

Please cite the following article at:

Sharifzadeh M (2013). Integration of process design and control: a review, *Chemical Engineering Research and Design*, 91 (12), 2515–2549 ([Link](#)).

Integration of process design and control: a review

Mahdi Sharifzadeh ^{a,1}

^a Centre for Process System Engineering (CPSE), Department of Chemical Engineering, Imperial College London,

Abstract

There is a large variety of methods in literature for process design and control, which can be classified into two main categories. The methods in the first category have a sequential approach in which, the control system is designed, only after the details of process design are decided. However, when process design is fixed, there is little room left for improving the control performance. Recognizing the interactions between process design and control, the methods in the second category integrate some control aspects into process design. With the aim of providing an exploration map and identifying the potential areas of further contributions, this paper presents a thematic review of the methods for integration of process design and control. The evolution paths of these methods are described and the advantages and disadvantages of each method are explained. The paper concludes with suggestions for future research activities.

Keywords

Integrated design and control, Sequential design and control, Control structure selection, Simultaneous optimization of a process and its controllers, inversely controlled process model, controllability.

¹ Correspondence concerning this article should be addressed to Mahdi Sharifzadeh; email: mahdi@imperial.ac.uk; address: Department of Chemical Engineering, Imperial College London, South Kensington Campus, London, UK, SW7 2AZ.

1. Introduction

This paper presents a thematic review of the relevant research into process design and control as a starting point and exploration map for the researchers in the field. In addition, the paper aims at encouraging the industrial application of these methods and identifying the challenging research frontiers with great potential impacts. This paper is organized in three parts. The first part discusses the incentives and barriers for integrated design and control and presents the industrial perspective about the subject. The second part provides the review of the research in the field. There are two categories of the methods. The methods in the first category have a sequential approach in which the process is designed first, and then the design of its control system is decided. However, recognizing the interactions between process design and control, the methods in the second category integrate process design and control. The third part of the paper provides summary and discussions of the reviewed methods and suggests future research activities.

2. Incentives for integrated design and control

The incentives for integrated design and control can be attributed to:

1. Shared decision-making domains,

Since the dynamic performance of a process strongly depends on its design, the decision-making domains of process and control engineers overlap, (Stephanopoulos and Reklaitis, 2011).

2. Conflicts and competitions between economic and controllability objectives,

The common perception is that process design is dominated by steady-state economic measures (e.g., total annual costs). However, many researchers have recognized the conflicts and competitions between economy and controllability of chemical processes. For instance, Luyben (2004) gave a list of examples where improving controllability conflicts with process economy.

In order to achieve high energy efficiency, thermodynamically reversible processes are favourable (i.e., no entropy production). However, reversible processes require negligible driving forces, (e.g. temperature difference in a heat exchanger). These driving forces are crucial for control systems to be able to reject disturbances or switch between steady states. Therefore, isolated decision-making for process and control design would result in, if not infeasible, a sub-optimal solution.

3. Characteristics of modern processes,

Modern chemical processes employ less in-plant inventories and they are highly integrated. These processes operate near operational constraints and should meet a larger variety of product specifications. Consequently, the perception of the role of control systems has changed to an integrated element of business planning in order to simultaneously ensure feasibility and optimality of process operation, (Edgar 2004; Stephanopoulos and Reklaitis, 2011).

4. Enabling skill-sets of process systems engineering (PSE),

Recent advances in mathematical skills of process systems engineers have equipped them with a portfolio of analysis tools (modelling, optimization, identification, diagnosis, and control) which enables them to consider the plant-wide interactions of process design and control, (Stephanopoulos and Reklaitis, 2011).

3. Industrial perspective

Industrialists had recognized the benefits of integrated process and control design even before academic researchers. Page Buckley (1964) was among the pioneer industrial engineers who recognized the importance of integrated design and control. He achieved this integration by transferring to Design Division of DuPont's Engineering Department and coordinating the efforts of process and instrumentation engineers. However, despite the large variety of methods developed thereafter, the industrial practice has conservatively maintained its traditional practice to design control systems for individual unit operations. For example, in Eastman Chemical Company, the procedure for designing a control system is still to set the throughput by the feed flowrate and then designing the control systems for individual units, sequentially, (Downs and Skogestad, 2011). The barriers against commercialization of the integrated approach are:

1. Different mind-sets,

Control engineers and process engineers usually have different mind-sets and for cultural reasons, it is difficult to encourage the integrated approach, (Downs and Skogestad, 2011).

2. Simplicity requirements,

Industrial incentives for simplicity and conceivability of control systems discourage the application of highly complex control systems such as real-time optimizations, (Downs and Skogestad, 2011).

3. Modelling efforts and computational costs,

Developing rigorous models and controllability analysis during the design stage can be time-consuming and expensive and requires a high level of expertise, (Chachuat, 2010; Downs and Skogestad, 2011).

The aforementioned barriers suggest that efficient methodologies are needed in order to capture the interactions between process design and control. Such methodologies should be able to systematically manage the conceptual as well as numerical complexities of the problem and encourage large-scale industrial applications.

4. Overview of the research in the field

The hierarchical tree in Fig. 1 gives an overview of the research in the field, and serves as a roadmap for the subsequent sections. It consists of two main branches. The methods in the left branch have a sequential/iterative approach in which the process is designed first and then a control system is designed for that process. However, the methods in the right branch have an integrated approach in which the effects of the process design on the control performance are also considered. The nodes are numbered and will be referenced in the subsequent sections. Other reviews of integrated design and control are presented by Sakizlis, et al. (2004), Seferlis and Georgiadis (2004), Ricardez-Sandoval, et al. (2009a) and Yuan, et al. (2011, 2012).

The review part of this paper is organized as follows. Firstly, the sequential approach for process design and control (the left branch in Fig. 1) is discussed and reviewed. In this branch, the process insights and heuristics, developed over decades of engineering practice enable conceptual as well as temporal and spatial decomposition of the problem. Another important decomposition technique is based on causality analysis, as discussed later.

Furthermore, the interactions between design and control strongly depend on the characteristics of the elements of control systems, i.e., controllers and control structures. Here, the focus is on the degree of centralization, the economic implications of set-point policies (e.g., self-optimizing control strategy), and the desired properties of controlled and manipulated variables. Furthermore, the causes of control imperfection are discussed because they limit process controllability. They are (1) interactions between control loops, (2) the constraints on manipulated variables, (3) model uncertainties and disturbance scenarios, (4) right-half-plane zeros, and (5) time delays. Based on the causes of control imperfection, a variety of methods is developed, which characterizes process controllability from different perspectives. Different definitions of operability, switchability, observability and controllability are presented and the methods for quantification of control imperfection and sensor placement based on process observability are reviewed briefly. The last methods on the left branch are based on passivity/dissipativity properties which characterize stability and controllability of the individual elements of a decentralized control system.

Please cite the following article at:

Sharifzadeh M (2013). Integration of process design and control: a review, *Chemical Engineering Research and Design*, 91 (12), 2515–2549 ([Link](#)).

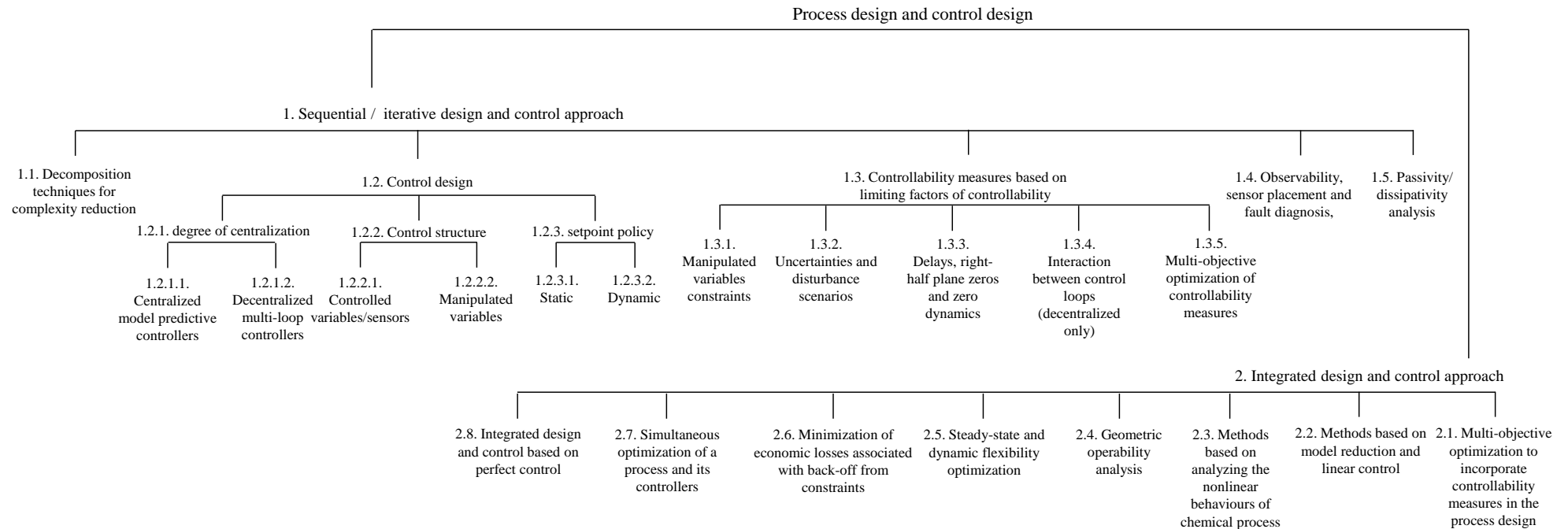


Fig. 1. An overview of the research in the field

Please cite the following article at:

Sharifzadeh M (2013). Integration of process design and control: a review, *Chemical Engineering Research and Design*, 91 (12), 2515–2549 ([Link](#)).

The disadvantage of the methods in the sequential approach is that they consider only individual subproblems (i.e., control structure selection, controllability analysis or controller design) and do not consider the interactions between them and process design. Some of the sequential methods have a qualitative approach and some others have yes/no or evaluation/ranking attitudes. The incentives to integrate controllability and control performance criteria into process design have motivated new studies which are shown on the right branch of Fig. 1. One way forward is to incorporate controllability and economic objective functions into a multi-objective optimization framework. Other researchers focused on reducing the first principles models to a linear model and applying the measures used in linear control theory. However, the disadvantage of linear methods is that the solution is only valid locally. Therefore, a group of researchers studied the undesirable nonlinear behaviour of chemical processes such as steady-state multiplicity. Other relevant research activities aim at quantification of the extent process nonlinearity. In addition, a nonlinear process model can be applied in order to map the available inputs into the output space and determine whether the process operation remains feasible for the expected disturbance scenarios. This idea resulted in the geometrical methods for operability analysis. Alternatively, flexibility analysis can be conducted using optimization. The early versions of flexibility optimization were based on a steady-state formulation, which identifies whether for a range of the values of uncertain variables, the process operation is feasible. Later, flexibility optimization was extended to include controllers and dynamic operability. Other researchers suggested minimization of the economic losses associated with disturbances. These losses were formulated in term of the required back-off from active constraints to ensure a feasible operation. In addition, the advancement of computational tools and optimization algorithms encouraged the researchers to optimize the process and controllers simultaneously. However, the resulted mathematical formulation is very large and limited to a certain type of controllers. The final part of this part of the paper studies the solution algorithms for the aforementioned optimization methods. The features of interest are MINLP and MIDO algorithms, simulation-optimization programming and global optimization.

Table 1 provides a representative sample of the research activities in the field, corresponding to the abovementioned methods. This table illustrates that researchers have applied a wide-spectrum range of methods and criteria for decision-making, which varies from steady-state nonlinear open-loop analysis to mixed integer dynamic optimization methods. In addition, the research focuses vary from individual elements of the problem such as controlled variables to stability of the whole process network. Another feature of interest is the various methods for control design which varies from self-optimizing control to optimizing a superstructure of proportional integral controllers and to advanced model predictive control systems. Finally and most importantly, the researchers have addressed a very diverse array of case studies including heat-integrated processes, processes with recycle streams, processes with potential steady-state multiplicities, and processes with discontinuous and periodic operations. The details of these methods will be presented later in this paper.

Table 1. Representative studies from different methods for process design and control

Author(s)	Optimization/ mathematical method	Contribution, merit, approach	Control design method	Case study
Hori and Skogestad (2008)	Maximum gain rule	Optimal linear combination of measurements as a controlled variable	self-optimizing control	Distillation column
Kariwala and Cao, (2010a)	Bidirectional branch and bound	Fast convergence, Local, average loss minimization	self-optimizing control	Distillation column
Rojas, et al. (2009)	Network representation of process, dissipativity analysis	Determination of the open-loop plant-wide, stabilizability and achievable disturbance attenuation	Open-loop	Heat Exchanger Network (HEN)
Georgakis and Li (2010)	Steady-state analysis	Incorporating design of experience for geometric analysis of steady-state operability	Open-loop	Tennessee Eastman process
Malcolm, et al. (2007)	Sequential integration	embedded control design- steady-state and dynamic flexibility optimization	Sequential Least Square (identification), Kalman filter (observation), LQR (regulation)	Polymerization reactor, binary distillation column
Ricardez-Sandoval, et al. (2011)	Identification of an uncertain model, Structured Singular Value (SSV) analysis	Avoiding computationally expensive dynamic modelling	PI control structure by Ricker (1996)	Tennessee Eastman process
Asteasuain, et al. (2006)	Decomposition technique based on sequential integration using gPROMS/gOPT	Systematic optimization of initial and final steady states, and the grade-transitions path	Optimization of a super structure of a multivariable feedforward–feedback controllers and their tuning parameters	Grade transition of a CSTR for styrene polymerization with potential steady-state multiplicity
Sakizlis, et al. (2004)	variant-2 of the generalized benders decomposition (GBD) algorithm	Multi-parametric programming, dynamic flexibility optimization	Explicit multi-parametric model predictive controller	Distillation column
Khajuria and Pistikopoulos (2011)	Partial differential algebraic (PDAE) equations, gPROMS	Multi-parametric programming, model identification	Explicit multi-parametric model predictive controller	Pressure swing adsorption (PSA) with periodic and discontinuous operations
Flores-Tlacuahuac and Biegler, (2007)	Full discretization	Comparison of the methods for translating the generalized disjunctive programming (GDP) formulation to convex and non-convex MINLP formulations	A superstructure of PI controllers	Two series heat-integrated CSTRs

4.1. Decomposition techniques for complexity reduction

The following subsections discuss the methods of the first node 1.1 on the left branch of Fig. 1, which concern decomposition techniques for complexity reduction. The fact that the process design problem needs to be resolved and decomposed into more manageable subproblems is not new in the area of process systems engineering. For example, Douglas (1988) presented a hierarchical view of a plant to make the problem of process design tractable. The methodology of Douglas employs different resolutions of the plant details, for example evaluation of the interactions of the plant and surroundings and then evaluation of the interactions of process components with each other and so on. The same is true for the control design problem and many authors suggested a hierarchical approach or a decomposition technique to reduce the problem complexities, as discussed in the following.

4.1.1. Complexity reduction based on process components

The early attempts for complexity reduction involved design of control structures for individual unit operations such as heat exchangers, reactors, and distillation columns and then interconnecting them in order to develop the overall plant-wide control structure. Here, engineering insights have to be employed to resolve the conflicts (e.g. two control valves on the same stream) that arise by adding individual control structures, (Ng, 1997). The inspiration for this approach is that comprehensive knowledge and experiences are available for controlling the major unit operations. Control design for special unit operations has been the subject of academic and industrial research, (e.g., Ward, et al. 2007; Ward, et al. 2010; Skogestad 1988, 2007; Hori and Skogestad, 2007).

Although the approach based on combining unit-wise control structures for individual unit operations ignores the plant-wide interactions among them, still this method has wide industrial applications. Downs and Skogestad (2011) attributed this practice to the “*overriding issues of reliable operation*”. This is because unit-wise control systems are simple and understandable to operators and plant engineers, and any malfunctioning unit operation can be treated without a need for intervention of control experts.

A criticism about the unit-wise approach is that combining the optimal control structures of individual unit operations does not guarantee the optimality of the overall plant-wide control structure. In addition, the heuristic methods used for eliminating the arising conflicts become more and more complicated and impractical as the number of process components increases, (Kookos 2001; Ng 1997).

4.1.2. Complexity reduction based on temporal decomposition

Temporal decomposition is another strategy to reduce the complexities of control structures. It employs differences in various time scales in which the control structure is performing.

Buckley (1964) recognized that control systems have a high frequency control layer for quantity control (material balance) and a low frequency control layer for quality control (e.g. specifications of products). As another example, it is well-known that in multi-loop control systems, interactive loops with a significant difference in their time constants may demonstrate a decoupled performance, and can operate separately, (Ogunnaike and Ray 1994).

Similarly, Morari et al., (1980a; 1980b; 1980c) categorized a control system into regulatory and optimizing parts. Those parts, which are responsible for regulation of the process, handle fast disturbances with a zero expected value in long-term. However, longstanding disturbances with significant economic effects are treated by the optimizing control systems.

4.1.3. Complexity reduction based on prioritization of control objectives

Several researchers attempted to reduce the complexities of the problem by prioritizing control objectives in order to decompose control structures into smaller parts, so each part pursues an individual objective. For example, Luyben (1996) presented a survey of the control structures developed for the Tennessee Eastman problem (a benchmark problem presented by Downs and Vogel 1993). He discussed the pros and cons of his own solution in addition to a list of other schemes such as those presented by Lyman and Georgakis (1995), McAvoy and Ye (1994), and Ricker (1996).

Luyben (1996) argued that different control structures developed for the Tennessee Eastman problem are the results of different rankings of the control objectives: “...diversity of structures is a very nice example of one of the basic process control principles that says that the “best” control structure depends on the control objectives”. In the following a brief review of these methods is presented.

McAvoy, and Ye (1994) suggested considering the overall mass balance through control of the flowrates, first. Then, the energy balances must be regulated by controlling temperatures and pressures. Later, the product quality and component mass balances are considered. Finally, the remaining degrees of freedom and setpoints of the regulatory control layer are employed for optimizing the operational costs.

The tiered framework approach suggested by Price, et al. (1993, 1994) firstly meet the targets of overall inventory and throughput regulations. Then, product specifications are treated. Later, operational constraints are considered, leaving the optimal operation to be the last target. They called their methodology a “direct descendant” of the Buckley’s method.

Ponton and Laing (1993) recommended developing the control structure for controlling the flowrates of products and feed first. Then, recycle flow must be regulated and the compositions of intermediate streams should be treated. Energy and temperature stabilization are the next targets and finally inventory control will be addressed.

Luyben, et al. (1997) proposed a framework for control structure selection in which firstly the control objectives and available degrees of freedom are determined. Then, energy management control system is designed. The decisions regarding control of the production flowrate are made and the product quality specifications and safety constraint satisfaction are ensured. It must be checked that the overall mass balance will be met for all the components. Then, the control systems of individual unit operations are designed. Finally, the remaining manipulated variables are assigned for optimizing the economic objective or dynamic controllability.

Larsson and Skogestad (2000) and Skogestad (2004a) developed an iterative top-down/bottom-up algorithm for control structure selection. The design approach in the top-down direction features steady-state economic analyses such as meeting the operational objectives, optimizing the process variables for important disturbances and determining active constraints with emphasize on throughput/efficiency constraints. However, the bottom-up design concerns dynamic issues such as designing the control structure for the regulatory layer, paring/partitioning the manipulated and controlled variables, and designing the supervisory control layer.

As discussed also by Edgar (2004), in evolution of the above methods, the priorities of the objectives have been reversed. In the early approaches, the control system was simply a tool to achieve the predetermined goals of production, which were set in the process design stage. The operation personnel did not think of the control system as an optimization tool to improve profitability of the process. Therefore, economic optimization had the lowest priority. However nowadays, business planning of process industries has become online and much less limited by the early decisions at the design stage. Consequently, the new control systems have also inputs in terms of economic parameters and translate them into operational decisions. This has encouraged designers to consider the highest priority to process profitability and the roles of other control tasks are to realize the targeted economic objectives.

4.1.4. Complexity reduction based on the production rate and the inventory control systems

The process inventories refer to gaseous, liquid, and solid materials accumulated within the process. Since inventory control systems have a dynamic nature and do not appear in a steady-state analysis, they have received special attentions in literature. Furthermore, inventory control has a priority in control structure design, because many instability modes such as emptying/overflowing of vessels or flooding/weeping of distillation columns are related to the inconsistencies or failures of inventory control systems. In addition, modern process plants tend to have less material inventories due to efficiency, safety and environmental considerations, which makes the control of their inventories more challenging. Therefore, developing general rules and methods which enable design of inventory

control systems without the aid of costly rigorous dynamic models is highly desirable. In the following, a brief review of the corresponding methods is presented.

Buckley (1964) emphasized the requirement for consistency of the flow controls, upstream and downstream of the throughput manipulation point (TMP). He suggested that in order to develop a consistent control structure, flows must be controlled in the opposite direction at the upstream of the throughput manipulation point and in the same direction at the downstream of this point.

Later, Price, et al. (1993, 1994) emphasized the existence of a primary path from feeds to products in most chemical processes. They suggested that the inventory control should be designed in the direction of the flow if the feed flowrate is chosen to be the throughput manipulation point and in the opposite direction of the flow if the product flowrate is chosen as the throughput manipulation point and in general radiates from the throughput manipulation point. This requirement is shown in Fig. 2 for a series of liquid inventories with throughput located inside the process.

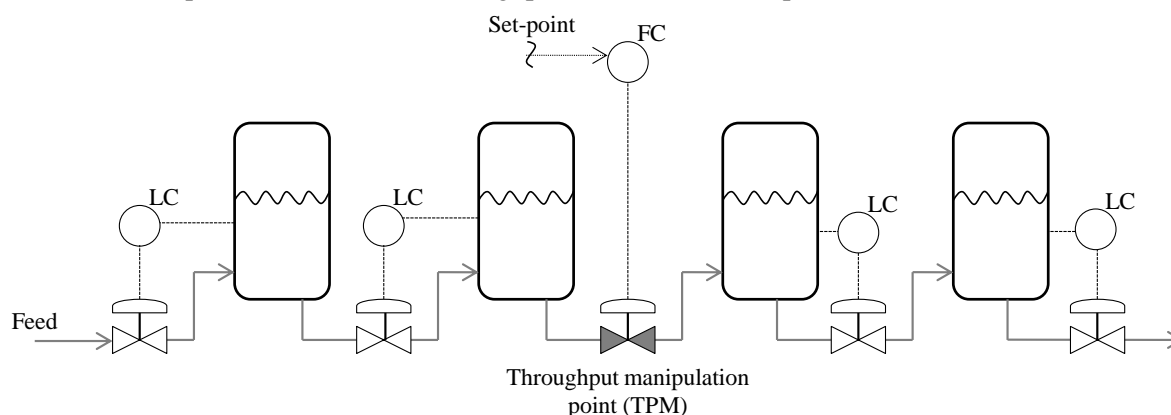


Fig. 2. Design of an inventory control system; the inflows are used for design of inventory control in the upstream of the throughput manipulation point. However, the outflows are used in the downstream of this point.

In a series of articles, Luyben and his co-workers (Luyben 1993a,b,c; Tyreus and Luyben 1993; Luyben 1994; Luyben, M. L., and Luyben, W. L., 1995), using examples of reaction-separation processes, explained how reaction kinetics and economic factors might result in different control structures. They recommended one flow control in the liquid recycle loop, but setting gas recycle at the maximum circulation rate. It is notable that the effects of recycle streams are not limited to material inventories, and energy inventories are also important. Luyben, et al. (1999) using the example of an exothermic reactor, showed that positive feedback of energy could lead to the loss of control action and may pose the risk of runaway reactions.

Aske and Skogestad (2009a, b) investigated the consistency requirements for inventory control systems. Their suggested rules can be summarized to firstly assign an inflow or outflow controller to each inventory and secondly to check whether inventory of each component is consistently regulated by at least a degree of freedom or a chemical reaction. Each phase inventory also needed to be controlled by the inflow or outflow or via phase change.

Recent studies (Skogestad 2004a; Aske and Skogestad 2009a, b; Ashe 2009; Downs and Skogestad 2011) have focused on the relation of inventory control and profitability. Chemical processes can be classified according to which constraints become active earlier, during economic optimization: (i) throughput constraints or (ii) efficiency constraints. In the case of new plants, economic objectives are often driven by optimizing the efficiencies regarding reaction yields, waste treatment, and energy consumptions. Therefore, after the optimal production rate is reached, any change in the throughput will result in economic losses and is treated as a disturbance. Conversely, when there are economic incentives to increase the production rate, for example because of high demand or high price of the products, the throughput constraints become active before the efficiency constraints. Therefore, in the second scenario, the process operation will be constrained by the throughput bottlenecks. The instances of these capacity constraints are limitations in the liquid flow to a vessel, the pressure difference of a distillation column or the temperature constraint of a reactor.

While dynamic degrees of freedom are assumed to have less economic importance, it has been shown that they are critical when process economy is constrained by the maximum throughput. In this case, the losses of process throughput can be avoided by temporary reductions in the in-plant material inventories. Aske (2009c) studied two cases of a coordinated MPC and a ratio control structure to show how dynamic degrees of freedom (which apparently have no steady-state economic significance) can be employed to increase the economic profitability.

4.1.5. Complexity reduction based on causality analysis

Causality analysis using graph theory reduces the first principles model to a *signed directed graph* (SDG). A signed directed graph represents the causal relationship between variables of a system. Often a sign or a weighting factor is added to an arc to characterize the direction or the intensity of that causality relation. The advantage of this methodology is that it extracts only necessary information and makes the model interrogation easier than the equivalent first principles model.

