d^y W_d \quad W_n\right] + G_d^z \left[W_d \; 0\right]. \tag{93}$$

Under the assumption of a uniform distribution of $[d' \quad n']$ the elements of z assume their largest values when the elements of $[d' \quad n']$ are at their bounds, i.e. either +1 or -1. Thus, the constraints (92) can be rewritten as

$$\|B_i\|_1 \leq 0 \quad i = 1 \ldots n_g \tag{94}$$

where B_i and b_i denote the i-th row of B and b, respectively.

The problem of finding H that minimizes the average loss $L_{av} = \frac{1}{2} \left\| J_{uu}^{1/2} (HG^y)^{-1} HY \right\|_F^2$ can be formulated as:

$$\begin{aligned} \min_H \; & \frac{1}{2} \left\| J_{uu}^{1/2} HY \right\|_F^2 \\ & \text{s.t.} \\ & HG^y = I \\ & \left\| \underbrace{\left(-G^z H \left[G_d^y W_d \quad W_n\right] + G_d^z \left[W_d \; 0\right]\right)_i}_{= B_i} \right\|_1 \leq 0, \quad i = 1..n_g \end{aligned} \tag{95}$$

This is a convex problem which is either infeasible, (then there is no variable combination that satisfies the constraint over the whole region), or has a unique solution, and the CV that minimizes the loss without violating the constraints is $\Delta c = H\Delta y$.

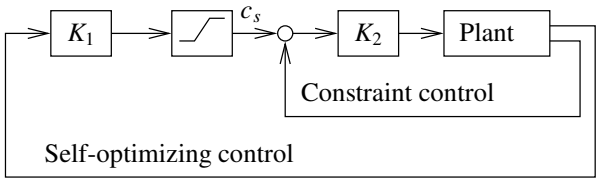


Figure 6: Cascade structure for handling active set changes (Adapted from Cao (2004))

Remark 9. *The problem* (95) *does not minimize the loss exactly when constraints that are nominally active become inactive. This is because J_{uu} depends on the active set of the linearization point, and the curvature will change when constraints become active or inactive. However, when there are no active constraints at the linearization point (or they do not become inactive), this approach can be expected to represent the true local loss well.*

Remark 10. *Note that in* (95) *we have a slightly different cost function as previously in the convex formulation* (63)*. That is, because in* (63) *the additional degree of freedom was used to make $HG^y = J_{uu}^{1/2}$. In the case presented in this section, we use the additional degree of freedom for setting $HG^y = I$. Therefore the J_{uu} term does not cancel, and J_{uu} remains in the objective of the optimization problem for finding H.*

5.3. Cascade control approach

The cascade control approach by Cao (2004, 2005) is an option when only few constraints can become active. It is based on the assumption that the process is operated in an unconstrained region most of the time, and that the number of constraints that can become active, is lower or equal to the number of CVs. The control structure is implemented in a cascade fashion, Fig. 6. The setpoint c_s of the inner loop acts as manipulated variable for the outer loop. As long as the constraint does not become active the outer loop with controller K_1 will manipulate the setpoint of the inner loop such that a self-optimizing CV is kept at its setpoint. However, when a new constraint becomes active, the saturation block will limit the setpoint of the inner loop such that the constraint is not violated.

Since the self-optimizing control variables are selected for the unconstrained nominal case, this approach will lead to losses when the constraints become active. If the process is operated in the nominal region most of the time, then the loss may be negligible. However, when the set of active constraints is likely to change frequently, other approaches, like the integrated approach or the parametric programming approach are better suited.

5.4. Constraint matching

Approximating nonlinear plant constraints linearly, as done in (82) and (90) can lead to feasibility problems, because the nonlinearity of the plant is not taken into account for the control structure selection (Manum and Skogestad, 2012). One way of addressing this problem is to measure the real constraint values, and to treat the distance to the constraints as a measured disturbance. In effect, this adapts the constraints in the model to match the real bounds.

5.5. Conclusion

Handling changing sets of active constraints is still one of the difficult issues when using self-optimizing control. The cascade and the multiparametric programming approaches are limited in applicability when the number of constraint regions is large. The number of potential regions grows exponentially with the number of constraints, and although the regions may be tracked by scalar functions, it can easily become infeasible in practice. The cascade approach, although very simple and pragmatic, is limited because the number of constraints, which can become active, has to be less than the number of CVs.

The integrated approach, because of its simplicity, is very much in the spirit of self-optimizing control. However, it may be impossible to find a single control structure, which is feasible for all disturbances. If a feasible H matrix can be found, the corresponding loss can be used to judge if it is indeed self-optimizing. If this is not the case, then one may divide the disturbance space, and design two separate control structures. This can be done until the loss is acceptable. A similar procedure can be used when the overall H is infeasible.

6. Self-optimizing CVs for larger operating windows

Most approaches reviewed so far are based on linearization around a nominally optimal operating point, and result in locally optimal control structures. If disturbances cause the plant to operate far away from the nominal point, the resulting loss of a local self-optimizing control scheme may become very large. CVs that perform well only locally may not be desirable in this case. Below we summarize some of the approaches that address this issue.

6.1. Polynomial zero loss-method

Using CVs that are linear combinations of measurements may result in an unacceptable loss because of strong curvature at the optimum. To handle these cases, the null-space method has been generalized by Jäschke and Skogestad (2012a) to systems described by polynomial equations. This method gives optimal CVs which are independent of a linearization point, and gives zero loss under the assumption of no measurement noise and no active constraint changes. As the method does not rely on a selected linearization point, this may be considered a "global" approach.

The point of origin is the steady-state optimization problem (1). Assuming that the active set does not change, the problem can be written as an equality constrained optimization problem

$$\begin{aligned} &\min \bar{J}(\bar{u}, x, d) \\ &\text{s.t. } g(\bar{u}, x, d) = 0, \end{aligned} \tag{96}$$

together with the corresponding measurement relations (assuming zero noise)

$$y = m(\bar{u}, x, d). \tag{97}$$

Furthermore, we assume that $\bar{J}(\bar{u}, x, d)$ and $g(\bar{u}, x, d)$ are polynomials in the polynomial ring $\mathbb{R}[x, d]$. Loosely speaking this can be considered as polynomials in the variables x and d with coefficients in $\mathbb{R}$. Moreover, it is assumed that Problem (96) has an optimal solution in the region, and the linear independence constraint qualifications (LICQ) and sufficient secondary

conditions for optimality (Nocedal and Wright, 2006) are satisfied in the region. Then the first order necessary optimality conditions are

$$\nabla \bar{J}(\bar{u}, x, d) + \nabla g(\bar{u}, x, d)^T \lambda \quad = 0 \tag{98}$$

$$g(\bar{u}, x, d) \quad = 0, \tag{99}$$

where the variables $\lambda \in \mathbb{R}^{n_g}$ denote the Lagrangian multipliers, and $\nabla = \frac{\partial}{\partial \left[\begin{smallmatrix} u \\ x \end{smallmatrix}\right]}$.

For obtaining the CVs, the Lagrange multipliers λ may be eliminated analytically by pre-multiplying (98) by a matrix $N(\bar{u}, x, d) \in \mathbb{R}^{n_{\bar{u}} \times (n_{\bar{u}} - n_g)}$ which is in the null-space of $\nabla g(\bar{u}, x, d)$, such that $N(\bar{u}, x, d)^T \nabla g(\bar{u}, x, d)^T = 0$. Then (98) becomes

$$r_{red} = N(\bar{u}, x, d)^T \nabla \bar{J}(\bar{u}, x, d) = 0. \tag{100}$$

r_{red} is called the reduced gradient. Unlike in the null-space method, the matrix $N(\bar{u}, x, d)$ is not constant, it is a function of the operating point $\bar{u}, x, d$, and has to be calculated analytically.

Having eliminated λ, it remains to eliminate the unknown states x and disturbances d. It is shown in Jäschke and Skogestad (2012a) that if

$$\begin{aligned} g(\bar{u}, x, d) &= 0 \\ y - m(\bar{u}, x, d) &= 0 \end{aligned} \tag{101}$$

has a finite number of solutions when considered as equations in the variables x and d, and $d_i \neq 0$ and $x_j \neq 0$, then it is possible to find functions $\mathcal{R}_k(\bar{u}, y) \in \mathbb{R}$, such that for $k = 1..n_u$ we have

$$\mathcal{R}_k(\bar{u}, y) = 0 \Leftrightarrow \begin{cases} [N(\bar{u}, x, d)^T \nabla J(\bar{u}, x, d)]_k & = 0 \\ g(\bar{u}, x, d) & = 0 \\ y - m(\bar{u}, x, d) & = 0. \end{cases} \tag{102}$$

Here, the notation $[\cdot]_k$ denotes the k-th element of a vector. In words, it is possible to eliminate the disturbances d and internal states x from the reduced gradient without explicitly solving (101) for x and d. The function $\mathcal{R}_k(\bar{u}, y)$ is called sparse resultant or toric resultant (Cox et al., 2005) of the system (102). It can be calculated using computer algebra software. A freely available software package for Maple (Waterloo Maple Inc., 2014) is described in Busé and Mourrain (2003).

The polynomial zero loss method gives a CV, which is zero if and only if the reduced gradient is zero. Since all computations are analytical, the resulting CV expressions can become very complex, rendering them useless in practice. However, in some practical cases they have a very simple structure (Jäschke and Skogestad, 2012b; Jäschke and Skogestad, 2014).

Example 7. *We consider again the cost function $J = \frac{1}{2}(u-d)^2$, but this time we assume nonlinear measurements:*

$$\begin{aligned} y_1 &= \frac{d^2 + u}{d} \\ y_2 &= ud, \end{aligned} \tag{103}$$

For this case, the gradient is $J_u = u - d$, and we can write the measurements as polynomials $y_1 d - d^2 - u = 0$ and $y_2 - ud = 0$. Applying the above method to eliminate u and d we obtain the CV

$$c = y_2 + 2y_1 - y_1^2 - 1. \tag{104}$$

The reader may verify that for a given $d \neq 0$, controlling $c = 0$ leads to $u = d$, which makes the gradient zero, $J_u = 0$.

6.2. Regression approach

Another approach for finding self-optimizing CVs was proposed by Ye et al. (2013a). This approach can be considered a surrogate model approach, where first a model is used to generate "gradient measurements" for a wide range of disturbances, and then a linear or polynomial function is fitted to this data. If a good fit is obtained, this fitted function will approximate the gradient well over all the operating envelope, and may be used as a self-optimizing CV.

