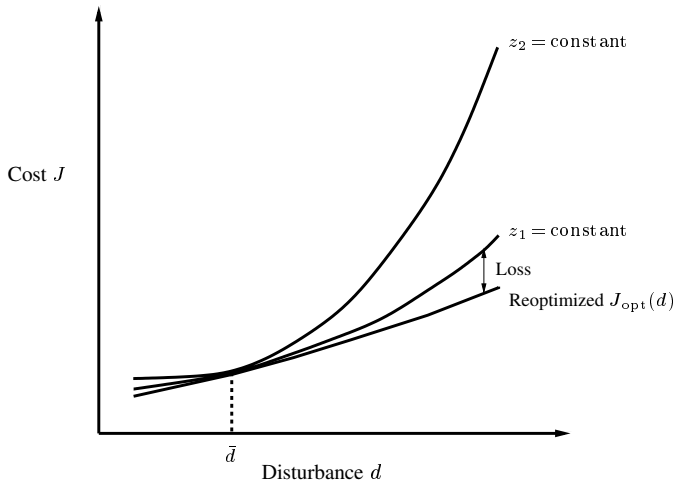

Figure 10.4: Loss imposed by keeping constant setpoint for the controlled variable. In this case z_1 is a better "self-optimizing" controlled variable than z_2.

10.3.2 Selecting controlled outputs: local analysis

We use here a local second-order accurate analysis of the loss function. From this, we derive the useful minimum singular value rule, and an exact local method; see Halvorsen et al. (2003) for further details. Note that this is a local analysis, which may be misleading; for example, if the optimum point of operation is close to infeasibility.

Consider the loss $L = J(u,d) - J_{\text{opt}}(d)$, where d is a fixed (generally non-zero) disturbance. We here make the following additional assumptions:

1. The cost function J is smooth, or more precisely twice differentiable.
2. As before, we assume that the optimization problem is unconstrained. If it is optimal to keep some variable at a constraint, then we assume that this is implemented ("active constraint control") and consider the remaining unconstrained problem.
3. The dynamics of the problem can be neglected when evaluating the cost; that is, we consider steady-state control and optimization.
4. We control as many variables z as there are available degrees of freedom, i.e. $n_z = n_u$.

For a fixed d we may then express $J(u,d)$ in terms of a Taylor series expansion in u around the optimal point. We get

$$J(u,d) = J_{\text{opt}}(d) + \underbrace{\left(\frac{\partial J}{\partial u}\right)^T_{\text{opt}}}_{=0}(u - u_{\text{opt}}(d)) + \frac{1}{2}(u - u_{\text{opt}}(d))^T \underbrace{\left(\frac{\partial^2 J}{\partial u^2}\right)_{\text{opt}}}_{=J_{uu}}(u - u_{\text{opt(d)}}) + \cdots \tag{10.2}$$

We will neglect terms of third order and higher (which assumes that we are reasonably close to the optimum). The second term on the right hand side in (10.2) is zero at the optimal point

for an unconstrained problem. Equation (10.2) quantifies how a non-optimal input $u - u_{\rm opt}$ affects the cost function. To study how this relates to output selection we use a linearized model of the plant

$$z = Gu + G_d d \tag{10.3}$$

where G and G_d are the steady-state gain matrix and disturbance model respectively. For a fixed d, we have $z - z_{\rm opt} = G(u - u_{\rm opt})$. If G is invertible we then get

$$u - u_{\rm opt} = G^{-1}(z - z_{\rm opt}) \tag{10.4}$$

Note that G is a square matrix, since we have assumed that $n_z = n_u$. From (10.2) and (10.4) we get the second-order accurate approximation

$$L = J - J_{\rm opt} \approx \frac{1}{2}\left(z - z_{\rm opt}\right)^T G^{-T} J_{uu} G^{-1}\left(z - z_{\rm opt}\right) \tag{10.5}$$

where the term $J_{uu} = (\partial^2 J/\partial u^2)_{\rm opt}$ is independent of z. Alternatively, we may write

$$L = \frac{1}{2}\|\widetilde{z}\|_2^2 \tag{10.6}$$

where $\widetilde{z} = J_{uu}^{1/2} G^{-1}(z - z_{\rm opt})$. These expressions for the loss L yield considerable insight. Obviously, we would like to select the controlled outputs z such that $z - z_{\rm opt}$ is zero. However, this is not possible in practice because of (1) varying disturbances d and (2) implementation error e associated with control of z. To see this more clearly, we write

$$z - z_{\rm opt} = z - r + r - z_{\rm opt} = e + e_{\rm opt}(d) \tag{10.7}$$

where

$$e_{\rm opt}(d) \triangleq r - z_{\rm opt}(d)$$
$$e \triangleq z - r$$

First, we have an optimization error $e_{\rm opt}(d)$ because the algorithm (e.g. the cook book for cake baking) gives a desired r which is different from the optimal $z_{\rm opt}(d)$. Second, we have a control or implementation error e because control is not perfect; either because of poor control performance or because of an incorrect measurement (steady-state bias) n^z. If we have integral action in the controller, then the steady-state control error is zero, and we have $e = n^z$. If z is directly measured then n^z is its measurement error. If z is a combination of several measurements y, $z = Hy$, see Figure 10.2(b), then $n^z = Hn^y$, where n^y is the vector of measurement errors for the measurements y.