The application of this method for fault detection in process industries has gained great interests. Maurya, et al. (2003a, 2003b, 2004) presented a review, including detailed evaluations of the advantages and disadvantages of the applications of these graphs for representing dynamic models. Yim, et al. (2006) and Thambirajah, et al. (2009) used signed directed graphs and connectivity matrices to extract causality relation from process topology. Then, they used these data for evaluation of the performance of control loops and disturbance propagation. In addition, transfer entropy was applied by Bauer, et al. (2004, 2007) as a probabilistic tool to extract causal relationship between process variables from plant data. Hangos and Tuza (2001) applied the signed directed graphs for optimal control structure selection in a decentralized control system. They demonstrated a one to one correspondence between linearized state space model and the weighted digraph. They used the graph-based method of maximum weight matching for determining the best control structure. Similarly, Bhushan and Rengaswamy (2000) applied signed directed graphs for designing sensor networks. The disadvantage of signed directed graphs is that they cannot describe time propagation properties. Recently, Fan and De-yun, (2007) applied dynamic signed directed graphs in which time parameters are considered for the branches of the graph. They addressed the problem of optimal sensor location for fault detection and diagnosis. Furthermore, signed directed graphs are often built bases on process connectivity described by piping and instrumentation diagrams and need validation. Yang et al., (2012) proposed two methods for validation of the extracted signed directed graphs, based on transfer entropy and cross correlation. Similarly, Maurya, et al., (2007) and Dong et al., (2010) combined qualitative trend analysis with signed directed graphs in order to improve the diagnostic resolution of their methods.

Daoutidis and Kravaris, (1992) introduced the concept of relative order as a causal measure for selection of control structures. The relative order is based on differential geometry and can be defined as the number of times that a controlled variable should be differentiated in order to generate an explicit relationship between that controlled variable and a manipulated variable. In other words, relative order represents the initial sluggishness of the response of the controlled variable to the corresponding manipulated variable. Similarly, Soroush (1996) employed similar concept for selection of controlled variables based on the relative order of these variables with respect to disturbances.

4.2. Control design

This section discusses control design. The corresponding node in the hierarchical tree of Fig. 1 is node 1.2 on the left branch. In this section, temporal and spatial decentralizations of controllers are discussed. In addition, conventional multi-loop controllers and their counterparts, i.e., model predictive controllers are explained. This section will provide supporting arguments for the following sections where the properties of control structures, implications of setpoint policies and interactions between control loops are discussed.

4.2.1. Degree of centralization

The degree of centralization can be defined as the level of independency of individual controllers within a control structure. Centralization can be spatial or temporal which is discussed in the following. In addition, the characteristics of multi-loop controllers (an example of decentralized controllers) and model predictive controllers (an example of centralized controllers) are reviewed in brief.

4.2.1.1. Degree of centralization: spatial

Rawlings and Stewart (2008) classified the control systems, with respect to their spatial degree of centralization, into four groups:

1. Centralized control structures in which a centralized controller employs a single objective function and a single model of the whole system for decision-making,
2. Decentralized control structures in which the controllers are distributed and the interactions between subsystems are totally ignored,
3. Communication-based control structures in which each distributed controller employs a model for its sub-process and an interaction model for communicating with other sub-systems. However, the distributed controllers have their own objective functions. The disadvantage of communication-based structures is that controllers with individual objective functions may compete rather than cooperate with each other and make the whole system unstable.
4. Cooperative control structures in which the distributed subsystems employ an objective function for the whole system, and the prediction of the last iteration of other controllers are available to each controller. The improvement is not in awareness of the local controllers from each other, but in the same objective function that is employed by all of them. This framework is plant-wide stable with no offset and by convergence of the control calculations provides centralized optimal decision.

The decision regarding the degree of centralization significantly influences the design of control structures. In conventional multi-loop control systems (an examples of decentralized controllers), the designer examines the alternative pairings between manipulated variables and controlled variables, often based on analysis of the interactions between candidate control loops. However, in model predictive control (MPC) systems (an examples of centralized controllers) these interactions are of no concern, because all manipulated and controlled variables are interconnected to each other through the control algorithm. However, neither an entirely decentralized control structure nor a fully centralized one is desirable, and it is often favourable to employ some degree of centralization which locates the control structure between these two extremes. The reason is that while a pure decentralized control structure does not necessarily ensures an optimal operation, there are many concerns regarding computational load, reliability, and the costs of implementation and maintenance of a large-scale centralized control structure.

Rawlings and Stewart (2008) also discussed that a fully connected communication strategy is unnecessary at least regarding plant stability. However, the penalty of reducing communications is synchronization of state calculations. In addition, reduction in communication between local MPCs causes problems in the systems with recycle streams (e.g. systems 1 and 2 in Fig. 3), because it requires iterative calculations or one subsystem must do the calculations for the others. Therefore, a hybrid communication strategy is recommended, in which a total communication scheme is considered for each recycle loop and a reduced communication scheme is considered for the rest of the process, (Rawlings and Stewart 2008).

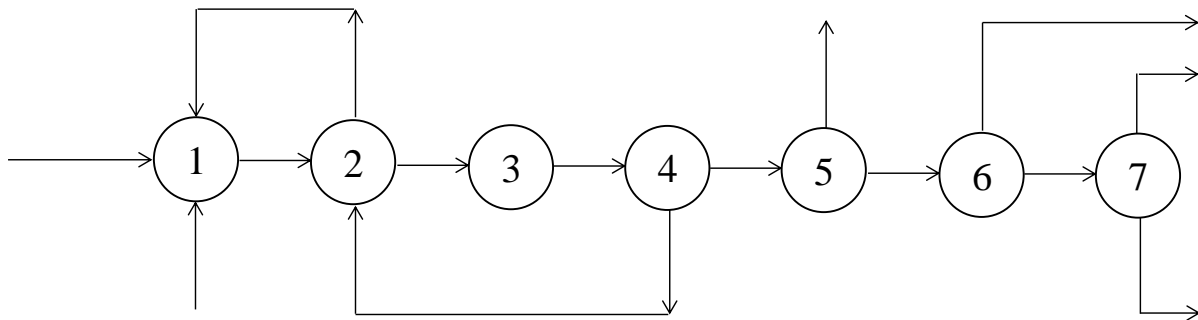


Fig. 3. Ethylene glycol flowsheet: (1) Feed tank, (2) preheater, (3) reactor, (4) evaporator, (5) light end columns, (6) mono ethylene glycol column, (7) higher glycol recovery, (adapted from Rawlings and Stewart, 2008).

4.2.1.2. Degree of centralization: temporal

The classification discussed in the last section suggests a spatial centralization. However, centralization of controllers can be temporal, as shown in Fig. 4 (adapted from Qin and Badgwell 2003). In the shown control structures, the decision-making process is decentralized vertically (top-down) through different time scales from days and weeks in the highest optimizing layer to seconds in the lowest regulating layer. While the left structure shows a decentralized control structure, the right structure suggests a higher degree of centralization.

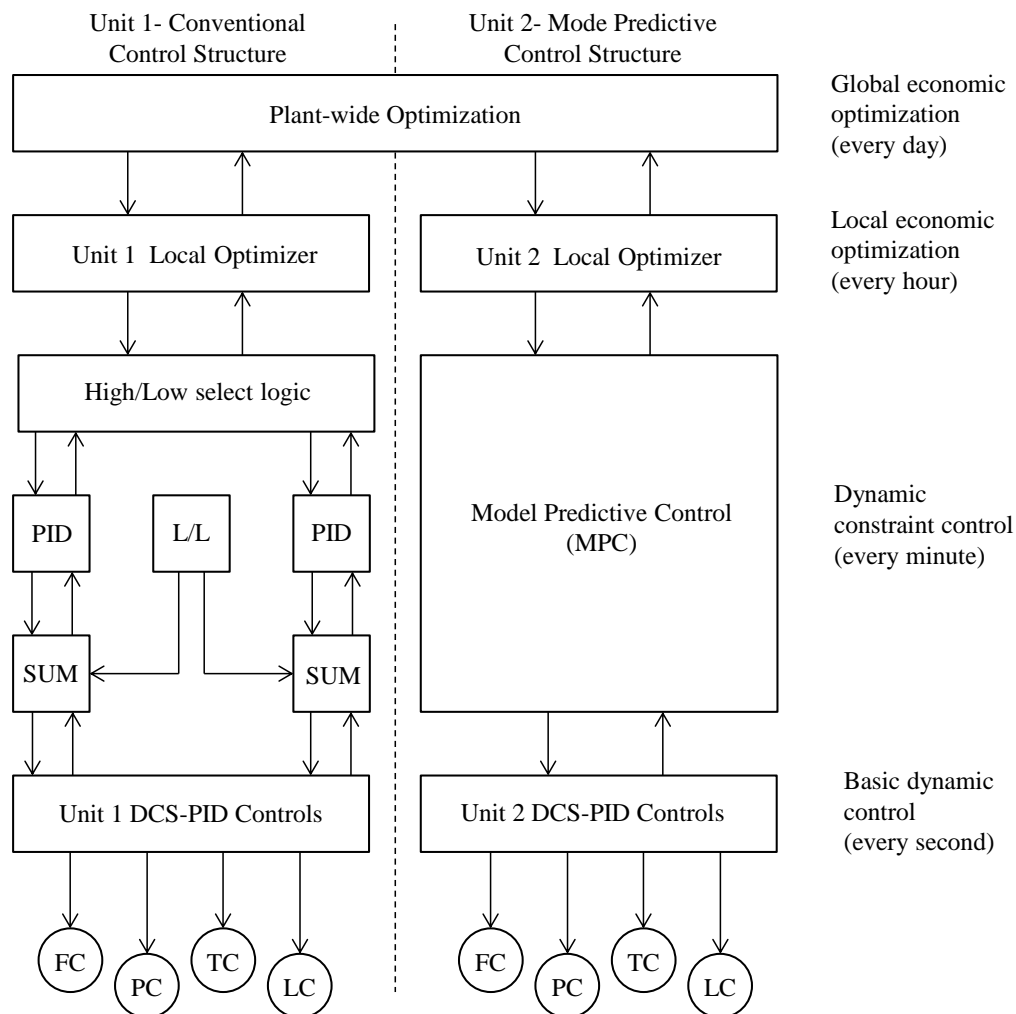


Fig. 4. Temporal decentralization of controllers. The hierarchy of conventional multi-loop and MPC structures are shown at the left and right respectively, (adapted from Qin and Badgwell 2003).

The top layer often employs a steady-state optimization for determining the setpoints. This information will be sent to the localized optimizers which may employ more detailed models and run more frequently. Detailed information will be sent to the constraint-handling control system which is responsible for moving the process from one constrained steady-state to another one while minimizing the violation of the constraints. In the right control structure, a model predictive controller is responsible for constraint handling, while in the left control structure, a combination of PIDs, lead-lag (L/L) blocks and logic-based elements are responsible for constraint handling. The regulatory layer which runs at much higher frequency, is responsible for maintaining the controlled variables at their setpoints, (Qin and Badgwell 2003).

Fig. 5 adapted from Harjunoski et al. (2009) shows the control system in a broader context which conforms to the *automation* paradigm. The lowest layer is responsible for process control including regulatory control systems, as well as monitoring and fault diagnosing systems. The middle layer is responsible for production scheduling, quality assurance and more advanced production control algorithms. On the top layer, the long-term production strategies are decided and the whole supply-chain including feedstock procurements, product warehousing, distributions and sales are coordinated. More details on automation can be found in ANSI/ISA-95 (2000, 2001, and 2005) standards which provide guidelines for the communication and information exchange between different sections of an enterprise.

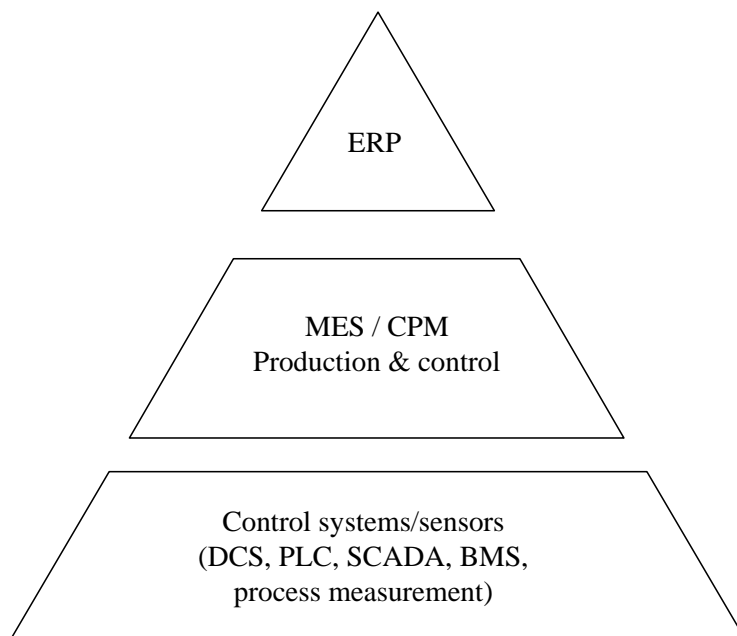


Fig. 5. Automation pyramid (adapted from Harjunoski et al. 2009).

4.2.1.3. Conventional multi-loop controllers

Chemical processes have some characteristics which make their control difficult. For example, when Qin and Badgwell (2003) were explaining the reasons for little impact of linear quadratic Gaussian (LQG)-based technologies on process industries (despite their success in electronics and aerospace areas), they emphasized that chemical processes are nonlinear, constrained, and multivariable systems and their behaviours change over the time (e.g. ageing of catalysts). By contrast, conventional multi-loop controllers are proved efficient in controlling chemical processes, because they have reliable operation and are understandable to plant people, (Downs and Skogestad 2011).

However, conventional multi-loop controllers have a significant drawback; i.e., leaving setpoints at constant values is a poor economic policy, because disturbances and the changes in economic parameters can change the optimal setpoints and in extreme, (e.g. moving bottlenecks) require control structure reconfiguration, (Downs and Skogestad 2011). The treatment of economic losses due to constant setpoint policy will be discussed later in this paper. Other drawbacks of multi-loop controllers are the required convoluted override logics for constraint handling, and interactions among control loops, (Stephanopoulos and Reklaitis, 2011).

4.2.1.4. Model predictive controllers

This section discusses model predictive controllers (MPCs) briefly. A detailed review of the common MPC technologies and their characteristics is presented by Qin and Badgwell, (2003) and Darby and Nikolaou (2012).

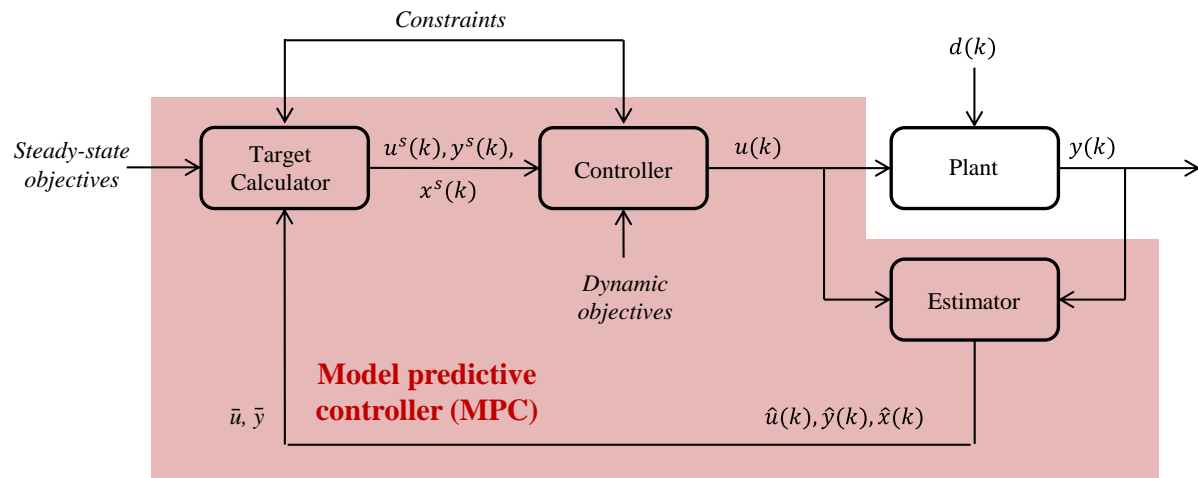


Fig. 6. The block diagram representation of an MPC system: estimator, target calculator, regulator, (adapted from Darby and Nikolaou 2012).

The concept is shown in Fig. 6 adapted from Darby and Nikolaou (2012). The estimator block enquires the manipulated and controlled variables and then using a model estimates the unmeasured states. Then, the target calculator calculates the target values of the manipulated and controlled variables. Finally, this information is used by a dynamic model (shown by the controller block) to bring the process from the current state to the targeted state. The outcomes of these calculations are the decisions regarding adjustment of the manipulated variables. Richalet, et al. (1978) emphasized that the economic advantages of model predictive control systems derive from manipulation of the setpoints by the target calculator rather than minimizing the variations of the controlled variables (i.e. controller error) using the dynamic model.

The capability for systematic constraint handling is another important advantage of MPC systems over multi-loop control systems. The modern MPC systems apply three types of constraint-handling methods. They are hard, soft and setpoint approximation constraint-handling methods. The hard constraints are those which are not allowed to be violated such as the constraints on the maximum, minimum, and the rate of the changes of the manipulated variables. The soft constraints (e.g. the constraints on some controlled variables) are permitted to be violated to some extent and their violations will be minimized by penalizing the objective function. Another way of handling soft constraints is the setpoint approximation method. In this method, a setpoint is assigned to a soft constraint and the deviations on both sides of the constraint are penalized. However, the penalty weights are assigned dynamically, so the penalty function becomes significant only when the constraint is likely to be violated, (Qin and Badgwell 2003).

4.2.2. Control structures

This section discusses control structures. A control structure consists of controlled variables (CVs) and manipulated variables (MVs). Manipulated variables, also known as process inputs, are selected from the available degrees of freedom with desired properties for performing a controlling action. Controlled variables are those process variables which are selected to be maintained constant at their setpoints by controllers. If direct measurement of a controlled variable is not possible then its value must be *inferred* or estimated from other process variables, (Qin and Badgwell 2003). These *inferential* controlled variables together with direct controlled variables are known as the measured variables. While selection of manipulated variables is the subject of degree of freedom analysis, controlled variables and their setpoints are strongly related to process profitability. The following subsections explore the characteristics of control structures and desirable properties of manipulated variables and controlled variables. The implications of controlled variables and setpoint policy for process profitability are also discussed.

4.2.2.1. Control structure reconfiguration

A comparison between the populations of manipulated variables and controlled variables provides insights about feasibility of a control problem. Fig. 7, adapted from Qin and Badgwell (2003), depicts the alternative scenarios. In the design stage, the population of manipulated variables often exceeds the population of controlled variables and the control problem is under-determined (right-hand side of Fig. 7). In this case, extra manipulated variables are available for economic optimization. During the process operation, the population of the manipulated variables may decrease for example because of activation of constraints, saturation of control valves, failures of control signals, or intervention of operation people, which make the control problem over-determined (left-hand side of Fig. 7), and consequently it becomes infeasible. The middle control problem in Fig. 7 represents a square problem with a deterministic solution. All these three scenarios may happen in the same control system. However, still it is expected to perform the best possible control action.

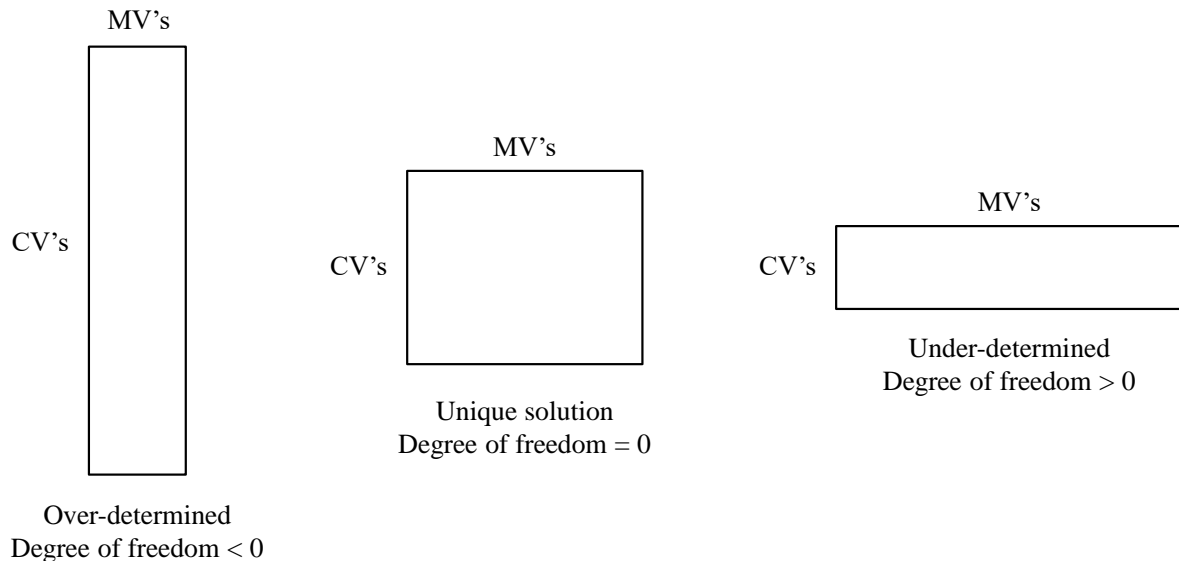


Fig. 7. Different configurations of a control structure, (Adapted from Froisy 1994).

For the case of conventional multi-loop control structures, drastic changes in economic parameters may necessitate control structure reconfiguration. These scenarios are mostly concerned with the movements in active constraints. An example of necessary control reconfiguration is when the inventory control structure and throughput manipulation point must be reconfigured due to movement of the economic bottleneck(s), (Aske 2009c). However, MPC systems are subject to dynamic changes in the dimension of the control problem during control execution. The reason is that the manipulated and controlled variables may disappear due to valve saturations, signal failures, or operator interventions in each control execution and return on the next one. These changes sometimes make

the control configuration over-determined and therefore perfect control (i.e., maintaining controlled variables at their desired values) would be infeasible. However, it is still desirable to have the best possible control action through the remaining manipulated variables. Unfortunately, these changes have a combinatorial nature and it is not possible to evaluate all of the alternative subspaces of a control problem at the design stage. Therefore, MPC systems have an online monitoring agent that is responsible for subproblem conditioning. The strategy is to meet the control objectives based on their priorities, (Qin and Badgwell 2003). In MPC systems in order to avoid saturation of the manipulated variables, their nominal values are treated as additional controlled variables with lower priorities. In addition, when a manipulated variable disappears from the control structure (e.g. because of operator intervention), it may be treated as a measured disturbance. Similarly, saturated valves are treated as one-directional manipulated variables. By contrast to manipulated variables, when a controlled variable is lost for instance because of signal failure or delay in measurements, the practical approach is to use the predicted value for it. However, if the faulty situation persists for an unreasonable number of execution steps, in some MPC algorithms, the contribution of the missing controlled variable will be omitted from the objective function, (Qin and Badgwell 2003).

