Bibliography

Abkowitz M (1964) Lecture notes on ship hydrodynamics–steering and manoeuvrability. Technical report Hy-5, Hydro and Aerodynamics Laboratory, Lyngby

Arnold R, Maunder L (1961) Gyrodynamics and its engineering applications. Academic, New York/London

Baitis E, Woollaver D, Beck T (1983) Rudder roll stabilization of coast guard cutters and frigates. Nav Eng J 95(3):267–282

Blanke M, Haals P, Andreasen KK (1989) Rudder roll damping experience in Denmark. In: Proceedings of IFAC workshop CAMS'89, Lyngby

Chalmers T (1931) The automatic stabilisation of ships. Chapman and Hall, London

Crossland P (2003) The effect of roll stabilization controllers on warship operational performance. Control Eng Pract 11:423–431

Donaire A, Perez T (2013) Energy-based nonlinear control of ship roll gyro-stabiliser with precession angle constraints. In: 9th IFAC conference on control applications in marine systems, Osaka

Fedyaevsky K, Sobolev G (1964) Control and stability in ship design. State Union Shipbuilding, Leningrad

Fossen TI (2011) Handbook of marine craft hydrodynamics and motion control. Wiley, Chichester

Fossen T, Perez T (2009) Kalman filtering for positioning and heading control of ships and offshore rigs. IEEE Control Syst Mag 29(6):32–46

Gaillarde G (2002) Dynamic behavior and operation limits of stabilizer fins. In: IMAM international maritime association of the Mediterranean, Creta

Grimble M, Katebi M, Zang Y (1993) $\mathcal{H}_\infty$–based ship fin-rudder roll stabilisation. In: 10th ship control system symposium SCSS, Ottawa, vol 5, pp 251–265

Haverre S, Moan T (1985) On some uncertainties related to short term stochastic modelling of ocean waves. In: Probabilistic offshore mechanics. Progress in engineering science. CML

Hearns G, Blanke M (1998) Quantitative analysis and design of rudder roll damping controllers. In: Proceedings of CAMS'98, Fukuoka, pp 115–120

Holden C, Fossen TI (2012) A nonlinear 7-DOF model for U-tanks of arbitrary shape. Ocean Eng 45: 22–37

Källström C, Wessel P, Sjölander S (1988) Roll reduction by rudder control. In: Spring meeting-STAR symposium, 3rd IMSDC, Pittsburgh

Lloyd A (1989) Seakeeping: ship behaviour in rough weather. Ellis Horwood

Marzouk OA, Nayfeh AH (2009) Control of ship roll using passive and active anti-roll tanks. Ocean Eng 36:661–671

Oda H, Sasaki M, Seki Y, Hotta T (1992) Rudder roll stabilisation control system through multivariable autoregressive model. In: Proceedings of IFAC conference on control applications of marine systems–CAMS

Perez T (2005) Ship motion control. Advances in industrial control. Springer, London

Perez T, Blanke M (2012) Ship roll damping control. Ann Rev Control 36(1):1367–5788

Perez T, Steinmann P (2009) Analysis of ship roll gyrostabiliser control. In: 8th IFAC international conference on manoeuvring and control of marine craft, Guaruja

Roberts G, Sharif M, Sutton R, Agarwal A (1997) Robust control methodology applied to the design of a combined steering/stabiliser system for warships. IEE Proc Control Theory Appl 144(2):128–136

Sellars F, Martin J (1992) Selection and evaluation of ship roll stabilization systems. Mar Technol SNAME 29(2):84–101

Stoustrup J, Niemann HH, Blanke M (1994) Rudder-roll damping for ships- a new $\mathcal{H}_\infty$ approach. In: Proceedings of 3rd IEEE conference on control applications, Glasgow, pp 839–844

van Amerongen J, van der Klugt P, van Nauta Lemke H (1990) Rudder roll stabilization for ships. Automatica 26:679–690

Control Structure Selection

Sigurd Skogestad
Department of Chemical Engineering,
Norwegian University of Science and
Technology (NTNU), Trondheim, Norway

Abstract

Control structure selection deals with selecting what to control (outputs), what to measure and what to manipulate (inputs), and also how to split the controller in a hierarchical and decentralized manner. The most important issue is probably the selection of the controlled variables (outputs), CV = Hy, where y are the available measurements and H is a degree of freedom that is seldom treated in a systematic manner by control engineers. This entry discusses how to find H for both for the upper (slower) economic layer and the lower (faster) regulatory layer in the control hierarchy. Each layer may be split in a decentralized fashion. Systematic approaches for input/output (IO) selection are presented.

