Plantwide control: The search for the self-optimizing control structure

Sigurd Skogestad*
Department of Chemical Engineering
Norwegian University of Science and Technology
N–7034 Trondheim Norway

October 15, 1999

Abstract

The following important question is frequently overlooked: Which variables should we control? The answer is related to finding a simple and robust way of implementing the optimal operating policy. The task is to translate the economic operating objectives into control objectives, and it is argued that the controlled variables should be those that give "self-optimizing control", which is when acceptable operation is achieved with constant setpoints for the controlled variables. A systematic procedure for finding controlled variables based on only steady-state information is presented and tested on a CSTR and a distillation case study.

Prepared for JPC (extended version of paper from IFAC World Congress, July 1999)

*E-mail: skoge@chembio.ntnu.no; phone: +47-7359-4154; fax: +47-7359-4080

1 Introduction

If we consider the control system in a chemical plant, then we find that it is structured hierarchically into several layers, each operating on a different time scale. Typically, layers include include scheduling (weeks), site-wide optimization (day), local optimization (hour), supervisory/predictive control (minutes) and regulatory control (seconds); see Figure 1. The task in each layer, e.g. optimization or control, may be performed by a computer algorithm or it may be performed manually (by an operator or engineer). The layers are interconnected through the controlled variables. More precisely,

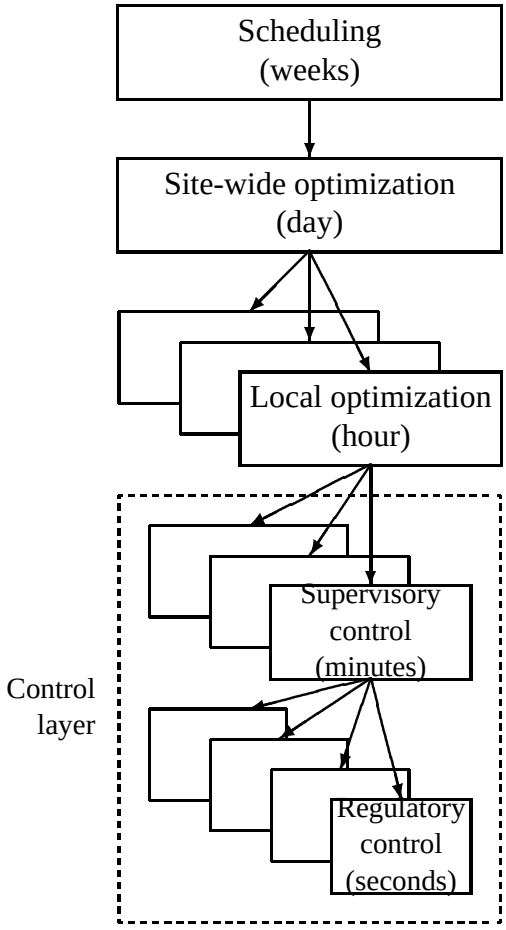


Figure 1: Typical control hierarchy in a chemical plant.

> The **controlled variables** (c) are the (internal) variables that link two layers in a control hierarchy, whereby the upper layer computes the setpoint (c_s) to be implemented by the lower layer.

We usually assume **time-scale separation** which for our purposes implies that the setpoints c_s can be assumed to be immediately implemented by the layers below.

Which should these internal controlled variables c be? That is, what should we control? Or to phrase the question in another way: Why do we in a chemical plant select to control a lot of internal variables (e.g. compositions, pressures, temperatures, etc.) for which there are noe explicit control requirements?

More generally, the issue of selecting controlled variables is the first subtask in the **control structure design** problem (Foss 1973); (Morari 1982); (Skogestad and Postlethwaite 1996) :

1. *Selection of controlled variables* c (variables with setpoints c_s)
2. *Selection of manipulated variables*

3. *Selection of measurements* v (for control purposes including stabilization)
4. *Selection of a control configuration* (structure of the controller that interconnects measurements/setpoints and manipulated variables)
5. *Selection of controller type* (control law specification, e.g., PID, decoupler, LQG, etc.).

Even though control engineering is well developed in terms of providing optimal control algorithms, it is clear that most of the existing theories provide little help when it comes to making such structural decisions.

The method presented in this paper for selecting controlled variables follows the ideas of Morari *et al.* (1980), Skogestad and Postlethwaite (1996) and Zheng *et al.* (1999) and is very simple. The basis is to define mathematically the quality of operation in terms of a scalar cost function J to be minimized. To achieve truly optimal operation we would need a perfect model, we would need to measure all disturbances, and we would need to solve the resulting dynamic optimization problem on-line. This is unrealistic in most cases, and the question is if it is possible to find a simpler implementation which still operates satisfactorily (with an acceptable loss). More preciely, the **loss** L is defined as the difference between the actual value of the cost function obtained with a specific control stragegy, and the truly optimal value of the cost function, i.e. $L = J - J_{\text{opt}}$. The simplest operation would result if we could select controlled variables such that we obtained acceptable operation with constant setpoints, thus effectively turning the complex optimization problem into a simple feedback problem and achieve "self-optimizing control":

> **Self-optimizing control** is when we can achieve an acceptable loss with constant setpoint values for the controlled variables

(The reader is probably familiar with the term self-regulation, which is when acceptable dynamic control performance can be obtained with constant manipulated inputs. Self-optimizing control is a direct generalization to the case where we can achieve acceptable (economic) performance with constant controlled variables.) The term "self-optimizing control" is short and descriptive, but also other terms have been used to describe the same idea, such as "feedback optimizing control" (Morari *et al.* 1980), "indirect optimizing control (through setpoint control)"(Halvorsen and Skogestad 1998), and "partial control"(Arbel *et al.* 1996).

Inspired by the work of Findeisen *et al.* (1980)), Morari *et al.* (1980) gave a clear description of what we here denote self-optimizing control, including a procedure for selecting controlled variables based on evaluating the loss. However, it seems that nobody followed up on their ideas. One reason was probably that no good example was given in the paper for how to choose controlled variables for the unconstrained case.

The main objective of this paper is to demonstrate, with a few examples, that the issue of selecting controlled variables is very important and to provide a systematic procedure for selecting controlled variables. A discussion of the literature is given at the end of the paper.

2 Optimization and control

When controlling a chemical plant our first concern is to stabilize the plant and keep the operation within given constraints. These issues may consume some degrees of freedom (e.g. to stabilize levels with no steady-state effect and to satisfy exact product specifications), but there will generally be many degrees of freedom u left. What should these be used for?

Loosely speaking, they should be used to "optimize the operation". There may be many issues involved, and to trade them off against each other in a systematic manner we usually quantify a scalar performance (cost) index J which should be minimized. In many cases this index is an economic measure, e.g. the operation cost. For example, J could be of the form

$$J(u, d) = \int_0^T \phi(u, d) dt \tag{1}$$

where u are degrees of freedom, d are time-varying disturbances, and T is the total operation time.

In this paper we will for simplicity use steady-state models and the integration in (1) may be replaced by time-averaging over the various steady-states. The main justification for using a steady-state analysis is that the economic performance is primarily determined by steady-state considerations. The effect of the dynamic control performance can be partly included in the economic analysis by introducing a control error term as an additional disturbance.

2.1 The optimization problem

The optimizing control problem can be formulated as

$$\min_u J(u, d) \tag{2}$$

subject to the inequality constraints

$$g(u, d) \leq 0 \tag{3}$$

where u are the N_u independent variables we can affect (degrees of freedom), and d are independent variables we can not affect (disturbances). Here the constraints for instance may be

- product specifications (e.g. $x_D \geq 0.95$)
- manipulated variable constraints (e.g. $0 \leq V \leq V_{max}$)
- other operational limitations (e.g. maximum pressure or avoid flooding)

Conflicting constraints may result in a problem without a feasible solution. For example, if we make a product by blending two streams then we cannot achieve a product specification outside the range of feed compositions.

We have assumed that all dependent (state) variables have been eliminated such that the cost function and constraints are in terms of the independent variables u and d. However, in some cases it is more convenient to keep the state variables x and corresponding model equations (equality constraints) and formulate the optimization problem as

$$\min_u J(x, u, d) \tag{4}$$

subject to the constraints

$$g_1(x, u, d) = 0 \tag{5}$$

$$g_2(x, u, d) \leq 0 \tag{6}$$

Although we use a steady-state analysis in this paper, it may be extended to truly unsteady-state-processes, like during a grade transition or for a batch process, by using a dynamic model and letting the setpoints c_s be precomputed trajectories as a function of time or of state variables.

2.2 Implementing the optimal solution

There are two main issues when it comes to optimizing control. The first is the mathematical and numerical problem of solving the optimization problem in (2) to obtain the optimal operating point. The optimization problem may be very large, with hundreds of thousands of equations and hundreds of degrees of freedom (e.g. for a complete ethylene plant), but with todays computers and optimization methods this problem is solvable, and it is indeed solved routinely today in some plants. The second issue, the focus of this paper, is how the optimal solution should be implemented in practice. Surprisingly, this issue has received much less attention.

To better understand the issues consider the three alternative structures for optimizing control shown in Figure 2:

(a) Open-loop implementation.