4.2.2.2. Degrees of freedom analysis

Konada and Rangaiah (2012) presented a recent review of the methods for degree of freedom (DOF) analysis. Degrees of freedom can be evaluated as:

$$\text{DOF} = \text{number of unknown variables} - \text{number of independent equations} \quad (1)$$

However, in the context of control engineering, external variables such as disturbances also need to be considered, (Stephanopoulos 2003):

$$\text{CDOF} = \text{number of unknown variables} - (\text{number of independent equations} + \text{number of external variables}) \quad (2)$$

In which CDOF stands for control degree of freedom and concerns the number of manipulated variables. The above approach has been applied by Seider, et al., (2010) for a number of processes. However, for large processes counting all the equations and variables may not be practical and is prone to mistakes. In addition, the focus of CDOF is mostly extensive variables. This is because manipulated variables are in principle defined as the flowrates of energy and materials (e.g., control valves, pump speeds, electricity streams). Therefore, researchers tried to develop methodologies which do not require first principles modelling and still are able to accurately determine the available degrees of freedom. Dixon (1972) introduced the notion of boundary variables. These are the variables which cross the predefined boundaries of a system. Furthermore, steady-state control degrees of freedom, CDOF_{ss} , were distinguished from dynamic control degrees of freedom.

$$\text{CDOF}_{ss} = N_{bv} - N_{bes} \quad (3a)$$

$$\text{CDOF} = \text{CDOF}_{ss} + N_0 \quad (3b)$$

N_{bes} represents boundary equations and N_{bv} represents boundary variables. N_0 is the number of independent holdups. Equation (3b) suggests that CDOF_{ss} is a subset of CDOF. Later, Pham (1994) introduced the concept of output control degrees of freedom:

$$\text{CDOF}_{ss,out} = \sum_{k=1}^K (P_k S_k) + M + E - N \quad (4)$$

where K is the number of circuits (a circuit is a set of streams connected inside the process), S is the number of stream split, P is the number of phases in the output stream of a circuit, M is the number of influential variables (e.g. a control valve), E is the number of energy streams, N is the number of phase constraints. Pham (1994) argued that the important implication of Equation (4) is that degrees of freedom analysis can be conducted assuming a single-component system. The focus of Pham's method was the operational degrees of freedom, i.e., the manipulated variables available when the process is built and is in operation. Konada and Rangaiah (2012) showed that Pham's method may result in wrong results because it assumes that the places of control valves are known in advance. However, Pham's method was a step forward; because it recognized that in order to evaluate the correct number of degrees of freedom it is not needed to write the equations. In an independent study, Ponton (1994) derived the general equations:

$$\text{CDOF}_{ss} = n_i + n_e + n_o - P + 1 \quad (5)$$

where n_i is the number of inlet material streams, n_o is the number of outlet material streams, n_e is the number of energy streams and P is the number of phases. However, equation (5) is of limited practicality because it is not possible to manipulate all the streams simultaneously. This issue has been addressed by Konda, et al. (2006) and Vasudevan, et al. (2008) who recently proposed and examined a method which is flowsheet-oriented, and requires only the information of process flow diagrams and general knowledge of important unit operations. The idea is to identify the streams that are redundant or restrained from being manipulated. Then, this number can be subtracted from the total number of streams in order to identify the available degrees of freedom. They argued that the restraining streams are mostly the characteristics of individual unit operations and not the process flowsheet and therefore, once they are calculated they can be used in any complicated process flowsheet. They proposed the following correlation:

$$\text{CDOF} = N_{\text{streams}} - \sum_{\text{all the units}} (N_{\text{restraining}}) - N_{\text{redundant}} \quad (6)$$

In above, N_{streams} is the total number of material and energy streams, $N_{\text{restraining}}$ is the number of streams that cannot be controlled, and $N_{\text{redundant}}$ is the number of streams that are not efficient to be manipulated (e.g., a material stream with small pressure drop). They further classified restraining streams based on the units with and without material holdups. The number of restraining streams is equal to total independent material balances in units without holdups. This is because each mass balance imposes a constraint and reduces one degree of freedom. However, in the case of unit operations with material inventories, there is additional flexibility and all the streams can be manipulated provided that not all of them are used for controlling extensive variables. Therefore, the number of restraining streams is equal to the number of independent material balances which are not associated with any mass inventory. Since the number of restraining streams is the inherent characteristics of a unit operation and is constant regardless of a flowsheet configuration, Konda, et al. (2006) presented a table for the number of restraining variables of major unit operations. They also demonstrated their method for distillation columns and a few complex flowsheets. Details of their methods and analyses can be found in Konda, et al. (2006) or Konada and Rangaiah (2012).

4.2.2.3. Manipulated variables (MVs)

Manipulated variables are those degrees of freedom which are used for inserting the control action to the system. The manipulated variables can be classified into two categories of steady-state and dynamic. Manipulated variables used for controlling material inventories are in the category of dynamic degrees of freedom (Skogestad 2004a). The steady-state degrees of freedom affect the ultimate state of the process and have more economic significance than dynamic degrees of freedom. The desired properties of manipulated variables are to be consistent with each other, far from saturation, reliable, and able to affect controlled variables with reasonable dynamics. Two manipulated variables may be inconsistent when they cannot be adjusted simultaneously. The example of inconsistency is when two control valves adjust the flowrate of the same material stream. Reliability is defined as the probability of failure to perform the desired action. Reliability of manipulated variables is important because it is not desirable to select a manipulated variable which is likely to fail for example due to corrosion or erosion.

If the available degrees of freedom are not sufficient to meet the controllability requirements, there are some limited opportunities for adding degrees of freedom to the process for example by inserting bypass streams, heat exchangers or buffer tanks into the process flowsheet, (Skogestad 2004a). In addition, van de Wal and de Jager, (2001) explained that in some scenarios, it might not be desirable to employ all the available manipulated variables. The reason can be in order to reduce the complexity of the control system. In addition, different operational modes (e.g., normal continuous operation versus start-up/shutdown) may require different control structures.

4.2.2.4. Controlled variables (CVs)

Selection of controlled variables is more complicated compared to manipulated variables. This is because controlled variables can be categorized based on two different tasks. Firstly, these variables are responsible for detection of disturbances and stabilizing processes within their feasible operational boundaries. Secondly, selection of controlled variables and their setpoints provide the opportunities to optimize profitability. The first category of controlled variables is selected for treatment of instability modes such as snowball effects (i.e., an instability mode concerned with materials inventories inside a recycle loop), or emptying/overflowing liquid holdups. The second category of controlled variables should be selected by economic criteria.

The controlled variables can be selected by engineering insights and practiced heuristics, especially when the control structure is developed for a unit operation. In addition, controllability measures, to be discussed later, can be applied for selecting controlled variables. Luyben (2005; 2006) listed five methods for selecting the location of temperature sensors within a distillation column, i.e., controlling the temperature of which tray inferentially ensures the desired compositions of product streams. They are:

1. Slope criterion,

In this method, a tray is selected, which has the largest temperature difference, compared to the neighbour trays.

2. Sensitivity criterion,

In this method, a tray is selected, which its temperature changes the most for a change in a manipulated variable.

3. Singular value decomposition (SVD) criterion,

This method is based on calculating the process gain matrix and its singular values as described by Moor (1992).

4. Invariant temperature criterion,

In this method, a tray is selected which its temperature does not change when the feed composition is changed and the compositions of the products are fixed.

5. Minimum product variability criterion,

In this method, a tray is selected which maintaining its temperature constant, results in least variability in product compositions.

The other common approach for selection of controlled variables, in particular for decentralized systems is to minimize the interactions between control loops using relative gain arrays (RGAs), as will be discussed later. However, none of the abovementioned methods ensures minimum economic losses in the presence of disturbances. The subsequent sections explain that optimal selection of controlled variables can ensure profitability.

4.2.2.5. Setpoint policy

When a control structure is selected for a process, the objectives for controlling that process such as stabilizing, safety concerns, environmental criteria, and profitability will be translated to maintaining a specific set of controlled variables at their setpoints. However, some of the targets of the abovementioned objectives may need to be updated time to time. This can be due to disturbances, the changes in environmental or safety policies, the changes in the specifications of products or feedstocks, or even because of the changes in the process behaviour over time (e.g. ageing of catalysts). The ability of the control structure to keep pace with these changes is crucial for feasibility and profitability of process operation.

As will be discussed in the subsequent subsections and shown in Fig. 8 adapted from Chachuat, et al. (2009), two strategies are possible for ensuring process feasibility and profitability. They are (i) static setpoint policy: off-line optimization and (ii) dynamic setpoint policy: on-line optimization. These are shown by red dotted vertical envelopes in Fig. 8. The methods for dynamic setpoint policy may apply two approaches. In the first approach, the measurements are used to update the model parameters (shown by model parameter adaptation). In the second approach, the measurements are used for updating modifier terms which are added to the objective function of the online optimizer, (shown by modifier adaptation).

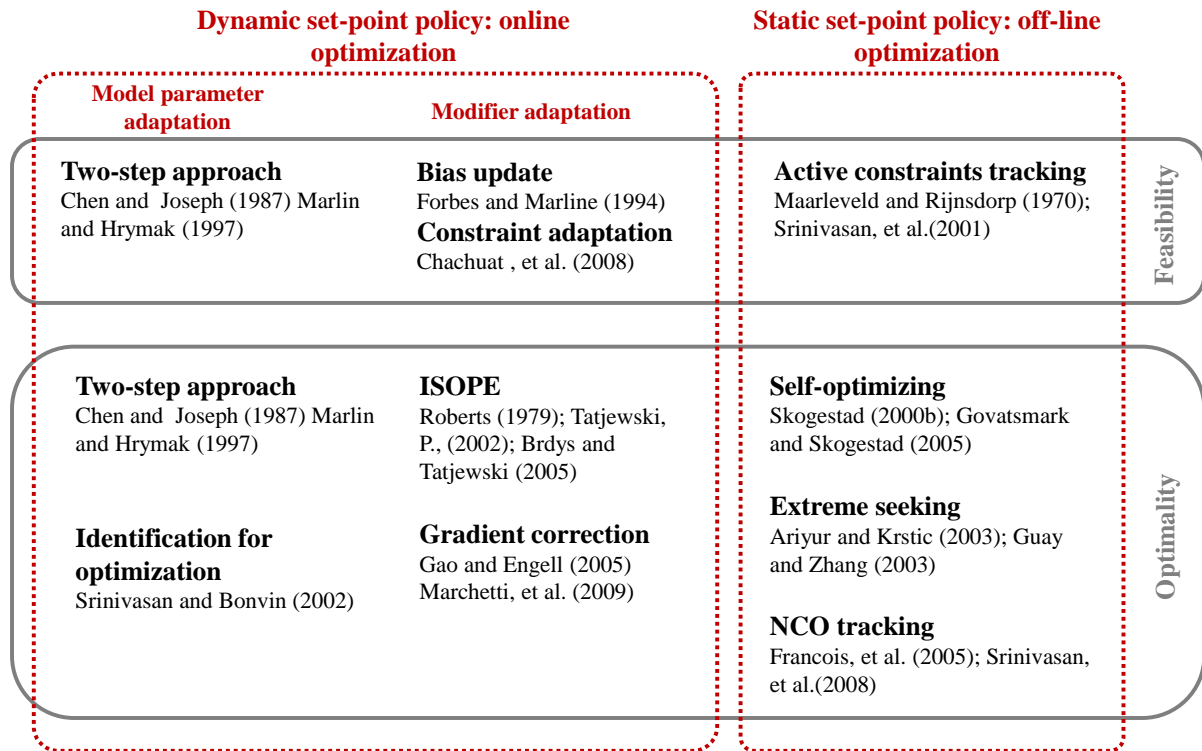


Fig. 8. Setpoint policies: the methods for static and dynamic setpoint policies are shown by the red dotted envelopes. The other classification is according to optimality and feasibility criteria, shown by the grey horizontal envelopes, (Adapted from Chachuat, et al. 2009).

The other classification, shown by grey horizontal envelopes, is according to (a) feasibility and (b) optimality criteria. Chachuat, et al. (2008) showed that the results of variational analysis in the presence of small parametric errors conform to the common sense that feasibility is of a higher priority than optimality. The references in the figure highlight the active researchers in the area. The dynamic and static setpoint policies are discussed in the subsequent subsections.

4.2.2.5.1. Static setpoint policy

The motivation for the static setpoint policy is that, while the costs of development and maintenance of a model-based online optimizer are relatively high, selection of the controlled variables which guarantee a feasible and near optimal operation is by no means trivial. Static setpoint policy has a direct relation to the optimal selection of controlled variables. In this approach, online optimization of setpoints is substituted by maintaining optimal controlled variables constant. This approach is also consistent with the culture of industrial practitioners who would like to counteract the model mismatches and the effects of disturbances by feedback control, (Chachuat, et al. 2009).

Morari, et al. (1980a) introduced the idea of optimal selection of controlled variables:

“In attempting to synthesize a feedback optimizing control structure, our main objective is to translate the economic objective into process control objectives. In other words we want to find a function c of the process variables [...] which when held constant, leads automatically to the optimal adjustment of the manipulated variables, and with it, the optimal operating conditions.”

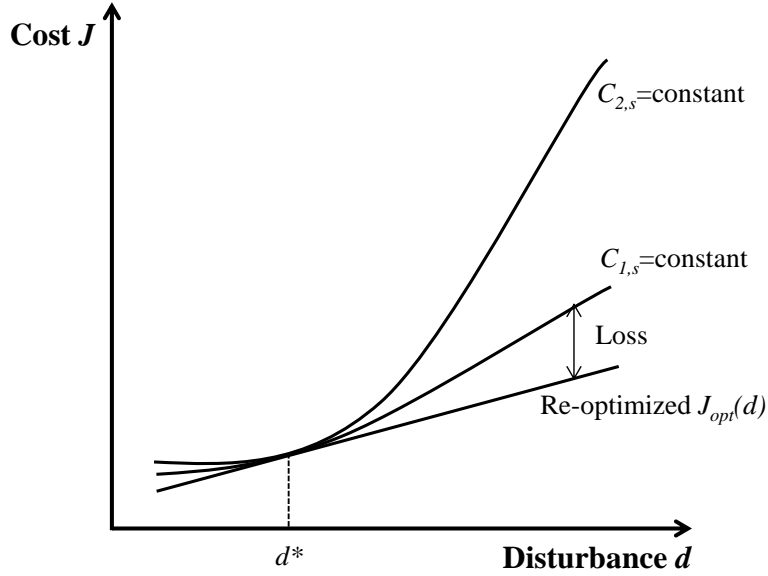


Fig. 9. Maintaining the setpoints at constant values results in an economic loss (distance between the re-optimized curve and the actual curve) due to a disturbance. However, the associated costs strongly depend on the selected controlled variable, (adapted from Skogestad 2000b).

Later, researchers (e.g. Skogestad 2000a, 2000b, 2004b; Kariwala 2007) investigated the notion of self-optimizing control. The concept is shown in Fig. 9, adapted from Skogestad (2000b). It shows that the costs (i.e. losses or decreases in profitability) associated with disturbances, are not the same for two different controlled variables. These controlled variables were maintained constant at their corresponding setpoints and the corresponding losses are compared to the scenario in which the objective function is re-optimized. As can be seen, in the presence of disturbance d , the cost associated with maintaining $C_{1,s}$ at its setpoint is significantly lower than $C_{2,s}$. This observation suggests that selection of controlled variables can be employed as a method for off-line optimization of process profitability.

Optimal controlled variables can be selected using *brute-force* optimization and direct calculations of the losses for different sets of controlled variables, which can be computationally expensive. Halvorsen, et al. (2003) presented a local method for optimal selection of controlled variables based on maximization of the minimum singular value. In that method, it was assumed that the setpoint error of different controlled variables (i.e., the difference between the constant setpoint and the re-optimized setpoint) are independent of each other, which does not often hold. Later, Alstad, et al. (2009) showed that an optimal linear combination of controlled variables is more likely to minimize the losses. This local method, called *null space* method, is based on the idea that the setpoints of optimal controlled variables must be insensitive to disturbances. This method ignores the measurement error. The work of Alstad, et al. (2009) also extends the methodology to the cases in which measurements are in excess or are fewer than the available inputs and the expected disturbances. The above methods are based on a quadratic objective function and linearization of the model. Therefore, the results are local and must be checked by a nonlinear model.

Kariwala (2007) proposed a computationally efficient method using singular value decomposition and Eigen-values for selection of optimal controlled variables. Later, this method was extended (Kariwala, et al. 2008) to use *average losses* instead of *worst-case losses*. The justification for this modification is that the worst-case scenario may not happen frequently and designing based on this scenario would result in the unreasonable losses of the control performance. Kariwala, et al. (2008) also showed that minimization of average losses had already minimized worst-case losses and was superior when the actual disturbance differs significantly from the average value.

Although maintaining controlled variables or a linear combination of them is convenient, there is no guarantee that the optimal operation is reached by the convergence. The reason is that in the presence of disturbances, the gradient of the cost function may changes from zero. In addition, the gradient of

the cost function may have a nonzero value for a constrained solution. Therefore, Cao, (2005) suggested that the sensitivity of the reduced gradient function to disturbances and implementation errors is a reliable measure for selection of controlled variables. Alternatively, some researchers chose to directly control the elements of the necessary condition for optimality. It can be shown (Chachuat, et al. 2009) that by determining the set of active constraints, the elements of the necessary condition for optimality can be decomposed into two categories. The first category ensure that the process operation remains feasible (i.e., the constraints are satisfied). The second category ensures an optimal operation (i.e., the reduced gradient is equal to zero).

However, the main difficulty associated with the methods based on static setpoint policy, is that active constraints may change. The methods for constraint handling proposed by researchers are split-range control (for the constraints on the manipulated variables), parametric programming, cascade control approach, and explicit constraint handling. Details of these methods can be found in literature (e.g., Umara, et al. 2012).

4.2.2.5.2. Dynamic setpoint policy

The methods in the second category (shown by the left red envelope in Fig. 8) apply an online optimizer to update the setpoints. The main challenge in the application of online optimizing control systems is the inability to develop accurate and reliable models with a manageable degree of complexity and uncertainty. The reason is that online optimization using an inaccurate model may result in a suboptimal or even infeasible operation, (Chachuat, et al. 2009). The two main approaches are (i) the methods for model parameter adaptation, in which the available measurements are used to refine the process model parameters; then this model is used for optimization, (Chen and Joseph, 1987; Marlin and Hrymak, 1997), and (ii) the methods for modifier adaptation in which modifier terms are added to the objective function and constraints and these modifiers are updated using available measurements, (Forbes and Marlin, 1994; Gao and Engell, 2005; Roberts, 1979; Tatjewski, 2002). The details and comparison of these methods are available in literature, (e.g., Chachuat, et al. 2009).

4.3. Controllability measures

Many research activities have been devoted to understanding the controllability characteristics of chemical processes. In the following, the definitions of operability and controllability are presented. and the limiting factors of controllability are reviewed. The corresponding node is node 1.3 on the left branch of Fig. 1.

4.3.1. Operability: Flexibility, Switchability and Controllability

The operability of a chemical process strongly depends on its operational mode, i.e. whether it deals with a constant load, or the load is time-dependent. A process with continuous operation spends most of its life cycle within a narrow envelope of steady states. Therefore, the control task is posed as regulation (i.e., disturbance rejection). By contrast, shutdowns, start-ups, and the operations of semi-continuous or periodic processes involve transient conditions along the desired time trajectories, and servo control is needed, (Pedersen, Jørgensen, and Skogestad 1999).

Operability is defined as the ability of input (manipulated) variables to meet the desired steady-state and dynamic performance criteria defined in the design stage, in the presence of expected disturbances, without violating any constraint, (Georgakis, et al. 2004). The mathematical descriptions of dynamic operability and steady-state operability are presented in Section 4.8.

Flexibility is defined as the ability to achieve a feasible operation over a range of uncertainties, (Dimitriadis and Pistikopoulos 1995). The mathematical programming of steady-state and dynamic flexibility optimizations is presented in Section 4.9.

A comparison between the definitions of operability and flexibility reveals some similarities and some differences. Both criteria emphasize the importance of ensuring a feasible operation by avoiding constraint violation. However, the criteria for flexibility also include the uncertainties in the model parameters, while in evaluating operability the focus is on disturbance scenarios. Furthermore, as will be seen later, the methods for flexibility optimization are able to identify the *worst-case scenario*

within the range of uncertain parameters and disturbances, while the methods for operability analysis assume that disturbances are known in advance.

Switchability is defined as the ability to move between operating points, (Pedersen, Jørgensen, and Skogestad 1999).

In addition, a variety of qualitative and quantitative definitions is available in literature for *controllability*, which reflects the experience of researchers. From the early studies, Ziegler, Nichols and Rochester (1942) suggested that their proposed test for finding tuning parameters can be used for classification of processes. Morari (1983) introduced the term *resiliency* that includes both switchability and controllability and is defined as the ability to move smoothly and rapidly between operating conditions and to effectively reject disturbances. He recognized that controllability is the inherent property of the process and does not depend on the controller design.

Kalman (1960) introduced the concept of *state controllability*. A state x is controllable, if for an initial condition $x_0 = x(t_0)$ and a final state x_1 , there exist a manipulated variable $u_1(t)$ and a final time t_1 , $0 < t < t_1$, such that $x_1 = x(t_1)$. In other words, the state controllability is the ability to bring the system from the initial state to the final state in a finite time.

Another important concept is *input-output controllability*. It is the ability to maintain the controlled variables $y(t)$, within their desired bounds or displacements from their setpoints r , in the presence of unknown but bounded disturbances d , using the available manipulated variables u , (Skogestad and Postlethwaite 2005).

A process is *functionally controllable* if for the desired trajectories of the output variables, $y(t)$, defined for $t > 0$, there exist some trajectories of the input variables, $u(t)$, defined for $t > 0$, which generates the desired controlled variables from the initial states $x(t_0)$, (Rosenbrock 1970).

It is notable that functional controllability depends on the structural properties of the system, i.e., a system that is functionally controllable with respect to a particular set of controlled variables may be rendered uncontrollable for another set. Furthermore, functional controllability is defined with respect to a set of desired trajectories of controlled variables. Therefore, a system may be functionally controllable for a set of controlled variable trajectories and becomes uncontrollable for another set. Furthermore, functional controllability has a clear relationship with perfect control, i.e., the controlled variables are maintained constant at their setpoints (or desired trajectories) and the manipulated variables are adjusted accordingly. This relationship provides the opportunity for evaluating the causes of control imperfection. For example, Russel and Perkins (1987) applied the concept of functional controllability and process inversion for discussing the causes of control imperfection in linear systems.

Furthermore, the concept of relative order can be applied in order to establish the necessary and sufficient condition for functional controllability and dynamic invertibility of nonlinear process models. Hirschorn, (1979) showed that in order for a nonlinear system to be invertible, the relative order of the controlled variables with respect to the manipulated variables should be finite. In addition, the functional controllability conditions requires that in order for function $f(t)$ to be selected as a desired output trajectory, its initial value and the initial values of its first $\alpha - 1$ derivatives should be equal to the corresponding values of the outputs trajectories, where α is the relative order. Later, McLellan (1994) showed that the index of a nonlinear inversion problem is equal to $\alpha + 1$ where α is the relative order of the process.

In addition, a comparison between the definitions of different controllability criteria suggests that functional controllability is more constraining compared to input-output controllability. This is because for a system to be functionally controllable the controlled variables should take the values of the desired trajectories. Therefore, their values are necessarily bounded, and the system features input-output controllability. However, the reverse is not true, because in the case of input-output controllability, although the system is required to have bounded outputs, it is not necessarily capable of following a certain desired trajectories of the controlled variables.

Functional controllability and input-output controllability concern only manipulated and controlled variables. On the other hand, state controllability additionally considers the initial and final conditions of the internal states. However, there is not a requirement for the controlled variables to follow a certain set of trajectories and a system which is state controllable may not be functionally controllable. However, a state controllable system has bounded inputs and outputs and is input-output

controllable. Finally it is notable that a system which is functional or input-output controllable is not necessarily capable of ensuring certain initial and final values for the internal states because functional controllability and input-output controllability do not consider internal states. Therefore, functional controllability and input-output controllability do not ensure state controllability.

4.3.2. Causes of control imperfection

Early studies in this research field involved *evaluation* studies, i.e. “*if the process is controllable at all?*” Later, the viewpoint of these studies evolved to address the question of “*how controllable the process is?*”, (Downs and Skogestad 2011). Several measures were introduced based on understanding of what limits process controllability. The idea is to apply the controllability measures iteratively in design stages to screen and eliminate solutions with undesirable properties. Moaveni and Khaki-Sedigh (2009) presented a recent and comprehensive review of these methods.

The limiting factors of process controllability can be classified to be (1) the interactions between control loops, (2) the manipulated variable constraints, (3) the delays and right-half-plane zeros, (4) the model uncertainties, and (5) the effects of disturbances. A variety of methods for quantifications of these deficiencies is available in literature, which with exception of few, all of them rely on linear models. The limiting factors of process controllability are discussed in the following.

4.3.2.1. Interactions between control loops

Bristol (1966) introduced relative gain arrays (RGAs) as the measure for the interactions between control loops, which has received significant industrial and academic attentions and is applied for pairing controlled and manipulated variables. An element of a relative gain array, $\Lambda = [\lambda_{ij}]$, represents the ratio of the open loop gain from the manipulated variable j to the controlled variable i , in which all the control loops are open, to the closed-loop gain in which all control loops, except the loop $i - j$, are perfectly controlled (Ogunnaike and Ray 1994):

$$\lambda_{ij} = \frac{(\partial y_i / \partial m_j)_{\text{all loops open}}}{(\partial y_i / \partial m_j)_{\text{loop } i-j \text{ open; all other loops closed with perfect control}}} \quad (7)$$

Since then, the Bristol's method has been extended by many researchers in order to capture the different characteristics of control loop interactions. Since RGA may not reflect one-way coupling in processes with triangular model matrix, Hovd and Skogestad (1992) introduced performance relative gain array (PRGA). In addition, often the number of inputs and outputs are not equal. Therefore, Chang and Yu (1990) introduced non-square relative gain array (NRGA) as the ratio of the open-loop gain to the closed-loop gain in which all the loops except the loop being studied, are perfectly controlled. Notice that for a non-square system with more outputs than inputs, it is not feasible to have zero steady-state off-sets. Therefore, the control objective was defined to minimize these off-sets in a least-square sense. Furthermore, since static RGA methods do not consider dynamic information, dynamic relative gain arrays (DRGAs) were introduced, in which transfer functions replace steady-state gains, (Tung and Edgar, 1981). The numerator is open loop transfer function but the denominator is perfectly controlled for all frequencies. The DRGAs rely on *a priori* decision about the type of controllers (McAvoy, et al. 2003). Similarly, Xiong et al., (2005) introduced effective relative gain array (ERGA). This measure combines the static RGA and bandwidth of the process model in order to provide a comprehensive measure of control loop interactions. In addition, the applications of the RGA methods have not been limited to single-input single-output (SISO) control systems. Manousiouthakis and Nikolaou (1989) introduced static nonlinear block relative gain arrays (NBGA) and dynamic nonlinear block relative gain arrays (DNBGA) as measures for the interactions between different blocks of a decentralized control structure. Furthermore, Chang and Yu (1992) argued that disturbances may be suppressed or amplified through control system interactions. Therefore, a non-interacting control structure is not necessarily the best choice and sometimes interactions are necessarily for disturbance rejection. To this end, they introduced a new measure called relative disturbance gain array (RDGA), which quantifies the capabilities of a control system

for disturbance rejection. Details of these methods can be found in Moaveni and Kariwala (2012) and Moaveni and Khaki-Sedigh (2009).