Keywords

Control configuration; Control hierarchy; Control structure design; Decentralized control;

Economic control; Input-output controllability; Input/output selection; Plantwide control; Regulatory control; Supervisory control

Introduction

Consider the generalized controller design problem in Fig. 1 where P denotes the generalized plant model. Here, the objective is to design the controller K, which, based on the sensed outputs v, computes the inputs (MVs) u such that the variables z are kept small, in spite of variations in the variables w, which include disturbances (d), varying setpoints/references (CV_s) and measurement noise (n),

$$w = [d, CV_s, n]$$

The variables z, which should be kept small, typically include the control error for the selected controlled variables (CV) plus the plant inputs (u),

$$z = [CV - CV_s; u]$$

The variables v, which are the inputs to the controller, include all known variables, including measured outputs (y_m), measured disturbances (d_m) and setpoints,

$$v = [y_m; d_m; CV_s].$$

The cost function for designing the optimal controller K is usually the weighted control error,
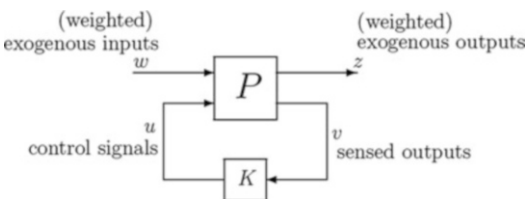
Control Structure Selection, Fig. 1 General formulation for designing the controller K. The plant P is controlled by manipulating u, and is disturbed by the signals w. The controller uses the measurements v, and the control objective is to keep the outputs (weighted control error) z as small as possible

$J' = ||W'z||$. The reason for using a prime on J (J'), is to distinguish it from the economic cost J which we later use for selecting the controlled variables (CV).

Notice that it is assumed in Fig. 1 that we know what to measure (v), manipulate (u), and, most importantly, which variables in z we would like to keep at setpoints (CV), that is, we have assumed a given control structure. The term "control structure selection" (CSS) and its synonym "control structure design" (CSD) is associated with the overall *control philosophy* for the system with emphasis on the **structural decisions** which are a prerequisite for the controller design problem in Fig. 1:

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

Decisions 2 and 3 (selection of u and y) are sometimes referred to as the input/output (IO) selection problem. In practice, the controller (K) is usually divided into several layers, operating on different time scales (see Fig. 2), which implies that we in addition to selecting the (primary) controlled variables ($CV_1 \equiv CV$) must also select the (secondary) variables that interconnect the layers (CV_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 (Fig. 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.
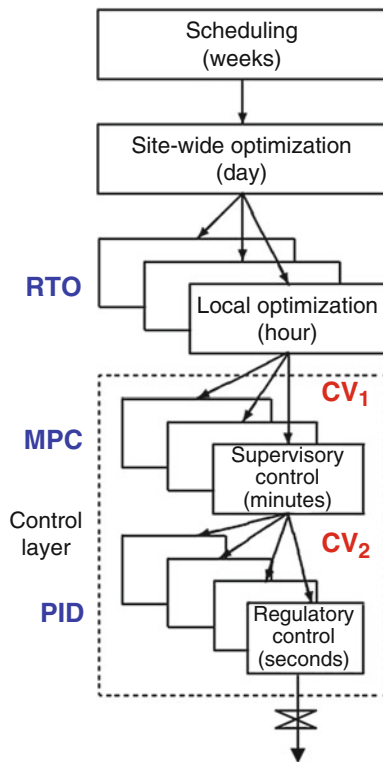
Control Structure Selection, Fig. 2 Typical control hierarchy, as illustrated for a process plant

The term "plantwide control," which is a synonym for "control structure selection," is used in the field of process control. Control 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. Yet, control structure selection is hardly covered in most control courses. This is probably related to the complexity of the problem, which requires the knowledge from several engineering fields. In the mathematical sense, the control structure selection problem is a formidable combinatorial problem which involves a large number of discrete decision variables.