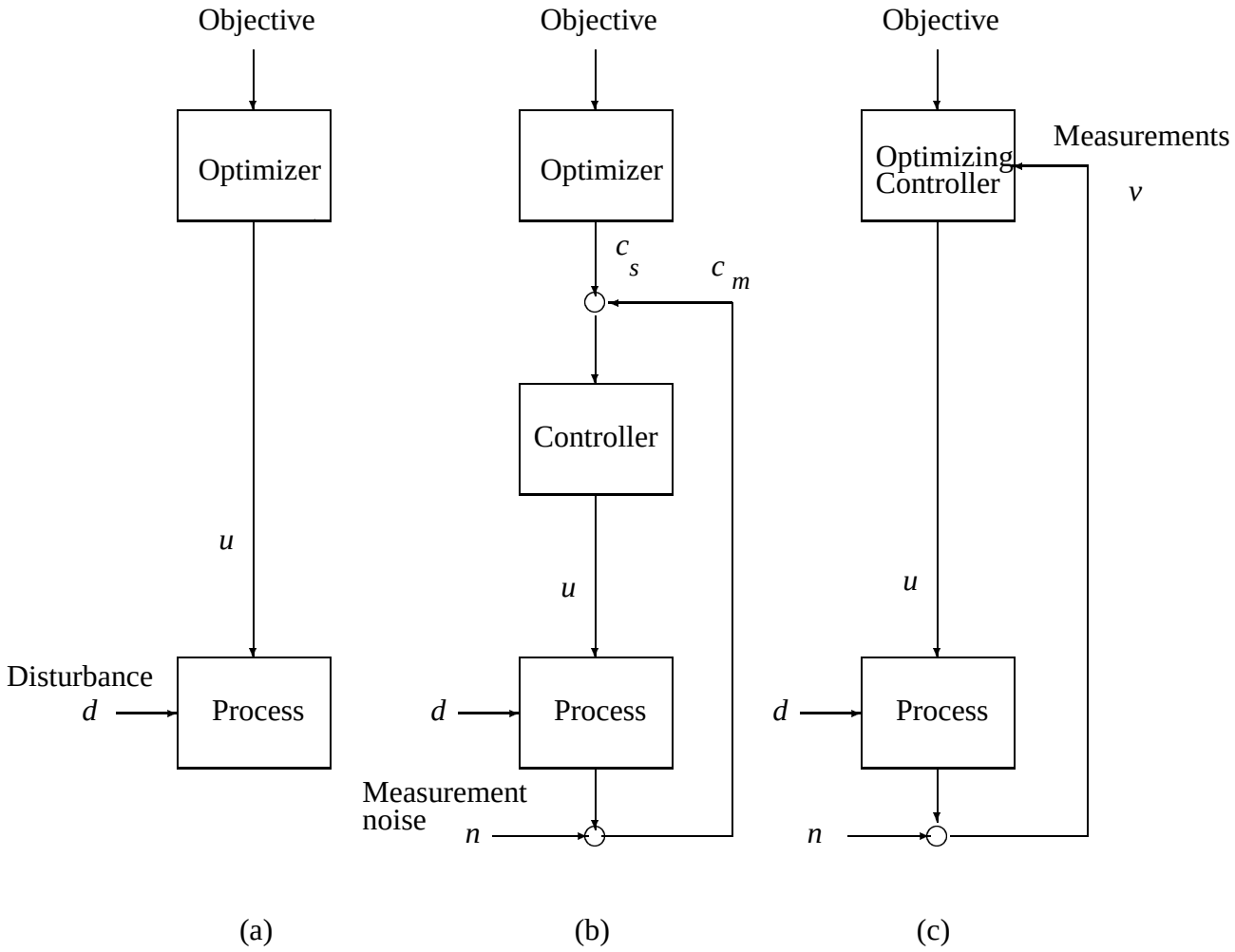


Figure 2: Alternative structures for optimizing control. Structure (b) is studied in this paper

(b) Closed-loop implementation with a separate control layer.

(c) Integrated optimization and control.

In the figure the block "process" denotes the process as seen from the optimization layer, so it may actually be a partially controlled plant which includes, for example, stabilizing level loops. This means that the independent variables u for the optimization may include some of the "original" manipulated variables (m) as well as the setpoints (e.g. level setpoints) used in the lower-layer controllers.

If there were no unmeasured disturbances d then the three implementations would give the same result. The key to understand why the three structures differ, is therefore to consider what happens to the degrees of freedom u in response to a disturbance d (more generally d may include any uncertainty, including errors in the model used for the optimization and control).

(a) For the open-loop implementation there is no feedback, and u remains unchanged when there is a disturbance d.

(b) For the closed-loop implementation with a separate control layer, the disturbance d affects the measurement c_m, and the controller adjust u so that c_m returns approximately to its setpoint, c_s.

(c) For the integrated optimization and control, all available measurements v (including c_m) are used to identify the disturbance and update the model, and then dynamic on-line optimization is used to recompute a new optimal value for u

In general, the open-loop implementation (a) is not acceptable because there is no attempt to correct for disturbances d

If we formulate the optimizing control problem in the usual mathematical fashion as in (2), then we find that the centralized implementation (c) is the optimal choice. Here there is one "big" controller, which based on all available measurements and other given information (including a model of the system and expected uncertainty), continuouosly computes the optimal values of all manipulated variables. However, there are fundamental reasons why such a solution is not the best, even with todays and tomorrows computing power. One important reason is the cost of obtaining a dynamic model; in the centralized controller it is critical that this model is accurate since there are no predetermined links, and the controller must rely entirely on the model to take the right action.

Therefore, in practice, we almost always use the closed-loop implementation (b) where we decompose the control system into separate parts and layers. In the simplest case shown in Figure 2b we may have two layers:

- A steady-state optimization layer which computes the optimal setpoints c_s for the controlled variables, and
- A feedback control layer which implements the setpoints, to get $c \approx c_s$.

In process control applications, the feedback control layer usually operates continuously, whereas the optimization layer (which may be an engineer) recomputes new setpoints c_s only quite rarely; maybe once an hour or once a day (when the plant has settled to a new steady-state). The idea is that by locally controlling the right variables c, we can take care of most of the disturbances, and thus reduce the need for continouos reoptimization. This also reduces the need for model information and tends to make the implementation more robust. On the other hand, it usually implies a performance loss compared to the "true" optimal (centralized) solution, and the challenge is to find a "self-optimizing" control structure (i.e. to find the right controlled variables c) for which the loss L is acceptable.

Active constraint control

In some cases there is no performance loss with the hierarchical structure in Figure 2b with a separate optimization and control layer. This is when the optimum lies at some constraints, and we use *active constraint control* where we choose the constrained variables as the controlled variables c (Maarleveld and Rijnsdrop 1970) (Arkun and Stephanopoulos 1980) (Fisher *et al.* 1988). However, in many cases the constraints move depending on the operating point, and a change in the active constraints requires reconfiguration of the loops. To avoid such an often complicated logic system, we may use in the lower layer a multivariable controller that can handle constraints. In particular, model predictive control (MPC), which has gained widespread use in industry over the last 20 years, provides a simple and efficient tool for tracking active constraints.

2.3 Degrees of freedom for control and optimization

A starting point for control and optimization is to establish the number of degrees of freedom for operation; both dynamically (for control, N_m) and at steady-state (for optimization, N_u). Fortunately, it is in most cases relatively straightforward to establish these numbers from process insight, e.g. see Ponton and Liang (1993) and Luyben (1996). The basis is that the number of independent variables for control (N_m) equals the number of variables that can be manipulated by external means (which in process control is the number of number of adjustable valves plus the number of other adjustable electrical and mechanical variables).

The number of degrees at freedom at steady-state (N_u) is generally less than the number of control degrees of freedom (N_m). We have

$$N_u = N_m - N_0 \tag{7}$$

where $N_0 = N_{m0} + N_{y0}$ is the number of variables with no steady-state effect (on the cost function). Here

N_{m0} - number of manipulated inputs (u's), or combinations thereof, with no steady-state effect.

N_{y0} - number of controlled output variables with no steady-state effect.

A simple example where N_{m0} is non-zero is a heat exchanger with bypass on both sides, (i.e. $N_m = 2$). However, at steady-state $N_u = 1$ since there is really only one operational degree of freedom, namely the heat transfer rate Q (which at steady-state may be achieved by many combinations of the two bypasses), so we have $N_{m0} = 1$.

N_{y0} usually equals the number of liquid levels with no steady-state effect (including most buffer tank levels). However, note that some liquid levels *do* have a steady-state effect, such as the level in a non-equilibrium liquid phase reactor, and levels associated with adjustable heat transfer areas. Also, we should *not* include in N_{y0} any liquid holdups that are left uncontrolled, such as internal stage holdups in distillation columns.

Remark on design degrees of freedom. In this paper we are concerned with *operational* degrees of freedom. The *design* degrees of freedom (which is not a concern in this paper) includes all the operational degrees of freedom plus all parameters related to the size of the equipment, such as the number of stages in column sections, area of heat exchangers, etc.

2.4 Introductory example

To give the reader som appreciation of the issues, we here consider a distillation plant. With a given feed stream and a specified pressure[1], a conventional two-product distillation column, as shown in Figure 3, has two degrees of freedom at steady state ($N_u = 2$). (From a control point of view the column has $N_m = 5$ degrees of freedom, but two degrees of freedom are needed to stabilize the reboiler and condenser holdups, which have no steady-state effect, and one degree of freedom is used to control the pressure at its given value). The two remaining degrees of freedom, e.g. selected to be the reflux flow L and the distillate flow D,

$$u = \begin{pmatrix} L \\ D \end{pmatrix}$$

(this is not a unique choice) may be used to optimize the operation of the plant. However the question is: How should the optimal solution be implemented, that is, which two variables c should be specified and controlled during operation?

To answer this question in a quantitative manner, we need to define the constraints for the operation and the cost function J to be minimized.

Constraints. We assume that the distillate product must contain at least 95% light component and that to avoid flooding the capacity of the column is limited by a maximum allowed vapor load,

$$x_D \geq x_{D,min} = 0.95$$

$$V \leq V_{max}$$

Cost function. Rather than minimzing the cost J, it is more natural in this case to maximize the profit $P = -J$, which is the product value minus the feed costs and the operational (energy) costs which are proportional to the vapor flow V,

$$P = p_D D + p_B B - p_F F - p_V V \tag{8}$$

Constrained operation. Let us first consider a case where

- distillate is the more valueable product ($p_D \gg p_B$)
- energy costs are low ($p_V \approx 0$)

[1] If column pressure is free we often find that the optimal choice is to have maximum cooling corresponding to minimum pressure ("floating pressure control" as suggested by Shinskey (1984)). The reason is that in most cases the relative volatility is improved when pressure is lowered.

