Preprints

16th Nordic Process Control Workshop

Lund, Sweden

25–27 August 2010

25 August	**Workshop Tutorial** **Department of Automatic Control** **Lund University**
26–27 August	**Nordic Process Control Workshop** **AF-Borgen**

Editor: Tore Hägglund

The 16th Nordic Process Control Workshop

Welcome the 16th Nordic Process Control Workshop (NPCW), and to Lund. The aim of the NPCW is to bring the Nordic process control community together, and to provide a rather informal forum for presenting recent and ongoing work in the process control area. The workshops are arranged with a period of one and a half year, circulating between Denmark, Finland, Norway, and Sweden.

The workshop is organized by the Nordic Working Group on Process Control, where the current members are: Hans Aalto, Neste Jacobs, Finland; Jan Peter Axelsson, Pfizer, Sweden; John Bagterp Jørgensen, DTU, Denmark; Claes Breitholtz, CTH, Sweden; Bjarne Foss, NTNU, Norway; Bjørn Glemmestad, Borealis, Norway; Kurt Erik Häggblom, Åbo Akademi, Finland; Tore Hägglund, LU, Sweden; Alf Isaksson, ABB, Sweden; Elling W. Jacobsen, KTH, Sweden; Sirkka-Liisa Jämsä-Jounela, Aalto Univ., Finland; Kaj Juslin, VTT, Finland; Annika Leonard, Vattenfall, Sweden; Bernt Lie, Telemark Univ. College, Norway; Tommy Mølback, Dong Energy, Denmark; Gürkan Sin, DTU, Denmark; and Sigurd Skogestad, NTNU, Norway.

Another responsibility of the Nordic Working Group on Process Control is to appoint recipients of the Nordic Process Control Award, one at each workshop. The award should be given to someone who has made "lasting and significant contributions to the field of process control". This year, professor Graham C. Goodwin has received the award, and we are happy that he will participate in the workshop and give a plenary talk with the title *Architectural Issues in Control System Design*. The award has been given to the following recipients during the last fifteen years.

1995	Howard H. Rosenbrock
1997	Karl Johan Åström
1998	Greg Shinskey
2000	Jens G. Balchen
2001	Charles R. Cutler
2003	Roger W. H. Sargent
2004	Ernst D. Gilles
2006	Manfred Morari
2007	Jasques Richalet
2009	John F. MacGregor
2010	Graham C. Goodwin

I hope this meeting in Lund will follow the tradition from the previous workshops and bring the Nordic process control community together and stimulate further research in our field and contacts between the participants.

Lund, August 2010

Tore Hägglund

Program

Thursday, August 26 2010

8.00 Registration

9.00 Opening

9.10 Award ceremony

Award presented to Professor Graham C Goodwin

Award lecture:

Architectural Issues in Control System Design
   Professor Graham C Goodwin

Abstract: Recent control literature places heavy emphasis on optimal design. For example, MPC uses on-line optimization to achieve optimal performance in the face of constraints, disturbances and modeling errors. This talk will make the point that optimization is the final step in a multi-stage design sequence. The first, and arguably the most important, stage is deciding on the control architecture. This stage is usually the most important and can make the difference between success and failure.

The ideas will be illustrated by several practical case studies including: centre line thickness control in reversing mills, networked control over communication channel, inner loop power control in WCDMA telecommunication systems, and reference tracking in model predictive control.

10.00 Break

10.30 Session 1: Modeling and Simulation

Accurate Dynamic Models for Type 1 Diabetes Identified from Novel   9
Clinical Data
   Daniel A. Finan, Signe Schmidt, John Bagterp Jørgensen
   Niels Kjølstad Poulsen, Kirsten Nørgaard, and Henrik Madsen

Commissioning a Distillation Column Simulator   11
   Ramkrishna Ghosh and Kurt-Erik Häggblom

Dynamic Modeling of Combustion in a BioGrate Furnace: a   12
Sensitivity Analysis on the Fuel Quality and Combustion Air Supply
   Alexandre Boriouchkine, Alexey Zakharov, and
   Sirkka-Liisa Jämsä-Jounela

Data-Based Uncertainty Modeling of MIMO Systems   18
   Hamed Jafarian and Kurt-Erik Häggblom

Modeling for Control of Beer Quality   20
   D.R. Warnasooriya, P.G. Rathnasiri, and Bernt Lie

12.10 Lunch

Friday, August 27 2010

7

Accurate Dynamic Models for Type 1 Diabetes Identified from Novel Clinical Data

Daniel A. Finan,[1] Signe Schmidt,[2] John Bagterp Jørgensen,[1] Niels Kjølstad Poulsen,[1] Kirsten Nørgaard,[2] and Henrik Madsen[1]

[1]*Department of Informatics and Mathematical Modeling, Technical University of Denmark, Kongens Lyngby, Denmark*
[2]*Department of Endocrinology, Hvidovre University Hospital, Hvidovre, Denmark*

The practical, day-to-day treatment regimen for people with type 1 diabetes (T1DM) entails self-administration of exogenous insulin in order to regulate blood glucose concentrations as close to normal levels as possible. The most efficacious dosing is achieved by measuring, as often as possible, blood glucose concentration through finger-sticks, and delivering insulin accordingly. To manage T1DM properly, therefore, is painful and requires constant decision making. By contrast, an *artificial pancreas* is a biomedical device in development that will automatically regulate blood glucose concentration while freeing patients of the daily burden of self-management.

A critical component of T1DM treatment is dosing insulin so as to offset the carbohydrate content of meals. By current standards, this often involves administration of insulin injections (called boluses) coincident with the meals. Moreover, these factors are commonly taken in a prescribed ratio, known as the insulin-to-carbohydrate ratio.

The control algorithm for an artificial pancreas may well be based on a mathematical model of the patient's glucose-insulin dynamics, as in a model predictive control framework. In such a model-based algorithm, it is advantageous to use an accurate, but simple, model. Linear dynamic models may provide sufficiently accurate predictions, and have other inherent advantages like straightforward, computationally tractable identification, and potential to be re-estimated online, thereby adapting to the evolving dynamics of the patient.

Unfortunately, the simultaneity and uniform proportionality of the meals and the insulin boluses confounds accurate estimation of the model parameters. To avoid this pitfall, we have devised a new in-clinic protocol based on design-of-experiment considerations that yields information-rich data for model identification. The protocol involves separating key factors that influence glucose concentration: meals, insulin boluses, and bouts of exercise.

The novel data give rise to more accurately identified models. A variety of empirical models was identified from the data: difference-equation models like autoregressive exogenous-input (ARX) and autoregressive moving-average exogenous-input (ARMAX) models, transfer-function (TF) models, and state-space models. In addition, "gray" forms of these models were identified which incorporate simple physiological elements such as estimates of subcutaneous-to-intravenous insulin absorption and appearance rate of glucose in the blood from a carbohydrate meal.