Another relevant criterion is the *integrity* of a decentralized control structure, which ensures whether a system remains stable while individual control loops are brought in and out. Niederlinski (1971) presented the integrity measure as:

$$NI = \det[\mathbf{G}(0)] / \prod_{i=1}^n g_{ii}(0) \quad (8)$$

where $\mathbf{G}[g_{ij}]_{n \times n}$ is the transfer matrix of a process. It is proved that if under closed loop condition the Niederlinski Index is negative, ($NI < 0$), the multi-loop control structure will be unstable for all values of the controller tuning parameters. This result is necessary and sufficient for 2×2 systems. However, for higher order systems it is a sufficient condition, i.e., if < 0 , the system will be unstable, (Ogunnaike and Ray 1994). It is notable that the interactions between control loops limit controllability of decentralized control systems and is not of concern for centralized control systems, (Ogunnaike and Ray 1994).

4.3.2.2. Manipulated variable constraints and the effects of disturbances

The effects of manipulated variable constraints can be measured using the methods for singular value decomposition (SVD). Consider the linear transfer function model below:

$$\mathbf{y}(s) = \mathbf{G}(s)\mathbf{u}(s) + \mathbf{G}_d(s)\mathbf{d}(s) \quad (9)$$

The gain matrix, \mathbf{G} , should be firstly scaled as $\mathbf{G}'(s) = \mathbf{S}_1 \mathbf{G}(s) \mathbf{S}_2$ in which \mathbf{S}_1 and \mathbf{S}_2 are output and input scaling vectors respectively. The importance of input scaling is sometimes neglected. However, Hori and Skogestad (2008) showed that for ill-conditioned processes such as distillation columns, input scaling is crucial. Then, the scaled gain matrix, $\mathbf{G}'(s)$, is decomposed into the products of two rotational matrices and a diagonal matrix of singular values. The smallest and largest singular values (σ_{min} and σ_{max} respectively) and their ratio (called condition number, γ) have implications for the constraints on the manipulated variables and hence process controllability. The singular value decomposition method relies on the property that the bounds on the reproducible output region depend on the minimum and maximum singular values and their ratio, (Cao, Biss and Perkins 1996):

$$\sigma_{min}(\omega) \|\mathbf{u}(j\omega)\|_2 \leq \|\mathbf{y}(j\omega)\|_2 \leq \sigma_{max}(\omega) \|\mathbf{u}(j\omega)\|_2 \quad (10)$$

$$\gamma = \sigma_{min}/\sigma_{max} \quad (11)$$

Therefore, it is desirable that σ_{min} and σ_{max} have large values to minimize the influence of manipulated variable constraints. However, the ratio of them, i.e., the condition number (CN), is also important because large γ implies strong dependency of output amplitude on the direction of input amplitude (Morari and Zafiriou 1989). Therefore, γ close to one is desirable. Furthermore, a large minimum singular value, σ_{min} , is desirable because it ensures that larger disturbances can be handled by the manipulated variables. In other words, the magnitudes of the disturbances that can be rejected depend on the manipulated variable constraints.

4.3.2.3. Delays and right-half-plane zero

Delays, right-half-plane zeros, and non-minimum phase behaviours have implications for closed loop performances, as discussed in the following.

Holt and Morari (1985) showed that in a multi-variable closed loop system, the minimum bound on the settling time for a controlled variable i is $\tau_i = \min p_{ij}$, where p_{ij} is time delay in the numerator of element g_{ij} in the transfer function matrix, \mathbf{G} . In addition, based on functional controllability, Perkins and Wong (1985) characterized a multi-variable system based on parameter Δ_{min} , which is the period that must be waited before the output trajectory can be specified independently, otherwise perfect control is not achievable. This period is bounded by the smallest and largest time delays in the process transfer function.

When process model is inverted, right-half-plane zeros become poles. It is well understood that right-half-plane zeros cannot be moved by any feedback controller and similar to time delays, they are the characteristics of the process (Yuan, et al. 2011). In particular, right-half-plane zeros limit the control performance of feedback controllers, (Skogestad and Postlethwaite, 2005).

4.3.2.4. Model uncertainties

Skogestad and Morari (1987) studied the effects of model uncertainties on control performances. Uncertainties in the process model require that the actual controller be detuned and hence degrade the control performance. If relative errors of transfer matrix elements are independent and have similar magnitude bounds, they concluded the relative gain array can be an indicator of closed loop sensitivity to uncertainties. Other contributions to quantify the effects of uncertainties, which are not limited to linear models, have been made by optimization-based methods, namely back-off (Narraway et al., 1991) and flexibility optimization (Swaney and Grossmann 1985a, b) methods, as will be discussed in this paper.

4.3.2.5. Multi-objective optimization methods based on controllability measures

One of the disadvantages of controllability measures is that each measure only considers a single cause of control imperfection. To address this issue, Cao and Yang (2004) proposed a multi-objective framework based on linear matrix inequalities (LMIs), which considers different controllability measures such as control error and control input effort.

The other issue about methods based on controllability measures is that enumeration and evaluation of all possible alternative solutions can lead to an intractable problem. In order to overcome this difficulty, researchers (Cao and Kariwala 2008; Kariwala and Cao 2009, 2010a, 2010b) proposed optimization frameworks based on a bi-directional branch and bound algorithm for screening alternative solutions, in which the nodes which do not lead to the optimal solution are eliminated faster and a smaller number of nodes need to be evaluated.

4.4. Observability and sensor placement

The operation of modern processes requires satisfaction of a variety of constraints including product quality, safety, and environmental constraints. Maximizing profit and at the same time meeting these criteria requires accurate knowledge of the process state. However utilization of a large number of sensors (i.e., controlled variables at regulatory control layer) may drastically increase the required investment and maintenance costs. Therefore, optimal placement of sensors has been the focus of researchers, as discussed in the following. The corresponding node in the hierarchical tree of Fig. 1 is node 1.4.

The initial research into sensor placement was based on state observability and was limited to linear systems. A system is state observable if for any time $t > 0$, the initial states $\mathbf{x}_0 = \mathbf{x}(t_0)$ can be estimated from the time trajectories of controlled variables $\mathbf{y}(t)$, and manipulated variables $\mathbf{u}(t)$. State observability is the dual notion of state controllability. While state controllability is the ability to drive the state variables to the desired values using manipulated variables, state observability is the ability to infer the values of the state variables from the manipulated and controlled variables.

For the linear time-invariant system of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \quad (12)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u} \quad (13)$$

controllability Gramian $\mathbf{W}_c(t)$ and observability Gramian, $\mathbf{W}_o(t)$ are defined as

$$\mathbf{W}_c(t) = \int_0^t e^{\mathbf{A}\tau} \mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T\tau} d\tau \quad (14)$$

$$\mathbf{W}_o(t) = \int_0^t e^{\mathbf{A}\tau} \mathbf{C}\mathbf{C}^T e^{\mathbf{A}^T\tau} d\tau \quad (15)$$

The time-invariant linear system of (12, 13) is state controllable and state observable if and only if for any $t > 0$, $\mathbf{W}_c(t)$ and $\mathbf{W}_o(t)$ are positive definite, respectively (Zhou, et al. 1996). These criteria can be applied for selection of manipulated (actuators) and controlled variables (sensors). However, they have a binary yes/no attitude to the problem and the number of candidate controlled and manipulated variables which pass these tests are often so large that makes ranking the alternative choices difficult. Therefore, as discussed by van de Wal and de Jager, (2001) and Singh and Hahn (2005), more quantitative measures were introduced by researchers about the degree of observability. Müller

&Weber, (1972) proposed three measures of (1) minimum Eigen values of $\mathbf{W}_c(t)$ and $\mathbf{W}_o(t)$, (2) determinants of $\mathbf{W}_c(t)$ and $\mathbf{W}_o(t)$, and (3) the reciprocal of the trace of $\mathbf{W}_c^{-1}(t)$ and $\mathbf{W}_o^{-1}(t)$ for selection of manipulated (actuators) and controlled (sensors) variables. Dochain et. al., (1997), Waldruff et al (1998) and van den Berg et. al., (2000) applied condition number and minimum or maximum singular values and trace of observability Gramian for optimal sensor location. The minimum Eigen value implies how far the system is from being unobservable. While minimum singular value is useful for identifying the worst observability direction, it is unlikely that all the states of the system need to be observed. On the other hand maximum Eigen value indicates the dominant observability direction and ensures that the most important states are easily observable. An alternative measure is the trace which represents sum of the singular values. While the above methods consider linear models, recently, researchers have addressed the problem of sensor placement for nonlinear processes using covariance matrices known as empirical observability Gramian (Hahn and Edgar 2002; Singh and Hahn 2005; Singh and Hahn 2006; Serpas, et al, 2013). Moreover, other researchers considered measurement costs and sensor reliability as the selection criteria (Bagajewicz, 1997, Chmielewski, et al., 2002, Muske and Georgakis 2003). In principle, the problem has a multi-objective function and its solution is a Pareto front which demonstrates the trade-off between costs, reliability and observability. Furthermore, in order to address large scale problems, Chamseddine and Noura (2012) proposed a method for decomposition of a complex system into interconnected sub-systems, and apply reduced order observers for these subsystems. Furthermore, design of experiment (DOE), a method for deciding the experimental setting in order to minimize the experiment costs, is applied by the researchers (Joshi and Boyd 2009; Alaña, 2010) as a natural tool for identifying optimal sensors. This is of particular interest to real-time optimization (RTO) systems because sensor location has direct effects on the quality of parameter estimation. The idea is to maximize the quality of the data acquired by sensors for updating the model parameters required for RTO execution, (Fraleigh, et al., 2003).

4.5. Passivity/dissipativity

The focus of the methods based on passivity/dissipativity analysis, is stability of decentralized control systems. The corresponding node is node 1.5 on the left branch of Fig. 1.

A comprehensive review of the methods for passivity analysis is presented in a book written by Bao and Lee (2007), for which a review is also provided by Ydstie, (2010). By definition, a dissipative system cannot deliver energy more than stored in it. This constraint can be formulated by the following inequality:

$$W(t_1) \leq W(t_0) + \int_{t_0}^{t_1} \varphi(\mathbf{y}(t), \mathbf{u}(t)) dt \quad (16)$$

in which, $\varphi(\mathbf{y}(t), \mathbf{u}(t))$ is energy supply rate (energy/time), and $W(t) \geq 0$ is the stored energy at time t . The above correlation is called dissipation inequality. $W(t_1)$ can be any generalized energy function and $\varphi(\mathbf{y}(t), \mathbf{u}(t))$ can be any abstract power function. In the following, three functions for energy supply rate are discussed, (Rojas, et al. 2009).

A system is called passive if $\varphi(\mathbf{y}(t), \mathbf{u}(t)) = \mathbf{y}(t)^T \mathbf{u}(t)$. Then, for $\mathbf{y}(t) = -k\mathbf{u}(t)$:

$$k \int_{t_0}^{t_1} \mathbf{y}^2(t) dt \leq W(t_0) \quad (17)$$

where $k > 0$ is the gain. Therefore by choosing an appropriate value for k , any passive system can be controlled using proportional controllers. Furthermore, it is possible to prove that two passive systems connected by a feedback control are also passive. The other systems of interest are *input feedforward passive* (IFP) systems in which:

$$\varphi(\mathbf{y}(t), \mathbf{u}(t)) = \mathbf{y}(t)^T \mathbf{u}(t) - v \mathbf{u}(t)^T \mathbf{u}(t), \quad v > 0 \quad (18)$$

and *output feedback passive* (OFP) systems in which:

$$\varphi(\mathbf{y}(t), \mathbf{u}(t)) = \mathbf{y}(t)^T \mathbf{u}(t) - \rho \mathbf{y}(t)^T \mathbf{y}(t), \quad \rho > 0 \quad (19)$$

A nonlinear input feedforward passive system is minimum phase (i.e., it has stable zero dynamics) and an output feedback passive system has bounded gains (i.e., it is input-output stable). If the dissipative inequality holds but with $\rho < 0$ or $v < 0$, then it is in shortage of IFP or OFP,

respectively. The shortage of IFP (or OFP) of a subsystem can be compensated with the excess of IFP (or OFP) of another subsystem in the same process network.

These properties serve as the foundations for studying stability of process networks and evaluation of controllability of decentralized and block-decentralized multivariable systems. The methodology is also extended to analyse the system integrity, i.e. whether the system remains stable if a control loop fails, and what back-up control loops are required in order to design a fault tolerant system. The advantage of passivity methods is that their representations are not limited to linear models. However, the required modelling efforts have limited their application to small problems, (Yuan, et al. 2011). Furthermore, the focus of these methods is feasibility rather than optimality. However, establishing a trade-off between competing control and process objectives is the key requirement for integrated design and control.

4.6. Multi-objective optimization methods to incorporate controllability measures into the process design

The methods using controllability measures suffer from several disadvantages. They have a yes/no attitude and each of these measures only concerns a certain limiting factor of controllability and at most can be used for highlighting the situations in which process controllability is lost. Acknowledging these limitations, some research activities have focused on defining multi-objective criteria to incorporate controllability measures into the process design. The corresponding node is node 2.1 on the right branch of Fig. 1.

Luyben and Floudas (1994) employed a multi-objective function for incorporating controllability measures and economic objectives. The economic objectives such as capital costs and operating costs were calculated using a steady-state model while bounds on controllability objectives were calculated using measures such as relative gain array, minimum singular value, condition number, and disturbance condition number. The resulting MINLP formulation was solved using generalized benders decomposition (GBD) algorithm (Geoffrion, 1972). Similarly, Chacon-Mondragon and Himmelblau (1996) proposed a bi-objective optimization in which costs and flexibility were optimized simultaneously. Later, Alhammad and Romagnoli (2004) proposed an optimization framework in which process economy, controllability, and environmental measures were incorporated into a multi-objective function.

4.7. Methods based on model reduction and linear control theory

Several research activities are devoted to use model identification and model reduction techniques in order to reduce the numerical complexities of the underlying mixed integer nonlinear dynamic optimization problem. The corresponding node is node 2.2 on the right branch of Fig. 1. Two research paths can be distinguished in this area; model reduction can be performed on the whole process-controller model (*bound worst-case approach*) or the problem is decomposed into a bi-level optimization and the reduced model is applied for control design only (*embedded control optimization*). These methods are discussed in the following.

Douglas and co-workers proposed a method based on model reduction, (Chawanku, et al. 2005; Ricardez-Sandoval, et al, 2008, 2009a, 2009b). The idea is to perform process identification on the nonlinear first principles model. The results of identification are a linear model and a model for uncertainties, which represents the difference between the full nonlinear model and the linear model. Then, the measures commonly used in robust control (e.g., structured singular value) are applied to estimate the bounds on process variables and to evaluate flexibility, stability and controllability of the process. These bounds give evaluations of the worst variability and the violations of constraints. For this reason, this methodology is termed *bound worst-case approach*. The advantage of this method is that the application of the reduced model eliminates the need for computationally expensive dynamic optimization. The disadvantage of this method is that it is based on a worst-case scenario which is not necessarily the most common scenario, and the method could be too conservative resulting in unnecessarily degradation of the objective function. To overcome this difficulty, Ricardez-Sandoval, et. al, (2011) suggested to calculate the worst disturbance scenario using structured singular value but the process variability should be calculated using a closed loop first principles model, resulting in a

less conservative solution. They called the new method *hybrid worst-case approach*, because in the new method both linear and nonlinear first principles models are involved.

Malcolm, et al. (2007), Moon, et al. (2011), and Patel, et al. (2008) pursued similar idea. However, they decomposed the problem into a bi-level optimization, in which control design was performed using a reduced (adaptive state-space) model, and process design was performed using the original first principles model. The linear state-space model is used for deciding control action in each optimization iteration, in order to disentangle the numerical complexities of feedback control. Malcolm, et al. (2007) applied sequential least square method for identification of the state-space model. Their method employs three layers (identifier/observer/regulator) at the control optimization level. In addition, the process optimization level consists of two optimization loops for steady-state and dynamic flexibility tests. The justification is that if a process does not feature a feasible steady-state operation, further investigation of dynamic flexibility is not needed. Patel, et al. (2008) applied similar idea with modified linear quadratic regulator (mLQR). The applied mLQR method incorporated an additional penalty term on the movement of the manipulated variables in order to add integrating action to the controller. They considered the corners of hyper-rectangular disturbance space rather than dynamic flexibility test.

The advantage of the aforementioned methods is that the linear model benefits from analytical solutions and the computationally expensive dynamic nonlinear optimization is avoided. The disadvantage of these methods is that due to application of a linear model, the solution is local. In addition, in the case of highly nonlinear processes, application of nonlinear identification and observation methods may further augment the required computation expenses, (Yuan et al., 2012).

4.8. Methods based on analysing nonlinear behaviour of chemical processes

Controllability measures often apply a linearized model. Such a linear assumption can be justified by designing a nonlinear compensator which removes some nonlinearities of the process behaviour. However, this approach is only efficient for regulatory control in the vicinity of a steady-state point, and may not be sufficient for highly nonlinear processes, (Bogle, et. al, 2004).

Chemical processes demonstrate a variety of nonlinear behaviour from which non-minimum phase behaviour, and input/output multiplicity have received significant attentions. The node corresponding to these research activities is node 2.3 on the right branch of Fig. 1.

The definition of non-minimum behaviour is based on the concept of zero dynamics. Zero dynamics are defined to be the internal dynamics of a nonlinear system when the deviations of the controlled variables (process outputs) are maintained at zero using the manipulated variables (process inputs). Unstable zero dynamics are the nonlinear analogues of right-half-plane zeros, and imply instability of process inversion, called non-minimum phase behaviour (Isidori 1989, Slotine and Li, 1991). Input multiplicity, i.e., a scenario in which for a give output several steady-states exist, can cause non-minimum phase behaviour, (Bogle, et al. 2004).

In order to determine steady-state multiplicities, the mathematical model of the process should be condensed into an algebraic equation (Silva-Beard and Flores-Tlacuahuac, 1999):

$$\mathbf{F}(\mathbf{x}, \mathbf{p}) = 0 \quad (20)$$

where \mathbf{x} is the state variables and \mathbf{p} is the vector of design parameters. Then, the necessary condition of input-multiplicity is given by the implicit function theorem (Poston and Stewart 1996):

$$\mathbf{F}(\mathbf{x}, \varphi) = \frac{\partial \mathbf{F}(\mathbf{x}, \varphi)}{\partial \varphi} = 0 \quad (21)$$

in which $\varphi \in \mathbf{p}$. Therefore, the maximum number of multiplicity points, k , is given by:

$$\frac{\partial \mathbf{F}^k(\mathbf{x}, \varphi)}{\partial \varphi^k} = 0, \text{ and } \frac{\partial \mathbf{F}^{k+1}(\mathbf{x}, \varphi)}{\partial \varphi^{k+1}} \neq 0 \quad (22)$$

Similarly, the necessary condition for output-multiplicity is given by:

$$\mathbf{F}(\mathbf{x}, \varphi) = \frac{\partial \mathbf{F}(\mathbf{x}, \varphi)}{\partial \mathbf{x}} = 0 \quad (23)$$

Then, *isolas*, i.e., the points where isolated solutions originate and disappear, can be found by:

$$\mathbf{F}(\mathbf{x}, \varphi) = \frac{\partial \mathbf{F}(\mathbf{x}, \varphi)}{\partial \varphi} = \frac{\partial \mathbf{F}(\mathbf{x}, \varphi)}{\partial \mathbf{x}} = 0 \quad (24)$$

A thorough review of the methods for bifurcation analysis is beyond the scope of this review. However, several interesting results for integrated design and control are reviewed in the following. Silva-Beard and Flores-Tlacuahuac (1999) studied the regions of nonlinear behaviour of a free-radical CSTR polymerization reactor using continuation algorithm and global multiplicity diagrams. They showed that closed loop control in the optimal point of operation could be difficult because steady-state multiplicities would introduce positive zeros into the transfer function and limit the speed of closed loop control. Pavan Kumar and Kaistha (2008a, b) showed, depending on the control structure, input steady-state multiplicity might cause state transition and wrong control action in a generic ideal reactive distillation column. They recommended a three-point temperature control structure for addressing large deviations in the throughput. Lewin and Bogle, (1996) showed that in the case of a polymerisation reactor, economic steady-state optimization degrades the dynamic performance because, the optimizer chose an operating point closer to the bifurcation point. In addition, the effects of input multiplicity on degrading switchability due to non-minimum phase behaviour are studied by Kuhlmann and Bogle (2001; 2004).

Kiss, et al. (2002, 2003, 2007) studied the effects of recycle streams on product selectivity and steady-state multiplicity of a reactor-separator process. They identified two types of inventory control; *self-regulatory inventory control* in which the reactants are fed according to stoichiometry of reactions and is characterized by a minimum reactor volume required to avoid snowball effects; and *regulation by feedback inventory control* in which the reactants inventories are controlled by manipulating fresh feed. They argued that although the latter method is more difficult to implement, it eliminates the risk of instability and steady-state multiplicity.

The early methods for analysis of the nonlinear behaviour of chemical process rely extensively on the analytical solution of the process model. Marquardt and Mönnigmann (2005) applied the underlying theory for synthesis rather than analysis. They defined a critical manifold as the stability boundary which separates the design parameter space of feasible steady states from unstable oscillatory states. Then, an operational point should back-off from the critical manifolds in order to ensure a safe operation due to uncertainties and disturbances. A signal function was applied for testing whether the manifold is crossed. This function enabled identifying unknown critical manifolds. Then, the constraints for maintaining distance from these new critical manifolds were added and the optimization was repeated until no new critical manifold is found. This method has been successfully applied to a system of hundreds of equations.

The significant advantage of the methods described above is that they apply a steady-state nonlinear model, and are able to diagnose some undesirable characteristics that may limit the dynamic performance. However, the limitation of these methods is that, considering the size and combinatorial characteristics, developing analytical solution of a nonlinear process model is of limited practicality, and numerical solution rely extensively on the nonlinear solvers, that is it is not clear whether the solver failed to find the solution or a solution with physical significance does not exist. Furthermore, due to the inherent non-convexities of these systems, global optimization methods are required in order to construct the proof that the results are not due to a sub-optimal solution.

Another relevant research area concerns the measures of nonlinearity. In general, nonlinearity of a process is defined by contradiction, i.e., a process is nonlinear if it does not fulfil the additively and homogeneity properties (the principles of superposition) of a linear system, (Choudhury, et. al, 2008). Nevertheless, the aforementioned definition does not give a measure for the extent of nonlinearity. Therefore, extensive research is devoted to quantification of process nonlinearities. The nonlinearity measures are broadly classified into model-based and data-driven measures. An example of the measures in the first category is the normalized maximum difference between the nonlinear process and the best linear approximation of this process (Helbig, et al, 2000). Another example is the curvature of the process response introduced by Guay (1997). The application of the methods in the first category requires identification of the process model. By contrast, the methods in the second category apply signal processing statistical tools to the time series of the process output measurements in order to quantify the process nonlinearities. Examples of the methods in the second category are the

application of the bispectrum and bicoherence in order to measure nonlinearity and non-gaussianity of the time series. Details of these methods can be found in Choudhury, et al., (2008). The measures of process nonlinearity give an early indication, during process design stage, in order to decide whether a linear controller is sufficient or a nonlinear controller is needed for optimal control performance. Furthermore, these measures have gained extensive applications for fault diagnosis during process operation. However, little work has been done in order to incorporate these measures in a systematic framework for integrated design and control, which suggests a potent research area.

4.9. Geometric operability analysis

The definition of operability was mentioned earlier in Section 4.3.1. The geometric measures for steady-state and dynamic operability were introduced in order to quantify the area in which the process remains operable, (Vinson and Georgakis 2000; Uztürk and Georgakis 2002). The corresponding node is node 2.4 on the right branch of Fig. 1. Here, the discussion is based on the following state-space representation of the process model (Georgakis, et al. 2004):

$$\begin{aligned} \text{M: } \quad & \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{d}) \\ & \mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{d}) \\ & \mathbf{h}_1(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{y}, \dot{\mathbf{u}}, \mathbf{u}, \mathbf{d}) = 0 \\ & \mathbf{h}_2(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{y}, \dot{\mathbf{u}}, \mathbf{u}, \mathbf{d}) \leq 0 \end{aligned} \quad \text{Process model}$$

In the above, $\mathbf{x} \in \mathbf{R}^{n_x}$ is the vector of state variables, $\mathbf{u} \in \mathbf{R}^{n_u}$ is the vector of input (manipulated) variables, $\mathbf{d} \in \mathbf{R}^{n_d}$ is the vector of disturbance variables, $\mathbf{y} \in \mathbf{R}^{n_y}$ is the vector of output (controlled) variables. The method for steady-state operability analysis utilizes a steady-state process model that maps process inputs to process outputs. The process inputs are able to take the values in the available input set (AIS). Using the process model and AIS, it is possible to calculate the achievable output set (AOS). Notice that (AOS) is a function of \mathbf{u} and \mathbf{d} . A comparison between the desired output set (DOS) and the achievable output set (AOS) can be quantified as the operability index (OI):

$$\text{OI} = \frac{\mu(\text{DOS} \cap \text{AOS})}{\mu(\text{DOS})} \quad (25)$$

where μ is a measure of the size of each set, e.g., in a two-dimensional space, it represents the area and in a three-dimensional space, it represents the volume, (Georgakis and Li 2010). However, there are different definitions for operability index depending on whether the setpoints are constant or they are controlled in intervals (i.e., equivalent to setpoint tracking).