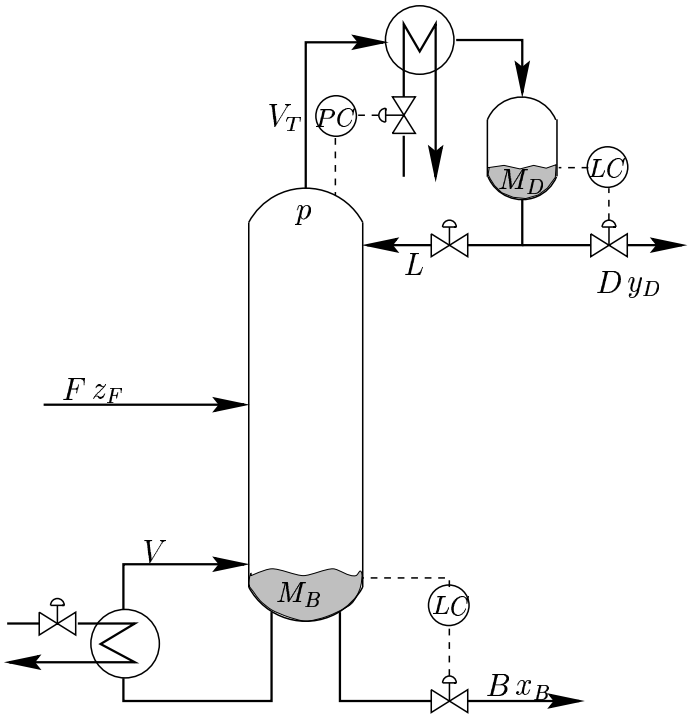


Figure 3: Typical distillation column controlled with the LV-configuration

In this case, it is optimal to operate the column at maximum load (to reduce the loss of light component in the bottom) and with the distillate composition at its specification (to maximize distillate flow by including as much heavy component as possible) (Gordon 1986), i.e.

$$V_{opt} = V_{max}$$

$$x_{D,opt} = x_{D,min} = 0.95$$

Thus, the optimum lies at constraints and implementation is obvious: We should select the vapor rate V and the distillate composition x_D as the controlled variables,

$$c = \begin{pmatrix} V \\ x_D \end{pmatrix}; \quad c_s = \begin{pmatrix} V_{max} \\ x_{D,min} \end{pmatrix}$$

In practice, we implement this using a lower-level feedback control system where we adjust the boilup V to keep the pressure drop over the column (an indicator of flooding) below a certain limit, and adjust the reflux L (or some other flow, depending on how the level and pressure control system is configured) so that x_D is kept constant.

Unconstrained operation. Next, consider a case where

- bottoms product is the more valuable product ($p_B \gg p_D$)
- energy costs are relatively high (the term $p_V V$ contributes significantly to J)

In this case the optimum may be unconstrained in both variables, and assume for the discussion that

$$V_{opt} = 0.76 V_{max}$$

$$x_{D,opt} = 0.973 > x_{D,min}$$

Implementation in this case is *not* obvious. Some candidate sets of controlled variables are

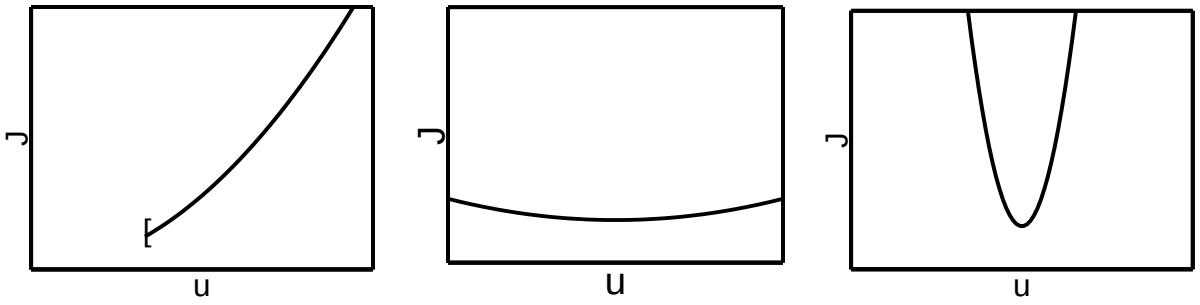
$$c_1 = \begin{pmatrix} x_D \\ x_B \end{pmatrix};\ c_2 = \begin{pmatrix} T_{top} \\ T_{btm} \end{pmatrix};\ c_3 = \begin{pmatrix} V \\ x_D \end{pmatrix};\ c_4 = \begin{pmatrix} L \\ V \end{pmatrix};\ c_5 = \begin{pmatrix} L/D \\ V/B \end{pmatrix}$$

and there are many others. Controlled variables c_1 and c_2 will yield a "two-point" control system where we close two loops for quality control; c_3 yields a "one-point" control system where only one quality loop is closed; whereas c_4 and c_5 are "open-loop" policies which require no additional feedback loops (except for the level and pressure loops already mentioned). All of these choices of controlled variables will have different self-optimizing control properties.

At the end of the paper, we study another distillation example, where the optimum is constrained in one variable and unconstrained in the other.

2.5 Summary so far

In summary, as illustrated in Figure 4, we have three classes of problems when implementing the optimal solution:



(a) Easy: Active constraint control

(b) Easy: Manual control

(c) Difficult.

Figure 4: Implementing the optimal solution

(a) *Constrained optimum.* In the figure is shown the case where the solution is constrained in u, i.e. the minimum J is obtained for u_{min}, but it could be in any other variable. As discussed above, implementation of the optimal solution in this case is easy; we implement the active constraint (this could involve a feedback loop). This case is very common, as the optimum is always constrained if we have a linear model with a linear objective function. Model predictive control (MPC) or a similar method may be used if the active constraint moves dependeing on the operating point.

(b) *Unconstrained flat optimum.* In this case the solution (operation) is insensitive to value of the independent variable u, and implementation is again easy; we may use the open-loop strategy with a constant value for u.

(c) *Unconstrained sharp optimum.* The more difficult problems for implementation is when the solution (operation) is sensitive to value of the independent variable u. In this case, we want to find a controller variable c in which the optimum is flatter.

The latter "difficult" problems are the focus of this paper.

3 Selection of controlled variables

In this section we present our procedure, but let us first formulate the problem a bit clearer.

3.1 Problem formulation

Let the "base set" for the N_u available degrees of freedom be denoted u (this is not a unique set), and let d denote the (important) disturbances (this is a unique set). Any other variable is then a function of these variables, for example, we can write for the controlled variables

$$c = f(u, d) \tag{9}$$

Assuming that the set of c's are independent, we can also derive the inverse relationship

$$u = f^{-1}(c, d) \tag{10}$$

where we assume the function f^{-1} exists and is unique.

For a given disturbance d we can solve the optimization problem (2) with constraints (3), and if a feasible solution exists obtain the optimal value $u_{opt}(d)$,

$$\min_{\substack{u \\ g(u,d)\leq 0}} J(u, d) = J(u_{opt}(d), d) = J_{\text{opt}}(d) \tag{11}$$

However, in actual operation the value of u will differ from the the optimal value $u_{\text{opt}}(d)$, and this results in a **loss** [2] L between the actual operating costs and the optimal operating costs,

$$L(u, d) = J(u, d) - J(u_{opt}(d), d) \tag{12}$$

The magnitude of the loss will depend on the control strategy used for adjusting u during operation, and to understand this better consider the "open-loop" and "closed-loop" strategies. Let d_0 denote the nominal value of the disturbance for which the optimization was performed, and let d denote its actual value (during operation). In this paper we use for simplicity the nominal optimal values as setpoints, i.e.

$$u_s = u_{\text{opt}}(d_0); \quad c_s = c_{\text{opt}}(d_0)$$

- In the "open-loop" strategy we attempt to keep u constant at its setpoint u_s (more precisely, $u = u_s + d_u$ where d_u is the implementation error)
- In the "closed-loop" strategy we adjust $u = f^{-1}(c, d)$ in a feedback fashion in an attempt to keep c constant at its setpoint c_s (more precisely, $c = c_s + d_c$, where d_c is the implementation error for control of c)

The open-loop policy is often poor; both because the optimal input $u_{opt}(d)$ depends strongly on the disturbance (so it is not a good policy to keep u_s constant), and because we are not able to implement u accurately (so the implementation error d_u is large).

The question we want to answer is therefore: *What is best choice for the controlled variables c to use in the closed-loop policy?* If we allow for combinations of measurements, then there are infinitely many choices. Specifically, if y_m represent all the candidate measured variables, then we can write

$$c = h(y_m, u) \tag{13}$$

where the function h is free to choose (note that the open-loop policy is included as the special case $h(y_m, u) = u$).

We next present two approaches for selecting controlled variables. We first consider the error, $u - u_{opt}(d)$, and based on this propose four requirements of the controlled variable. We then consider a more exact stepwise procedure based on explicitly evaluating the loss.

[2] It is not really necessary to introduce the loss function, and we may instead work directly with the actual cost $J(u, d)$. However, the loss provides a better "absolute scale" on which to judge whether a given set of controlled variables c is "good enough", and thus is self-optimizing.

3.2 Requirements for the controlled variable

Consider an closed-loop implementation where we attempt to keep c constant at the value c_s. With this implementation the operation may be non-optimal (with a positive loss) due to the presence of a setpoint error and an implementation error.

1. The setpoint error, $e_{cs} = c_s - c_{opt}(d)$, is the difference between the setpoint and truly optimal value
2. The implementation error, $d_c = c - c_s$, is the difference between the actual value and the setpoint.

These two errors are generally independent; the setpoint error is caused by disturbances (changes in the operating point), whereas the implementation error is caused by measurement error and poor control. The overall error $e_c = c - c_{opt}(d)$ (which causes a positive loss), is then the sum of the two,

$$e_c = e_{cs} + d_c \tag{14}$$

Clearly, we want e_c to be small. In addition, we would like that a large value of e_c results in only a small value of the "base set" error e_u, that is, we want u to be insensitive to changes in c (or equivalently, we want c to be sensitive to changes in u).