In most cases, the errors e and $e_{\rm opt}(d)$ can be assumed independent. The maximum value of $|z - z_{\rm opt}|$ for the expected disturbances and implementation errors, which we call the "expected optimal span", is then

$$\mathrm{span(z)} = \max_{\mathrm{d,e}} |\mathrm{z - z_{opt}}| = \max_{\mathrm{d}} |\mathrm{e_{opt}(d)}| + \max_{\mathrm{e}} |\mathrm{e}| \tag{10.8}$$

Example 10.1 Cake baking continued. *Let us return to the question: why select the oven temperature as a controlled output? We have two alternatives: a closed-loop implementation with $z = T$ (the oven temperature) and an open-loop implementation with $z = u = Q$ (the heat input). From experience, we*

know that the optimal oven temperature $T_{\rm opt}$ is largely independent of disturbances and is almost the same for any oven. This means that we may always specify the same oven temperature, say $r = T_s = 190^\circ C$, as obtained from the cook book. On the other hand, the optimal heat input $Q_{\rm opt}$ depends strongly on the heat loss, the size of the oven, etc., and may vary between, say, 100 W and 5000 W. A cook book would then need to list a different value of $r = Q_s$ for each kind of oven and would in addition need some correction factor depending on the room temperature, how often the oven door is opened, etc. Therefore, we find that it is much easier to get $e_{\rm opt} = T_s - T_{\rm opt}$ [°C] small than to get $e_{\rm opt} = Q_s - Q_{\rm opt}$ [W] small. Thus, the main reason for controlling the oven temperature is to minimize the optimization error. In addition, the control error e is expected to be much smaller when controlling temperature.

From (10.5) and (10.7), we conclude that we should select the controlled outputs z such that:

1. G^{-1} is small (i.e. G is large); the choice of z should be such that the inputs have a large effect on z.
2. $e_{\rm opt}(d) = r - z_{\rm opt}(d)$ is small; the choice of z should be such that its optimal value $z_{\rm opt}(d)$ depends only weakly on the disturbances (and other changes).
3. $e = z - r$ is small; the choice of z should be such that it is easy to keep the control or implementation error e small.
4. G^{-1} is small, which implies that G should not be close to singular. For cases with two or more controlled variables, the variables should be selected such that they are independent of each other.

By proper scaling of the variables, these four requirements can be combined into the "maximize minimum singular value rule" as discussed next.

10.3.3 Selecting controlled outputs: maximum scaled gain method

We here derive a very simple method for selecting controlled variables in terms of the steady-state gain matrix G from inputs u (unconstrained degrees of freedom) to outputs z (candidate controlled variables).

Scalar case. In many cases we only have one unconstrained degree of freedom (u is a scalar and we want to select one z to control). Introduce the scaled gain from u to z:

$$G' = G/\text{span}(z)$$

Note form (10.8) that $\text{span}(z) = \max_{d,e} |z - z_{\rm opt}|$ includes both the optimization (setpoint) error and the implementation error. Then, from (10.5), the maximum expected loss imposed by keeping z constant is

$$L_{max} = \frac{|J_{uu}|}{2} \left(\frac{\max_{d,e} |z - z_{\rm opt}|}{G} \right)^2 = \frac{|J_{uu}|}{2} \frac{1}{|G'|^2} \tag{10.9}$$

Here $|J_{uu}|$, the Hessian of the cost function, is independent of the choice for z. From (10.9), we then get that *the "scaled gain" $|G'|$ should be maximized to minimize the loss*. Note that the loss decreases with the square of the scaled gain. For an application, see Example 10.6 on page 398.

Multivariable case. Here u and z are vectors. Introduce the scaled outputs $z' \triangleq S_1 z$ and the scaled plant $G' = S_1 G$. Similar to the scalar case we scale with respect to the span,

$$S_1 = \text{diag}\{\frac{1}{\text{span}(z_i)}\} \tag{10.10}$$

where

$$\mathrm{span(z_i)} = \max_{\mathrm{d,e}} |\mathrm{z_i - z_{i,opt}}| = \max_{\mathrm{d}} \mathrm{e_{i,opt}(d)} + \max_{\mathrm{e}} |\mathrm{e_i}|$$

From (10.6), we have $L = \frac{1}{2}\|\widetilde{z}\|_2^2$ where $\widetilde{z} = J_{uu}^{1/2} G^{-1}(z - z_{\mathrm{opt}})$. Introducing the scaled outputs gives $\widetilde{z} = J_{uu}^{1/2} G'^{-1}(z' - z'_{\mathrm{opt}})$. With the assumed scaling, the individual scaled output deviations $z'_i - z'_{i,\mathrm{opt}}$ are less than 1 in magnitude. However, the variables z_i are generally correlated, so any combinations of deviations with magnitudes less than 1 may not possible. For example, the optimal values of both z_1 and z_2 may change in the same direction when there is a disturbance. Nevertheless, we will here assume that the expected output deviations are uncorrelated by making the following assumption:

(A1) The variations in $z'_i - z'_{i_{\mathrm{opt}}}$ are uncorrelated, or more precisely, the "worst-case" combination of output deviations $z'_i - z'_{i_{\mathrm{opt}}}$, with $\|z' - z'_{\mathrm{opt}}\|_2 = 1$, can occur in practice. Here $z' = S_1 z$ denotes the scaled outputs.