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 usually two main objectives for control:

1. Longer-term economic operation (minimize economic cost J subject to satisfying operational constraints)
2. Stability and short-term regulatory control

The first objective is related to "making the system operate as intended," where economics are an important issue. Traditionally, control engineers have not been much involved in this step. The second objective is related to "making sure the system stays operational," where stability and robustness are important issues, and this has traditionally been the main domain of control engineers. In terms of designing the control system, the second objective (stabilization) is usually considered first. An example is bicycle riding; we first need to learn how to stabilize the bicycle (regulation), before trying to use it for something useful (optimal operation), like riding to work and selecting the shortest path.

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 then be given as constraints.

In theory, the optimal strategy is to combine the control tasks of optimal economic operation and stabilization/regulation 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 involves only a few variables. Such control systems can be designed and tuned with much less effort, especially when it comes to the modeling and tuning effort.

So how are large-scale systems controlled in practise? Usually, the controller K is decomposed

into several subcontrollers, using two main principles

- *Decentralized (local) control.* This "horizontal decomposition" of the control layer is usually based on separation in space, for example, by using local control of individual units.
- *Hierarchical (cascade) control.* This "vertical decomposition" is usually based on time scale separation, as illustrated for a process plant in Fig. 2. The upper three layers in Fig. 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.

In accordance with the two main objectives for control, the control layer is in most cases divided hierarchically in two layers (Fig. 2):

1. A "slow" supervisory (economic) layer
2. A "fast" regulatory (stabilization) layer

Another reason for the separation in two control layers, is that the tasks of economic operation and regulation are fundamentally different. Combining the two objectives in a single cost function, which is required for designing a single centralized controller K, is like trying to compare apples and oranges. For example, how much is an increased stability margin worth in monitory units [$]? 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.

Notation and Matrices H_1 and H_2 for Controlled Variable Selection

The most important notation is summarized in Table 1 and Fig. 3. To distinguish between the two control layers, we use "1" for the upper supervisory (economic) layer and "2" for the regulatory layer, which is "secondary" in terms of its place in the control hierarchy.

There is often limited possibility to select the input set (u) as it is usually constrained by the
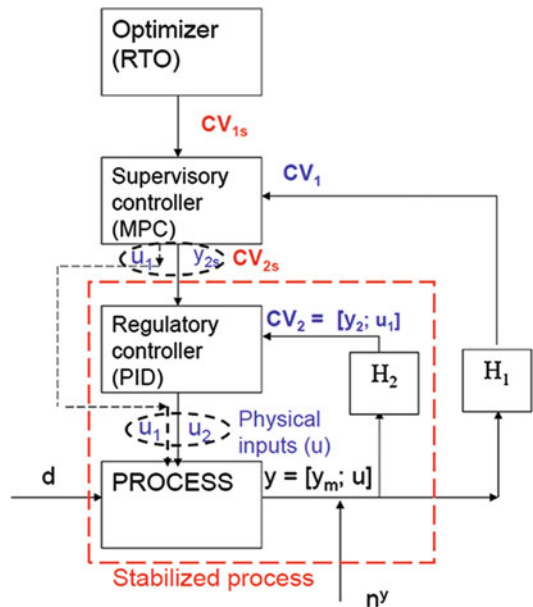
Control Structure Selection, Fig. 3 Block diagram of a typical control hierarchy, emphasizing the selection of controlled variables for supervisory (economic) control ($CV_1 = H_1 y$) and regulatory control ($CV_2 = H_2 y$)

