Control structure selection and plantwide control[1]

Sigurd Skogestad
Department of Chemical Engineering
Norwegian University of Sciience at Technolog (NTNU)
Trondheim, Norway

1. Introduction

Consider the generalized control design problem in Figure 1. Here, the objective is to design the controller K, which, based on the measurements y, computes the inputs (MVs) u such that the controlled outputs (CVs) z are kept close to their desired setpoints, in spite of unknown disturbances, varying setpoints and measurement noise(w). However, note that it assumed that we know what to measure (y), manipulate (u), and, most importantly, what variables we would like to control (z), that is, we assume a given control structure.

The term ”constrol structure selection” (CSS) and its synonym ”control structure deign” (CSD) is associated with the overall *control philosophy* for the system with emphasis on the ***structural decisions****:*

1. *Selection of controlled variables (CVs, "outputs", z in Figure 1)*
2. *Selection of manipulated variables (MVs, "inputs", u in Figure 1)*
3. *Selection of (extra) measurements (v in Figure 1)*
4. *Selection of control* ***configuration*** (structure of overall controller that interconnects the controlled, manipulated and measured variables; structure of K in Figure 1)
5. *Selection of type of controller K* (PID, MPC, LQG, H-infinity, etc.) and objective function (norm) used to design and analyze it.

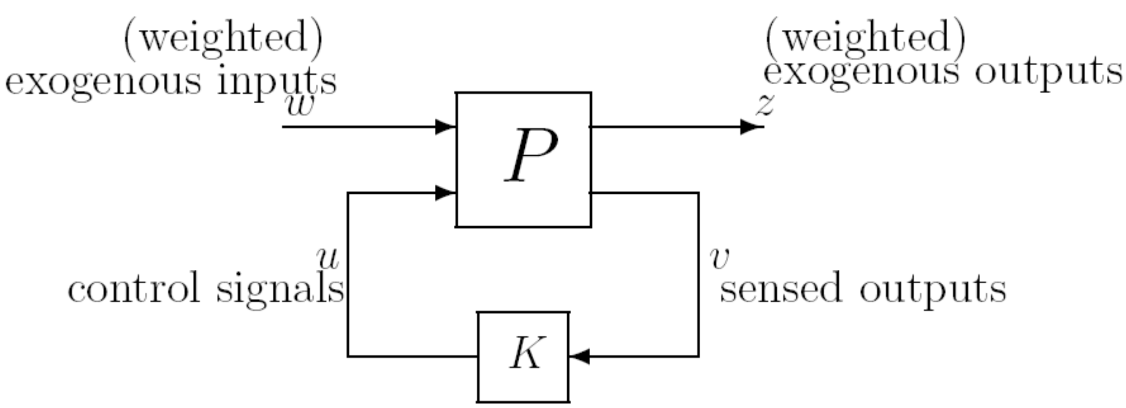

Figure 1. General control problem.

[1] Draft chapter for Springer Encyclopedia on Control Systems (15 July 2013)

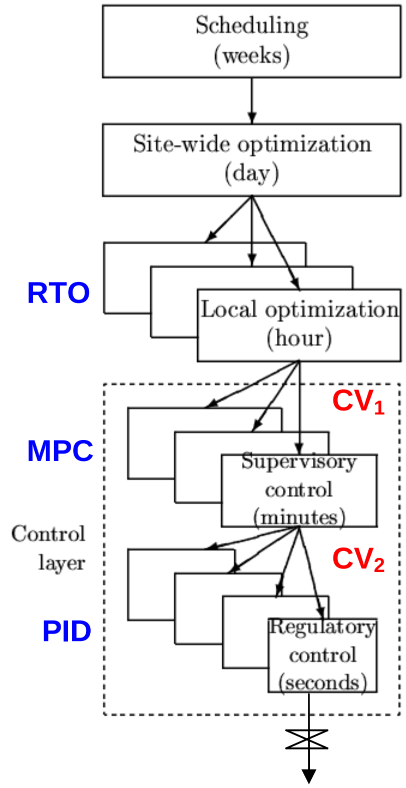

Figure 2: *Typical control hierarchy for a chemical plant (plantwide control)*

Decisions 2 and 3 (selection of u and y) are sometimes referred to as the input/output-selection (IO) problem. In practice, the control system (K) is usually divided into several layers, separated by time scale (see Figure 2).

Control structure selection includes all the *structural* decisions that the engineer needs to make when designing a control system, but it does not involve the actual design of each individual controller block. Thus, it involves the decisions necessary to make a block diagram (Figure 1; used by control engineers) or process & instrumentation diagram (used by process engineers) for the entire plant, and provides the starting point for a detailed controller design

The term "plantwide control", which is a synonym for "control structure selection", is used in the field of process control. Controls structure selection is particularly important for process control because of the complexity of large processing plants, but it applies to all control applications, including vehicle control, aircraft control, robotics, power systems, biological systems, social systems and so on.

