Self-optimizing control: Local Taylor series analysis *

Ivar J. Halvorsen, Sigurd Skogestad† and John C. Morud
Department of Chemical Engineering
Norwegian University of Science and Technology
N–7491 Trondheim Norway

Submitted to Ind.Eng.Chem.Res.
This version: October 11, 2002

Abstract

Which variables should we select to control? More precisely, in this paper, the issue is how to best *implement* the optimal operation policy in the presence of uncertainty (disturbances and implementation errors) by selecting the right set of controlled variables c to be kept at constants setpoints c_s. The most difficult is to select the unconstrained controlled variables, and we here use a second-order Taylor approximation around the optimal point of operation. This results in an exact local method, which may be used to find the optimal linear combination of measurements to use as controlled variables. We also derive the approximate singular value rule, which is very useful for quick screening and elimination of poor candidate variables.

1 Introduction

In this paper we consider optimal steady-state operation. The first step is to quantify the desired operation by defining a scalar cost function. The second step is to optimize the operation by minimizing the cost with respect to the available degrees of freedom. The third step, equally important in our view (but not considered to be an important an issue by most people), is the actual *implementation* of the optimal policy in the plant by use of its control system. Obviously, form a purely mathematical point of view, it would be optimal to use a centralized on-line optimizing controller with continuous update of its model parameters. However, for a number of reasons, we almost always decompose the control system into several layers, which in a chemical plant typically include scheduling (weeks), site-wide optimization (day), local optimization (hour), supervisory/predictive control (minutes) and regulatory control (seconds); see Figure 1. In this paper we consider the interaction between the local optimization layer and the feedback control layer. The two layers interact through the *controlled variables* c, whereby the optimizer computes their optimal setpoints c_s (updating them about every hour), and the control layer attempts to implement them in practice, i.e. to get $c \approx c_s$. The issue to be considered in this paper is then: What variables c should we control? So far very little theory has been available to analyze this important problem, and the objective of this paper is to study the problem from a local (linear) point of view.

*An early version of this paper was presented at the AIChE Annual meeting, Miami Beach, 1998

†Corresponding author. E-mail: skoge@chemeng.ntnu.no; Phone: +47-7359-4154

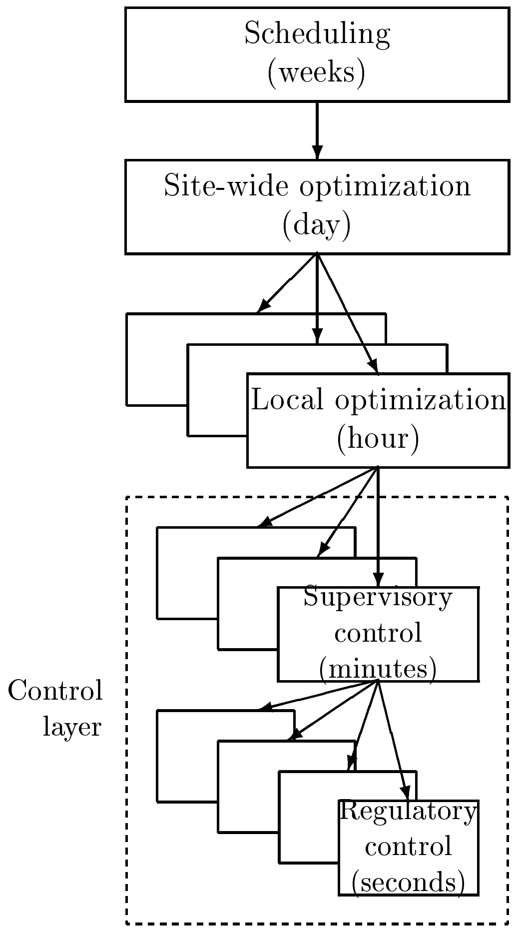

Figure 1: Typical control hierarchy in a chemical plant.

Let us describe the issues in a little more detail. The setpoints c_s for the controlled variables c are kept constant between each update. However, this constant setpoint policy will, for example due to disturbances and implementation errors, result in a loss, $L = J - J_{\text{opt}}$, when compared to the truly optimal operation. If this loss is acceptable, then we have a "self-optimizing" control system:

> **Self-optimizing control** (Skogestad 2000) is when we can achieve acceptable loss with constant setpoint values for the controlled variables

We assume in this paper that the optimally active constraints are always controlled (implemented), the remaining unconstrained degrees of freedom. It is for such cases that the choice of controlled variables is an important issue. Skogestad (2000) presents for this case three approaches for selecting controlled variables for self-optimizing control:

1. Select variables c that satisfy certain properties which are derived below. The properties are useful for identifying candidate controlled variables, but do not provide a quantitative means of selecting the best variables.

2. *Minimum singular value rule:* Select variables that maximize the minimum singular value of the appropritately scaled steady-state gain matrix G' from inputs (u) to the selected controlled variables (c). This rule is derived below. It is not exact, but very simple and often works surpisingly well. It is especially useful for eliminating poor candidates.

3. Exact method based on "brute-force" evaluation of the loss with alternative sets of controlled variables kept at their setpoints.

The last method is normally very expensive numerically. However, in this paper we derive an exact *local* method with a small numerical load. Importantly, this local method may also be used to find the optimal measurement combination to control.

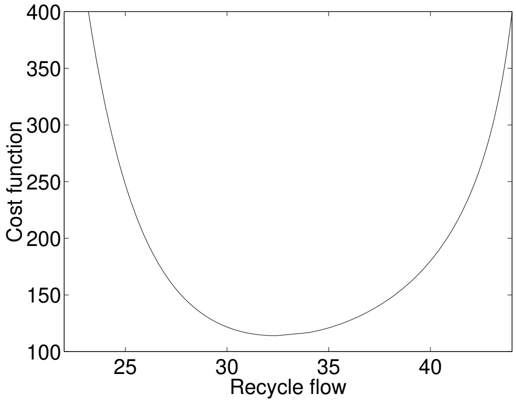

Figure 2: Cost function $J(u, d)$ as a function of independent variable u.
The local analysis holds in the region where the second derivative (Jacobian) $\frac{\partial^2 J}{\partial u^2}$ is constant.

The main limitation with the results in this paper is that the are based on a local second-order analysis, and thus may only apply for small deviations from the optimum, see Figure 2. The range of validity of a local analysis is highly problem-dependent, but several process control applications, not reported in this paper, has shown the agreement with a nonlinear analysis is good.