The reason for using the vector 2-norm, and not the max-norm, is mainly for mathematical comvenience. With assumption (A1) and (A.104), we then have from (10.6) that the maximum (worst-case) loss is

$$L_{max} = \max_{\|z' - z'_{\mathrm{opt}}\|_2 \le 1} \frac{\|\widetilde{z}\|_2}{2} = \frac{1}{2}\bar{\sigma}^2(J_{uu}^{1/2} G'^{-1}) = \frac{1}{2}\frac{1}{\underline{\sigma}^2(G' J_{uu}^{-1/2})} \tag{10.11}$$

where $G' = S_1 G$ and the last equality follows from (A.40). The result may be stated as follows

> **Maximum gain (minimum singular value) rule.** *Let G denote the steady-state gain matrix from inputs u (unconstrained degrees of freedom) to outputs z (candidate controlled variables). Scale the outputs using S_1 in (10.10) and assume that A1 holds. Then to minimize the steady-state loss select controlled variables z that maximize $\underline{\sigma}(S_1 G J_{uu}^{-1/2})$.*

The rule may stated as minimizing the scaled minimum singular value, $\underline{\sigma}(G')$, of the scaled gain matrix $G' = S_1 G S_2$, where the output scaling matrix S_1 has the inverse of the spans along its diagonal, whereas the input "scaling" is generally a full matrix, $S_2 = J_{uu}^{-1/2}$. This important result was first presented in the first edition of this book (Skogestad and Postlethwaite, 1996) and proven in more detail by Halvorsen et al. (2003).

Example 10.5 *The aero-engine application in Chapter 13 (page 500) provides a nice illustration of output selection. There the overall goal is to operate the engine optimally in terms of fuel consumption, while at the same time staying safely away from instability. The optimization layer is a look-up table, which gives the optimal parameters for the engine at various operating points. Since the engine at steady-state has three degrees of freedom we need to specify three variables to keep the engine approximately at the optimal point, and six alternative sets of three outputs are given in Table 13.3.2 (page 503). For the scaled variables, the value of $\underline{\sigma}(G'(0))$ is $0.060, 0.049, 0.056, 0.366, 0.409$ and 0.342 for the six alternative sets. Based on this, the first three sets are eliminated. The final choice is then based on other considerations including controllability.*

Remark 1 In the maximum gain rule, the objective function and the magnitudes of the disturbances and measurement noise enter indirectly through the scaling S_1 of the outputs z. To obtain $S_1 = \mathrm{diag}\{\frac{1}{\mathrm{span}(z_i)}\}$ we need to obtain for each candidate output $\mathrm{span}(z_i) = \max_d e_{i,\mathrm{opt}}(d) + \max|e_i|$. The

second contribution to the span is simply the expected measurement error, which is the measurement error plus the control error. The first contribition, $e_{i,\mathrm{opt}}$, may be obtained from a (nonlinear) model as follows: Compute the optimal values of the unconstrained z for the expected disturbances (with optimally constrained variables fixed). This yields a "look-up" table of z_{opt} for various expected disturbance combinations. From this data obtain for each candidate output, the expected variation in its optimal value, $e_{i_{\mathrm{opt}}} = (z_{i_{\mathrm{opt,max}}} - z_{i_{\mathrm{opt,min}}})/2$.

Remark 2 Our desire to have $\underline{\sigma}(G')$ large for output selection is *not* related to the desire to have $\underline{\sigma}(G)$ large to avoid input constraints as discussed in Section 6.9. In particular, the scalings, and thus the matrix G', are different for the two cases.

Remark 3 We have in our derivation assumed that the nominal operating point is optimal. However, it can be shown that the results are independent of the operating point, provided we are in the region where the cost can be approximated by a quadratic function as in (10.2) (Alstad, 2005). Thus, it is equally important to select the right controlled variables when we are nominally non-optimal.

Exercise 10.1 *Recall that the maximum gain rule requires that the minimum singular value of the (scaled) gain matrix be maximized. It is proposed that the loss can simply be minimized by selecting the controlled variables as $z = \beta y$, where β is a large number. Show that such a scaling does not affect the selection of controlled variables using the singular value method.*

10.3.4 Selecting controlled outputs: exact local method

The maximum gain rule is based on assumption A1 on page 395, which may not hold for some cases with more than one controlled variable ($n_z = n_u > 1$). This is pointed out by Halvorsen et al. (2003), who derived the following exact local method.

