PERGAMON

Automatica 37 (2001) 487–510

automatica

www.elsevier.com/locate/automatica

Survey Paper

A review of methods for input/output selection☆

Marc van de Wal[a,1], Bram de Jager[b,*]

[a]Philips CFT, Mechatronics Motion, P.O. Box 218, SAQ-2116, 5600 MD Eindhoven, Netherlands
[b]Faculty of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, Netherlands

Received 11 June 1998; revised 3 July 2000; received in final form 6 September 2000

Abstract

Control system design involves input/output (IO) selection, that is, decisions on the number, the place, and the type of actuators and sensors. The choice of inputs and outputs affects the performance, complexity, and costs of the control system. Due to the combinatorial nature of the selection problem, systematic methods are needed to complement one's intuition, experience, and physical insight. This paper reviews the currently known IO selection methods, which aids the control engineer in picking a suitable method for the problem at hand. The methods are grouped according to the control system property that is addressed and applications are grouped according to the considered control systems. A set of criteria is proposed that a good IO selection method should possess. It is used to assess and compare the methods and it could be used as a guideline for new methods. The state of the art in IO selection is sketched and directions for further research are mentioned.

Keywords: Control system design; Structural properties; Input signals; Output signals; Controllability; Observability; Relative gain array; Robust control

1. Introduction

Control system design could be split into the following six steps. First, the control goals are formulated. This involves choosing and characterizing the exogenous variables w and choosing and imposing requirements on the controlled variables z, see Fig. 1. The control goals should be quantified in the time and/or frequency domain. The choice of z may be affected by the outcome of the other steps of control system design, like the choice of the plant model G in the second step of control system design. The techniques used in other steps also determine the model type, e.g., linear or nonlinear, time-invariant or time-varying, physical principles or black box. In the third step, the control structure is selected. Fourth, the controller K is designed. The choice of the design method (PID, LQG, adaptive control, $\mathscr{H}_\infty$ optimization, etc.) depends on aspects like the control goal, model accuracy, and restrictions on the implementation. Fifth, the closed-loop system is evaluated via simulations or pilot plant experiments. In the last step, the controller and hardware like sensors, actuators, and control processors are implemented in the real plant. Iterative refinements of these steps are often necessary, e.g., meeting the control goal might call for a more accurate model or a modification of controller parameters.

The third step of control system design, i.e., control structure selection, is usually split into two separate problems which are then solved successively: input/output (IO) selection, which is the focus of this paper, and control configuration (CC) selection. Approaches to solve IO and CC selection jointly are seldomly encountered in the literature. Here, the IO selection problem is posed as follows:

> Select suitable variables u to be manipulated by the controller (plant inputs) and suitable variables y to be supplied to the controller (plant outputs).

Both the inputs and the outputs of G are divided into two classes, but in this paper the terms "inputs" and "outputs" are reserved for u and y. Each combination of

☆This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Editor Manfred Morari.

* Corresponding author. Tel.: +31-40-2472784; fax: +31-40-2461418.

E-mail addresses: m.m.j.van.de.wal@philips.com (M. van de Wal), a.g.de.jager@wfw.wtb.tue.nl (B. de Jager).

[1] http://www.wfw.wtb.tue.nl/english/research/control/marcw/.

0005-1098/01/$ - see front matter
PII: S0005-1098(00)00181-3

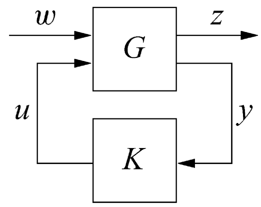


Fig. 1. General control system set-up; plant G, controller K, exogenous variables w, controlled variables z, manipulated variables (inputs) u, and measured variables (outputs) y.

inputs and outputs is called an IO set. CC selection is only relevant for decentralized control, i.e., for controllers with one or more entries that are structurally zero. Establishing which outputs in y are used to determine each input in u is often called partitioning or pairing in case of a diagonal controller. Surveys of CC selection methods can be found in Cao (1995, Chapter 1), Kinnaert (1995), Van de Wal and De Jager (1995), and Van de Wal (1995).

IO selection involves selecting an appropriate number, place, and type of actuators and sensors. In general, IO selection is performed prior to the physical realization of the plant, based on a plant model and a proposed set of candidate actuators and sensors. Sometimes it may be desirable to perform IO selection for a plant that is already equipped with actuators and sensors. Reasons for not using all the available actuators and sensors could be the reduction of the control system complexity and the costs of operation and maintenance. Also, devices may be present that cannot be used at the same time, e.g., two valves that influence exactly the same flow. This can occur because different plant configurations, e.g., at startup and at full load, require different instrumentation. It is also possible to study the benefits of sensors and actuators besides or instead of those available. This situation occurs when new sensor or actuator technology becomes available, when an existing plant needs a refit, or when a plant has to meet tighter specifications.

IO selection involves more than actuator and sensor selection. Apart from physically sensed variables, y may include variables derived from the sensed variables (e.g., a differentiated version of a sensed position) and commanded setpoint or tracking variables. In analogy, apart from physically actuated variables, u may include reference variables for other control system parts. Despite a slight abuse in terminology, the outputs y will sometimes be called measured variables, whereas the inputs u will often be called manipulated variables. IO selection may also be used to decide on particular controller restrictions which will be called "controller schemes". These reflect the decomposition of the controller into different units, like feedback, feedforward, and cascade loops. In case of the general control system set-up, distinct controller schemes result in distinct plants G and distinct IO sets. Distinct candidate controller schemes can also be incorporated simultaneously into G. The majority of the currently known IO selection methods cannot be used for systems giving rise to a decentralized K in Fig. 1, due to the restriction to centralized controllers. An IO selection method that also handles decentralized controllers K in Fig. 1 could be used to solve IO and CC selection jointly, but such methods are rare in the literature.

Sensible IO selection is important. First, the IO set may limit the performance, which may not be overcome by advanced controller design. For instance, the IO set may give zeros in the right half-plane, imposing an upper bound on the bandwidth. Second, the IO set determines aspects like reliability and the expenses of hardware, implementation, operation, and maintenance. In this respect, a small IO set (i.e., an IO set with a small number of inputs and outputs) is often preferred to a large IO set. Especially in the context of fault-tolerance, redundancy of actuators and/or sensors may also be desired. This should then be incorporated in the selection. Normally, the selection can be used to find IO sets that show no redundancy at all, providing information about the performance losses due to instrumentation faults.

The number of candidate IO sets grows exponentially with the number of candidate inputs and outputs. Suppose there are N_u candidate inputs and N_y candidate outputs. Assume that these can all be used together if desired, i.e., assume that there is a well-defined full IO set including all N_u inputs and all N_y outputs. If certain devices are not compatible, the problem can sometimes be embedded in a larger problem where the devices are allowed to be used together. During IO selection, a subset of n_u inputs and n_y outputs, i.e., an $n_u \times n_y$ IO set, is generated from the full IO set (n_z will be used to denote the number of controlled variables z). The total number of distinct IO sets is given by $(2^{N_u} - 1)(2^{N_y} - 1) + 1$, with "$+1$" due to the empty IO set with $n_u = 0$ and/or $n_y = 0$ (open-loop system). The exponential growth of the problem motivates the need for systematic IO selection methods to quickly and easily assess a large number of candidate IO sets. Though IO selection could be performed by controller design and closed-loop evaluation for each candidate IO set, this is not feasible for anything but a small number of candidates.

This paper provides a comprehensive review of the currently known IO selection methods. As far as we know, the only reviews in the 1990s that are related to this topic and that are also rather extensive were given by Morari (1992) and Skogestad and Postlethwaite (1996, Chapter 10). A decade earlier, the articles by Morari, Arkun, and Stephanopoulos (1980) and Nishida, Stephanopoulos, and Westerberg (1981) should be mentioned. The main contributions are the following. First, a list of properties is proposed that a good IO selection method should possess (Section 2). This list is used to assess known methods and it could be used as a guideline to develop new ones. Second, a comprehensive review of IO selection methods is given that is believed to be rather complete (Section 3). To aid the control engineer who faces an IO selection problem in choosing the right method, these methods are grouped according to the control system property that they address. Third, an overview of applications of systematic IO selection methods is given, revealing that such methods are important for a wide variety of problems (Section 4). These applications are grouped according to the class of systems that is considered, so one can pick the method that best matches one's problem. Fourth, a qualitative assessment and comparison of the reviewed methods is given (Section 5). This sketches the state of affairs and gives rise to some future research topics.

2. Desirable properties of IO selection methods

Partly inspired by Nett (1989) and Reeves (1991), eight properties are proposed which we believe are the main properties of the ideal IO selection method. These properties will be used as the basis for a qualitative assessment of the IO selection methods to be reviewed. Desirably, an IO selection method is:

1. *Well-founded*: The theory behind an IO selection method must be sound and complete. The method should be easy to use and transparent, i.e., bearing the basic idea of the method in mind, the way in which the outcome is affected by a change in the control goals must be understandable. At least one convincing application should prove the method's practical relevance.

2. *Efficient*: An IO selection method should make it possible to quickly evaluate a large number of candidate IO sets. Algorithms are commonly called efficient if they solve problems in time polynomial in a measure of the problem size; if not, they are called inefficient. Based on De Jager and Toker (1998), it is unlikely that the majority of the methods to be reviewed can solve the IO selection problem in time polynomial in N_u and N_y. Therefore, the term efficiency is used in a less formal context here, namely to express the expected analytical and computational effort related to a problem which is not solvable in polynomial time.

3. *Effective*: Effectiveness implies that those candidate IO sets are eliminated for which the considered selection criterion cannot be achieved ("*nonviable* IO sets"), while those candidates are kept for which it can be achieved ("*viable* IO sets"). Necessary *or* sufficient conditions for the existence of a controller achieving a particular criterion are often used. In both cases, the IO selection method may be ineffective: a necessary condition may lead to the faulty acceptance of nonviable IO sets, while a sufficient condition may lead to the faulty rejection of viable IO sets. Hence, effectiveness calls for conditions which are necessary *and* sufficient.

4. *Generally applicable*: An IO selection method should deal with a wide variety of control problems. For instance, a method is preferably suitable or easily generalized to handle classes of nonlinear systems. General applicability requires a set-up which can describe a wide variety of control problems. To a large extent, this is possible with the set-up of Fig. 1.

5. *Rigorous*: Viability should be addressed rigorously to cover a wide variety of issues that are important for control system design. For instance, an IO selection criterion based on robust stability is more rigorous than a criterion based on nominal stability. In general, a more rigorous criterion selects a smaller number of viable IO sets which may be manageable for more detailed further analysis.

6. *Quantitative*: An IO selection method preferably employs a quantitative criterion for IO set viability to clearly distinguish between the prospects of candidate IO sets. For instance, a qualitative criterion like state controllability only provides a "yes" or "no" answer to input set viability, while a quantitative controllability measure provides additional information on "how strongly" an input set affects the state.

7. *Controller independent*: An IO selection method should eliminate IO sets for which there does not exist any controller meeting the intended control goal. Usually, it is undesirable to impose restrictions on the controller design method, because this yields biased conclusions on IO set viability. On the other hand, if restrictions on the controller design method or the maximum controller order do play a role, a controller-dependent IO selection method may be advantageous. For efficiency reasons, IO selection should not involve complete controller design.

8. *Direct*: For the purpose of efficiency or if the list of candidates is infinite (as is often the case for flexible structures), it is desired that an IO selection method directly characterizes the viable IO sets, instead of performing a candidate-by-candidate test for a particular criterion. The latter, brute-force approach is indirect and not solvable in time polynomial in N_u and N_y (De Jager & Toker, 1998).

3. Description of IO selection methods

This section describes the key ideas of various IO selection methods and conditions. Most of the ideas stem from the process industry, but many of them can also be applied in other disciplines, like mechanical systems. Only those methods will be reviewed that are useful for fairly general problems, even though they might originally have been developed for a specific application. In this sense, the review is believed to be rather complete. For the sake of brevity, formulas and derivations are omitted if clarity is not endangered.

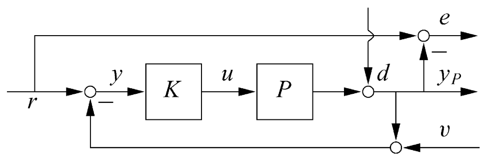


Fig. 2. One degree-of-freedom (DOF) control system set-up often assumed for IO selection; plant P, controller K, reference r, disturbance d, and sensor noise v.

To obtain a better overview of the many methods, they are divided into different groups. To aid the control engineer in picking a suitable method, the methods are grouped according to the selection criterion they address, like state controllability or the occurrence of right half-plane zeros. Section 3.8 is devoted to methods that can be used independently of the selection criteria. The order in which the different groups are discussed is according to ascending level of rigor.

Unless explicitly noted, all reviewed IO selection methods apply to (1) finite dimensional, (2) linear, (3) time-invariant, and (4) continuous-time plants and controllers. IO selection for other systems is largely unexplored. Besides these four limitations, many IO selection methods exhibit two additional ones and it will be clear from the text whether any of these play a role for the considered method. First, it is often assumed that $n_u = n_y$, leading to square controllers and IO sets. Second, the methods are often tied to the restrictive set-up of Fig. 2. If this set-up were transformed into the more general one of Fig. 1, the reference r, the disturbance d, and the sensor noise v would make up the entries of w, while z usually consists of $e = y_P - r$ and/or u. Fig. 2 is only useful if the outputs y (or y_P) and the controlled variables z are directly related and the control goals can be formulated in terms of y. This is the case if z can be measured directly ($z = y$) or if an explicit relationship is known between y and the immeasurable variables in z ($z = f(y)$). Direct control of the measurable variables in y may then be satisfactory. This is called inferential control and the measurements are referred to as secondary measurements. If y and z are not directly related, it is not always possible nor desired to transform the specifications for the controlled variables into equivalent specifications for the measured variables. A separate treatment of y and z, as in Fig. 1, would then be welcome.

3.1. Accessibility

Govind and Powers (1982) propose a (mainly) qualitative technique for IO selection based on cause-and-effect graphs. Such a graph shows the relationships between various variables and can be generated for linear and nonlinear systems. The key idea for IO selection is that a causal path must exist between the manipulated and the controlled variables on the one hand (see Fig. 3) and the measured and the controlled variables on the other hand: with the manipulated variables it must be possible to affect the controlled variables and with the measured variables it must be possible to obtain the values of the controlled variables, i.e., accessibility is required. A large number of candidate IO sets may be termed viable if they are only assessed for accessibility. So, additional criteria should be invoked. For input selection and a *linear* plant model, Govind and Powers (1982) use the steady-state gains, time constants, and time delays along candidate cause-and-effect paths as additional quantitative accessibility measures.

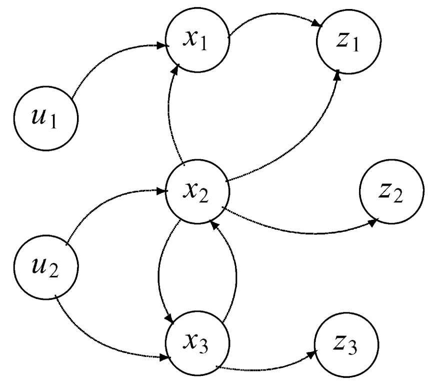


Fig. 3. Simple cause-and-effect graph.