The achievable output set (AOS) can be calculated for a given available input set (AIS) and by fixing disturbances at their nominal values \mathbf{d}^N . Then a comparison between the desired output set (DOS) and the achievable output set (AOS) leads to quantification of the steady-state servo-operability index, as follows:

$$\text{s-OI} = \frac{\mu(\text{AOS}_u(\mathbf{d}^N) \cap \text{DOS})}{\mu(\text{DOS})} \quad (26)$$

in which

$$\text{AOS}_u(\mathbf{d}^N) = \{\mathbf{y} \mid \mathbf{M}(\dot{\mathbf{x}} = 0, \dot{\mathbf{u}} = 0, \mathbf{d} = \mathbf{d}^N); \forall \mathbf{u} \in \text{AIS}\} \quad (27)$$

Similarly the regulatory steady-state operability index will be:

$$\text{r-OI} = \frac{\mu(\text{AIS} \cap \text{DIS}_d(\mathbf{y}^N))}{\mu(\text{DIS}_d(\mathbf{y}^N))} \quad (28)$$

in which desired input set is defined as:

$$\text{DIS}_d(\mathbf{y}^N) = \{\mathbf{u} \mid \mathbf{M}(\dot{\mathbf{x}} = 0, \dot{\mathbf{u}} = 0, \mathbf{y} = \mathbf{y}^N); \forall \mathbf{d} \in \text{EDS}\} \quad (29)$$

where EDS is the expected disturbance space. Later, this method was developed to include dynamic operability (Uztürk and Georgakis 2002; Georgakis, et al. 2003). The set of values over which inputs can move is called dynamic available input space (dAIS). The dynamic desired operating space (dDOpS) is a function of desired output set (DOS), expected disturbance space (EDS) and the maximum allowable response time t_f^d as follows:

$$\text{dDOpS} = \{(t_f, \mathbf{y}_{sp}, \mathbf{d}) \mid t_f \leq t_f^d, \forall \mathbf{y}_{sp} \in \text{DOS}, \forall \mathbf{d} \in \text{EDS}\} \quad (30)$$

Similarly, the dynamic achievable operating space (dAOpS) is defined as

$$dAOpS = \{(t_f, \mathbf{y}_{sp}, \mathbf{d}) \mid t_f \geq t_f^*, \forall \mathbf{y}_{sp} \in DOS, \forall \mathbf{d} \in EDS, \forall \mathbf{u} \in dAIS, \} \quad (31)$$

t_f^* is the minimum time that is required for optimal control and its value can be calculated using dynamic optimization, (Georgakis, et al. 2003). In order to define the dynamic operability two other spaces are needed:

$$S_1 = \{(\mathbf{y}_{sp}, \mathbf{d}) \mid \forall \mathbf{y}_{sp} \in DOS, \forall \mathbf{d} \in EDS\} \quad (32)$$

$$S_2 = \{(\mathbf{y}_{sp}, \mathbf{d}) \mid t_f^* \leq t_f^d, \forall \mathbf{y}_{sp} \in DOS, \forall \mathbf{d} \in EDS\} \quad (33)$$

Then

$$dOI = \frac{\mu(S_2)}{\mu(S_1)} \quad (34)$$

The first operating space, S_1 , is the combination of the setpoints (DOS) and disturbances (EDS). The second space, S_2 , is the projection of intersection of $dDOpS$ and $dAOpS$, and represents the operating space that can be achieved. Therefore, the dynamic operability represents the fraction of operating space that can be achieved by the available inputs during the desirable response time. More details on these methods can be found in (Georgakis, et al. 2004).

It is notable that in the case of input multiplicity, additional interior points of AIS also need to be imaged in order to calculate the complete boundaries of AOS, (Subramanian and Georgakis 2001). This method is nonlinear and multi-variable. However, it has no implication for the regulatory control structure or inventory control system, (Vinson and Georgakis 2000). In addition, the problem suffers from the curse of dimensionality, i.e., the dimensions of the abovementioned sets increase sharply and the problem becomes intractable. To overcome this difficulty, Georgakis and Li (2010) introduced a method based on the techniques used in *design of experience* (Montgomery 2005) which selects a finite number of points to perform the input-output mapping.

4.10. Steady-state and dynamic flexibility optimization

A variety of methods for steady-state and dynamic flexibility optimization has been proposed by the researchers (e.g., Swaney and Grossman 1985a; Grossmann and Floudas 1987; Dimitriadis and Pistikopoulos 1995). The corresponding node is node 2.5 on the right branch of Fig. 1. The steady-state process model can be represented by the following equations, (Dimitriadis and Pistikopoulos 1995):

$$\mathbf{h}(\mathbf{d}, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = 0 \quad (35a)$$

$$\mathbf{g}(\mathbf{d}, \mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) \leq 0 \quad (35b)$$

$$\boldsymbol{\theta} \in \mathbf{T} = \{\boldsymbol{\theta} \mid \boldsymbol{\theta}^L \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}^U\} \quad (35c)$$

$$\mathbf{z} \in \mathbf{Z} = \{\mathbf{z} \mid \mathbf{z}^L \leq \mathbf{z} \leq \mathbf{z}^U\} \quad (35d)$$

where $\dim\{\mathbf{h}\} = \dim\{\mathbf{x}\}$. In above, \mathbf{x} is the vector of the state variables, \mathbf{z} is the vector the control (input) variables, $\boldsymbol{\theta}$ is the vector of the uncertain parameters, \mathbf{d} is the vector of the design variables. The design variables are decided during the process design stage and remain unchanged during the process operation. In the above set of equations, the state variables can be eliminated between equations (35a) and (35b), resulting in the following concise representation of the process model:

$$\mathbf{g}(\mathbf{d}, \mathbf{x}(\mathbf{d}, \mathbf{z}, \boldsymbol{\theta}), \mathbf{z}, \boldsymbol{\theta}) = \mathbf{f}(\mathbf{d}, \mathbf{z}, \boldsymbol{\theta}) \leq 0 \quad (36)$$

As shown by Halemane and Grossman (1983), for evaluating the steady-state flexibility the following optimization problem need to be solved:

$$\chi(\mathbf{d}) = \max_{\boldsymbol{\theta} \in \mathbf{T}} \min_{\mathbf{z} \in \mathbf{Z}} \max_{j \in J} f_j(\mathbf{d}, \mathbf{z}, \boldsymbol{\theta}) \quad (37)$$

where j is the index of inequalities of equation (36). If $\chi(\mathbf{d}) \leq 0$, the design is feasible for all $\boldsymbol{\theta} \in \mathbf{T}$, otherwise a set of critical values for uncertain parameters $\boldsymbol{\theta}^c$ is determined which causes the worst violation of constraints.

Swaney and Grossman (1985a), proposed a scalar flexibility index in order to quantify the area in which the process operation remains feasible, under uncertain conditions.

$$\begin{aligned}
F &= \max \delta && \text{steady – state flexibility optimization} \\
\text{Subject to} \\
\max_{\theta \in T(\delta)} \min_{z \in Z} \max_{j \in J} f_j(\mathbf{d}, \mathbf{z}, \boldsymbol{\theta}) &\leq 0 \\
\delta &\geq 0 \\
T(\delta) &= \{\boldsymbol{\theta} | \boldsymbol{\theta}^N - \delta \Delta \boldsymbol{\theta}^- \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}^N + \delta \Delta \boldsymbol{\theta}^+\}
\end{aligned}$$

In above $\boldsymbol{\theta}^N$ is the nominal value of uncertain parameters. In addition, $\Delta \boldsymbol{\theta}^+$ and $\Delta \boldsymbol{\theta}^-$ are the expected deviations from this value. The implication of the above formulation is that F is the largest scaled deviation of the uncertain parameters that can be accommodated by the process before the operation is rendered infeasible. In addition, Swaney and Grossman, (1985a) showed that under certain convexity conditions the infeasible operating points lie on the vertices of the uncertainty space and the problem simplifies to identifying the active constraints whose intersections limit feasible operation.

The early versions of flexibility optimization employed a steady-state formulation. The optimization variables were process design parameters and process inputs which could be optimized to compensate the losses associated with realization of uncertainties. The steady-state version of flexibility analysis did not consider transient states. However, in some important applications such as batch processes, shutdown and start-up procedures, disturbance rejection, or product changeover the dynamic operability of the process is of vital importance. Therefore, Dimitriadis and Pistikopoulos (1995) extended this method to include dynamic process optimization under time-varying uncertainties. In the dynamic formulation, the following system consisting of ordinary differential equations was considered to describe the process model:

$$\mathbf{h}(\mathbf{d}, \dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{z}(t), \boldsymbol{\theta}(t), t) = 0; \mathbf{x}(0) = \mathbf{x}_0 \quad (38)$$

$$\mathbf{g}^{path}(\mathbf{d}, \mathbf{x}(t), \mathbf{z}(t), \boldsymbol{\theta}(t), t) \leq 0 \quad (39)$$

$$\mathbf{g}_k^{point}(\mathbf{d}, \mathbf{x}(t^k), \mathbf{z}(t^k), \boldsymbol{\theta}(t^k), t^k) \leq 0, \quad k = 1, \dots, NP \quad (40)$$

Compared to the steady-state formulation, in the above states, inputs and uncertain parameters are time-dependent. The constraints \mathbf{g}^{path} and \mathbf{g}_k^{point} represent path and point constraints respectively. Then, the dynamic flexibility can be tested by solving the following optimization problem:

$$\chi(\mathbf{d}) = \max_{\theta(t) \in T(t)} \min_{z(t) \in Z(t)} \max_{j \in J, t \in [0, H]} g_j(\mathbf{d}, \mathbf{x}(t), \mathbf{z}(t), \boldsymbol{\theta}(t), t) \quad (41)$$

Subject to

$$\mathbf{h}(\mathbf{d}, \dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{z}(t), \boldsymbol{\theta}(t), t) = 0; \mathbf{x}(0) = \mathbf{x}_0 \quad (42)$$

$$T(t) = \{\boldsymbol{\theta}(t) | \boldsymbol{\theta}^L(t) \leq \boldsymbol{\theta}(t) \leq \boldsymbol{\theta}^U(t)\} \quad (43)$$

$$Z(t) = \{\mathbf{z}(t) | \mathbf{z}^L(t) \leq \mathbf{z}(t) \leq \mathbf{z}^U(t)\} \quad (44)$$

where H is the maximum time over which the flexibility of the dynamic system is considered. Similar to steady-state test, if $\chi(\mathbf{d}) \leq 0$, the system is flexible. Otherwise, at least for one $\boldsymbol{\theta}(t)$, there is no control action which can make the process operation feasible over the considered time horizon.

Similar to the steady-state case, the dynamic operability index problem can be formulated as follows:

$$DF = \max \delta \quad \text{dynamic flexibility optimization}$$

Subject to

$$\chi(\mathbf{d}) = \max_{\theta(t) \in T(t)} \min_{z(t) \in Z(t)} \max_{j \in J, t \in [0, H]} g_j(\mathbf{d}, \mathbf{x}(t), \mathbf{z}(t), \boldsymbol{\theta}(t), t) \leq 0$$

$$\mathbf{h}(\mathbf{d}, \dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{z}(t), \boldsymbol{\theta}(t), t) = 0; \mathbf{x}(0) = \mathbf{x}_0$$

$$\delta \geq 0$$

$$T(\delta, t) = \{\boldsymbol{\theta}(t) | \boldsymbol{\theta}^N(t) - \delta \Delta \boldsymbol{\theta}(t)^- \leq \boldsymbol{\theta}(t) \leq \boldsymbol{\theta}^N(t) + \delta \Delta \boldsymbol{\theta}(t)^+\}$$

$$Z(t) = \{\mathbf{z}(t) | \mathbf{z}^L(t) \leq \mathbf{z}(t) \leq \mathbf{z}^U(t)\}$$

The concept is shown in Fig. 10, (adapted from Sakizlis, et al. 2004). In this method, firstly multi-period optimization is performed for an initial set of uncertain scenarios. This gives intermediate values for the design variables. Then, a feasibility test (another optimization) is performed in which the design variables are fixed and the violations of the constraints are maximized using the uncertain parameters. This gives a critical scenario of the uncertain parameters with worst violation of constraints. The current set of the uncertain scenarios is updated and the two optimization problems are solved iteratively, until the second optimization fails to find a realization of the uncertain parameters, which violates the constraints and therefore, the design is feasible for the whole range of uncertain parameters.

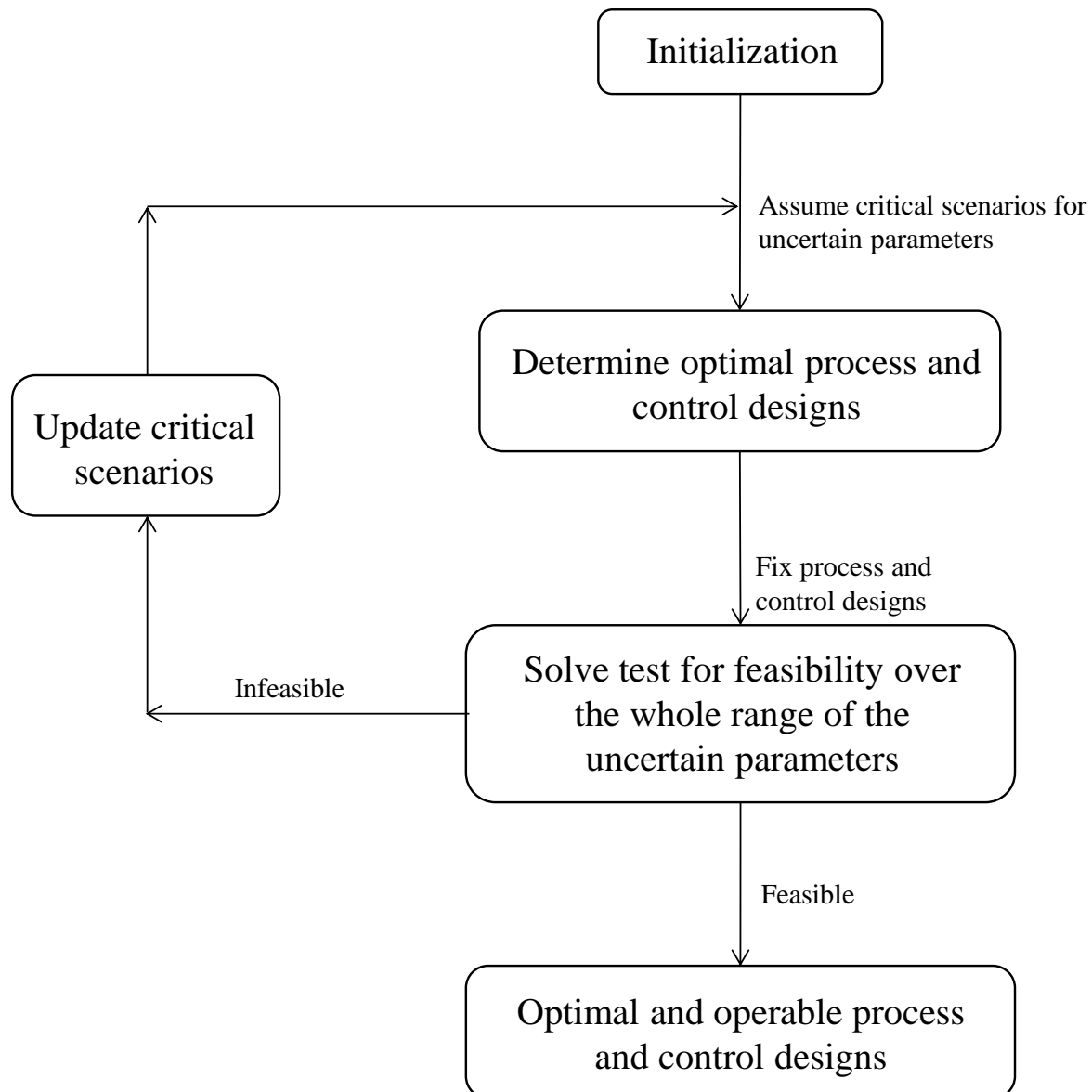


Fig. 10. The algorithm for flexibility optimization, (adapted from Sakizlis, et al 2004).

4.11. Economic optimization based on minimization of the economic losses associated with back-off from active constraints

Professor Perkins and his students (Narraway, et al. 1991; Narraway and Perkins 1993) proposed a method based on minimization of economic penalties associated with back-off from active constraints. The importance of this contribution was the recognition and integration of economic objectives into the problem formulation. The corresponding node is node 2.6 on the right branch of Fig. 1.

The idea is shown in Fig. 11, adapted from (Kookos and Perkins 2004). In many processes, the optimal steady-state economic solution lies on the intersection of the constraints. However, these constraints may be violated due to disturbances and uncertainties. Therefore, in order to ensure a safe and feasible operation, the nominal operating point must be moved away from the active constraints. Minimization of economic penalties associated with *back-off* from active constraints leads to identification of the optimal dynamic economic solution, shown in Fig. 11.

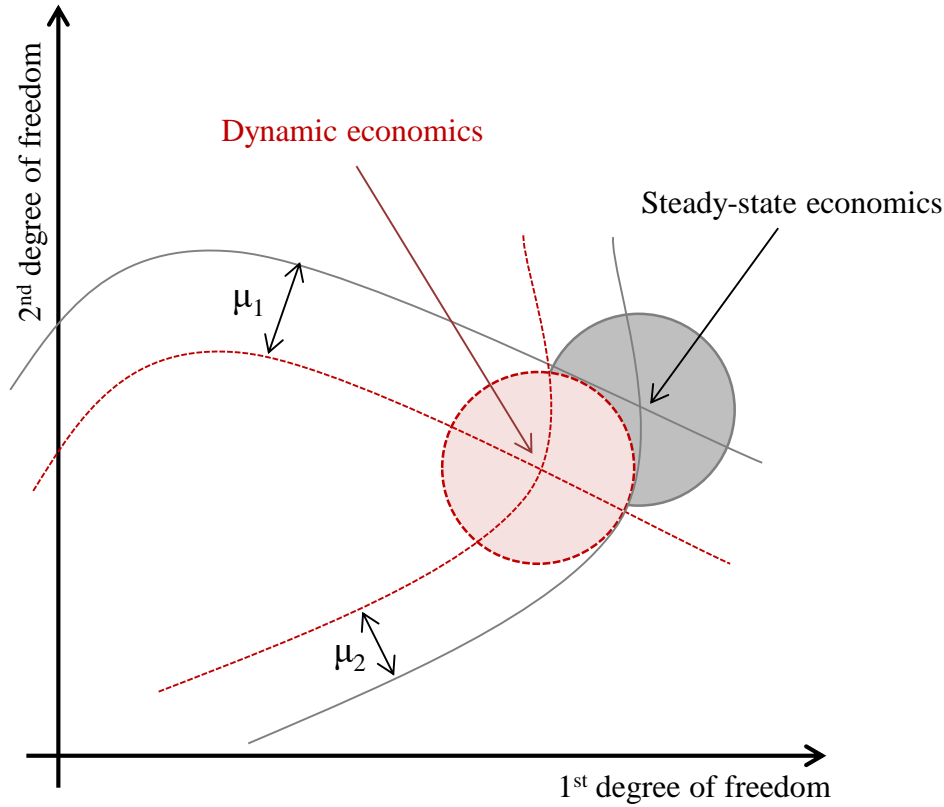


Fig. 11. Optimal steady-state and dynamic economic solutions, (adapted from Kookos and Perkins 2004).

The early versions of the back-off method were based on frequency analysis and perfect control (Narraway, et al. 1991; Narraway and Perkins 1993). Later, they developed a general formulation which included any linear time-invariant output feedback controller. The back-off method was also extended to time domain by considering decentralized (Heath, et al. 2000) and centralized (Kookos and Perkins 2001) proportional integral controllers.

4.12. Simultaneous optimization of a process and its controllers

In this approach, a superstructure of the process, its control structure, and its controllers is developed and optimized in order to systematically generate and screen alternative solutions. Here, the mathematical formulation of the problem is posed as a mixed-integer dynamic optimization (MIDO) problem. These methods are discussed in the following. The corresponding node is node 2.7 on the right branch of Fig. 1.

Since the underlying mathematical formulation for simultaneous optimization of a process and its controllers may result in large-scale MIDO problems, several research activities were devoted to develop new solution strategies in order to reduce the computational costs.

Samsatli, et al. (1998) proposed smooth approximation of binary variables to reformulate the MIDO problem using continuous variables, as follows.

$$Y = \frac{1}{2} \times [\tanh \{\xi(z)\} + 1] \quad (45)$$

Here ξ is a large number, then $Y = 1$ when $z \gg 0$ and $Y = 0$ when $z \ll 0$. However, the proposed approximating function results in errors when $z \approx 0$, because $Y = 0.5$.

Several research activities investigated decomposition of the problem into a primal subproblem and a master subproblem, with of the aim of reducing the computational costs. In all these methods, the primal subproblem is performed in a reduced space in which the binary variables are fixed. The primal subproblem gives an upper bound on the solution. However, as discussed in the following, different methods are used to formulate the master subproblem which determines the new realizations for the binary variables and gives lower bound on the solution. These two subproblems are solved

iteratively until the difference of the upper and lower bounds lies within the desirable tolerance. Avraam, et al. (1998, 1999) and Sharif, et al. (1998) applied linearization to construct an MILP master subproblem which was solved using outer approximation (OA) method. By comparison, Mohideen, et al. (1997), Schweiger and Floudas, (1997) and Bansal, et al. (2000a, 2003) applied dual information and generalized benders decomposition (GBD) algorithm (Geoffrion 1972) to construct the master subproblem. The former method based on outer approximation requires less evaluation of the primal subproblem because its master subproblem gives tighter lower bounds. However, the application of outer approximation algorithms required that the binary variables appear linearly and separated in the objective function and constraints. It is notable that new OA algorithms (e.g., applied by the recent versions of DICOPT) are extended to overcome this deficiency.

The method of full discretization based on orthogonal collocation was also applied by Cervantes and Biegler (2000b; 2002), and Flores-Tlacuahuac and Biegler (2005; 2007; 2008). Flores-Tlacuahuac and Biegler (2007) also studied the effects of the convexities of the problem formulation on the results. In that study, firstly the problem formulation was presented using generalized disjunctive programming (GDP) method (Biegler, et al. 1997) and then it was translated into several equivalent mixed integer formulations such as Big M, disaggregation or nonconvex formulations.

From the application point of view, these methods are applied to a number of case studies: a double effect distillation column (Bansal, et al. 2000b), a high purity distillation column (Ross, et al. 2001), and a multi-component distillation column (Bansal, et al. 2002). Later, Sakizlis, et al. (2003; 2004) and Khajuria and Pistikopoulos (2011) extended this method by including multi-parametric model predictive controllers. Asteasuain, et al. (2006) studied simultaneous process and control system design of styrene polymerization CSTR reactor. They considered a superstructure of feedback and forward controllers, and the optimization included the determination of optimal initial and final steady states and the time trajectories between them. Recently, Terrazas-Moreno, et al. (2008) studied a methyl-methacrylate continuous polymerization reactor. In this research, the design decisions (equipment size and steady-state operating conditions), the scheduling decisions (grade productions sequence, cycle duration, production quantities, inventory levels) and the optimal control decisions (grade transition time and profile) were made simultaneously.

4.13. A step forward: integrated design and control based on perfect control

This part of the paper reports the contribution of the author. The corresponding node is node 2.8 on the right branch of Fig. 1.

As discussed in the last sections, in integrated design and control, conventionally the models of process and controllers are combined and optimized simultaneously, as shown in Fig. 12. However, many researchers recognized that simultaneous optimization of a superstructure of the process, its control structure and controllers is a tough challenge for current dynamic optimization technologies. Therefore, some researchers attempted to develop more efficient solution algorithms and some others incorporated model reduction algorithms into the optimization framework.

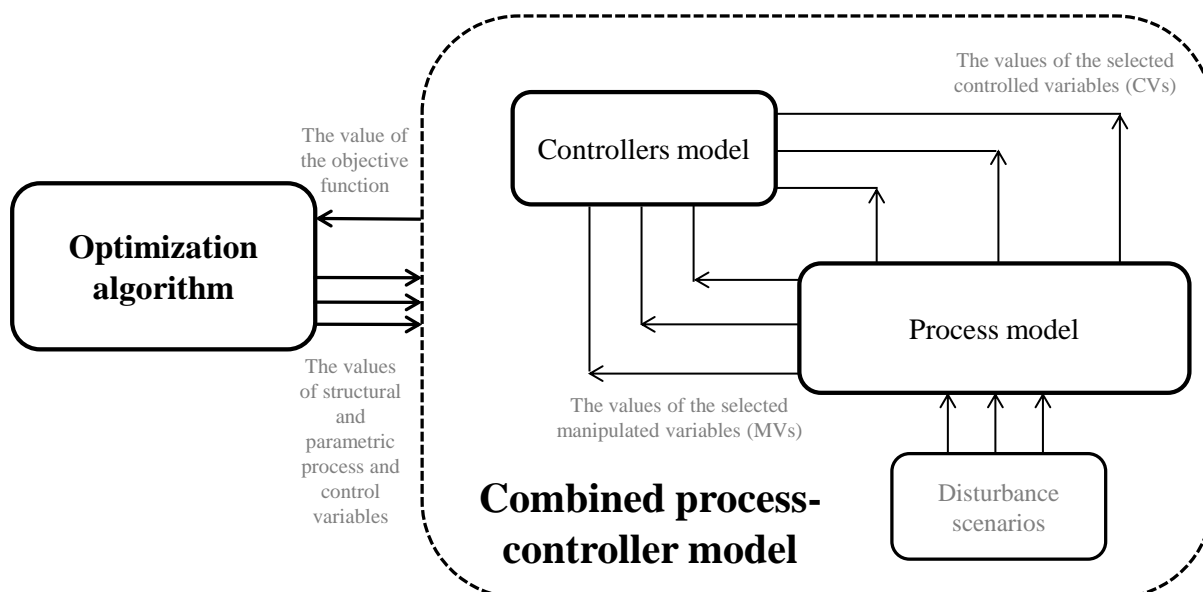


Fig. 12. The conventional optimization framework for integrated design and control of chemical processes, (adapted from Sharifzadeh and Thornhill 2013).