It may be argued that control structure selection is more important than the controller design itself, but still control structure selection is hardly covered in most control courses. This is probably related to the complexity of the problem, which requires the input from several domains of knowledge. In any mathematical sense, the control structure selection problem is a formidable combinatorial problem which involves a large number of discrete decision variables, and this is probably why the progress in the area has been relatively slow. In addition, the problem has been poorly defined in terms of its objective.

2. Overall objectives for control and structure of the control layer

The starting point for control system design is to define clearly the operational objectives. There are two main objectives for control:

- Stability and shorter-term regulatory control

- Longer-term optimal operation (minimize economic cost J subject to satisfying operational constraints)

The first objective is related to "making sure the system runs (operates)", where stability and robustness are important issues, and is usually the main domain of control engineers. The second objective is related to "making the system operate as intended", where economics are an important issue. An example is bicycle riding; we first need to learn how to stabilize the bicycle (regulation), before trying to use for something useful (optimal operation), like riding to work. We use the term "economic" cost, because usually the cost function J can be given a monetary value, but more generally, the cost J could be any scalar cost. For example, the cost J could be the "environmental impact" and the economics could be given as a constraint.

In theory, the optimal strategy is to combine the control tasks of regulation and optimal economic operation in a single centralized controller K, which at each time step collects all the information and computes the optimal input changes. In practice, simpler controllers are used. The main reason for this is that in most cases one can obtain acceptable control performance with simple structures, where each controller block only involves a few variables, and such control systems can be designed and tuned with much less effort, especially when it comes to the modelling and tuning effort.

So how are systems controlled in practise? The main simplification is to decompose the overall control problem into many simpler control problems, using two two main principles

- *Decentralized (local) control.* This "horizontal decomposition" of the control layer is mainly based on separation in space, for example, by using local control of individual process units.
- *Hierarchical control.* This "vertical decomposition" is mainly based on time scale separation, as illustrated for a process plant in Figure 2. The upper three layers in Figure 2 deal explicitly with economic optimization and are not considered here. We are concerned with the two lower *control layers*, where the main objective is to track the setpoints specified by the layer above.

As shown in Figure 2, the control layer is usually divided in two parts; a faster regulatory (stabilization) layer and a slower supervisory (economic) layer . The main justification for separating into two layers is that the two tasks of regulation and economically optimal operation are fundamentally different, and that the benefit of combining them is usually limited. Only if there is a reasonable benefit in combining the two layers, for example, because there is limited time scale separation between the tasks of regulation and optimal economics, should one consider combining them into a single controller.

3. Matrices H and H2 for controlled variable selection

The most important notation is summarized in Table 1. To distinguish between the two control layers, we use "1" for the upper supervisory layer and "2" to denote the regulatory layer, which is "secondary" in terms of its place in the control hierarchy.

Table 1. Important notation (see also Figure 3)

u = [u1; u2] = set of all available physical inputs (degrees of freedom)
u_1 = inputs used directly by supervisory control layer
u_2 = inputs used by regulatory layer
y_m = set of all candidate measured variables
$y = [y_m; u]$ = combined set of measurements and inputs
y_2 = outputs in regulatory layer (subset or combination of measurements y_m); $\dim(y_2)=\dim(u_2)$
$CV_1 = H\, y$ = controlled variables in supervisory layer; $\dim(CV_1)=\dim(u)$

$CV_2 = [y_2; u_1] = H_2$ y = controlled variables in regulatory layer; dim (CV_2)=dim(u)
$MV_1 = [y_{2s}; u_{1s}] = CV_{2s}$ = manipulated variables in supervisory layer; dim(MV_1)=dim(u)
$MV_2 = u_2$ = manipulated variables in regulatory layer; dim (MV_2)≤dim(u)

There is usually limited flexibility with respect to the set of all available inputs (u) as it is usually given by the system design. However, there may be a possibility to add inputs (e.g. add an extra valve) or to move it to another location, for example, to reduce the time delay and thus improve the input-output controllability.

However, there is much more flexibility in terms of output selection, and the most important structural decision is related to the selection of controlled variables in the two control layers, as given by the decision matrices H and H_2 (see Figure 3).

$CV_1 = H\ y$

$CV_2 = H_2\ y$

Note from the definition that y includes, in addition to the candidate measured outputs (y_m), also the physical inputs u. This allows for the possibility of selecting an input u as a ”controlled” variable, which means that this input is kept constant (and left ”unused” for control).

In general, H and H_2 are ”full” matrices, allowing for measurement combinations as controlled variables. However, for simplicity, especially in the regulatory layer, we often pefer to control individual measurements, that is, H_2 is often a “selection matrix”, where each row in H_2 contains one 1-element (to identify the selected variable) and the rest 0's.