The reader is referred to Skogestad (2000) for references to related work on selection of controlled variables. Of the earlier work, Morari *et al.* (1980) come closest to the ideas presented in this paper. Morari *et al.* (1980) considered a local second-order expansion of the cost, but did not consider the expansion of the loss in terms of $e_c = c(d) - c_{\text{opt}}(d)$ (with a "moving" disturbance d), which is a key element in this paper.

2 Preliminaries

Summary of important notation

u - "base set" for the unconstrained degrees of freedom
d - disturbance variables
$J(u, d)$ - cost function to be minimized
$L = J - J_{\text{opt}}(d)$ - loss
$u_{\text{opt}}(d)$ - optimal value of u for given d
$e_u = u - u_{\text{opt}}(d)$ - deviation from optimal u (that results in a positive loss) $y = f_y(u, d)$ - measured variables
$c = h(y) = f_c(u, d)$ - selected controlled variables (function h free to select)
$c_{\text{opt}}(d) = f_c(u_{\text{opt}}(d), d)$ - optimal value of c for given d
$n = c - c_s$ - implementation error
$v = c_s - c_{\text{opt}}(d)$ - setpoint error
$e_c = c - c_{\text{opt}}(d) = v + d$ - deviation from optimal c (resulting in loss)
* - nominal operating point (normally assumed optimal)

2.1 Minimizing the cost J

We assume that the optimal operation problem can be quantified in terms of a scalar performance index (cost) J to be minimized, and here consider the (remaining) *unconstrained* optimization problem

$$\min_u J(u, d) \tag{1}$$

Since any real operational problem is constrained, the word "remaining" needs to be emphasized. It means that if it is optimal (for the original problem with constraints) to keep some variable at constraints, then we assume these are implemented ("active constraint control"), and consider the remaining unconstrained problem. The reason for this is that it is for the unconstrained degrees of freedom that the selection of controlled variables is an issue (the other degrees of freedom are simply used to satisfy the active constraints). We further assume that the (economic) cost is determined by the steady-state behavior.

For the optimization itself it does not really matter which "base set" variables u we select for the (remaining) unconstrained degrees of freedom as long as they form an independent set. They may consist, for example, of a subset of the physical manipulators (valves) that effect the steady-state behavior.

For any value of the disturbances d we can solve the (remaining) unconstrained optimization problem and obtain $u_{\text{opt}}(d)$ for which

$$\min_u J(u, d) = J(u_{\text{opt}}(d), d) \stackrel{\text{def}}{=} J_{\text{opt}}(d)$$

Instead of evaluating the cost, it may be better to evaluate the *non-negative* loss function. The loss function expresses the difference between the actual operating costs (e.g. obtained when we adjust u in order to keep c at a given setpoint) and the optimal operating cost (obtained with $u = u_{\text{opt}}(d)$),

$$L(u, d) = J(u, d) - J_{\text{opt}}(d) \tag{2}$$

The objective of the operation is to minimize J (or some average of J), or equivalently to minimize the loss L. The loss function is zero if we use the optimal policy $u = u_{\text{opt}}(d)$.

Remarks:

1. If the original problem is linear (linear cost function with linear constraints), then it is well known that from Linear Programming theory that there will be no remaining unconstrained variables, so we are here studying nonlinear problems.

2. We are also implicitly assuming that the set of active constraints for the original problem does not change (for example, when we have disturbances d).

2.2 Implementation and setpoint errors

Let us first consider an *"open-loop" implementation* where we attempt to keep the "base set" independent variables u constant at the values u_s. With this implementation, operation may be non-optimal (with a positive loss) due to the presence of an setpoint error, $v_u(d) = u_s - u_{\text{opt}}(d)$, and an implementation error, $n_u = u - u_s$. The overall error, the difference between the actual and optimal input which causes a positive loss, is the sum of these two errors

$$e_u \stackrel{\text{def}}{=} u - u_{\text{opt}}(d) = v_u(d) + n_u \tag{3}$$

The open-loop policy is often poor; both because the optimal input value often depends strongly on the disturbance (so v_u is large), and because we are not able to measure and implement u accurately (so n_u is large).

Let us next consider the *"closed-loop" implementation* where we attempt to keep the variable(s) c constant at the setpoint c_s. Here c may be selected from the (candidate) measured variables y. We normally also include in the candidate measured variables the "base set" variables u used in the open-loop policy:

$$y \text{ also includes the variables } u$$

The reason is that the variables u are normally physical variables (e.g. valve positions or flows) which can be set at given values. The measured variables are a function of u and d,

$$y = f_y(u, d) \tag{4}$$

and the controlled variables c are combinations of these, that is,

$$c = h(y) \tag{5}$$

where the function h is free to choose. Examples of possible controlled variables are the difference between two measurements, $h(y) = y_1 - y_2$, the ratio between two inputs, $h(y) = u_1/u_2$ (assuming that the inputs are measured), and so on. An open-loop policy is obtained with $h(y) = u$. Substituting (4) into (5) gives

$$c = h(f_y(u, d), u) = f_c(u, d) \tag{6}$$

We assume that the number of controlled variables c equals the number of independent variables u, or more exactly that we starting from $c = f_c(u, d)$ can derive the inverse relationship

$$u = f_c^{-1}(c, d) \tag{7}$$

where the function f_c^{-1} exists and is unique.

To compare the alternative choices for c we consider the loss with constant setpoints c_s. The actual value of the controlled variable c will differ from its optimal value $c_{\text{opt}}(d)$ due to the presence of

1. **Setpoint error:**

$$v(d) = c_s - c_{\text{opt}}(d) \tag{8}$$

The setpoint c_s is generally different from the optimal value $c_{\text{opt}}(d)$, for example, due to disturbances d.

2. **Implementation error:**

$$n = c - c_s \tag{9}$$

The actual value of c is generally different from c_s, for example, due to measurement error or imperfect control.

The total error that results in a positive loss is then

$$e_c \stackrel{\text{def}}{=} c - c_{\text{opt}}(d) = v(d) + n \tag{10}$$

The two errors $v(d)$ and n are generally independent. Clearly, we would like to have both $v(d)$ and n small.