The ideas by Daoutidis and Kravaris (1992) could be useful for input selection for *nonlinear* plants. They define the relative degree r_{ij} of a controlled variable z_i with respect to a manipulated variable u_j ($y = z$ is assumed). In that way, r_{ij} is a measure of the effect of each one input on each one output, such that it can also be used for CC selection. The generic computation of r_{ij} only requires structural information on the system. The relative degree is proposed as a measure of the dynamic interaction between manipulated and controlled variables or as a measure of the sluggishness of the response of the controlled variables to steps in the manipulated variables. Intuitively, the relative degree is related to the "physical closeness" of manipulated and controlled variables, or to the "direct effect" of manipulated variables on controlled variables. These heuristics are often used for IO selection. In a cause-and-effect graph, $r_{ij} + 1$ is the minimum number of edges connecting u_j to z_i, provided r_{ij} is finite; an infinite r_{ij} indicates nonaccessibility of u_j to z_i. In Fig. 3, $r_{12} = 1$ for the two edges $u_2 \rightarrow x_2$ and $x_2 \rightarrow z_1$, while $r_{21} = \infty$. The relative degree is thus related to the number of states in the shortest cause-and-effect path. This is consistent with the interpretation of the relative degree as the number of integrations the input has to go through before it affects the output.

The relative degree could form the basis for a quantitative accessibility measure for input selection: the lower

r_{ij}, the better the accessibility of u_j to z_i. Thus, compute

$$r_{zu} := \sum_{i=1}^{n_z} \min(r_{i1}, \dots, r_{in_u}) \tag{1}$$

for a given input set, then candidate input sets with small r_{zu} are preferred. Daoutidis and Kravaris (1992) propose the relative degree also as a tool for CC selection, since it measures dynamic interaction between (groups of) inputs and outputs. Soroush (1996) proposes output selection based on the relative degree of a measured variable with respect to an exogenous variable.

3.2. State controllability and state observability

This section deals with linear plant models P in the state-space description:

$$\dot{x} = Ax + Bu, \tag{2}$$

$$y = Cx + Du, \tag{3}$$

where n_u need not be equal to n_y. These equations represent the lower right part of G in Fig. 1. It is clear that potential IO selection methods rely on controllability and observability. These terms are broadly interpreted, as will become clear later. In this section *state* controllability and *state* observability are the focus.

State controllability. *System* (2) *or the pair* (A, B) *is called state controllable if, for any initial state* $x(0) = x_0$, *any time* $t_e > 0$, *and any final state* x_e *there exists an input* $u(t)$ *such that* $x(t_e) = x_e$.

State observability. *System* (2)–(3) *or the pair* (C, A) *is called state observable if, for any time* $t_e > 0$, *the initial state* $x(0) = x_0$ *can be determined from the time history of the input* $u(t)$ *and the output* $y(t)$ *in the interval* $[0, t_e]$.

These are "binary" concepts: either a plant exhibits the property or not. More rigorous quantitative controllability and observability measures could also be invoked. The binary and quantitative approaches are considered both.

3.2.1. Controllability and observability in a binary sense

A straightforward IO selection method would be to reject candidate IO sets for which (A, B) is uncontrollable or (C, A) is unobservable. Various simple tests are available (Zhou, Doyle, & Glover, 1996, Section 3.2). Nevertheless, such IO selection methods have not been encountered in the literature, probably due to the lack of rigor. Hovd and Skogestad (1992a) use stabilizability (i.e., controllability of the unstable modes) and detectability (i.e., observability of the unstable modes) for the selection of inputs and outputs, respectively.

Morari and Stephanopoulos (1980a) propose *structural* state controllability and observability as IO selection criteria. The plant is represented in a structural model which requires only information related to whether a variable is involved in a particular system equation or not. A structural version of (2)–(3) would have only two types of matrix entries: entries which are fixed at zero and entries which can take any numerical value, including zero. A numerical model also depends on the values of the parameters which may be uncertain (typically at the early stages of process design) or which may vary with the considered linearization point. Structural controllability and observability thus provide generic information about the system, but they are only necessary for controllability and observability in the numerical sense. Necessary and sufficient conditions for structural state controllability and observability consist of (1) accessibility conditions (the states should be accessible from the inputs and the outputs should be accessible from the states) and (2) rank conditions on the structural pairs (A, B) and (C, A), see Morari and Stephanopoulos (1980a). So, structural controllability and observability imply (but are not implied by) accessibility. Structural models are also used to describe *nonlinear* plants (Georgiou & Floudas, 1989; Morari & Stephanopoulos, 1980a). By "linearizing" a nonlinear model, a structural linear model can be obtained. For instance, $\dot{x}_1 = x_1 x_2$, $\dot{x}_2 = x_2^2$ would yield

$$A = \begin{bmatrix} \times & \times \\ 0 & \times \end{bmatrix}.$$

Structural controllability and observability thus provide prospects for IO selection for nonlinear systems.

In Lin, Tade, and Newell (1991) and Lin, Newell, Douglas, and Mallick (1994), the key idea for IO selection is structural *output* controllability. This means that each output in y can be influenced by at least one input in u independently (so, at least $n_u \geq n_y$). Structural output controllability and cause-and-effect are thus strongly linked. The main motivation to introduce structural output controllability is that, in practice, it may not be necessary to (structurally) control and observe all states. For meaningful IO selection, controlled variables z should be properly represented by any of the candidate output sets y. More on structural controllability and observability can be found in Georgiou and Floudas (1989) and Russell and Perkins (1987).

3.2.2. Controllability and observability in a quantitative sense

Due to the binary nature of (structural) state controllability and state observability, they are unlikely to provide IO selection conditions which are selective enough. The accepted number of IO sets may be too large for controller design and closed-loop evaluation. It should be possible to draw conclusions on the strength of the coupling between inputs, states, and outputs. Various quantitative measures for state controllability and observability have been proposed for this.

In the upcoming IO selection methods, the controllability Gramian $W_c(t)$ and the observability Gramian $W_o(t)$ play a role, see, e.g., Zhou et al. (1996, Chapter 3). If A in (2) is stable, $L_c := \lim_{t\to\infty} W_c(t)$ and $L_o := \lim_{t\to\infty} W_o(t)$ are the unique, positive semi-definite solutions to the following Lyapunov equations:

$$AL_c + L_c A^T + BB^T = 0, \tag{4}$$

$$A^T L_o + L_o A + C^T C = 0. \tag{5}$$

Georges (1995) selects optimal actuator and sensor locations (n_u and n_y are fixed) based on maximizing the minimum eigenvalues of $W_c(T)$ and $W_o(T)$, respectively, with T a given, possibly infinite, time. The idea is to minimize the input energy to reach a given state and to maximize the output energy generated by a given state. Georges (1995) extends this idea to nonlinear systems $\dot{x} = f(x) + Bu, y = Cx$. IO selection then requires the solution of nonlinear, partial differential equations.

Müller and Weber (1972) embed the minimum eigenvalues of $W_c(t)$ and $W_o(t)$ in an *infinite set* of quantitative controllability and observability measures. Among those are the determinants of $W_c(t)$ and $W_o(t)$ and the reciprocals of the traces of $W_c^{-1}(t)$ and $W_o^{-1}(t)$. IO selection could aim at maximizing any of these measures.

Hać and Liu (1993) and Ko, Tongue, and Packard (1994) define performance indices that involve the eigenvalues of L_c and L_o. The IO selection criteria proposed by Hać and Liu (1993) are based on the energy of the input and the output in case of transient responses or persistent disturbances. The criteria are especially useful for flexible structures. They provide a balance between the importance of the lower- and the higher-order modes. For systems with small damping ratios and well-separated natural frequencies, the locations of force actuators and velocity sensors tend to be collocated (also for many other criteria). The selection criterion for point actuators in Hać and Liu (1993) is extended to distributed actuators in Hać (1995).

Assume that the plant is controllable, observable, and stable. A *balanced realization* of (2)–(3) can then be obtained, with L_c and L_o equal and diagonal (Moore, 1981). The diagonal entries of these Gramians are called the Hankel singular values (HSVs), reflecting the *joint* controllability and observability of the balanced states. Candidate IO sets with large HSVs are preferred in Samar and Postlethwaite (1994). HSVs also form the basis for IO selection in Gawronski and Lim (1996) and Lim (1997), but they employ special properties of flexible structures. As in Fig. 1, Lim (1997) explicitly defines exogenous and controlled variables and performance weightings to improve the IO selection criterion of Gawronski and Lim (1996).

Another quantitative controllability measure could be the size of a region of initial states from which the state can return to the origin in a fixed time t_e with an admissible input signal. Schmitendorf (1984) terms u admissible if it obeys a magnitude bound, while Vander Velde and Carignan (1984) call u admissible if it obeys an energy bound. Both measures may be helpful for actuator selection. Vander Velde and Carignan (1984) also discuss a dual approach to sensor selection and they address the effect of possible actuator and sensor failures from a statistic viewpoint.

3.3. Right half-plane zeros

Distinct IO sets may give rise to distinct numbers and distinct locations of (multivariable) system zeros, as defined in, e.g., Zhou et al. (1996, Section 3.11). Considering plant (2)–(3), the zeros can be interpreted as those values of s where the rank of the corresponding transfer function matrix (TFM) $P(s) := C(sI - A)^{-1}B + D$ from u to y is smaller than its normal rank, i.e., its maximally possible rank for at least one value of s. In general, the zeros of $P(s)$ have no direct relation with the zeros of the individual elements of $P_{ij}(s)$.

Zeros in the right half-plane (RHP) limit the closed-loop performance. For instance, for SISO systems as in Fig. 2 the performance specification is often in terms of a magnitude bound on the sensitivity defined as

$$S(s) := (I + P(s)K(s))^{-1}. \tag{6}$$

In case of tracking (r) or disturbance (d) rejection, $|S(j\omega)|$ is required small for low frequencies. The bandwidth is desirably large, but RHP zeros impose an upper bound. For a *stable* plant with a single RHP zero, this bound is smaller if the RHP zero is closer to the origin (Skogestad & Postlethwaite, 1996, Section 5.6). For a stable MIMO plant, the implications of RHP zeros are similar to those for stable SISO plants (Skogestad & Postlethwaite, 1996, Section 6.5). For an *unstable* plant, the restrictions due to RHP zeros are even more severe. Performance limitations due to RHP zeros are extensively discussed by, e.g., Freudenberg and Looze (1985, SISO), Havre and Skogestad (1996, MIMO), Sidi (1997, SISO), Skogestad and Postlethwaite (1996, Chapters 5 (SISO) and 6 (MIMO)), and Zhou et al. (1996, Chapter 6 (MIMO)). If magnitude bounds are imposed on u and y, Zafiriou and Chiou (1996) show that RHP zeros of the *individual elements* of $P(s)$ may be detrimental, in contrast to the case without these requirements. Goodwin and Seron (1995) initiate efforts towards quantifying performance limitations due to "RHP zeros" in nonlinear systems. Though zeros in the left half-plane do not impose fundamental limitations on control, they may cause practical problems if they are close to the origin (Skogestad & Postlethwaite, 1996, Section 5.6).

A guideline for IO selection in the set-up of Fig. 2 is to reject IO sets which introduce RHP zeros with magnitudes below the desired bandwidth. This is employed by, e.g., Biss and Perkins (1993), Hovd and Skogestad (1993), Samar and Postlethwaite (1994), and Wolff, Skogestad, Hovd, and Mathisen (1992). For unstable plants, also IO sets should be avoided with RHP zeros close to RHP poles. Exact cancellation of RHP poles and zeros causes the unstable modes to become uncontrollable and/or unobservable and so a stabilizing controller does not exist. Lee and Speyer (1993) show that a sensible placement of single actuators and sensors along a flexible beam is important to avoid cancellation of zeros and poles in the origin.

3.4. Input–output controllability

A large amount of literature on IO selection is devoted to quantitative measures for IO controllability. The treatment is usually restricted to control problems in the set-up of Fig. 4, where $P_d(s)$ models the disturbance at the output of the plant and $\Delta(s)$ models the uncertainty (multiplicative input uncertainty is depicted, but sometimes multiplicative output uncertainty is considered). Unless noted otherwise, it is assumed in this section that any candidate output set y properly represents the control objectives. The concept of IO controllability is, crudely:

Input–output controllability. *The plant in Fig.* 4 *is called IO controllable if acceptable performance can be achieved*, *i.e.*, *if the outputs* y *and the inputs* u *can be kept acceptably small*, *in the presence of bounded uncertainties* Δ, *references* r, *disturbances* d, *and sensor noises* v.

The difference with the definition by Skogestad and Postlethwaite (1996, Section 5.1) is that uncertainties are taken into account. In contrast to state controllability, IO controllability tries to capture aspects which are relevant *in practice*. State controllability requires transition from one state to any other within a finite time interval, which may be irrelevant. Still, state controllability does not imply IO controllability, since it does not address the system behavior during this time interval. The IO controllability measures in this section usually capture only one aspect of IO controllability. Combining measures is thus required for rigorous IO selection.

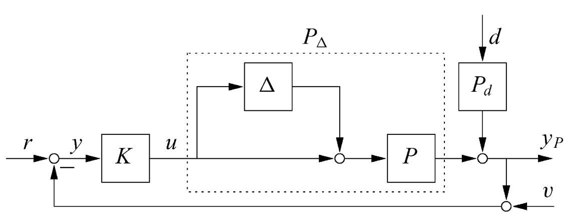


Fig. 4. Control system set-up for assessing IO controllability.

Morari (1983) states that the performance is limited due to nonminimum-phase elements (RHP zeros and time delays), restrictions on the inputs u, and model uncertainties. Most IO controllability measures are based on input restrictions and uncertainties which are usually not addressed simultaneously. All measures are simple and give a rough idea of how easy the plant is to control, irrespective of the controller. It could be argued that the controller-independent IO selection methods and conditions in other sections are also based on controllability measures, since controllability is a property of the plant, the control goals, and the IO set alone.

Most IO controllability measures assume a suitable *scaling* of the involved variables, because the results critically depend on it. A proper scaling expresses the relative significance of the variables that determine the performance. One way to scale the variables is to divide them by their allowed (u and y) or expected (r, d, and v) magnitudes, see Skogestad and Postlethwaite (1996, Section 1.4). Some IO controllability measures are scaling-independent. This is often seen as an advantage, but this contradicts the importance of a suitable scaling for the reason mentioned above. It does not matter for a scaling-independent measure whether the plant is badly or well scaled: the results are the same and so the performance specifications are not well addressed. We therefore believe that for IO selection it is desired to have scaling-dependent measures and to apply them to a plant for which the candidate inputs and outputs have been suitably scaled a priori.