Let the diagonal matrix W_d contain the magnitudes of expected disturbances and the diagonal matrix W_e contain the expected implementation errors associated with the individual controlled variables. We assume that the combined disturbance and implementation error vector has norm 1, $\| \begin{bmatrix} d' \\ e' \end{bmatrix} \|_2 = 1$. Then, it may be shown that the worst-case loss is (Halvorsen et al., 2003)

$$\max_{\| \begin{bmatrix} d' \\ e' \end{bmatrix} \|_2 \le 1} L = \frac{1}{2} \bar{\sigma}([\, M_d \quad M_e \,])^2 \tag{10.12}$$

where

$$M_d = J_{uu}^{1/2} \left(J_{uu}^{-1} J_{ud} - G^{-1} G_d \right) W_d \tag{10.13}$$

$$M_e = J_{uu}^{1/2} G^{-1} W_e \tag{10.14}$$

Here $J_{uu} = \left(\partial^2 J / \partial u^2\right)_{\mathrm{opt}}$, $J_{ud} = \left(\partial^2 J / \partial u \partial d\right)_{\mathrm{opt}}$ and the scaling enters through the weights W_d and W_e.

10.3.5 Selecting controlled outputs: direct evaluation of cost

The local methods presented in Sections 10.3.2-10.3.4 are very useful. However, in many practical examples nonlinear effects are important. In particular, the local methods may not

be able to detect feasibility problems. For example, in marathon running, selecting a control strategy based on constant speed may be good locally (for small disturbances). However, if we encounter a steep hill (a large disturbance), then operation may not be feasible, because the selected reference value may be too high. In such cases, we may need to use a "brute force" direct evaluation of the loss and feasibility for alternative sets of controlled variables. This is done by solving the nonlinear equations, and evaluating the cost function J for various selected disturbances d and control errors e, assuming $z = r + e$ where r is kept constant (Skogestad, 2000). Here r is usually selected as the optimal value for the nominal disturbance, but this may not be the best choice and its value may also be found by optimization ("optimal back-off") (Govatsmark, 2003). The set of controlled outputs with smallest worst-case or average value of J is then preferred. This approach may be time consuming because the solution of the nonlinear equations must be repeated for each candidate set of controlled outputs.

10.3.6 Selecting controlled outputs: measurement combinations

We have so far selected z as a subset of the available measurements y. More generally, we may consider *combinations* of the measurements. We will restrict ourselves to *linear* combinations

$$z = Hy \tag{10.15}$$

where y now denotes all the available measurements, including the inputs u used by the control system. The objective is to find the measurement combination matrix H.

Optimal combination. Write the linear model in terms of the measurements y as $y = G^y u + G_d^y d$. Locally, the optimal linear combination is obtained by minimizing $\bar{\sigma}([\,M_d \quad M_e\,])$ in (10.12) with $W_e = HW_{n^y}$, where W_{n^y} contains the expected measurement errors associated with the individual measured variables; see Halvorsen et al. (2003). Note that H enters (10.12) indirectly, since $G = HG^y$ and $G_d = HG_d^y$ depend on H. However, (10.12) is a nonlinear function of H and numerical search-based methods need to be used.

Null space method. A simpler method for finding H is the *null space method* proposed by Alstad and Skogestad (2004), where we neglect the implementation error, i.e., $M_e = 0$ in (10.14). Then, a constant setpoint policy ($z = r$) is optimal if $z_{\rm opt}(d)$ is independent of d, that is, when $z_{\rm opt} = 0 \cdot d$ in terms of deviation variables. Note that the optimal values of the individual measurements $y_{\rm opt}$ still depend on d and we may write

$$y_{\rm opt} = Fd \tag{10.16}$$

where F denotes the *optimal* sensitivity of y with respect to d. We would like to find $z = Hy$ such that $z_{\rm opt} = Hy_{\rm opt} = HFd = 0 \cdot d$ for all d. To satisfy this, we must require

$$HF = 0 \tag{10.17}$$

or that H lies in the left null space of F. This is always possible, provided $n_y \geq n_u + n_d$. This is because the null space of F has dimension $n_y - n_d$ and to make $HF = 0$, we must require that $n_z = n_u < n_y - n_d$. It can be shown that when (10.17) holds, $M_d = 0$. If there are too many disturbances, i.e. $n_y < n_u + n_d$, then one should select only the important disturbances (in terms of economics) or combine disturbances with a similar effect on y (Alstad, 2005).

In the presence of implementation errors, even when (10.17) holds such that $M_d = 0$, the loss can be large due to non-zero M_e. Therefore, the null space method does not guarantee that the loss L using a combination of measurements will be less than using the individual measurements. One practical approach is to select first the candidate measurements y, whose sensitivity to the implementation error is small (Alstad, 2005).

10.3.7 Selecting controlled outputs: examples

The following example illustrates the simple "maximize scaled gain rule" (mimimum singular value method).