From this, we can derive the following *requirements for of a good candidate controlled c variable* (also see Skogestad and Postlethwaite (1996), page 404):

Requirement 1. Its optimal value is insensitive to disturbances (so that the setpoint error $e_{cs} = c_s - c_{opt}(d)$ is small)

Requirement 2. It is easy to measure and control accurately (so that the implementation error d_c is small)

Requirement 3. Its value is sensitive to changes in the manipulated variables u, that is, the gain from u to y is large (so that even a large error in the controlled variable c results in only a small error in u).

Requirement 4. For cases with two or more controlled variables, the selected variables should not be too closely correlated.

In short, we should select variables c for which the variation in optimal value and implementation error is small compared to their adjustable range (the range c may reach by varying u) (Skogestad and Postlethwaite (1996), page 408).

For small variations we may use a linearized relationship between the "base set" and controlled variables, $c = Gu$, where G is the gain matrix. If we assume that each controlled variable c has been scaled such that the sum of its optimal range and its implementation error is 1 (this takes care of requirements 1 and 2), and that each "base variable" u has been scaled such that a unit change has the same effect on the cost function J, then we should prefer sets of controlled variable for which $\underline{\sigma}(G)$ is maximized (maximizing the gain takes care of requirement 3 and using the minimum singular value $\underline{\sigma}(G)$ takes care of requirement 4) (Skogestad and Postlethwaite (1996), page 406).

Note that requirement 1 says that its *optimal* value should be insensitive to disturbances, and *not* that its value should be insensitive to disturbances. Actually, we usually want its value to be sensitive to disturbances so that we can detect them (and thus correct for them).

All four requirements should be satisfied. For example, assume we have a mixture of three components, and we have a measurement of the sum of the three mole fractions, $c = x_A + x_B + x_C$. This measurement is always 1 and thus independent of disturbances (so requirement 1 is satisfied), but it is of course not a suitable controlled variable because it is also insensitive to the manipulated variables u (so requirement 3 is not satisfied). Requirement 3 also eliminates variables that have an extremum (maximum or minimum) when the cost has its minimum, because the variable is then insensitive to changes in u.

The requirements listed above are very useful for identifying good candidate controlled variables (also called "dominant" variable by some authors). However, for a more exact evaluation one should use

the procedure descibed next. It based on evaluating the loss imposed by keeping the selected controlled variables constant.

3.3 Stepwise procedure for evaluating the loss

To compare alternative choices for c it is probably simplest to directly evaluate the cost function (or equivalently the loss function) for expected values of the disturbance d and the implementation error d_c. A good choice for of controlled variables c (i.e. good choice of the function h) is then the one that with constant values of c (more precicely, $c = c_s + d_c$) gives an acceptable value of the loss L for the expected set of disturbances $d \in \mathcal{D}$, and expected set of implementation (control) errors $d_c \in \mathcal{D}_c$.

This results in the following procedure for selecting controlled variables c:

Step 1: Degree of freedom analysis. Determine the number of degrees of freedom (N_u) available for optimization, and identify a base set (u) for the degrees of freedom.

Step 2: Cost function and constraints. Define the optimal operation problem by formulating a scalar cost function J to be minimized for optimal operation, and specify the constraints that need to be satisfied.

Step 3: Identify the most important disturbances (uncertainty). These may be caused by

- Errors in the assumed (nominal) model used in the optimization (including the effect of incorrect values for the nominal disturbances d_0 used in the optimization)
- Disturbances ($d - d_0$) (including parameter changes) that occur after the optimization
- Implementation errors (d_c) for the controlled variables c (e.g. due to measurement error or poor control)

From this one defines the set of disturbances $\mathcal{D}$ and set of implementation errors $\mathcal{D}_c$ to be considered. Often it is a finite set of disturbance combinations, for example, consisting of the extreme values for the individual disturbances. In addition, one must determine how to evaluate the mean cost function J_{mean}. There are many possibilities, for example

1. Average cost for a finite set of disturbances
2. Mean cost from *Monte-Carlo* evaluation of a distribution of d and d_c.
3. Worst-case loss

Step 4: Optimization.

1. First solve the nominal optimization problem, that is, find $u_{opt}(d_0)$. From this may one also obtain a table with the nominal optimal values for all other variables (including the candidate controlled variables).
2. In most cases (unless it involves too much effort) we then solve the optimization problem for the disturbances d in question (defined in step 3). This is needed to check whether there exists a feasible solution $u_{opt}(d)$ for all disturbances d, and to find the optimal cost $J(u_{opt}, d)$ needed if we want to evaluate the loss. It may also be used in step 5 to identify controlled variables.

Step 5: Identify candidate controlled variables. Typically, these are measured variables or simple combinations thereof. The requirements given above are useful in identifying good candidates. For example, based on the optimization in step 4, one may look for variables which optimal value is only weakly dependent of disturbances (requirement 1). The variable should also be easy to control and measure (requirement 2), and it should be sensitive to changes in the manipulated inputs (requirement 3). Insight and experience may also be helpful at this stage, because the possible number of combinations may be extremely large.

Step 6: Evaluation of loss. We compute the mean value of the loss for alternative sets of controlled variables c. This is done by evaluating the loss

$$L(u,d) = J(u,d) - J(u_{opt}(d),d); \quad u = f^{-1}(c_s + d_c, d) \tag{15}$$

with fixed setpoints c_s for the defined disturbances $d \in \mathcal{D}$ and implementation errors $d_c \in \mathcal{D}_c$. We usually select the setpoints as the nominal optimal values, $c_s = c_{opt}(d_0)$, but it is also possible to let the value of c_s be subject to an optimization.

Step 7: Further analysis and selection. We select for further consideration the sets of controlled variables with acceptable loss (and which thus yield self-optimizing control). These could then be analyzed to see if they are adequate with respect to other criteria that may be relevant, such like the region of feasibility and the expected dynamic control performance (input-output controllability)

3.4 Toy example

To give a simple "toy example", let $J = (u-d)^2$. For this problem we always have $J_{opt}(d) = 0$ corresponding to $u_{opt}(d) = d$. Let is now consider three alternative choices for the controlled output (e.g. we can assume they are three alternative measurements)

$$c_1 = 0.1(u-d); \quad c_2 = 20u; \quad c_3 = 10u - 5d$$

Since $u_{opt}(d) = d$, we have that the optimal value for the three alternative controlled outputs as a function of the disturbance are

$$c_{1opt}(d) = 0; \quad c_{2opt}(d) = 20d; \quad c_{3opt} = 5d$$

In the nominal case $d_0 = 0$ and we have in all three cases that $c_{opt}(d_0) = 0$, so we select in all three cases $c_s = 0$. We assume that the variables have been scaled such that $|d| \leq 1$ (disturbance) and $|d_{ci}| \leq 1, i = 1,2,3$ (i.e. same implementation error for all three variables).

Let us first evaluate how the three candidate variables meet the above requirements.

1. *Its optimal value is insensitive to disturbances.* From this point of view, the preferred controlled variable is c_1 (zero sensitivity), followed by c_3 (sensitivity 5) and c_2 (sensitivity 20).
2. *It is easy to control accurately.* There is no difference here since the implementation error d_c is the same for the three variables.
3. *Its value is sensitive to changes in* u. This favors c_2 (gain 20), followed by c_3 (gain 10), whereas variable c_1 (gain 0.1) is very poor in this respect.

Let us next evaluate the losses. For this simple example the losses can be evaluated analytically, and we find for the three alternatives

$$L_1 = (10d_{c1})^2; \quad L_2 = (0.05d_{c2} - d)^2; \quad L_3 = (0.1d_{c3} - 0.5d)^2$$

(For example, for c_3 we have $u = (c_3 + 5d)/10$ and with $c_3 = c_{3s} + d_{c3} = d_{c3}$ we get $J = (u-d)^2 = (0.1d_{c3} + 0.5d - d)^2$). With $|d| = 1$ and $|d_c| = 1$ the worst-case values of the losses are

$$L_1 = 100; \quad L_2 = 1.05^2 = 1.1025; \quad L_3 = 0.6^2 = 0.36$$

and we find that *output c_3 is the best overall choice for self-optimizing control* (with the smallest loss), and c_1 is the worst. (This is the same conclusion that followed by considering the three requirements.) We see that with no implementation error ($d_c = 0$) c_1 would be the best, and with no disturbance ($d = 0$) c_2 would be the best.

4 Reactor case study

We here illustrate the stepwise procedure for selecting controlled variables with a simple example that can easily be reproduced by the reader. Consider a continuously stirred tank reactor (CSTR) where two irreversible first-order reactions take place

$$A \to B; \quad r_A = k_A x_A [\mathrm{s}^{-1}]$$

$$B \to C; \quad r_B = k_B x_B [\mathrm{s}^{-1}]$$

Component B is the desired product and its concentration as a function of the residence time has a peak value, at which we want to operate the reactor.

Model. Let z_i and x_i denote mole fractions of component i in the feed and reactor, respectively, and let F [mol/s] be the feed rate and M [mol] the reactor holdup. There are only three components, A, B and C, and steady-state material balances yield

$$z_A F - x_A F - k_A x_A M = 0$$

$$z_B F - x_B F + k_A x_A M - k_B x_B M = 0$$

$$x_C = 1 - x_A - x_B$$

The feed contains only components A and C, i.e. $z_C = 1 - z_A$. We consider the following nominal data:

$$z_A = 0.8; \; k_A = 1 s^{-1}; \; k_B = 1 s^{-1}; \; F = 1 \mathrm{mol/s}$$

Step 1: Degree of freedom analysis

With a given feed the reactor has one degree of freedom at steady-state, which may be selected as the reactor holdup, i.e

$$u = M \quad [\mathrm{mol}]$$

The value of M should be adjusted to optimize the operation.

Step 2: Cost function and constraints

In this example component B is the desired product and the objective is to maximize the concentration of B, i.e. we choose the cost function

$$J = -100 \cdot x_B$$

(in most cases we would recycle unreacted A, but this is not the case in this example). We would like to find a controlled variable which results in a mean loss of less than 0.5 when there are disturbances.

There are no extra constraints, except for physical constraints such as $0 \leq M < \infty$.