Different groups of controllability measures will be considered. The foundation is often laid by the singular value decomposition (SVD). A complex $l \times m$ matrix F can be factorized by an SVD as follows:

$$F = Y \Sigma U^*, \tag{7}$$

where $\{\cdot\}^*$ denotes the complex conjugate transpose. The $l \times l$ matrix Y and the $m \times m$ matrix U are unitary, i.e., $Y^* = Y^{-1}$ and $U^* = U^{-1}$. The matrices Y and U form orthonormal bases for the column (output) space of F and the row (input) space of F, respectively. The columns Y_i of Y will be called the "left (output) singular vectors", while the columns U_i of U will be called the "right (input) singular vectors". The $l \times m$ matrix Σ contains a diagonal matrix Σ_1 of real nonnegative *singular values* σ_i, arranged in descending order:

$$\Sigma = \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} \quad \text{if } l \geq m \quad \text{or} \quad \Sigma = [\Sigma_1 \ 0] \quad \text{if } l \leq m, \tag{8}$$

where

$$\Sigma_1 = \mathrm{diag}(\bar{\sigma} := \sigma_1, \sigma_2, \ldots, \underline{\sigma} := \sigma_k) \quad \text{with } k = \min(l, m). \tag{9}$$

The σ_i are the square roots of the k largest eigenvalues λ of F^*F and FF^*:

$$\sigma_i(F) = \sqrt{\lambda_i(F^*F)} = \sqrt{\lambda_i(FF^*)}, \quad i = 1, \ldots, k. \tag{10}$$

Based on the magnitudes of the singular values, the directions of Y_1 and U_1 are often referred to as the "most important directions", while the directions of Y_2 and U_2 are often called the "second most important directions", and so on. Provided $\underline{\sigma}(F) \neq 0$, the Euclidean *condition number* of the matrix F is

$$\kappa(F) := \frac{\bar{\sigma}(F)}{\underline{\sigma}(F)}. \tag{11}$$

Finally, the pseudo-inverse (or Moore–Penrose generalized inverse) of a nonsquare, complex matrix F can be written as

$$F^{\dagger} = \sum_{i=1}^{r} \frac{1}{\sigma_i(F)} U_i Y_i^* \quad \text{with } r = \text{rank}(F). \tag{12}$$

3.4.1. The minimum singular value

A first IO controllability measure that is often used for IO selection is the frequency-dependent minimum singular value of the plant: $\underline{\sigma}(P(j\omega)) \geq 0$ (usually, "$(j\omega)$" will be omitted). In Tzouanas, Luyben, Georgakis, and Ungar (1990), $\underline{\sigma}(P)$ is suggested for *on-line* selection of the input set, to adapt to changing operating conditions. The general rule is that IO sets should be selected which yield a large $\underline{\sigma}(P)$. Three reasons for this are given below.

The *first* one is related to independent control of all the outputs y (Skogestad & Postlethwaite, 1996, Section 6.3). Apart from the prerequisite $n_u \geq n_y$, the IO set should guarantee $\underline{\sigma}(P) > 0$ (except at possible zeros on the imaginary axis), which implies the rank of P to be equal to n_y. If $\underline{\sigma}(P) \approx 0$, independent control is probably difficult to achieve.

The *second* reason is related to input constraints. Morari (1983) argues that for a plant to have good tracking (r) and disturbance rejection (d) in case of input magnitude limitations, $\underline{\sigma}(P)$ should be large. Yu and Luyben (1986) call $\underline{\sigma}(P)$ the "Morari Resilience Index" (MRI). Dynamic resilience is synonymous to IO controllability. Yu and Luyben (1986) suggest to select the input set with the largest MRI over the frequency range of interest, which is adopted by Havre, Morud, and Skogestad (1996) and Wolff et al. (1992). This quantifies the rule "choose inputs which have a large effect on the output", see Morari (1983) and Seborg, Edgar, and Mellichamp (1989, Chapter 28).

The *third* reason is from Havre et al. (1996) and Skogestad and Postlethwaite (1996, Section 10.3). Assume that the plant in Fig. 4 has additional manipulated inputs u_0 that are not used for controlling y and additional outputs y_0 that are not fed back, but that are important for performance. The reference r (assumed constant) and the input u_0 are generated by a controller at a higher level. For a given disturbance d, assume that there exist u_0 and u that optimize the performance. The inputs will not be equal to the optimal values in practice, due to imperfect control, model uncertainties, disturbances, and sensor noise. An approach to output selection is thus to select outputs y that keep u_0 and u close to the optimal values. Havre et al. (1996) and Skogestad and Postlethwaite (1996, Section 10.3) show that this corresponds with large $\underline{\sigma}(P)$, at least for low frequencies where feedback is effective.

3.4.2. The maximum singular value

Contrary to the other IO controllability measures, the maximum singular value $\bar{\sigma}$ usually applies to different TFMs than $P(s)$. Depending on the objective, $\bar{\sigma}$ may either be required small or large.

Havre et al. (1996) and Skogestad and Postlethwaite (1996, Section 10.3) select secondary measurements for inferential control. The plant P in Fig. 4 has additional outputs z which are not fed back. The outputs z should be kept at a given reference value r_z which is transformed into a reference value r_y for y. Assuming close-to-perfect control of y via $u = P^{-1}(r_y + v - P_d d)$, the closed-loop control error $r_z - z$ is given by $r_z - z = M_d d + M_v v$, with M_d and M_v depending on open-loop TFMs relating d, z, u, and y. For a nonsquare IO set, P^{-1} is replaced by the pseudo-inverse $P^{\dagger}$. In case of a square IO set and no sensor noise ($v = 0$), the control law would yield perfect control and $M_d = 0$ if $y = z$. The rule for selecting secondary measurements y is to keep $\bar{\sigma}(M_d)$ and $\bar{\sigma}(M_v)$ small in the frequency ranges of interest. This may involve a trade-off which could be addressed by combining M_d and M_v and selecting y such that $\bar{\sigma}([M_d \;\; M_v])$ is small. A related, less general method is discussed by Bequette and Edgar (1986). They consider square IO sets and the regulator problem ($z_r = 0, y_r = 0$) without sensor noise ($v = 0$). Under perfect steady-state control of y, output sets with small $\bar{\sigma}(M_d(0))$ are preferred. This minimization of the sensitivity of the regulated variables z to the disturbances d is then balanced against the maximization of the effect of the inputs u on the outputs y. The latter involves $\bar{\sigma}(P)$: output sets with large $\bar{\sigma}(P)$ are preferred.

The input selection method of Cao and Biss (1996) is one of the few *direct* methods. With N_y fixed outputs and N_u candidate inputs, the plant for the full input set, $P = [P_1, \ldots, P_{N_u}]$, is constructed. Cao and Biss (1996) suggest to retain those n_u inputs ($N_y \leq n_u \leq N_u$) with the largest "single-input gain":

$$\sigma_{u_j}(P) := \sqrt{P_j^* P_j} = \sigma(P_j) = \bar{\sigma}(P_j) \tag{13}$$

at the relevant frequency (note that there is only one singular value for the column P_j). This approach may not be effective, since the *combination* of inputs is not assessed. An input with all entries in P_j equal to zero, except for a single large entry, could have a large $\sigma_{u_j}(P)$, whereas an input with all entries in P_j small could have a small $\sigma_{u_j}(P)$. So, an input set may be selected which is unable to affect all outputs. The related "single-input disturbance gain" $\sigma_{d,u_j}(P, P_d)$ is defined in Cao (1995, Section 6.2). Consider Fig. 4 with $\Delta = 0$, $r = 0$, and $v = 0$. For perfect disturbance rejection ($y = 0$), $u = -P^{\dagger}P_d d$ must be applied. With P corresponding to the full input set, input u_j, and $(P^{\dagger}P_d)_j$ the jth *row* of $P^{\dagger}P_d$:

$$\sigma_{d,u_j}(P, P_d) := \sqrt{(P^{\dagger}P_d)_j (P^{\dagger}P_d)_j^*} = \sigma((P^{\dagger}P_d)_j)$$
$$= \bar{\sigma}((P^{\dagger}P_d)_j). \qquad (14)$$

Cao (1995) suggests to retain those inputs for which $\sigma_{d,u_j}(P, P_d)$ is *large*, since these would be important for disturbance rejection. This heuristic rule may not be effective: P in (14) corresponds to the full input set, so $\sigma_{d,u_j}(P, P_d)$ may state something about the effectiveness of u_j as part of the full input set, but not about its effectiveness in a smaller input set. For effective yet indirect input selection, $\sigma_{d,u_j}(P, P_d)$ should be recomputed for distinct input sets and plants P. One may wonder why Cao (1995) does not suggest to retain those inputs for which $\sigma_{d,u_j}(P, P_d)$ is *small*, since these inputs would require the least control effort for rejection of a given set of disturbances. This rule would be in line with the use of the MRI $\underline{\sigma}(P)$ when $P_d = I$ (note the fact $\bar{\sigma}(F^{-1}) = 1/\underline{\sigma}(F)$ and the "inverse" in $\sigma_{d,u_j}(P, P_d)$).

3.4.3. The condition number

The frequency-dependent condition number is another common IO controllability measure. In general, an IO set should be chosen which results in a small condition number. The condition number of a column or row vector (which have only one singular value) equals one and so it is not selective if $n_u = 1$ or $n_y = 1$.

Morari (1983) shows that a small condition number $\kappa(P)$ corresponds to good robustness against full-block (unstructured) multiplicative uncertainty. This would justify selecting IO sets giving rise to a small $\kappa(P)$. Skogestad and Postlethwaite (1996, Section 6.10) state that the performance of plants (IO sets) with a small $\kappa(P)$ is robust to both diagonal and unstructured multiplicative uncertainty Δ at the input of the plant. Section 3.6 sophisticates the use of $\kappa(P)$ to address robustness.

Reeves (1991, Section 5.2) proposes a method to reduce the number of candidate inputs and outputs prior to applying the computationally more involved method in Section 3.6. Starting with the full $N_u \times N_y$ IO set, *that* single input or single output is eliminated which produces the reduced $(N_u - 1) \times N_y$ or $N_u \times (N_y - 1)$ IO set with the smallest $\kappa(P)$, computed at a relevant frequency. In this way, the size of the IO set is reduced gradually, until it is manageable for other techniques. The final IO set is not guaranteed to be the IO set of that dimension for which $\kappa(P)$ is smallest. For that, $\kappa(P)$ should have been computed for *all* IO sets of that dimension.

Skogestad and Morari (1987a) consider the *disturbance* condition number. For the kth disturbance d_k in Fig. 4, it is

$$\kappa_{d_k}(P, P_d) := \frac{\|P^{-1}P_{d_k}\|_2}{\|P_{d_k}\|_2}\bar{\sigma}(P)$$
$$\text{with } 1 \le \kappa_{d_k}(P, P_d) \le \kappa(P). \qquad (15)$$

Here, P_{d_k} is the kth column of P_d corresponding to d_k and $\|\cdot\|_2$ denotes the Euclidean vector norm (the length) at a given frequency. For nonsquare P, P^{-1} should be replaced by $P^{\dagger}$. The disturbance condition number is a measure of the input magnitude which is needed to reject a disturbance in the direction P_{d_k}, relative to rejecting a disturbance with the same magnitude, but in the direction requiring the least control effort. Input sets yielding a small $\kappa_{d_k}(P, P_d)$ are most effective for disturbance rejection. The essential difference between input selection with $\underline{\sigma}(P)$ and with $\kappa_{d_k}(P, P_d)$ is that $\underline{\sigma}(P)$ aims at rejecting disturbances with large magnitudes by using a given input magnitude, whereas $\kappa_{d_k}(P, P_d)$ aims at small input magnitudes for a given disturbance *direction*, irrespective of the disturbance's magnitude. The definition of the "input disturbance alignment" in Cao and Rossiter (1996) shows close resemblance with (15):

$$\eta_{d_k}(P, P_d) := \frac{\|PP^{\dagger}P_{d_k}\|_2}{\|P_{d_k}\|_2} \quad \text{with } 0 \le \eta_{d_k}(P, P_d) \le 1. \qquad (16)$$

Input sets with $\eta_{d_k}(P, P_d)$ close to one are preferred. The projection norm of P_{d_k} on the column space of P is then large, which serves good disturbance rejection.

3.4.4. Singular vectors

The next group of IO controllability measures involves the left (Y) or right (U) singular vectors from the SVD in (7).

Moore, Hackney, and Carter (1987) suggest three output selection rules. The key idea is to find the best compromise between measurements y that are mutually independent and measurements that are sensitive to changes of the inputs u. For this purpose, the SVD of P at steady state is computed. Square IO sets are considered. The *first* rule relies on the SVD of $P(0)$ for the *full output set*. Those outputs are selected that correspond to the entries of each left singular vector in Y with the largest absolute values (note that outputs cannot be distinguished if the same entry occurs multiple times). This is based on the notion that the left singular vectors point into the direction of the first (Y_1), the second (Y_2), etc., most sensitive combination of outputs. The selected outputs are stated to be sensitive to the inputs and, due to orthogonality of the vectors in Y, relatively independent. A similar procedure could be proposed for input selection, involving the largest absolute values of U. This output selection method is *direct*, since the SVD is not recomputed for each candidate. For the same reason, however, the method may be ineffective. The *second* rule is a modified version of the first one. Suppose that the entry in Y_2 corresponding to the entry of the largest absolute value of Y_1 is large and vice versa. Then there would still be significant interaction between the outputs selected by the first rule. To arrive at an output set with reduced interaction, output selection can be based on the differences between the absolute-valued entries of the left singular vectors, possibly at the price of reduced sensitivity to u. The *third* rule resolves the possible ineffectiveness of the previous ones by recomputing the SVD for all candidate output sets. Moore et al. (1987) state that a large value of $\underline{\sigma}(P(0))$ indicates a good sensitivity to inputs and that a small value of $\kappa(P(0))$ indicates a good mutual independence of the outputs. An output set for which the measure:

$$\varrho(P(0)) := \frac{\underline{\sigma}(P(0))}{\kappa(P(0))} \tag{17}$$

is large should exhibit a good compromise between sensitivity to inputs and mutual independence of outputs.

In Keller and Bonvin (1987), an input selection method is proposed that aims at selecting the n_u inputs with "the strongest and most orthogonal effect" on the (controlled) outputs y. This is quantified via the n_u largest singular values and the corresponding right singular vectors of the input matrix B in the plant's state-space description in a scaled modal basis. The input selection method is *direct* if B is generated for the full input set.

In Cao and Biss (1996) a *direct* approach is suggested to select the $n_u \geq N_y$ inputs from the N_u candidates with the largest effect on the fixed number of N_y outputs y. The SVD of P for the *full input set* is computed at a relevant frequency. Assuming that P has full row rank N_y, the "single-input effectiveness" for u_j is given by

$$v_{u_j}(P) = \sqrt{\sum_{i=1}^{N_y} U_{ji}^* U_{ji}}. \tag{18}$$

The n_u inputs that yield the largest values $v_{u_j}(P)$ should be selected. It can be expected, but not guaranteed, that this input set gives the largest ratio $\|u_1\|_2/\|u\|_2$, with u_0 the part of $u = u_1 + u_0$ that is in the null space of P, i.e., $Pu_0 = 0$. In Cao (1995, Section 6.3), the SVD of $P^\dagger P_d$ is used to define the "single-input disturbance effectiveness", as a similar measure for the disturbance rejection effectiveness.

3.4.5. The relative gain array

The final IO controllability measure that is often employed for IO selection is the frequency-dependent relative gain array (RGA) for a nonsingular, square matrix:

$$\Lambda(P(j\omega)) := P(j\omega) .* (P^{-1}(j\omega))^T, \tag{19}$$

with ".*" denoting element-by-element multiplication (Hadamard or Schur product). For a nonsquare P, the inverse in (19) is again replaced by the pseudo-inverse (Chang & Yu, 1990). The RGA is independent of the scaling of the plant's inputs and outputs if $n_u = n_y$, while it is independent of output scaling if $n_u > n_y$ and independent of input scaling if $n_u < n_y$. The RGA was introduced by Bristol (1966) as a measure for interactions in decentralized control systems. As a result, there is a bulk of literature proposing RGA-related conditions for CC selection, but the RGA is also useful for IO selection. A frequently encountered rule is that IO sets causing large RGA elements should be avoided, since the corresponding plants would be difficult to control, see, e.g., Chen, Freudenberg, and Nett (1994).