To have a simple control structure, with as few regulatory loops as possible, it is desirable that H_2 contains many 1's in the right part of the matrix, meaning that the corresponding input u is left unused for regulatory control. As an example, assume there are 3 candidate output measurements (temperatures T) and 2 inputs (flowrates q),

$y_m' = [T_1\ T_2\ T_3]$, $u' = [q_1\ \ q_2]$

Then the choice

$H_2 = [0\ 1\ 0\ 0\ 0;\ 0\ 0\ 0\ 0\ 1]$

means that we have $CV_2 = H_2\ y = [T_2; q_2]$. From the definition in Table 1, $CV_2 = [y_2; u_1]$, so we have $y_2=T_2$ and $u_1=q_2$, and the latter implies that we in the regulatory layer use $u_2=q_1$ (to control $y_2=T_2$) and leave the “unused” input $u_1=q_2$ for the supervisory control layer. If we instead select

H2 = [1 0 0 0 0; 0 0 1 0 0]

then we have $CV_2= [T_1; T_3]$. None of these are inputs, so u_1 is an empty set in this case. This means that we close two regulatory loops, using $u_2 = [q_1; q_2]$ to control $y_2 = [T_1; T_3]$. The degrees of freedom for the supervisory layer will then be the two temperature setpoints, $MV_1= CV_{2s}$.

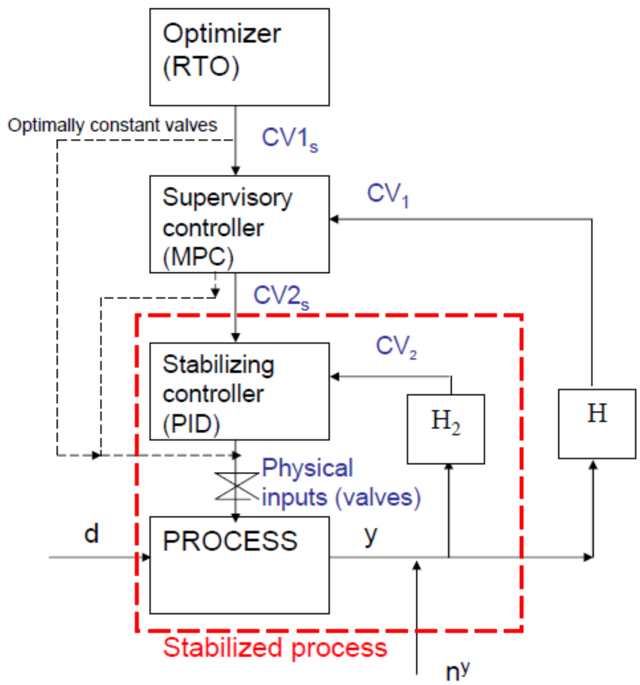

Figure 3: *Block diagram of process control hierarchy illustrating the selection of controlled variables (H and H_2) for optimal economic operation (CV_1) and stabilization (c_2=CV_2).*

4. Supervisor control layer and selection of economic controlled variables (CV_1)

Table 2. Objectives of supervisory control

O1. Control primary "economic" variables CV_1 at setpoint using as degrees of freedom CV_{2s}, which includes the setpoints to the regulatory layer (y_{2s}) as well as any "unused" degrees of freedom (u_1).

O2. Switch controlled variables (CV_1) depending on operating region, for example, because of change in active constraints.

O3. Supervise the regulatory layer, for example, to avoid input saturation (u_2), which may destabilize the system

O4. Coordinate control loops (multivariable control) and reduce effect of interactions (decoupling)

O5. Provide feedforward action from measured disturbances

O6. Make use of extra inputs, for example, to improve the dynamic performance (valve position control) or to extend the operating range (split range control)
O7. Make use of extra measurements, for example, to estimate the primary variables CV_1.

Some objectives for the supervisory control layer are given in Table 2. The main structural issue for the supervisory control layer, and probably the most important decision in the design of any control system, is the selection of the primary (economic) controlled variable CV_1. In many case, a good engineer can make a reasonable choice based on process insight and experience. However, the control engineer must realize that this a critical decision that someone has to make. The main rules and issues for selecting CV_1 are

CV_1-Rule 1. Control active constraints (almost always)

- Active constraints may often be identified by engineering insight. For example, consider the problem of minimizing the driving time between to cities (cost J=T). There is a single input (power P) and the optimal solution is often constrained. When driving a fast car, the active constrainit may be the speed limit (CV1 = v with setpoint v_{max}). When driving an old car it may be the maximum power (CV1 = P with setpoint P_{max}). The latter corresponds to an input constraint (u) which is trivial to implement; the former corresponds to an output constraint ($y_{max} = v_{max}$) which requires a controller ("cruise control").

- For *"hard" output constraints*, which cannot be violated at any time, we need to introduce a *backoff* the guarantee feasibility. The backoff is defined as the different between the optimal value and actual setpoint, for example, we need to back off from the speed limit because of the possibility for measurement error and imperfect control

$$y_s = y_{max} - \text{backoff}$$

CV_1-Rule 2. For the remaining unconstrained degrees of freedom, look for "self-optimizing" variables which when held constant, indirectly lead to close-to-optimal operation, in spite of disturbances.