Control Structure Selection, Table 1 Important notation

$u = [u_1; u_2] =$ set of all available physical plant inputs
$u_1 =$ inputs used directly by supervisory control layer
$u_2 =$ inputs used by regulatory layer
$y_m =$ set of all measured outputs
$y = [y_m; u] =$ combined set of measurements and inputs
$y_2 =$ controlled outputs in regulatory layer (subset or combination of y); $\dim(y_2) = \dim(u_2)$
$CV_1 = H_1 y =$ controlled variables in supervisory layer; $\dim(CV_1) = \dim(u)$
$CV_2 = [y_2; u_1] = H_2 y =$ controlled variables in regulatory layer; $\dim(CV_2) = \dim(u)$
$MV_1 = CV_{2s} = [y_{2s}; u_1] =$ manipulated variables in supervisory layer; $\dim(MV_1) = \dim(u)$
$MV_2 = u_2 =$ manipulated variables in regulatory layer; $\dim(MV_2) = \dim(u_2) \leq \dim(u)$

plant design. However, there may be a possibility to add inputs or to move some to another location, for example, to avoid saturation or to reduce the time delay and thus improve the input-output controllability.

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_1 and H_2 (see Fig. 3).

$$CV_1 = H_1 y$$
$$CV_2 = H_2 y$$

Note from the definition in Table 1 that $y = [y_m; u]$. Thus, 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 (or, more precisely, the input is left "unused" for control in this layer).

In general, H_1 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 usually a "selection matrix," where each row in H_2 contains one 1-element (to identify the selected variable) with the remaining elements set to 0. In this case, we can write $CV_2 = H_2 y = [y_2; u_1]$, where y_2 denotes the actual controlled variables in the regulatory layer, whereas u_1 denotes the "unused" inputs (u_1), which are left as degrees of freedom for the supervisory layer. Note that this indirectly determines the inputs u_2 used in the regulatory layer to control y_2, because u_2 is what remains in the set u after selecting u_1. To have a simple control structure, with as few regulatory loops as possible, it is desirable that H_2 is selected such that there are many inputs (u_1) left "unused" in the regulatory layer.

Example. Assume there are three candidate output measurements (temperatures T) and two inputs (flowrates q),

$$y_m = [T_a T_b T_c], \quad u = [q_a q_b]$$

and we have by definition $y = [y_m; u]$. Then the choice

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

means that we have selected $CV_2 = H_2 y = [T_b; q_b]$. Thus, $u_1 = q_b$ is an unused input for regulatory control, and in the regulatory layer we close one loop, using $u_2 = q_a$ to control $y_2 = T_b$. If we instead select

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

then we have $CV_2 = [T_a; T_c]$. None of these are inputs, so u_1 is an empty set in this case. This means that we need to close two regulatory loops, using $u_2 = [q_a; q_b]$ to control $y_2 = [T_a; T_c]$.

Supervisory Control Layer and Selection of Economic Controlled 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 cases, a good engineer can make a reasonable choice based on process insight and experience. However, the control engineer must realize that this is a critical decision. 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, but more generally requires optimization based on a detailed model.

   For example, consider the problem of minimizing the driving time between two cities (cost $J = T$). There is a single input ($u = $ fuel flow $f\ [l/s]$) and the optimal solution is often constrained. When driving a fast car, the active constraint may be the speed limit ($CV_1 = v\ [km/h]$ with setpoint v_{max}, e.g., $v_{max} = 100\ km/h$). When driving

Control Structure Selection, Table 2 Objectives of supervisory control layer

O1. Control primary "economic" variables CV_1 at setpoint using as degrees of freedom MV_1, which includes the setpoints to the regulatory layer ($y_{2s} = CV_{2s}$) 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 additional inputs, for example, to improve the dynamic performance (usually combined with input midranging control) or to extend the steady-state operating range (split range control)
O7. Make use of extra measurements, for example, to estimate the primary variables CV_1

an old car, the active constraint maybe the maximum fuel flow ($CV_1 = f [l/s]$ with setpoint f_{max}). The latter corresponds to an input constraint ($u_{max} = f_{max}$) which is trivial to implement ("full gas"); 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* (safety margin) to guarantee feasibility. The backoff is defined as the difference between the optimal value and the actual setpoint, for example, we need to back off from the speed limit because of the possibility for measurement error and imperfect control

$$CV_{1,s} = CV_{1,max} - \text{backoff}$$

For example, to avoid exceeding the speed limit of 100 km/h, we may set backoff = 5 km/h, and use a setpoint $v_s = 95$ km/h rather than 100 km/h.

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 ($CV_1 = H_1y$) are variables which when kept constant, indirectly (through the action of the feedback control system) lead to close-to optimal adjustment of the inputs (u) when there are disturbances (d).