Example 10.6 Cooling cycle. *A simple cooling cycle or heat pump consists of a compressor (where work W_s is supplied and the pressure is increased to p_h), a high-pressure condenser (where heat is supplied to the surroundings at high temperature), an expansion valve (where the fluid is expanded to*

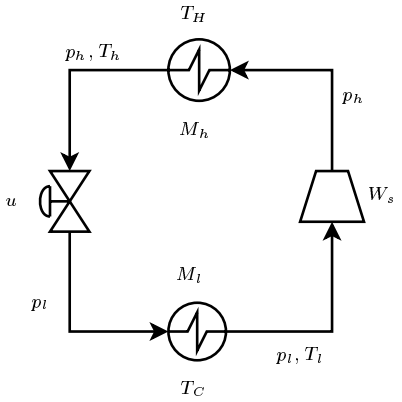

Figure 10.5: Cooling cycle

a lower pressure p_l such that the temperature drops) and a low-pressure evaporator (where heat is removed from the surroundings at low temperature); see Figure 10.5. The compressor work is indirectly set by the amount of heating or cooling, which is assumed given. We consider a design with a flooded evaporator where there is no super-heating. In this case, the expansion valve position (u) remains as an unconstrained degree of freedom, and should be adjusted to minimize the work supplied, $J = W_s$. The question is: what variable should we control?

Seven alternative controlled variables are considered in Table 10.1. The data is for an ammonia cooling cycle, and we consider Δy_{opt} for a small disturbance of 0.1 K in the hot surroundings ($d_1 = T_H$). We do not consider implementation errors. Details are given in Jensen and Skogestad (2005). From (10.9), it follows that it may be useful to compute the scaled gain $G' = G/\mathrm{span}(z(d_i))$ for the various disturbances d_i and look for controlled variables z with a large value of $|G'|$. From a physical point of view, two obvious candidate controlled variables are the high and low pressures (p_h and p_l). However, these appear to be poor choices with scaled gains $|G'|$ of 126 and 0, respectively. The zero gain is because we assume a given cooling duty $Q_C = UA(T_l - T_C)$ and further assume saturation $T_l = T^{\mathrm{sat}}(p_l)$. Keeping p_l constant is then infeasible when, for example, there are disturbances in T_C. Other obvious candidates are the temperatures at the exit of the heat exchangers, T_h and T_l. However,

Table 10.1: Local "maximum gain" analysis for selecting controlled variable for cooling cycle

Variable (y)	$\Delta z_{\text{opt}}(d_1)$	$G = \frac{\Delta z}{\Delta u}$	$\lvert G' \rvert = \frac{\lvert G \rvert}{\lvert \Delta z_{\text{opt}}(d_1) \rvert}$
Condenser pressure, p_h [Pa]	3689	-464566	126
Evaporator pressure, p_l [Pa]	-167	0	0
Temperature at condenser exit, T_h [K]	0.1027	316	3074
Degree of sub-cooling, $T_h - T^{\text{sat}}(p_h)$ [K]	-0.0165	331	20017
Choke valve opening, u	8.0×10^{-4}	1	1250
Liquid level in condenser, M_h [m^3]	6.7×10^{-6}	-1.06	157583
Liquid level in evaporator, M_l [m^3]	-1.0×10^{-5}	1.05	105087

the temperature T_l at the evaporator exit is directly related to p_l (because of saturation) and also has a zero gain. The open-loop policy with a constant valve position u has a scaled gain of 1250, and the temperature at the condenser exit (T_h) has a scaled gain of 3074. Even more promising is the degree of subcooling at the condenser exit with a scaled gain of 20017. Note that the loss decreases in proportion to $|G'|^2$, so the increase in the gain by a factor $20017/1250 = 16.0$ when we change from constant choke valve opening ("open-loop") to constant degree of subcooling, corresponds to a decrease in the loss (at least for small perturbations) by a factor $16.0^2 = 256$. Finally, the best single measurements seem to be the amount of liquid in the condenser and evaporator, M_h and M_l, with scaled gains of 157583 and 105087, respectively. Both these strategies are used in actual heat pump systems. A "brute force" evaluation of the cost for a (large) disturbance in the surrounding temperature ($d_1 = T_H$) of about 10 K, confirms the linear analysis, except that the choice $z = T_h$ turns out to be infeasible. The open-loop policy with constant valve position ($z = u$) increases the compressor work by about 10%, whereas the policy with a constant condenser level ($z = M_h$) has an increase of less than 0.003%. Similar results hold for a disturbance in the cold surroundings ($d_2 = T_C$). Note that the implementation error was not considered, so the actual losses will be larger.

The next simple example illustrates the use of different methods for selection of controlled variables.