Some aspects of this formidable computational complexity, however, should be attributed to controllers. Optimization of controllers requires decision-making regarding the type of controllers (e.g., PI, MPC), the degree of centralization (and in the case of decentralized controllers, paring/partitioning between the manipulated and controlled variables), and controller parameters. Considering the large number of candidate manipulated and controlled variables, simultaneous optimization of controllers adds several orders of magnitude to the size of the problem. In addition to the computational complexities, there are several involved conceptual complexities. It is widely recognized that controllability is the inherent property of the process and does not depend on the controller design, (Morari 1983), i.e., it is not possible to resolve the uncontrollability issues of a process by designing more sophisticated controllers. Furthermore, there is no general agreement between researchers on the criteria for selection of the controller type. Some researchers (Luyben 2004; Skogestad 2009) emphasize simplicity and robustness of the conventional multi-loop control systems and criticize the reliability and costs of modern types. On the other side of this discussion, other researchers (Stephanopoulos, and Ng 2000; Rawlings and Stewart 2008) argue the economic advantages of model-based control systems and their systematic approach for handling constraints. In addition, they criticize the economic disadvantages of the constant-setpoint policy in decentralized control systems. Furthermore, in practice, advanced controllers (e.g. MPCs) are designed using commercial packages, often during process commissioning stages (Sakizlis, et al. 2010; Qin and Badgwell 2003), which may not be available at the process design stages. Therefore, disentangling the conceptual and computational complexities associated with controllers from the integrated design and control problem is highly desirable.

Recently, Sharifzadeh and Thornhill (2012a, b) and Sharifzadeh (2013) proposed integrated design and control, based on perfect control. The implication of perfect control is that the best achievable control performance can be determined by the inverse solution of the process model, (Garcia and Morari1982; Morari and Zafiriou 1989), in which manipulated variables taking account of disturbances such that the controlled variables are precisely at their desired setpoints. This is a well-known concept that has resulted in development of a class of controllers which use the inverse of the process model as an internal element, (Skogestad and Postlethwaite 2005). However, no attempt has been made to incorporate the concept of perfect control into integrated design and control using first principles modelling. Sharifzadeh and Thornhill (2012a, b) and Sharifzadeh (2013) addressed this opportunity. The concept is shown in Fig. 13. The controller model is eliminated and the process model is inverted. Here, the direction of information flow is reversed compared to Fig. 12, and the required values of the manipulated variables are calculated according to the desired values of the controlled variables. In the proposed optimization framework, the complexities associated with

controllers are disentangled from the problem formulation, while the process and its control structure are still optimized simultaneously. In addition, while the proposed framework is independent of the type of controllers, it provides the guidelines, in term of the best achievable performance, for control practitioners in order to design the actual controllers. Both steady-state and dynamic representations of this framework are formulated and demonstrated using case studies by Sharifzadeh and Thornhill (2012a, b), and Sharifzadeh (2012, 2013).

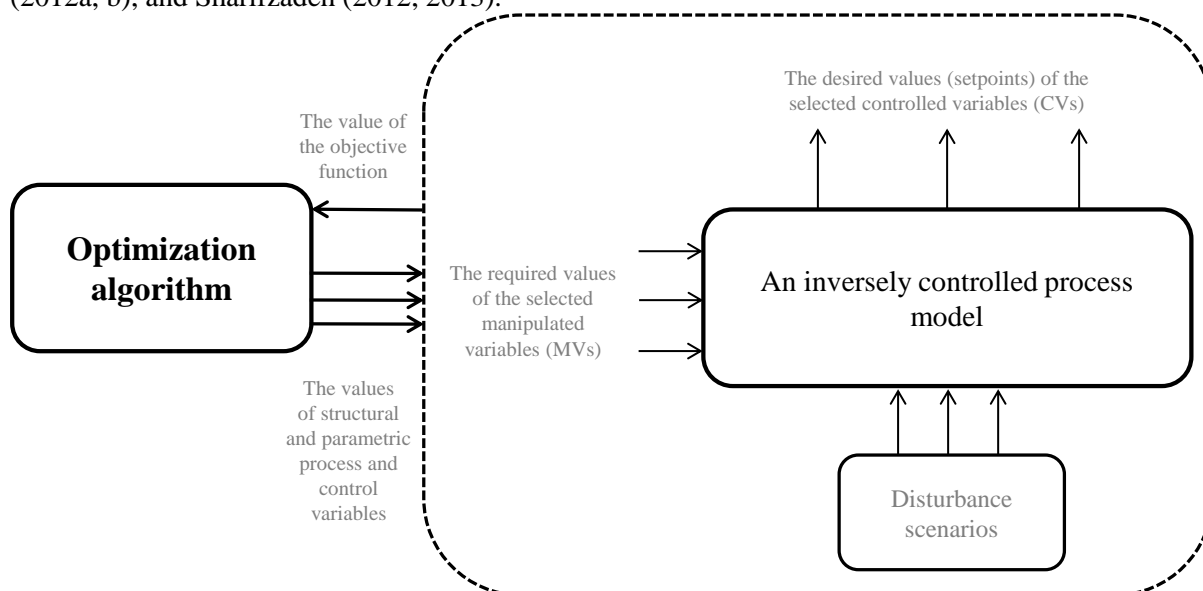


Fig. 13. The new integrated design and control framework using an inversely controlled process model, (adapted from Sharifzadeh and Thornhill 2013).

4.14. Optimization programming

This section briefly reviews the relevant methods for solving optimization-based algorithms of the last sections. The features of interest are the methods for solving MINLP and MIDO problems, the methods for global optimization, and simulation-optimization programming.

4.14.1. MINLP solution algorithms

As discussed earlier, in design and control of chemical processes, two categories of variables are involved, structural variables and parametric variables. The examples of structural variables are process configuration and selection of controlled and manipulated variables. The examples of parametric variables are the size of process equipment, process operating conditions and the setpoints of controlled variables. Simultaneous optimization of structural and parametric variables, involved in integrated design and control, requires mixed integer nonlinear programming (MINLP). The main MINLP solution algorithms can be explained using four subproblems. They are:

Subproblem NLP 1: the relaxation subproblem.

In this subproblem, the discrete variables are relaxed to have non-integer values. In general, the solution of Subproblem NLP1 results in non-integer values for discrete variables and gives a lower bound on the objective function of the main MINLP problem.

Subproblem NLP2: the subproblem with fixed discrete variables.

The solution of this subproblem gives an upper bound on the objective function of the main MINLP problem.

Subproblem NLPF: the feasibility subproblem with fixed discrete variables.

The Subproblem NLPF can be thought as minimization of the infeasibilities of the corresponding NLP2 subproblem.

Subproblem M-MILP: the cutting planes subproblem

The Subproblem M-MILP exploits the convexity of the objective function and the constraints, as they are replaced by the corresponding supporting hyper-planes. Due to the convexity of the feasible region, these hyper-planes are outer approximations of the nonlinear feasible region. Subproblem M-MILP may include linearization of all the constraints or only the violated constraints. The hyper-planes in Subproblem M-MILP provide new values for discrete variables, and a non-decreasing lower bound for the objective function. In other words, Subproblem M-MILP over estimates the feasible region and underestimates the objective function.

The mathematical formulation of the above subproblems can be found in (Grossmann 2002). The main MINLP algorithms are branch and bound (BB), outer approximation (OA), generalized benders decomposition (GBD), and extended cutting planes (ECP) which can be explained using the above sub-problems NPL1, NP2, NLPF, and M-MILP, as explained in the following, and shown in Fig. 14, (adapted from Grossmann 2002).

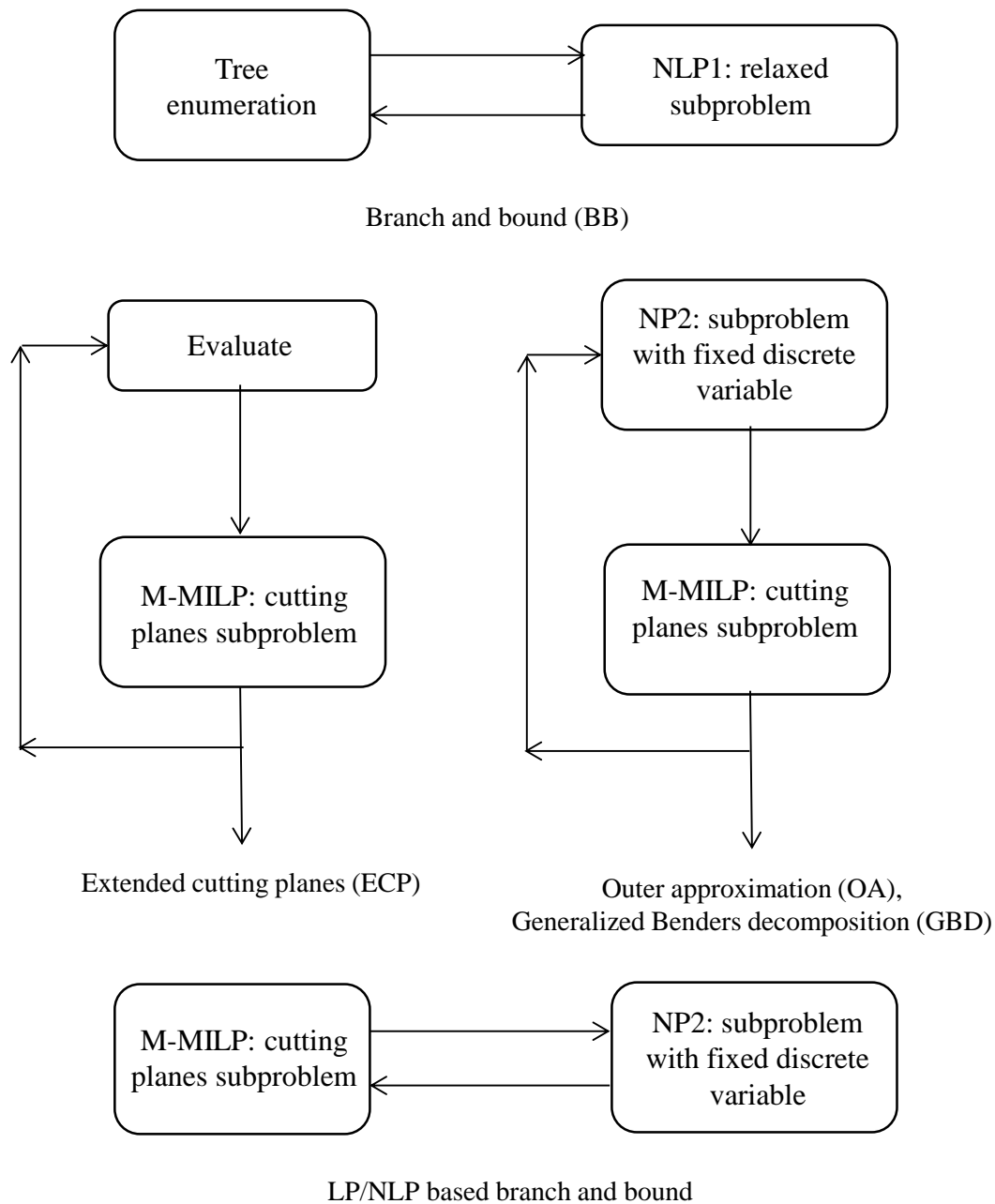


Fig. 14. Different MINLP algorithms represented as a combination of NLP and M-MILP subproblems, (adapted from Grossmann 2002).

Branch and bound (BB)

The branch and bound algorithm successively enumerates the nodes of the tree (constructed according to the integer variables) by fixing the discrete variables corresponding to the current node, and solving the relaxed Subproblem NLP1 for the rest of discrete variables. If all the discrete variables take integer values, the algorithm stops. Otherwise, the nodes of the tree are enumerated. The relaxed Subproblem NLP1 gives a lower bound for the subproblems in the descendant nodes. Fathoming is performed when the lower bound exceeds the current upper bound, when the Subproblem is infeasible, or when all discrete variables take integer values.

Outer approximation (OA)

In outer approximation algorithm, NLP2 (subproblem with fixed discrete variables) and M-MILP (subproblem with cutting planes) are solved iteratively. If the solution of NLP2 is feasible, it is used for constructing the cutting planes in M-MILP. Otherwise, the feasibility Subproblem, NLPF, is solved to generate the corresponding feasible solution. NLP2 and M-MILP subproblems give the upper and lower bounds respectively. The iterations continue until the difference of the lower and upper bounds lies within the allowable tolerance.

Generalized Benders decomposition (GBD)

Generalized benders decomposition (GBD) is similar to outer approximation (OA) in that subproblems M-MILP and NLP2 are solved iteratively. However, in GBD only active constraints are linearized for constructing the cutting planes.

Extended cutting planes (ECP)

The extended cutting planes algorithm does not require the abovementioned NLP sub-problems. M-MILP Subproblem is solved iteratively by adding the linearization of the most violated constraints. The algorithm converges when the violation of constraints lies within the allowable tolerance.

The algorithms based on branch and bound are only attractive when NLP subproblems are not computationally expensive or when due to the small dimension of discrete variables, the number of NLP subproblems is small. In general, outer approximation (OA) methods converge in fewer iterations. It can be shown that in extreme when the objective function and the constraints are linear, OA finds the solution in one iteration. In fact, as explained by Grossmann (2002), the M-MILP Subproblem does not even need to be solved to optimality. The generalized benders decomposition (GBD) algorithm can be thought as a special case of OA algorithm. Since the lower bounds of the GBD algorithm are weaker than OA algorithms, a larger number of iterations is required. For the case of extended cutting planes (ECP), since the discrete and continuous variables are treated simultaneously, a larger number of iterations is required. There are other variants and extensions of the above-mentioned algorithms such as branch and cut, LP/NLP branch and bound, and so on, which are not the focus of this discussion. The interested reader may refer to literature, (e.g., Biegler and Grossmann 2004). In general, branch and bound methods perform well when relaxation of MINLP is tight. Outer approximation methods are better when the NLP subproblems are computationally expensive. GDB methods are more favourable for problems with a large number of discrete variables and ECP methods are preferred for linear problems, (Biegler and Grossmann 2004).

The off-the-shelf commercial solvers for mixed integer nonlinear problems are available within modelling systems such as GAMS and AMPL. The common computer codes for nonlinearly constrained MINLPs are DICOPT, SBB, α -ECP and BARON. DICOPT is developed by Viswanathan and Grossmann (1990) at Carnegie Mellon University, based on OA. According to the recent manual of software, the algorithm is extended to include integer variables which appear nonlinearly in the problem formulation, (DICOPT documentation 2012). BARON is developed by Sahinidis (1996) and implements a global optimization method. This solver is based on a branch and reduce algorithm. SBB applies a branch and bound method and α -ECP is based on an extended cutting plane method. All these methods are based on the assumption of convexity of the objective function and constraints and may converge to a local solution in the presence of non-convexities. The methods for global optimization will be discussed later.

4.14.2. Dynamic optimization

In general, the solution algorithms for dynamic optimization problems can be classified into variational, sequential, full discretization and multiple-shooting methods. These methods are discussed in the following.

The *variational methods* use the first order optimality necessary conditions based on Pontryagin's Maximum Principle (Cervantes and Biegler 2000a). The resulted formulation conforms to a boundary value problem which can be solved using methods such as single shooting, and invariant embedding. If the analytical solution is found, these methods have the advantage that the solution is achieved in the original infinite dimensional space. However, analytical solution is often not possible and numerical solution features combinatorial characteristics in the presence of constraints. Therefore, the application of variational methods is limited to small problems.

In the *sequential integration* methods, also called *partial discretization* or *control vector parameterization*, only the control input variables (i.e., manipulated variables) are discretized. When initial conditions, time-independent variables and the parameters of the partial discretization are fixed, the resulted differential algebraic equations (DAEs) can be solved using a DAE solver. This produces the required objective function and gradients for an NLP solver. The solver determines the optimal values of the time-independent variables and the parameters of partial discretization. The special feature of sequential methods is that they generate a feasible solution in each iteration, (Biegler and Grossmann 2004).

In *full discretization* methods, also called *simultaneous methods*, all time-dependent variables are discretized which results in a large-scale nonlinear problem. The main technique for discretization is collocation based on finite elements, in which the profiles of the time-dependent variables are approximated by a family of polynomials. These methods follow an infeasible path and the differential algebraic equations are solved at the optimum point, only. Therefore, the execution time is significantly shorter than the sequential method. The full discretization methods are advantageous when state variables are (path) constrained or unstable modes exists, (Biegler and Grossmann 2004). In addition, the control input (manipulated) variables are discretized at the same level of accuracy as the state variables and the output (controlled) variables. However, the reformulated discretized problem could be very large which requires careful initialization of the optimization algorithm.

A method that should be categorized between the two extremes of the sequential integration and full discretization methods is called *multiple shooting*. In this method, the time horizon is divided to several intervals and in each interval a partial discretization problem, based on sequential approach is solved. The continuities between time intervals are established using additional equality constraints. The main advantage of the multiple-shooting methods over the sequential methods is that the (path) constraints on state variables can be imposed at the points between the time intervals.

4.14.3. Global optimization

The motivation for the research into global optimization is that the conventional nonlinear optimization methods do not guarantee to find the global solution in non-convex problems. The methods for global optimization can be broadly classified into stochastic methods and deterministic methods. The stochastic optimization methods, in general, apply an algorithm in analogy to physical systems (e.g. evolution in genetic algorithm) in order to generate trial points which approach an equilibrium point. The common examples of stochastic optimization methods are genetic algorithm, simulated annealing, and Tabu search algorithms. Stochastic optimization methods are widely applied. For example Low and Sorensen, (2003a-b, 2005), and Wongrat, and Younes (2011) applied genetic algorithms and Exler, et al. (2008) applied Tabu Search for mixed integer dynamic optimization. Furthermore, these methods do not require calculation of the gradients and are convenient for simulation-optimization programming. However, the stochastic global optimization methods do not construct any proof for the solution to be the global optimum and tend to be inefficient in highly constrained problems.

Recently, a variety of methods for deterministic global optimization methods is proposed by researchers. In summary, the main idea is to use convex envelopes and under-estimators in order to

construct the equivalent lower bounding convex problem. Consider the following mixed integer programming (MIP) problem:

$$\min Z = f(x, y) \quad (\text{MIP})$$

Subject to

$$g(x, y) \leq 0$$

In which x and y are continuous and discrete variables respectively. In addition, $f(x, y)$ and $g(x, y)$ are generally non-convex. Then, the equivalent lower bounding mixed integer programming (LBMIP) problem has the general form of:

$$\min Z = \bar{f}(x, y) \quad (\text{LBMIP})$$

Subject to

$$\bar{g}(x, y) \leq 0$$

where, \bar{f} and \bar{g} , are valid convex under-estimator, $\bar{f}(x, y) \leq f(x, y)$ and $\bar{g}(x, y) \leq 0$ holds if $g(x, y) \leq 0$. As discussed by Biegler and Grossmann (2004), the differences between different methods for deterministic global optimization are based on the way that the under-estimator problem is constructed and the way that branching is performed on discrete and continuous variables. The spatial tree enumeration can be done for both continuous and discrete variables. Alternatively, the spatial branch and bound can be performed on continuous variables and the resulted LBMIP can be solved by conventional MIP methods at each node. Branching on continuous variables is performed by dividing the feasible region, and comparing the upper and lower bounds for fathoming each sub-region, (Fig. 15). The sub-region which contains the global optimal solution is found by eliminating the sub-regions which are proved not to contain the global optimal solution. Finally, some methods branch on discrete variables of the LBMIP problem and switch to spatial branch and bound on the nodes where feasible values for discrete variables are found. For constructing the under-estimator some special structures such as bilinear, linear fractional, or concave separable may be assumed for continuous variables. Alternatively, in some methods a quadratic large term is added to the original function. Nevertheless, in all these methods, the quality of the under-estimator depends on the method for tightening the upper and lower bounds. The details of these methods and the way that the convex envelopes and under-estimators are constructed are reviewed by Grossmann and Biegler (2004) and Tawarmalani and Sahinidis (2004), and Floudas, et al. (2005). Recently researchers have extended the global optimization methods for dynamic optimization. Barton and Lee, (2004) solved MIDO problems with embedded linear time-varying dynamic systems to global optimality. Later Chachuat, et al. (2005; 2006) proposed a decomposition method based on outer approximation, which is able to address a wider range of problems with embedded ordinary differential equations (ODEs) without enumerating the discrete variables.

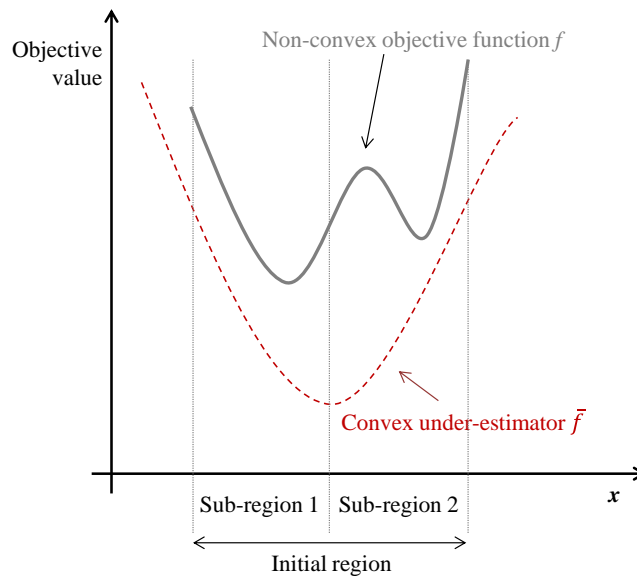


Fig. 15. The concept of constructing the convex under-estimator for a non-convex function, (adapted from Biegler and Grossmann, 2004).

4.14.4. Optimization with implicit constraints: Simulation-optimization programming

Simulation-optimization programming techniques conform to the optimization with implicit constraints, and have been proved efficient in process optimization using simulators (e.g., Caballero et al 2007; Sharifzadeh et al., 2011). In simulation-optimization programming, the process simulator has an input-output black-box relationship to the optimizer. Optimization is performed in the outer loop and the simulation is solved in the inner loop. The advantage of this method is that it provides an opportunity to apply off-the-shelf simulation software tools with advanced thermodynamic property packages. In addition, the number of optimization variables is limited to the required specifications of the simulation program, (i.e., the variables which should be specified independently to run the simulator). For fixed values of the optimization variables, the equation solver of the simulator is able to calculate the values of the remaining variables. By convergence of the equation solver, the value of the objective function is evaluated and reported to the optimizer. The disadvantage of these methods is that evaluation of the objective function is computationally expensive and time-consuming because for each evaluation, the equation solver needs to converge.

Recently, Cozad, et al. (2011) proposed a method for simulation-optimization in which, a surrogate models is constructed and its parameters are optimized against the maximum error between the rigorous simulation and the lean surrogate model. The advantage of the new approach is that the surrogate model provides cheap evaluations of the gradients and can be optimized using standard optimization algorithms.

5. Suggestions for future research

In the following, based on the reviewed materials, several directions for future research activities are suggested.

Suggestion 1. Developing high fidelity models is still the main barrier toward integrated design and control. Knowledge of many of the processes is still of an empirical nature and there is no guarantee that a controller that is designed using a simplified model at the design stage, will achieve the desired performance at the operational stage. The author suggests further research into developing methodologies which are independent of the detailed controller design and are robust to uncertain process parameters. To this end, both conceptual and computational complexities should be addressed simultaneously, i.e., model fidelity and integrity should not be compromised for the sake of numerical complexity reduction.

Suggestion 2. The author suggests developing methods which systematically capture the process insights and engineering judgments (e.g., Section 4.1) for decomposition and complexity reduction of integrated design and control problems as a potential research area with great potential impacts. Such a desirable methodology may also act as preconditioning module of any integrated design and control framework and will mostly concern the structure of the problem formulation rather than fine details. The required developments potentially will benefit from “*expanding the scope of modeling options...*” to new modelling techniques such as “*graph theoretical models, Petri nets, rule-based systems, semantic networks, ontology’s, agents,...*” as discussed by Stephanopoulos and Reklaitis (2011) and Venkatasubramanian (2009, 2011).

Suggestion 3. The characteristic of different controller types such as multi-loop controllers and model predictive controllers were discussed earlier in this paper. Unfortunately, as discussed earlier, there is no general agreement between researchers on the best type of controllers. Some researchers emphasize simplicity and robustness of the conventional multi-loop control systems and criticize the reliability and costs of modern types. On the other side of this discussion, other researchers argue the economic advantages of model-based control systems and their systematic approach for constraint-handling. In addition, they criticize the economic disadvantages of the constant-setpoint policy in decentralized control systems. The author suggests developing a systematic decision-making framework for deciding the best controller type, control law, and the degree of centralization. In particular, it should be investigated that in the presence of the limiting factors of controllability (discussed in Section 4.3.2), which type of controllers (e.g., feedback, feedforward, or model-based) is more capable of approaching perfect control in term of the best achievable control performance. The

outcome can also include a set of qualitative guidelines that in the case of a process with some specific characteristics (e.g. delays), the controllers should employ some advantageous elements (e.g., feedforward).