2.3 Linearized models

As already mentioned, the outputs c are related to the inputs and disturbances by the relationship $c = f_c(u, d)$ in (6). For small devations from the nominal point (denoted *) we may use a linearized relationship between the unconstrained degrees of freedom u and any candidate set of controlled variables c. The linearized model becomes

$$\Delta c = G\Delta u + G_d \Delta d \tag{11}$$

where $\Delta u = u - u^*$, $\Delta d = d - d^*$, $\Delta c = c - c^*$, $G = (\partial f_c/\partial u)^{*T}$ and $G_d = (\partial f_c/\partial d)^{*T}$. Similarly, a linearization of the relationship $y = f_y(u, d)$ in (4) yields

$$\Delta y = G^y \Delta u + G_d^y \Delta d \tag{12}$$

where $G^y = (\partial f_y/\partial u)^{*T}$ and $G_d^y = (\partial f_y/\partial d)^{*T}$, and linearization of $c = h(y)$ in (5) yields

$$\Delta c = H_y \Delta y \tag{13}$$

where the matrix $H_y = (\partial h/\partial y)^T$ is free to choose. Note that H_y is not a function of the operating point. Combining these equations yields

$$G = H_y G^y \tag{14}$$

$$G_d = H_y G_d^y \tag{15}$$

From a linear point of view, the issue of selecting controlled variables is then to find the optimal choice for the matrix H_y.

3 Taylor series analysis

3.1 Expansion of cost function (around fixed nominal point)

We assume that the cost function J is smooth, or more precisely twice differentiable, at the operating point we are considering. A second order Taylor series expansion of the cost function about the nominal point(u^*, d^*) gives

$$\begin{aligned} J(u,d) \quad &= \quad J^* + J_u^{*T}(u-u^*) + J_d^{*T}(d-d^*) + \frac{1}{2}(u-u^*)^T J_{uu}^*(u-u^*) \\ &\quad + \frac{1}{2}(d-d^*)^T J_{dd}^*(d-d^*) + (d-d^*)^T J_{du}^*(u-u^*) + \mathcal{O}^3 \end{aligned} \tag{16}$$

where

$$J^* = J(u^*, d^*)$$

$$J_u^* = \left(\frac{\partial J}{\partial u}\right)^*$$

$$J_d^* = \left(\frac{\partial J}{\partial d}\right)^*$$

$$J_{uu}^* = \left(\frac{\partial^2 J}{\partial u^2}\right)^*$$

$$J_{dd}^* = \left(\frac{\partial^2 J}{\partial d^2}\right)^*$$

$$J_{du}^* = \left(\frac{\partial^2 J}{\partial d \partial u}\right)^*$$

We can write the expansion in (16) more compactly as

$$J(u,d) = J^* + \begin{pmatrix} J_u^* & J_d^* \end{pmatrix}^T \begin{pmatrix} \Delta u \\ \Delta d \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \Delta u \\ \Delta d \end{pmatrix}^T \mathcal{H}^* \begin{pmatrix} \Delta u \\ \Delta d \end{pmatrix} \tag{17}$$

where $\mathcal{H}^*$ is the Hessian matrix of J with respect to $\begin{pmatrix} \Delta u \\ \Delta d \end{pmatrix}^T$,

$$\mathcal{H}^* = \begin{pmatrix} J_{uu}^* & J_{ud}^* \\ J_{du}^* & J_{dd}^* \end{pmatrix}$$

The Hessian matrix is always symmetric, so J_{uu}^* and J_{dd}^* are symmetric and $J_{ud}^* = {J_{du}^*}^T$.

If the nominal point is optimal (so J is at a minimum) then

1. $J_u^* = 0$ because the Jacobian with respect to the independent variables must be zero at the optimum.
2. $\Delta u^T J_{uu}^* \Delta u$ is positive for any nonzero vector Δu, i.e. J_{uu}^* is positiv definite: $J_{uu}^* > 0$ (if the minimum is a saddle the $\Delta u^T J_{uu}^* \Delta u$ is zero in some direction and J_{uu}^* is positive semidefinite, $J_{uu}^* \geq 0$).

3.2 Optimal input as a function of d

We assume here that the nominal operating point (u^*, d^*) is optimal so we have $d = d^*$ and $u^* = u_{\text{opt}}(d^*)$, and the Jacobian must be zero,

$$J_u^* = \left(\frac{\partial J}{\partial u}\right)^* = 0$$

Next, consider a disturbance and input change so that the new operating point is (u, d) with the new Jacobian

$$J_u = \frac{\partial J}{\partial u}$$

An first-order expansion of the Jacobian gives

$$J_u = J_u^* + J_{uu}^*(u - u^*) + {J_{ud}^*}^T (d - d^*)$$

We assume that we change the input so that also the new operating point is optimal, i.e. $u = u_{\text{opt}}(d)$. Then also the new Jacobian is zero, i.e. $J_u = 0$, and we get

$$0 = J_{uu}^*(u_{\text{opt}}(d) - u_{\text{opt}}(d^*)) + {J_{ud}^*}^T (d - d^*) \tag{18}$$

and we derive the following first-order accurate approximation of the optimal input when there is a disturbance change

$$\boxed{u_{\text{opt}}(d) - u_{\text{opt}}(d^*) = -{J_{uu}^*}^{-1} J_{ud}^* (d - d^*)} \tag{19}$$

We may also express the setpoint error $v(d)$ directly in terms of the disturbance. Using the linearized model in (11) we get

$$c_{\text{opt}}(d) - c_{\text{opt}}(d^*) = G(u_{\text{opt}}(d) - u_{\text{opt}}(d^*)) + G_d(d - d^*)$$

and assuming the setpoints are nominally optimal, $c_s = c_{\text{opt}}(d^*)$, (19) gives a first-order accurate approximation of the setpoint error

$$v(d) = c_{\text{opt}}(d^*) - c_{\text{opt}}(d) = \left(G J_{uu}^{*\,-1} J_{du}^* - G_d\right)\ (d - d^*) \tag{20}$$

3.3 Expansion of loss function (around moving optimal point)

Above we expanded the *cost function* at the *nominal* point (u^*, d^*), and this required a term for the deviation of the disturbance from its nominal value, $d - d^*$. Here, we instead expand the *loss function* at the *optimal* point $(u_{\text{opt}}(d), d)$ for each given disturbance d, that is, the point about which we expand moves with the disturbance, and no term $d - d^*$ is needed. For a given disturbance d, a second-order Taylor expansion of the cost function $J(u, d)$ about the optimal point gives

$$J(u, d) \approx J(u_{\text{opt}}(d), d) + {J_u}^T(u - u_{\text{opt}}(d)) + \frac{1}{2}(u - u_{\text{opt}}(d))^T J_{uu}(u - u_{\text{opt}}(d)) \tag{21}$$

where

$$J_u = \left(\frac{\partial J}{\partial u}\right)^{\text{opt}} = 0$$

$$J_{uu} = \left(\frac{\partial^2 J}{\partial u^2}\right)^{\text{opt}}$$

The loss function is defined as

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

and we derive the following very useful second-order expansion of the loss function (Skogestad and Postlethwaite 1996)

$$\boxed{L = \frac{1}{2}(u - u_{\text{opt}}(d))^T J_{uu}(u - u_{\text{opt}}(d)) = \frac{1}{2}e_u^T J_{uu} e_u} \tag{22}$$