- 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 rarely can measure the gradient J_u, we select $CV_1 = H_1y$. The selection of a good H_1 is a nontrivial task, but some quantitative approaches are given below.

For example, consider again the problem of driving between two cities, but assume that the objective is to minimize the total fuel, $J = V$ [liters]., Here, driving at maximum speed will consume too much fuel, and driving too slow is also nonoptimal. This is an unconstrained optimization problem, and identifying a good CV_1 is not obvious. One option is to maintain a constant speed ($CV_1 = v$), but the optimal value of v may vary depending on the slope of the road. A more "self-optimizing" option, could be to keep a constant fuel rate ($CV_1 = f [l/s]$), which will imply that we drive slower uphill and faster downhill. More generally, one can control combinations, $CV_1 = H_1y$ where H_1 is a "full" matrix.

CV_1 Rule 3. For the unconstrained degrees of freedom, one should *never* control a variable that reaches its maximum or minimum value at the optimum, for example, never try to control directly the cost J. Violation of this rule gives either infeasibility (if attempting to control J at a lower value than J_{min}) or nonuniqueness (if attempting to control J at higher value than J_{min}).

Assume again that we want to minimize the total fuel needed to drive between two cities, $J = V$ [l]. Then one should avoid fixing the

total fuel, $CV_1 = V$ [l], or, alternatively, avoid fixing the fuel consumption("gas mileage") in liters pr. km ($CV_1 = f$ [l/km]). Attempting to control the fuel consumption[l/km] <u>*below*</u> *the car's minimum value is obviously not possible (infeasible). Alternatively, attempting to control the fuel consumption* <u>*above*</u> *its minimum value has two possible solutions; driving slower or faster than the optimum. Note that the policy of controlling the fuel rate f [l/s] at a fixed value will never become infeasible.*

For CV_1-Rule 2, it is always possible to find good variable combinations (i.e., H_1 is a "full" matrix), at least locally, but whether or not it is possible to find good individual variables (H_1 is a selection matrix), is not obvious. To help identify potential "self-optimizing" variables ($CV_1 = c$), the following requirements may be used:

Requirement 1. The *optimal* value of c is insensitive to disturbances, that is, $dc_{opt}/dd = H_1F$ is small. Here $F = dy_{opt}/dd$ is the optimal sensitivity matrix (see below).

Requirement 2. The variable c 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. Here $G^y = dy/du$ is the measurement gain matrix (see below).

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. If the optimal active constraints vary depending on the disturbances, new controlled variables (CV_1) must be identified (offline) for each active constraint region, and online switching is required to maintain optimality. In practise, it is easy to identify when to switch when one reaches a constraint. It is 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 CV_1 in some regions of active constraints. In this case, the setpoint for CV_1 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 may either be centralized, e.g., using model predictive control (MPC), or decomposed into simpler subcontrollers using standard elements, like decentralized control (PID), cascade control, selectors, decouplers, feedforward elements, ratio control, split range control, and input midrange control (also known as input resetting, valve position control or habituating control). In theory, the performance is better with the centralized approach (e.g., MPC), but the difference can be small when designed by a good engineer. The main reasons for using simpler elements is that (1) the system can be implemented in the existing "basic" control system, (2) it can be implemented with little model information, and (3) it can be build up gradually. However, such systems can quickly become complicated and difficult to understand for other than the engineer who designed it. Therefore, model-based centralized solutions (MPC) are often preferred because the design is more systematic and easier to modify.

Quantitative Approach for Selecting Economic Controlled Variables, CV_1

A quantitative approach for selecting economic controlled variables is to consider the effect of the choice $CV_1 = H_1y$ on the economic cost J when disturbances d occur. One should also include noise/errors (n^y) related to the measurements and inputs.