Step 3: Disturbances

We will consider the following disturbances (errors):

- d_1: Feed rate reduced by 30%
- d_2: Feed fraction of A reduced from 0.8 to 0.6
- d_3: Feed fraction of A increased from 0.8 to 10.0
- d_4: Rate constant k_A increased by 50%
- d_5: Rate constant k_B increased by 50%
- d_c: Implementation error for the controlled variable, e.g., due to measurement error: see step 5.

The mean loss is here chosen as the average of the loss resulting from each of these six disturbances (one at a time).

Step 4: Optimization

In many cases it is optimal to operate the reactor with maximum holdup M, for example, this would be the case if the objective were to maximize the production or concentration of component C. However, in our case we want to maximize the concentration of the intermediate product B goes through a maximum as we increase the holdup M (see Figure 5 which shows the cost function $J = -100x_B$ as a function of M).

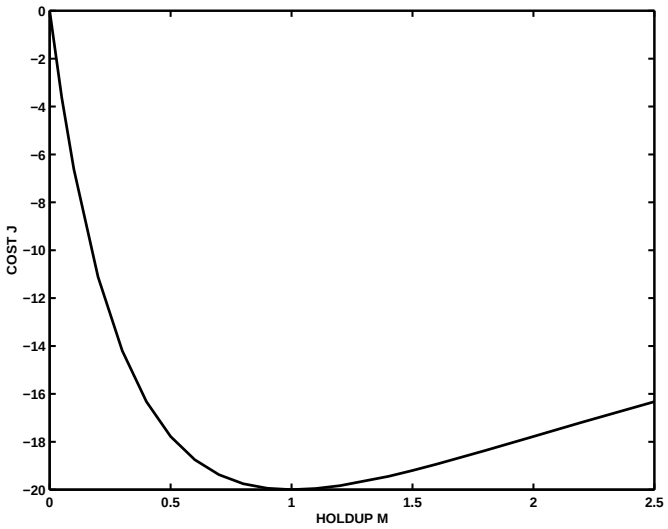


Figure 5: Reactor case study: Cost function J as a function of holdup M

The optimal holdup and corresponding optimal compositions in the nominal case are:

$$\text{Nominal optimum}: \quad M = 1.0\text{mol}; \quad x_A = 0.4, x_B = 0.2, x_C = 0.4$$

corresponding to $J = -100x_B = -20$. When there are disturbances, the optimal values change as given in Table 1.

Step 5: Candidate controlled variables

As mentioned, the reactor has one steady-state degree of freedom during operation. How should this degree of freedom be set? The following candidates for the controlled variable c are suggested

- $c_1 = M$ (holdup)
- $c_2 = M/F$ (residence time)
- $c_3 = x_A$
- $c_4 = x_B$
- $c_5 = x_C$
- $c_6 = x_B/x_A$
- $c_7 = \theta_1 = x_A + 2x_B + 3x_C$ (a property variable)
- $c_8 = \theta_2 = x_A + 3x_B + 2x_C$ (a property variable)

These are essentially the available measurements in the reactor. In alternatives 7 and 8 the property variables may represent a boiling temperature, a viscosity, a refraction index or similar. We select the setpoint for each variable as the nominally optimal value.

We need to identify the implementation error d_c for each of the candidate controlled variables. Assuming that the implementation error is mainly to due to measurement errors and that the measurement error is 10% for M and F, and 5% for the mole fractions, we use the following values for d_c:

- M: 10%
- M/F: 20%
- x_A, x_B and x_C: 5%
- x_A/x_B: 10%
- θ_1 and θ_2: 0.1 unit (about 5%)

Which controlled variable is preferred? It seems clear that it will be better to keep M/F rather than M constant, because the optimal residence time M/F is independent of the feed rate, whereas the optimal value of the holdup M clearly depends on the feedrate. It is also rather obvious that a policy based on keeping x_B constant is most likely to fail, because x_B goes through a maximum as we increase M, and if we specify a value of x_B above this maximum, then operation is infeasible. However, otherwise it is not at all clear, even in this simple case, what the best choice of the controlled variable is.

Step 6: Evaluation of loss

To compare the alternatives we compute the loss $L = J - J_{\text{opt}}$ with each of the candidate variables kept constant at its nominal optimal value. The results are given in Table 2 for the 8 candidate variables and the 6 disturbances. Disturbance d_c represents the implementation error and the loss in this case is evaluated with $c = c_s + d_c$. For example, for $c = M/F$ we have $c_s = 1.0$ and $d_c = 0.2 \cdot 1.0 = 0.2$ (20% error) and fixing $M/F = 1.0 + 0.2 = 1.2$ (rather than the optimal value of 1.0) results in a loss of 0.17 (see Table 2).

We see from Table 2 that the loss is quite small in most cases, but in some cases there is no feasible solution (marked as inf. in the table). As expected, this is the case if we specify $x_B = 0.2$, and this is higher than its maximum achievable value. But note that infeasibility may occur for most choices of controlled variables if the disturbance is sufficiently large. For example, if we specify $x_A = 0.4$ then we obviously get infeasibility when $z_A < 0.4$. Note that there is usually no "warning", in terms of a large value of the loss, as we approach infeasibility.

With the numbers given above, the implementation error is not very important. However, in many cases it may be a critical factor which eliminates an otherwise good candidate controlled variable.

Step 7: Selection of controlled variable

Most engineers would probably attempt to control some composition if such measurements were available, and indeed we find that the variable $c_6 = x_B/x_A$ is the ideal variable to keep constant when there are disturbances in z_A (see Table 1). However, keeping $x_B/x_A = 0.5$ results in rather large losses when there are disturbances in the rate constants, resulting in an average loss of 0.74. Keeping $x_A = 0.40$ gives a large loss when z_A changes, and the average loss is 0.52 (just above the acceptable). As expected, keeping $x_B = 0.2$ results in infeasibility in some cases, and also keeping $x_C = 0.4$ results in infeasibility when z_A is too low. Controlling measurement θ_1 gives a large loss when there is a disturbance in z_A, whereas θ_2 is somewhat sensitive to implementation errors. In summary, from Table 2 we see that none of the composition measurements ($x_A, x_B, x_C, x_B/x_A$) or property measurements (θ_1, θ_2) result in an acceptable avarage loss of 0.5 or less.

Somewhat surprisingly, the "open-loop" policy where we keep the holdup $M = 1.0$ results in very small losses for all disturbances, except when there is a disturbance in the feedrate F where the loss is 0.6. The sensitivity to feedrate disturbances is eliminated if we include "feedforward" action from the feedrate F and instead keep the residence time M/F constant, but there is a penalty in terms of a larger implementation error. In conclusion, the simplest and best strategy for the example is to keep constant the holdup M (average loss 0.18) or the residence time M/F (average loss 0.10).

Disturbance	J	M	$\frac{M}{F}$	x_A	x_B	x_C	$\frac{x_B}{x_A}$	θ_1	θ_2
Nominal	-20.00	1.00	1.00	0.40	0.20	0.40	0.50	2.00	1.80
$d_1 : F = 0.7$	-20.00	0.70	1.00	0.40	0.20	0.40	0.50	2.00	1.80
$d_2 : z_A = 0.6$	-15.00	1.00	1.00	0.30	0.15	0.55	0.50	2.25	1.85
$d_3 : z_A = 1.0$	-25.00	1.00	1.00	0.50	0.25	0.25	0.50	1.75	1.75
$d_4 : k_A = 1.5$	-24.24	0.82	0.82	0.36	0.24	0.40	0.67	2.04	1.88
$d_5 : k_B = 1.5$	-16.16	0.82	0.82	0.44	0.16	0.40	0.37	1.96	1.72

Table 1: Optimal values for reactor case study

Disturbance	Loss for $M = 1.0$	Loss for $\frac{M}{F} = 1.0$	Loss for $x_A = 0.4$	Loss for $x_B = 0.2$	Loss for $x_C = 0.4$	Loss for $\frac{x_B}{x_A} = 0.5$	Loss for $\theta_1 = 2.0$	Loss for $\theta_2 = 1.80$
Nominal	0	0	0	0	0	0	0	0
$d_1 : F = 0.7$	0.62	0	0	0	0	0	0	0
$d_2 : z_A = 0.6$	0.00	0.00	1.67	inf.	15.0*	0	3.35	0.36
$d_3 : z_A = 1.0$	0.00	0.00	1.00	5.00	1.75	0	1.39	0.28
$d_4 : k_A = 1.5$	0.24	0.24	0.24	4.24	0	1.39	0.06	0.82
$d_5 : k_B = 1.5$	0.16	0.16	0.16	inf.	0	2.83	0.04	0.72
d_c : impl. error	0.05	0.17	0.05	inf.	0.05	0.20	0.29	1.72
Average loss	0.18	0.10	0.52	inf.	2.80*	0.74	0.86	0.65
Ranking	2	1	3	8	7	5	6	4

inf. denotes infeasible operation

*) At the limit to infeasibility ($M/F = 0$) for z_A=0.6.

With impl.error: $M = 1.1, M/F = 1.2, x_A = 0.42, x_B = 0.21, x_C = 0.42, x_B/x_A = 0.55, \theta_1 = 2.1, \theta_2 = 1.9$

Table 2: Loss for reactor case study

Another case: No C in feed

Above we assumed that the feed contained components A and C, but consider next a different case where the feed contains component B rather than component C. Otherwise all the data is the same. Variable x_A here replaces x_B/x_A as the ideal variable with respect to disturbances in z_A (this can be proven analytically), and keeping x_A constant also yields small losses when there are other disturbances. Therefore, in this case the loss is smallest (0.18) when x_A is constant; it is 0.20 with θ_1 constant; 0.32 with x_C constant; 0.81 with M/F constant; 0.89 with M constant; 1.09 with θ_2 constant; whereas keeping x_B constant or x_B/x_A constant results in infeasibility (Skogestad *et al.* 1999). Thus we find, somewhat surprisingly, that the ranking is almost reversed compared to that found above.

It is not easy to explain physically why the particular variables are preferred in the two cases. This shows that it may be difficult to rely on physical insight when selecting the "dominant" variable to control.