This expression quantifies how a (suffciently small) input deviation $e_u eqdef u - u_{\text{opt}}(d)$ affects the loss. From (11) we have, for a given disturbance d, that the corresponding deviation in the controlled variables c is $c - c_{\text{opt}}(d) = G\,(u - u_{\text{opt}}(d))$. Here G is assumed to be invertible, see (7), and using (10) we derive

$$\boxed{e_u = G^{-1}(v(d) + n)} \tag{23}$$

where v is the setpoint error and n the implementation error for c.
Remarks:

1. Alternatively, we can express the loss in terms of $e_c \stackrel{\text{def}}{=} c - c_{\text{opt}}(d) = v(d) + n$:

$$L = \frac{1}{2}e_c^T J_{cc} e_c \tag{24}$$

where

$$J_{cc} = {G^{-1}}^T J_{uu} G^{-1}$$

2. The Hessian matrix J_{uu} is independent of the choice of controlled variables c.
3. Expressions (22)-(24) do not require that the setpoints c_s are nominally optimal.
4. Note that the Jacobian J_{uu} in (21) is evaluated at the point $(u_{\mathrm{opt}}(d), d)$ which is different from the nominal point (u^*, d^*) considered in (16). However, if the nominal point is optimal and the disturbance change is small ($\Delta d = d - d^*$ is small) then $J_{uu} \approx J^*_{uu}$. This may be used in an **alternative proof of (22)**: Assume that the nominal point is optimal such that $J^*_u = 0$. Using (16) we derive second-order expansions for $J(u,d)$ and $J(u_{\mathrm{opt}}, d)$ about the nominal value $J^* = J(u^*, d^*)$. Subtracting these expansions gives a second-order expansion for the loss

$$L(u,d) = J(u_{\mathrm{opt}}, d) - J(u,d) \approx \frac{1}{2}(u-u^*)^T J^*_{uu}(u-u^*) - \frac{1}{2}(u_{\mathrm{opt}}-u^*)^T J^*_{uu}(u_{\mathrm{opt}}-u^*) + (u-u_{\mathrm{opt}})^T J^*_{ud}(d-d^*) \quad (25)$$

Here from (18) for the optimal input we have that

$$J_{ud}(d-d^*) = -J^*_{uu}(u_{\mathrm{opt}} - u^*)$$

and substituting this into (25) gives after some algebra the desired answer

$$L(u,d) = \frac{1}{2}(u - u_{\mathrm{opt}})^T J^*_{uu}(u - u_{\mathrm{opt}})$$

where $J_{uu} \approx J^*_{uu}$, and we have rederived (22).

3.4 Loss with constant setpoints

We here assume that the setpoints c_s are nominally optimal, $c_s = c_{\mathrm{opt}}(d^*)$. Substitution of (20) into (23) then gives the following expression for the input error $e_u \stackrel{\mathrm{def}}{=} u - u_{\mathrm{opt}}(d)$:

$$\boxed{e_u = \left(J^{*\,-1}_{uu} J^*_{ud} - G^{-1}G_d\right)(d - d^*) + G^{-1}n} \quad (26)$$

and we find that the loss $L = \frac{1}{2}e_u^T J^*_{uu} e_u$ may be expressed as a function of d and n. Note that this expression assumes that the setpoints are nominally optimal.

4 Selection of controlled variables: Simple methods

4.1 Requirements for controlled variables

To minimize the loss $L = \frac{1}{2}e_u^T J_{uu} e_u$ we want a small input error e_u. From $e_u = G^{-1}(v(d) + n)$ in (23) we then derive the following **four requirements for of a good candidate controlled c variable** (Skogestad and Postlethwaite 1996):

Requirement 1. Its optimal value is insensitive to disturbances (so that the setpoint error v is small)

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

Requirement 3. Its value is sensitive to changes in the manipulated variables u, that is, the gain G from u to c is large (so that G^{-1} is small). (Equivalently, from (24) $J_{cc} = G^{-1}J_{uu}G$ should be small, that is the optimum should be "flat" with respect to the variable c.)

Requirement 4. For cases with two or more controlled variables, the selected variables are not closely correlated (such that G is not close to singular, resulting in a large G^{-1}).

All four requirements need be satisfied. *In short, these requirements tell that we should select variables c for which their 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).

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).

4.2 Minimum singular value rule

From (23) we have that

$$L = \frac{1}{2} e_u^T J_{uu} e_u = \frac{1}{2} \|z\|_2^2 \tag{27}$$

where

$$z = J_{uu}^{1/2} e_u = J_{uu}^{1/2} G^{-1} e_c \tag{28}$$

where $e_c = c - c_{\text{opt}}(d) = v + n$ and and $\|z\|_2$ denotes the 2-norm of the vector. We assume that each controlled variable c_i is scaled such that the sum of its optimal range (v_i) and its implementation error (n_i) is unity, and that for combined errors the 2-norm is less than 1, i.e. $\|e'_c\|_2 = \|c' - c'_{\text{opt}}\|_2 = 1$ We also assume that each "base variable" u is scaled such that a unit change in each input has the same effect on the cost function J (actually, we must require that the Hessian $J'_{uu} = \left(\frac{\partial^2 J}{\partial u'^2}\right)$ is a scalar times identity matrix, i.e. $J'_{uu} = \alpha I$). We use primes ($'$) to show that the variables u and y have been scaled, and G' denotes the scaled steady-state gain matrix from u' to c'. Then the resulting worst-case loss is

$$\max_{\|e'_c\|_2 \le 1} L = \max_{\|e'_c\|_2 \le 1} \frac{1}{2} \|z'\|_2^2 = \frac{1}{2} (\bar{\sigma}({J'_{uu}}^{1/2} G^{-1})^2 = \frac{1}{2} (\bar{\sigma}(\alpha^{1/2} G'^{-1})^2 = \frac{\alpha}{2} \frac{1}{\underline{\sigma}(G')^2} \tag{29}$$

where $\alpha = \bar{\sigma}(J'_{uu})$ is a constant (independent of the choice of c), and $\underline{\sigma}(G')$ denotes the minimum singular value of the matrix G'. The second equality follows since $\bar{\sigma}$ is the induced 2-norm of a matrix. The last equality follows since $\bar{\sigma}(G'^{-1}) = 1/\underline{\sigma}(G')$.