The quality of the identified models was based on several measures including:
- Accuracy with which the model was fit to the training data
- Values of parameters and/or combinations of parameters (e.g., steady-state insulin-to-glucose gain)
- Accuracy of predictions for independent (test) data

Identification results for one experiment are shown in Fig. 1. The experiment included three inputs: a meal, an episode of exercise, and a correction bolus, as depicted in the figure. The particular model shown was a transfer function model that was identified from the experimental data. With the exception of the initial, minor glycemic excursion between trial time -60 min and trial time 0 min, the model prediction is very accurate.
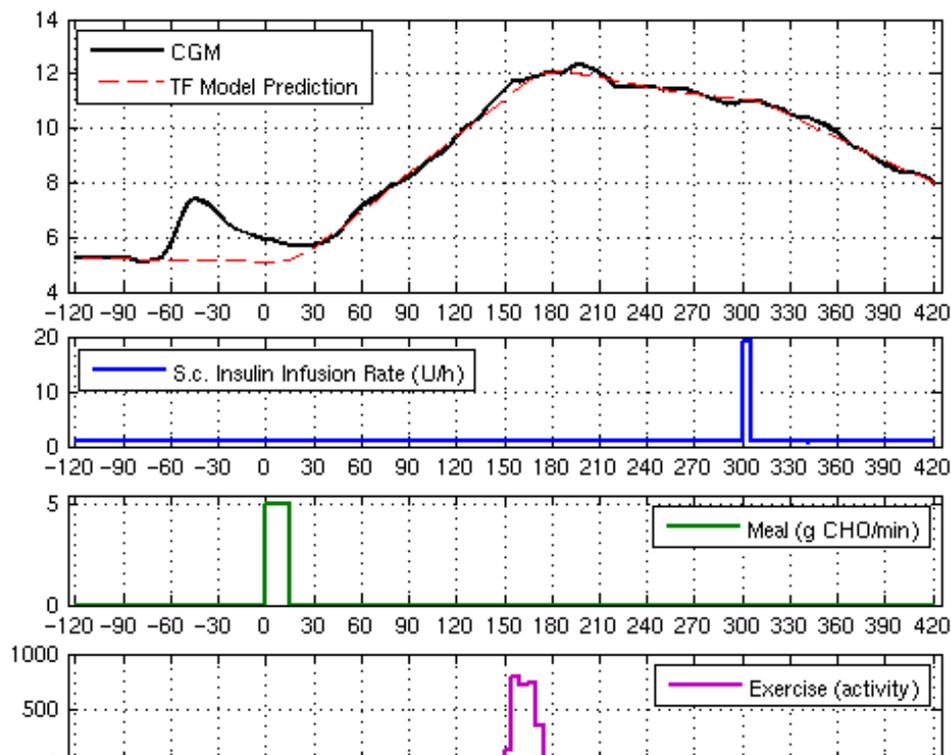
Fig. 1. Identification results for one subject. The TF model prediction is infinite-step. The experimental inputs included an unbolused meal at trial time 0 min, a bout of moderate exercise at trial time 150 min, and a correction bolus at trial time 300 min.

Commissioning a Distillation Column Simulator

Ramkrishna Ghosh and Kurt-Erik Häggblom

Process Control Laboratory, Department of Chemical Engineering,
Åbo Akademi University, Åbo, Finland

Abstract

Energy saving has become an extremely important issue in the chemical process industries. Distillation columns, in particular, consume huge amounts of energy. One way of minimizing the energy consumption is improved control, which enables operation closer to certain constraints.

It has been estimated that 75% of the cost associated with an advanced control project typically goes into model development (Gevers, 2005). Hence, efficient modeling and system identification techniques suited for industrial use and tailored for control design applications are needed. This task is especially difficult for "ill-conditioned" MIMO systems such as distillation columns. Even for a linear system, this ill-conditioning makes the system behavior resemble that of a (strongly) nonlinear system. Because of this, identification, modeling and control of ill-conditioned systems are demanding tasks.

In an industrial environment, the identification usually has to be carried out while the plant is in normal operation. It is then essential to keep the variation of inputs and outputs within specified limits and to limit the duration of the identification experiments. However, this also limits the information available for system identification. Thus, there is a trade-off between how much one is prepared to "pay" for the information and the information needed for system identification.

These kinds of identification issues can be investigated by means of a pilot-scale distillation column at Åbo Akademi. However, in order to enable more effective identification studies, a distillation column simulator has been constructed using MathWork's Simulink as programming environment. Because the simulator is to be used in conjunction with the real distillation column, it is desired that the behavior of the simulator is close to that of the real column. A significant number of previously performed identification experiments with the distillation column are available for the tuning of the simulator.

In order facilitate the simulator tuning, we study the effects of column and mixture property parameters appearing in the simulator model on observable static and dynamic properties. In particular, we want to find out which parameters most strongly affect these properties in various input-output relationships. Besides the tuning issue, this information also has more general interest.

Dynamic modeling of combustion in a BioGrate furnace: a sensitivity analysis on the fuel quality and combustion air supply

Alexandre Boriouchkine, Alexey Zakharov, Sirkka-Liisa Jämsä-Jounela

Aalto University, School of Science and Technology, Department of Biotechnology and Chemical Technology, Process Automation Research Group, 00076 Aalto, Finland. e-mail: aboriouc@cc.hut.fi

Abstract: This paper considers dynamic modeling of the bed combustion in a furnace of the BioGrate boiler. The developed dynamic model is heterogeneous, including solid and gas phases. Furthermore, the model considers chemical reactions in both, gas and solid phases. In addition, fuel movement on the grate is included in to the model. The energy required by the process is employed through a radiation function validated by industrial data from a BioGrate boiler at Trolhättan, Sweden. The model is implemented in MATLAB environment and tested with the industrial data. The results are presented and discussed.

1. Introduction

Increasing utilization of renewable energy has created new energy efficiency challenges for industry. Biomass is one of the most important raw material for renewable energy. All the available biomass sources have to be considered for energy production. Fuel properties of biomass vary a lot depending on its origin, on processing and handling for fuel. Variable properties cause fluctuations in combustion and set challenges to develop new combustion control strategies.

One of the latest successful processes developed, which use wood waste as a fuel, is a BioGrate-boiler technology developed by MWBiopower. The combustion of wood waste is, however, a very complex process involving several highly coupled chemical reactions. Furthermore, operational conditions of the furnace greatly affect the yields of chemicals produced during the combustion process, i.e., fractions of tars, gases and char. Moreover, not only the yields of chemicals differ under different combustion conditions, but also their reactivity in succeeding reactions. As a result of such complexity, optimization of a boiler control strategy requires a detailed process model [3].

The most recent publications, considering modeling the combustion of solid fuel on a grate, concentrate on the combustion of either straw or municipal waste. Shin and Choi [21] have developed a 1-D model of waste incineration to understand better phenomena occurring inside a municipal solid waste (MSW) incinerator. Van der Lans et al. [18] have developed a two-dimensional, homogeneous model for design and operation parameter optimization of a straw combustion process. Goh et al. [19] have developed the model of grate combustion of municipal solid waste in order to study the process, since efficient incinerator design requires extensive knowledge of the combustion process. Later, Yang et al. [20] developed a 2-D model of a MSW incinerator which was then verified using experimental data obtained from a pot reactor. Kær [22] developed a one-dimensional model to describe a fixed bed combustion of straw and using walking grate concept extended it to cover the moving bed of a straw boiler. Similarly, Zhou et al. [23] assumed insignificant horizontal temperature gradients in order to simplify the thermal conversion model on the grate of a straw boiler to one dimension.

This paper describes the developed model of a BioGrate furnace. The purpose for the modeling work was to construct a dynamic model providing an insight into the chemical and physical phenomena occurring inside the process. This paper is organized as follows: Section 2 describes the structure of a BioGrate boiler process, Section 3 presents the model and its aspects, Section 4 discusses the implementation details of the model, Section 5 presents the simulation results, Section 6 summarizes the results of simulations.

2. Process description of a BioGrate boiler

A BioGrate consists of following parts: a water filled ash space below the grate, while the grate itself is located above the reservoir. The BioGrate is covered with a heat insulating brick wall, which reflects the heat radiation back to the grate [3].

The grate consists of several ring zones. These zones are further divided into two types of rings: rotating and fixed. A half of the grate rings are rotating and the rest are fixed. Every second rotating ring rotates clockwise and the others rotate counterclockwise. This structure helps spreading fuel evenly upon the surface of the conical grate [3].

Fuel is fed into the center of the grate from below. In the middle of the cone the fuel dries as a result of heat radiation, which is emitted by the combusting flue gas and reflected back to the grate by the grate walls. The dry fuel then proceeds to the outer shell of the grate, where pyrolysis and char combustion occur. The ash and carbon residues fall off the edge of the grate into the water-filled ash pit [3].