Example 10.7 Selection of controlled variables. *As a simple example, consider a scalar unconstrained problem, with the cost function $J = (u-d)^2$, where nominally $d^* = 0$. For this problem we have three candidate measurements,*

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

We assume the disturbance and measurement noises are of unit magnitude, i.e. $|d| \leq 1$ and $|n_i^y| \leq 1$. For this problem, we always have $J_{\text{opt}}(d) = 0$ corresponding to

$$u_{\text{opt}}(d) = d, \quad y_{1,opt}(d) = 0, \quad y_{2,opt}(d) = 20d \quad \text{and} \quad y_{3,opt}(d) = 5d$$

For the nominal case with $d^* = 0$, we thus have $u_{\text{opt}}(d^*) = 0$ and $y_{\text{opt}}(d^*) = 0$ for all candidate controlled variables and at the nominal operating point we have $J_{uu} = 2, J_{ud} = -2$. The linearized models for the three measured variables are

$$\begin{array}{lll} y_1: & G_1^y = 0.1, & G_{d1}^y = -0.1 \\ y_2: & G_2^y = 20, & G_{d2}^y = 0 \\ y_3: & G_3^y = 10, & G_{d3}^y = -5 \end{array}$$

Let us first consider selecting one of the individual measurements as a controlled variable. We have

$$\begin{array}{lll} \text{Case } 1: & z = y_1, & G = G_1^y \\ \text{Case } 2: & z = y_2, & G = G_2^y \\ \text{Case } 3: & z = y_3, & G = G_3^y \end{array}$$

The losses for this example can be evaluated analytically, and we find for the three cases

$$L_1 = (10e_1)^2; \quad L_2 = (0.05e_2 - d)^2; \quad L_3 = (0.1e_3 - 0.5d)^2$$

(For example, with $z = y_3$, we have $u = (y_3 + 5d)/10$ and with $z = n_3^y$, we get $L_3 = (u - d)^2 = (0.1n_3^y + 0.5d - d)^2$.) With $|d| \leq 1$ and $|n_i^y| \leq 1$, the worst-case losses (with $|d| = 1$ and $|n_i^y| = 1$) are $L_1 = 100$, $L_2 = 1.052 = 1.1025$ and $L_3 = 0.62 = 0.36$, and we find that $z = y_3$ is the best overall choice for self-optimizing control and $z = y_1$ is the worst. We note that $z = y_1$ is perfectly self-optimizing with respect to disturbances, but has the highest loss. This highlights the importance of considering the implementation error when selecting controlled variables. Next, we compare the three different methods discussed earlier in this section.

A. *Maximum scaled gain (singular value rule): For the three choices of controlled variables we have without scaling $|G_1| = \underline{\sigma}(G_1) = 0.1$, $\underline{\sigma}(G_2) = 20$ and $\underline{\sigma}(G_3) = 10$. This indicates that z_2 is the best choice, but this is only correct with no disturbances. Let us now follow the singular value procedure.*

   1. *The input is scaled by the factor $1/\sqrt{(\partial^2 J/\partial u^2)_{\text{opt}}} = 1/\sqrt{2}$ such that a unit deviation in each input from its optimal value has the same effect on the cost function J.*
   2. *The maximum setpoint error due to variations in disturbances is given as $e_{\text{opt},i} = G_i^y J_{uu}^{-1} J_{ud} - G_{di}^y$. Then, for $z = y_1$, $e_{\text{opt},1} = 0.1 \cdot \frac{1}{2} \cdot (-2) - (-0.1) = 0$ and similarly, $e_{\text{opt},2} = -20$ and $e_{\text{opt},3} = 5$.*
   3. *For each candidate controlled variable the implementation error is $n^z = 1$.*
   4. *The expected variation ("span") for $z = y_1$ is $|e_{\text{opt},i}| + |n_1^y| = 0 + 1 = 1$. Similarly, for $z = y_2$ and $z = y_3$, the spans are $20 + 1 = 21$ and $5 + 1 = 6$, respectively.*
   5. *The scaled gain matrices and the worst-case losses are*

$$\begin{aligned} z = y_1: &\quad |G_1'| = \tfrac{1}{1} \cdot 0.1/\sqrt{2} = 0.071; &\quad L_1 &= \tfrac{1}{2|G'|^2} = 100 \\ z = y_2: &\quad |G_2'| = \tfrac{1}{21} \cdot 20/\sqrt{2} = 0.67; &\quad L_2 &= \tfrac{1}{2|G'|^2} = 1.1025 \\ z = y_3: &\quad |G_3'| = \tfrac{1}{6} \cdot 10/\sqrt{2} = 1.18; &\quad L_3 &= \tfrac{1}{2|G'|^2} = 0.360 \end{aligned}$$

We note from the computed losses that the singular value rule (= maximize scaled gain rule) suggests that we should control $z = y_3$, which is the same as found with the "exact" procedure. The losses are also identical.