Thus, to minimize the loss L we should maximize $\underline{\sigma}(G')$. In summary, we have derived the following rule:

> *Assume that the base variables u are selected and scaled such that they all have a similar effect on the cost (or more precisely such that $J'_{uu} = \alpha I$), assume that we have scaled each candidate controlled variable c such that the expected variation in $c' - c'_{\text{opt}}$ is of magnitude 1 (including the effect of both disturbances and control error), and let G' denote the resulting scaled gain matrix from u' to y' ($\Delta y' = G' \Delta u'$). Then select controlled variables c that maximize the minimum singular value of G', $\underline{\sigma}(G')$.*

Interestingly, we note that this rule does not depend on the actual expression for the objective function J, but it does enter indirectly through the variation of c_{opt} with d, which enters into the scaling. A more detailed procedure with scaling is given in the Appendix.

There are two limitations with the singular value rule when applied to multivariable systems (with more than one controlled variable):

1. We have assumed that it is possible to scale the inputs such that $J'_{uu} = (\partial^2 J / \partial u'^2)^{\text{opt}}$ is close to a constant times identity matrix. If this is not the case then one should search for another set of "base case" inputs. This may not always be possible.

2. We have assumed that any output deviation $e_c = c' - c'_{\rm opt}$ satisfying $\|e_c\|_2 \le 1$ is allowed, including the "worst-case" combination corresponding to $\underline{\sigma}(G')$. For this to be true, we must assume the variations $c'_i - c'_{i_{\rm opt}}$ for the each variable c_i are independent. This generally holds for the implementation error n_i, but not for the setpoint error v_i since their optimal values as a function of d are generally correlated.

Remark. We here use the minimum singular value rule to select controlled variables (outputs). The minimum singular value can also be used as a tool for selecting manipulated inputs variables (inputs) (Morari 1983), but this is actually an unrelated condition which requires a different scaling of the variables.

5 Selection of controlled variables: Exact local method

In this section we assume that the setpoints are nominally optimal,

$$c_s = c_{\rm opt}(d^*)$$

Upon substitution of (26) into (22) we can write the loss as

$$L = \frac{1}{2}\|z\|_2^2 \tag{30}$$

where

$$z = J_{uu}^{1/2}\left[\left(J_{uu}^{-1}J_{ud} - G^{-1}G_d\right)(d-d^*) + G^{-1}n\right] \tag{31}$$

Let the elements in the positive diagonal matrices W_d and W_n represent the expected magnitudes of the individual disturbances and implementation errors, i.e. let

$$d - d^* = W_d d'$$

$$n = W_n n'$$

where d' and n' are normalized to have magnitude less than 1. More precisely, we assume that the *combined* disturbances and implementation errors are two-norm bounded (see the discussion section for a justification on why to use the two-norm),

$$\| \begin{pmatrix} d' \\ n' \end{pmatrix} \|_2 \le 1 \tag{32}$$

Under these assumptions the worst-case loss is

$$L = \frac{\bar{\sigma}(M)^2}{2}) \tag{33}$$

where

$$M = \begin{pmatrix} M_d & M_n \end{pmatrix} \tag{34}$$

$$M_d = J_{uu}^{1/2}\left(J_{uu}^{-1}J_{ud} - G^{-1}G_d\right)W_d$$

$$M_n = J_{uu}^{1/2}G^{-1}W_n$$

and $J_{uu}^{1/2}J_{uu}^{1/2} = J_{uu}$. This follows since the induced (worst-case) 2-norm of a matrix is equal to its maximimum singular value. Minimization of the loss L for any (worst-case) combination of disturbances and implementation errors is then equivalent to minimizing $\bar{\sigma}(M)$.

Note that this method does not suffer from the limitations of the singular value rule. This is because the disturbances d and implementation errors n are generally independent variables, so that any combination satisfying (32) is allowed.

5.1 Procedure for exact evaluation of loss

The exact local procedure for analyzing candidate sets of controlled variables c then becomes:

1. Define the optimal operation problem (specify the cost function J).
2. Solve the nominal optimization problem with respect to the degrees of freedom u and and find the second-order derivatives of the cost, J_{uu} and J_{ud}.
3. For each candidate set c of controlled variables obtain the linear model $\Delta c = G\Delta u + G_d \Delta d$.
4. Define the uncertainty:
   (a) The elements in the diagonal matrix W_d represents the magnitude of each disturbance (this is independent of c).
   (b) The elements in the diagonal matrix W_n represent the magnitude of the implementation error for a candidate set c (e.g. due to measurement error).
5. For each *candidate set* c compute the singular value $\bar{\sigma}(M)$, of the matrix M in (34).
6. Select the set c with the smallest $\bar{\sigma}(M)$, as this minimizes the loss $L = \frac{1}{2}\bar{\sigma}(M)^2$.

Some comments:

1. It is clear from M_n in (34) that G^{-1} should be small to minimize effect of implementation errors.
2. Yi and Luyben (1995) suggest to select controlled variables that minimize the input usage in response to disturbances, $(\partial u/\partial d)_c = G^{-1}G_d$. This makes some sense as it is certainly clear from M_d that we do not want to make this term very large. However, from M_d in (34) we see that we do not want to make $G^{-1}G_d$ zero (unless $J_{uu}^{-1}J_{ud}$ is zero, which it rarely is) , so this certainly is not a general rule.

5.2 Optimal linear combination of measurements

Assume that we from a given a set of measurements y (which generally also includes the variables u), want to find the best linear combination to control,

$$\Delta c = H_y \Delta y$$

This may be done be finding the matrix H_y that minimizes the loss $L = \bar{\sigma}(M)$ in (33). Here the matrix H_y is free to choose, except that we make the restriction that the number of controlled variables (c's) equals the number of independent inputs (u's) (recall the comments following (13)). We also need to identify the implementation error (measurement noise) associated with controlling each of the candidate measurements y and collect these in the diagonal matrix W_n^y. The matrices that enter in M in (34) then become

$$G = H_y G^y; \quad G_d = H_y G_d^y; \quad W_n = H_y W_n^y$$

and we can numerically search for the matrix H_y that minimizes $\bar{\sigma}(M)$. This approach gives the optimal local measurement combination in terms of minimizing the loss when there are disturbances and implementation errors bounded as in (32). Note that H_y generally is a "full" matrix, that is, it is generally optimal to combine all available measurements y in the controlled variables n.