Hovd and Skogestad (1992b) and Skogestad and Morari (1987b) show that plants with large absolute-valued RGA elements (1) are very sensitive (especially around the bandwidth) to diagonal multiplicative input uncertainty if an inverse-based controller "$K = P^{-1}$" is used and (2) are very sensitive to element-by-element uncertainty in P. In practice, it is often plausible to assume the first type of uncertainty, e.g., in case of neglected actuator dynamics. However, it is disputable to assume the second type, because the individual elements of $P(s)$ are usually coupled in some way. Motivated by (1), one step of the IO selection procedure in Samar and Postlethwaite (1994) is based on the magnitude of RGA elements.

Reeves (1991, Section 5.2) proposes two RGA-based heuristics to reduce the $N \times N$ full IO set ($N = N_u = N_y$) to a smaller, $n \times n$ IO set ($n = n_u = n_y$). These heuristics may not produce the IO set that is optimal with respect to the intended objective. The *first* heuristic employs the following lower bound to the condition number at a given frequency (Nett & Manousiouthakis, 1987):

$$\kappa_\Lambda(P) := 2\max(\|\Lambda(P)\|_{i1}, \|\Lambda(P)\|_{i\infty}) - 1 \leq \kappa(P) \tag{20}$$

with $\|\Lambda(P)\|_{i1} := \max_j \sum_i |\Lambda_{ij}|$ the maximum absolute-valued row sum and with $\|\Lambda(P)\|_{i\infty} := \max_i \sum_j |\Lambda_{ij}|$ the maximum absolute-valued column sum (the subscript "i" denotes the *induced* matrix norm). Actually, $\kappa_\Lambda(P)$ is a lower bound to the *minimum condition number* $\tilde{\kappa}(P)$ under the best-possible IO scalings:

$$\kappa_\Lambda(P) \leq \tilde{\kappa}(P) := \inf_{D_1, D_2} \kappa(D_1 P D_2) \leq \kappa(P). \tag{21}$$

D_1 and D_2 are diagonal matrices with real, positive entries. Computation of $\tilde{\kappa}(P)$ involves a convex optimization and so is rather easy, but computation of $\kappa_\Lambda(P)$ is easier. Based on the fact that $\kappa_\Lambda(P) \leq \kappa(P)$ and the notion that IO sets yielding a small plant condition number are preferred, Reeves (1991) suggests to discard *that* single input and output from the $N \times N$ IO set for which the $(N-1) \times (N-1)$ reduced IO set exhibits the smallest $\kappa_\Lambda(P)$. The *second* heuristic relies on the aforementioned notion that a large RGA element Λ_{ij} only permits a small uncertainty in the corresponding plant element P_{ij}. Reeves (1991) proposes to identify the row and column indices of the largest RGA element and to discard the corresponding input and output from the $N \times N$ IO set. For the $(N-1) \times (N-1)$ IO sets obtained with either of these heuristics, the RGA is recomputed and additional inputs and outputs may be eliminated. This procedure is repeated until the IO set with the intended dimension is reached.

As shown in Cao and Biss (1996), Chang and Yu (1990), and Skogestad and Postlethwaite (1996, Section 10.5), to some extent the RGA may be used for a *direct* approach to IO selection. Consider P corresponding to the full IO set and suppose $N_u \neq N_y$. The RGA for the nonsquare P could then be used for squaring down the plant to have dimension $\min(N_u, N_y) \times \min(N_u, N_y)$. If $N_u > N_y$, it is suggested to eliminate the inputs u_j corresponding to an RGA *column* sum $\sum_{i=1}^{N_y} \Lambda_{ij} \leq 1$ that is much smaller than one. The elements in each row and column of the RGA sum up to one. If $N_y > N_u$, it is suggested to eliminate the outputs y_i corresponding to an RGA *row* sum $\sum_{j=1}^{N_u} \Lambda_{ij} \leq 1$ that is much smaller than one. In this way, the squared IO set consists of the inputs with the largest effect on the outputs and of the outputs which can best be affected by the inputs. In Cao (1995, Section 6.4), the RGA-like matrix $\Lambda_d := (P^\dagger P_d) .* ((P^\dagger P_d)^\dagger)^{\mathrm{T}}$ is used in a similar way to reject inputs which are less effective for disturbance rejection.

3.5. Efficiency of manipulation and estimation

The objective of actuators is to manipulate the system such that it behaves as desired. This should be achieved with limited energy. A reasonable approach to actuator selection is thus the minimization of an input-set-dependent cost function J_u in terms of the input energy. This "efficiency of manipulation" is the topic of Section 3.5.1. The objective of sensors is to gain the best-possible information on the system's behavior. Hence, sensor selection could be based on the minimization of an output-set-dependent cost function J_y involving the estimation errors of relevant variables, like states. This "efficiency of estimation" is treated in Section 3.5.2. In Section 3.5.3, the cost functions J_{uy} combine efficiency of manipulation and estimation. All methods assume the plant to be described by (2)–(3), possibly extended with process disturbances w_x in (2) and sensor noise w_y in (3).

3.5.1. Efficiency of manipulation

In Al-Sulaiman and Zaman (1994), the cost function J_u takes essentially the same form as in the well-known linear quadratic regulator (LQR) problem:

$$J_u = \int_0^{t_e} (x(t)^{\mathrm{T}} Q x(t) + u(t)^{\mathrm{T}} R u(t))\,\mathrm{d}t. \tag{22}$$

$Q = Q^{\mathrm{T}} \geq 0$ and $R = R^{\mathrm{T}} > 0$ are weights. The input set yielding the smallest J_u is the most appropriate one. Al-Sulaiman and Zaman (1994) only evaluate J_u *after* a state feedback has been designed by pole placement and a closed-loop simulation for a disturbance has been run over the time interval $[0, t_e]$. The input selection thus depends on the choice of the disturbance signal. Xu, Warnitchai, and Igusa (1994) use a similar cost function with $t_e = \infty$. They look for the IO set yielding the smallest J_u under noise-free static output feedback, by solving a nonlinear programming problem.

For *nonlinear* systems, Cao, Biss, and Perkins (1996) consider selection of inputs with magnitude constraints. The cost function J_u takes the form

$$J_u = \int_0^{t_e} (z(t) - z_r)^{\mathrm{T}} Q (z(t) - z_r)\,\mathrm{d}t \tag{23}$$

with z_r a specified setpoint for the controlled variables z and with $Q > 0$ a diagonal weighting matrix. For a given input set, J_u is minimized as a function of the input signal $u(t)$ and the final time t_e, subject to the constraints $u_l \leq u(t) \leq u_u$ and subject to the nonlinear system behavior $f(\dot{x}, x, z, u, t) = 0$ and a given initial and final system state. The input set that yields the smallest J_u is preferred. Input selection based on this optimization requires large computational effort (Cao, 1995, Section 8.2) and the results may depend on the choices of z_r and the initial conditions.

3.5.2. Efficiency of estimation

Morari and Stephanopoulos (1980b) select (secondary) measurements aimed at minimizing the errors in the estimates of relevant variables, like the controlled variables z. This aim is transformed into the minimization of cost functions J_y. The error sources are model uncertainties, process disturbances, and sensor noise. These are treated from a stochastic point of view by introducing colored process noise w_x and white sensor noise w_y. Four output selection criteria are proposed. These are all based on a *static estimator*, derived for the steady-state model. The first aims at minimizing the static estimation error, the second aims at minimizing the effect of model uncertainties on the estimates, and the third and fourth aim at minimizing the estimation errors if the static estimator is used for the dynamic system. The first criterion is essentially a multivariable extension of the "projection error criterion" derived by Joseph and Brosilow (1978) who estimate single variables. According to this criterion, the results tend to improve if measurements are added. The second criterion is essentially the same as the "condition number criterion" of Weber and Brosilow (1972), involving the open-loop steady-state TFM from the disturbances w_x to the measurements y. According to this criterion, the results tend to degrade if n_y increases, which is quite surprising. Hence, a trade-off must be made when the projection error criterion and condition number criterion are used concurrently for output selection. The counter-intuitive result for the condition number criterion can (partially) be attributed to an uncertainty model that is physically inconsistent for the candidate output sets, see also Lee (1991, Section 3.5).

The output selection method proposed by Kumar and Seinfeld (1978) aims at minimizing state-estimation errors using a *dynamic* estimator (Kalman filter) instead of a static one. Model uncertainties, process disturbances, and sensor noise are again approached from a stochastic point of view by adding white noise to (2) and (3). The cost function J_y to be minimized as a function of the output set involves the trace of the estimation error covariance matrix $\Theta(t)$:

$$J_y = \alpha\,\mathrm{trace}(\Theta(t_e)) + \beta\int_0^{t_e} \mathrm{trace}(\Theta(t))\,\mathrm{d}t,$$
$$\Theta := \mathrm{E}((x-\hat{x})^{\mathrm{T}}(x-\hat{x})) \tag{24}$$

with $\alpha > 0$ and $\beta > 0$ weighting parameters. The matrix $\Theta(t)$ is obtained by solving a differential Riccati equation. For a fixed number of sensors, an iterative algorithm identifies the optimal ones among all candidates.

Rhodes and Morari (1995) make an attempt towards output selection for a *nonlinear* autonomous plant $\dot{x} = f(x)$, $y = g(x)$. The basic idea is not minimization of estimation errors, but minimization of "modeling errors". The aim is to designate the smallest number of (secondary) measurements y that enables an accurate recreation of the nonlinear system dynamics. This only involves sampled system data $y(k)$. After output selection, the outputs that are needed to accurately describe the system are known, but a model still has to be derived. As in Kammer (1996), the derived output sets may be well suited for modeling, but not for control where the outputs may have different tasks. This is also recognized by Roh and Park (1997), who consider actuator selection for control and for identification, using different criteria depending on the task. In Kammer (1996), the criterion for selecting sensors for modal identification of flexible structures is maximization of the determinant of the observability Gramian, which is also among the criteria proposed by Müller and Weber (1972) for selecting control sensors.

3.5.3. Joint efficiency of manipulation and estimation

In Norris and Skelton (1989), IO selection is based on a similar cost function as in linear quadratic Gaussian (LQG) control:

$$J_{uy} = E\left(\int_0^{\infty} (z^{\mathrm{T}}(t)Qz(t) + u^{\mathrm{T}}(t)Ru(t))\,\mathrm{d}t\right) \tag{25}$$

with $z = Fx$ the variables to be kept small. Both the actuators and the sensors may have dynamics like (2)–(3). These dynamics may in turn be disturbed with additive, possibly correlated, white noise. One approach to IO selection would be to recompute J_{uy} (i.e., to recompute the estimator and feedback gains) for each IO set and to identify the IO set(s) yielding the smallest J_{uy}. However, to reduce computational effort, Norris and Skelton (1989) compute J_{uy} only for the full IO set. Under retention of the corresponding estimator and feedback gains, the effectiveness of each actuator and sensor is expressed as the change in J_{uy} if an actuator or sensor is eliminated (the larger the change, the more effective). These effectiveness measures are then used for IO selection. However, the definitions of the measures are disputable: they may not be physically meaningful and they are not consistent for actuators on the one hand and sensors on the other.

In Mellefont and Sargent (1977), minimization of a slightly adapted version of the LQG cost function in (25) forms the basis for *on-line* switching between output sets. The switching policy is determined off-line. By switching, the number of employed measurements is made as small as possible, to reduce the costs (computer time) of processing measurements.

3.6. Combined robust stability and nominal performance

This section considers combined robust stability (RS) and nominal performance (NP) as the IO selection criterion. RS guarantees stability in the presence of uncertainties, whereas NP guarantees stability and performance in the absence of uncertainties. Section 3.6.1 summarizes an IO selection method which is restricted to the one DOF control system set-up in Fig. 5, while Section 3.6.2 describes methods for the set-up in Fig. 1. Combined RS and NP does not imply the more rigorous property of robust performance (RP), i.e., guaranteed stability and performance in the presence of uncertainties, though the converse holds.

3.6.1. One DOF control system set-up

Reeves (1991) discusses a method to reject those IO sets for which there does not exist a controller achieving joint RS and NP. The original idea is due to Nett (1989). The method is also summarized by Banerjee and Arkun (1995) and implemented in a MATLAB toolbox (Reeves,

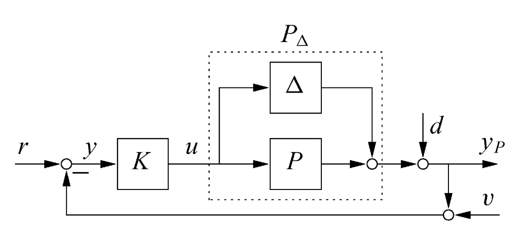


Fig. 5. One DOF control system set-up with unstructured, additive uncertainty.

Nett, & Arkun, 1991). The starting point is RS for the control system in Fig. 5, where P and K are *square*. The unstructured, additive uncertainty Δ obeys the relative, additive uncertainty bound:

$$\frac{\bar{\sigma}(\Delta(\mathrm{j}\omega))}{\bar{\sigma}(P(\mathrm{j}\omega))} \leq \delta_{ra}(\omega). \tag{26}$$

Reeves (1991) uses δ_{ra} to specify the amount of uncertainty to be tolerated. A necessary and sufficient condition for RS is: there exists a controller that stabilizes all $P_\Delta = P + \Delta$ with the same number of unstable poles as P if and only if

$$\bar{\sigma}(P)\bar{\sigma}(P^{-1}(I-S)) < \frac{1}{\delta_{ra}} \quad \forall \omega. \tag{27}$$

The nominal sensitivity S is defined in (6). In the set-up of Fig. 5, S arises in tracking (r) and disturbance (d) rejection problems. Performance specifications could thus be imposed in terms of S. Note that (27) depends on K via S, while for IO selection a controller-independent criterion is desirable. This can be achieved by transforming (27) into a necessary condition. This involves some substitutions. The following necessary condition for combined RS and NP results: there exists a controller that stabilizes all P_Δ with the same number of unstable poles as P (RS) and that achieves $\bar{\sigma}(S) \leq \sigma_S$, with $\sigma_S < 1\ \forall \omega \leq \omega_S$ (NP), only if

$$\kappa(P) < \frac{1}{\delta_{ra}(1-\sigma_S)} \quad \forall \omega \leq \omega_S. \tag{28}$$

$\kappa(P) = 1$ for all 1×1 IO sets and so these cannot be compared. Qualitatively, (28) states that for IO sets causing large plant condition numbers (1) only limited performance (large σ_S and/or small ω_S) can be achieved and (2) only small uncertainty (small δ_{ra}) is permitted. The second part was already noted for unstructured, multiplicative uncertainty in Section 3.4.3. Candidate IO sets that do not meet (28) are rejected. However, IO sets may be incorrectly accepted, due to necessity.

Banerjee and Arkun (1995) and Reeves (1991) use the same δ_{ra} and σ_S for all candidate IO sets, implying the right-hand side of (28) to be the same for all candidates. However, from randomly generated matrices it was observed that $\kappa(P)$ tends to increase if P is extended with columns and rows. Hence, in terms of (28), achieving RS and NP would become more difficult if inputs and outputs were added and larger IO sets would be more easily rejected. This is also illustrated by the examples in Banerjee and Arkun (1995), Hoskin, Nett, and Reeves (1991), and Reeves (1991, Section 3.3). This counter-intuitive result can be traced back to the unstructured uncertainty representation: if an IO set is extended with inputs and outputs, the dimension of Δ grows accordingly, introducing additional uncertainty associated with *all* inputs and outputs. Such uncertainties are unlikely to occur in practice. To avoid this inconsistency when using (28), δ_{ra} should be determined for each IO set separately. Moreover, the NP specification σ_S is directly related to the outputs and using a distinct output set may call for a distinct σ_S. However, these approaches to specifying δ_{ra} and σ_S are infeasible for a large number of candidates.