The air required in combustion is fed into the grate from the bottom of the grate (primary air) and from the grate walls (secondary air). In addition, in order to ensure clean combustion, additional air can be fed from the top of the grate (tertiary air). Burning produces heat that is absorbed in several steps. First, the evaporator absorbs the energy in the flue gases. Next, part of the energy of the flue gases is transferred to superheaters. In the third phase, the heat is transferred to the convective evaporator. Finally, economizers remove the remaining flue-gas energy [3].

The operation principle of a power plant is based on the steam generation. As any other bio power plant, a BioGrate power plant comprises several parts, including a boiler, a turbine generator, a feed-water tank, a water treatment plant and a flue gas-cleaning system. Solid fuel is fed into the furnace of the boiler, where it is combusted, generating heat and flue gases. Flue gases contain fly ash

which comprises several harmful components. Therefore, flue gases are purified of fly ash before the release into the atmosphere. Therefore, the flue gases are subjected to several steps of the cleaning procedure and then emitted into the atmosphere. Instead, the heat acquired from the fuel is used for the steam production.

The steam produced in the boiler is led to a generator turbine, which converts its mechanical energy into electricity. As steam performs mechanical work, its pressure decreases, steam with decreased pressure is then used for heating utility streams, such as water [2]. After steam has released enough energy it condenses. The condensed steam is called condensate which along with pretreated feed water is fed into a feed-water tank. Inside the tank, liquid is heated with a bled steam from the turbine. This procedure increases energetic efficiency of the process [1].

3. Dynamic modeling

The current model of a BioGrate uses walking grate concept modified for a BioGrate furnace. In addition, the chemical reaction kinetics were selected, especially, to fit the operational conditions of the BioGrate. Furthermore, an experimental model was used to model the radiation distribution inside the furnace.

Biomass bed reacts in a series of four different chemical reactions: drying, pyrolysis, char gasification and char combustion [5]. Active drying starts when temperature of a particle reaches the boiling point of water. Then, the high temperature of a furnace initiates a pyrolysis reaction. The pyrolysis reaction produces three products: gases, char and tar. Gases are mainly composed of CO, CO_2, H_2, and C_1-C_3 hydrocarbons. Tar contains many organic components, such as levoglucosan, furfural, furan derivatives and phenolic compounds [6]. Next, each reaction will be discussed in details.

3.1. Continuity equation models, their parameters and assumptions

3.1.1. Assumptions

Several assumptions are made in order to simplify the modeling work. The assumptions are listed in descending order of importance:

1. The system is one dimensional, because the length of the grate is significantly longer than the height. Therefore, the temperature gradient in the horizontal direction is insignificant compared that in the vertical direction.
2. Plug-flow gas assumption [9]. The gas phase is assumed to be ideal [9], [10].
3. The solid is assumed to be a porous material [11].
4. Diffusion in the gas phase is neglected, since the effect of convection on transportation of the gas is significantly greater [5].
5. Pressure dynamics are ignored, because the release of gaseous species is negligible compared to the primary air flow and, as a result, pressure evolution can be neglected [9]
6. Heat produced in char combustion is assumed to be retained in the solid phase [9].

7. No volume reduction (shrinkage) occurs during drying, pyrolysis and combustion [9], [7]
8. The temperature of the gas released from solids is the same as that of the solids [9]
9. The temperature of solids in a discretized block is uniform [9]
10. The heat capacity of the wood is assumed to be constant [9]
11. No heat loss

Next, the simplified continuity equations are presented.

3.1.2. Solid Phase continuity equation

Solid phase reacts through drying, pyrolysis and char combustion reactions:

$$\frac{\partial \rho_s}{\partial t} = -R_f \tag{1}$$

where ρ_s is the density of the solid phase, and R_f the overall reaction rate of the solids.

3.1.3. Energy continuity equation of the solid phase

Energy equation for the solid phase considers heat conduction, heat exchange between phases, energy lost in drying and pyrolysis reactions, and energy gained in char combustion:

$$\frac{\partial T_s}{\partial t} C_s \rho_s = \frac{\partial}{\partial x}\left(k_{cond}\frac{\partial T_s}{\partial x}\right) + k_{conv}v_p(T_f - T_s) - ...$$
$$R_{evap}\Delta H_{evap} - R_{pyr}\Delta H_{pyr} + R_{comb,C}\Delta H_{comb,C} - ... \tag{2}$$
$$R_{gasi,CO2}\Delta H_{gasi,CO2} - R_{gasi,H2O}\Delta H_{gasi,H2O}$$

where T_s is the temperature of solid phase, C_s the heat capacity of the solid phase, ρ_s the density of the solid phase, x the vertical coordinate, k_{cond} the heat conduction coefficient of the solid phase, k_{conv} the heat convection coefficient between the gas and solid phases, v_p is the density number, T_f the temperature of the gas phase, and R_{evap} and R_{pyr} the reaction rates of drying, pyrolysis. Reaction rates $R_{comb,C}$, $R_{gasi,CO2}$ and $R_{gasi,H2O}$ correspond to reaction rates of char combustion, gasification with carbon dioxide and gasification with water steam, respectively. ΔH_{evap} and ΔH_{pyr} are the reaction enthalpies of drying, pyrolysis. Reaction enthalpies $\Delta H_{comb,C}$, $\Delta H_{gasi,CO2}$ and $\Delta H_{gasi,H2O}$ correspond to reaction enthalpies of char combustion, gasification with carbon dioxide and gasification with water steam, respectively.

The radiation reflected from the grate walls to the fuel bed is described through boundary conditions. The boundary conditions are defined as follows:

At the surface of the fuel bed, $x = a$

$$k_{cond}\frac{\partial T_s}{\partial x}\bigg|_{x=a} = I_{in} - e\sigma T_s^4 \tag{3}$$

where I_{in} is the energy flux into the system, and $e\sigma T_s^4$ the energy flux out of the system.

To describe the energy flux, I_{in}, an experimental model was used. The model was defined from the experimental data of a BioGrate boiler located in Trolhättan, Sweden.

$$k_{cond} \frac{\partial T_s}{\partial x}\bigg|_{x=0} = e\sigma T_s^4 \qquad (4)$$

where $e\sigma T_s^4$ is the energy flux out of the system.

According to Yagi and Kunii [15], flowing fluid improves the heat conduction of the bed due to effect of axial dispersion; therefore. the overall heat conduction coefficient will become:

$$k_{cond} = k_{cond,0} + \alpha\beta \cdot \text{Re}\,\text{Pr} \qquad (5)$$

where $k_{cond, 0}$ is the heat conduction of the bed with a stagnant fluid, and the right hand term represents the heat conduction due to axial dispersion. Re is the Reynolds number, Pr the Prandatl number, and $\alpha\beta$ is a geometrical constant for cylinders and spheres. The geometrical constant is reported typically to take values of between 0.1 and 0.13 [15]. In the model, the average $\alpha\beta = 0.115$ was used.

Heat conduction coefficient from the study of Yagi and Kunii [15] was used to describe heat conduction in the bed while heat conduction coefficient for wood particles was based on the study [14].

3.1.4. Gas phase continuity equation

Reacted solid components of wood are transferred to gas phase, in addition, gas phase continuity equation considers gas flow:

$$\frac{\partial}{\partial t}(\rho_f \varepsilon_b Y_i) - \frac{\partial}{\partial x}(v_f \rho_f \varepsilon_b Y_i) = R_i \qquad (6)$$

where ρ_f is the density of gas phase, ε_b the bed porosity, Y_i the mass fraction of the gaseous component i, v_f the gas flow velocity and R_i the rate of formation of gaseous component i.

3.1.5. Energy continuity equation of the gas phase

Assuming no heat loss will occur, the energy continuity equation can be denoted as follows:

$$\frac{\partial h_f}{\partial t}\rho_f = -\frac{\partial}{\partial x}\left(\varepsilon_b v_f h_f\right) - k_{conv} v_p (T_f - T_s) + ... $$
$$R_{comb,CO}\Delta H_{comb,CO} + R_{comb,H2}\Delta H_{comb,H2} \qquad (7)$$

where h_f is an enthalpy of the gas phase, ρ_f the density of the gas phase, ε_b the bed porosity, v_f the gas flow velocity, and R_i the rate of formation of gaseous component i, k_{conv} is the heat convection coefficient between the gas and solid phases, v_p is the density number, T_f the temperature of the gas phase and T_s is the temperature of the solid phase.