In Section 4 it was shown that under the assumption that the correct active set is known and does not change, the local loss can be expressed as the weighted norm of the gradient (20), where the weighting factor is the inverse square root of the reduced Hessian, $J_{uu}^{-1/2}$. By using a controller with integral action, the CV will be controlled to its setpoint $h(y) = c_s$, or equivalently $h(y) - c_s = 0$. The loss for a given operation point is then given by (Jäschke and Skogestad, 2011)

$$L = \frac{1}{2} \left\| J_{uu}^{-1/2} \left(J_u - (h(y) - c_s) \right) \right\|_2^2 . \tag{105}$$

This expression is used in the regression method proposed by Ye et al. (2013a), which consists of the following steps:

1. The process is simulated for N different values of inputs u, disturbances d, and noise n^y over the whole operating range. The sample points should be chosen such that they are representative of the noise and disturbance distribution. At each of these $i = 1, \ldots, N$ sample points, the gradient $J_u^{(i)}$ and the corresponding measurement values $y^{(i)}$ are recorded together with the reduced Hessians $J_{uu}^{(i)}$.
2. The sampled gradient values are used to fit a regression model of form $\hat{J}_u = h(y)$, which describes the relationship between the measurements and the gradient. In its simplest form $h(y)$ is linear, i.e. $h(y) = Hy$, and the regression procedure calculates the elements in the H-matrix and the optimal setpoint c_s. (Alternatively a nonlinear function such as a polynomial function may be used to capture nonlinear effects.) The objective to be minimized by the regression is based on (105):

$$\phi = \frac{1}{2N} \sum_{i=1}^{N} \left\| J_{uu}^{(i)-1/2} (J_u^{(i)} - (Hy^{(i)} - c_s)) \right\|_2^2 . \tag{106}$$

   The regression gives the matrix H and the setpoints c_s, such that the average loss is minimized.
3. The resulting CV is then an approximation of the gradient, $c = Hy = \hat{J}_u \approx J_u$, which is to be controlled at a constant setpoint c_s.

In order to obtain the model of the gradient in terms of the measurements, Ye et al. (2013a) use least squares regression and a neural network. However, any other regression that gives a good fit may be used as well.

The evaluation of the gradient at the different samples in the operating space can be done analytically, if the model is very simple. In more complicated cases, the gradient must be evaluated using finite differences, or possibly automatic differentiation procedures. Large-scale systems with many inputs and disturbances will require a lot of sampling points. To reduce the amount of data that has to be generated and saved, one may replace J_{uu} in (106) by the identity matrix I. This approximation was found to give good results, while simplifying the procedure (Ye et al., 2013a). However, the challenge with this approach is still that the number of sampling points grows exponentially with number of disturbances and noisy measurements.

In principle, all measurements y could be used in the regression model. However, as it is typically desired not to include more measurements than necessary into the CV, the regression based approach can also be combined with a branch and bound method (Kariwala et al., 2013) for finding best subsets of measurements. The idea of fitting a measurement function to the optimality conditions has also been applied for finding CVs for a dynamic optimization problem (Ye et al., 2013b).

6.3. Controlled variable adaptation

In the regression based approach from Section 6.2, a simple regression function is generally preferable in order to avoid over-fitting. Therefore, a certain level of regression errors is inevitable, which may make the operational range with an acceptable economic loss relatively narrow even though the approach is "global". To address this issue, a CV adaptation scheme was proposed by Ye et al. (2014), where the CVs are adjusted depending on where the process is operated.

In this scheme, a plant model is simulated off-line on sample points distributed over the entire operating region. The resulting measurements, manipulated variables and gradient variables are collected and stored in a database. Based on the collected data set, a non-optimality monitoring model is built up using statistical process monitoring approaches, where the non-optimality status is treated as a special process "fault". When applying this scheme online, the new measurements are first checked by the monitoring model to determine whether current operation is optimal. If a non-optimal status is identified, the current measurements are then used to find a subset of neighbourhood sampling points in the database to apply the so called "just-in-time" regression. The resulting CVs together with their setpoints determined as coefficients of regression functions are then applied to the control system for CV adaptation. Numerical case studies show this scheme can significantly reduce economic loss in a wide operational range even with simple linear CVs of a few measurements. However, the controllers may have to be re-tuned in order to obtain a good performance over the whole operating region.

6.4. Global approximation of controlled variables

Recently, another approach for solving the self-optimizing control problem over the whole operation range was proposed by Ye et al. (2015a). The idea is to separate the loss contribution due to disturbances L_d, and the contribution due to noise L_n, such that the total loss is $L = L_d + L_n$. As in the regression approach, the disturbance space is sampled, and for each disturbance realization $d^{(i)}$, the values of $y^{*(i)}$, $J_{uu}^{(i)}$ and $G^{y(i)}$ are stored.

Then an optimization problem is formulated to minimize the average loss over all disturbance realizations $i = 1 \ldots, N$,

$$\begin{aligned} \min_{H} \; & \frac{1}{N} \sum_{i=1}^{N} (L_d^{(i)} + L_n^{(i)}) \\ \text{s.t.} \quad & HG_{nom}^{y} = J_{uu,nom}, \end{aligned} \tag{107}$$

where $L_d^{(i)} = \frac{1}{2} y^{*(i)T} H^T J_{cc}^{(i)} H y^{*(i)}$ denotes the loss due to disturbance and $L_n^{(i)} = \frac{1}{2} \text{trace}(W^2 H^T J_{cc}^{(i)} H)$ the loss resulting from measurement noise. For independent noise $W = E(nn^T)$ is a diagonal matrix, and $J_{cc}^{(i)}$ is defined as (Halvorsen et al., 2003)

$$J_{cc}^{(i)} = (HG^{y,(i)})^{-1} J_{uu}^{(i)} (HG^{y,(i)})^{-1}. \tag{108}$$

Further, G^y_{nom} and $J_{uu,nom}$ are the gain and the Reduced Hessian corresponding to a chosen nominal operating point.

However, even with the simplified loss evaluation derived above, to solve (107) the optimal CV as linear combinations of measurements is still not easy. The optimization problem is non-convex due to the dependence of J_{cc} on the combination matrix, H as shown in (108). In addition to a direct optimization approach, Ye et al. proposed a short-cut approach by fixing J_{cc} at a nominal point. It is interesting to note that the short-cut approach leads to an analytic solution of H similar to the local approaches, although matrices involved in the global solution have to be constructed based optimization data obtained from the whole operation space. For further details of this derivation, readers are referred to the original paper by Ye et al. (2015a).

6.5. Conclusion

There have been a number of attempts to find CVs that give good operation over a wide operating range. The polynomial zero loss method is a generalization of the null-space method, that is capable of handling higher order curvature in the system. However, it is only suitable for small problems, as the complexity of the measurement combination tends to grow with variable number and degree of the polynomials. The complexity of the invariant is difficult to know a-priori. Depending on the problem structure, it may be very simple, or very complicated. Especially when high order terms are present in the resultant, the CV may become very sensitive to model error and measurement noise.

The regression based approach makes it possible to design CVs that are obtained from model data from the whole operating range, as long as the active constraints do not change. Moreover, it allows for finding polynomial measurement combinations, that are simpler than the exact ones found by the polynomial method described in Section 6.1. Since the data is obtained from offline simulations, it is possible to run it with a large number of sample points, such that a good approximation of the gradient can be found in terms of measurements.

The CV adaptation scheme adopts a just-in-time regression approach on-line in order to enlarge the operational range with acceptable economic loss particularly associated with relatively simple CV functions. It is worth pointing out that in a traditional hierarchical process automation scheme, an optimization layer only updates set-points of feedback control loops in the low layer to achieve real-time optimization. In such a scheme, without adapting the CVs, such updating has to be executed very regularly in order to achieve optimal operation. The CV adaptation scheme, however, may require less frequent updates, because the CVs are designed such that they are self-optimizing around the current operating point, and they are only updated when necessary.

Finally, for finding global self-optimizing CVs it simplifies the calculations significantly to separate the loss contributions due to noise (L_n) and due to disturbances (L_d). Thus only the disturbance space must be sampled, while the noise contribution can be calculated without sampling all noise realizations. This derivation may facilitate further development in this field. Moreover, the similarity of the results obtained by the 'global' approach and the local methods may be worth further investigation.

7. Model-free approaches for self-optimizing control

The previously described methods all rely on an accurate model for finding self-optimizing CVs. This may be considered a drawback, because the cost of modelling can be very high. To address this issue, model-free approaches have been developed. These methods rely on plant

measurement data instead of a-priori model information. Often historical data are readily available, and can be analyzed to study the behavior of the plant. Both approaches described below are local methods that rely on the same assumptions as the developments in Section 4.

7.1. Historical data approach (using non-optimal data)

This approach was proposed by Jäschke and Skogestad (2013), and the idea is to use historical measurement data to find a quadratic model of the operating cost. In particular, historical data of the plant measurements is used to fit parameters J_{nom}, J_y and J_{yy} to obtain the following model of the operating cost

$$J = J_{nom} + J_y \Delta y + \frac{1}{2} \Delta y^T J_{yy} \Delta y. \tag{109}$$

Then, considering the previously derived quadratic approximation of the cost (13),

$$J \approx J_{nom} + \begin{bmatrix} J_u & J_d \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \Delta u^T & \Delta d^T \end{bmatrix} \begin{bmatrix} J_{uu} & J_{ud} \\ J_{du} & J_{dd} \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}, \tag{110}$$

and assuming that the measurement equations (50) are invertible, such that

$$\begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} = \left[\tilde{G}^y\right]^{\dagger} \Delta y, \tag{111}$$

we can eliminate Δu and Δd from (110) and obtain

$$J \approx J_{nom} + \underbrace{\begin{bmatrix} J_u & J_d \end{bmatrix} \left[\tilde{G}^y\right]^{\dagger}}_{=J_y} \Delta y + \frac{1}{2} \Delta y^T \underbrace{\left[\tilde{G}^y\right]^{\dagger T} \begin{bmatrix} J_{uu} & J_{ud} \\ J_{du} & J_{dd} \end{bmatrix} \left[\tilde{G}^y\right]^{\dagger}}_{=J_{yy}} \Delta y. \tag{112}$$

Considering the term J_{yy} more closely it can be seen that

$$J_{yy} = \left[\tilde{G}^y\right]^{\dagger T} \begin{bmatrix} J_{uu} & J_{ud} \\ J_{du} & J_{dd} \end{bmatrix} \left[\tilde{G}^y\right]^{\dagger} \tag{113}$$

$$= \left[\tilde{G}^y\right]^{\dagger T} \begin{bmatrix} \begin{bmatrix} J_{uu} & J_{ud} \end{bmatrix} \tilde{G}^{y\dagger} \\ \begin{bmatrix} J_{du} & J_{dd} \end{bmatrix} \tilde{G}^{y\dagger} \end{bmatrix} \tag{114}$$

$$= \left[\tilde{G}^y\right]^{\dagger T} \begin{bmatrix} H \\ \begin{bmatrix} J_{du} & J_{dd} \end{bmatrix} \tilde{G}^{y\dagger} \end{bmatrix}, \tag{115}$$

where the matrix $H = \begin{bmatrix} J_{uu} & J_{ud} \end{bmatrix} \left[\tilde{G}^y\right]^{\dagger}$ is the same as in the expression for the null-space method (52). Therefore, if the matrix J_{yy} is known, the H-matrix from the null-space method can be recovered by multiplication from the left with G^{yT},

$$H = G^{yT} J_{yy}. \tag{116}$$

The above results can be used for finding CVs from historical data by the following steps:

1. Run plant experiments to obtain G^y.
2. Analyze historical plant data and the plant experiment data to determine J_{nom}, J_y, and J_{yy}.
3. Compute the optimal measurement combination as $H = G^{yT} J_{yy}$.

In general Step 1 will require plant test experiments where the inputs u are changed, and G^y is approximated using finite differences. Step 2 requires to measure and record historical the operation cost J together with all available measurements at different sample times, and to fit the parameters J^{nom}, J_y, and J_{yy} to the measured cost. Also closed loop operation data should be included in the data set for finding J_{yy}, because the relationship between the cost and the measurements will not change when a feedback-loop is closed. However, in order to obtain a good estimate of J_{yy}, it is necessary to also include open-loop test data (for example the data that is used to find G^y) as this also includes directions in y that may not occur when using only closed-loop data. In contrast to the method described in Section 7.2, it is not required that the data is obtained from optimal operation. It is rather desirable that the data comes from many suboptimal points that are not very far from the optimal point. Since only disturbances that are present in the measurement data will be accounted for in the fitted model, it is beneficial to collect the data over a long operation period.