Due to the assumption of inferential control, each of the n_z controlled variables must be represented by at least one of the outputs y and the square IO sets should at least have dimension $n_z \times n_z$ (Banerjee & Arkun, 1995). Besides meeting (28), additional requirements are thus imposed on the number and type of outputs. This must be accounted for prior to checking (28), e.g., by physical insight, which endangers the systematics and possibly the effectiveness of the IO selection method. This drawback applies to all IO selection methods assuming inferential control.

In Reeves (1991), two lower bounds to the condition number $\kappa(P)$ (in fact, to the *minimum* condition number $\tilde{\kappa}(P)$ in (21)) are used: the first one, $\kappa_\Lambda(P)$, is given by (20) and the second one is (Nett & Manousiouthakis, 1987):

$$\kappa_{\bar{\sigma}}(P) := \bar{\sigma}(\Lambda(P)) \leq \tilde{\kappa}(P) \leq \kappa(P) \tag{29}$$

with the RGA $\Lambda(P)$ according to (19). Reeves (1991) conjectures that $\kappa_\Lambda(P)$ is generally closer to $\tilde{\kappa}(P)$ than $\kappa_{\bar{\sigma}}(P)$. Both bounds are independent of the IO scaling. Replacing $\kappa(P)$ in (28) by either of the lower bounds relaxes the IO selection condition and more IO sets may be accepted. A reasonable motivation for introducing the scaling-independent quantities is that IO selection with (28) may give incorrect results if scaling is not or incorrectly performed. However, as noted before, appropriate scaling, and hence dependence on scaling, is crucial for a meaningful comparison of the variables on which uncertainty and performance specifications are imposed, as well as for the physical meaning of the variables. According to Reeves (1991), δ_{ra} and σ_S are specified under the assumption of proper scaling.

3.6.2. General control system set-up

The IO selection methods in this section use the set-up in Fig. 6. The uncertainties Δ are isolated from the nominal plant model G and w and z are split into parts related to uncertainty ($w_\mathrm{u}, z_\mathrm{u}$) and performance ($w_\mathrm{p}, z_\mathrm{p}$). Many control system set-ups (including those in Figs. 2, 4, and 5) can be cast into the set-up of Fig. 6. Δ may act upon the plant in a multiplicative, additive, or many other ways and it may be unstructured or structured.

Ross and Swartz (1997) consider combined RS and NP. In Fig. 6, w and z are required to have bounded amplitudes and the control problem is formulated in the l_1 system norm setting. The uncertainties are treated as nonlinear and/or time varying, which may be conservative if they are linear and time invariant. The performance is in terms of minimizing tracking errors to

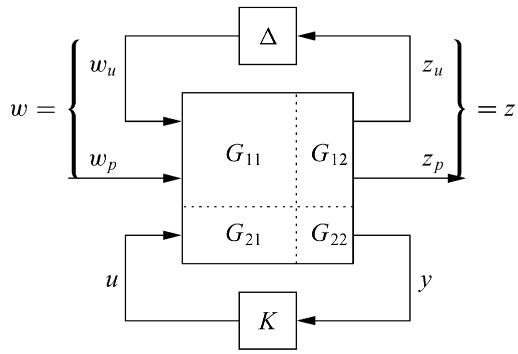


Fig. 6. General control system set-up; plant G, controller K, and uncertainty Δ.

setpoint changes. A controller-independent condition is derived that could be used to check if, for a given IO set, there exists any linear time-invariant controller meeting NP and RS. For unstructured Δ (as for NP only), this test involves solving a convex, constrained optimization problem. For structured Δ, the optimization is not convex and an approximation is invoked. The global optimum may not be found, so the IO selection may not be effective.

In Van de Wal and De Jager (1996) and Van de Wal (1998, Section 4.1), an IO selection method is proposed for separate or combined RS and NP. Δ is assumed to be unstructured, linear, and time invariant. An $\mathscr{H}_\infty$ control problem is formulated, with the $\mathscr{H}_\infty$ norm of a TFM $H(s)$ defined as

$$\|H(s)\|_\infty := \sup_\omega \bar{\sigma}(H(\mathrm{j}\omega)). \tag{30}$$

Weighting filters imposing specifications on the uncertainties and the exogenous and controlled variables are incorporated into G. For a given IO set and corresponding G, it can be checked whether there exists a stabilizing controller achieving $\|M\|_\infty < \gamma$. Here, γ is a specified value and $M := \mathscr{F}_l(G, K)$ is the nominal closed-loop system, denoted as a lower linear fractional transformation (Zhou et al., 1996, Chapter 10). This could be read as "G closed by K". In case of RS (NP), w_p and z_p (w_u and z_u) in Fig. 6 are absent and the $\mathscr{H}_\infty$ norm requirement is imposed on the relevant nominal closed-loop system denoted by M_RS (M_NP). In case of combined RS and NP, the $\mathscr{H}_\infty$ norm requirement becomes $\|[M_\mathrm{RS}\ \ M_\mathrm{NP}]\|_\infty < \gamma$. The IO selection method amounts to checking the set of controller existence conditions proposed by Glover and Doyle (1988). These involve requirements on the solutions of two algebraic Riccati equations. For a given IO set, the conditions are checked in succession, up to failure. For IO selection aimed at RP or RS against *structured* uncertainties, G can be modified such that the same method can be used, see Section 3.7.

In Van de Wal et al. (1997) and Van de Wal (1998, Section 4.1), the IO selection method sketched above is used for nonlinear systems via linearizations in a grid of stationary operating points. For the nonlinear $\mathscr{H}_\infty$ problem to be locally solvable, it is sufficient that the corresponding linear $\mathscr{H}_\infty$ problem is solvable (Van der Schaft, 1996, Chapter 7). Due to this, IO sets may be incorrectly rejected. On the other hand, for the nonlinear $\mathscr{H}_\infty$ problem to be globally solvable in the operating region, it is necessary that the linear $\mathscr{H}_\infty$ problems are solvable for all equilibria. This is caused by the fact that the sizes of the controllers' regions of attraction around the equilibria are unknown during IO selection and may not be large enough to overlap. Due to this, IO sets may be incorrectly accepted. Nevertheless, the IO selection method is useful for initial screening of many IO sets. The IO sets accepted for the first operating point are checked for the second operating point and so on, until all grid points have been checked. If the number of accepted IO sets is small enough, (nonlinear) controller design and closed-loop evaluation could be invoked to rigorously address IO set viability.

3.7. Robust performance

This section reviews IO selection methods and conditions with RP as the selection criterion. As in Section 3.6, methods for a one DOF control system set-up and methods for the more general set-up of Fig. 6 are treated separately.

3.7.1. One DOF control system set-up

The necessary conditions for IO selection in Braatz (1993) and Braatz, Lee, and Morari (1996) and the necessary conditions for secondary measurement selection in Lee (1991) and Lee and Morari (1991) are derived in the context of a particular controller design method. In Braatz (1993, Chapter 6) this is robust loopshaping. This technique is based on magnitude bounds for TFMs which are of special interest and which are used to parameterize the controller (Skogestad & Morari, 1988). For instance, for a tracking or disturbance rejection problem in the set-up of Fig. 5, K could be parameterized in terms of the sensitivity S in (6). If the RP control goal is appropriately quantified, the following is a necessary condition for existence of a robustly performing controller designed via loopshaping of S:

$$\mu(G_{11}(0) + G_{12}(0)P^{-1}(0)G_{21}(0)) < 1 \tag{31}$$

with μ the structured singular value of a matrix. A formal definition of μ can be found in many textbooks (Zhou et al., 1996, Chapter 11). Here it suffices to see μ as an extension of the maximum singular value $\bar{\sigma}$ that permits (structured) uncertainty characterizations and performance specifications to be captured simultaneously and nonconservatively. In (31), G_{11}, G_{12}, and G_{21} correspond to the partitioning of G in Fig. 6, while G_{22} equals $-P$ in, e.g., Fig. 5. Condition (31) is thus applicable to control problems which are first formulated in a one DOF control system set-up and which, after defining w and z, are transformed into the set-up of Fig. 6. Braatz (1993) and Braatz et al. (1996) show that (31) is also necessary for existence of an RP controller with integral action, i.e. $K(s) = (1/s)\hat{K}(s)$, with $\hat{K}(0)$ nonsingular and $S(0) = 0$. Braatz et al. (1996) propose additional necessary conditions in terms of μ which can be used for *joint* IO and CC selection. RP is actually a *dynamic* system property and hence the steady-state condition (31) may not be very useful for IO selection. Lee and Morari (1991) discuss conditions which are in essence equivalent to (31), but which are derived for loopshaping in estimation-based inferential control systems where estimates of the controlled variables are used. A TFM like S may not be relevant for performance and other TFMs are used instead. Application of all the above-mentioned IO selection conditions is restricted to square IO sets and steady state.

Rivera (1989) specifies NP in terms of magnitude constraints on the inputs u and on variables in z. The IO selection method involves three steps: the candidates are successively tested for (1) NP in terms of satisfaction of the magnitude constraints (using $\bar{\sigma}$), (2) RS for uncertainties bounded by the $\mathscr{H}_\infty$ norm (using μ), and (3) satisfaction of the constraints in the face of the uncertainties, i.e., RP (using μ). The constraints cannot be incorporated directly into the $\mathscr{H}_\infty$ norm setting that is common for RP. The $\mathscr{L}_1$ system norm setting would be more appropriate for handling constraints. However, by employing an inequality relating the $\mathscr{H}_\infty$ and $\mathscr{L}_1$ norm, Rivera (1989) accounts for magnitude constraints, though possibly quite conservatively. The IO selection conditions are restricted to steady state and square IO sets and they assume integral control also in the presence of uncertainty, i.e., $S_\Delta(0) = 0$, with $S_\Delta := (I + (P + \Delta)K)^{-1}$ for the system in Fig. 5 with unstructured Δ.

In Trierweiler and Engell (1997) and Trierweiler (1997), the "robust performance number" $\varsigma(P, T_s)$ is proposed as a potential IO selection tool which is not restricted to steady state. For a control system in the set-up of Fig. 2, $\varsigma(P, T_s)$ is

$$\varsigma(P, T_s) := \sup_\omega \sqrt{\bar{\sigma}((I - T_s)T_s)\left(\tilde{\kappa}(P) + \frac{1}{\tilde{\kappa}(P)}\right)}. \tag{32}$$

This measure is based on the μ measure for RP in Zhou et al. (1996, Section 11.3.3). The latter is valid in case of a plant-inverting controller and the conservative, *unstructured*, multiplicative input uncertainty in Fig. 4. The μ measure also involves the plant condition number $\kappa(P)$. To reduce conservatism, Trierweiler and Engell (1997) replace $\kappa(P)$ by the minimum plant condition number $\tilde{\kappa}(P)$ in (21). In case of the less conservative, *structured*, multiplicative input uncertainty, $\tilde{\kappa}(P)$ is more appropriate, as illustrated by Chen et al. (1994). Furthermore, the first term below the root in (32) replaces an expression in the original μ measure that represents the RS and NP requirements. Here, T_s is a *specification* for the complementary sensitivity function that is defined as

$$T(s) := P(s)K(s)(I + P(s)K(s))^{-1} = I - S(s). \tag{33}$$

The choice of T_s is determined by (1) the performance specifications and (2) the occurrence of RHP zeros, RHP poles, and time delays in P. Due to (2), the choice of T_s is not completely free, since closed-loop stability must be guaranteed. Trierweiler and Engell (1997) state that $\bar{\sigma}((I - T)T)$ is largest in the crossover region and so $\varsigma(P, T_s)$ automatically emphasizes the frequency region where suppressing the effect of uncertainties is usually the most important. The IO sets yielding a small $\varsigma(P, T_s)$ are preferred, since these will give good RP if an inverse-based controller "$K = P^{-1}$" is used. Trierweiler and Engell (1997) also propose an extension to $\varsigma(P, T_s)$ that applies to polytopic models, i.e., sets of linear system models which together represent the original model. This extension makes it possible to consider various types of uncertainties explicitly and to examine nonlinear systems with varying operating conditions. However, IO selection may be conservative if the nonlinear system is not accurately described by the polytopic model.

Three drawbacks of IO selection with the RP number $\varsigma(P, T_s)$ are mentioned. First, the derivation of $\varsigma(P, T_s)$ is mainly heuristic and lacks a sound theoretical foundation. It seems an arbitrary combination of different numbers, which makes $\varsigma(P, T_s)$ difficult to assess. Second, due to the use of the condition number larger IO sets are more likely to be eliminated (see Section 3.6) and in case of the optimally scaled condition number the physical meaning of the inputs and outputs of P may be lost. Third, it may be necessary to re-compute T_s for distinct IO sets, e.g., if the RHP zeros are not the same for all IO sets.

3.7.2. General control system set-up

Lee (1991, Chapter 3) suggests necessary conditions for RP-based IO selection which are tied to LQG control and model predictive control (MPC) or to integral controllers. The conditions are in terms of a μ bound on weighted plant data. Their application is restricted to square IO sets and steady state. This is circumvented for the methods in the rest of this section, but these are computationally more involved.

Consider Fig. 6 and denote the nominal closed-loop system relating w to z by M. The upcoming IO selection methods are posed in the $\mathscr{H}_\infty$ system norm setting and they aim at selecting those IO sets for which there exists a stabilizing controller achieving RP or RS against *structured* uncertainties. The RP requirement is expressed as a requirement on the μ value of a TFM $H(s)$, defined as

$$\|H(s)\|_\mu := \sup_\omega \mu(H(j\omega)). \tag{34}$$

Strictly speaking, $\|H\|_\mu$ is not a norm. In contrast to the $\mathscr{H}_\infty$ norm requirement $\|M\|_\infty < \gamma$, there is no set of conditions that can be used to verify whether there exists a stabilizing controller achieving $\|M\|_\mu < \gamma$. Controller design aimed at minimizing $\|M\|_\mu$ or achieving a particular value of $\|M\|_\mu$ could be invoked instead. This μ-synthesis problem is still unresolved, but an approximate design, called D–K iteration (Zhou et al., 1996, Chapter 11), is possible. The aim is to make the following μ upper bound small enough:

$$\|H(s)\|_{\bar{\mu}} := \sup_{\omega} \bar{\mu}(H(\mathrm{j}\omega)) \geq \|H(s)\|_\mu \tag{35}$$

with

$$\bar{\mu}(H(\mathrm{j}\omega)) := \inf_{D(\mathrm{j}\omega)} \bar{\sigma}(D(\mathrm{j}\omega)H(\mathrm{j}\omega)D^{-1}(\mathrm{j}\omega)). \tag{36}$$

The D-scales $D(\mathrm{j}\omega)$ belong to a set determined by the structure of Δ (Zhou et al., 1996, Chapter 11). For notational simplicity, it is assumed in (36) that each individual block in Δ is square and that w_p and z_p have the same dimension. Each iterative step in D–K iteration alternates between computing $\bar{\mu}(M)$ and the D-scales for a frequency grid, fitting a TFM through $D(\mathrm{j}\omega)$, and performing an $\mathscr{H}_\infty$ optimization of G augmented with the D-scale approximations. The iteration could be stopped if $\|M\|_{\bar{\mu}}$ is small enough or does not further converge. This is the basic idea of IO selection in Van de Wal (1998, Section 6.6). The IO selection method in De Jager, Van de Wal, and Kamidi (1998) combines D–K iteration with $\mathscr{H}_\infty$ controller existence checks (Section 3.6.2) for a scaled plant. Due to the fact that achieving $\|M\|_{\bar{\mu}} < \gamma$ is sufficient for achieving $\|M\|_\mu < \gamma$, IO sets may be incorrectly rejected. The methods in the remainder of this section circumvent the time-consuming D–K iterations for distinct candidate IO sets, at the price of reduced effectiveness.