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

$$J = \text{cost feed} + \text{cost utilities (energy)}$$
$$- \text{value products [\$/s]}$$

Note that the economic cost function J is used to *select* the controlled variables (CV_1), and another cost function (J'), typically involving the deviation in CV_1 from their optimal setpoints CV_{1s}, is used for the actual controller design (e.g., using 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: $g(u, x, d) \leq 0$

In many cases, the economics are determined by the steady-state behavior, so we can set $dx/dt = 0$. The optimization problem should be resolved for the expected disturbances (d) to find the truly optimal operation policy, $u_{opt}(d)$. The nominal solution (d_{nom}) may be used to obtain the setpoints (CV_{1s}) for the selected controlled variables. In

practise, the optimum input $u_{opt}(d)$ cannot be realized, because of model error and unknown disturbances d, so we use a feeback implementation where u is adjusted to keep the selected variables CV_1 at their nominally optimal setpoints.

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 supervisory (economic) controlled variables, CV_1*

CV_1-Rule 1: Control Active Constraints

A primary goal for solving the optimization problem 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 (d) and market conditions (prices).

CV_1-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 *control "self-optimizing" variables* c.

1. **"Brute force" approach.** Given a set of controlled variables $CV_1 = c = H_1y$, one computes the cost J(c,d) when we keep c constant ($c = c_s + H_1n^y$) 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.

$$y = G^yu + G_d^yd$$

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 $G = H_1G^y$.**

In other words, the maximum gain rule, which essentially is a quantitative version of Requirements 1, 3 and 4 given above, says that one should control "sensitive" variables, with a large scaled gain G from the inputs (u) to $c = H_1 y$. 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 matrix F, defined as

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

Each column in $\mathbf{F}$ 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

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

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 $c = H_1 y$ such that

$$H_1 \mathbf{F} = 0$$

Note that H_1 is a nonsquare matrix, so $H_1 F = 0$ does not require that $H_1 = 0$ (which is a trivial uninteresting solution), but rather that H_1 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_1 to solve the following problem

$$\min\_H_1 ||M(H_1)||_2$$

where 2 denotes the Frobenius norm and

$$M(H_1) = J_{uu}^{1/2}(H_1 G^y)^{-1} H_1 Y, Y$$
$$= \begin{bmatrix} FW_d & W_{ny} \end{bmatrix}.$$

Note here that the optimal choice with $W_{ny} = 0$ (no noise) is to choose H_1 such that $H_1 F = 0$, which is the nullspace method. For the general case, when H_1 is a "full" matrix, this is a convex problem and the optimal solution is $H_1^T = (YY')^{-1} G^y Q$ where Q is any nonsingular matrix.

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 sense 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 primary 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 controlled outputs y_2 (among all candidate measurements y_m).
2. Selection of inputs $MV_2 = u_2$ (a subset of all available inputs u) to control the outputs y_2.
3. Pairing of inputs u_2 and outputs y_2 (since decentralized control is normally used).

Decisions 1 and 2 combined (IO selection) is equivalent to selecting H_2 (Fig. 3). Note that we do not "use up" any degrees of freedom in the regulatory layer because the set points (y_{2s})

become manipulated variables (MV_1) for the supervisory layer (see Fig. 3). Furthermore, since the set points are set by the supervisory layer in a cascade manner, the system eventually approaches the same steady-state (as defined by the choice of economic variables CV_1) regardless 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 a backoff for the input when entering this operational region, and doing so will have an economic penalty.

In the regulatory layer, the outputs (y_2) are usually selected as individual measurements and they are often not important variables in themselves. Rather, they are "extra outputs" that are controlled in order to "stabilize" the system, and their setpoints (y_{2s}) are changed by the layer above, to obtain economical optimal operation. For example, in a distillation column one may control a temperature somewhere in the middle of the column ($y_2 = T$) in order to "stabilize" the column profile. Its setpoint ($y_{2s} = T_s$) is adjusted by the supervisory layer to obtain the desired product composition ($y_1 = c$).

Input-Output (IO) Selection for Regulatory Control (u_2, y_2)

Finding the truly optimal control structure, including selecting inputs and outputs for regulatory control, requires finding also the optimal controller parameters. This is an extremely difficult mathematical problem, at least if the controller K is decomposed into smaller controllers. In this section, we consider some approaches which does not require that the controller parameters be found. This is done by making assumptions related to achievable control performance (controllability) or perfect control.