The measurements y are not independent data, therefore a simple least squares regression will give poor results. However, by using partial least squares (PLS) regression which can handle dependent and collinear data, better results can be achieved and the cost function parameters in (109) can be estimated quite well.

Remark 11. *Note the similarity between the expressions* $H = G^{y^T}(YY^T)^{-1}$ *from* (64) *and* $H = G^{y^T} J_{yy}$ *from* (116)*. From this, the matrix* YY^T *may be interpreted as an approximation to the inverse of* J_{yy}*.*

7.2. Optimal data approach

In contrast to the more general approach described in the previous section, the optimal data approach (Jäschke and Skogestad, 2011b) requires measurements from optimal data points. One way to obtain these data points is to apply an extremum seeking method (Krstic and Wang, 2000). All the measurements that has been collected from periods of *optimal operation* at sample times $t_1, t_2, \ldots, t_N$ is collected into a large data matrix

$$Y = \begin{bmatrix} \Delta y(t_1) & \Delta y(t_2) & \ldots & \Delta y(t_N) \end{bmatrix}. \tag{117}$$

The exact cost function or its value need not be known, as long as it is known that the data is optimal. We further assume a linear relationship from the inputs Δu and disturbances Δd to the data as in (21), and that the measurement noise n^y is negligible. Moreover, as in the null-space method, it is assumed that $n_y \geq n_u + n_d$. We then have

$$\begin{aligned} Y &= \begin{bmatrix} \frac{\partial y^{opt}}{\partial d}\Delta d(t_1) & \frac{\partial y^{opt}}{\partial d}\Delta d(t_2) & \cdots & \frac{\partial y^{opt}}{\partial d}\Delta d(t_N) \end{bmatrix} \\ &= F \begin{bmatrix} \Delta d(t_1) & \Delta d(t_2) & \ldots & \Delta d(t_N) \end{bmatrix}, \end{aligned} \tag{118}$$

where $\Delta d(t_1), \ldots, \Delta d(t_N)$ are the disturbances, which where acting on the process when the samples where taken. Since $n_y \geq n_u + n_d$, and we assume optimal data with no measurement noise, we may simply select H in the left null-space of F, which happens to be the null-space of Y.

Because in practice the data will contain some degree of noise, the left null-space of Y will be empty. However, we may find measurement combinations that are only minimally changing under optimal operation, by performing a singular value decomposition $Y = U\Sigma V^T$, and selecting H as the n_u columns corresponding to the smallest singular values. If the optimal data is consistent in the sense that disturbances have been rejected optimally, then the method can give

quite good measurement combinations, which can be used for a feedback implementation. Of course, this approach will only reject disturbances which also are present in the "learning data" Y, other disturbances will not be handled optimally.

7.3. Conclusions

The data methods have the advantage that they do not rely on a model. The *optimal data approach* is relative restrictive, since it requires that the process has been brought to optimality before recording the data. This may be done by plant perturbations and a gradient descent method (Experimental optimization (Box and Draper, 1998; Jäschke and Skogestad, 2011a)), or other model-free optimization approaches such extremum seeking (Krstic and Wang, 2000).

The general historical data approach does not require optimal data. Its main advantage is that it requires limited plant testing, and that readily available historical data can be used to obtain more information about the optimal behavior of the plant. As with all data-based methods, the quality of the results will be very dependent on the quality of the data.

8. Towards self-optimizing control for dynamic problems

Considering the progress of self-optimizing control approaches in the context of steady-state optimization, it is a natural question to ask if the concepts can be extended towards dynamic optimization problems such as e.g. batch processes or product grade transitions. This problem is found to be much more complicated than finding self-optimizing variables for steady-state optimization, because of several complicating issues:

- The system is constantly changing, and finding variables (or variable combinations) that remain constant under optimal operation, even without disturbances affecting the system, can be very difficult.
- The time scale separation between optimization of the economic cost and control is not as clear as for steady-state optimization. That is, optimization and control cannot be considered separately as in steady-state optimization, where transients are assumed to have negligible impact on the operating cost.
- The CVs need to be controllable, during the overall operation period. For instance, in batch processes, the gain from inputs to outputs can change orders of magnitude over the batch, making it difficult to control these variables.
- Active set changes in dynamic optimization are more difficult to handle than in steady-state optimization. For instance, endpoint constraints require action ahead of time, and also switching between different active path constraint sets must be done at the right point in time.
- The active set is dependent on both, initial conditions and disturbances that may happen at any time during the process. In contrast to the steady-state case, where only the disturbance magnitude has an effect on the active set, also the timing of the disturbance will affect active set changes.

Nevertheless there has been work reported in the literature to find simple implementation strategies for optimal operation in batch processes. However, as the main focus of this paper is on steady-state optimal operation, we only give a brief overview of the main approaches. Generally, the self-optimizing control methods for dynamic systems are very close to the concept of necessary conditions of optimality (NCO) tracking (Srinivasan et al., 2003b,a; Srinivasan and Bonvin, 2004, 2007; Srinivasan et al., 2008), where the necessary optimality conditions of the process are either measured directly or approximated by a "solution model". Because handling active set changes in dynamic optimization is more challenging than in steady-state optimization, previous work on self-optimizing control for dynamic optimization problems has focused on cases where the active set does not change. However, this is not as restrictive as may seem at first sight, because the optimal trajectories are typically composed of continuous arcs (Srinivasan et al., 2003b), during which the active set remains constant. Therefore, the self-optimizing variables can be controlled within each arc, if a suitable switching strategy can be found.

One of the first to consider self-optimizing control for dynamic optimization problems was Dahl-Olsen et al. (2008). An approach by Ye et al. (2013b) uses the regression method described in Section 6.2 to find an approximation to the optimality conditions in terms of measurements: First the model is simulated and optimized over the entire expected disturbance space, and the data samples are saved. Then a measurement function that approximates the optimality conditions is fitted to this data. This kind of approach was applied by Grema and Cao (2016) to the case of waterflooding in oil reservoirs.

Another paper by Jäschke et al. (2011) uses the polynomial method described in Section 6.1. The method is applicable to simple input-affine dynamic systems that are described by polynomials. First the optimality conditions are formulated analytically, and then all unknown state and disturbance variables are eliminated using measurement relations.

A third method has been reported by Hu et al. (2012a), where candidate CVs were evaluated by the solution of a bi-objective optimization problem. In this approach, also a controller is designed, and the bi-objective optimization program attempts to simultaneously find the best CVs and minimize the tracking error of the controller.

Self-optimizing control concepts for dynamic optimization problems are still an immature research field, the time-varying nature of the optimal solution, and the difficulties in handling active set changes remain the biggest challenges. Although strictly speaking it is not possible to separate control and optimization, in practical cases this can often be done with a negligible loss. For example, a batch temperature can be controlled in a time scale of minutes, while the complete batch runs for hours. This has been successfully exploited also in the NCO tracking framework (Srinivasan et al., 2008) , where a "solution model" is developed, and a simplified near-optimal solution is implemented using simple controllers.

9. Applications of self-optimizing control

9.1. Overview of self-optimizing control applications reported in literature

Methods and concepts from self-optimizing control have been successfully applied to many processes reported in the literature. In Table 1 we have compiled an overview of processes for which a self-optimizing control structure has been used.

Table 1: Overview over applications of self-optimizing control in the literature

Process	References	Comments
Tennessee Eastman process	Larsson et al. (2001)	Large-scale case study, heuristics
Heat exchanger networks	Jäschke and Skogestad (2012c); Jäschke and Skogestad (2014); Jäschke and Skogestad (2014); Girei et al. (2014)	Linear, polynomial method and regression methods
Heat integrated distillation	Engelien et al. (2003); Engelien and Skogestad (2004)	With and without a pre-fractionator
Lee & Neville Evaporator	Cao (2005); Kariwala et al. (2008); Heldt (2010); Umar et al. (2012); Hu et al. (2012b)	Frequently studied test case
Reactor/separator with recycle	Larsson et al. (2001, 2003); Skogestad (2004a); Baldea et al. (2008); Seki and Naka (2008)	
Exothermic reactor	Kariwala (2007); Jäschke and Skogestad (2011a); Ye et al. (2013a)	Small case study
Gas lift allocation	Alstad and Skogestad (2003)	
Gasoline blending	Skogestad (2004b)	
HDA process	de Araújo et al. (2007b,a); Cao and Kariwala (2008)	
Distillation columns	Hori and Skogestad (2008); Kariwala and Cao (2010); Yelchuru and Skogestad (2013); Graciano et al. (2015)	Also regulatory layer selection
Petlyuk column (3 products)	Alstad and Skogestad (2007); Khanam et al. (2014)	
Kaibel column (4 products)	Ghadrdan et al. (2012)	
Refrigeration cycles	Jensen and Skogestad (2007a,b); Verheyleweghen and Jäschke (2016)	
HVAC system	Yin et al. (2015)	Cooling system, lab experiments
Ammonia Process	Araújo and Skogestad (2008); Manum and Skogestad (2012); Graciano et al. (2015)	Constraint handling Manum and Skogestad (2012)
Waste water treatment	de Araujo et al. (2013); Francisco et al. (2015)	
Catalytic Naphta reformer	Lid and Skogestad (2008)	
Waste incineration plant	Jäschke et al. (2009)	
Off-gas system	de Araujo and Shang (2009)	
CO_2-capturing process	Panahi and Skogestad (2011); Zaman et al. (2012); Schach et al. (2011, 2013)	
Advanced flash stripper	Walters et al. (2016)	
LNG process	Michelsen et al. (2010); Husnil et al. (2014)	
Stryrene Monomer plant	Vasudevan et al. (2009)	
Natural Gas to liquids process	Panahi and Skogestad (2012)	
IRCC HMR unit	Michelsen et al. (2012)	
IRCC acid gas removal unit	Jones et al. (2014)	
IRCC air separation unit	Roh and Lee (2014)	
Diesel hydrosulfurization plant	Sayalero et al. (2012)	
Cumene process	Gera et al. (2011, 2013)	
Sewer system	Mauricio-Iglesias et al. (2015)	Combined with Stochastic programming
Batch-reactor-optimization	Ye et al. (2015b); de Oliveira et al. (2016)	Batch-to-batch optimization, regression method
Solid oxide fuel cell	Chatrattanawet et al. (2015)	
Oxy-fuel circulating fluidized bed combustion	Niva et al. (2015)	
Water flooding of oil reservoirs	Grema and Cao (2016)	Regression method

9.2. Related applications

The idea of keeping a certain variable constant in order to achieve a superior goal is also found elsewhere. For example, in indirect control (Skogestad and Postlethwaite, 2005) the objective is to operate the process such that some unmeasured primary variables y are at given setpoints. Here, only the secondary measured variable is controlled, while the primary variable is not controlled directly. In a self-optimizing control framework, one may consider the secondary variables in indirect control as self-optimizing CVs. These should be kept at their optimal setpoints such that the primary variables assume their optimal values. Using self-optimizing control concepts, Ghadrdan et al. (2013) designed a static estimator that minimizes the closed-loop prediction error for such a case.