- Self-optimizing variables (CV_1=Hy) are variables which when kept constant, indirectly (through the action of the feedback control system) lead to close-to optimal adjustment of the inputs when there are disturbances.

- An ideal self-optimizing variable is the gradient of the cost function with respect to the unconstrained input. $CV_1 = dJ/du = J_u$

- More generally, since we can directly measure the gradient, we select CV_1 = Hy. The selection of a good H is a non-trivial task. For example, consider again the problem of minimizing the driving timeT, but assume this time that we only have a limited amount of fuel, and that driving at maximum power or maximum speed will use too much fuel. This is an unconstrained optimization problem, and identifying a good CV1 is not obvious. One possibility could be to keep a constant speed (CV_1= v), but the optimal value of v will vary depending on the slope of the road. A better option, could be to keep a constant fuel flow (CV_1 = q). More generally, one can control combinations, CV_1 = Hy where H is a "full" matrix.

CV_1-Rule 3. For the unconstrained degrees of freedom, one should *never* control a variable that reaches its unconstrained maximum or minimum value at the optimum, for example, never try to control directly the cost J.

Violation of this rule gives either infeasiblity (if attempting to control J at a lower value than J_{min}) or non-uniqueness (if attempting to control J at higher value than J_{min})

For CV1-Rule 2, at least locally, it is always possible to good variable combinations (i.e., H is a "full" matrix), but whether or not it is possible to find good individual variables (H is a selection matrix), is not obvious. To help identify potential "self-optimizing" variables, the following requirements may be used:

Requirement 1. The *optimal* value of c is insensitive to disturbances, that is, $dc_{opt}/dd = HF$ is small.

Requirement 2. The variable c t is easy to measure and control accurately

Requirement 3. The value of c is sensitive to changes in the manipulated variable, u; that is, the gain, $G=HG^y$, from u to c is large (so that even a large error in controlled variable, c, results in only a small variation in u.) Equivalently, the optimum should be 'flat' with respect to the variable, c.

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

All four requirements should be satisfied. For example, for the operation of a marathon runner, the heart rate may be a good "self-optimizing" controlled variable c (to keep at constant setpoint). Let us check this against the four requirements. The optimal heart rate is weakly dependent on the disturbances (requirement 1) and the heart rate is easy to measure (requirement 2). The heart rate is quite sensitive to changes in power input (requirement 3). Requirement 4 does not apply since this is a problem with only one unconstrained input (the power). In summary, the heart rate is a good candidate.

Regions and switching. Note that new controlled variables (CV1) must be identified (offline) for each active constraint region, and on-line switching is required for optimality. In practise, it is easy to identify when to switch when one encounters a constraint. It seems less obvious when to switch out of a constraint, but actually one simply has to monitor the value of the unconstrained CVs from the neighbouring regions and switch out of the constraint region when the unconstrained CV reaches its setpoint.

In general, one would like to simplify the control structure and reduce need for switching. This may require using a suboptimal CV1 in some regions of active constraints. In this case the setpoint for CV1 may not be its nominally optimal value (which is the normal choice), but rather a "robust setpoint" (with backoff) which reduces the loss when we are outside the nominal constraint region.

Structure of supervisory layer. The supervisory layer ("advanced control") may either be centralized (e.g. using model predictive control) or decomposed into simpler subcontrollers using standard elements, like decentralized control (PID), cascade control, selectors, decouplers, feedforward elements, ratio control, split range control, valve position control (more generally, known as habituating control, midrange control or input resetting), and so on. In theory, the performance is better with the centralized approach (e.g. MPC), but the difference is usually small if designed by a good engineer. The main reason for using simpler elements, is that the system can be implemented in the existing "basic" control system, that it can be tuned with little model information and that it can be build up gradually. However, such systems can become complicated and difficult to understand for other that the engineer who designed it. Therefore, model-based centralized solutions (MPC) are often preferred because the design is more systematic and easier to modify.

5. Quantitative approach for selecting economic controlled variables, CV1

A quantitative approach is to consider the effect of the choice CV_1=Hy on the economic cost J when disturbances d occur. One should also include noise and error (n^y) related to the easurements and inputs (y).

Step S1. *Define operational objectives (economic cost function J and constraints)*
We first quantify the operational objectives in terms of a scalar cost function J [$/s] that should be minimized (or equivalently, a scalar profit function, P = -J, that should be maximized). For process control applications, this is usually easy, and typically we have

J = cost feed + cost utilities (energy) – value products [$/s]

Note that the cost function J is used to *select* the controlled variables (CV1), and another cost function, typically involving the deviation in CV_1 from their optimal setpoints CV_{1s}, is used for the actual controller design (MPC).

Step S2. *Find optimal operation for expected disturbances*

Mathematically, the optimization problem can be formulated as

$$\min_u J(u,x,d)$$

subject to:

Model equations: dx/dt=f(u,x,d)

Operational constraints: g(u,x,d) ≤ 0

This problem should be solved for expected disturbances (d) to find the truly optimal operation policy, uopt(d). The nominal solution may be used to obtain setpoints (CV_{1s}) for the selected controlled variables.