3.2 Chemical reactions of the model

The thermal decomposition of wood comprises three main chemical reactions: drying, pyrolysis and char gasification with char combustion. In general, the chemical reactions can be depicted using experimental or semi-experimental models. However, since Arrhenius dependence equations

are simple to use, and also accurate; therefore, they have been used in this work.

3.2.1 Moisture evaporation

Usually, fuels used in combustion processes contain moisture. Depending on the type of fuel, a fuel particle can contain various amounts of moisture. According to Thunman et al. [4], fuel particles can contain up to 60 wt% of moisture while char residue being as low as 10 wt% of the wet wood. Water can be bound to the structure of a wood particle or reside in its pores.

Di Blasi et al. [12] presented a simple, yet accurate model to describe drying kinetics in the updraft gasifiers, which use countercurrent combustion conditions:

$$R_{evap} = 5.6 \cdot 10^8 \exp\left(-88[kJ/mol]/\Re/T_s\right)\rho_{Water} \qquad (8)$$

where ρ_{Water} is the density of water, T_s the temperature of the solid phase, and $\Re$ the gas constant.

3.2.2 Pyrolysis

After a particle has dried, the next reaction occurring is pyrolysis. In the pyrolysis reaction, a dry wood particle is decomposed into tar, volatile organic components and char. However, fractions of tar, gas and char in the product yield are strongly dependent on the reaction conditions of a combustion process.

Alves and Figueiredo [13] presented a mathematical model of wet wood pyrolysis. The simulation results of this model were experimentally validated in temperature range of 298 – 780 °C with a wet pine cylinder having the radius of 18.5 mm, with water content being 45-49 wt-%. Experimental and simulated results agreed.

Cellulose is reported to react with the following reaction kinetics [13]:

$$R_{devol, cel} = 2 \cdot 10^9 \cdot \exp(-146[kJ/mol]/\Re/T_s) \cdot \rho_{cel} \qquad (9)$$

where ρ_{cel} is the density of cellulose, T_s the temperature of the solid phase, and $\Re$ the gas constant.

Hemicellulose is reported to react with the following rection kinetics [13]:

$$R_{devol, Hemi} = 7 \cdot 10^4 \cdot \exp(-83[kJ/mol]/\Re/T_s) \cdot \rho_{Hemi} \qquad (10)$$

where ρ_{Hemi} is the density of hemicellulose, T_s the the temperature of the solid phase, and $\Re$ the gas constant.

3.2.3 Combustion of pyrolysis gases

The yield of pyrolytic gases is around 85 wt. % under the operation conditions of a BioGrate boiler, since under these conditions the gasifying pyrolysis is the dominant pyrolysis mode. Therefore, significant amount of energy, used by the boiler, comes from the combustion of gases; this fact poses the combustion of pyrolytic gases as the most important energy source. However, the composition of the gaseous products of pyrolysis reported in the study

of Dupont et al. [25], suggests that carbon monoxide has the highest concentration in the pyrolytic gas, while the fraction of other combustible gases remains under 10 wt. %. Therefore, in order to ensure the acceptable accuracy of the model, while keeping the model simple, only the oxidation of carbon monoxide to carbon dioxide is considered.

In addition to the oxidation of carbon monoxide, also the combustion hydrogen is included in the model because of three facts. First, char is known to react with water steam producing carbon monoxide and hydrogen. Second, significant amount of water steam is released during the drying reaction. Consequently, also the amount of produced hydrogen in char gasification reaction can become significant. Finally, hydrogen reacts rapidly in the presence of oxygen, producing significant amounts of energy.

In the presence of water steam and oxygen, carbon monoxide is known to follow the following kinetics [26]:

$$R_{comb,CO} = 1.3 \cdot 10^{14} \exp(-30[kcal/mole]/\Re/T_g) \cdot ... \\ \rho_{CO} \cdot \rho_{O2}{}^{0.5} \cdot \rho_{H2O}{}^{0.5} \tag{11}$$

where ρ_{CO} is the density of carbon monoxide, ρ_{O2} is the density of oxygen, ρ_{H2O} is the density of water steam, T_g is the temperature of the solid phase, and $\Re$ the gas constant.

Hydrogen is reported to react through the following kinetics [26]:

$$R_{comb,H2O} = 2.96 \cdot 10^{11} \exp(-6900[K]/T_g) \cdot ... \\ \rho_{H2} \cdot \rho_{O2}{}^{0.92} \tag{12}$$

where ρ_{H2} is the density of hydrogen, ρ_{O2} is the density of oxygen, T_g is the temperature of the solid phase, and $\Re$ the gas constant.

3.2.4 Char conversion reactions

Char combustion in the model is accounted for with the model presented in Janse et al. [24]. This model is chosen because it is valid over the temperature range 573-773K, which corresponds to the temperature of char combustion in BioGrate. In addition, the pyrolysis conditions of char particles, which were used to obtain the model parameters, are similar to the pyrolysis conditions inside a BioGrate.

$$R_{comb} = 5.3 \cdot 10^5 \exp(-125[kJ/mole]/(\Re T_s)) \cdot ... \\ P_{O2}^{0.53}(1-X)^{0.49}\rho_C \tag{13}$$

where ρ_c is the density of the char, T_s is the temperature, of the solid phase, $\Re$ is the gas constant, and P_{O2} the pressure of oxygen, and X is the degree of the conversion of char.

Matsumoto et al. [27] have reported, in their study, that the random pore rate equations is the best option to describe the gasification reactions of char with carbon dioxide and water steam. Therefore, the rate equations presented in the study of Matsumoto et al. [27] for char gasification with carbon dioxide and water steam are used in the model.

Gasification reaction of char with carbon dioxide [27]:

$$R_{gasi,CO2} = P_{CO2}^{0.22} \cdot 2.24 \cdot 10^3 \exp(-93.9[kJ/mole]/\Re/T_s) \\ \cdot (1-X)\sqrt{1-10\ln(1-X)} \cdot \rho_C \tag{14}$$

where ρ_c is the density of the char, T_s the temperature, of the solid phase, $\Re$ is the gas constant, and P_{CO2} the pressure of carbon dioxide, and X is the degree of the conversion of char.

Gasification reaction of char with water steam [27]:

$$R_{gasi,H2O} = P_{H2O}^{0.22} \cdot 9.99 \cdot 10^4 \cdot ... \\ \exp(-136[kJ/mole]/\Re/T_s) \cdot ... \\ \cdot (1-X)\sqrt{1-10\ln(1-X)} \cdot \rho_C \tag{15}$$

where ρ_c is the density of the char, T_s the temperature of the solid phase, $\Re$ is the gas constant, and P_{H2O} the pressure of carbon dioxide, and X is the degree of the conversion of char.

4 Implementation of the models and the description of the testing environment

The model was implemented in the MATLAB environment, in which a set of finite difference methods was used to solve the continuity equations. The overall solving algorithm is presented in Figure 1.
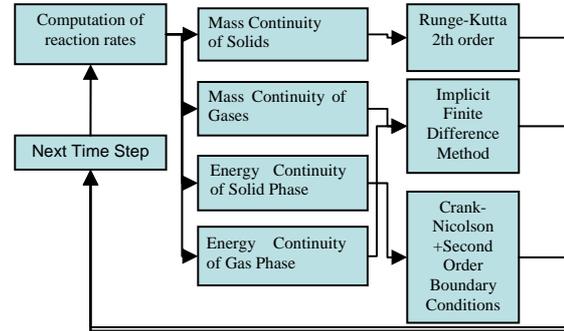
Figure 1. Model solving scheme.

5 Simulation Results

This section presents simulation results obtained with different fuel parameter values.