The self-optimizing control ideas have also been applied for designing regulatory control structures (Yelchuru and Skogestad, 2013). However, in contrast to the approaches described above, the objective for the regulatory layer is not to minimize the economic loss, but rather to minimize state drift under the effect of uncertainties. That is, the control structure is designed in such a way that the process states are affected as little as possible when disturbances enter the plant, and the control layer above is affected minimally.

10. Discussion on the relationship of self-optimizing control to other approaches

To achieve self-optimizing control, one needs a methodology for designing the *control structure*, i.e. selecting which variables to control with the available degrees of freedom. Once a control structure has been decided on (for example using the approaches and tools presented previously in this paper), we may control the CVs to their setpoints by any suitable controller. Thus, the problem of designing the controller and optimizing the plant can be considered separately. Indeed, the idea of self-optimizing control may be considered as an approach to separate the time-scales of the economic optimization and the controller design, such that the tasks of optimization and controller design can be decoupled as much as possible. This paradigm distinguishes self-optimizing control from other methods for achieving optimal operation, such as neighboring extremal control (Jazwinski, 1964; Bryson and Ho, 1975), extremum seeking control (Ariyur and Krstic, 2003), or economic model predictive control (Diehl et al., 2011; Ellis et al., 2014), where control and optimization are done simultaneously.

Neighboring extremal control (see e.g. Jazwinski (1964); Bryson and Ho (1975)) was introduced to obtain an optimal feedback law to compute the input trajectories for perturbed optimal control problems. Neighboring extremal control shares with the *local* self-optimizing control methods from Section 4 that it is based on a linearization of the nonlinear model and a quadratic approximation of the cost function. However, in neighboring extremal control this information is used to determine a state feedback law (or output feedback law (Gros et al., 2009)), while the local methods for self-optimizing control in Section 4 find an optimal approximation of the optimality conditions in terms of the measurements y. The controller design is then considered at a later stage.

We note that the term "neighboring extremal control" has also been used to describe the way of calculating an approximation of the gradient (François et al., 2012). However, going back to the original references on neighboring extremal control (Breakwell et al., 1963; Jazwinski, 1964; Bryson and Ho, 1975), it seems that the goal is to compute a feedback law, which happens to be based on a linearization of the optimality conditions.

Extremum seeking control (Krstic and Wang, 2000) was also introduced as a *controller design* methodology. The goal is to design a controller which generates an input that maximizes

the value of a given measurement. A small sinusoidal perturbation is continuously added to the input signal in order to obtain gradient information which is used by the controller to generate an input that drives the measurement to its maximum value. The large body of literature that studies the stability of extremum seeking control schemes (Krstic and Wang, 2000; Ariyur and Krstic, 2003; Guay and Zhang, 2003; Guay and Dochain, 2017) reflects the fact that extremum seeking control is a *controller design methodology*, and not a control structure design methodology.

Moreover, to design an extremum seeking controller the cost must be measured online, while this is not necessary for most of the self-optimizing control approaches. Note also that the measurement to be maximized would be a very poor choice of a self-optimizing variable, because it has zero gain at the optimum. Moreover, if the setpoint is chosen below the optimum, there are multiple solutions, while if the setpoint is chosen too high, it is infeasible.

NCO-tracking (Necessary conditions of optimality tracking) (Srinivasan and Bonvin, 2004, 2007; Srinivasan et al., 2008) is similar in spirit to self-optimizing control. The idea is simply to select the optimality conditions as controlled variables, and to make sure they are tracked under plant operation. Here, too this approach was developed initially for dynamic optimization problems, where the controller design and the optimality cannot be strictly separated, but was later also applied for optimizing steady state processes (François et al., 2005). While the optimality conditions are the ideal self-optimizing controlled variables, they are typically not measured, such that it is necessary to express them in terms of measurements. This is indeed the approach used in the method for systems described by polynomials (Jäschke and Skogestad, 2012b). However, it is not always possible to express the optimality conditions exactly in terms of measurements, and in this case the loss L used in Section 4 is a natural criterion for assessing the quality of the approximation. The local self-optimizing control methods (Halvorsen et al., 2003; Alstad and Skogestad, 2007; Alstad et al., 2009) were developed directly based on the loss L. The connection to the optimality conditions was established later by Jäschke et al. (2011).

Economic MPC (Diehl et al., 2011; Ellis et al., 2014) has been proposed to integrate the layers of a control structure into a single monolithic controller. The economic MPC controller is not designed to track setpoints of CVs, but rather to manipulate the inputs in such a way that the economic cost is optimized directly. This has the advantage that also transients can be optimized, which is important for e.g. batch systems or expensive product grade transitions. However, unlike controlling self-optimizing CVs, which can often be done by simple PID controllers, economic MPC comes with much more complexity. It requires accurate models that reflect the true plant behavior over several time-scales. For large plants, these models are expensive to develop and maintain. Moreover, the optimal solution may be difficult for the operators to interpret, and if parts of the plant are operated manually, this may require to change the whole MPC controller, or take it offline. On the other hand, the nature of a self-optimizing control structure, lends itself easily to a decentralized control structure, such that parts of the plant can be operated manually, without requiring to re-engineer the whole control system.

11. Current status and future work

11.1. Summary of the status of self-optimizing control

In recent years, the problem of finding self-optimizing controlled variables has been approached rigorously, and mathematical formulations for local methods have been developed. These methods are based on an approximation of the nonlinear problem around a nominal operating point, such that they are essentially local results. Nevertheless, they are useful for finding

good candidate control structures, that then can be evaluated on a larger operation region. Based on the local methods, tailor-made branch and bound methods were proposed to efficiently search all possible combinations of variables for the best subset of measurements (Kariwala and Cao, 2009).

The local methods were extended towards model-free, data-based approaches, where the optimal self-optimizing CVs are found by analyzing measurement data. Moreover, there have been significant efforts to obtain variables that are near-optimal in a larger operating region. These include possibilities for handling active set changes, but also smooth nonlinearities. Connections to related methods such as NCO tracking were established (Jäschke and Skogestad, 2011a), and were used to develop approaches for finding CVs that are not based on linearization around a nominal operating point. In particular the polynomial approach (Jäschke and Skogestad, 2012a) and data driven approaches (Ye et al., 2013a) may be used for finding CVs that can be used within a larger operational envelope. A different approach to cover larger operation regions is described by Manum and Skogestad (2012), where the authors propose switching the control structure for each active set region.

Self-optimizing control ideas were also applied to dynamic optimization problems. However, they apply to very specific cases, and more general results are currently not available.

11.2. Future work – open issues

Despite the significant progress in the field of designing self-optimizing control structures, there remain challenges and open questions that need to be addressed in future research:

Industrial applications. Considering that self-optimizing control has been applied in many simulation studies, it seems surprising that there has been little application in experiments and industry. The only experimental work that we are aware of was was performed by Yin et al. (2015). While many industrial processes are controlled by self-optimizing CVs, it seems that the predominant way to design a control structure is still based on engineering insight. A successful industrial application and verification of the approaches presented in this paper could pave the way for more wide-spread industrial acceptance, and operational savings. In addition this may lead to interesting practical and theoretical challenges, as well as new insights.

Integration of self-optimizing control ideas into the process design phase. Many decisions that affect the operation of the plant are made during the plant design phase. In order to reach the maximum potential it is necessary to study how the design will affect later performance during operation. Although there has been work on integrating process control and design, see e.g. Sharifzadeh (2013), there has been no work on systematically integrating self-optimizing control concepts into the plant design phase. Results in this area could have a large impact, because modern plant design is generally based on process models, and including operational aspects into the plant design can lead to significant savings later during operation of the plant.

Structural constraints on the H-matrix. Beside not using all available measurements in a CV, see Section 4.7, it is often desirable to design the control structure such that measurements of a certain part of the plant are controlled by inputs that belong to the same part. This imposes a structure on the measurement selection matrix. Although some special structures (the ones that are preserved upon premultiplying H by a non-singular invertible matrix) can be handled with the present approaches, finding control structures with general structural constraints still is a difficult problem that requires finding a H matrix of the desired structure that minimizes the Frobenius or 2-norm of $M = J_{uu}^{1/2}(HG^y)^{-1}HY$. This results in non-convex optimization problems that are difficult to solve.

Robust self-optimizing control. Most current methods for selecting self-optimizing CVs are based on availability of a model that reflects the true plant behavior. In particular, all uncertainty of the plant is assumed to be modelled as parametric uncertainty. In reality there will be uncertainty in the model and its structure, as well as unmodelled parametric disturbances. If the models are poor, this may lead to an unacceptable economic loss. Currently there are no systematic methods for addressing uncertainty in the model structure within the self-optimizing control framework. Future research may address this for example by probing the system to handle unmodelled disturbances. Similar to the ideas in Jäschke and Skogestad (2011a), one may combine self-optimizing control ideas with extremum seeking methods (Krstic and Wang, 2000) or dual control approaches (Heirung et al., 2015).

Self-optimizing CVs from operating data. Although there has been preliminary work on finding self-optimizing controlled variables based on plant data (Section 7) there are still open questions. One issue is how to systematically handle noise in the data. Other topics of interest are methods for analyzing the data, and for extracting information that is relevant for finding self-optimizing CVs. Finally, it would be desirable to develop methods for combining model information with measurement data in order to obtain improved performance. Potential research directions may include combining model based methods with data, using e.g. modifier adaptation techniques (Marchetti et al., 2010).

Active Set changes. An issue that often makes it difficult to apply self-optimizing control ideas to real systems is that the optimal solution may become very complex with many active set regions and complicated measurement combinations involving many variables. Possible future research may therefore consider ways for further simplifying the results that have been outlined in this paper. For example, if we allow only p number of partitions of the operating range (i.e. p control structures), what are the p optimal measurement combinations?

Short-cut methods. At present, finding a self-optimizing control structure requires modelling the plant and optimizing the model to find the nominal operating point. This can be very cumbersome, and it would be useful to be able to suggest "standard" solutions that can be shown to be near-optimal in most practically relevant cases. This could take the shape of a "control structure library" which contains building blocks for a self-optimizing control structure. These building blocks could then be put together to yield an overall self-optimizing control structure.

Self-optimizing variables for dynamic optimization problems. For dynamic optimization problems, the time-dependency of the optimal operation strategy may result in extremely many possible scenarios and active constraint regions, and finding ways to deal with this complexity is an open question. Moreover, within a region, finding variables or variable combinations that give near-optimal operation when kept at their constant setpoints is not straightforward for dynamic optimization problems. Also, handling measurement noise has not been considered systematically for such cases.

Including controllability considerations in the CV selection. Typically, only the steady state economics are taken into account in the present methods for CV selection. However, in practice the dynamic behavior of the CVs and controllability are important selection criteria, too. A potential research direction is to include these dynamic controllability considerations into the self-optimizing framework in a systematic manner.

Self-optimizing control with continuous and integer variables. In practice there often occur situations, where units are turned on or shut down. This gives rise to optimization problems with both integer and continuous variables. There are currently no methods to address such kind of problem systematically.

General nonlinear combinations of measurements as CVs Although there has been work

on finding polynomial measurement combinations as CVs for polynomial systems, there are no systematic ways for finding CVs that consist of more general nonlinear measurement combinations. Moreover, it would be desirable to have tools to help the engineer decide if a nonlinear measurement combination is required or not.