In practise, the optimum input $u_{opt}(d)$ can not be realized, because of model error and unknown disturbances, and we look for simple control implementations where u is adjusted to keep CV1 at the nominally optimal setpoints. In many cases, the economics are determined by the steady-state behaviour and we can set dx/dt=0.

Together with obtaining the model, the optimization step S2 is often the most time consuming step in the entire plantwide control procedure.

Step S3. ***Select "economic" (primary) controlled variables, CV1***

5.1 CV1-rule 1. *Control active constraints*

A major objective of the optimization is to find the expected regions of active constraints, and a constraint is said to be "active" if g=0 at the optimum.. The optimally active constraints will vary depending on disturbances and market conditions (prices).

5.2 CV1-rule 2. *Control self-optimizing variables*

After having identified (and controlled) the active constraints, one should consider the remaining lower-dimension unconstrained optimization problem, and for the remaining unconstrained degrees of freedom one should search for c*ontrol "self-optimizing" variable*s c.

1. "**Brute force" approach.** Given a set of controlled variables CV1=c=Hy, one computes the cost J(c,d) when we keep c constant ($c = c_s + Hn^y$) for various disturbances (d) and measurement errors (n^y). In practise, this is done by running a large number of steady-state simulations to try to cover the expected future operation.

2."Local" approaches based on a quadratic approximation of the cost J. Linear models are used for the effect of u and d on y.

$y = G^y u + G^y_d d$ (xa)

This is discussed in more detail in Alstad et al. (2009) and references therein. The main local approaches are:

2A. **Maximum gain rule; maximize the minimum singular value of $G=HG^y$.** In words, the maximum gain rule, which essentially is a quantitative version of Requirements 1 and 3 given above, says that one should control "sensitive" variables, with a large scaled gain G from the inputs (u) to c=Hy. This rule is good for pre-screening and also yields good insight.

2B. **Nullspace method.** This method yields optimal measurement combinations for the case with no noise, $n^y=0$. One must first obtain the optimal measurement sensitivity

$$\mathbf{F} = dy^{opt}/dd.$$

Each column in **F** expresses the optimal change in the y's when the independent variable (u) is adjusted so that the system remains optimal with respect to the disturbance d. Usually, it is simplest to obtain F numerically by optimizing the model. Alternatively, we can obtain F from a quadratic approximation of the cost function

$$F = G^y_d - G^y J_{uu}^{-1} J_{ud}$$

Then, assuming that we have at least as many (independent) measurements y as the sum of the number of (independent) inputs (u) and disturbances (d), the optimal is to select c=Hy such that

$$H\mathbf{F}=0$$

Note that H is a nonsquare matrix, so HF=0 does not require that H=0 (which is a trivial uninteresting solution), but rather that H is in the nullspace of F^T.

2C. **Exact local** method (loss method). This extends the nullspace method to include noise (n^y) and allows for any number of measurements. The noise and disturbances are normalized by introducing weighting matrices W_{ny} and W_d (which have the expected magnitudes along the diagonal) and then the expected loss, $L = J - J_{opt}(d)$, is minimized by selecting H to solve the following problem

Min_H $||M(H)||_2$

(Frobenius norm) where $M(H) = J^{1/2}_{uu} (HG^y)^{-1} HY$ and $Y = [FW_d \;\; W_{ny}]$. Note here that the optimal choice with Wny=0 (no noise) is to choose H such that HF=0, which is the nullspace method. For the general case, when H is a "full" matrix, this is a convex problem and the optimal solution is $H' = (YY')^{-1} G^y Q$ where Q is any non-singular matrix.

6. Regulatory control layer

The main purpose of the regulatory layer is to "stabilize" the plant, preferably using a *simple* control structure (e.g. single-loop PID controllers) which does not require changes during operation. "Stabilize" is here used in a more extended meaning to mean that the process does not "drift" too far away from acceptable operation when there are disturbances. The regulatory layer should make it possible to use a "slow" supervisory control layer that does not require a detailed model of the high-frequency dynamics. Therefore, in addition to track the setpoints given by the "supervisory layer" (e.g., MPC), the regulatory layer may directly control variables (CV_1) that require fast and tight control, like economically important active constraints.

In general, the design of the regulatory layer involves the following structural decisions:

1. Selection of outputs y_2 to control (among all candidate measurements y_m).

2. Selection of inputs u_2 (a subset of all available inputs u) to control the outputs y_2.

3. Pairing of inputs and outputs (since decentralized control is normally used).

Note that we do not "use up" any degrees of freedom in the regulatory layer because the set points (y_{2s}) are left as manipulated variables (MV_1) for the supervisory layer (see Figures 2 and 3). Furthermore, since the set points are set by the supervisory layer in a cascade manner , the system approaches on a long time scale the same steady-state (as defined by the choice of economic variables CV_1) irrespective of the choice of controlled variables in the regulatory layer.