5.1 Simulation case I, studying the effect of fuel quality on the combustion of wood chips

Wood chips possess several quality properties, including the fuel bed porosity, the bed density and the moisture content. However, the moisture content is the most important property, since it varies significantly between different batches of the fuel. In contrast to the moisture content, the bed porosity and bed density, in case of wood chips, vary only insignificantly; therefore, they are of no interest in the current study. The simulation results with moisture contents of 40, 50 and 60 wt. % are shown in Figure 2.
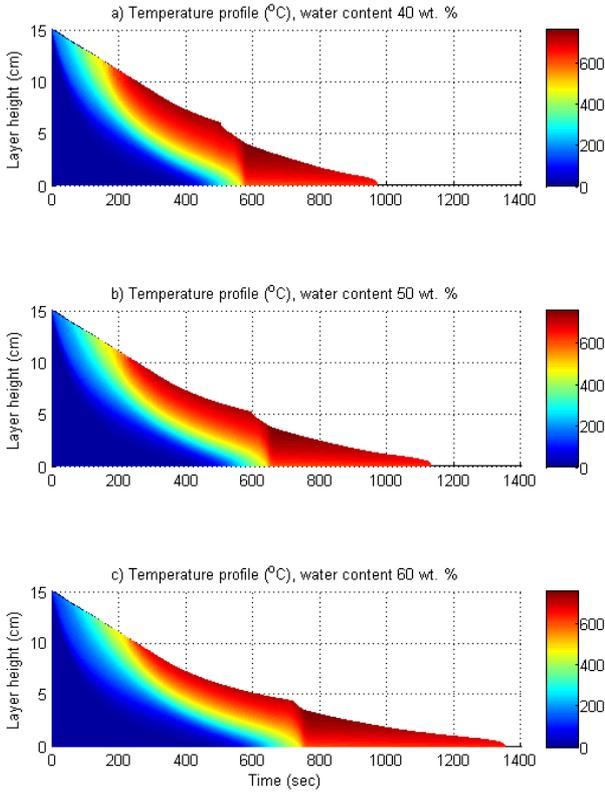
Figure 2. Temperature profiles of the fuel bed with fuel with a) 40 wt.%

b) 50 wt. % and c) 60 wt. % moisture content

The results presented in Figure 2 suggest that, although the temperature of the fuel beds with a different moisture content remains within the same range, there are considerable differences in the combustion process. Not only the combustion time increases with increasing moisture content, but also the reaction front becomes narrower. The increase in combustion time can be explained by the fact that the higher the moisture content the longer time it takes for the water to evaporate. Furthermore, high moisture content prevents the heat flux from penetrating deeper into the fuel bed, because a larger amount of heat is required to dry the same amount of fuel. Similar behavior was also observed in the study of Yang et al. [28]. In the study [28], it was found that burning rate is inversely proportional to the moisture content of a fuel bed.

5.2 Simulation case II, studying the effect of the air flow on the combustion of wood chips

The air flow also has a significant effect on the combustion process, since it provides the oxygen required for chemical reactions, such as char oxidation. Simulations were conducted with the air flows of 0.5, 0.75, 1, 2 and 5 m^3/s, while particle sizes were 20, 35 and 50 mm. Table 1 presents the result obtained from the simulation, while Figure 3 visualizes the results presented in Table 1.

Table 1. Combustion time of fuels with different air flows.

Air flow m^3/s	Combustion time (s)		
	20 mm	35 mm	50 mm
0.5	1621	1590	1558
0.75	1398	1357	1336
1	1379	1320	1286
2	1373	1279	1230
5	1435	1269	1202

Figure 3 shows that, as the volume of the air flow increases from 0.5 to 0.75 m^3/s, the combustion times for all particle sizes decrease significantly from around 1600 s at 0.5 m^3/s, to 1350 – 1400 second at 1 m^3/s.
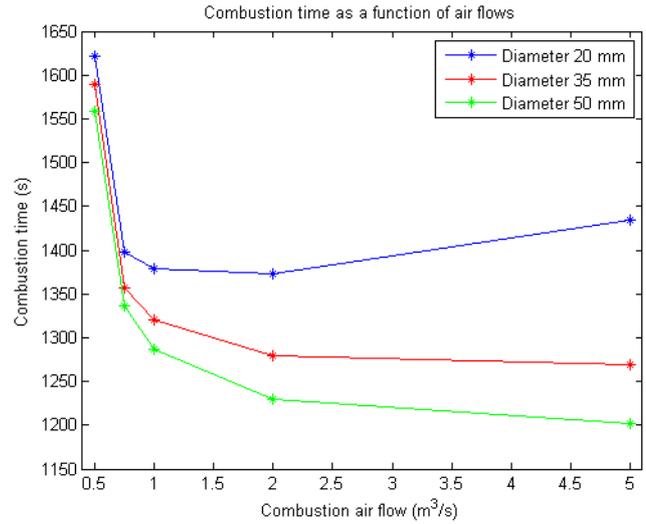
Figure 3. Combustion times as a function of the volumetric air flow.

This phenomenon is the result of oxygen deficiency at low air supplies in the combustion process, i.e., not enough oxygen is supplied to the process to burn the char at its maximum rate.

Figure 3 indicates that, indeed, with low air flows char burns slowly. However, as shown in Figure 3 also high air flows can slow down the combustion of the fuel. This phenomenon can be explained by the cooling property of the air flow. The air supplied to the process has a significantly lower temperature, especially, at the combustion front where the temperature of the solid phase is high. In addition to the significant temperature difference, air is supplied in an opposite direction to the reaction front, thus making heat conduction less efficient and narrowing the reaction front. This finding is also supported by the study of Thunman and Leckner [8]. Figure 4 shows the difference in reaction front thicknesses between airflows of 0.5 and 5 m^3/s.
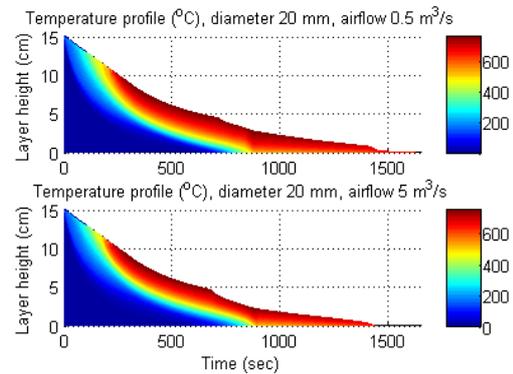
Figure 4. Temperature profiles of beds with 0.5 and 5 m^3/s airflows

Nevertheless, the cooling effect is not significant with large fuel particle diameters, compared to that of 20 mm particles. Figure 3 indicates that in case of 20 mm large particles the combustion time decreases while air flow is increased from 0.5 to 0.75 m^3/s. However, when the air flow is further increased above 2 m^3/s, the combustion

time starts decreasing. In contrast to 20 mm large particles, larger particles seem to be less affected by the cooling of the air flow. Furthermore, the combustion time continues decreasing as the air flow is increased above 2 m^3/s. These findings can be explained by the fact that for smaller particle sizes the density number, the ratio of particle area to unit volume, is larger than that of large particles. In addition, for small particle sizes, also the heat convection coefficient, which is responsible for heat exchange between the gas and solid phases, has a larger value than the coefficient of large particles. Therefore, in case of small particles the heat exchange between phases is more efficient, compared to large particles, thus the cooling property of the air flow affects small particles more than the larger ones. Horttanainen et al. [29] have concluded in their study that combustion air flow rate can be increased as particle size is increased. This finding is in agreement with the result obtained from simulation case II.

6 Conclusions

The investigation on the boiler furnace model was started by studying phenomena occurring in the BioGrate furnace. The furnace model was decomposed in to reaction rates and governing equations for mass and energy conservation.

The developed BioGrate model was than used to study the process phenomena occurring inside the BioGrate furnace with varying process conditions. In addition, a sensitivity analysis was made for different parameters, which affect the combustion process.