Applications in other disciplines. Self-optimizing control has been developed in the process control community. However, the idea of keeping certain key variables constant at a given setpoint when disturbances occur, is applied in other fields as well (Skogestad, 2004b). For example in portfolio management, where a constant percentage of the value is invested in stocks, and the remaining part is invested in government bonds. Another example is the central bank, which aims to maximize welfare by adjusting the interest rate such that the inflation rate remains at a constant value. It would be interesting to apply the self-optimizing control approaches described in this paper to rigorous studies in other disciplines, such as economy or social sciences. This could open up new research fields, and at the same time provide support for decision and policy makers.

Acknowledgements

J.J. wishes to thank S. Skogestad for introducing him to self-optimizing control, and for many good discussions on the topic.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Approximated loss expression & Minimum Singular Value Rule (Maximum gain rule)

The minimum singular value rule was described by Skogestad and Postlethwaite (1996) and may be used as an approximation to evaluating the local loss described in Section 4. It gives a good intuition about what variables are good CV candidates. The loss L associated with a given control strategy is caused by two factors: The first factor is referred to as the *setpoint error*,

$$v(d) = c_s - c^*(d), \tag{A.1}$$

where c_s is the actual setpoint, and $c^*(d)$ denotes the *optimal* setpoint. The setpoint error can be directly related to Requirement 2) listed in Section 1.4, because a CV whose optimal value is very sensitive to disturbances will have a large setpoint error.

The second factor causing suboptimal operation is referred to as *implementation error* e. It accounts for the fact that it is not possible to keep the CVs at their optimal values because of measurement noise n^y or the use of a controller without integral action,

$$e = c - c_s. \tag{A.2}$$

Here c denotes the actual value of the CV, and c_s denotes the setpoint. Note that if controllers with integral action are used, we have that

$$e = n^y. \tag{A.3}$$

The implementation error e is related to Requirements 1) and 3) in Section 1.4. Requirement 1) states that the inputs should have a sufficiently strong effect, so that c can be kept close to its

optimal value. Requirement 3) states that the CV should be easy to keep at its setpoint. This in turn requires a small implementation error.

The total error e_c which causes the loss L during operation, is thus

$$\begin{aligned} e_c &= c - c^*(d) \\ &= \nu(d) + e. \end{aligned} \tag{A.4}$$

A self-optimizing structure will aim to reduce the control error e_c such that the loss is acceptable.

Now, for every disturbance direction, we may calculate the setpoint error $\nu(d_j)$ by

$$\nu(d_j) = HF\Delta d_j \tag{A.5}$$

where Δd_j is a vector with the maximum value of the disturbance component d_j on the j-th position, and zeros otherwise. Using (23) and (A.4), and assuming integral action (i.e. $e = n^y$) we have that the total error for a given control structure given by the matrix H is

$$\begin{aligned} e_c &= \Delta c - \Delta c^*(d) \\ &= HG^y\Delta u + HG_d^y\Delta d + Hn^y \\ &\quad - HG^y\Delta u^* - HG_d^y\Delta d - Hn^y \\ &= HG^y(\Delta u - \Delta u^*). \end{aligned} \tag{A.6}$$

Assuming that HG^y is invertible, we can solve for $(\Delta u - \Delta u^*)$, and inserting into (18) yields the following expression for the loss variable:

$$z = J_{uu}^{1/2}(HG^y)^{-1}e_c. \tag{A.7}$$

To obtain an approximation of the loss, it is assumed that a scaling matrix S_c exists, such that

$$e_c = S_c e_c', \tag{A.8}$$

where $\left\|e_c'\right\|_2^2 = 1$. Note that the diagonal elements of S_c satisfy

$$S_{c,ii} = \operatorname{diag}\left(\sum_{j=1}^{n_d} |\nu(d_j)| + |n_i^y|\right). \tag{A.9}$$

Using the scaling matrix S_c, the loss variable z from (A.7) can be written as

$$z = J_{uu}^{1/2}(HG^y)^{-1}S_c e_c'. \tag{A.10}$$

Defining the "scaled gain" G' as

$$G' = S_c^{-1}HG^y J_{uu}^{-1/2}, \tag{A.11}$$

we have

$$z = G'^{-1}e_c'. \tag{A.12}$$

Now the worst-case loss corresponding to a given set of CVs represented by H can be estimated by

$$L_{\text{wc}} = \max_{\|e_c\|\leq 1} \frac{1}{2}\|z\|_2^2 = \frac{1}{2}\bar{\sigma}^2(G'^{-1}) = \frac{1}{2}\frac{1}{\underline{\sigma}^2(G')}. \tag{A.13}$$

Based on (A.13) the maximum gain rule states that good CVs should be selected such as to maximize the scaled gain matrix G', or more precisely to maximize the minimum singular value of the scaled gain.

Remark 12. *The minimum singular value rule combines nicely the requirements for a good CV. Selecting CVs corresponding to a large scaled gain will be easy to control (large gain from the inputs to CV), and will have a low optimal variation (since we are scaling by the small optimal variation).*

Remark 13. *In the original work (Skogestad and Postlethwaite, 1996; Halvorsen et al., 2003), the scaled gain is defined as* $\hat{G}' = S_c^{-1} H G^y$*, and it is assumed that* J_{uu} *can be expressed as* αU*, where* α *is scalar and* U *is an orthogonal matrix. Then the worst-case loss can be written as* $L_{\text{wc}} = \frac{\alpha}{2} \frac{1}{\underline{\sigma}^2(\hat{G}')}$*. However, the assumption that* J_{uu} *can be written as a constant times an orthogonal matrix is not valid in general, and may lead to an additional loss (Hori and Skogestad, 2008).*

Example 8 (Minimum Singular Value Rule). *We continue considering the process from Example 1. Here we have* $v(d) = F W_d d$ *given for a disturbance with maximum value* $(d = 1)$ *as*

$$v(d=1) = \begin{bmatrix} 0 & 20 & 5 & 1 \end{bmatrix}^T, \tag{A.14}$$

and the scaling matrix $S_c = diag(F W_d) + W_n = diag\left(\begin{bmatrix} 1 & 21 & 6 & 2 \end{bmatrix}\right)$ *Thus, the scaled gains for the four measurements become*

$$G_1' = 0.1/(1 \cdot \sqrt{2}) = 0.071 \tag{A.15}$$

$$G_2' = 20/(21 \cdot \sqrt{2}) = 0.673 \tag{A.16}$$

$$G_3' = 10/(6 \cdot \sqrt{2}) = 1.179 \tag{A.17}$$

$$G_4' = 1/(2 \cdot \sqrt{2}) = 0.354. \tag{A.18}$$

The scaled gain is largest for y_3*, so according to the minimum singular value rule, this is the best single measurement that should be selected as a CV. This is in correspondence with the results from Example 1, where* y_3 *was shown to give the smallest loss.*

A key limitation for the maximum gain rule is the assumption of the existence of S_c, with the property that $\left\|e_c'\right\| = 1$. The basis for this assumption is that the errors for each element in $e_c = v(d) + n^y$ are independent of each other. However, this assumption is generally not satisfied, because the setpoint errors for different CVs are not independent. Therefore the maximum gain rule may lead to misleading results (Hori and Skogestad, 2008).

References

Vidar Alstad. *Studies on selection of controlled variables*. PhD thesis, Norwegian University of Science and Technology, Department of Chemical Engineering, 2005.

Vidar Alstad and Sigurd Skogestad. Combination of measurements as controlled variables for self-optimizing control. In Andrzej Kraslawski and Ilkka Turunen, editors, *European Symposium on Computer Aided Process Engineering-13 36th European Symposium of the Working Party on Computer Aided Process Engineering*, volume 14 of *Computer Aided Chemical Engineering*, pages 353 – 358. Elsevier, 2003. doi: http://dx.doi.org/10.1016/S1570-7946(03)80140-2. URL `http://www.sciencedirect.com/science/article/pii/S1570794603801402`.

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.

Antonio Araújo and Sigurd Skogestad. Control structure design for the ammonia synthesis process. *Computers & Chemical Engineering*, 32(12):2920 – 2932, 2008. ISSN 0098-1354. doi: http://dx.doi.org/10.1016/j.compchemeng.2008.03.001. URL http://www.sciencedirect.com/science/article/pii/S0098135408000392.

Arnon Arbel, Irven H. Rinard, and Reuel Shinnar. Dynamics and control of fluidized catalytic crackers. 3. designing the control system: choice of manipulated and measured variables for partial control. *Industrial & Engineering Chemistry Research*, 35(7):2215–2233, 1996.

Kartik B. Ariyur and Miroslav Krstic. *Real-Time Optimization by Extremum-Seeking Control*. Wiley-Interscience, 2003.

Karl J. Åström and Björn Wittenmark. *Adaptive control*. Dover Publications Inc., 2008.

Michael Baldea, Antonio Araujo, Sigurd Skogestad, and Prodroms Daoutidis. Dynamic considerations in the synthesis of self-optimizing control structures. *AIChE Journal*, 54(7):1830–1841, 2008. ISSN 1547-5905. doi: 10.1002/aic.11470. URL http://dx.doi.org/10.1002/aic.11470.

Atanu Banerjee and Yaman Arkun. Control configuration design applied to the tennessee eastman plant-wide control problem. *Computers & Chemical Engineering*, 19(4):453 – 480, 1995. ISSN 0098-1354. doi: http://dx.doi.org/10.1016/0098-1354(94)00058-V. URL http://www.sciencedirect.com/science/article/pii/009813549400058V.

Mato Baotić, Francesco Borrelli, Alberto Bemporad, and Manfred Morari. Efficient on-line computation of constrained optimal control. *SIAM Journal on Control and Optimization*, 47:2470–2489, 2008.

George E. P. Box and Norman R. Draper. *Evolutionary Operation. A Statistical Method for Process Improvement*. John Wiley & Sons, 1998.

John V. Breakwell, Jason L. Speyer, and Arthur E. Bryson. Optimization and control of nonlinear systems using the second variation. *Journal of the Society for Industrial and Applied Mathematics Series A Control*, 1(2):193–223, 1963. doi: 10.1137/0301011. URL http://dx.doi.org/10.1137/0301011.

Arthur E. Bryson and Yu-Chi Ho. *Applied Optimal Control, revised edition*. Taylor & Francis, 1975.

Laurent Busé and Bernarad Mourrain. Using the maple multires package, 2003. URL http://www-sop.inria.fr/galaad/software/multires.

Yi Cao. Constrained self-optimizing control via differentiation. In *Proceedings of the 7th International Symposium on Advanced Control in Chemical Processes (ADCHEM 2003), Honkong*, pages 445–453, July 2004.

Yi Cao. Direct and indirect gradient control for static optimisation. *International Journal of Automation and Computing*, 2:60–66, 2005.

Yi Cao and Vinay Kariwala. Bidirectional branch and bound for controlled variable selection: Part I. principles and minimum singular value criterion. *Computers & Chemical Engineering*, 32(10):2306 – 2319, 2008.

Narissara Chatrattanawet, Sigurd Skogestad, and Amornchai Arpornwichanop. Control structure design and dynamic modeling for a solid oxide fuel cell with direct internal reforming of methane. *Chemical Engineering Research and Design*, 98:202–211, June 2015. ISSN 02638762. doi: 10.1016/j.cherd.2015.04.029. URL http://www.sciencedirect.com/science/article/pii/S0263876215001434.

Panagiotis D. Christofides and Nael H. El-Farra, editors. *Special issue on Economic nonlinear model predictive control, Journal of pocess control*, volume 24 (8), 2014.