The inputs for the regulatory layer (u_2) are selected as a subset of all the available inputs (u). For stability reasons, one should avoid input saturation in the regulatory layer. In particular, one should avoid using inputs (in the set u_2) that are optimally constrained in some disturbance region. Otherwise, in order to avoid input saturation, one needs to include backoff for the input when one enters this operation region, which will have an economic penalty.

There is usually much more flexibility in terms of output selection (y_2). In the regulatory layer, the outputs (y_2) are usually individual measurements and they are often not important variables in themselves. Rather, they are "extra measurements" that are controlled in order to "stabilize" the system, and their setpoints (y_{2s}) are changed by the layer above, in a cascade manner. For example, in a distillation column one may control a temperature somewhere in the middle of the column (y_2=T). This is to "stabilize" the column profile, and its setpoint (y_{2s}=T_s) is set by the layer above in order to obtain a desired product composition (y_1=c).

7. Input-output selection for regulatory control

Interestingly, the selection of inputs and output for regulatory control (decisions 1 and 2) may be combined into a single decision, by considering the selection of CV2 = [y2; u1]. This follows because we want to use all inputs u for control, so assuming that that the set u is given, "selection of inputs u_2" (decision 2) is by elimination equivalent to "selection of inputs u_1". Note that CV_2 include all variables that we keep at desired (constant) values within the fast time horizon of the regulatory control layer, including the "unused" input u_1

7.1 Survey by Van de wal and Jager

Van de Wal and Jager provide an overview of methods for input-output selection, some of which include:

1. "Accessability" based on guaranteeing a cause-effect relationship between the selected inputs (u_2) and outputs y_2). Use of such measures may eliminate unworkable control structures.

2. "State controllability and state observability" to ensure that any unstable modes can be stabilized using the selected inputs and outputs.

3. "Input-output controllability" analysis to ensure that y_2 can be acceptably controlled using u_2. This is based on scaling the system, and then analysing the transfer matrices $G_2(s)$ (from u_2 to y_2) and G_{d2} (from expected disturbances d to y_2). Some important controllability measures are right half plane zeros (unstable dynamics of the inverse), condition number, singular values, relative gain array, etc. One problem here is that there are many different measures, and it is not clear which should be given most emphasis.

4. "Achievable robust performance". This may be viewed as a more detailed version of input-output controllability, where several relevant issues are combined into a single measure. However, this requires that the control problem can actually be formulated clearly, which may be very difficult, as already mentioned. In addition, it requires finding the optimal robust controller for the given problem, which may be very difficult.

Most of these methods are useful for analysing a given structure (u_2, y_2) and less suitable for selection. Also, the list of methods is also incomplete, as disturbance rejection, which is probably the most important issue for the regulatory layer, is hardly considered.

7.2 A systematic approach for IO-selection based on minimizing state drift caused by disturbances

The objectives of the regulatory control layer are many, and Yelchuru and Skogestad list 13 partly conflicting objectives. To have a truly systematic approach to regulatory control design, which includes also the selection of inputs and outputs, we would need to quantify all these partially conflicting objectives in terms of a scalar cost function J_2. We here consider one simple measure, namely the state drift

$J_2 = ||Wx||$,

Note that this cost will be used to *select* controlled variables (CV_2), and not to design the controllers. The justification for considering the state drift, is that the regulatory layer should ensure that the system, as measured by the weighted states Wx, does not drift too far away from the desired state and thus stays in the "linear region" when there are disturbances.

As noted above, the two decisions related to input-output selection can be combined into a single decision by introducing the "combined" controlled variables in the regulatory layer,

$CV_2= [y_2; u_1]$

where u_1 denotes the inputs that are <u>not</u> used by the regulatory control layer.. The input-output selection problem for the regulatory layer is then to select the matrix H_2,

$CV_2 = H_2 y$

where $y = [y_m; u]$ and y_m contains all candidate measurements. The cause for changes in x are disturbances d. To this effect, consider the linear model (in deviation variables)

$y= G^y u + G^y_d d$ (xa)

$x= G^x u+ G^x_d d$ (xb)

where the G-matrices are transfer matrices. Here, G^x_d gives the effect of the disturbances on the states with no control, and the idea is that this effect should be reduced by closing the regulatory control loops. Within the "slow" time scale of the supervisory layer, we can assume that CV2 is perfectly controlled and thus constant, or CV2 = 0 in terms of deviation variables. This gives

$CV_2 = H_2 G^y u + H_2 G^y_d d = 0$

and solving with respect to u gives

$u = -(H_2G^y)^{-1}(H_2G^y_d)d$

and we have

$x = P^x_d(H_2) d$

where

$P^x_d = G^x_d - G^x(H_2G^y)^{-1}H_2G^y_d$ (xx)

is the disturbance effect for the "partially" controlled system with only the regulatory loops closed. Note that it is not generally possible to make $P^x_d = 0$ because we have more states that we have available inputs.To have a small "state drift", we want $J_2 = ||W P_d d||$ to be small, and to have a simple regulatory control system we want to close as few regulatory loops as possible. Assume that we have normalized the disturbances so that the norm of d is 1, then we can solve the following problem

For 0, 1, 2… etc. loops closed solve: min_H_2 $|| M_2(H_2)||$ (x)

where $M_2 = W P^x_d$ and

dim(u2)=dim(y2)=no. of loops closed

By comparing the value of $|| M_2(H_2)||$ with different number of loops closed (i.e., with different H_2), we can then decide on an appropriate regulatory layer structure. For example, assume that we find that the value of J_2 is 110 (0 loops closed), 0.2 (1 loop), and 0.02 (2 loops), and assume we have scaled the disturbances and states such that a value less than about 1 is acceptable, then closing 1 regulatory loop is probably the best choice.