Before we look at the approaches, note again that the IO-selection for regulatory control may be combined into a single decision, by considering the selection of

$$CV_2 = [y_2; u_1] = H_2 y$$

Here u_1 denotes the inputs that are not used by the regulatory control layer. This follows because we want to use all inputs u for control, so assuming 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" inputs u_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. "Accessibility" 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 analyzing a given structure (u_2, y_2) but 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.

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 (2013) list 13 partly conflicting objectives. To have a truly systematic approach to regulatory control design, including IO-selection, we would need to quantify all these partially conflicting objectives in terms of a scalar cost function J_2. We here consider a fairly general cost function,

$$J_2 = ||Wx||$$

which may be interpreted as the weighted state drift. One 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. Note that the cost J_2 is used to *select* controlled variables (CV_2) and not to design the controller (for which the cost may be the control error, $J_2' = ||CV_2 - CV_{2s}||$).

Within this framework, the IO-selection problem for the regulatory layer is then to select the nonsquare matrix H_2,

$$CV_2 = H_2y$$

where $y = [y_m; u]$, such that the cost J_2 is minimized. The cause for changes in J_2 are disturbances d, and we consider the linear model (in deviation variables)

$$y = G^y u + G^y_d d$$
$$x = G^x u + G^x_d d$$

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 to reduce the disturbance effect by closing the regulatory control loops. Within the "slow" time scale of the supervisory layer, we can assume that CV_2 is perfectly controlled and thus constant, or $CV_2 = 0$ 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

$$u = -\left(H_2 G^y\right)^{-1} \left(H_2 G^y_d\right) d$$

and we have

$$x = P^x_d (H_2) d$$

where

$$P^x_d (H_2) = G^x_d - G^x \left(H_2 G^y\right)^{-1} H_2 G^y_d$$

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 than 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||$M_2 (H_2)$ ||

where M_2 = WP^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 J_2-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_d^x 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 $||M_2||$ 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 ($\omega = 0$), that is, we can use the steady-state gain matrices when computing P_d^x. 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, which we should, then the expected value of J_2 is minimized by solving the problem $min\_H_2 \; ||M_2(H_2)||_2$ where (Yelchuru and Skogestad 2013)

$$M_2(H_2) = J_{2uu}^{1/2}(H_2 G^y)^{-1} H_2 Y_2$$

$$Y_2 = [F_2 W_d \;\; W_n]; \quad F_2 = \frac{\partial y_{opt}}{\partial d}$$
$$= G^y J_{2uu}^{-1} J_{2ud} - G_{dy}$$

where $J_{2uu} \triangleq \frac{\partial^2 J_2}{\partial u^2} = 2 G^{x^T} W^T W G^x$, $J_{2ud} \triangleq \frac{\partial^2 J_2}{\partial u \partial d} = 2 G^{x^T} W^T W G_{dx}$,

Note that this is the same mathematical problem as the "exact local method" presented for selecting $CV_1 = H_1 y$ for minimizing the economic cost J, but because of the specific simple form

for the cost J_2, it is possible to obtain analytical formulas for the optimal sensitivity, F_2. Again, W_d and W_{ny} are diagonal matrices, expressing the expected magnitude of the disturbances (d) 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

$$H_2^T = (Y_2 Y_2^T)^{-1} G^y (G^{y^T}(Y_2 Y_2^T)^{-1} G^y)^{-1} J_{2uu}^{1/2}$$

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 2012). The approach of Yelchuru and Skogestad 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.

Comments 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 $G_{2d} = H_2 G_d^y$. 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 $J_2 = ||Wx||$ 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, CV_1, 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 CV_1.

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.

Summary and Future Directions

Control structure design involves the structural decisions that must be made before designing the actual controller, and it is in most cases a much more important step than the controller design. In spite of this, the theoretical tools for making the structural decisions are much less developed than for controller design. This chapter summarizes some approaches, and it is expected, or at least hoped, that this important area will further develop in the years to come.

The most important structural decision is usually related to selecting the economic controlled variables, $CV_1 = H_1y$, and the stabilizing controlled variables, $CV_2 = H_2y$. However, control engineers have traditionally not used the degrees of freedom in the matrices H_1 and H_2, and this chapter has summarized some approaches.