The sensitivity analysis showed that the combustion time increased linearly with the increase of moisture content. A study on the air flow effect indicated that oxygen deficiency slowed down the combustion process, however, excess air, on the other hand, increased the combustion time by cooling the solid phase. The results obtained from the simulator were found to be in agreement with the results found in literature.

7 References

1 Kiameh, P., Power Generation Handbook, McGraw-Hill, 2002, 557 p.

2 Joronen, T., Kovács, J., Majanne, Y., *Voimalaitosautomaatio*, Suomenautomaatioseura Oy, Helsinki 2007, 1st edition, 276 p.

3 WWW, Anon, Wärtsilä Oyj, Brochure,

http://service.wartsila.com/Wartsila/global/docs/en/power/media_publications/brochures/bioenergy_fi.pdf, 03.08.2009

4 Thunman, H., Davidsson, K., Leckner, B., *Separation of drying and devolatilization during conversion of solid fuels*, Combustion and Flame **137** (2004), pp. 242 - 250

5 Peters, B., Bruch, C., *A flexible and stable numerical method for simulating the thermal decomposition of wood particles*, Chemosphere **42** (2001), pp. 481 – 490

6 Di Blasi, C., Heat, *Momentum and Mass Transport Through a Shrinking Biomass Particle Exposed to Thermal Radiation*, Chemical Engineering Science **5** (1996), pp. 1121-1132

7 Di Blasi, C., *Combustion and gasification rates of lignocellulosic chars*, Progress in Energy and Combustion Science **35** (2009), pp. 121 – 140

8 Thunman, H., Leckner, B., Ignition and propagation of a reaction front in cross-current bed combustion of wet biofuels, Fuel **80** (2001), pp. 473-481

9 Zhou H, Jensen, A. D., Glaborg, P., Jensen, P., A., Kavaliauskas, A., *Numerical modeling of straw combustion in a fixed bed*, Fuel **84** (2005), pp. 389-403

10 Kær S., K., *Straw combustion on slow moving grates-a comparison of model predictions with experimental data*, Biomass and Bioenergy **28** (2005), pp. 307-320

11 Yang, Y., B., Yamauchi, H., Nasserzadeh, V., Swithenbank, J., *Effects of fuel devolatilization on the combustion of wood chips and incineration of simulation municipal solid wastes in a packed bed*, Fuel **82** (2003), pp. 2205-2221

12 Di Blasi, C., Branca, C., Sparano, La Mantia, B., *Drying Characteristics of wood cylinders for conditions pertinent to fixed-bed countercurrent gasification*, Biomass and Bioenergy **25** (2003), pp. 45-58

13 Alves, S., S., Figueiredo, J., L., *A model for pyrolysis of wet wood*, Chemical Engineering Science **44** (1989), pp. 2861-2869

14 Janssens, M., Douglas, B., *Wood and Wood Products*, Handbook of Building Materials for Fire Protection, Edited by Harper, C., A., McGraw-Hill 2004, 542 p.

15 Yagi, S., Kunii, D., *Studies on Effective Thermal Conductivities in Packed Beds*, A.I.Ch.E. Journal **3** (1957), pp. 373-381

16 Fjellerup, F., Henriksen, U., *Heat Transfer in a Fixed Bed of Straw Char*, Energy & Fuels **17** (2003), pp. 1251-1258

17 Horttanainen M., Saastamoinen, J., Sarkomaa, P., *Operational Limits of Ignition Front Propagation against Airflow in Packed Beds of Different Wood Fuels*, Energy & Fuels **16** (2002), pp. 676-686

18 R.P. van der Lans, L. T. Pedersen, A. Jensen, P. Glarborg and K. Dam-Johansen, *Modelling and expirements of straw combustion in a grate furnace*, Biomass and Bioenergy **19** (2000) , pp. 199-208

19 Y. R. Goh, Y. B. Yang, R. Zakaria, R. G. Siddall, V. Nasserzadeh, J. Swithenbank, *Development of an Incinerator Bed Model for Municipal Solid Waste Incineration*, Combustion Science and Technology **162** (2001), pp. 37-58

20 Yang, Y., B., Goh,, Y.,R., Zakaria, R., Nasserzadeh, V., Swithenbank, J*., Mathematical modelling of MSW incineration on a traveling bed*, Waste Management 22 (2002), pp. 369-380

21 Shin, D, Choi, S, *The Combustion of Simulated Waste Particles in a Fixed Bed*, Combustion and Flame **121** (2000), pp. 167-180

22 Kær S., K., *Straw combustion on slow moving grates-a comparison of model predictions with experimental data*, Biomass and Bioenergy **28** (2005), pp. 307-320

23 Zhou H, Jensen, A. D., Glaborg, P., Jensen, P., A., Kavaliauskas, A., *Numerical modeling of straw combustion in a fixed bed*, Fuel **84** (2005), pp. 389-403

24. Janse, A., M., C., de Jonge, H., G., Prins, W., van Swaaij, W., P., M., Combustion Kinetics of Char Obtained by Flash Pyrolysis of Pine Wood, Ind. Eng. Chem. Res. 37 (1998), pp. 3909-3918

25. Dupont, C., Chen, Li., Cances, J., Commandre, J.-M., Cuoci, A., Pierucci, S., Ranzi, E., *Biomass pyrolysis: Kinetic modeling and expiremental validation under high temperature and flash heating rate conditions*, Journal of Analytical and Applied Pyrolysis **85** (2009), pp. 260-267

26. Babushik, V., I., Dakdancha, A., N., *Global kinetic parameters for high-temperature gas-phase reactions*, Combustion, Explosion, and Shock Waves **29** (1993), pp. 464-489

27. Matsumoto, K., Takeno, K., Ichinose, T., Ogi, T., Nakanishi, M., *Gasification reaction kinetics on biomass char obtained as a by-product of gasification in an entrained-flow gasifier with steam and oxygen at 900 – 1000 °C*, Fuel **88** (2009), pp. 519 – 527

28 Yang, Y., B., Sharifi, V., N., Swithenbank J., *Effect of air flow rate and fuel moisture on the burning behavioursof biomass and simulated municipal solid wastes in packed beds*, Fuel **83** (2004), pp. 1553-1562

29. Horttanainen M., Saastamoinen, J., Sarkomaa, P., *Operational Limits of Ignition Front Propagation against Airflow in Packed Beds of Different Wood Fuels*, Energy & Fuels **16** (2002), pp. 676-686

Data-Based Uncertainty Modeling of MIMO Systems

Hamed Jafarian and Kurt-Erik Häggblom

Process Control Laboratory, Department of Chemical Engineering,
Åbo Akademi University, Åbo, Finland

Abstract

Many robust control design methods require a linear model consisting of nominal model augmented by an uncertainty description. A general form for such a model is

$$G = G_0 + H_{21}\Delta(I - H_{11}\Delta)^{-1}H_{12} \qquad (1)$$

where G is the transfer function of the true system, G_0 is a nominal model, and Δ is a perturbation causing uncertainty about the true system. Depending on the particular type of uncertainty (additive, input or output multiplicative, inverse types of uncertainty, combinations of various types of uncertainty), H_{11}, H_{12} and H_{21} can contain combinations of (known) constant matrices and the nominal model.

Assume that we have information about the true system in the form of a number of possible transfer functions G_k, $k = 1,\ldots,N$. The nominal model and the perturbation Δ_k associated with G_k are unknown, but they have to satisfy

$$G_k = G_0 + H_{21}\Delta_k(I - H_{11}\Delta_k)^{-1}H_{12}, \quad k = 1,\ldots,N. \qquad (2)$$

How should G_0 be determined? It has been shown that $\left\|\Delta\right\|_\infty$ is a control relevant measure of the distance between G and G_0 for models of the form (1) and that the achievable stability margin by feedback control is inversely proportional to this distance. For a given type of uncertainty model, this suggests that G_0 should be determined by solving the optimization problem

$$\min_{G_0} \max_k \left\|\Delta_k\right\|_\infty \qquad (3)$$

subject to the appropriate data matching condition (2). Obviously, the type of uncertainty model giving the smallest minimum is the best one according to this measure.