David Cox, John Little, and Donal O'Shea. *Using Algebraic Geometry*. Springer, 2005.

Håkon Dahl-Olsen, Sridharakumar Narasimhan, and Sigurd Skogestad. Optimal output selection for control of batch processes. In *Proceedings of the American Congrol Conference, Seattle, WA*, 2008.

Antonio C. B. de Araujo and Helen Shang. Enhancing a smelter off-gas system using a plant-wide control design. *Industrial & Engineering Chemistry Research*, 48(6):3004–3013, 2009. doi: 10.1021/ie800254s. URL http://dx.doi.org/10.1021/ie800254s.

Antonio C. B. de Araújo, Eduardo S. Hori, and Sigurd Skogestad. Application of plantwide control to the HDA process. II–Regulatory control. *Industrial & Engineering Chemistry Research*, 46(15):5159–5174, 2007a. doi: 10.1021/ie061393z. URL http://dx.doi.org/10.1021/ie061393z.

Antonio C. B. de Araujo, Simone Gallani, Michela Mulas, and Sigurd Skogestad. Sensitivity analysis of optimal operation of an activated sludge process model for economic controlled variable selection. *Industrial & Engineering Chemistry Research*, 52(29):9908–9921, 2013. doi: 10.1021/ie4006673. URL http://dx.doi.org/10.1021/ie4006673.

Antonio C.B. de Araújo, Marius Govatsmark, and Sigurd Skogestad. Application of plantwide control to the HDA process. I–Steady-state optimization and self-optimizing control. *Control Engineering Practice*, 15(10): 1222 – 1237, 2007b. ISSN 0967-0661. doi: http://dx.doi.org/10.1016/j.conengprac.2006.10.014. URL http://www.sciencedirect.com/science/article/pii/S0967066106001997.

Vinicius de Oliveira, Johannes Jschke, and Sigurd Skogestad. Null-space method for optimal operation of transient processes. *IFAC-PapersOnLine*, 49(7):418 – 423, 2016. ISSN 2405-8963. doi: http://dx.doi.org/10.1016/j.ifacol.2016.07.378. URL //www.sciencedirect.com/science/article/pii/S2405896316305857. 11th {IFAC} Symposium on

Dynamics and Control of Process SystemsIncluding Biosystems DYCOPS-CAB 2016Trondheim, Norway, 68 June 2016.

Moritz Diehl, Rishi Amrit, and James B. Rawlings. A lyapunov function for economic optimizing model predictive control. *IEEE Transactions on Automatic Control*, 56(3):703–707, 2011. ISSN 0018-9286. doi: 10.1109/TAC.2010.2101291.

Matthew Ellis, Helen Durand, and Panagiotis D. Christofides. A tutorial review of economic model predictive control methods. *Journal of Process Control*, 24(8):1156 – 1178, 2014. ISSN 0959-1524. doi: http://dx.doi.org/10.1016/j.jprocont.2014.03.010. URL http://www.sciencedirect.com/science/article/pii/S0959152414000900.

Satish Enagandula and James B. Riggs. Distillation control configuration selection based on product variability prediction. *Control Engineering Practice*, 14(7):743 – 755, 2006. ISSN 0967-0661. doi: http://dx.doi.org/10.1016/j.conengprac.2005.03.011. URL http://www.sciencedirect.com/science/article/pii/S0967066105000912.

Hilde Kathrine Engelien and Sigurd Skogestad. Selecting appropriate control variables for a heat-integrated distillation system with prefractionator. *Computers & Chemical Engineering*, 28(5):683 – 691, 2004.

Hilde Kathrine Engelien, Truls Larsson, and Sigurd Skogestad. Implementation of optimal operation for heat integrated distillation columns. *Chemical Engineering Research and Design*, 81(2): 277 – 281, 2003. ISSN 0263-8762. doi: http://dx.doi.org/10.1205/026387603762878755. URL http://www.sciencedirect.com/science/article/pii/S0263876203723089.

Sebastian Engell. Feedback control for optimal process operation. *Journal of Process Control*, 17(3):203 – 219, 2007.

Sebastian Engell, Tobias Scharf, and Marten Völker. A methodology for control structure selection based on rigorous process models. In *Proc. 16th IFAC World Congress 2005*, pages 878–884, 2005.

Anthony V Fiacco. *Introduction to sensitivity and stability analysis in nonlinear programming*, volume 226. Academic press New York, 1983.

W. Findeisen, F.N. Nailey, M. Brdys, K. Malinowski, P. Tatjewski, and A. Wozniak. *Control and Coordination in Hierarchical Systems*. John Wiley & Sons, 1980.

Alan Stuart Foss. Critique of chemical process control theory. *AIChE Journal*, 19(2):209–214, 1973. ISSN 1547-5905. doi: 10.1002/aic.690190202.

Gregory François, Bala Srinivasan, and Dominique Bonvin. Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty. *Journal of Process Control*, 15(6):701 – 712, 2005.

M. Francisco, S. Skogestad, and P. Vega. Model predictive control for the self-optimized operation in wastewater treatment plants: Analysis of dynamic issues. *Computers & Chemical Engineering*, 82:259–272, November 2015. ISSN 00981354. doi: 10.1016/j.compchemeng.2015.07.003. URL http://www.sciencedirect.com/science/article/pii/S0098135415002343.

Grégory François, Balasubrahmanya Srinivasan, and Dominique Bonvin. Comparison of six implicit real-time optimization schemes. *Journal Européen des Systèmes Automatisés*, 46:291–305, 2012.

Vivek Gera, Nitin Kaistha, Mehdi Panahi, and Sigurd Skogestad. Plantwide control of a cumene manufacture process. 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 522 – 526. Elsevier, 2011. doi: http://dx.doi.org/10.1016/B978-0-444-53711-9.50105-X. URL http://www.sciencedirect.com/science/article/pii/B978044453711950105X.

Vivek Gera, Mehdi Panahi, Sigurd Skogestad, and Nitin Kaistha. Economic plantwide control of the cumene process. *Industrial & Engineering Chemistry Research*, 52(2):830–846, 2013. doi: 10.1021/ie301386h. URL http://dx.doi.org/10.1021/ie301386h.

Maryam Ghadrdan, Sigurd Skogestad, and Ivar Johan Halvorsen. Economically optimal operation of kaibel distillation column: Fixed boilup rate. In *Proceedings of the IFAC Symposium on Advanced Control of Chemical Processes*, pages 744–749, 2012.

Maryan Ghadrdan, Chriss Grimholt, and Sigurd Skogestad. A new class of model-based static estimators. *Industrial & Engineering Chemistry Research*, 52(35):12451–12462, 2013. doi: 10.1021/ie400542n. URL http://dx.doi.org/10.1021/ie400542n.

Salihu Adamu Girei, Yi Cao, Alhaji Shehu Grema, Lingjian Ye, and Vinay Kariwala. *Data-driven Self-optimizing Control*, volume 33 of *Computer Aided Chemical Engineering*. Elsevier, 2014. ISBN 9780444634344. doi: 10.1016/B978-0-444-63456-6.50109-5. URL http://www.sciencedirect.com/science/article/pii/B9780444634566501095.

Marius Govatsmark. *Integrated Optimization and Control*. PhD thesis, Norwegian University of Science and Technology, NTNU, 2003.

José Eduardo A. Graciano, Johannes Jäschke, Galo A.C. Le Roux, and Lorenz T. Biegler. Integrating self-optimizing control and real-time optimization using zone control MPC. *Journal of Process Control*, 34:35–48, October 2015. ISSN 09591524. doi: 10.1016/j.jprocont.2015.07.003. URL

http://www.sciencedirect.com/science/article/pii/S0959152415001523.

A. S. Grema and Yi Cao. Optimal feedback control of oil reservoir waterflooding processes. *International Journal of Automation and Computing*, 13(1):73–80, 2016. ISSN 1751-8520. doi: 10.1007/s11633-015-0909-7. URL http://dx.doi.org/10.1007/s11633-015-0909-7.

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

Martin Grötschel, Sven O. Krumke, and Jörg Rambau, editors. *Online Optimization of Large Scale Systems*. Springer, September 2001.

Martin Guay and Denis Dochain. A proportional-integral extremum-seeking controller design technique. *Automatica*, 77:61 – 67, 2017. ISSN 0005-1098. doi: http://dx.doi.org/10.1016/j.automatica.2016.11.018. URL //www.sciencedirect.com/science/article/pii/S0005109816304538.

Martin Guay and Tao Zhang. Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties. *Automatica*, 39(7):1283 – 1293, 2003.

Inc Gurobi Optimization. Gurobi optimizer reference manual, 2015. URL http://www.gurobi.com.

Ivar J. Halvorsen, Sigurd Skogestad, John C. Morud, and Vidar Alstad. Optimal selection of controlled variables. *Industrial & Engineering Chemistry Research*, 42(14):3273–3284, 2003.

Jonathan A. Heath, Ioannis K. Kookos, and John D. Perkins. Process control structure selection based on economics. *AIChE Journal*, 46(10):1998–2016, 2000. ISSN 1547-5905. doi: 10.1002/aic.690461012. URL http://dx.doi.org/10.1002/aic.690461012.

Tor Aksel N. Heirung, Bjarne Foss, and B. Erik Ydstie. Mpc-based dual control with online experiment design. *Journal of Process Control*, 32:64 – 76, 2015. ISSN 0959-1524. doi: http://dx.doi.org/10.1016/j.jprocont.2015.04.012.

Steffen Heldt. Dealing with structural constraints in self-optimizing control engineering. *Journal of Process Control*, 20 (9):1049 – 1058, 2010.

Eduardo Shigueo Hori and Sigurd Skogestad. Selection of controlled variables: Maximum gain rule and combination of measurements. *Industrial & Engineering Chemistry Research*, 47(23):9465–9471, 2008. doi: 10.1021/ie0711978. URL http://dx.doi.org/10.1021/ie0711978.

Wuhua Hu, Jianfeng Mao, Gaoxi Xiao, and Vinay Kariwala. Selection of self-optimizing controlled variables for dynamic processes. In *Proc IFAC Symposium on Advanced Control of Chemical Processes (AdChem 2012), Singapore, July 10-13*, pages 774–779, 2012a.

Wuhua Hu, Lia Maisarah Umar, Gaoxi Xiao, and Vinay Kariwala. Local self-optimizing control of constrained processes. *Journal of Process Control*, 22(2):488 – 493, 2012b.

Yuli Amalia Husnil, GyeongCheol Yeo, and Moonyong Lee. Plant-wide control for the economic operation of modified single mixed refrigerant process for an offshore natural gas liquefaction plant. *Chemical Engineering Research and Design*, 92(4):679–691, April 2014. ISSN 02638762. doi: 10.1016/j.cherd.2013.11.009. URL http://www.sciencedirect.com/science/article/pii/S026387621300498X.

International Business Machines. Ibm ilog cplex optimization studio documentation, 2014. URL http://www.ibm.com.

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.

Johannes Jäschke and Sigurd Skogestad. Optimal operation by controlling the gradient to zero. *Proc.IFAC World Congress, Milano, Italy*, pages 593–598, 2011.

Johannes Jäschke and Sigurd Skogestad. Optimal controlled variables for polynomial systems. *Journal of Process Control*, 22(1):167 – 179, 2012a.

Johannes Jäschke and Sigurd Skogestad. Parallel heat exchanger control, EU/UK patent application pct/ep2013/059304 and gb1207770.7, 2012b.