B. *Exact local method: In this case, we have $W_d = 1$ and $W_{e_i} = 1$ and for y_1*

$$M_d = \sqrt{2}\left(2^{-1} \cdot (-2) - 0.1^{-1} \cdot (-0.1)\right) \cdot 1 = 0 \quad \text{and} \quad M_e = \sqrt{2} \cdot 0.1^{-1} \cdot 1 = 10\sqrt{2}$$

which give

$$L_1 = \frac{\bar{\sigma}([\, M_d \quad M_e \,])2}{2} = \frac{1}{2}(\bar{\sigma}(0 \quad 10\sqrt{2})) = 100$$

Similarly, we find with z_2 and z_3

$$L_2 = \frac{1}{2}(\bar{\sigma}(-\sqrt{2} \quad \sqrt{2}/20)) = 1.0025 \quad \text{and} \quad L_3 = \frac{1}{2}(\bar{\sigma}(-\sqrt{2}/2 \quad \sqrt{2}/10)) = 0.26$$

Thus, the exact local method also suggests selecting $z = y_3$ as the controlled variable. The reason for the slight difference from the "exact" nonlinear losses is that we assumed d and n^y individually to be less than 1 in the exact nonlinear method, whereas in the exact linear method we assumed that the combined *2-norm of d and n^y was less than* 1.

C. *Combinations of measurements: We now want to find the best combination $z = Hy$. In addition to y_1, y_2 and y_3, we also include the input u in the set y, i.e.*

$$y = [\, y_1 \quad y_2 \quad y_3 \quad u \,]^T$$

We assume that the implementation error for u is 1, i.e. $n^u = 1$. We then have $W_n^y = I$, where W_n^y is a 4×4 matrix. Furthermore, we have

$$G^y = [\,0.1 \quad 20 \quad 10 \quad 1\,]^T \quad G_d^y = [\,-0.1 \quad 0 \quad -5 \quad 0\,]^T$$

Optimal combination. We wish to find H such that $\bar{\sigma}([\,M_d \quad M_e\,])$ in (10.12) is minimized, where $G = HG^y$, $G_d = HG_d^y$, $W_e = HW_n^y$, $J_{uu} = 2$, $J_{ud} = -2$ and $W_d = 1$. Numerical optimization yields $H_{\text{opt}} = [\,0.0209 \quad -0.2330 \quad 0.9780 \quad -0.0116\,]$; that is, the optimal combination of the three measurements and the manipulated input u is

$$z = 0.0209y_1 - 0.23306y_2 + 0.9780y_3 - 0.0116u$$

We note, as expected, that the most important contribution to z comes from the variable y_3. The loss is $L = 0.0405$, so it is reduced by a factor 6 compared to the previous best case ($L = 0.26$) with $z = y_3$.

Null space method. In the null space method we find the optimal combination without implementation error. This first step is to find the optimal sensitivity with respect to the disturbances. Since $u_{\text{opt}} = d$, we have

$$\Delta y_{opt} = F\Delta d = G^y \Delta u_{\text{opt}} + G_d^y \Delta d = \underbrace{(G^y + G_d^y)}_{F} \Delta d$$

and thus the optimal sensitivity is

$$F = [\,0 \quad 20 \quad 5 \quad 1\,]^T$$

To have zero loss with respect to disturbances we need to combine at least $n_u + n_d = 1 + 1 = 2$ measurements. Since we have four candidate measurements, there are an infinite number of possible combinations, but for simplicity of the control system, we prefer to combine only two measurements. To reduce the effect of implementation errors, it is best to combine measurements y with a large gain, provided they contain different information about u and d. More precisely, we should maximize $\underline{\sigma}([\,G^y \quad G_d^y\,])$. From this we find that measurements 2 and 3 are the best, with $\underline{\sigma}([\,G^y \quad G_d^y\,]) = \underline{\sigma}\begin{bmatrix} 20 & 0 \\ 10 & -5 \end{bmatrix} = 4.45$. To find the optimal combination we use $HF = 0$ or

$$20h_2 + 5h_3 = 0$$

Setting $h_2 = 1$ gives $h_3 = -4$, and the optimal combination is $z = y_2 - 4y_3$ or (normalizing the 2-norm of H to 1):

$$z = -0.2425y_2 + 0.9701y_3$$

The resulting loss when including the implementation error is $L = 0.0425$. We recommend the use of this solution, because the loss is only marginally higher (0.0425 instead of 0.0405) than that obtained using the optimal combination of all four measurements.

Maximizing scaled gain for combined measurements. For the scalar case, the "maximize scaled gain rule" can also be used to find the best combination. Consider a linear combination of measurements 2 and 3, $z = h_2y_2 + h_3y_3$. The gain from u to z is $G = h_2G_2^y + h_3G_3^y$. The span for z, $\text{span}(z) = |e_{\text{opt,z}}| + |e_z|$, is obtained by combining the individual spans

$$e_{\text{opt,z}} = h_2 e_{\text{opt},2} + h_3 e_{\text{opt},3} = h_2 f_2 + h_3 f_3 = 20h_2 + 5h_3$$

and $|e_z| = h_2|e_2| + h_3|e_3|$. If we assume that the combined implementation errors are 2-norm bounded, $\| \begin{bmatrix} e_2 \\ e_3 \end{bmatrix} \|_2 \leq 1$, then the worst-case implementation error for z is $|e_z| = \| \begin{bmatrix} h_2 \\ h_3 \end{bmatrix} \|_2$. The resulting scaled gain that should be maximized in magnitude is

$$G' = \frac{G}{\text{span}} = \frac{h_2G_2^y + h_3G_3^y}{|h_2e_{\text{opt},2} + h_3e_{\text{opt},3}| + |e_z|} \tag{10.18}$$