In principle, this is straightforward, but there are three remaining *issues*: 1) We need to choose an appropriate norm, 2) we should include measurement noise to avoid selecting insensitive measurements and 3) the problem must be solvable numerically.

Issue 1. The norm of M_2 should be evaluated in the frequency range between the "slow" bandwidth of the supervisory control layer (ω_{B1}) and the "fast" bandwidth of the regulatory control layer (ω_{B2}). However, since it is likely that the system sometimes operates without the supervisory layer, it is reasonable to evaluate the norm of P^x_d in the frequency range from 0 (steady state) to ω_{B2}. Since we want H_2 to be a constant (not frequency-dependent) matrix, it is reasonable to choose H_2 to minimize the norm of M_2 at the frequency where ||M2|| is expected to have its peak. For some mechanical systems, this may be at some resonance frequency, but for process control applications it is usually at steady state (ω=0), that is , we can use the steady-state gain matrices in (xx). In terms of the norm, we use the 2-norm (Frobenius norm), mainly because it has good numerical properties, and also because it has the interpretation of giving the expected variance in x for normally distributed disturbances.

Issues 2 and 3. If we include also measurement noise n^y, then expected value of J_2 is minimized by solving the problem min_H_2 $||M_2(H_2)||_2$ where

$$\mathbf{M}_2(\mathbf{H}_2) = \mathbf{J}_{2uu}^{1/2}(\mathbf{H}_2\mathbf{G}^y)^{-1}\mathbf{H}_2\mathbf{Y}_2 \quad (11)$$

$$\mathbf{Y}_2 = [\mathbf{F}_2\mathbf{W}_d \quad \mathbf{W}_n]; \quad \mathbf{F}_2 = \frac{\partial \mathbf{y}_{opt}}{\partial \mathbf{d}} = \mathbf{G}^y J_{2uu}^{-1} J_{2ud} - \mathbf{G}^y_d \quad (12)$$

$$\text{where} \quad \mathbf{J}_{2uu} \triangleq \frac{\partial^2 J_2}{\partial \mathbf{u}^2} = 2\mathbf{G}^{x^T}\mathbf{W}^T\mathbf{W}\mathbf{G}^x, \quad \mathbf{J}_{2ud} \triangleq \frac{\partial^2 J_2}{\partial \mathbf{u}\partial \mathbf{d}} = 2\mathbf{G}^{x^T}\mathbf{W}^T\mathbf{W}\mathbf{G}^x_d,$$

Note that this is the same mathematical problem as presented for selecting CV_1=Hy, but the objective function J_2 is different from J . Again, W_d and W_n are diagonal matrices, expressing the expected magnitude of the disturbances and noise (for y). For the case when H_2 is a "full" matrix, this can be reformulated as a convex optimization problem and an explicit solution is

$$\mathbf{H}_2{}^T = (\mathbf{Y}_2\mathbf{Y}_2{}^T)^{-1}\mathbf{G}^y(\mathbf{G}^{y^T}(\mathbf{Y}_2\mathbf{Y}_2{}^T)^{-1}\mathbf{G}^y)^{-1}\mathbf{J}_{2_{uu}}^{1/2}$$

and from this we can find the optimal value of J_2. It may seem restrictive to assume that H_2 is a "full" matrix, because we usually want to control individual measurements, and then H_2 should be a selection matrix, with 1's and 0's. Fortunately, since we in this case want to control as many measurements (y_2) as inputs (u_2), we have that H_2 is square in the selected set, and the optimal value of J_2 when H_2 is a selection matrix is the same as when H_2 is a full matrix. The reason for this is that specifying (controlling) any linear combination of y_2, uniquely determines the individual y_2's, since $\dim(u_2)=\dim(y_2)$. Thus, we can find the optimal selection matrix H_2, by searching through all the candidate square sets of y. This can be effectively solved using the branch and bound approach of Kariwala and Cao, or alternatively it can be solved as a mixed-integer problem with a quadratic program (QP) at each node (Yelchuru and Skogestad). The approach of Yelchuri and Sjkogestad can also be applied to the case where we allow for disjunct sets of measurement combinations, which may give a lower J_2 in some cases.