Johannes Jäschke and Sigurd Skogestad. Control structure selection for optimal operation of a heat exchanger network. In *Control (CONTROL), 2012 UKACC International Conference on*, pages 148 –153, sept. 2012c.

Johannes Jäschke and Sigurd Skogestad. Using process data for finding self-optimizing controlled variables. In *Proc IFAC International Symposium on Dynamics and Control of Process Systems (DYCOPS 2013), Mumbai, India, July 10-13*, pages 451–456, 2013.

Johannes Jäschke and Sigurd Skogestad. Optimal operation of heat exchanger networks with stream split: Only temperature measurements are required. *Computers & Chemical Engineering*, 70: 35–49, November 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2014.03.020. URL http://www.sciencedirect.com/science/article/pii/S009813541400101X.

Johannes Jäschke and Sigurd Skogestad. A self-optimizing strategy for optimal operation of a preheating train for a crude oil unit. In Jiri Jaromir Klemes, Petar Sabev Varbanov, and Pen Yen Liew, editors, *24th Eu-*

ropean Symposium on Computer Aided Process Engineering, volume 33 of *Computer Aided Chemical Engineering*, pages 607 – 612. Elsevier, 2014. doi: http://dx.doi.org/10.1016/B978-0-444-63456-6.50102-2. URL http://www.sciencedirect.com/science/article/pii/B9780444634566501022.

Johannes Jäschke, Helge Smedsrud, Sigurd Skogestad, and Henrik Manum. Optimal operation of a waste incineration plant for district heating. In *Proc. American Control Conference, ACC, St. Louis*, pages 665–670, June 2009.

Johannes Jäschke, Miroslav Fikar, and Sigurd Skogestad. Self-optimizing invariants in dynamic optimization. In *Proceedings of the 50th CDC and ECC*, pages 7753–7758, Orlando, Florida, 2011.

Johannes Jäschke, Xue Yang, and Lorenz T. Biegler. Fast economic model predictive control based on nlp-sensitivities. *Journal of Process Control*, 24(8):1260–1272, 2014. ISSN 0959-1524. doi: http://dx.doi.org/10.1016/j.jprocont.2014.04.009. URL http://www.sciencedirect.com/science/article/pii/S0959152414001085.

Andrew H. Jazwinski. Optimal trajectories and linear control of nonlinear systems. *AIAA Journal*, 2(8):1371 – 1379, 1964.

Jørgen Bauck Jensen and Sigurd Skogestad. Optimal operation of simple refrigeration cycles: Part i: Degrees of freedom and optimality of sub-cooling. *Computers & Chemical Engineering*, 31(5-6):712 – 721, 2007a.

Jørgen Bauck Jensen and Sigurd Skogestad. Optimal operation of simple refrigeration cycles: Part ii: Selection of controlled variables. *Computers & Chemical Engineering*, 31(12):1590 – 1601, 2007b.

Dustin Jones, Debangsu Bhattacharyya, Richard Turton, and Stephen E. Zitney. Plant-wide control system design: Primary controlled variable selection. *Computers & Chemical Engineering*, 71: 220–234, December 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2014.08.004. URL http://www.sciencedirect.com/science/article/pii/S0098135414002464.

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

Vinay Kariwala and Yi Cao. Bidirectional branch and bound for controlled variable selection. part II: Exact local method for self-optimizing control. *Computers & Chemical Engineering*, 33(8):1402 – 1412, 2009.

Vinay Kariwala and Yi Cao. Bidirectional branch and bound for controlled variable selection part III: Local average loss minimization. *Industrial Informatics, IEEE Transactions on*, 6(1):54 –61, 2010.

Vinay Kariwala, Yi Cao, and S. Janardhanan. Local self-optimizing control with average loss minimization. *Industrial & Engineering Chemistry Research*, 47:1150–1158, 2008.

Vinay Kariwala, Lingjian Ye, and Yi Cao. Branch and bound method for regression-based controlled variable selection. *Computers & Chemical Engineering*, 54(0):1 – 7, 2013.

Ambari Khanam, Mohammad Shamsuzzoha, and Sigurd Skogestad. *Optimal Operation and Control of Divided Wall Column*, volume 33 of *Computer Aided Chemical Engineering*. Elsevier, 2014. ISBN 9780444634344. doi: 10.1016/B978-0-444-63456-6.50113-7. URL http://www.sciencedirect.com/science/article/pii/B9780444634566501137.

Steinar Kolås, Bjarne Foss, and Tor Steinar Schei. State estimation is the real challenge in nmpc. In *International workshop on assessment and future directions of nonlinear model predictive control, Pavia, Italy*, 2008.

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

Truls Larsson, Kristin Hestetun, Espen Hovland, and Sigurd Skogestad. Self-optimizing control of a large-scale plant: The tennessee eastman process. *Industrial & Engineering Chemistry Research*, 40(22):4889–4901, 2001.

Truls Larsson, Marius Govatsmark, Sigurd Skogestad, and Cheng-Ching Yu. Control structure selection for reactor, separator, and recycle processes. *Industrial & Engineering Chemistry Research*, 42(6):1225–1234, 2003. doi: 10.1021/ie0200860. URL http://dx.doi.org/10.1021/ie0200860.

Tore Lid and Sigurd Skogestad. Data reconciliation and optimal operation of a catalytic naphtha reformer. *Journal of Process Control*, 18(3-4):320 – 331, 2008.

William L. Luyben. Steady-state energy conservation aspects of distillation column control system design. *Industrial & Engineering Chemistry Fundamentals*, 14(4):321–325, 1975. doi: 10.1021/i160056a007. URL http://pubs.acs.org/doi/abs/10.1021/i160056a007.

William L. Luyben. The concept of "eigenstructure" in process control. *Industrial & Engineering Chemistry Research*, 27(1):206–208, 1988.

Henrik Manum and Sigurd Skogestad. Self-optimizing control with active set changes. *Journal of Process Control*, 22 (5):873 – 883, 2012.

Alejandro Marchetti, Benoit Chachuat, and Dominique Bonvin. A dual modifier-adaptation approach for real-time optimization. *Journal of Process Control*, 20(9):1027 – 1037, 2010.

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.

Miguel Mauricio-Iglesias, Ignacio Montero-Castro, Ane L. Mollerup, and Gürkan Sin. A generic methodology for the optimisation of sewer systems using stochastic programming and self-optimizing control. *Journal of Envi-*

ronmental Management, 155:193–203, May 2015. ISSN 03014797. doi: 10.1016/j.jenvman.2015.03.034. URL http://www.sciencedirect.com/science/article/pii/S0301479715001711.

Mihajlo D. Mesarović, D. Macko, and Yasuhiko Takahara. *Theory of Hierarchical Multilevel Systems*. Academic press, 1970.

Finn Are Michelsen, Berit Floor Lund, and Ivar Johan Halvorsen. Selection of optimal, controlled variables for the tealarc lng process. *Industrial & Engineering Chemistry Research*, 49(18):8624–8632, 2010. doi: 10.1021/ie100081j. URL http://dx.doi.org/10.1021/ie100081j.

Finn Are Michelsen, Lei Zhao, and Bjarne Foss. Operability analysis and control design of an {IRCC} {HMR} unit with membrane leakage. *Energy Procedia*, 23(0):171 – 186, 2012. ISSN 1876-6102. doi: http://dx.doi.org/10.1016/j.egypro.2012.06.041. URL http://www.sciencedirect.com/science/article/pii/S1876610212010818. The 6th Trondheim Conference on {CO2} Capture, Transport and Storage.

Manfred Morari, George Stephanopoulos, and Yaman Arkun. Studies in the synthesis of control structures for chemical processes. Part I: Formulation of the problem. Process decomposition and the classification of the control task. Analysis of the optimizing control structures. *AIChE Journal*, 26(2):220–232, 1980.

Lawrence T. Narraway and John D. Perkins. Selection of process control structure based on linear dynamic economics. *Industrial & Engineering Chemistry Research*, 32(11):2681–2692, 1993. doi: 10.1021/ie00023a035. URL http://pubs.acs.org/doi/abs/10.1021/ie00023a035.

L.T. Narraway, J.D. Perkins, and G.W. Barton. Interaction between process design and process control: economic analysis of process dynamics. *Journal of Process Control*, 1(5):243 – 250, 1991. ISSN 0959-1524. doi: http://dx.doi.org/10.1016/0959-1524(91)85015-B. URL http://www.sciencedirect.com/science/article/pii/095915249185015B.

Laura Niva, Enso Ikonen, and Jen Kovcs. Self-optimizing control structure design in oxy-fuel circulating fluidized bed combustion. *International Journal of Greenhouse Gas Control*, 43: 93 – 107, 2015. ISSN 1750-5836. doi: http://dx.doi.org/10.1016/j.ijggc.2015.10.012. URL http://www.sciencedirect.com/science/article/pii/S1750583615300992.

Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer, 2006.

Mehdi Panahi and Sigurd Skogestad. Economically efficient operation of {CO2} capturing process part i: Self-optimizing procedure for selecting the best controlled variables. *Chemical Engineering and Processing: Process Intensification*, 50(3):247 – 253, 2011. ISSN 0255-2701. doi: http://dx.doi.org/10.1016/j.cep.2011.02.005. URL http://www.sciencedirect.com/science/article/pii/S0255270111000444.

Mehdi Panahi and Sigurd Skogestad. Selection of controlled variables for a natural gas to liquids process. *Industrial & Engineering Chemistry Research*, 51(30):10179–10190, 2012.

Le Chi Pham and Sebastian Engell. Control structure selection for optimal disturbance rejection. In *Control Applications, (CCA) Intelligent Control, (ISIC), 2009 IEEE*, pages 707–712, July 2009. doi: 10.1109/CCA.2009.5280963.

Hans Pirnay, Rodrigo López-Negrete, and Lorenz T Biegler. Optimal sensitivity based on ipopt. *Mathematical Programming Computation*, 4(4):307–331, 2012.

Gade Pandu Rangaiah and Vinay Kariwala, editors. *Plantwide Control: Recent Developments and Applications*. John Wiley & Sons Ltd, 2012.

John E. Rijnsdorp. *Integrated process control and automation*. Elsevier, Amsterdam, 1991.

Kosan Roh and Jay H. Lee. Control Structure Selection for the Elevated-Pressure Air Separation Unit in an IGCC Power Plant: Self-Optimizing Control Structure for Economical Operation. *Industrial & Engineering Chemistry Research*, 53(18):7479–7488, May 2014. ISSN 0888-5885. doi: 10.1021/ie402909j. URL http://dx.doi.org/10.1021/ie402909j.

Mario E. Salgado and Arthur Conley. Mimo interaction measure and controller structure selection. *International Journal of Control*, 77(4):367–383, 2004. doi: 10.1080/00207170420001976 31. URL http://dx.doi.org/10.1080/00207170420001976 31.

Pär Samuelsson, Björn Halvarsson, and Bengt Carlsson. Interaction analysis and control structure selection in a wastewater treatment plant model. *IEEE Transactions on Control Systems Technology*, 13(6):955–964, Nov 2005. ISSN 1063-6536. doi: 10.1109/TCST.2005.854322.