5 Distillation case study

We consider a binary mixture with constant relative volatility $\alpha = 1.12$ to be separated in a distillation column with 110 theoretical stages and the feed entering at stage 39 (counted from the bottom with the reboiler as stage 1). Nominally, the feed contains 65 mole% of light component ($z_F = 0.65$) and is saturated liquid ($q_F = 1.0$). This is "column D" of Skogestad and Morari (1988) and represents a propylene-propane splitter where propylene (light component) is taken overhead as a final product with at least 99.5% purity, whereas unreacted propane (heavy component) is recycled to the reactor for reprocessing. We assume that the feed rate is given at 1 kmol/min and that there is no capacity limit in the column. (We will later study separately the case where $V_{max} = 10$ kmol/min and what implications this has on the operation of the column.)

Step 1: Degree of freedom analysis

As mentioned, for a given column feed rate the column has two operational degrees of freedom at steady state. These may for instance be selected as the reflux and distillate flows,

$$u = \begin{pmatrix} L \\ D \end{pmatrix}$$

Step 2: Cost function and constraints

Ideally, the optimal operation of the column should follow from considering the overall plant economics. However, to be able to analyze the column separately, we introduces prices for all streams enetering and exiting the column and consider the following profit function P which should be maximized (i.e. $J = -P$)

$$P = p_D D + p_B B - p_F F - p_V V \tag{16}$$

We use the following prices [$/kmol]

$$p_D = 20, \quad p_B = 10 - 20x_B, \quad p_F = 10, \quad p_V = 0.1$$

The price $p_V = 0.1$ [$/kmol] on boilup includes the costs for heating and cooling which both increase proportionally with the boilup V. The price for the feed is $P_F = 10$ [$/kmol], but its value has no significance on the optimal operation when the feed rate is given. The price for the distillate product is 20 [$/kmol], and its purity specification is

$$x_D \geq 0.995$$

Disturbance	x_D	x_B	D/F	L/F	V/F	L/D	P/F
Nominal	0.995	0.040	0.639	15.065	15.704	23.57	4.528
$F = 1.3$	0.995	0.040	0.639	15.065	15.704	23.57	4.528
$z_F = 0.5$	0.995	0.032	0.486	15.202	15.525	31.28	2.978
$z_F = 0.75$	0.995	0.050	0.741	14.543	15.284	19.62	5.620
$q_F = 0.5$	0.995	0.040	0.639	15.133	15.272	23.68	4.571
$x_D = 0.996$	0.996	0.042	0.637	15.594	16.232	24.47	4.443
$p_D = 30$	0.995	0.035	0.641	15.714	16.355	24.51	
$p_V = 0.5$	0.995	0.138	0.597	11.026	11.623	18.47	

Nominal values: $F = 1, z_F = 0.65, q_F = 1.0, p_D = 20, p_V = 0.1$

Table 3: Optimal operating point (with maximum profit P/F) for distillation case study

There is no purity specification on the bottoms product, but we note that its price is reduced in proportion to the amount of light component (because the unneccessary reprocessing of light component reduces the overall capacity of the plant).

With a nominal feedrate $F = 1$ [kmol/min], the profit value of the column is of the order 4 [$/min], and we would like to find a controlled variable which results in a loss L less than 0.04 [$/min] for each disturbance (corresponding to a yearly loss of about $20000).

Step 3: Disturbances

We consider five disturbances:

d_1: An increase in feedrate F from 1 to 1.3 kmol/min.

d_2: A decrease in feed composition z_F from 0.65 to 0.5

d_3: An increase in feed composition z_F from 0.65 to 0.75

d_4: A decrease in feed liquid fraction q_F from 1.0 (pure liquid) to 0.5 (50% vaporized)

d_c: An increase of the purity of distillate product x_D from 0.995 (its desired value) to 0.996

The latter is an implementation error (safety margin) for x_D. In addition, we will consider implementation errors for the other selected controlled variable (see below).

Step 4: Optimization

In Table 3 we give the optimal operating point for the five disturbances; larger feedrate ($F = 1.3$), less and more light component in the feed ($z_F = 0.5$ and $z_F = 0.65$), a partly vaporized feed ($q_F = 0.5$), and a purer distillate product ($y_D = 0.98$). In addition, we have considered the effect of a higher price for the distillate product ($p_D = 30$) and a five times higher energy price ($p_V = 0.5$).

As expected, the optimal value of all the variables listed in the table ($x_D, x_B, D/F, L/F, V/F, P/F$) are insensitive to the feedrate, since the columns has no capacity constraints, and the efficiency is assumed independent of the column load.

We find that, except in the last case with a much higher energy price, the optimal bottom composition stays fairly constant around $x_B = 0.04$. This indicates that a good strategy for implementation may be to control x_B at a constant value. On the other hand, the value of D/F varies considerably, so we expect this to be a poor choice for the controlled variable.

Disturbance	$x_B = 0.04$	$D/F = 0.639$	$L = 15.065$	$L/F = 15.065$	$V/F = 15.704$	$L/D = 23.57$
Nominal	0	0	0	0	0	0
$F = 1.3$	0	0	0.514	0	0	0
$z_F = 0.5$	0.023	inf.	0.000	0.000	0.001	1.096
$z_F = 0.75$	0.019	2.530	0.006	0.006	0.004	0.129
$q_F = 0.5$	0.000	0.000	0.001	0.001	0.003	0.000
$x_D = 0.996$	0.086	0.089	0.091	0.091	0.091	0.093
20% impl.error	0.012	inf.	0.119	0.119	0.127	0.130

inf. denotes infeasible operation

Nominal values: $x_D = 0.995, z_F = 0.65, q_F = 1.0$

20% impl.error: $x_B = 0.048, D/F = 0.766, L = 18.08, L/F = 18.08, V/F = 18.85, L/D = 28.28$

Table 4: Loss for distillation case study.

Step 5: Candidate controlled variables

It is clear that one of the controlled variables should be the distillate composition, x_D. This follows since the optimal solution is always obtained when the product purity specification for the most expensive product is "active", i.e. in our case when $x_D = 0.995$. We are then left with one "unconstrained" degree of freedom which we want to specify by keeping a controlled variable at a constant value.

We found above that the optimal value of the bottom composition is relatively insensitive to disturbances and other changes, and this indicates that x_B is a good candidate controlled variable. However, there at least two practical problems associated with this choice. First, on-line composition measurements are often unreliable and expensive. Second, dynamic performance may be poor because it is generally difficult to control both product compositions ("two-point control") due to strong interactions (e.g. SM and Shinskey). Thus, if possible, we would like to control some other variable.

The following six alternative controlled variables are considered (in addition to x_D which should always be kept constant at its lower limit of 0.995):

$$x_B;\ D/F;\ L;\ L/F;\ V/F; L/D$$

We consider implementation errors of about 20% in all variables, including x_D.

Step 6: Evaluation of loss

In Table 4 we show for $F = 1$ [lmol/min] the loss $L = P_{opt} - P$ [\$/min] when each of the six candidate controlled variables are kept constant at their nominally optimal values. Recall that we would like the loss to be less than 0.04 [\$/min] for each disturbance.

We have the following comments to the results for the loss given in Table 4:

1. As expected, we find that the losses are small when we keep x_B constant.
2. Somewhat surprisingly, for disturbances in feed composition it is even better to keep L/F or V/F constant
3. Not surprisingly, keeping D/F (or D) constant is not an acceptable policy, e.g., operation is infeasible when z_F is reduced from 0.65 to 0.5.
4. All alternatives are insensitive to disturbances in feed enthalpy (q_F).
5. L/D is not a good controlled variable, primarily because its optimal value is rather sensitive to feed composition changes.

6. For a implementation error (overpurification) in x_D where x_D is 0.996 rather than 0.995 all the alternatives give a loss of about 0.09. For a larger overpurification where $x_D = 0.998$ (not shown in Table) the loss ranges from 0.43 (x_D constant) to 0.79 (L/D constant). The loss is rather large, so we conclude that we should try to control x_D close to its specification.
7. For reflux L and boilup V one should include "feedforward" action from F, and keep L/F and V/F constant (e.g., the loss is 0.514 if we keep L constant and F increases by 30%).
8. Using L/F or V/F as controlled variables is rather sensitive to implementation errors.
9. Other controlled variables have also been considered (not shown in Table). For example, a constant composition (temperature) on stage 19 (towards the bottom), $x_{19} = 0.20$, gives a loss of 0.064 when z_F is reduced to 0.5, but otherwise the losses are similar to those with x_B constant.
10. We have not computed the effect of changes in prices in Table 4, because these do not effect column behavior, so all alternatives behave the same (with the same loss). Thus, if there are price changes, then one must recompute new optimal values for the variables.

Step 7: Selection of controlled variables

From Table 4 the following three candidate sets of controlled variables yield the lowest losses

$$c_1 = \begin{pmatrix} x_B \\ x_D \end{pmatrix}; \quad c_2 = \begin{pmatrix} L/F \\ x_D \end{pmatrix}; \quad c_3 = \begin{pmatrix} V/F \\ x_D \end{pmatrix};$$

The "two-point" control structure c_1 where both compositions are controlled, is known to result in a difficult control problem due to strong interaction, and we may not be able to keep the compositions at their specifications. The loss will then be larger than indicated, and it is probably better to keep L/F or V/F constant.

It is usually simpler to keep a liquid flow L/F rather than a vapor flow V/F constant (less implementation error). On the other hand, we have already noted that it is important to control x_D close to its specification, and this is probably more easily done using reflux L. In conclusion, we have:

Proposed control system.

- V is used[3] to keep $x_D = 0.995$.
- $L/F = 15.07$ is kept constant.

Alternative control system.

- L is used to keep $x_D = 0.995$.
- $V/F = 15.70$ is kept constant.

If it turns out to be difficult to keep L/F (or V/F) constant, then we may instead manipulate L (or V) to keep a temperature towards the bottom of the column constant.

Column with capacity limitations

Above we assumed that the feedrate was given and that the column had no capacity limit. However, all columns have a capacity limit, To understand better the implications of this, consider a column with $V_{max} = 10$ [kmol/min]. Otherwise the specifications and cost data are as above.