6 Example

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

$$y_1 = 0.1(u-d); \quad y_2 = 20u; \quad y_3 = 10u - 5d$$

We further assume that the magnitude of the disturbance and the measurement noise is 1, that is, $|d| \leq 1$ and $|n_i| \leq 1$. Which controlled variable should we select?

Optimization. For the nominal case with $d^* = 0$ we have in all three cases that $y_{\rm opt}(d^*) = 0$ so we select in all three cases $c_s = y_{\rm opt}(d^*) = 0$. For a disturbance d we have $u_{\rm opt}(d) = d$, which confirms (19), and the optimal value of the controlled variable for the three cases are

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

Exact nonlinear evaluation of loss. The losses can for this example be evaluated analytically, and we find for the three cases

$$L_1 = (10n_1)^2; \quad L_2 = (0.05n_2 - d)^2; \quad L_3 = (0.1n_3 - 0.5d)^2$$

(For example, in case 3, we have $u = (y_3 + 5d)/10$ and with $y_3 = y_{3s} + n_3 = n_3$ we get $J = (u-d)^2 = (0.1n_3 + 0.5d - d)^2$). With $|d| \leq 1$ and $|n_i| \leq 1$ the worst-case values of the losses (with $|d| = 1$ an $|n_i| = 1$) are $L_1 = 100$, $L_2 = 1.05^2 = 1.1025$ and $L_3 = 0.6^2 = 0.36$, and we find that the $c = y_3$ is the best overall choice for self-optimizing control and $c = y_1$ is the worst. (However, with no implementation error y_1 would be the best, and with no disturbances y_2 would be the best.)

We will now compare this result with the linear methods derived in this paper.

Linearized models. The linearized models for the three choices of controlled outputs ($c = y_1$, $c = y_2$, $c = y_3$) are

$$G_1 = 0.1, G_{d1} = -0.1$$

$$G_2 = 20, G_{d2} = 0$$

$$G_3 = 10, G_{d3} = -5$$

At the nominal operating point we have (independent of the choice of controlled variable)

$$J_u^* = 0, \; J_d^* = 0$$

$$\mathcal{H}^* = \begin{pmatrix} J_{uu}^* & J_{ud}^* \\ J_{ud}^* & J_{dd}^* \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$$

1. Singular value rule. For the three choices of controlled outputs we have and $\underline{\sigma}(G_1) = 0.1$, $\underline{\sigma}(G_2) = 20$ and $\underline{\sigma}(G_3) = 10$. This would indicate that $c = y_2$ is the best choice, but this is only correct with no disturbances. The reason for the error is that we have not scaled the output variables properly; in particular, we have not take into account the effect of the disturbances on the magnitude of $c - c_{\rm opt}(d)$.

Let us now follow the singulat value procedure given in the Appendix. We use procedure (Method B) since the Hessian matrix is available.

1. Scale each input u_j such that a unit deviation in each input from its optimal value has the same effect on the cost function J:

$$D_u = u_{scl} = \left[J_{uu}^{1/2}\right] = \sqrt{2}$$

2. For each candidate output variable obtain its maximum setpoint error $v_{max,i}$ due to variation in disturbances:

$$v_i = c_{i,opt}(d^*) - c_{i,opt}(d) = \left(G_i J_{uu}^{-1} J_{ud} - G_{d,i)}\right)\ (d_{max} - d^*)$$

Here, $(d_{max} - d^*) = 1$ and we get

$$v_1 = 0.1 \cdot \frac{1}{2} \cdot (-2) - (-0.1) = 0$$

$$v_2 = 20 \cdot \frac{1}{2} \cdot (-2) - -0 = -20$$

$$v_3 = 10 \cdot \frac{1}{2} \cdot (-2) - (-5) = 5$$

which confirms that with $u_{\text{opt}} = d$, we have $y_{1,opt} = 0, y_{2,opt} = 20d$ and $y_{3,opt} = 5d$.

3. For each candidate output variable obtain its expected implementation error n_i (sum of measurement error and control error): It is given that $n_i = 1$.

4. Scale the candidate controlled variables such that for each variable i the sum of the magnitudes of v_i and the implementation error n_i is similar, that is, use $c_{scl,i} = |v_i| + |n_i|$. We get

$$D_{c1} = c_{scl,1} = 0 + 1 = 1$$

$$D_{c2} = c_{scl,2} = 20 + 1 = 21$$

$$D_{c3} = c_{scl,3} = 5 + 1 = 6$$

5. Obtain from (37) the scaled gain matrix, $G' = D_c^{-1} G D_u$, and from (29) the worst-case losses:

$$G'_1 = \frac{1}{1} \cdot 0.1 \cdot \sqrt{2} = 0.14; \quad L = \frac{1}{2|G'_1|^2} = 100$$

$$G'_2 = \frac{1}{21} \cdot 20 \cdot \sqrt{2} = 1.35; \quad L = \frac{1}{2|G'_2|^2} = 1.1025$$

$$G'_3 = \frac{1}{6} \cdot 10 \cdot \sqrt{2} = 2.36; \quad L = \frac{1}{2|G'_3|^2} = 0.360$$

6. Select as candidates those sets of controlled outputs that correspond to a large value of the minimum singular value $\underline{\sigma}(G')$: For a scalar G, $\underline{\sigma}(G) = |G|$, and the singular value rule tells us (as expected) that y_3 is the best choice for the controlled variable, followed by y_2 and y_1.