Suggestion 4. Furthermore, as discussed also by other researchers (e.g., Klatt and Marquardt, 2009), evolving new computational technologies have changed the perceptions of process systems engineers of their problem-solving capabilities. It is expected that if a problem (here, integrated design and control) is presented in its formal statement, the problem-solving strategy can be reformulated into an algorithmic procedure and solved by the means of computer programming tools. Such tools for computer-aided design assist designer to accelerate the process of decision-making and to enhance the fidelity of the results based on rigorous analysis. Examples of these programs are the simulation software tools by AspenTech² (e.g. Aspen Plus and Aspen HYSYS), PSE³ (gPROMS) and MathWorks⁴ (e.g. Simulink).

The author suggests that the methods discussed in this review (e.g., the controllability measures, the methods for analysing nonlinear behaviour of chemical processes, the geometric operability analysis, the method based on an inversely controlled process model and so on) should be incorporated as built-in modules and library functions into the process systems engineering software tools. Then, after the process is modelled, the software tool would provide the option to the designer to evaluate the control performance of the designed process using an automated procedure. The abovementioned built-in modules would enhance the computational capabilities, available to the industrial practitioners, in order to efficiently consider the controllability characteristics of the process at early stages of process design.

6. Conclusion

In this paper, a thematic review of literature regarding process design and control was presented. Fig. 1 gave an overview of research in the field. The main approaches for process design and control can be classified into sequential methods and integrated design and control methods. The sequential methods have a yes/no attitude to the problem while the integrated design and control methods incorporate some control aspects into the process design. All the above methods use mathematical modelling. However, the methods using first principles modelling are more successful in integrating design and control.

Due to high dimensionality of the problem, a variety of methods addresses the problem by decomposing it into several smaller subproblems. Decomposition can be based on individual unit operations, different time-scales, prioritization of control objectives, or heuristics for the design of inventory control systems.

This paper also reviewed the characteristics and desired properties of the elements of control systems. Spatial and temporal decentralizations of control systems were explained and conventional multi-loop controllers and centralized model predictive controllers were discussed. This paper also discussed the desirable properties of manipulated and controlled variables. The economic implications of static and dynamic setpoint policies were discussed and the importance of selection of controlled variables for process profitability was emphasized.

The causes of control imperfection, also limit process controllability. Different definitions for operability, flexibility, and controllability were presented and the causes of control imperfection namely the interactions between control loops, delays and right-half-plane zeros, manipulated variable constraints and model uncertainties were discussed in this paper. Moreover, it was explained that the methods based on passivity, exploit the process model to evaluate the stability and integrity of decentralized control structures.

Then, the discussions moved to the methods in which process design and control are integrated to some extents. A category of optimization-based methods uses a multi-objective function for screening

² Web address: <http://www.aspentech.com/>

³ Web address: <http://www.psenterprise.com/gproms/>

⁴ Web address: <http://www.mathworks.co.uk/>

alternative solutions. This also provides the opportunity for incorporating controllability measures into economic optimization. These methods were reviewed in this paper.

A variety of methods is devoted to flexibility analysis, i.e., the question of whether for a range of uncertain scenarios, the process operation remains feasible. As discussed in this paper, the optimization methods for steady-state and dynamic operability analyses are developed by researchers. In addition, it is possible to evaluate the feasibility of the process operation by mapping the bounds of the input variables into the output spaces. This idea resulted in the geometric methods for operability analysis.

It was also discussed that minimizing the economic losses associated with disturbances, in terms of back-off from active constraints, can be applied as an economic measure for integrated design and control. By development of computational capabilities, some researchers optimized the process and its controllers simultaneously. However, the underlying formulation features combinatorial nature and is limited to smaller problems.

The comparisons between the methods on the right branch of the Fig. 1 are illustrative. All these methods try to establish criteria for evaluating and screening the performances of alternative decisions in designing process and control systems. Some methods employ the controllability measures, and incorporate them into a multi-objective function. In the methods based on model reduction, robust control measures are used instead. In the methods for analysing the nonlinear behaviour of chemical processes, the aim is to avoid undesirable characteristics such as steady-state multiplicity. The geometric methods for operability analysis are trying to ensure that, using available inputs, for all disturbance scenarios, the desired outputs are achievable. Similarly, the methods for flexibility optimization, try to evaluate and quantify the effects of uncertain parameters on feasible process operation. In some research, the decision-making criterion is the economic losses associated with retreat from active constraints. Finally, the methods for simultaneous optimization of controller and process directly measure the controller error and incorporate it to a multi-objective function.

Furthermore, investigating the evolution path of the methods for integrated design and control suggests that the methods which have a direct link to the underlying first principles are more successful in integrating process design and control. This is the reason that almost all of the methods on the right branch of Fig. 1 are nonlinear. Furthermore, simultaneous optimization of process and controllers pose a tough challenge for the current optimization methods, and requires efficient complexity reduction methods. The requirement for complexity reduction should address both numerical and conceptual complexities, in terms of the required computational costs, reliability of the solution and the desirable properties such as controllability, operability, and flexibility.

7. Acknowledgement

The author gratefully acknowledges partial financial support from the 2010 and 2011 ISA Educational Foundation scholarships, the Burkett Scholarship and Ure bursary award of Chemical Engineering Department, Imperial College London.

8. Bibliography

- Alaña, J. E., 2010. Optimal measurement locations for parameter estimation of non linear distributed parameter systems. *Brazilian Journal of Chemical Engineering*, 27(4), 627 – 642.
- Alhammadi, H., Romagnoli, J., 2004. Process Design and Operation: Incorporating Environmental, Profitability, Heat Integration and Controllability Considerations. In *The Integration of Process Design and Control*. Volume 17, Edited by Seferlis, P., Georgiadis, C. M., Elsevier Science, Amsterdam, 264-306.
- Alstad, V., Skogestad, S., Hori, E. S., 2009. Optimal measurement combinations as controlled variables. *J. Process. Contr.* 19 (1), 138-148.
- ANSI/ISA-95, 2000. Enterprise-Control System Integration Part 1: Models and Terminology. Online: <http://www.isa-95.com/subpages/technology/isa-95.php> accessed March 2012.
- ANSI/ISA-95, 2001. Enterprise-Control System Integration Part 2: Object Model Attributes. Online: <http://www.isa-95.com/subpages/technology/isa-95.php> , accessed March 2012.
- ANSI/ISA-95, 2005. Enterprise Control System Integration Part 3: Activity Models of Manufacturing

- Operations Management. Online: <http://www.isa-95.com/subpages/technology/isa-95.php> accessed March 2012.
- Ariyur, K. B., Krstic, M., 2003. Real-time optimization by extremum seeking. John Wiley, New York.
- Aske, E. M. B., Skogestad, S., 2009a. Dynamic degrees of freedom for tighter bottleneck control. 10th International Symposium on Process Systems Engineering, Salvador-Bahia, Brazil, August 16-20.
- Aske, E. M. B., Skogestad, S., 2009b. Consistent inventory control. *Ind. Eng. Chem. Res.* 48 (24), 10892-10902.
- Aske, E. M. B., 2009c. Design of Plantwide Control Systems with Focus on Maximizing Throughput. PhD Thesis. Department of Chemical Engineering, Norwegian University of Science and Technology.
- Asteasuain, M., Bandoni, A., Sarmoria, C., Brandolin, A., 2006. Simultaneous process and control system design for grade transition in styrene polymerization. *Chem. Eng. Sci.* 61 (10), 3362-3378.
- Avraam, M.P., Shah, N., Pantelides, C.C., 1998. Modelling and optimisation of general hybrid systems in the continuous time domain. *Comput. Chem. Eng.* 22 Supplement, S221-S228.
- Avraam, M.P., Shah, N., Pantelides, C.C., 1999. A decomposition algorithm for the optimisation of hybrid dynamic processes. *Comput. Chem. Eng.* 23 Supplement, S451-S454.
- Bagajewicz, M. J. 1997. Design and retrofit of sensor networks in process plants. *AIChE J.* 43 (9), 2300-2306.
- Bansal, V., Perkins, J. D., Pistikopoulos, E. N., Ross, R., van Schijndel, J. M. G., 2000a. Simultaneous design and control optimisation under uncertainty. *Comput. Chem. Eng.* 24 (2-7), 261-266.
- Bansal V., Ross R, Perkins J. D., Pistikopoulos E. N. 2000b. The interactions of design and control; double-effect distillation. *J. Process. Contr.* 10 (2-3), 219-227.
- Bansal, V., Perkins, J. D., Pistikopoulos, E. N., 2002. A case study in simultaneous design and control using rigorous, mixed-integer dynamic optimization models. *Ind. Eng. Chem. Res.* 41 (4), 760-778.
- Bansal, V., Sakizlis, V., Ross, R., Perkins, J.D., Pistikopoulos, E.N., 2003. New algorithms for mixed-integer dynamic optimization. *Comput. Chem. Eng.* 27, 647-668.
- Barton, P.I., Lee, C.K., 2004. Design of process operations using hybrid dynamic optimization. *Comput. Chem. Eng.* 28, 955-969.
- Bao J., Lee P. L., 2007. *Process Control: The Passive Systems Approach*. Advances in Industrial Control, Springer-Verlag, London.
- Bauer M., Thornhill N. F., Meaburn, A., 2004. Specifying the directionality of fault propagation paths using Transfer entropy. 7th international symposium on DYNamics and CONtrol of Process Systems DYCOPS 7, Boston, USA, July 1-4, 2004.
- Bauer, M., Cox, J. W., Caveness, M. H., Downs J. J., Thornhill, N. F., 2007. Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE T. Contr. Syst.* 15 (1), 12-21.
- Bhushan, M., Rengaswamy, R., 2000. Design of sensor network based on the signed directed graph of the process for efficient fault diagnosis. *Ind. Eng. Chem. Res.* 39, 999-1019.
- Biegler, L.T., Grossmann, I. E., Westerberg, A.W., 1997. *Systematic Methods for Chemical Process Design*. Prentice-Hall, London.
- Biegler, L. T., Grossmann, I. E., 2004. Retrospective on optimization. *Comput. Chem. Eng.* 28 (8), 1169-1192.
- Bogle, I.D.L., Ma K., Hagemann, J., Fraga E.S., 2004. Analysing the controllability of nonlinear process systems. In *The Integration of Process Design and Control*. Volume 17, Edited by Seferlis, P., Georgiadis, C. M., Elsevier Science, Amsterdam, pp. 168-186.
- Brdys, M., Tatjewski, P., 2005. *Iterative algorithms for multilayer optimizing control*. Imperial College Press, London.
- Bristol, E., 1966. On a new measure of interaction for multivariable process control. *IEEE T. Automat. Contr.* 11 (1), 133-134.
- Buckley, P. S., 1964. *Techniques of Process Control*, Wiley, New York.

- Caballero, J. A., Odjo, A., Grossmann, I. E., 2007. Flowsheet optimization with complex cost and size functions using process simulators. *AIChE J.* 53 (9), 2351–2366.
- Cao, Y., Biss, D., Perkins, J. D., 1996. Assessment of input-output controllability in the presence of control constraints. *Comput. Chem. Eng.* 20 (4), 337–346.
- Cao Y., Yang Z., 2004. Multiobjective process controllability analysis. *Comput. Chem. Eng.* 28 (1-2), 83–90.
- Cao, Y., 2005. Direct and indirect gradient control for static optimisation. *International Journal of Automation and Computing* 2 (1), 13–19.
- Cao, Y., Kariwala, V., 2008. Bidirectional branch and bound for controlled variable selection Part I. Principles and minimum singular value criterion. *Comput. Chem. Eng.* 32 10, 2306–2319.
- Cervantes, A.M., Biegler, L.T., 2000a. Optimization strategies for dynamic systems, in: Floudas, C., Pardalos, P., Eds., *Encyclopedia of Optimization*, Kluwer.
- Cervantes, A.M., Biegler, L.T., 2000b. A stable elemental decomposition for dynamic process optimization. *Journal of Applied Mathematics and Computing* 120 (1-2), 41–57.
- Cervantes, A.M., Biegler, L.T., 2002. A reduced space interior point strategy for optimization of differential algebraic systems. *Comput. Chem. Eng.* 24 (1), 39–51.
- Chachuat, B., Singer, A.B., Barton, P.I., 2005. Global mixed-integer dynamic optimization. *AIChE J.* 51 (8), 2235–2253.
- Chachuat, B., Singer, A.B., Barton, P.I., 2006. Global Methods for Dynamic Optimization and Mixed-Integer Dynamic Optimization. *Ind. Eng. Chem. Res.* 45 (25), 8373–8392.
- Chachuat, B., Marchetti, A., Bonvin, D., 2008. Process optimization via constraints adaptation. *J. Process. Contr.* 18 (3-4), 244–257.
- Chachuat, B., Srinivasan, B., Bonvin, D., 2009. Adaptation strategies for real-time optimization. *Comput. Chem. Eng.* 33 (10), 1557–1567.
- Chachuat, B., 2010. Introduction to the special issue on optimal process control. *Optim. Contr. Appl. Met.* 31 (5), 391–392.
- Chamseddine A. , Noura H., 2012. Sensor network design for complex systems. *Int. J. Adapt. Control.* 26, 496–515.
- Chang, J. W, Yu, Ch., 1990. The relative gain for non-square multivariable systems. *Chem. Eng. Sci.* 45 (5), 1309–1323.
- Chang, J. W, Yu, Ch., 1992. Relative disturbance gain array. *AIChE J.* 38 (4), 521–534.
- Choudhury, M.A.A.S., Shah, S.L., and Thornhill, N.F., 2008. *Diagnosis of Process Nonlinearities and Valve Stiction: Data Driven Approaches*, Springer, Berlin. ISBN: 978-3-540-79223-9.
- Chacon-mondragon, O. L., Himmelblau, D. M., 1996. Integration of flexibility and control in process design. *Comput. Chem. Eng.* 20 (4), 447–452.
- Chawanku, N., Budman, H., Douglas, P.L., 2005. The integration of design and control: IMC control and robustness. *Comput. Chem. Eng.* 29 (2), 261–271.
- Chmielewski, D. J., Palmer, T., Manousiouthakis, V., 2002. On the theory of optimal sensor placement. *AIChE J.* 48 (5), 1001–1012.
- Chen, C. Y., Joseph, B., 1987. On-line optimization using a two-phase approach: An application study. *Ind. Eng. Chem. Res.* 26 (9), 1924–1930.
- Cozad, A., Sahinidis, N., Miller, D., 2011. A computational methodology for learning low-complexity surrogate models of process from experiments or simulations. *AIChE Annual Meeting*, Minneapolis, October 16-21, 2011. Paper 679a.
- Daoutidis, P., Kravaris, C., 1992. Structural evaluation of control configurations for multivariable nonlinear processes. *Chem. Eng. Sci.* 47 (5), 1091–1107.
- Darby, M.L., Nikolaou M., 2012. MPC: Current practice and challenges. *Control Engineering Practice* (20) 328–342.
- Dimitriadis, V., Pistikopoulos, E. N., 1995. Flexibility analysis of dynamic systems. *Ind. Eng. Chem. Res.* 34 (12), 4451–4462.
- DICOPT, 2012. Documentation, (Online: www.gams.com/dd/docs/solvers/dicopt.pdf, accessed April 2012).
- Dixon, D.C., 1972. Degrees of Freedom in Dynamic and Static Systems. *Ind. Eng. Chem. Fund.* 11 (2), 198–205.

- Dochain, D., Tali-Mammar, N., Babary, J. P., 1997. On modeling, monitoring and control of fixed bed bioreactors. *Comput. Chem. Eng.*, 21, 1255.
- Dong, G., Chongguang, W. U., Beike., Z., Xin., M. A., 2010. Signed directed graph and qualitative trend analysis based fault diagnosis in chemical industry. *Chinese Journal of Chemical Engineering*, 18 (2) 265-276
- Douglas, J., 1988. *Conceptual Design of Chemical Processes*. McGraw-Hill, New York.
- Downs, J.J., Vogel, E.F., 1993. A plant-wide industrial process control problem. *Comput. Chem. Eng.* 17 (3), 245–255.
- Downs J. J., Skogestad S., 2011. An industrial and academic perspective on plantwide control. *Annual Reviews in Control* 35 (1), 99–110.
- Edgar, T. F., 2004. Control and operations: when does controllability equal profitability? *Comput. Chem. Eng.* 29 (1), 41-49.
- Exler, O., Antelo, L.T., Egea, J. A., Alonso, J A.A., Banga R., 2008. A Tabu search-based algorithm for mixed-integer nonlinear problems and its application to integrated process and control system design. *Comput. Chem. Eng.* 32 (8), 1877-1891.
- Fan, Y., De-yun, X., 2007. Sensor location problem for large-scale complex systems in the framework of the dynamic Signed Directed Graph. *IEEE International Conference on systems, Man and Cybernetics, ISIC*.
- Flores-Tlacuahuac, A., Biegler, L.T., 2005. Dynamic optimization of HIPS open loop unstable polymerization reactors. *Ind. Eng. Chem. Res.* 44 (8), 2659–2674.
- Flores-Tlacuahuac, A., Biegler, L.T., 2007. Simultaneous mixed-integrated dynamic optimization for integrated design and control. *Comput. Chem. Eng.* 31 (5-6), 588–600.
- Flores-Tlacuahuac A., Biegler L.T., 2008. Global optimization of highly nonlinear dynamic systems. *Ind. Eng. Chem. Res.* 47 (8), 2643–2655.
- Forbes, J. F., Marlin, T. E., 1994. Model accuracy for economic optimizing controllers: The bias update case. *Ind. Eng. Chem. Res.* 33 (8), 1919–1929.
- Floudas, C.A., Akrotirianakis, I.G., Caratzoulas, S., Meyer, C.A., Kallrath, J., 2005. Global optimization in the 21st century: Advances and challenges. *Comput. Chem. Eng.* 29, 1185–1202.
- Fraleigh, L.M., Guay, M., Forbes, J.F., 2003. Sensor selection for model-based real-time optimization: relating design of experiments and design cost, *Journal of Process Control*, 13 (7), 667–678.
- Francois, G., Srinivasan, B., Bonvin, D., 2005. Use of measurements for enforcing the necessary conditions of optimality in the presence of constraints and uncertainty. *J. Process. Contr.* 15 (6), 701–712.
- Gao, W., Engell, S., 2005. Iterative set-point optimization of batch chromatography. *Comput. Chem. Eng.* 29 (6), 1401–1409.
- Garcia, C. E., Morari, M., 1982. Internal model control. 1. A unifying review and some new results. *Ind. Eng. Chem. Proc. D. D.* 21 (2), 309-332.
- Geoffrion, A.M., 1972. Generalized benders decomposition. *J. Optimiz. Theory App.* 10 (4), 237-260.
- Georgakis, C., Uztürk, D., Subramanian, S., Vinson, D.R., 2003. On the operability of continuous processes. *Control Eng. Pract.* 11 (8), 859–869.
- Georgakis, C., Vinson, D. R., Subramanian, S., Uztiirk, D., 2004. A geometric approach for process operability analysis. In *The Integration of Process Design and Control*. Volume 17, Edited by Seferlis, P., Georgiadis, C. M., Elsevier Science, Amsterdam, pp. 96-126.
- Georgakis, C., Li, L., 2010. On the calculation of operability sets of nonlinear high-dimensional processes. *Ind. Eng. Chem. Res.* 49 (17), 8035–8047.
- Govatsmark, M. S., Skogestad, S., 2005. Selection of controlled variables and robust setpoints. *Ind. Eng. Chem. Res.* 44 (7), 2207–2217.
- Grossmann, I. E., Floudas, C. A., 1987. Active constraint strategy for flexibility analysis in chemical processes. *Comput. Chem. Eng.* 11 (6), 675–693.
- Grossmann, I. E., 2002. Review of nonlinear mixed- integer and disjunctive programming techniques. *Optimization & Engineering*, 3, 227-252.
- Guay M., 1997. Measurement of Nonlinearity in Chemical Process Control. PhD Thesis. Department

- of Chemical Engineering, Queen's University at Kingston.
- Guay, M., Zhang, T., 2003. Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainty. *Automatica* 39 (7), 1283–1294.
- Halemane, K. P., Grossman, I. E., 1983. Optimal process design under uncertainty. *AIChE J.* 29 (3), 425–433.
- Halvorsen, I. J., Skogestad, S., Morud, J. C., Alstad, V., 2003. Optimal selection of controlled variables. *Ind. Eng. Chem. Res.* 42 (14), 3273–3284.
- Hangos, K. M., Tuza, Zs., 2001. Optimal control structure selection for process systems. *Comput. Chem. Eng.* 25 (11–12), 1521–1536.
- Hahn J., Edgar T.F., 2002. An improved method for nonlinear model reduction using balancing of empirical Gramians *Computers and Chemical Engineering*, 26 (10), 1379–1397.
- Harjunkski, I., Nyström, R., Horch A., 2009. Integration of scheduling and control - Theory or practice? *Comput. Chem. Eng.* 33 (12), 1909–1918.
- Heath J. A., Kookos I. K., Perkins J. D., 2000. Process control structure selection based on economics. *AIChE J.*, 46 10, 1998–2016.
- Helbig, A., Marquardt, W., Allgower, F., 2000. Nonlinearity measures: definition, computation and applications. *Journal of Process Control*, 10, 113–123.
- Hirschorn, R., 1979. Invertibility of nonlinear control systems. *IEEE T. Automat. Contr.* 24 (6), 855 – 865.
- Holt, B. R., Morari, M., 1985. Design of resilient processing plants-V: The effect of deadtime on dynamic resilience. *Chem. Eng. Sci.* 40 (7), 1229–1237.
- Hori, E.S., Skogestad S., 2007. Selection of control structure and temperature location for two-product distillation columns. *Chem. Eng. Res. Des.* 85 (3), 293–306.
- Hori, E. S., Skogestad, S., 2008. Selection of controlled variables: Maximum gain rule and combination of measurements. *Ind. Eng. Chem. Res.* 47 (23), 9465–9471.
- Hovd, M., Skogestad, S., 1992. Simple frequency-dependent tools for control system analysis, structure selection and design. *Automatica*.28 (5), 989–996.
- Isidori, A., 1989. *Nonlinear Control Systems: An Introduction*. Second Edition, Springer Verlag, Berlin.
- Joshi, S. Boyd, S., 2009. Sensor Selection via Convex Optimization. *IEEE Transactions on Signal Processing*, 57 (2), 451 – 462.
- Kalman, R. E., 1960. On the general theory of control systems. In: *Proceedings of First IFAC Congress*, Butterworths, Moscow, 481–492.
- Kariwala, V. 2007. Optimal measurement combination for local self-optimizing control. *Ind. Eng. Chem. Res.* 46 (11), 3629–3634.
- Kariwala V., Cao, Y., Janardhanan. S., 2008. Local self-optimizing control with average loss minimization. *Ind. Eng. Chem. Res.* 47 (4), 1150–1158.
- Kariwala, V., Cao, Y., 2009. Bidirectional branch and bound for controlled variable selection. Part II: Exact local method for self-optimizing control. *Comput. Chem. Eng.* 33 (8), 1402–1412.
- Kariwala, V., Cao, Y., 2010a. Bidirectional branch and bound for controlled variable selection part iii: Local average loss minimization. *IEEE Trans. Ind. Inf.* 6 (1), 54 - 61.
- Kariwala V., Cao, Y., 2010b. Branch and bound method for multiobjective pairing selection. *Automatica* 46 (5), 932–936.
- Kiss, A.A., Bildea, C.S., Dimian, A.C., Iedema, P.D., 2002. State multiplicity in CSTR–separator–recycle polymerisation systems. *Chem. Eng. Sci.* 57 (4), 535–546.
- Kiss, A.A., Bildea, C.S., Dimian, A C., Iedema, P. D., 2003. State multiplicity in PFR–separator–recycle polymerization systems. *Chem. Eng. Sci.* 58 (13), 2973–2984.
- Kiss, A.A., Bildea, C.S., Dimian, A.C., 2007. Design and control of recycle systems by non-linear analysis. *Comput. Chem. Eng.* 31 (5–6), 601–611.
- Khajuria, H., Pistikopoulos, E. N., 2011. Dynamic modeling and explicit/multi-parametric MPC control of pressure swing adsorption systems. *J. Process. Contr.* 21 (1), 151–163.
- Klatt K.-U., Marquardt W., 2009. Perspectives for process systems engineering - Personal views from academia and industry. *Comput. Chem. Eng.* 33 (3), 536–550.
- Konda, N.V.S.N. M., Rangaiah, G.P., Krishnaswamy, P.R., 2006. A simple and effective procedure

- for control degrees of freedom. *Chem. Eng. Sci.* 61, (4), 1184–1194.
- Konda, N.V.S.N. M., Rangaiah, G.P., 2012. Control degree of freedom analysis for plantwide control of industrial processes. In *Plantwide Control: Recent Developments and Applications*. First Edition. Edited by Rangaiah, G.P., Kariwala, V., John Wiley, Chichester, pp. 21–42.
- Kookos, I. K., Perkins, J. D., 2001. An algorithm for simultaneous process design and control. *Ind. Eng. Chem. Res.* 40 (19), 4079–4088.
- Kookos, I. K., 2001. Control Structure Selection Based on Economics. PhD thesis, Department of Chemical Engineering and Chemical Technology, Imperial College of Science, Technology and Medicine.
- Kookos, I. K., Perkins, J. D., 2004. The back-off approach to simultaneous design and control. In *The Integration of Process Design and Control*. Volume 17, Edited by Seferlis, P., Georgiadis, C. M., Elsevier Science, Amsterdam, pp. 216–238.
- Kuhlmann, A. Bogle, I. D. L., 2001. Controllability evaluation for nonminimum phase-processes with multiplicity. *AIChE J.* 47 (11), 2627–2632.
- Kuhlmann, A. Bogle, I. D. L., 2004. Design of nonminimum phase processes for optimal switchability. *Chem. Eng. Process.* 43 (5), 655–662.
- Larsson, T., Skogestad, S., 2000a. Plantwide control - A review and a new design procedure. *Model. Ident. Control* 21 4, 209–240.
- Lewin, D.R., Bogle, D., 1996. Controllability analysis of an industrial polymerization reactor. *Comput. Chem. Eng.* 20 (2), S871–S876.
- Low K.H., Sorensen, E., 2003a. Simultaneous optimal design and operation of multivessel batch distillation. *AIChE J.* 49 (10), 2564–2576.
- Low K.H., Sorensen, E., 2003b. Simultaneous optimal design and operation of multipurpose batch distillation. *Chem. Eng. Process.* 43 (3), 273–289.
- Low, K.H., Sorensen, E., 2005. Simultaneous optimal configuration, design and operation of batch distillation. *AIChE J.* 51 (6) 1700–1713.
- Luyben, M. L., Floudas, C. A., 1994. Analyzing the interaction of design and Control - 1 . A multiobjective framework and application to binary distillation synthesis. *Comput. Chem. Eng.* 18 (10), 933–969.
- Luyben, M. L., Luyben, W. L., 1995. Design and control of a complex process involving two reaction steps, three distillation columns, and two recycle streams. *Ind. Eng. Chem. Res.* 34 (11), 3885–3898.
- Luyben, M. L., Tyreus, B. D., Luyben, W. L., 1997. Plantwide control design procedure. *AIChE J.* 43 (12), 3161–3174.
- Luyben W. L., 1993a. Dynamics and control of recycle systems. 1. Simple open-loop and closed-loop systems. *Ind. Eng. Chem. Res.* 32 (3), 466–475.
- Luyben W. L., 1993b. Dynamics and control of recycle systems. 2. Comparison of alternative process designs. *Ind. Eng. Chem. Res.* 32 (3), 476–486.
- Luyben W. L., 1993c. Dynamics and control of recycle systems. 3. Alternative process designs in a ternary system. *Ind. Eng. Chem. Res.* 32 (6), 1142–1153.
- Luyben, W.L., 1994. Snowball effects in reactor/separator processes with recycle. *Ind. Eng. Chem. Res.* 33 (2), 299–305.
- Luyben, W. L., 1996. Simple regulatory control of the Eastman process. *Ind. Eng. Chem. Res.* 35 (10), 3280–3289.
- Luyben, W. L., Luyben, M. L., Tyreus B. D., 1999. *Plantwide Process Control*. McGraw-Hill, New York.
- Luyben, W. L., 2004. The need for simultaneous design education. In *The Integration of Process Design and Control*. Volume 17, Edited by Seferlis, P., Georgiadis, C. M., Elsevier Science, Amsterdam, pp. 10–41.
- Luyben, W. L., 2005. Guides for the selection of control structures for ternary distillation columns. *Ind. Eng. Chem. Res.* 44 (18), 7113–7119.
- Luyben, W. L., 2006. *Distillation Design and Control using Aspen Simulation*. John Wiley, Hoboken.
- Lyman, P. R., Georgakis, C., 1995. Plant-wide control of the Tennessee Eastman problem. *Comput. Chem. Eng.* 19 (3), 321–331.