Elena G. Sayalero, Sigurd Skogestad, Csar de Prada, J. Miguel Sola, and Rafael Gonzlez. Self-optimizing control for hydrogen optimization in a diesel hydrodesulfurization plant. In Iftekhar A. Karimi and Rajagopalan Srinivasan, editors, *11th International Symposium on Process Systems Engineering*, volume 31 of *Computer Aided Chemical Engineering*, pages 1647 – 1651. Elsevier, 2012. doi: http://dx.doi.org/10.1016/B978-0-444-59506-5.50160-7. URL http://www.sciencedirect.com/science/article/pii/B9780444595065501607.

Marc-Oliver Schach, Rdiger Schneider, Henning Schramm, and Jens-Uwe Repke. Control structure design for CO_2-absorption processes with large operating ranges. In *Proceedings of the 1st Post Combustion Capture Conference, Abu Dhabi, 2011*, pages 283–288, 2011.

Marc-Oliver Schach, Rüdiger Schneider, Henning Schramm, and Jens-Uwe Repke. Control structure design for CO_2-

absorption processes with large operating ranges. *Energy Technology*, 1(4):233–244, 2013. ISSN 2194-4296. doi: 10.1002/ente.201200057. URL http://dx.doi.org/10.1002/ente.201200057.

Hiroya Seki and Yuji Naka. Optimizing control of CSTR/distillation column processes with one material recycle. *Industrial & Engineering Chemistry Research*, 47(22):8741–8753, 2008. doi: 10.1021/ie800183a. URL http://dx.doi.org/10.1021/ie800183a.

Hamid Reza Shaker and Mohammad Komareji. Control configuration selection for multivariable nonlinear systems. *Industrial & Engineering Chemistry Research*, 51(25):8583–8587, 2012. doi: 10.1021/ie301137k. URL http://dx.doi.org/10.1021/ie301137k.

Mahdi Sharifzadeh. Integration of process design and control: A review. *Chemical Engineering Research and Design*, 91(12):2515 – 2549, 2013. ISSN 0263-8762. doi: http://dx.doi.org/10.1016/j.cherd.2013.05.007. URL http://www.sciencedirect.com/science/article/pii/S0263876213001986.

Reuel Shinnar. Chemical reactor modelling for purposes of controller design. *Chemical Engineering Communications*, 9(1-6):73–99, 1981.

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

Sigurd Skogestad. Control structure design for complete chemical plants. *Computers & Chemical Engineering*, 28(1-2): 219 – 234, 2004a.

Sigurd Skogestad. Near-optimal operation by self-optimizing control: from process control to marathon running and business systems. *Computers & Chemical Engineering*, 29(1):127 – 137, 2004b. ISSN 0098-1354. doi: http://dx.doi.org/10.1016/j.compchemeng.2004.07.011. URL http://www.sciencedirect.com/science/article/pii/S009813540400184X. {PSE} 2003.

Sigurd Skogestad and Ian Postlethwaite. *Multivariable Feedback Control, 1st Edition*. John Wiley & Sons, 1996.

Sigurd Skogestad and Ian Postlethwaite. *Multivariable Feedback Control, 2nd Edition*. John Wiley & Sons, 2005.

Sigurd Skogestad, Ivar Halvorsen, and John C Morud. Self-optimizing control: The basic idea and Taylor series analysis. In *AIChE Annual Meeting 1998, Miami Beach, paper 229c*, 1998.

Bala Srinivasan and Dominique Bonvin. Dynamic optimization under uncertainty via nco tracking: A solution model approach. *BatchPro Symposium*, pages 17–35, 2004.

Bala Srinivasan and Dominique Bonvin. Real-time optimization of batch processes by tracking the necessary conditions of optimality. *Industrial & Engineering Chemistry Research*, 46(2):492–504, 2007.

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, 2003a.

Bala Srinivasan, Srinivas Palanki, and Dominique Bonvin. Dynamic optimization of batch processes: I. characterization of the nominal solution. *Computers & Chemical Engineering*, 27(1):1 – 26, 2003b.

Bala Srinivasan, Lorenz T. Biegler, and Dominique Bonvin. Tracking the necessary conditions of optimality with changing set of active constraints using a barrier-penalty function. *Computers & Chemical Engineering*, 32:572–279, 2008.

Lia Maisarah Umar, Wuhua Hu, Yi Cao, and Vinay Kariwala. Selection of controlled variables using self-optimizing control method. In Gade Pandu Rangaiah and Vinay Kariwala, editors, *Plantwide Control:Recent Developments and Applications*, pages 43–71. John Wiley & Sons, 2012.

Lia Maisarah Umar, Yi Cao, and Vinay Kariwala. Incorporating feedforward action into self-optimising control policies. *The Canadian Journal of Chemical Engineering*, 92(1):90–96, 2014. ISSN 1939-019X. doi: 10.1002/cjce.21783. URL http://dx.doi.org/10.1002/cjce.21783.

Marc van de Wal and Bram de Jager. A review of methods for input/output selection. *Automatica*, 37(4):487 – 510, 2001. ISSN 0005-1098.

Suraj Vasudevan and Gade Pandu Rangaiah. Integrated framework incorporating optimization for plant-wide control of industrial processes. *Industrial & Engineering Chemistry Research*, 50(13):8122–8137, 2011. doi: 10.1021/ie1022139. URL http://dx.doi.org/10.1021/ie1022139.

Suraj Vasudevan, Gade Pandu Rangaiah, N. V. S. N. Murthy Konda, and Wee Hwa Tay. Application and evaluation of three methodologies for plantwide control of the styrene monomer plant. *Industrial & Engineering Chemistry Research*, 48(24):10941–10961, 2009. doi: 10.1021/ie900022h. URL http://pubs.acs.org/doi/abs/10.1021/ie900022h.

Adriaen Verheyleweghen and Johannes Jäschke. Self-optimizing control of a two-stage refrigeration cycle. *IFAC-PapersOnLine*, 49(7):845 – 850, 2016. ISSN 2405-8963. 11th {IFAC} Symposium on Dynamics and Control of Process SystemsIncluding Biosystems DYCOPS-CAB 2016Trondheim, Norway, 68 June 2016.

Matthew S. Walters, Joshua O. Osuofa, Yu-Jeng Lin, Thomas F. Edgar, and Gary T. Rochelle. Process control of the advanced flash stripper for {CO2} solvent regeneration. *Chemical Engineering and Processing: Process Intensification*, 107:21 – 28, 2016. ISSN 0255-2701. doi: http://dx.doi.org/10.1016/j.cep.2016.06.007. URL //www.sciencedirect.com/science/article/pii/S0255270116301490.

Waterloo Maple Inc. Maple user manual. Maplesoft, a division of Waterloo Maple Inc., 2005-2014, 2014.

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, 2013a.

Lingjian Ye, Vinay Kariwala, and Yi Cao. Dynamic optimization for batch processes with uncertainties via approximating invariant. In *Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on*, pages 1786–1791, June 2013b.

Lingjian Ye, Yi Cao, Xiushui Ma, and Zhihuan Song. A novel hierarchical control structure with controlled variable adaptation. *Industrial & Engineering Chemistry Research*, 53(2):1469514711, 2014.

Lingjian Ye, Yi Cao, and Xiaofeng Yuan. Global approximation of self-optimizing controlled variables with average loss minimization. *Industrial & Engineering Chemistry Research*, 54:12040–12053, 2015a.

Lingjian Ye, Zhihuan Song, and Xiushui Ma. Batch-to-batch self-optimizing control for batch processes. *Huagong Xuebao/CIESC Journal*, 66(7):2573–2580, 2015b. ISSN 04381157. doi: 10.11949/j.issn.0438-1157.20150247. URL `http://www.scopus.com/inward/record.url?eid=2-s2.0-84938150229&partnerID=tZOtx3y1`.

Ramprasad Yelchuru and Sigurd Skogestad. Convex formulations for optimal selection of controlled variables and measurements using mixed integer quadratic programming. *Journal of Process Control*, 22 (6):995 – 1007, 2012. ISSN 0959-1524. doi: http://dx.doi.org/10.1016/j.jprocont.2012.04.013. URL `http://www.sciencedirect.com/science/article/pii/S0959152412000984`.

Ramprasad Yelchuru and Sigurd Skogestad. Quantitative methods for regulatory control layer selection. *Journal of Process Control*, 23(1):58 – 69, 2013. ISSN 0959-1524. doi: http://dx.doi.org/10.1016/j.jprocont.2012.10.006. URL `http://www.sciencedirect.com/science/article/pii/S0959152412002351`.

Chang K. Yi and William L. Luyben. Evaluation of plant-wide control structures by steady-state disturbance sensitivity analysis. *Industrial & Engineering Chemistry Research*, 34(7):2393–2405, 1995. doi: 10.1021/ie00046a023. URL `http://pubs.acs.org/doi/abs/10.1021/ie00046a023`.

Xiaohong Yin, Shaoyuan Li, and Wenjian Cai. Enhanced-efficiency operating variables selection for vapor compression refrigeration cycle system. *Computers & Chemical Engineering*, 80: 1–14, September 2015. ISSN 00981354. doi: 10.1016/j.compchemeng.2015.05.005. URL `http://www.sciencedirect.com/science/article/pii/S0098135415001453`.

Muhammad Zaman, Jay Hyung Lee, and Rafiqul Gani. Carbon dioxide capture processes: Simulation, design and sensitivity analysis. In *Control, Automation and Systems (ICCAS), 2012 12th International Conference on*, pages 539–544, Oct 2012.

Alex Zheng, Rama V. Mahajanam, and James M. Douglas. Hierarchical procedure for plantwide control system synthesis. *AIChE Journal*, 45(6):1255–1265, 1999. ISSN 1547-5905. doi: 10.1002/aic.690450611. URL `http://dx.doi.org/10.1002/aic.690450611`.

Vitae

Johannes Jäschke is an Associate Professor at the Dept. of Chemical Engineering at the Norwegian University of Science and Technology (NTNU). He received the degree Dipl.-Ing. in Mechanical Engineering at RWTH Aachen University in 2007, and his PhD in Chemical Engineering at NTNU in 2011. He worked as a postdoc at NTNU and Carnegie Mellon University before taking his current position in 2014. His interests lie within the field of process systems engineering, with a strong focus on modelling, optimization and control. By linking methods from chemical engineering, optimization and control theory, his goal is to systematically develop practically applicable solutions for operating processes in a safe, reliable, and economical way.

Yi Cao is a Reader in Control Systems Engineering, Cranfield University. He obtained PhD in Control Engineering from the University of Exeter, UK in 1996, MSc in Industrial Automation from Zhejiang University, China in 1985. His main research interest is in developing systematic approaches to solve various operational problems involved in industrial processes using both models and data.

Vinay Kariwala is the R&D Manager for the Business Unit Measurements & Analytics in ABB India. Before that he was Group Manager for Control & Optimization group in Corporate Re-

search Centre at ABB, Bangalore. Prior to joining ABB in 2012, he was an Assistant Professor in Department of Chemical Engineering at Nanyang Technological University, Singapore. He got his Ph.D. degree from University of Alberta, Edmonton, Canada in 2004. His current research interests are in plantwide Control, process monitoring and particulate processes.



https://dspace.org/

Self-optimizing control – A survey

Jäschke, Johannes; Cao, Yi; Kariwala, Vinay

2017-04-04
Elsevier Science B.V., Amsterdam.
http://dx.doi.org/10.1016/j.arcontrol.2017.03.001
Jäschke J, Cao Y, Kariwala V, Self-optimizing control – A survey, Annual Reviews in Control, Vol. 43, 2017, pp. 199-223

Downloaded from a DSpace repository. https://dspace.org/