If information about the system is obtain through identification, input-output data are available. An attractive way of removing noise from the output is to fit a model G_k to the data and to calculate a noise-free output y_k by $y_k = G_k u_k$, where u_k is the input in experiment k. Since the purpose of the experiments in this context is to excite the system in various ways, the inputs do not tend to be persistently exciting in all individual experiments. Thus, G_k only applies to the particular input u_k, and the relevant information is input-output data $\left\{u_k, y_k\right\}$, $k = 1,\ldots,N$. This means that the model matching condition (2) should be replaced by the input-output matching condition

$$y_k = G_0 u_k + H_{21}\Delta_k(I - H_{11}\Delta_k)^{-1}H_{12}u_k. \qquad (4)$$

It can be shown that the use of (4) instead of (2) results in a less conservative uncertainty model.

Our modeling approach is to model G_0 in the frequency domain using sampled frequency responses of the input-output data. Because of the availability of G_k, these are easy to calculate for standardized inputs. The available information is thus $\{u_k(\mathrm{j}\omega), y_k(\mathrm{j}\omega), \omega \in \Omega\}$, $k = 1, \ldots, N$, and we solve the optimization problem frequency-by-frequency, i.e.

$$\min_{G_0(\mathrm{j}\omega)} \max_k \left\| \Delta_k(\mathrm{j}\omega) \right\|_2 \quad \text{s.t. (4),} \quad \forall \omega \in \Omega. \tag{5}$$

The uncertainty Δ_k is assumed to be unstructured.

For some types of uncertainty, the optimization problem can easily be formulated as a convex optimization problem. Additive uncertainty, for example, which in its basic form is described by

$$y_k = G_0 u_k + \Delta_k u_k, \quad k = 1, \ldots, N \tag{6}$$

results in the optimization problem

$$\min_{G_0} \gamma \quad \text{s.t.} \quad \begin{bmatrix} \gamma I & y_k - G_0 u_k \\ (y_k - G_0 u_k)^* & u_k^* u_k \end{bmatrix} \succeq 0, \quad k = 1, \ldots, N, \quad \forall \omega \in \Omega \tag{7}$$

which is a convex optimization problem. Here A^* denotes the complex conjugate transpose of A and $P \succeq 0$ denotes that P is positive semidefinite.

Many types of uncertainty descriptions do not readily give a convex optimization problem. For example, an output multiplicative uncertainty described by

$$y_k = G_0 u_k + \Delta_k G_0 u_k, \quad k = 1, \ldots, N \tag{8}$$

results in the optimization problem

$$\min_{G_0} \gamma \quad \text{s.t.} \quad \begin{bmatrix} \gamma I & y_k - G_0 u_k \\ (y_k - G_0 u_k)^* & u_k^* G_0^* G_0 u_k \end{bmatrix} \succeq 0, \quad k = 1, \ldots, N, \quad \forall \omega \in \Omega \tag{9}$$

which is non-convex due to the appearance of $G_0^* G_0$. An iterative solution by keeping $G_0^* G_0$ fixed during each iteration tends to produce local minima which are non-global. However, we show how this optimization problem, and similar ones for some other types of uncertainty, can be reformulated as a convex optimization problem.

For control design, G_0 is needed as a transfer function or a state-space model. In principle, we should determine such a model by replacing G_0 in the appropriate consistency relations, like those appearing in (7) and (9), by a suitable parameterization of G_0. However, so far we have not been able to obtain a satisfactory solution in that way. Instead, we have fitted a model to the calculated frequency responses $G_0(\mathrm{j}\omega)$, $\omega \in \Omega$. A drawback of this approach is that $\min \left\| \Delta(\mathrm{j}\omega) \right\|_2$ will increase, usually also $\min \left\| \Delta \right\|_\infty$, sometimes even drastically. We have studied various approaches of reducing the effects of this drawback.

Modeling for Control of Beer Quality

D.R. Warnasooriya[1], P.G. Rathnasiri[1], Bernt Lie[2]

[1]*Department of Chemical and Process Engineering, University of Moratuwa, Kattubedda, Moratuwa, Sri Lanka.*

E-Mail; dilantha1981@gmail.com, ratnasiri@cheng.mrt.ac.lk

[2]*Telemark University College, P.O. Box 203/Postboks 203, N-3901 Porsgrunn, Norway.*

E-Mail; Bernt.Lie@hit.no

Abstract

Beer is the most common alcoholic drink around the world. When talking about the beer quality, flavor of the beer is more concerned. Most of the brewers in Sri Lanka are using traditional methods to brewing beer. Most of brewers using pre identified recipe to produce mass production of beer. Therefore, beer quality i.e. flavor is varying brand by brand. It is important to study the variation of temperature how will effect to the final alcohol production and the flavor compound formation. Beer manufacturing industry can be used this knowledge to increase the production efficiency and the product quality.

It is very important to know about dynamics of forming flavor compounds. In this work the fermentation process is concerned since all the flavor compounds are formed during the fermentation. The mechanistic model is developed by based on the knowledge of biochemical processes in the yeast cell and previously developed mathematical models which are available in the literature. There are many beer models can be found such as Engasser et al.,(1981);Growth kinetic model, Gee at al.,(1988), Phisalaphong et al.(2006);Growth kinetic model and effect on temperature, W.Fred Ramirez and Jan Maciejowski.,(2007);Optimal beer fermentation, etc. The beer fermentation process is modeled and simulated in MATLAB environment.

Growth model, nutrient model, and the flavor model are considered and developed. Growth model consists of sugar consumption, biomass growth and ethanol formation models. Those models are developed with temperature dependant parameters to observe the effect of temperature. Three amino acids which are Valine, Leucine and Iso Leucine are considered for the Nutrient model. Consumption of these three amino acids is considered during fermentation. Flavor model is developed based on the growth model and the nutrient model. Flavor compounds are categorized into three groups which are Fusel alcohols, esters and vicinal diketones. Altogether nine parameters are considered as flavor compounds and the effects of temperature on those are simulated with MATLAB. Industrial temperature profile is obtained and applied for the developed model and simulated in MATLAB and the results are analyzed.

PI controller is applied to get identified temperature profile to obtain optimal flavor formation and the dynamic model is used for find suitable controller parameters for best control.

A two phase MPC and its application to a grinding process

Alexey Zakharov*, Alexandre Boriouchkine*, Sirkka-Liisa Jämsä-Jounela*

* Aalto University, P.O. box 16100, 00076, Aalto, Finland. e-mail: alexey.zakharov@ tkk.fi

Abstract: The growing complexity of the control systems and the increased use of nonlinear models cause a dramatic increase in the computational requirements of MPCs. Therefore, more computationally efficient MPC are needed. This paper presents a two-phase MPC approach for decreasing computational demand without sacrificing its efficiency. The first phase of the MPC treats the input variables as independent decision variables of the objective optimization, since the largest part of the objective value arises from a few earliest sampling intervals. In contrast, the second phase combines input variables, defining the rest of the MPC objective value, in an open-loop control which is specified by a few independent decision variables. The method is compared against the traditional Quadratic Programming implementation of an MPC for the Grinding Plant control problem. The two-phase MPC demonstrates a better performance compared with the traditional controller with the same control horizon.

Keywords: MPC, dynamic optimization, grinding process, industrial application, stability

1. INTRODUCTION

Current economic conditions has set new challenges for productivity and keeping the operational conditions of processes within required boundaries. These challenges have resulted in the growing complexity of industrial control systems. Consequently, the increasing complexity of control systems requires more efficient advanced level (higher-level) control strategy which almost universally utilizes model predictive controllers (MPCs). However, the use of MPCs can be computationally heavy because of several factors, such as complex process models, as well as a greater number of manipulated and controlled variables, and constraint management. Nevertheless, a limited computational power forces to make a compromise between efficiency and computational time of a control strategy. However, in several applications the control strategy efficiency is critical. To shorten the computational time without losing efficiency, a modification must be done to existing MPC techniques. Several options exist for improving the efficiency of an MPC. One of such techniques is the dynamic programming, since it provides many useful insights into the MPC performance problems. However, dynamic programming reduces the problem to Hamilton-Jacobi-Bellman (HJB) equation that in most cases can not be solved analytically, because it is a partial differential equation, and solving HJB equation is computationally unattractive due to the high dimensionality. For this reason, Dynamic Programming is not applied directly to practical control problems, but instead, different simplified methods are used.