We note from the computed losses, that for this scalar case, the singular value rule gives identically the same results as the "exact" procedure, but this will not generally be the case when u is a vector.
2. Exact local method. Minimization of the loss L for any (worst-case) 2-norm combination of disturbances and implementation errors is equivalent to minimizing $\bar{\sigma}(M)$ where

$$M = \begin{pmatrix} M_d & M_n \end{pmatrix} \tag{35}$$

$$M_d = J_{uu}^{1/2} \left(J_{uu}^{-1} J_{ud} - G^{-1} G_d\right) W_d$$

$$M_n = J_{uu}^{1/2} G^{-1} W_n$$

where we in this case have

$$W_d = 1,\ W_n = 1$$

We find for $c = y_1$:

$$M_{d1} = \sqrt{2}\left(2^{-1} \cdot (-2) - 0.1^{-1} \cdot (-0.1)\right) \cdot 1 = 0$$

$$M_{n1} = \sqrt{2} \cdot 0.1^{-1} \cdot 1 = 10\sqrt{2}$$

which gives

$$L_1 = \frac{\bar{\sigma}(M_1)^2}{2} = \frac{1}{2}(\bar{\sigma}(0 \quad 10\sqrt{2})) = 100$$

Similarly, for $c = y_2$ and $c = y_3$:

$$L_2 = \frac{\bar{\sigma}(M_2)^2}{2} = \frac{1}{2}(\bar{\sigma}(-\sqrt{2} \quad \sqrt{2}/20)) = 1.0025$$

$$L_3 = \frac{\bar{\sigma}(M_3)^2}{2} = \frac{1}{2}(\bar{\sigma}(-\sqrt{2}/2 \quad \sqrt{2}/10)) = 0.26$$

The reason for the slight difference from the "exact" nonlinear results presented initially ($L_1 = 100, L_2 = 1.1025, L_3 = 0.36$) is not due to nonlinearity, but that we in the nonlinear evaluation allowed d and n individually to be less than 1, wheras we in the linear method assume that the *combined* 2-norm of d and n is less than 1. (For example, in the nonlinear evaluation the worst case was $|d| = 1$ and $|n| = 1$ which has a combined 2-norm of $\| \begin{pmatrix} 1 \\ 1 \end{pmatrix} \|_2 = 1.414$, which is not allowed in the exact local method where the combined 2-norm must be less than 1.) As argued in the discussion section, the use of the combined 2-norm is more reasonable from a practical point of view.

3. Optimal combination of measurements. Above we considered control of the invidual measurements, and considered the choices $c = y_1$, $c = y_2$ and $c = y_3$. We now want to find c as the the best linear combination of all candidate measurements y, where

$$y = (\, y_1 \quad y_2 \quad y_3 \quad u\,)^T$$

Note that we have included also the "base case " input u as a candidate measurement (assuming that this a manipulable variable). We assume that the implemention error n_i related to the three measurements (y_1, y_2, y_3) are all 1, and that implementation error n_u related to the manipulated input is also 1, so

$$W_n^y = I$$

where W_n^y is a 4×4 matrix. Furthermore,

$$G^y = \begin{pmatrix} G_1 \\ G_2 \\ G_3 \\ G_4 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 20 \\ 10 \\ 1 \end{pmatrix}$$

$$G_d^y = \begin{pmatrix} G_{d1} \\ G_{d2} \\ G_{d3} \\ G_{d4} \end{pmatrix} = \begin{pmatrix} -0.1 \\ 0 \\ -5 \\ 0 \end{pmatrix}$$

The matrices used to obtain M in (35) then become

$$G = H_y G^y$$

$$G_d = H_y G_d^y$$
$$W_n = H_y W_n^y$$

and we want to find $H_y = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 \end{pmatrix}$ that minimizes $\bar{\sigma}(M)$.

Numerical optimization yields $H_y = \begin{pmatrix} 0.02085 & -0.23933 & 0.97064 & -0.01197 \end{pmatrix}$, that is, the optimal combination of the three measurements and the manipulated input u is

$$c = 0.02085 y_1 - 0.23933 y_2 + 0.97064 y_3 - 0.01197 u$$

We note, as expected, that the most important contribution to c comes from the variable y_3. The loss is $L = 0.010494$, so it is reduced by a factor 25 compared to the previously best case ($L = 0.26$) with $c = y_3$. The reason for this large reduction is because we are able to reduce contribution from M_d to almost zero. It should also be noted that in the optimal matrix H_y we have $h_2/h_4 = 20$, which is not surprising, since this is the ratio between the noise level for these otherwise identical variables ($y_2 = 20u$ and $y_4 = u$).

7 Discussion

7.1 "Ideal" choice of controlled outputs

If we for the moment disregard the implementation error n, then the ideal choice of controlled outputs would be to have $v(d) = c_s - c_{\text{opt}}(d) = 0$ for any value of d. Here $c_s = c^*$ is constant, so to achieve this, we need the optimal value of the controlled variable to be independent of the disturbance. From (26) it is clear that a controlled variable that gives $G = J_{uu}$ and $G_d = J_{ud}$ satisfies this requirement.

An example of such an ideal output would be to have a direct measurement of the gradient of the cost function with respect to the input (since it is optimal for any disturbance to have this gradient zero, we could directly specify its setpoint at zero). In particular, consider the following output

$$c = f(u, d) = \alpha_1 \frac{\partial J(u, d)}{\partial u} + \alpha_0 = \alpha_1 J_u + \alpha_0 \tag{36}$$

where α_0 and α_1 are constants. To see that this output would be "ideal", we linearize (36) to get

$$\Delta c = \alpha_1 \left(J_{uu} \Delta u + J_{ud}^T \Delta d \right)$$

i.e. we find that $G = c_1 J_{uu}$ and $G_d = c_1 J_{ud}^T = J_{ud}$, which upon substitution into (20) gives $v = 0$.

However, as we see when studying, for example, selection of measurement locations in a distillation column, the implementation error n may be a very important factor, and the "ideal" output may not be the best after all.

7.2 Use of 2-norm for noise and disturbances

Instead of using the 2-norm in (32) we could have used the infinity-norm, $\| \begin{pmatrix} d' \\ n' \end{pmatrix} \|_\infty$ which would have allowed for the "extreme" case with all variables simultaneously at their "worst-case" values ($|d_k| = 1$, $|n_i| = 1$).

However, two reasons for using the combined 2-norm in (32) are:

1. From (30) the loss is given by the 2-norm of the vector z, and it is therefore mathematically convenient to use the 2-norm also for the disturbance and noise.

2. With the 2-norm it is *not* possible to have all the individual disturbance and noise variables simultaneously at their "worst-case" values ($|d_k| = 1$, $|n_i| = 1$). We argue that this is reasonable from a physical point of view, since it is unlikely that this extreme worst case will occur in practice. For example, for the case with three variables, the following three variable combinations have 2-norm equal to 1:

$$\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \|_2 = 1; \quad \| \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \|_2 = 1; \quad \| \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix} \|_2 = 1$$

We note that we can have only one variable at a time equal to 1, and if all three variables are equal in magnitude then their maximum value is $1/\sqrt{3} = 0.577$.