Comment on the state drift approach

1. We have assumed that we perfectly control y_2 using u_2, at least within the bandwidth of the regulatory control system. Once one has found a candidate control structure (H_2), one should check that it is possible to achieve acceptable control. This may be done by analyzing the input-output controllability of the system $y_2 = G_2 u_2 + G_{2d} d$, based on the transfer matrices $G_2 = H_2 G^y$ and $G_{2d} = H_2 G^y{}_d$. If the controllability of this system is not acceptable, then one should consider the second-best matrix H_2 (with the second-best value of the state drift J_2) and so on.

2. The state drift cost drift J2 = $||Wx||$ is in principle independent of the economic cost (J). This is an advantage because we know that the economically optimal operation (e.g., active constraints) may change, whereas we would like the regulatory layer to remain unchanged. However, it is also a disadvantage, because the regulatory layer determines the initial response to disturbances, and we would like this initial response to be in the right direction economically, so that the required correction from the slower supervisory layer is as small as possible. Actually, this issue can be included by extending the state vector x to include also the economic controlled variables, CV1, which is selected based on the economic cost J. The weight matrix W may then be used to adjust the relative weights of avoiding drift in the internal states x and economic controlled variables CV1.

3. The above steady-state approach does not consider input-output pairing, for which dynamics are usually the main issue. The main pairing rule is to "pair close" in order to minimize the effective time delay between the selected input and output. For a more detailed approach, decentralized input-output controllability must be considered.

8. Plantwide control (process control)

Skogestad has suggested a procedure for control structure design of processing plants (economic plantwide control) consisting of the following steps:

I Top Down Part

<u>Step S1</u>: Define operational objective (cost J) and constraints

<u>Step S2:</u> Identify degrees of freedom and optimize operation for disturbances

<u>Step S3</u>: Implementation of optimal operation: What to control ? Select primary CV1's (self-optimizing control)

<u>Step S4:</u> Where set the production rate? Location of throughput of manipulator (TPM)

II Bottom Up Part

Step S5: Structure of regulatory control layer: What more to control (secondary CV2's) ?

Step S6: Structure of supervisory control layer

Step S7: Structure of Real-time optimization layer (if any)

Luyben has proposed a similar procedure for plantwide control, with more emphasis on regulatory control,.

Note here Step S4 on locating "throughput manipulator" (TPM) which is special for process control. This decision should be based on the overall economics, and the main rule is to locate it at the expected production bottleneck of the plant. The location is very important because it determines the structure of the entire inventory control system (pressure and level control), which should be "radiating" around the location of the TPM in order to get "local" input-output pairings.

In process control, typical variables that are selected for regulatory control (CV2) are inventories such as liquid levels, accumulating components, pressures and other "drifting" variables such as certain temperatures. All of these variables are related to the "drift" in the process and thus may be captured with the state drift

REFERENCES and FURTHER READING

V. Alstad and S. Skogestad, "Null Space Method for Selecting Optimal Measurement Combinations as Controlled Variables", *Ind. Eng. Chem. Res.*, **46** (3), 846-853 (2007).

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

J.J. Downs and S. Skogestad, ``An industrial and academic perspective on plantwide control", *Annual Reviews in Control*, **17,** 99-110 (2011).

Engell, S. (2007). Feedback control for optimal process operation. J. Proc. Control 17, 203–219.

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

V. Kariwala, Y. Cao, " Bidirectional branch and bound for controlled variable selection. Part III. Local average loss minimization", IEEE Transactions on Industrial Informatics 6 (2010) 54–61.

I.K. Kookos and Perkins, J.D., "An Algorithmic method for the selection of multivariable process control structures", *J. Proc. Control*, 12, 85-99 (2002).

Luyben, M.L., Tyreus, B.D. & Luyben, W.L. **1997**. Plant-wide Control Design Procedure. *AIChE Journal*, *43*, 3161-3174.

M. Morari, Y. Arkun, and G. Stephanopoulos, "Studies in the synthesis of control structures for chemical processes. Part I.", *AIChE Journal*, **26**, 209-214 (1973).

L.T. Narraway, J.D. Perkins, “Selection of control structure based on economics”, *Computers and Chem. Eng.*, **18**, S511–S515 (1993).

F.G. Shinskey, *Process Control Systems*, 3rd ed., McGraw-Hill, 1988

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

S. Skogestad, “Control structure design for complete chemical plants”, *Computers and Chem. Eng.*, **28** (1-2), 219-234 (2004).

S. Skogestad and I. Postlethwaite, *Multivariable feedback control*, Wiley, 2nd edition (2005).

Sigurd Skogestad ''Economic Plantwide Control'', Chapter 11 in: *Plantwide control. recent Developments and Applications*, G.P. Rangaiah and V. kariwala (Editors). Wiley, Chichester, 2012. ISBN 978-0-470-98014-9. pp. 229-251.

M. van de Wal and B. de Jager, “Review of methods for input/output selection”, Automatica 37 (2001) 487-510

R. Yelchuru and S. Skogestad, ``Quantitative methods for regulatory layer selection'', *Journal of Process Control*, **23** 58-69 (2013).