The expression (10.18) gives considerable insight into the selection of a good measurement combination. We should select H (i.e. h_2 and h_3) in order to maximize $|G'|$. The null space method corresponds to selecting H such that $e_{\text{opt}} = h_2 e_{\text{opt},2} + h_3 e_{\text{opt},3} = 0$. This gives $h_2 = -0.2425$ and $h_3 = 0.9701$, and $|e_z| = \| \begin{bmatrix} h_2 \\ h_3 \end{bmatrix} \|_2 = 1$. The corresponding scaled gain is

$$G' = \frac{-20 \cdot 0.2425 + 10 \cdot 0.9701}{0 + 1} = -4.851$$

with a loss $L = \alpha/(2|G'|^2) = 0.0425$ (as found above). (The factor $\alpha = J_{uu} = 2$ is included because we did not scale the inputs when obtaining G'.)

Some additional examples can be found in Skogestad (2000), Halvorsen et al. (2003), Skogestad (2004b) and Govatsmark (2003).

Exercise 10.2* *Suppose that we want to minimize the LQG-type objective function, $J = x^2 + ru^2$, $r > 0$, where the steady-state model of the system is*

$$x + 2u - 3d = 0$$

$$y_1 = 2x, \quad y_2 = 6x - 5d, \quad y_3 = 3x - 2d$$

Which measurement would you select as a controlled variable for $r = 1$? How does your conclusion change with variation in r? Assume unit implementation error for all measurements.

Exercise 10.3 *In Exercise 10.2, how would your conclusions change when u (open-loop implementation policy) is also included as a candidate controlled variable? First, assume the implementation error for u is unity. Repeat the analysis, when the implementation error for u and each of the measurements is* 10.

10.3.8 Selection of controlled variables: summary

When the optimum coincides with constraints, optimal operation is achieved by controlling the active constraints. It is for the remaining unconstrained degrees of freedom that the selection of controlled variables is a difficult issue.

The most common "unconstrained case" is when there is only a single unconstrained degree of freedom. The rule is then to select a controlled variable such that the (scaled) gain is maximized.

Scalar rule: "maximize scaled gain $|G'|$"

- G = unscaled gain from u to z
- Scaled gain $G' = G/\text{span}$
- span = optimal range ($|e_{\text{opt}}|$) + implementation error ($|e|$)

In words, this "maximize scaled gain rule" may be expressed as follows:

Select controlled variables z with a large controllable range compared to their sum of optimal variation and implementation error. Here

- controllable range = range which may be reached by varying the inputs (as given by the steady-state gain)
- optimal variation: due to disturbance (at steady-state)

- implementation error = sum of control error and measurement error (at steady-state)

For cases with more than one unconstrained degree of freedom, we use the gain in the most difficult direction as expressed by the minimum singular value.

> **General rule:** *"maximize the (scaled) minimum singular value $\underline{\sigma}(G')$ (at steady-state)"*

We have written "at steady-state" because the cost usually depends on the steady-state, but more generally it could be replaced by "at the bandwidth frequency of the layer above (which adjusts the setpoints for z)".

10.4 Regulatory control layer

In this section, we are concerned with the regulatory control layer. This is at the bottom of the control hierarchy and the objective of this layer is generally to "stabilize" the process and facilitate smooth operation. It is *not* to optimize objectives related to profit, which is done at higher layers. Usually, this is a decentralized control system of "low complexity" which keeps a set of measurements at given setpoints. The regulatory control layer is usually itself hierarchical, consisting of cascaded loops. If there are "truly" unstable modes (RHP-poles) then these are usually stabilized first. Then, we close loops to "stabilize" the system in the more general sense of keeping the states within acceptable bounds (avoiding drift), for which the key issue is local disturbance rejection.

The most important issues for regulatory control are what to measure and what to manipulate. Some simple rules for these are given on page 405. A fundamental issue is whether the introduction of a separate regulatory control layer imposes an inherent performance loss in terms of control of the primary variables z. Interestingly, the answer is "no" provided the regulatory controller does not contain RHP-zeros, and provided the layer above has full access to changing the reference values in the regulatory control layer (see Theorem 10.2 on page 415).

10.4.1 Objectives of regulatory control

Some more specific objectives of the regulatory control layer may be:

O1. Provide sufficient quality of control to enable a trained operator to keep the plant running safely without use of the higher layers in the control system.

This sharply reduces the need for providing costly backup systems for the higher layers of the control hierarchy in case of failures.

O2. Allow for simple decentralized (local) controllers (in the regulatory layer) that can be tuned on-line.

O3. Take care of "fast" control, such that acceptable control is achievable using "slow" control in the layer above.

O4. Track references (setpoints) set by the higher layers in the control hierarchy.