This property of the vector 2-norm is also reasonable from a statistical point of view, from which we expect that the use of many independent measurements reduces the variance. For example, assume that we have p independent measurements y_i of the same variable y, and we take their average $\bar{y} = \frac{1}{p}\sum_{i=1}^{p} y_i$ The noise (implementation error) on each individual measurement y_i is n_i, i.e. $y_i = y + n_i$, and we assume that the noise satisfies $\|n\|_2 \leq 1$, where n is a vector of the individual noises n_i. We assume that the true variable is equal to zero ($y = 0$). We then have $y_i = n_i$, and $\bar{y} = \frac{1}{p}\sum_{i=1}^{p} y_i = \frac{1}{p}\,(\underline{1}\; n)$, where $\underline{1} = (\,1 \quad 1 \quad \cdots \quad 1\,)$ is a row vector of length p, and n is a column vector of the invidual noises. The worst-case value of the measured average is then

$$|\bar{y}| = \frac{1}{p} \max_{\|n\|_2 \leq 1} (\underline{1}\; n) = \frac{1}{p}\bar{\sigma}(\underline{1}) = \frac{1}{p}\sqrt{p} = 1/\sqrt{p}$$

That is, the worst-case deviation is reduced as we increase the number p of measurements, and, importantly, this reduction $1/\sqrt{p}$ is equal to the (estimated) reduction in statistical variance when we have p measurements (samples) with the same mean. Thus further justifies the use of the vector 2-norm to bound n (On the other hand, if we used the vector infinity-norm, then we would not get any benefit of adding extra measurement, which is unreasonable).

8 Conclusion

We have derived an exact and numerically efficient local method for evaluating the loss when controlling the variables c a constant setpoints, by evaluating $\bar{\sigma}(M)$ where M is given in (34). This approach may also be used to numerically search for the optimal measurement combination, $\Delta c = H^y \Delta y$, to use as controlled variables.

Evaluation of the minimimum singular value $\underline{\sigma}(G')$, of the scaled gain matrix G' from the "inputs" u to the controlled variables c, provides a simple, but for multivariable cases generally not exact, means for evaluating the loss. However, it provides a simple method for quickly screening candidate sets of controlled variables. This follows since a small value of $\underline{\sigma}(G')$ is always bad, as it implies that there is a combination of implementation errors (n) that result in a large loss.

References

Morari, M. (1983). Design of Resilient Processing Plants -III. *Chemical Engineering Science* **38**(11), 1881–1891.

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* **26**(2), 220–232.

Skogestad, S. (2000). Plantwide control: the search for the self-optimizing control structure. *J. Proc. Control* **10**, 487–507.

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

Yi, C.K. and W.L. Luyben (1995). Evaluation of plant-wide control structures by steady-state disturbance sensitivity analysis. *Ind. Eng. Chem. Res.* **34**, 2393–2405.

Acknowledgement

We thank Mr. Vidar Alstad for helpful comments and assistance with Matlab calculations.

Appendix. Singular value rule with scaling

A procedure for use of the minimum singular value to select controlled outputs is summarized below. A key part of this procedure is the scaling, and we let a prime (′) denote the scaled model and variables. We have in the original units $\Delta c = G\Delta u$. We scale each input j and controlled variable i by a scaling factor

$$c_i' = \frac{c_i}{c_{scl,i}}, \quad u_j' = \frac{u_j}{u_{scl,j}}$$

The model in scaled units is then $\Delta c' = G'\Delta u'$ with

$$G' = D_c^{-1} G D_u \tag{37}$$

where the diagonal scaling matrices are

$$D_c = \mathrm{diag}\{c_{scl,i}\}, \quad D_u = \mathrm{diag}\{u_{scl,j}\} \tag{38}$$

Note that the scaling changes the units, but the problem itself is unchanged. Two method for generating the scalings are given depending on whether (A) we directly use the nonlinear model to generate the data, or (B) the Hessian matrices are available.

1. Scale each input u_j such that a unit deviation in each input from its optimal value has the same effect on the cost function J (i.e. such that $(\partial^2 J/\partial u^2)_{\mathrm{opt}}$ has identical diagonal elements equal (c), and is close to c times a unitary matrix):

   (Method A) For each input find the change $u_{scl,j}$ that increases the loss from 0 til L_{max} where L_{max} is the largest acceptable loss.

   (Method B) Select $u_{scl,j} = \left[J_{uu}^{1/2}\right]_{jj}$ (the square root of the corresponding diagonal element of the Hessian).

2. For each candidate output variable obtain its maximum setpoint error $v_{max,i}$ due to variation in disturbances:

(Method A) From the nonlinear model compute the optimal parameters (inputs and outputs) for various conditions (disturbances, operating points). (This yields a "look-up" table of optimal parameter values as a function of the operating conditions.). From this identify

$$v_i = \max\left(|c_{s_i} - c_{i_{\mathrm{opt,max}}}|, |c_{s_i} - c_{i_{\mathrm{opt,min}}}|\right)$$

(Method B) Assuming that the setpoints are nominally optimal, i.e. $c_s = c_{\mathrm{opt}}(d^*)$, we have from (20)

$$v_i = c_{i,opt}(d^*) - c_{i,opt}(d) = \left(G_i J_{uu}^{-1} J_{ud} - G_{d,i)}\right)\ (d_{max} - d^*)$$

where $d_{max} - d^*$ is a vector consisting of the expected magnitude of each disturbance.

3. For each candidate output variable obtain its expected implementation error n_i (sum of measurement error and control error).

4. Scale the candidate controlled variables such that for each variable i the sum of the magnitudes of v_i and the implementation error n_i is similar, which corresponds to selectingthe scaling

$$c_{scl,i} = |v_i| + |n_i|$$

5. Compute the scaling matrices D_c and D_u from (38) and obtain the matrix $G' = D_c^{-1} G D_u$ from (37).

6. Select as candidates those sets of controlled outputs that correspond to a large value of the minimum singular value $\underline{\sigma}(G')$.

Some comments and insights on the singular value rule:

1. Since $u - u_{\mathrm{opt}} = G^{-1}(v + n)$ we always want always G^{-1} small, that is, candidate variables corresponding to a small $\underline{\sigma}(G)$ may be eliminated.

2. The minimum singular value rule applies also to the case with "back-off", since we do not assume optimal nominal setpoints. We find that we should *not* use back-off for variables that correspond to large directions of G^{-1} (small directions of G), because then $G^{-1}v$ and thus the loss L is large.

3. It seems difficult to identify from these results promising alternatives, as this required either G^{-1} or v (or preferable both) to be small.