Above we found $(V/F)_{opt} = 15.70$ (in the nominal case). This implies that the above results only apply for feedrate $F \leq V_{max}/(V/F)_{opt} = 10/15.70 = 0.637$ [kmol/min].

We now want to study how the column should be operated also for other (larger) desired feedrate. We use the term desired, rather than given, since we find that it is not economic to increase the feedrate above a certain value.

[3]There are other possible choices for controlling x_D, e.g. we could use the distillate flow D. However, V has a more direct effect.

Imagine, that it is your job to optimize the profit of the column. You will then find that it is not profitable to accept too large feedrates, because when V reaches V_{max}, you are forced to give up bottom purity and to put a larger fraction of the light component into the bottom product which has a lower price than the feed. Some more carful thinking, reveals that if you have reached $V = V_{max}$ and are free to decide on the feedrate, then you should try to optimize P/V (instead of P/F as done above). With the given column data, we find that $(P/V)_{opt} = 0.331$ [\$/kmol] is obtained when $x_B = 0.09$, $V/F = 12.77$ and $L/F = 12.15$. Thus, with $V_{max} = 10$ [kmol/min], the optimal profit is obtained with a feedrate of $F_{opt} = 10/12.77 = 0.783$ [kmol/min].

Since the distillate is the more valueable product, it is optimal in all cases, also with capacity limits, to keep the distillate product at its specification $x_D = 0.995$. We then have the following three cases depending on the desired feedrate:

1. **Low desired (given) feedrates**; $F \leq 0.637$ [kmol/min]). This is the case we studied above where the column operates below its capacity limit on the boilup V. The optimal solution is unconstrained in one variable, and the nominally optimal value of x_B is 0.04 in this case.

   Proposed control system.

   - L is used to keep $x_D = 0.995$.
   - V is adjusted to keep $V/F = 15.70$ constant

   This is the "alternative control system" proposed above, but is chosen here because we can use the same composition controller as for cases 2 and 3 (below).

2. **Intermediate desired (given) feedrates**; $0.637 < F < 0.783$ [kmol/min]. The column operates optimally at maximum capacity $V = V_{max} = 10$ [kmol/min], and there are no remaining degrees of freedom (i.e. we have "active constraint control").

   Proposed control system.

   - L is used to keep $x_D = 0.995$.
   - V is kept constant at V_{max}. (As F is increased from 0.637 to 0.783, the value x_B of increases from 0.04 and 0.09).

3. **Large *desired* feedrates**; $F > 0.783$ [kmol/min]). The column operates at maximum capacity $V = V_{max} = 10$ [kmol/min] and we should keep $F = 0.783$ to maximize P/V. In this case the column is a bottleneck for the plant as it is not economic to have higher feedrates because too much valueable product will be lost in the bottoms product. Thus, the feedrate becomes a third degree of freedom in the optimization, and the optimal solution is unconstrained in this variable (the nominally optimal value of x_B is 0.09 in this case).

   Proposed control system.

   - L is used to keep $x_D = 0.995$.
   - V is kept constant at $V_{max} = 10$.
   - F is adjusted to keep $V/F = 12.77$ constant (this is better than keeping F constant)

The three above cases for the desired feedrate can easily be implemented in a single control system using some simple logic.

In summary, the distillation case study shows the importance of selecting the right controlled variables when implementing the optimal solution. We note that the implementation error was not important in this case study, but we stress that it should be included in the analysis. For example, the implementation error is the main reason why we rarely select temperatures near the column ends as controlled variables (because the measurement error is too large compared to its sensitivity).

6 Discussion

6.1 Previous work

Inspired by the work of Findeisen *et al.* (1980), the basic idea of self-optimizing control was formulated almost twenty years ago by Morari *et al.* (1980). They write that "in attempting to synthesize a feedback optimizing control structure, our main objective is to translate the economic objectives into process control objectives. In other words, *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 optimal operating conditions.* [...] This means that by keeping the function $c(u, d)$ at the setpoint c_s, through the use of the manipulated variables u, for various disturbances d, it follows uniquely that the process is operating at the optimal steady-state." If we replace ithe term "optimal adjustments" by by "acceptable adjustments (in terms of the loss)" then the above is a precise description of a self-optimizing control structure. The only factor they fail to consider is the implementation error $d_c = c - c_s$. Morari *et al.* (1980) also propose to select the best set of controlled variables based on minimizing the loss ("feedback optimizing control criterion 1"). They also propose that Monte Carlo simulations should be used to evaluate the loss if the disturbances have a probability distribution.

As a minor remark we mention that Morari *et al.* (1980) claim that "ideally one tries to select c in such a way such that some or all the elements in c are independent of the disturbances d." This statement is generally *not* true, because we need to be able to detect the disturbances through the variables c. A better requirement is that the *optimal* values of the elements in c are insensitive to disturbances d (this is requirement 1 presented earlier).

The ideas of Shinnar (1981), who makes use of "dominant variables", and Luyben (1988), who introduced the term "eigenstructure", are similar to the ones presented in this paper, but they have a more intuitive process-oriented approach.

Let us first consider the interesting work of Shinnar (1981) which may at first seem unrelated. He introduces the following sets of variables

- Y_p (the "primary" or "performance" or "economic" variables) is "the set of process variables that define the product and process specifications as well as process constraints"
- Y_d is the set of dynamically measured process variables
- Y_{cd} (a subset of Y_d) is the "set of process variables on which we base our dynamic control strategy"
- U_d is the dynamic input variables

The goal is to maintain Y_p within prescribed limits and to achieve this goal "we choose in most cases a small set Y_{cd} and try to keep these at a fixed set of values by manipulating U_d" (later, in Arbel *et al.* (1996), he introduced the term "partial control" to describe this idea). He writes that the overall control algorithm can normally be decomposed into a dynamic control system (which adjust U_d) and a steady-state control which determines the set points of Y_{cd} as well as the values of U_s [the latter are the manipulations which only can be changed slowly], and that we "look for a set Y_{cd}, U_d that contains variables that have a maximum compensating effect on Y_p". If one translates the words and notation, then one realizes that Shinnar's idea of "partial control" is very close to the idea of "self-optimizing control" presented in this paper and in Morari *et al.* (1980). Specifically, set $u = U_d$ (manipulated variables), $y_m = Y_d$ (measurements) and $c = Y_{cd}$ (controlled variables), and interpret J as consisting of some weighted sum of the (unconstrained) variables in Y_p, e.g. $J = \|Y_p - Y_{ps}\|$ where $\|\cdot\|$ represents some appropriate norm (possibly weighted) and Y_{ps} are the setpoints of the primary variables, and we see that this is the same problem as studied in this paper (the difference is that Shinnar assumes that there exist at the outset a set of "primary" variables Y_p that need to be controlled, whereas we in this paper assume that the starting point is an economic cost function that should be minimized).

In Shinnar (1981) a case study of a fluidized catalytic cracker (FCC) is presented where the controlled variables are selected mainly based on process insight ("our main concern is to control the heat balance and the set of Y_{cd} is chosen accordingly"). Thus, the approach for selecting the controlled variables is

mostly based on physical insight. The later paper by Arbel *et al.* (1996) is similar to Shinnar (1981), but it extends the FCC case study and introduces the concepts of "dominant variables" and "partial control"[4]. The dominant variables are the process variables that tend to dominate the process behavior, for example, the temperature in a reactor, and which therefore should be selected as controlled variables. The authors provide some intuitive ideas and examples for selecting dominant variables which may be useful in some cases, especially when no model information is available. However, it is not clear how helpful the idea of "dominant" variable is, since they are not really defined and no explicit procedure is given for identifying them. Indeed, Arbel *et al.* (1996) write that "the problems of partial control have been discussed in a heuritic way" and that "considerably further researh is needed to fully understand the problems is steady-state control of chemical plants". It is believed that the approach presented in this paper, based on using a steady-state model to evaluate the (economic) loss, provides an important part of this missing theoretical framework.

Tyreus (1999) provides some additional interesting ideas on how to select dominant variables, partly based on the extensive variable idea of Georgakis (1986) and the thermodynamic ideas of Ydstie, but again no procedure for selecting such variables are presented. Again, it is believed that a more systematic approach, based on applying the four requirement in Section 3.2, or (even better) evaluating the loss, will confirm the recommendations given by Tyreus (1999).

Luyben (1988) introduced the term "eigenstructure" to describe the inherently best control structure (with the best self-regulating and self-optimizing property). However, he did not really define the term, and also the name is unfortunate since "eigenstructure" has a another unrelated mathematical meaning in terms of eigenvalues. Apart from this, Luyben and coworkers (e.g. Luyben (1975), Yi and Luyben (1995)) have studied unconstrained problems, and some of the examples presented point in the direction of the selection methods presented in this paper. However, Luyben proposes to select controlled ouputs which minimizes the steady-state sensitive of the manipualted variable (u) to disturbances, i.e. to select controlled outputs (c) such that $(\partial u/\partial d)_c$ is small, whereas we really want to minimize the steady-state sensitivity of the economic loss (L) to disturbances, i.e. to select controlled outputs (c) such that $(\partial L/\partial d)_c$ is small.

Fisher *et al.* (1988) discuss selection of controlled variables, mainly focused towards active constraint control. However, somewhat hidden in their HDA example (p. 614) one finds statements about selecting controlled variables which optimal values are insensitive to disturbances (requirement 1 for variable selection presented in this paper).

In accordance with the approach presented in this paper, Narraway *et al.* (1991), Narraway and Perkins (1993) and Narraway and Perkins (1994)) strongly stress the need to base the selection of the control structure on economics, and they discuss the effect of disturbances on the economics. However, they do not formulate any rules or procedures for selecting controlled variables.

Marlin and Hrymak (1997) stress the need to find a good way of implementing the optimal solution in terms how the control system should respond to disturbances, "i.e. the key constraints to remain active, variables to be maximized or minimized, priority for adjusting manipulated variables, and so forth." They suggest that an issue for improvement in today's real-time optimization systems is to select the control system that yields the highest profit for a range of disturbances that occur between each execution of the optimization. This is similar to the ideas presented in this paper.