The starting point for the IO selection method proposed by Lee, Braatz, Morari, and Packard (1995) is to check whether there exists a stabilizing controller achieving $\|M\|_{\bar{\mu}} < \gamma$. To arrive at controller-independent IO selection conditions, the requirement of the controller being stabilizing is dropped, which is equivalent to dropping the controller's causality. According to Lee et al. (1995), the consequences are only significant in the crossover region, where joint stability and performance requirements are especially difficult to be met. Note that dropping the stabilizing property amounts to dropping the interdependence of gain and phase as imposed by Bode's gain-phase relationships: the performance specifications, expressed as magnitude requirements, are more easily met if the phase requirements can be neglected. The IO selection conditions are in terms of two LMIs that must be jointly feasible across frequency. Neglecting crucial frequencies in the grid is another source of necessity. Only for a few special cases, like RP with unstructured Δ, checking the *joint* feasibility of the LMIs is convex. For more general cases, only the feasibility of the individual LMIs can be checked. This implies that the input (output) set is checked under the assumption of a perfect output (input) set, but that the combination of the actual input and output sets cannot be checked. The pros and cons of this IO selection method are discussed in more detail by Van de Wal, Philips, and De Jager (1998).

The last two methods to be reviewed reduce the structured RP problem to an unstructured one, such that $\mathscr{H}_\infty$ controller existence conditions can be used for IO selection, as in Section 3.6.2. Both methods proceed along four steps that are mainly the same: (1) D–K iteration for the *full IO set* until $\|M\|_{\bar{\mu}}$ does not further decrease, (2) μ-analysis for the full IO set's optimal closed-loop system and generation of particular TFMs to be used in the next step, (3) modification of the candidate IO sets' plants G with these TFMs, and (4) subjecting the modified plants to the $\mathscr{H}_\infty$ controller existence conditions. The two IO selection methods use different TFMs in the second step. For both methods, using the TFMs related to the full IO set also for the other IO sets is based on the assumption that these TFMs are representative for IO sets which are almost as good as the full one.

The first method is described in Van de Wal (1998, Section 4.2.1) and Van de Wal and De Jager (1997). As in D–K iteration, G is augmented with approximations of the D-scale data resulting from the μ upper bound computation. These TFMs, called "D-scale estimates", are constructed for the full IO set, but they are also used for the other IO sets. The D-scale estimates may not be optimal for other IO sets than the full one, which is the main reason for the sufficiency of this method. So, IO sets may be incorrectly rejected. The second method is described in Van de Wal (1998, Section 4.2.2) and Van de Wal and De Jager (1998). A particular TFM $\hat{\Delta}(s)$, representing an uncertainty, is absorbed into G for the distinct candidate IO sets. The magnitude of $\hat{\Delta}$ is such that the RP level of the full IO set's optimal closed-loop system is just violated, at least for one frequency, and therefore it is called a "worst case uncertainty". This method has a necessary character and IO sets may be incorrectly accepted. This is mainly due to two sources. First, $\hat{\Delta}$ is not necessarily worst case for all frequencies. Second, even if it is, it may not be worst case for other IO sets than the full one for which it was constructed.

3.8. Search methods

Most IO selection criteria discussed are evaluated on a per candidate basis, so worst case all IO sets have to be checked. This implies an exhaustive search for a number of combinations that grows exponentially with $N_u + N_y$, which is practically intractable and inefficient. As will appear below, to avoid the combinatorial nature of the problem, in general one has to make a compromise, either in the choice of the selection criterion or in the choice of the search strategy.

The computational effort can be reduced significantly by realizing that often it is not necessary to check all candidate IO sets to discriminate viable from nonviable IO sets. For many IO selection problem formulations, the following two dual properties hold: (A) eliminating inputs and/or outputs does not improve control and (B) adding inputs and/or outputs does not worsen control. Realizing this, subsets (supersets) of nonviable (viable) IO sets need not be checked, but can directly be termed nonviable (viable). The validity of (A) or (B) depends on the considered system and the IO selection criterion. For instance, the properties may not apply if there are candidate actuators and sensors affecting the open-loop dynamics or if the IO selection criterion involves the plant's condition number.

Properties (A) or (B) can form the basis of a strategy of going through the candidate IO sets. Such a strategy is discussed by De Jager and Toker (1998). It is based on a novel algorithm to generate all maximal-independent or minimal-dependent sets, a standard problem in combinatorial optimization. The problem is still $\mathcal{NP}$-hard, i.e., there is no solution possible in polynomial time, although in practice (for $N_u + N_y$ not too large) the computing time is polynomial in $N_u + N_y$ and in the number of IO sets that characterize the complete solution. An example using a μ-based RP criterion for a problem with $N_u + N_y = 28$ has shown the algorithm to reduce significantly, i.e., by a factor 10^{-6}, the number of IO sets to be checked, compared to an exhaustive search (De Jager et al., 1998).

The algorithm proposed by De Jager and Toker (1998) generates all *minimal dependent sets*, i.e., sets that are viable but whose subsets are all nonviable. These IO sets do not contain redundant actuators or sensors. If any device fails, even if the failure is detected and the controller changed accordingly, the performance will be below par. But, (1) redundancy aspects can be built into the selection criterion and (2) combining two disjunct minimal IO sets will protect against single-device failures. The results are therefore also useful for adding fault-tolerance to a system. Contained in the minimal-dependent sets are sets with the smallest number of devices: the *minimum-dependent sets*. By providing all minimal-dependent sets (not only the smallest ones) the designer is given full information about the performance structure. He can use this knowledge to attend to aspects not included in the selection criterion and can limit a further search for an optimal IO set to the minimal-dependent sets.

Another approach to avoid exhaustive testing on a candidate-by-candidate basis is by using optimization. The minimal-dependent set algorithm generates all sets that meet a certain constraint using the selection criterion. Adding an optimization criterion to the constraint may yield a unique solution. Several algorithms in combinatorial optimization can be employed (Aarts & Lenstra, 1997). Hassibi, How, and Boyd (1998) combine an $\mathscr{L}_1$ norm optimization criterion on the controller coefficients with other design constraints. For closed-loop pole region constraints, this yields a linear programming problem that can be solved efficiently. If controller coefficients tend to be zero, it may be possible that some actuators and sensors are not used. This method combines IO and CC selection. It gives an approximate solution to a combinatorial optimization problem. It is not guaranteed to generate one of the optimum solutions with the lowest number of devices, but it may end up with a solution close to the optimal ones. It has the advantage of being very efficient, because it is a direct method and uses a simple convex optimization routine.

4. Applications of IO selection methods

This section illustrates that systematic IO selection is important for a wide variety of applications. For most references and applications, the employed IO selection method is assigned to one of the main groups in Section 3. Sometimes the method is application-specific and then it is not further characterized.

4.1. Process control

Systematic IO selection is important for systems with a large number of candidate inputs and outputs. Therefore, most literature on IO selection is related to the large-scale systems from process industry. Applications of all selection criteria in Sections 3.1–3.7 are encountered. The majority of the applications is based on IO controllability measures, since especially for the systems in process industry with many candidate IO sets simple criteria are preferred. The IO controllability measures are easily computed and well understood. They are also well suited to focus on a tight frequency range or a single frequency. Steady-state performance is crucial in process industry and so many IO controllability measures are studied for frequency zero. More rigorous IO selection criteria are often applied in a later stage when the number of candidates has already been reduced considerably.

4.1.1. Individual units

For *individual units*, the following IO selection applications are encountered:

Distillation columns: This application is often encountered:

- The selection of the number and placement of temperature sensors in Bequette and Edgar (1986, IO controllability), Chang and Yu (1990, IO controllability),

Havre et al. (1996, IO controllability), Joseph and Brosilow (1978, IO controllability, efficiency of manipulation and estimation), Lee (1991, Section 3.5 and Appendix B.3.3, efficiency of manipulation and estimation, IO controllability, RP), Lee and Morari (1991, RP), Lee et al. (1995, RP), and Morari and Stephanopoulos (1980b, efficiency of manipulation and estimation).
- The selection of the placement and type of sensors in Moore et al. (1987, IO controllability).
- Input selection in Cao (1995, Section 5.5, IO controllability), Figueroa, Romagnoli, and Barton (1992, RP), Morari and Stephanopoulos (1980a, state controllability and observability), Rivera (1989, NP, RS, RP), Rivera et al. (1993, IO controllability, NP, RS, RP), and Skogestad and Morari (1987b, IO controllability).
- IO selection in Hansen, Heath and Jørgensen (1996, RHP zeros, IO controllability) and in Yu and Luyben (1986, IO controllability).

Recalling the broad interpretation of inputs in the introduction of this paper, it is noteworthy that in Hansen, Heath, and Jørgensen (1996) two candidate inputs are setpoint variables instead of physically actuated variables.

Tubular reactors: The optimal location of temperature and concentration sensors in Kumar and Seinfeld (1978, efficiency of manipulation and estimation).

Boilers: Input selection in Keller and Bonvin (1987, IO controllability).

Mixer-blenders: IO selection in Morari and Stephanopoulos (1980a, state controllability and observability).

4.1.2. Small-scale integrated plants

For *small-scale integrated plants*, the following applications are mentioned:

FCC: Hovd and Skogestad (1993, RHP zeros) and Morari and Stephanopoulos (1980b, efficiency of manipulation and estimation) select temperature sensor sets for a fluid catalytic cracker (three units).

CSTR: Input selection for a series connection of two continuous stirred tank reactors and an intermediate mixer is considered in Cao (1995, IO controllability, efficiency of manipulation and estimation), Cao et al. (1996, efficiency of manipulation and estimation) and, without a mixer, in Daoutidis and Kravaris (1992, accessibility).

HID: Weitz and Lewin (1996, IO controllability) study input selection for two heat-integrated distillation columns.

DEE: Double-effect evaporators, i.e., two evaporators in series, are considered in Heath, Perkins, and Walsh (1996, IO controllability) and Narraway and Perkins (1993, IO controllability), where both inputs and outputs are selected, and in Morari and Stephanopoulos (1980b, efficiency of manipulation and estimation), where only outputs are selected.

MDHE: Govind and Powers (1982, accessibility) focus on IO selection for the combination of a mixer, a divider, and a heat exchanger.

HE: For a network of heat exchangers, Daoutidis and Kravaris (1992, accessibility) perform input selection and Reeves (1991, Section 3.3, combined RS and NP) performs IO selection.

4.1.3. Large-scale integrated plants

For *large-scale integrated plants*, the following applications can be found:

HDT: In Cao (1995, Chapter 7, RHP zeros, IO controllability), Cao and Biss (1996, IO controllability), and Cao and Rossiter (1996, IO controllability), inputs are selected for the hydrodealkylation of toluene process (ten units), while in Cao, Rossiter, and Owens (1997, IO controllability) outputs are selected for this process.

TEP: Banerjee and Arkun (1995, combined RS and NP) perform IO selection for the Tennessee–Eastman plant (five units).

WOP: Morari and Stephanopoulos (1980a, state controllability and observability) perform IO selection for the Williams–Otto plant (four units).

TID: In Lin et al. (1994, state controllability and observability), IO selection is applied for a thermally integrated distillation sequence (16 units).

FFC: In Narraway and Perkins (1993, IO controllability) this is done for a froth flotation circuit in mining industry.

Finally, Luyben, Tyréus, and Luyben (1998) suggest approaches for designing regulatory control structures for plantwide control problems. The Tennessee–Eastman plant, the HDA process, the isomerization process, and the vinyl acetate process serve as examples (individual units, such as reactors, heat exchangers, and distillation columns, are also discussed). Unlike the system-theoretic approaches that are the focus in this review, the approach by Luyben et al. (1998) relies completely on engineering heuristics, experience, and features that are unique for the considered problems.

4.2. Flexible structures

Control of flexible mechanical structures is another field where IO selection attracts much attention. Only part of the literature is referred to below. The selection methods are always based on quantitative measures for state controllability/observability or efficiency of manipulation/estimation. These measures lend themselves well to express the energy contained in a system's state, input, and output vector. Limitation of the energy stored in or supplied to a flexible system is an important performance indicator in the field of vibration reduction. Moreover, the well described and specific nature of the problem (e.g., the plants are usually stable and linear with complex poles and small damping factors) sometimes makes it possible to simplify the IO selection problem and to derive accurate approximate expressions for the effect of each candidate actuator and sensor.

Unlike most other applications, the number of candidate actuator and sensor locations is not always finite. Optimization algorithms can be invoked to find the optimal locations. Moreover, distributed actuators and sensors (like piezoelectric layers) are often applied, instead of the more common point actuators and sensors. Selecting among distributed actuators and sensors also involves decisions on the geometry:

- For point actuators and sensors, IO selection for a finite number of candidates is the topic in Andersson (1997), Balas and Young (1999, efficiency of estimation), Gawronski and Lim (1996, state controllability and observability), Georges (1995, state controllability and observability), Hać and Liu (1993, state controllability and observability), Lim (1997, state controllability and observability), Norris and Skelton (1989, efficiency of manipulation and estimation), Roh and Park (1997, efficiency of manipulation and estimation), Seto and Mitsuta (1994), and Vander Velde and Carignan (1984, state controllability and observability).
- IO selection for an infinite number of candidates is discussed in Lee and Chen (1994, efficiency of manipulation and estimation) and Xu et al. (1994, efficiency of manipulation and estimation).
- Selection of (single) distributed actuators and sensors is considered by Hać (1995, state controllability and observability) and Ko et al. (1994, state controllability and observability).

4.3. Flight control

Specialized IO selection methods have also been used for aircraft and spacecraft control purposes. There is no method that is clearly used the most frequently. Based on the complexity and the nature of the control problem, the preferred IO selection method should be considered for each application separately. For a process-control-like problem (e.g., controlling fluids, temperatures, and pressures in engines) IO controllability measures are a suitable starting point, whereas state controllability and observability and efficiency of manipulation and estimation may be preferred for mechanical structures (e.g., active vibration control of wings and antennas). The following applications of flight control are encountered:

- Hoskin et al. (1991, combined RS and NP), Reeves (1991, Section 7.2, combined RS and NP), and Samar and Postlethwaite (1994, state controllability and observability, RHP zeros, IO controllability) determine suitable actuators and sensors for controlling high-performance aircraft engines.
- Compressors are often used in aircraft engines and, in this context, IO selection for compression systems could also be mentioned, see Hendricks and Gysling (1994), Montazeri-Gh, Allerton, and Elder (1996, efficiency of manipulation and estimation), Simon, Valavani, Epstein, and Greitzer (1993), and Van de Wal et al. (1997, combined RS and NP).
- The optimal location of actuators for attitude control of satellites is considered by Müller and Weber (1972, state controllability and observability).