- Maarleveld, A., Rijnsdorp, J. E., 1970. Constraint control on distillation columns. *Automatica* 6 (1), 51–58.
- Malcolm, A., Polan, J., Zhang, L., Ogunnaike B. A., Linninger A. A., 2007. Integrating systems design and control using dynamic flexibility analysis. *AIChE J.* 53 (8), 2048–2061.
- Manousiouthakis, V., Nikolau, M., 1989. Analysis of decentralised control structures for nonlinear systems. *AIChE J.* 35 (4), 549–558.
- Marchetti, A., Chachuat, B., Bonvin D., 2009. Modifier-adaptation methodology for real-time optimization. *Ind. Eng. Chem. Res.* 48 (13), 6022–6033.
- Marlin, T. E., Hrymak, A. N., 1997. Real-time operations optimization of continuous processes. *AIChE Symposium Series-CPC-V*, 93, 156–164.
- Marquardt, W., Mönnigmann, M., 2005. Constructive nonlinear dynamics in process systems engineering. *Comput. Chem. Eng.* 29 (6), 1265–1275.
- Maurya, M. R., Rengaswamy, R., Venkatasubramanian V., 2003a. A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis. *Ind. Eng. Chem. Res.* 42 (20), 4789–4810.
- Maurya, M. R., Rengaswamy, R., Venkatasubramanian V., 2003b. A systematic framework for the development and analysis of signed digraphs for chemical processes. 2. Control loops and flowsheet analysis. *Ind. Eng. Chem. Res.* 42 (20), 4811–4827.
- Maurya, M. R., Rengaswamy, R., Venkatasubramanian V., 2004. Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets. *Eng. Appl. Artif. Intel.*, 17 (5), 501–518.
- Maurya, M. R., Rengaswamy, R., Venkatasubramanian, V., 2007. A signed directed graph and qualitative trend analysis-based framework for incipient fault diagnosis. *Trans IChemE, Part A, Chemical Engineering Research and Design*, 85 (A10): 1407–1422.
- McAvoy, T. J., Ye, N., 1994. Base control for the Tennessee Eastman problem. *Comput. Chem. Eng.* 18 (5), 383–413.
- McAvoy, T.J., Arkun, Y., Chen R., Robinson, D., Schnelle, P.D., 2003. A new approach to defining a dynamic relative gain. *Control Eng. Pract.* 11 (8), 907–914.
- Moaveni, B., Khaki-Sedigh, A., 2009. *Control Configuration Selection for Multivariable Plants*. Springer, Berlin.
- Moaveni, B., Kariwala, V., 2012. Input-Output Pairing Selection for Design of Decentralized Controller. In *Plantwide Control: Recent Developments and Applications*. First Edition. Edited by Rangaiah, G.P., Kariwala, V., John Wiley, Chichester, pp. 73–96.
- McLellan, P. J., 1994. A differential-algebraic perspective on nonlinear controller design methodologies. *Chem. Eng. Sci.* 49 (10), 1663–1679.
- Mohideen, M.J., Perkins, J.D., Pistikopoulos, E.N., 1997. Towards an efficient numerical procedure for mixed integer optimal control. *Comput. Chem. Eng.* 21 Supplement, S457–S462.
- Montgomery, D. C., 2005. *Design and Analysis of Experiments*. John Wiley, New York.
- Moon J., Kim S., Linninger, A.A., 2011. Embedded control for optimizing flexible dynamic process performance. *Ind. Eng. Chem. Res.* 50 (9), 4993–5004.
- Moore, C. F., 1992. Selection of controlled and manipulated variables. In *Practical Distillation Control*. Van Nostrand Reinhold, New York.
- Morari, M., Y. Arkun, Stephanopoulos G., 1980a. Studies in the synthesis of control structures for chemical processes. Part I: Formulation of the problem. Process decomposition and the classification of the control tasks. Analysis of the optimizing control structures. *AIChE J.* 26 (2), 220–232.
- Morari, M., Stephanopoulos G., 1980b. Studies in the synthesis of control structures for chemical processes. Part II: Structural aspects and the synthesis of alternative feasible control schemes. *AIChE J.* 26 (2), 232–246.
- Morari, M., Stephanopoulos G., 1980c. 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. *AIChE J.* 26 (2), 247–260.
- Morari, M., 1983. Design of resilient processing plants-III: A general framework for the assessment

- of dynamic resilience. *Chem. Eng. Sci.* 38 (2), 1881-1891.
- Morari, M., Zafiriou, E., 1989. *Robust Process Control*. Prentice-Hall International, Englewood Cliffs.
- Müller, P., Weber, H. 1972. Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems. *Automatica*, 8, 237–246.
- Muske, K. R., Georgakis, C., 2003. Optimal measurement system design for chemical processes. *AIChE J.* 49 (6), 1488–1494.
- Narraway L. T., Perkins, J. D., Barton, G. W., 1991, Interaction between process design and process control: economic analysis of process dynamics, *J. Process. Contr.* 1 (5), 243–250.
- Narraway, L. T., Perkins, J. D., 1993. Selection of process control structure based on linear dynamic economics. *Ind. Eng. Chem. Res.* 32 (11), 2681–2692.
- Ng, C., 1997. A systematic approach to the design of plant-wide control strategies for chemical processes. PhD thesis. Department of Chemical Engineering, Massachusetts Institute of Technology.
- Niederlinski A., 1971. A heuristic approach to the design of linear multivariable interacting control systems. *Automatica* 7 (6), 691-701.
- Ogunnaike, B. A., Ray, W. H., 1994. *Process Dynamics, Modeling, and Control*. Oxford University Press, New York.
- Patel, J., Uygun, K., Huang, Y., 2008. A path constrained method for integration of process design and control. *Comput. Chem. Eng.* 32 (7), 1373-1384.
- Pavan Kumar, M. V., Kaistha N., 2008a. Steady-state multiplicity and its implications on the control of an ideal reactive distillation column. *Ind. Eng. Chem. Res.* 2008, 47 (8), 2778–2787.
- Pavan Kumar, M. V., Kaistha N., 2008b. Decentralized control of a kinetically controlled ideal reactive distillation column. *Chem. Eng. Sci.* 63 (1), 228-243.
- Pedersen K., Jørgensen S. B., Skogestad, S., 1999. A review on process controllability. Unpublished internal paper, (Online: <http://130.203.133.150/viewdoc/download;jsessionid=71E5CA8CBC76B1E981251ACEB46ED923?doi=10.1.1.40.4192&rep=rep1&type=pdf> , Accessed January 2012).
- Perkins, J. D., Wong, M. P. F., 1985. Assessing controllability of chemical plants. *Chem. Eng. Res. Des.* 63, 358-362.
- Pham, Q.T., 1994. Degrees of freedom of equipment and processes. *Chemical Engineering Science*. 49 (15), 2507-2512.
- Ponton, J., Laing, D. M., 1993. A hierarchical approach to the design of process control systems. *Transaction of IChemE* 71, 181-188.
- Ponton J.W., 1994. Degrees of freedom analysis in process control. *Chem. Eng. Sci.* 49 (13), 2089-2095.
- Poston, T.; Stewart, I. 1996. *Catastrophe Theory and Its Applications*; Dover, London.
- Price R. M., Georgakis C., 1993. Plantwide regulatory control design procedure using a tiered framework. *Ind. Eng. Chem. Res.* 32 (11), 2693-2705.
- Price R. M., Lyman P. R., Georgakis C., 1994. Throughput manipulation in plantwide control structures. *Ind. Eng. Chem. Res.* 33 (5), 1197-1207.
- Qin, S. J., Badgwell, T. A., 2003. A survey of industrial model predictive control technology. *Control Eng. Pract.* 11 (7), 733-764.
- Rawlings, J. B., Stewart, B. T., 2008. Coordinating multiple optimization-based controllers: New opportunities and challenges. *J. Process. Contr.* 18 (9), 839-845.
- Ricardez-Sandoval, L. A., Budman, H. M., Douglas, P. L., 2008. Simultaneous design and control of processes under uncertainty: A robust modelling approach. *J. Process. Contr.*, 18 (7-8), 735-752.
- Ricardez-Sandoval, L. A., Budman, H. M., Douglas, P. L., 2009a. Integration of design and control for chemical processes: A review of the literature and some recent results. *Annual Reviews in Control* 33 (2), 158–171.
- Ricardez-Sandoval, L. A., Budman H. M., Douglas P. L., 2009b. Simultaneous design and control of chemical processes with application to the Tennessee Eastman process. *J. Process. Contr.* 19 (8), 1377–1391.
- Ricardez-Sandoval, L.A., Douglas, P.L., Budman, H.M., 2011. A methodology for the simultaneous

- design and control of large-scale systems under process parameter uncertainty. *Comput. Chem. Eng.* 35 (2), 307–318.
- Richalet, J., Rault, A., Testud, J. L., Papon, J., 1978. Model predictive heuristic control: Applications to industrial processes. *Automatica* 14 (5), 413–428.
- Ricker, N. L. 1996. Decentralized control of the Tennessee Eastman challenge process. *J. Process. Contr.* 6 (4), 205–221.
- Roberts, P. D., 1979. An algorithm for steady-state system optimization and parameter estimation. *Int. J. Syst. Sci.* 10 (7), 719–734.
- Rojas, O.J., Setiawan, R., Bao J., Lee, P.L., 2009. Dynamic operability analysis of nonlinear process networks based on dissipativity. *AIChE J.* 55 (4), 963–982.
- Rosenbrock, H. H., 1970. *State-space and Multivariable Theory*. Nelson, London.
- Ross R., Perkins, J. D., Pistikopoulos, E. N., Koot, G.L.M., van Schijndel, J.M.G., 2001. Optimal design and control of a high-purity industrial distillation system. *Comput. Chem. Eng.* 25 (1), 141–150.
- Russell, L.W., Perkins, J.D., 1987. Towards a method for diagnosis of controllability and operability problems in chemical-plants. *Chem. Eng. Res. Des.* 65 (6), 453–461.
- Sakizlis, V., Perkins, J.D., Pistikopoulos E.N., 2003. Parametric controllers in simultaneous process and design optimization. *Ind. Eng. Chem. Res.* 42 (20), 4545–4563.
- Sakizlis V., Perkins J. D., Pistikopoulos, E. N., 2004. Recent advances in optimization-based simultaneous process and control design. *Comput. Chem. Eng.* 28 (10), 2069–2086.
- Sakizlis V., Vakamudi K., Coward A. , Mermans I., 2010. Advanced process control in the plant engineering and construction phases. *Hydrocarb. Process.*, October 2010, 89 (10), 49–56.
- Sahinidis, N. V., 1996. BARON: A general purpose global optimization software package. *J. Global Optim.* 8 (2), 201–205.
- Samsatli, N. J., Papageorgiou, L. G., Shah, N., 1998. Robustness metrics for dynamic optimization models under parameter uncertainty. *AIChE J.* 449, 1993–2006.
- Seferlis P., Georgiadis M. C., 2004. *The Integration of Process Design and Control*, Elsevier Science, Amsterdam.
- Seider, W.D., Seader, J.D., Lewin, D.R, Widagdo, S., 2010. *Product and Process Design Principles: Synthesis, Analysis and Design*, 3rd Edition, John Wiley & Sons, NJ.
- Schweiger, C. A., Floudas, C. A., 1997. Interaction of design and control: Optimization with dynamic models. In *Optimal Control: Theory, Algorithms, and Applications*. Hager, W. W., Pardalos, P. M., Eds.; Kluwer Academic Publishers: Norwell, MA, 388–435.
- Sharif, M., Shah, N., Pantelides, C.C., 1998. On the design of multicomponent batch distillation columns. *Comput. Chem. Eng.* 22, S69–S76.
- Sharifzadeh M., Rashtchian D., Pishvaie M.R., Thornhill N.F., 2011. Energy induced separation network synthesis of an olefin compression section: a case study. *Ind. Eng. Chem. Res.* 50 (3), 1610–1623.
- Sharifzadeh, M., Thornhill, N.F., 2012a. Optimal selection of control structures using a steady-state inversely controlled process model. *Comput. Chem. Eng.* 38, 126–138.
- Sharifzadeh M., Thornhill N.F., 2012b. Integrated design and control using a dynamic inversely controlled process model, (accepted for publication at *Comput. Chem. Eng.*)
- Sharifzadeh, M., 2012. *Integrated Design and Control with a focus on Control Structures*. PhD Thesis, Department of Chemical Engineering, Imperial College London.
- Sharifzadeh, M., 2013. Implementation of a steady-state inversely controlled process model for integrated design and control of an ETBE reactive distillation. *Chem. Eng. Sci.* 92 (5) 21–39.
- Serpas M., Hackebeil G., Laird C., Hahn J., 2013. Sensor location for nonlinear dynamic systems via observability analysis and MAX-DET optimization, *Comput. Chem. Eng.* 105–112.
- Silva-Beard A., Flores-Tlacuahuac, A., 1999. Effect of Process Design/Operation on the Steady-State Operability of a Methyl Methacrylate Polymerization Reactor. *Ind. Eng. Chem. Res.* 38 (12), 4790–4804.
- Singh, A. K., Hahn, J., 2005. Determining optimal sensor locations for state and parameter estimation for stable nonlinear systems. *Ind. Eng. Chem. Res.*, 44, 5645–5659.

- Singh, A.K., Hahn, J. 2006. Sensor location for stable nonlinear dynamic systems: Multiple sensor case. *Ind. Eng. Chem. Res.*, 45, 3615–3623.
- Skogestad, S., Morari, M., 1987. Effect of disturbance directions on closed-loop performance. *Ind. Eng. Chem. Res.* 26 (10), 2029-2035.
- Skogestad, S., Morari, M., 1988. Understanding the dynamic behavior of distillation columns. *Ind. Eng. Chem. Res.* 27 (10), 1848-1862.
- Skogestad, S., 2000a. Plantwide control: the search for the self-optimizing control structure. *J. Process. Contr.* 10 (5), 487-507.
- Skogestad, S., 2000b. Self-optimizing control: the missing link between steady-state optimization and control. *Comput. Chem. Eng.* 24 (2-7), 569-575.
- Skogestad, S., 2004a. Control structure design for complete chemical plants. *Comput. Chem. Eng.* 28 (1-2), 219-234.
- Skogestad, S. 2004b. Near-optimal operation by self-optimizing control: from process control to marathon running and business systems. *Comput. Chem. Eng.* 29 (1), 127-137.
- Skogestad, S., Postlethwaite, I., 2005. *Multivariable Feedback Control: Analysis and Design*. Second Edition, John Wiley, Chichester.
- Skogestad, S. 2007. The dos and don'ts of distillation column control. *Chem. Eng. Res. Des.* 85 (1), 13-23.
- Skogestad, S. 2009. Feedback: still the simplest and best solution. *Model. Ident. Control*, 30 (3), 149-155.
- Slotine, J. E., Li, W., 1991. *Applied Nonlinear Control*. Prentice-Hall, Englewood Cliffs.
- Soroush, M., 1996. Evaluation of achievable control quality in nonlinear processes. *Comput. Chem. Eng.* 20 (4), 357–364.
- Srinivasan, B., Primus, C. J., Bonvin, D., Ricker, N. L., 2001. Run-to-run optimization via constraint control. *Control Eng. Pract.* 9 (8), 911–919.
- Srinivasan, B., Bonvin, D., 2002. Interplay between identification and optimization in run-to-run optimization schemes. In *Proceeding of American Control Conference ACC*. Anchorage, USA, May 8-10, 2174–2179.
- Srinivasan, B., Biegler, L.T., Bonvin D., 2008. Tracking the necessary conditions of optimality with changing set of active constraints using a barrier-penalty function. *Comput. Chem. Eng.* 32 (3), 572–579.
- Stephanopoulos, G., 2003. *Chemical Process Control: An Introduction to Theory and Practice*. Prentice-Hall, London.
- Stephanopoulos, G., Ng, C., 2000. Perspectives on the synthesis of plant-wide control structures. *J. Process. Contr.* 10 (2-3), 97-111.
- Stephanopoulos, G., Reklaitis, G., V., 2011. *Process systems engineering: from solvay to modern bio- and nanotechnology. A history of development, successes and prospects for the future.* *Chem. Eng. Sci.* 66, 4272–4306.
- Subramanian, S., Georgakis, C., 2001. Steady-state operability characteristics of idealized reactors. *Chem. Eng. Sci.* 56 (17), 5111–5130.
- Swaney, E., Grossmann I. E., 1985a. An index for operation flexibility in chemical process design. 1. Formulation and theory. *AIChE J.* 31 (4), 621–630.
- Swaney, R. E., Grossmann, I. E., 1985b. An index for operational flexibility in chemical process design. Part 11: Computational Algorithms. *AIChE J.* 31 (4), 631-641.
- Tatjewski, P., 2002. Iterative optimizing set-point control—The basic principle redesigned. In *Proceedings 15th World Congress of IFAC*, Volume 15 Part 1, Barcelona, July 2002.
- Tawarmalani M., Sahinidis N.V., 2004. Global optimization of mixed-integer nonlinear programmes: a theoretical and practical study. *Math. Program.* 99, 563–591.
- Terrazas-Moreno, S., Flores-Tlacuahuac A., Grossmann, I. E., 2008. Simultaneous design, scheduling, and optimal control of a methyl-methacrylate continuous polymerization reactor. *AIChE J.* 54 (12), 3160–3170.
- Thambirajah J., Benabbas L., Bauer M., Thornhill N. F., 2009. Cause-and-effect analysis in chemical processes utilizing XML, plant connectivity and quantitative process history. *Comput. Chem. Eng.* 33 (2), 503-512.

- Tung, L.S., Edgar, T.F., 1981, Analysis of control-output interactions in dynamic systems. *AIChE* . 27 (4) 690–693.
- Tyreus, B. D., Luyben, W. L., 1993. Dynamics and control of recycle systems. 4. Ternary systems with one or two recycle streams. *Ind. Eng. Chem. Res.* 32 (6), 1154-1162.
- Umara, L. M., Hub, W., Cao, Y., Kariwala, V., 2012. Selection of controlled variables using self-optimizing control method. In *Plantwide Control: Recent Developments and Applications*. First Edition. Edited by Rangaiah, G.P., Kariwala, V., John Wiley, Chichester, pp. 43-71.
- Uztürk, D., Georgakis, C., 2002. Inherent dynamic operability of processes: general definitions and analysis of siso cases. *Ind. Eng. Chem. Res.* 41 (3), 421–432.
- van den Berg, F. W. J., Hoefsloot, H. C. J., Boelens, H. F. M., Smilde, A. K. 2000. Selection of optimal sensor position in a tubular reactor using robust degree of observability criteria. *Chem. Eng. Sci.* 55 (4), 827–837.
- van de Wal, M., de Jager, B. 2001. A review of methods for input/output selection. *Automatica* 37 (4), 487–510.
- Vasudevan S., Konda, N.V.S.N. M., Rangaiah, G.P., 2008. Control degrees of freedom using the restraining number: further evaluation. *Asia Pac. J. Chem. Eng.* 3, 638–647.
- Venkatasubramanian, V., 2009. Drowning in data: Informatics and modeling challenges in a data-rich networked world. *AIChE J.* 55 (1), 2–8.
- Venkatasubramanian, V., 2011. Systemic failures: Challenges and opportunities in risk management in complex systems. *AIChE J.* 57 (1), 2–9.
- Vinson, D. R., Georgakis, C., 2000. A new measure of process output controllability. *J. Process. Contr.* 10 (2-3), 185-194.
- Viswanathan, J., Grossmann, I. E., 1990. A combined penalty function and outer-approximation method for MINLP optimization. *Comput. Chem. Eng.* 14 (7), 769-782.
- Ward, J. D., Doherty, M. F., Yu, C.-C., 2007. Plantwide operation and control of processes with crystallization. *AIChE J.* 53 (11), 2885–2896.
- Ward J. D., Yu C.-C., Doherty M. F., 2010. Plantwide dynamics and control of processes with crystallization. *Comput. Chem. Eng.* 34 (1), 112–121.
- Waldraff, W., Dochain, D.; Bourrel, S., Magnus, A. 1998. On the use of observability measures for sensor location in tubular reactor. *J. Process Control*, 8 (5-6), 497–505.
- Wongrat, W., Younes, A., 2011. Control vector optimization and genetic algorithms for mixed-integer dynamic optimization in the synthesis of rice drying processes. *J. Franklin. I.* 348 (7), 1318–1338.
- Xiong, Q., Cai, W.-J., He, M.-J., 2005. A practical loop pairing criterion for multivariable processes. *Journal of Process Control*. 15 (7), 741–747.
- Yang, F., Shah, S., Xiao, D., 2012. Signed directed graph based modeling and its validation from process knowledge and process data. *Int. J. Appl. Math. Comput. Sci.*, Vol. 22, No. 1, 41–53.
- Ydstie, B. E., 2010. Process Control: The Passive Systems Approach Review of, by Bao, J., Lee, P. L., *IEEE Contr. Syst. Mag.*, Feb. 2010, 78-80.
- Yim, S.Y., Ananthakumar, H.G., Benabbas, L., Horch A., Drath R., Thornhill, N. F., 2006. Using process topology in plant-wide control loop performance assessment. *Comput. Chem. Eng.* 31 (2), 86-99.
- Yuan, Z., Chen, B., Zhao, J., 2011. An overview on controllability analysis of chemical processes. *AIChE J.* 57 (5), 1185–1201.
- Yuan, Z., Chen, B., Sin, G., Gani, R., 2012. State-of-the-Art and Progress in the Optimization-based Simultaneous Design and Control for Chemical Processes. *AIChE J.* 58 (6), 1640–1659.
- Zhou, K., Doyle, J. C., Glover, K. 1996. *Robust and Optimal Control*. Englewood Cliff, New Jersey, Prentice-Hall.
- Ziegler, J.G., Nichols, N. B., Rochester, N. Y., 1942. Optimum settings for automatic controllers. *Transactions of ASME* 64, 759-768.