Finally, Zheng *et al.* (1999) present a procedure for selecting controlled variables based on economic penalties that is similar to the approach presented in this paper (apparently, the work has been performed independently), but they do not consider the implementation error. The procedure is applied to a reactor-separator-recycle system.

In summary, it is clear that many authors have been aware of the importance of the ideas presented in this paper. The main contribution of the present paper is to bring the ideas together and formulate them

[4]The term "partial control" or "partially controlled system" is used by other authors (Waller *et al.* 1988) (Skogestad and Postlethwaite 1996) in a more general sense, to mean the system as it is seen from some higher layer in the control hierarchy with some loops already closed (e.g. a plant where the level loops are closed). For a partially controlled system, the setpoints to the lower loops (e.g. level setpoints) replace the "consumed" manipulated variables (e.g. flows) as degrees of freedom.

more clearly, and to present some case studies.

6.2 Region of feasibility

In this paper we have evaluated the (economic) *loss* for disturbances of a given magnitude. Another important issue to consider is the region of feasibility (stability) (Zheng *et al.* 1999). We could evaluate *feasibility loss*, defined as the difference between the disturbance region where feasible operation is possible (using the optimal u) and the disturbance region we can handle with constant values of c, to be as small as possible.

For example, consider a case where there is an inequality constraint on an input variable which optimally is active only under certain conditions (disturbances), but this constrained input variable is *not* included as a controlled variable. Here one must be careful to avoid infeasibility during implementation, for example, there may be a disturbance such that the specified value of the controlled variable can only be achieved with a nonphysical value of the input (e.g. a negative flowrate).

The on-line optimization is usually for simplicity based on the nominal disturbance (d_0), and two approaches to avoid infeasibility in such a case are to

1. use "back-offs" for the setpoints during implementation (Narraway *et al.* 1991), or
2. add "safety margins" to the constraints during the (nominal) optimization

One approach for obtaining the values for the back-offs or safety margins may be to solve a "robust optimization problem" (Glemmestad 1997)) where one considers all possible disturbances. There is clearly a need for more research in this area.

A third, and better approach in terms of minimizing the loss, is to track the active constraint, but this requires a more complex control system. In particular, model predictive control is very well suited and much used for tracking active constraints.

6.3 Additional examples

Additional example for selecting controlled variables are available in a number of conference publications and Ph.D. theses, e.g.

- Ph.D. thesis of Morud (1995), chapter 8: CSTR with chemical reaction
- Ph.D. thesis of Glemmestad (1997): Application to heat exchanger networks; special emphasis on feasibility issues.
- Ph.D. thesis of Havre (1998): Application to selection of temperature location in distillation.
- Halvorsen and Skogestad (1997) and Halvorsen and Skogestad (1998): Application to integrated Petlyuk distillation columns.

The papers and theses are available over the internet.

The Petlyuk distillation example is particularly interesting because in this case the choices of controlled variables make a big difference (whereas the differences for the rather simple examples presented in this paper admittedly were quite small).

7 Conclusion

In this paper we have presented a systematic procedure for selecting controlled variables c based on evaluating with constant setpoints c_s the loss $J - J_{\text{opt}}$ for possible disturbances. If the loss is acceptable then we have "self-optimizing" control. The procedure requires a steady-state process model and a clear definition of the cost function J to be minimized during operation (obviously, without such a cost function one cannot judge what operation is the best). The procedure was applied to three example; to a

simple ”toy” example, to a somewhat academic CSTR example, and finally to a more realistic distillation column example.

To assist in selecting good candidate variables one should look for variables that satisfy the following requirements:

Requirement 1. Its optimal value is insensitive to disturbances

Requirement 2. It is easy to measure and control accurately

Requirement 3. Its value is sensitive to changes in the manipulated variables

Requirement 4. For cases with two or more controlled variables, the selected variables should not be too closely correlated.

One problem is that in general it not is clear offhand if a self-optimizing structure exists, and going through the various alternatives, for example using the given procedure, can be quite tedious. On the other hand, since the issue of finding good controlled variables is a *structural* problem, then we often find that a good structure obtained for a particular case, also works well on another similar process case with different parameter values. Thus, if we can actually find a self-optimizing structure for a process, then it is almost like an invention (and may pobably even be patented).

Acknowledgment. *The ideas in this paper have developed gradually over several years. Valuable discussions with J.C. Morud, I.J. Halvorsen, T. Larsson, K. Havre, B. Glemmestad and M.S. Govatsmark are gratefully acknowledged*

References

Arbel, A., I.H. Rinard and R. Shinnar (1996). Dynamics and control of fluidized catalytic crackers. 3. designing the control system: Choice of manipulated and measured variables for partial control. *Ind. Eng. Chem. Res.* pp. 2215–2233.

Arkun, Y. and G. Stephanopoulos (1980). Studies in the synthesis of control structures for chemical processes: Part iv. design of steady-state optimizing control structures for chemical process units. *AIChE Journal* **26**(6), 975–991.

Findeisen, W, F.N. Bailey, M. Brdys, K. Malinowski, P. Tatjewski and A. Wozniak (1980). *Control and coordination in Hierarchical Systems*. John Wiley & sons.

Fisher, W.R., M.F. Doherty and J.M. Douglas (1988). The interface between design and control. 3: Selecting a set of controlled variables.. *Ind. Eng. Chem. Res.* **27**(4), 611–615.

Foss, C.S. (1973). Critique of chemical process control theory. *AIChE Journal* **19**(2), 209–214.

Georgakis, C. (1986). On the use of extensive variables in process dynamics and control. *Chemical Engineering Science* **47**(6), 1471–1484.

Glemmestad, B. (1997). Optimal operation of integrated processes: Studies on heat recovery systems. PhD thesis. Norwegian University of Science and Technology. Available from http://www.chembio.ntnu.no/users/skoge/.

Gordon, M.L. (1986). Simple optimization for dual composition control. *Hydrocarbon Processing* pp. 59–62.

Halvorsen, I.J. and S. Skogestad (1997). Optimizing control of petlyuk distillation: Understanding the steady-state behavior. *PSE ’97-ESCAPE-7, 25-29 May, 1997, Trondheim, Norway, Supplement to Computers. chem. Engng.* **21**, S249–S254.

Halvorsen, I.J. and S. Skogestad (1998). Use of feedback for indirect optimizing control: Application to petlyuk distillation. *DYCOPS-5, 5th IFAC Symposium on Dynamics and Control of Process Systems, Corfu, Greece, June 8-10, 1998* pp. 399–404.

Havre, K. (1998). Studies on controllability analysis and control structure design. PhD thesis. NTNU Trondheim. Available from http://www.chembio.ntnu.no/users/skoge/.

Luyben, W.L. (1975). Steady-state energy conservation aspects of distillation column control system design. *Ind. Eng. Chem. Fundam.* pp. 321–325.

Luyben, W.L. (1988). The concept of eigenstructure in process control. *Ind. Eng. Chem. Res.* pp. 206–208.

Luyben, W.L. (1996). Design and control degrees of freedom. *Ind. Eng. Chem. Res.* pp. 2204–2214.

Maarleveld, A. and J.E. Rijnsdrop (1970). Constraint control of distillation columns. *Automatica* pp. 51–58.

Marlin, T.E. and A.N. Hrymak (1997). Real-time operations optimization of continuous processes. In: *Fifth international conference on chemical process control (CPC-5, Lake Tahoe, Jan. 1996).* Vol. 93 of *AIChE Symposium Series.* pp. 156–164.

Morari, M. (1982). Integrated plant control: A solution at hand or a research topic for the next decade?. In: *CPC-II*. pp. 467–495.

Morari, M., G. Stephanopoulos and Y. Arkun (1980). 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* pp. 220–232.

Morud, J. (1995). Studies on the dynamics and operation of integrated plants. PhD thesis. University of Trondheim. Available from http://www.chembio.ntnu.no/users/skoge/.

Narraway, L. and J. Perkins (1994). Selection of process control structures based in economics. *Computers chem. Engng* pp. S511–S515.

Narraway, L.T. and J.D. Perkins (1993). Selection of process control structure based on linear dynamic economics. *Ind. Eng. Chem. Res.* pp. 2681–2692.

Narraway, L.T., J.D. Perkins and G.W. Barton (1991). Interaction between process design and process control: economic analysis of process dynamics. *J. Proc. Cont.* pp. 243–250.

Ponton, J.W. and D.M. Liang (1993). Hierarchical approach to the design of process control systems. *Chemical engineering Research and Design* pp. 181–188.

Shinnar, R. (1981). Chemical reactor modelling for purposes of controller design. *Cheng. Eng. Commun.* pp. 73–99.

Shinskey, F.G. (1984). *Distillation Control*. 2 ed.. McGraw-Hill Book Company.

Skogestad, S. and I. Postlethwaite (1996). *Multivariable Feedback Control*. John Wiley & Sons.

Skogestad, S. and M. Morari (1988). Understanding the dynamic behavior of distillation columns. *Ind. Eng. Chem. Res.* **27**(10), 1848–1862.

Skogestad, S., I.J. Halvorsen, T. Larsson and M.S. Govatsmark (1999). Plantwide control: The search for the self-optimizing control structure. In: *Preprints 14th IFAC World Congress, Beijing*. IFAC. pp. 1–25.

Tyreus, B.D. (1999). Dominant variables for partial control; part 1 and 2. *Ind. Eng. Chem. Res.* **38**, 1432–1455.

Waller, K.V., D.H. Finerman, P.M. Sandelin, K.E. Häggblom and S.E. Gustafsson (1988). An experimental comparison of four control structures for two-point control of distillation. *Ind. Eng. Chem. Res.* **27**(4), 624–630.

Yi, C.K. and W.L. Luyben (1995). Evaluation of plant-wide control structures by steady-state disturbance sensitivity analysis. *Ind. Eng. Chem. Res.* pp. 2393–2405.

Zheng, A., R.V. Mahajannam and J.M. Douglas (1999). Hierchical procedure for plantwide control system synthesis. *AIChE Journal* **45**, 1255–1265.