4.4. Other applications

Some other IO selection applications that do not belong to any of the categories discussed above are listed below:

- Braatz et al. (1996, RP) consider actuator placement for a paper machine.
- IO selection for active suspensions is studied by Al-Sulaiman and Zaman (1994, efficiency of manipulation), De Jager et al. (1998, RP), Van de Wal et al. (1998, RP), Van de Wal and De Jager (1996, combined RS, and NP) Van de Wal and De Jager (1998, RP), and Van de Wal and De Jager (1997, RP).
- Demetriou and Fahroo (1999, efficiency of manipulation), Katsikas, Tsahalis, Manolas, and Xanthakis (1995), and Ruckman and Fuller (1995) study actuator placement for active noise control.
- Kosut and Kabuli (1995) focus on IO selection for a thermal processing chamber.

5. A bird's eye view on IO selection

Based on the desirable properties in Section 2, the IO selection methods reviewed in this paper are assessed and compared to obtain insight into the state of the art in IO selection. This gives rise to some suggestions for further research.

5.1. Qualitative assessment of IO selection methods

The assessment is qualitative and certainly not definitive, since this would call for a comparative, quantitative set of examples or a benchmark competition. The desirable properties in Section 2 form the basis for the evaluation. The extent to which the methods fulfill the properties is expressed by these symbols in Table 1:

+: The method exhibits this property to a satisfactory extent.
0: The method exhibits this property to a moderate extent.
−: The method does not exhibit this property, or only to a minor extent.

Table 1
Qualitative assessment of the reviewed classes of IO selection methods

Section	IO selection criterion	Desirable properties of IO selection methods								
		1	2	3	4	5	6	7	8	$\mathscr{Q}$
3.1	Accessibility	+	+	+	+	−	−	+	−	7.1
3.2	State controllability & observability	+	+	+	+	−	0	+	−	7.6
3.3	Right half-plane zeros	+	+	0	0	−	+	+	−	6.8
3.4	IO controllability	0	+	−	0	0	+	+	0	5.9
3.5	Efficiency of manipulation & estimation	0	0	0	+	+	+	−	−	6.2
3.6	Robust stability & nominal performance	0	0	0	0	+	+	+	−	6.2
3.7	Robust performance	0	0	−	0	+	+	0	−	5.0
3.8	Search methods & robust performance (μ)	+	0	+	+	+	+	0	0	8.5

Table 1 also provides a quality indicator $\mathscr{Q}$ for each *group* of IO selection methods. The eight desirable properties of IO selection were ranked according to descending importance. The following weights $\mathscr{W}$ are assigned to each property: $\mathscr{W} = 3$ for Properties 1–3, $\mathscr{W} = 2$ for Properties 4–6, and $\mathscr{W} = 1$ for Properties 7 and 8. If a property is assigned the symbol $+$, $\mathscr{W}$ is multiplied by $+1$, if it is assigned 0, $\mathscr{W}$ is multiplied by 0, and if it is assigned $-$, $\mathscr{W}$ is multiplied by -1. By averaging, scaling, and translating, $\mathscr{Q}$ then takes a value between 0 and 10. The ranking of the desirable properties is partially subjective and problem-dependent. So, $\mathscr{Q}$ only gives a rough indication of what to expect from each method.

Table 1 gives a fairly complete picture of the currently known IO selection methods and their pros and cons. Clearly, all methods show shortcomings and the development of an IO selection method resolving *all* shortcomings is probably too ambitious. The most remarkable issues in Table 1 are now discussed, following the order of the eight desirable properties of IO selection methods:

1. *Well founded*: Most IO selection methods lack a sound theoretical foundation. The method based on combined RS and NP (Reeves, 1991) uses the same performance and uncertainty specifications for all IO sets, but the assumption of unstructured uncertainty requires that the uncertainty quantification is reconsidered for each IO set. It seems as if for some IO selection methods the plant's condition number is quite arbitrarily applied as a robustness indicator, without noting that this implies a type of uncertainty which may not be appropriate. For the RP-based IO selection methods that reduce the control problem to an unstructured one, the theoretical foundation is moderate, since it partially relies on an engineering heuristic and no immediate prospects are foreseen for a theory in support of this heuristic. For the IO controllability measures, some of the definitions seem to conflict with the actually intended objectives.

2. *Efficient*: Regarding *computational effort*, the methods aimed at RP and efficiency of manipulation and estimation involve computations which require more effort than, e.g., computation of the IO controllability measures. When a μ-based RP criterion is used, D–K iteration is required. Due to the search method employed, the efficiency of the RP-based approach in Section 3.8 could be termed moderate. Regarding *analytical effort*, a less rigorous IO selection criterion usually requires less details and hence less analytical effort. For instance, the method based on accessibility only involves structural system data, whereas the methods based on RP require numerical data of the plant, the disturbances, and the uncertainties, as well as detailed performance specifications.

3. *Effective*: Many IO controllability measures are only crude representations of the actual objectives. The effectiveness of IO selection is then inherently poor. The methods based on RP and the method based on combined RS and NP proposed by Reeves (1991) rely on either necessary or sufficient conditions and IO sets may be incorrectly accepted or rejected. Apart from the least rigorous IO selection methods, efficiency and effectiveness do not go well together. Controller design becomes inevitable for an effective and rigorous IO selection, as illustrated by, e.g., the RP-based method involving μ-synthesis. The number of controller designs can be reduced in the last case, because only existence conditions for a (suboptimal) controller need to be checked.

4. *Generally applicable*: Section 3.4 shows that numerous IO selection conditions can be derived from the IO controllability measures. Most measures are developed for a one DOF control system set-up, which limits the general applicability. Usually, it is then assumed that the measured variables (outputs) properly represent the controlled variables, such that performance specifications can be imposed on the outputs. This is also assumed for most IO selection methods based on RHP zeros and some of the methods based on RP or combined RS and NP. Accessibility can be checked for nonlinear control systems, while state controllability and observability have also been extended to nonlinear systems (Nijmeijer & Van der Schaft, 1990, Chapter 3). This also holds for RHP zeros: unstable zero dynamics of a nonlinear system

plays a similar role to that of RHP zeros (Isidori, 1995, Section 4.3). For other IO selection methods, generalizations to nonlinear systems may also be possible.

5. *Rigorous*: A major disadvantage of the methods based on accessibility and state controllability/observability is the lack of rigor. The individual IO controllability measures do not address IO set viability very rigorously either, but this can be improved by sequential IO selections for distinct IO controllability measures. The RP-based IO selection methods are the most rigorous, though this is disputable for the methods that only apply at steady state. RP still neglects some important issues in control system design, like good reliability and low costs of purchasing, operating, and maintaining a control system. These issues could be used for the ultimate decision on which IO set is going to be implemented.

6. *Quantitative*: The majority of the IO selection methods employ quantitative criteria. Exceptions are structural state controllability and observability. For accessibility, the relative degree is a possible quantification, but in general it is not rigorous enough to indicate whether the intended control goal can be achieved.

7. *Controller independent*: Some methods based on efficiency of manipulation and estimation assume static state-feedback or output-feedback control, or static estimation, while some of the RP-based methods assume integral control, LQG control, MPC, or D–K iteration. To avoid controller dependence, most methods based on IO controllability assume perfect control, whereas some methods based on RP or combined RS and NP check the existence of any finite-dimensional, linear, time-invariant, and stabilizing controller achieving a specified norm bound of the closed-loop system.

8. *Direct*: Some IO controllability measures provide prospects for direct IO selection. They are computed only once, namely for the full IO set. The naturally occurring exponential growth of the IO selection problem is then avoided. None of these direct, often heuristic, methods are guaranteed to be effective. Checking the IO selection criteria on an indirect candidate-by-candidate basis is often suggested, but it may not be necessary to check all candidates, see Section 3.8.

It would be useful to have a practical guideline for choosing the "right" approach to IO selection. This should be a well-motivated one. A general guideline for picking the "right" approach to IO selection is hard to give. For a problem with a large number of candidate IO sets, IO selection may proceed by sequentially applying IO selection methods with gradually increasing rigor and hence the number of candidate IO sets is reduced gradually. Proposals for systematic, sequential IO selection procedures are given by, e.g., (from less detailed to more detailed) Lee (1991, Chapter 3), Trierweiler (1997, Chapter 2), and Van de Wal (1998, Chapter 7).

5.2. Directions for future research

Besides performance, other issues in IO selection are control system complexity and costs. An obvious way to perform IO selection is to first eliminate the candidate IO sets that cannot achieve the intended control objectives and then to assess control system complexity and costs to make the ultimate decision on the IO set. Future research should pay attention to the development of IO selection methods which better integrate these issues. Whereas addressing control system complexity and costs may not always call for system-theoretic concepts, addressing control objectives does. The developments in the field of LMIs have increased the possibilities to address control objectives in a more rigorous and general way.

Only a few IO selection methods are readily applicable to nonlinear systems. Two categories can be distinguished: methods that do require a linearized plant model and methods that do not. IO selection based on a linear model is the easiest and most efficient. However, it may be ineffective, since the nonlinear plant may loose desirable properties due to linearization, like controllability. For nonlinear systems that operate close to equilibria, methods for linear systems are quite useful, especially for initial screening. The conclusions on IO set viability are then valid locally around each investigated equilibrium, but the size of this region is not known at the stage of IO selection. To guarantee a certain size of the region, the linearization errors in a prescribed region around an equilibrium could be treated as uncertainties and the IO selection methods based on RS or RP could be applied. The development of IO selection methods that do not require linearization is a wide open field. Some of the selection criteria that have a nonlinear equivalent have already been mentioned in the assessment of general applicability in Section 5.1.

References

Aarts, E. H. L., & Lenstra, J. K. (Eds.) (1997). *Local search in combinatorial optimization*. Chichester: Wiley-Interscience.

Al-Sulaiman, & Zaman, S. (1994). Actuator placement in lumped parameter systems subjected to disturbance. *Computers & Structures*, *52*(1), 41–47.

Andersson, M. (1997). Noise robust actuator placement on flexible structures. *Proceedings of IEEE Conference on Decision and Control* (pp. 1958–1959).

Balas, G. J., & Young, P. M. (1999). Sensor selection via closed-loop control objectives. *IEEE Transactions on Control Systems Technology*, *7*(6), 692–705.

Banerjee, A., & Arkun, Y. (1995). Control configuration design applied to the Tennessee Eastman plant-wide control problem. *Computers & Chemical Engineering*, *19*(4), 453–480.

Bequette, B. W., & Edgar, T. F. (1986). The equivalence of non-interacting control system design methods in distillation. *Proceedings of American Control Conference*, vol. 1 (pp. 31–37).

Biss, D., & Perkins, J. D. (1993). Application of input-output controllability analysis to chemical processes. *Proceedings of European Control Conference*, vol. 2 (pp. 1056–1059).

Braatz, R. D. (1993). *Robust loopshaping for process control*. Ph.D. thesis, California Institute of Technology.

Braatz, R. D., Lee, J. H., & Morari, M. (1996). Screening plant designs and control structures for uncertain systems. *Computers & Chemical Engineering*, *20*(4), 463–468.

Bristol, E. H. (1966). On a new measure of interaction for multivariable process control. *IEEE Transactions on Automatic Control*, *11*, 133–134.

Cao, Y. (1995). *Control structure selection for chemical processes using input–output controllability analysis*. Ph.D. thesis, University of Exeter, UK.

Cao, Y., & Biss, D. (1996). New screening techniques for choosing manipulated variables. *Proceedings of IFAC World Congress*. vol. M (pp. 103–108).

Cao, Y., Biss, D., & Perkins, J. D. (1996). Assessment of input–output controllability in the presence of control constraints. *Computers & Chemical Engineering*, *20*(4), 337–346.

Cao, Y., & Rossiter, D. (1996). Input screening method for disturbance rejection. *UKACC International Conference on CONTROL*, vol. 1 (pp. 497–502).

Cao, Y., Rossiter, D., & Owens, D. (1997), Measurement selection for partially controlled systems. *Proceedings of IEEE Conference on Decision and Control*, vol. 1 (pp. 66–71).

Chang, J.-W., & Yu, C.-C. (1990). The relative gain for non-square multivariable systems. *Chemical Engineering Science*, *45*(5), 1309–1323.

Chen, J., Freudenberg, J. S., & Nett, C. N. (1994). The role of the condition number and the relative gain array in robustness analysis. *Automatica*, *30*(6), 1029–1035.

Daoutidis, P., & Kravaris, C. (1992). Structural evaluation of control configurations for multivariable nonlinear processes. *Chemical Engineering Science*, *47*(5), 1091–1107.

De Jager, B., & Toker, O. (1998). Complexity of input output selection. In A. Beghi, L. Finesso, & G. Picci (Eds.), *Proceedings of International Symposium on Mathematical Theory of Networks and Systems*, Il Polygrafo, Padova (pp. 597–600), Padova: Polygrafo.

De Jager, B., Van de Wal, M., & Kamidi, R. (1998). Large-scale rigorous actuator and sensor selection. *Proceedings of IEEE Conference on Decision and Control* (pp. 4314–4319).

Demetriou, M. A., & Fahroo, F. (1999). Optimal location of actuators for control of a 2-D structural acoustic model. *Proceedings of IEEE Conference on Decision and Control* (pp. 4290–4295).

Figueroa, J., Romagnoli, J., & Barton, G. (1992). Robust control indicators in process design and control. In J.D. Perkins (Ed.), *Preprints IFAC Workshop on Interactions Between Process Design and Process Control* (pp. 139–144). Oxford: Pergamon Press.

Freudenberg, J. S., & Looze, D. P. (1985). Right half plane poles and zeros and design tradeoffs in feedback systems. *IEEE Transactions on Automatic Control*, *30*(6), 555–565.

Gawronski, W., & Lim, K. B. (1996). Balanced actuator and sensor placement for flexible structures. *International Journal of Control*, *65*(1), 131–145.

Georges, D. (1995). The use of observability and controllability Gramians or functions for optimal sensor and actuator location in finite-dimensional systems. *Proceedings of IEEE Conference on Decision and Control*, vol. 4 (pp. 3319–3324).

Georgiou, A., & Floudas, C. A. (1989). Structural analysis and synthesis of feasible control schemes: Theory and applications. *Chemical Engineering Research and Design*, *67*, 600–618.

Glover, K., & Doyle, J. C. (1988). State-space formulae for all stabilizing controllers that satisfy an $\mathscr{H}_\infty$-norm bound and relations to risk sensitivity. *Systems & Control Letters*, *11*, 167–172.

Goodwin, G. C., & Seron, M. M. (1995). Complementarity constraints for nonlinear systems. *Preprints IFAC Nonlinear Control Systems Design Symposium*, vol. 2 (pp. 771–776).

Govind, R., & Powers, G. J. (1982). Control system synthesis strategies. *A.I.Ch.E Journal*, *28*(1), 60–73.

Hać, A. (1995). Distribution of actuators in vibration control of adaptive structures. *Proceedings of American control conference*, vol. 6 (pp. 4295–4299).

Hać, A., & Liu, L. (1993). Sensor and actuator location in motion control of flexible structures. *Journal of Sound and Vibration*, *167*(2), 239–261.

Hansen, J. E., Heath, J., & Jørgensen, S. B. (1996). Control structure selection for energy integrated distillation column. *Proceedings of UKACC International Conference on CONTROL*, vol. 1 (pp. 509–514).

Hassibi, A., How, J. P., & Boyd, S. P. (1998). Low-authority controller design via convex optimization. *Proceedings of IEEE Conference on Decision and Control*, vol. 1 (pp. 140–145).