There has been a belief that the use of "advanced control," e.g., MPC, makes control structure design less important. However, this is not correct because also for MPC must one choose inputs ($MV_1 = CV_{2s}$) and outputs (CV_1). The selection of CV_1 may to some extent be avoided by use of "Dynamic Real-Time Optimization (DRTO)" or "Economic MPC," but these optimizing controllers usually operate on a slower time scale by sending setpoints to the basic control layer ($MV_1 = CV_{2s}$), which means that selecting the variables CV_2 is critical for achieving (close to) optimality on the fast time scale.

Cross-References

▶ Control Hierarchy of Large Processing Plants: An Overview
▶ Industrial MPC of Continuous Processes
▶ PID Control

Bibliography

Alstad V, Skogestad S (2007) Null space method for selecting optimal measurement combinations as controlled variables. Ind Eng Chem Res 46(3): 846–853

Alstad V, Skogestad S, Hori ES (2009) Optimal measurement combinations as controlled variables. J Process Control 19:138–148

Downs JJ, Skogestad S (2011) An industrial and academic perspective on plantwide control. Ann Rev Control 17:99–110

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

Foss AS (1973) Critique of chemical process control theory. AIChE J 19(2):209–214

Kariwala V, Cao Y (2010) Bidirectional branch and bound for controlled variable selection. Part III. Local average loss minimization. IEEE Trans Ind Inform 6: 54–61

Kookos IK, Perkins JD (2002) An Algorithmic method for the selection of multivariable process control structures. J Proc Control 12:85–99

Morari M, Arkun Y, Stephanopoulos G (1973) Studies in the synthesis of control structures for chemical processes. Part I. AIChE J 26:209–214

Narraway LT, Perkins JD (1993) Selection of control structure based on economics. Comput Chem Eng 18:S511–S515

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

Skogestad S (2004) Control structure design for complete chemical plants. Comput Chem Eng 28(1–2):219–234

Skogestad S (2012) Economic plantwide control, chapter 11. In: Rangaiah GP, Kariwala V (eds) Plantwide control. Recent developments and applications. Wiley, Chichester, pp 229–251. ISBN:978-0-470-98014-9

Skogestad S, Postlethwaite I (2005) Multivariable feedback control, 2nd edn. Wiley, Chichester

van de Wal M, de Jager B (2001) Review of methods for input/output selection. Automatica 37:487–510

Yelchuru R, Skogestad S (2012) Convex formulations for optimal selection of controlled variables and measurements using Mixed Integer Quadratic Programming. J Process Control 22:995–1007

Yelchuru R, Skogestad S (2013) Quantitative methods for regulatory layer selection. J Process Control 23:58–69

Controllability and Observability

H.L. Trentelman
Johann Bernoulli Institute for Mathematics and
Computer Science, University of Groningen,
Groningen, AV, The Netherlands

Abstract

State controllability and observability are key properties in linear input–output systems in state-space form. In the state-space approach, the relation between inputs and outputs is represented using the state variables of the system. A natural question is then to what extent it is possible to manipulate the values of the state vector by means of an appropriate choice of the input function. The concepts of controllability, reachability, and null controllability address this issue. Another important question is whether it is possible to uniquely determine the values of the state vector from knowledge of the input and output signals over a given time interval. This question is dealt with using the concept of observability.

Keywords

Introduction

In the state-space approach to input–output systems, the relation between input signals and output signals is represented by means of two equations. In the continuous-time case, the first of these equations is a first-order vector differential equation driven by the input signal and is often called *the state equation*. The second equation is an algebraic equation, often called the *output equation*. The unknown in the differential equation is called the *state vector* of the system. Given a particular input signal and initial value of the state vector, the state equation generates a unique solution, called the state trajectory of the system. The output equation determines the corresponding output signal as a function of this state trajectory and the input signal. Thus, in the state space approach, the input–output behavior of the system is obtained using the state vector as an intermediate variable.

In the context of input–output systems in state-space form, the properties of controllability and observability characterize the interaction between the input, the state, and the output. In particular, controllability describes the ability to manipulate the state vector of the system by applying appropriate input signals. Observability describes the ability to determine the values of the state vector from knowledge of the input and output over a certain time interval. The properties of controllability and observability are fundamental properties that play a major role in the analysis and control of linear input–output systems in state-space form.