Havre, K., Morud, J., & Skogestad, S. (1996). Selection of feedback variables for implementing optimizing control schemes. *Proceedings of UKACC International Conference on CONTROL*, vol. 1 (pp. 491–496).

Havre, K., & Skogestad, S. (1996). Effect of RHP zeros and poles on performance in multivariable systems. *Proceedings of UKACC International Conference on CONTROL*, vol. 2 (pp. 930–935).

Heath, J., Perkins, J., & Walsh, S. (1996). Control structure selection based on linear dynamic economics: Multiloop PI structures for multiple disturbances. *Proceedings of IFAC World Congress*, vol. M (pp. 85–90).

Hendricks, G. J., & Gysling, D. L. (1994). Theoretical study of sensor-actuator schemes for rotating stall control. *Journal of Propulsion and Power*, *10*(1), 101–109.

Hoskin, R. F., Nett, C. N., & Reeves, D. E. (1991). Control configuration design for a mixed vectored thrust ASTOVL aircraft in hover. *AIAA Guidance, Navigation, and Control Conference* (pp. 1625–1642).

Hovd, M., & Skogestad, S. (1992a). Controllability analysis for unstable processes. In: J. D. Perkins (Ed.), *Preprints IFAC Workshop Interactions Between Process Design and Process Control* (pp. 49–54). Oxford: Pergamon Press.

Hovd, M., & Skogestad, S. (1992b). Simple frequency-dependent tools for control system analysis, structure selection and design. *Automatica*, *28*(5), 989–996.

Hovd, M., & Skogestad, S. (1993). Procedure for regulatory control structure selection with application to the FCC process. *A.I.Ch.E. Journal*, *39*(12), 1938–1953.

Isidori, A. (1995). *Nonlinear control systems* (3rd ed.). Berlin: Springer.

Joseph, B., & Brosilow, C. B. (1978). Inferential control of processes — Part I: Steady state analysis and design. *A.I.Ch.E. Journal*, *24*(3), 485–492.

Kammer, D. C. (1996). Optimal sensor placement for modal identification using system-realization methods. *Journal of Guidance, Control, and Dynamics*, *19*(3), 729–731.

Katsikas, S. K., Tsahalis, D., Manolas, D., & Xanthakis, S. (1995). A genetic algorithm for active noise control actuator positioning. *Mechanical Systems and Signal Processing*, *9*(6), 697–705.

Keller, J. P., & Bonvin, D. (1987). Selection of inputs for the purpose of model reduction and controller design. *Proceedings of IFAC World Congress* (pp. 209–214).

Kinnaert, M. (1995). Interaction measures and pairing of controlled and manipulated variables for multi-input multi-output systems: A survey. *Journal A*, *36*(4), 15–23.

Ko, B., Tongue, B. H., & Packard, A. (1994). A method for determining the optimal location of a distributed sensor/actuator. *Shock and Vibration*, *1*(4), 357–374.

Kosut, R. L., & Kabuli, M. G. (1995). On actuator-sensor selection in thermal processes. *Proceedings of IEEE Conference on Decision and Control* (pp. 2482–2483).

Kumar, S., & Seinfeld, J. H. (1978). Optimal location of measurements in tubular reactors. *Chemical Engineering Science*, *33*, 1507–1516.

Lee, A.-C., & Chen, S.-T. (1994). Collocated sensor/actuator positioning and feedback design in the control of flexible structure system. *ASME Journal of Vibration and Acoustics*, *116*(2), 146–154.

Lee, J. H. (1991). *Robust inferential control: A methodology for control structure selection and inferential control system design in the presence of model/plant mismatch*. Ph.D. thesis, California Institute of Technology.

Lee, J. H., Braatz, R. D., Morari, M., & Packard, A. (1995). Screening tools for robust control structure selection. *Automatica*, *31*(2), 229–235.

Lee, J. H., & Morari, M. (1991). Robust measurement selection. *Automatica*, *27*(3), 519–527.

Lee, Y. J., & Speyer, J. L. (1993). Zero locus of a beam with varying sensor and actuator locations. *Journal of Guidance, Control, and Dynamics*, *16*(1), 21–25.

Lim, K. B. (1997). Disturbance rejection approach to actuator and sensor placement. *Journal of Guidance, Control, and Dynamics*, *20*(1), 202–204.

Lin, X., Newell, R. B., Douglas, P. L., & Mallick, S. K. (1994). Application of output structural controllability to a thermally integrated distillation sequence. *International Journal of Systems Science*, *25*(9), 1437–1459.

Lin, X., Tade, M. O., & Newelll, R. B. (1991). Output structural controllability condition for the synthesis of control systems for chemical processes. *International Journal of Systems Science*, *22*(1), 107–132.

Luyben, W. L., Tyréus, B. D., & Luyben, M. L. (1998). *Plantwide process control*. New York: McGraw-Hill.

Mellefont, D. J., & Sargent, R. W. H. (1977). Optimal measurement policies for control purposes. *International Journal of Control*, *26*(4), 595–602.

Montazeri-Gh, M., Allerton, D. J., & Elder, R. L. (1996). Actuator placement for active surge control in a multi-stage compressor. *Proceedings of International Gas Turbine and Aeroengine Congress & Exhibition*, Paper no. 96-GT-241.

Moore, B. C. (1981). Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, *26*(1), 17–32.

Moore, C., Hackney, J., & Carter, D. (1987). Selecting sensor location and type for multivariable processes. *Proceedings of Shell Process Control Workshop* (pp. 291–308).

Morari, M. (1983). Design of resilient processing plants — III: A general framework for the assessment of dynamic resilience. *Chemical Engineering Science*, *38*(11), 1881–1891.

Morari, M. (1992). Effect of design on the controllability of chemical plants. In J. D. Perkins (Ed.), *Preprints IFAC Workshop Interactions Between Process Design and Process Control* (pp. 3–16). Oxford: Pergamon Press.

Morari, M., Arkun, Y., & Stephanopoulos, G. (1980). Studies in the synthesis of control structures for chemical processes. *A.I.Ch.E. Journal*, *26*(2), 220–231.

Morari, M., & Stephanopoulos, G. (1980a). Studies in the synthesis of control structures for chemical processes — Part II: Structural aspects and the synthesis of alternative feasible control schemes. *A.I.Ch.E. Journal*, *26*(2), 232–246.

Morari, M., & Stephanopoulos, G. (1980b). Studies in the synthesis of control structures for chemical processes — Part III: Optimal selection of secondary measurements within the framework of state estimation in the presence of persistent unknown disturbances. *A.I.Ch.E. Journal*, *26*(2), 247–260.

Müller, P. C., & Weber, H. I. (1972). Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems. *Automatica*, *8*(3), 237–246.

Narraway, L. T., & Perkins, J. D. (1993). Selection of process control structure based on linear dynamic economics. *Industrial and Engineering Chemistry Research*, *32*(11), 2681–2692.

Nett, C. N. (1989). A quantitative approach to the selection and partitioning of measurements and manipulations for the control of complex systems. *Presentation at Caltech Control Workshop*.

Nett, C. N., & Manousiouthakis, V. (1987). Euclidean condition and block relative gain: Connections, conjectures, and clarifications. *IEEE Transactions on Automatic Control*, *32*(5), 405–407.

Nijmeijer, H., & Van der Schaft, A. (1990). *Nonlinear dynamical control systems*. New York: Springer.

Nishida, N., Stephanopoulos, G., & Westerberg, A. W. (1981). A review of process synthesis. *A.I.Ch.E. Journal*, *27*(3), 321–351.

Norris, G. A., & Skelton, R. E. (1989). Selection of dynamic sensors and actuators in the control of linear systems. *Journal of Dynamic Systems, Measurement, and Control*, *111*(3), 389–397.

Reeves, D. E. (1991). *A comprehensive approach to control configuration design for complex systems*. Ph.D. thesis. Georgia Institute of Technology.

Reeves, D. E., Nett, C. N., & Arkun, Y. (1991). *Control configuration design toolbox for use with* MATLAB.

Rhodes, C., & Morari, M. (1995). Determining the model order of nonlinear input/output systems directly from data. *Proceedings of American Control Conference*, vol. 3 (pp. 2190–2194).

Rivera, D. E. (1989). A structured singular value test for variable selection and pairing. *Proceedings of American Control Conference*, vol. 1 (pp. 560–565).

Rivera, D. E., Desai, A., Beaumariage, T., & Roberts, C. (1993). A knowledge-based framework for controller structure selection in intelligent process control. *Proceedings of American Control Conference*, vol. 2 (pp. 1890–1894).

Roh, H.-S., & Park, Y. (1997). Actuator and exciter placement for flexible structures. *Journal of Guidance, Control, and Dynamics*, *20*(5), 850–856.

Ross, R., & Swartz, C. L. E. (1997). Inclusion of model uncertainty in a computational framework for dynamic operability assessment. *Computers & Chemical Engineering*, *21*, S415–S420.

Ruckman, C. E., & Fuller, C. R. (1995). Optimizing actuator locations in active noise control systems using subset selection. *Journal of Sound and Vibration*, *186*(3), 395–406.

Russell, L. W., & Perkins, J. D. (1987). Towards a method for diagnosis of controllability and operability problems in chemical plants. *Chemical Engineering Research and Design*, *65*, 453–461.

Samar, R., & Postlethwaite, I. (1994). Multivariable controller design for a high performance aero-engine. *Proceedings of International Conference on CONTROL*, vol. 2 (pp. 1312–1317).

Schmitendorf, W. E. (1984). An exact expression for computing the degree of controllability. *Journal of Guidance, Control, and Dynamics*, *7*(4), 502–504.

Seborg, D. E., Edgar, T. F., & Mellichamp, D. A. (1989). *Process dynamics and control*. New York: Wiley.

Seto, K., & Mitsuta, S. (1994). A new method for making a reduced-order model of flexible structures using unobservability and uncontrollability and its application in vibration control. *JSME International Journal, Series C*, *37*(3), 444–449.

Sidi, M. (1997). Gain-bandwidth limitations of feedback systems with non-minimum-phase plants. *International Journal of Control*, *67*(5), 731–743.

Simon, J. S., Valavani, L., Epstein, A. H., & Greitzer, E. M. (1993). Evaluation of approaches to active compressor surge stabilization. *ASME Journal of Turbomachinery*, *115*, 57–67.

Skogestad, S., & Morari, M. (1987a). Effect of disturbance directions on closed-loop performance. *Industrial and Engineering Chemistry Research*, *26*(10), 2029–2035.

Skogestad, S., & Morari, M. (1987b). Implications of large RGA elements on control performance. *Industrial and Engineering Chemistry Research*, *26*(11), 2323–2330.

Skogestad, S., & Morari, M. (1988). Some new properties of the structured singular value. *IEEE Transactions on Automatic Control*, *33*(12), 1151–1154.

Skogestad, S., & Postlethwaite, I. (1996). *Multivariable feedback control: Analysis and design*. Chichester, UK: Wiley.

Soroush, M. (1996). Evaluation of achievable control quality in nonlinear processes. *Computers & Chemical Engineering*, *20*(4), 357–364.

Trierweiler, J. O. (1997). *A systematic approach to control structure design*. Ph.D. thesis, Universität Dortmund.

Trierweiler, J. O., & Engell, S. (1997). The robust performance number: A new tool for control structure design. *Computers and Chemical Engineering*, *21*, S409–S414.

Tzouanas, V. K., Luyben, W. L., Georgakis, C., & Ungar, L. H. (1990). Expert multivariable control. 1. Structure and design methodology. *Industrial and Engineering Chemistry Research*, *29*(3), 382–389.

Van de Wal, M. (1995). *A survey of methods for control structure design*. Technical Report WFW 98.007. Faculty of Mechanical Engineering, Eindhoven University of Technology.

Van de Wal, M. (1998). *Selection of inputs and outputs for control*. Ph.D. thesis, Eindhoven University of Technology.

Van de Wal, M., & De Jager, B. (1995). Control structure design: A survey. *Proceedings of American Control Conference*, vol. 1 (pp. 225–229).

Van de Wal, M., & De Jager, B. (1996). Selection of sensors and actuators for an active suspension control problem. *Proceedings of IEEE International Conference Control Applications* (pp. 55–60).

Van de Wal, M., & De Jager, B. (1997). Selection of sensors and actuators based on a sufficient condition for robust performance. *Proceedings of European Control Conference*.

Van de Wal, M., & De Jager, B. (1998). Input/output selection for robust control. *Proceedings of UKACC International Conference on CONTROL*, vol. 2 (pp. 1023–1028).

Van de Wal, M., Philips, P., & De Jager, B. (1998). Actuator and sensor selection for an active vehicle suspension aimed at robust performance. *International Journal of Control*, *70*(5), 703–720.

Van de Wal, M., Willems, F., & De Jager, B. (1997). Selection of actuators and sensors for active surge control. *Proceedings of IEEE International Conference Control Aapplications* (pp. 121–126).

Van der Schaft, A. J. (1996). *$\mathscr{L}_2$-gain and passivity techniques in nonlinear control*. Heidelberg: Springer.

Vander Velde, W. E., & Carignan, C. R. (1984). Number and placement of control system components considering possible failures. *Journal of Guidance, Control, Dynamics*, *7*(6), 703–709.

Weber, R., & Brosilow, C. (1972). The use of secondary measurements to improve control. *A.I.Ch.E. Journal*, *18*(3), 614–623.

Weitz, O., & Lewin, D. R. (1996). Dynamic controllability and resiliency diagnosis using steady state process flowsheet data. *Computers & Chemical Engineering*, *20*(4), 325–335.

Wolff, E. A., Skogestad, S., Hovd, M., & Mathisen, K. W. (1992). A procedure for controllability analysis. In: J.D. Perkins (Ed.), *Preprints IFAC workshop interactions between process design and process control* (pp. 127–132). Oxford: Pergamon Press.

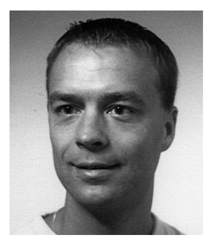
Xu, K., Warnitchai, P., & Igusa, T. (1994). Optimal placement and gains of sensors and actuators for feedback control. *Journal of Guidance, Control, and Dynamics*, *17*(5), 929–934.

Yu, C.-C., & Luyben, W. L. (1986). Design of multiloop SISO controllers in multivariable processes. *Industrial and Engineering Chemistry Process Design and Development*, *25*(2), 498–503.

Zafiriou, E., & Chiou, H.-W. (1996). On the dynamic resiliency of constrained processes. *Computers & Chemical Engineering*, *20*(4), 347–355.

Zhou, K., Doyle, J. C., & Glover, K. (1996). *Robust and optimal control*. Upper Saddle River, NJ: Prentice-Hall.

Bram de Jager received the Ir. degree from Delft University of Technology in 1983. Subsequently he was employed by Delft University of Technology and Stork Boilers. Currently he is at Eindhoven University of Technology, where he obtained his Ph.D. degree in 1992. His research interests are robust nonlinear control of mechanical systems, control structure design, and process control.



Marc van de Wal was born in 1970 in Eindhoven, The Netherlands. He received the Ir. degree in mechanical engineering from Eindhoven University of Technology in 1993. He then started working on a Ph.D. project on control structure selection. After receiving the Ph.D. degree in 1998, he joined the Philips Centre for Industrial Technology where he is currently working on high-precision electromechanical servo systems. His main research interests are robust control, gain-scheduling and LPV control, and advanced feedforward control.