In the 60s, the development of Dynamic Programming-employing control techniques led to the invention of the linear quadratic regulator (LQR). The LQR is a rare case when the HJB equation can be solved analytically. The LQR was found to be useful for many practical applications and it might be considered as the direct predecessor of modern MPCs (so called 'zero generation' of MPC).

Unfortunately, the linear-quadratic regulator is limited by linear dynamics, quadratic objective function and absence of constraints, thus leaving many industrial problems out of its scope. Moreover, it is well acknowledged that economic operating points in typical process units often lie at the intersection of constraints. As a result, a successful industrial controller must keep the system as close to the constraints as possible without actually violating them. Thus, the next generation of MPC appeared having the following main features: linear process constraints, a linear process model, a quadratic objective and a finite time horizon (see Richalet et al. (1978), Prett et al. (1982)). The finite horizon was used to approximate the infinite horizon problem, which hardly can be solved. Since in the presence of constraints even the solution of the finite horizon optimization problem cannot be derived analytically, the quadratic programming was employed to perform the optimization.

In the early 90s, it was discovered that the constrained optimization can cause feasibility problems, especially, when large disturbances appear. Therefore, the most of the modern MPC software products have been enforced to use soft output constraints (Qin and Badgwell 2003).

On the other hand, because of the finite horizon formulation, MPC faced stability problems. Attempts to achieve stability included different prediction and control horizon approaches and the introduction of a terminal cost to the MPC objective. These methods were criticized in the study of Bitmead et al. (1990) as 'playing games', because there were no clear conditions to guarantee stability. Thereby, the stability of MPC was studied actively during the early 90's (Keerthi and Gilbert (1988), Mayne and Michalska (1990) are among the first papers exploring this question) and a comprehensive review of these studies is provided in Mayne et al. (2000).

Briefly, the stability is almost universally established through the use of the value function of MPC as a natural Lyapunov function. On the other hand, the Dynamic Programming provided some useful insights concerning the MPC stability. One example of that is the 'inverse optimality principle', which is used to ensure the stability of MPC by utilizing the fake HJB equation (for details, see for example Bitmead et al 1990, Magni and Sepulchre 1997).

It is well known that the performance of MPC depends on the quality of underlying model: an MPC is as good as its model. For that reason during the last decade the focus was moved to the nonlinear MPC utilizing of a more accurate nonlinear process model. Basically, the implementation of such MPC cannot be based on QP anymore. Therefore, the convex optimization techniques are employed instead of QP.

Consequently, the nonlinearity of the models along with other factors, such as the complexity of control systems, increases the computational requirements for MPCs. However, the computational requirements of MPCs are critical for many applications, especially, for large and fast processes. Therefore, many researchers have concentrated their efforts on reducing the number of on-line computations (Bemporad et al. 2002, Pannocchia et al. 2007, Rao et al. 1998).

In contrast to computational requirements and stability, another important property of an MPC, namely, optimality did not attract so much attention in the literature, even though the finite-horizon MPCs do not provide the optimal solution of problems. In general, the researchers do not focus on the exploration of optimality because of the idea that a close-to-optimal solution may be found through increasing the control horizon. However, an example is given in Di Palma, Magni (2007), where MPC performance is not a non-decreasing function of the optimization horizon. In addition, a longer control horizon also requires more computations and a compromise must be made between the close-to-optimal properties of the controller and its computational demands.

Simultaneously with traditional MPC development, some attempts were made to estimate the solution of HJB equation indirectly. For example, an iterative approach was proposed in Sardis and Lee (1979). Unfortunately, until today, highly efficient methods based on HJB equation have not been developed and the conclusion was made in Cannon (2004) that 'compared with conventional NMPC the computational burden of currently available methods for the HJB successive approximation approach remains prohibitive'.

Even HJB equations are unattractive for numerical implementation, the Dynamic Programming appears still to provide a useful insight into the MPC optimality. One example is presented in the work of Grune and Pannek (2009), where HJB equation was employed to estimate the 'degree of suboptimality' of MPC solutions, which was further used for adaptive determining of the MPC horizon.

Another idea risen from Dynamical Programming, (which present research is focused on) is the desire to have the terminal cost of MPC as close to the value function of the

infinite horizon problem as possible (Mayne et al. 2000). If the value function is employed as a terminal cost, MPC provides the optimal solution even with the time horizon being unity. In particular, the solution of Riccati equation, which is the value function of the unconstrained infinite horizon problem, is proposed as the terminal cost for MPC objective in many papers (see Chmielewski and Manousiouthakis (1996), Sznaier and Damborg (1987)). Although the stability is attained within the approach, in general, it is not possible to expect that a 'good' approximation of the value function can be found. In particular, if the MPC setpoint lies on the border of constraints, quadratic functions cannot capture the essential asymmetry of the value function.

In the present paper, the emphasis is moved on the estimation of the value function of the infinite horizon problem, which provides close-to-optimal behaviour of the controller even with a short control horizon. Thus, the method achieves a decrease in computational demand without sacrificing its efficiency. An industrial application (a model of a Grinding process) is used to test the developed method.

The paper is organized as follows. Section 2 contains a description of the proposed MPC controller, and Section 3 a description of the grinding process. In Section 4, the results are presented and compared against a QP implementation of MPC, and Section 5 contains the conclusion.

2. DESCRIPTION OF THE TWO-PHASE MPC

2.1 The idea of the two-phase MPC

In this section, the two-phase method will be presented for a simple linear discrete state space dynamics:

$$x(k+1) = Ax(k) + Bu(k)$$
$$y(k+1) = Cx(k+1) \tag{1}$$

where $x = (x_1, x_2, ..., x_n)$ is the vector of the current state of the system, $y = (y_1, y_2, ..., y_m)$ is the vector of the system outputs, and $u = (u_1, u_2, ..., u_l)$ is the vector of the input variables. For the sake of simplicity, it is assumed that there is no noise in neither, the dynamics nor in the measurements, and the state of the system is exactly known. In addition, the process is assumed to have M linear constraints:

$$P_i y(k) \leq q_i, i = 1, 2, ..., M . \tag{2}$$

Under dynamics (1) and constraints (2), the optimal setpoint y^* is usually defined by a higher level of the control hierarchy (for example at the real time optimization layer). At this setpoint the steady state of the system x^* and the optimal steady state control u^* are defined using dynamics (1).

A typical objective function of a MPC with the control horizon equals N has the following form:

$$J_N(x(0), u) = \sum_{k=0}^{N-1} l(y(k), \Delta u(k)) + F(x(N), u(N-1)) \tag{3}$$

where different forms of $l(y, \Delta u)$ may be used in different controllers and the terminal cost $F(x, u)$ is needed to stabilize the controller. In fact the widely used approach, which spreads the control action at control horizons N until the prediction horizon K, employs the terminal cost of the following form:

$$F(x(N), u(N-1)) = \sum_{k=N}^{K-1} l(y(k), 0) \qquad (4)$$

Another popular option is the solution of Riccati equation is used as the terminal cost.

On the other hand, dynamic programming theory provides the ideal candidate for the role of the terminal cost, which is able to guarantee both stability and optimality of the solution even with the control horizon equals one. Indeed, the finite horizon formulation is just a simplification of the original infinite horizon optimization problem with dynamics (1), constraints (2) and the following objective:

$$\min J_\infty(x(0), u) = \sum_{k=0}^{\infty} l(y(k), \Delta u(k)) . \qquad (5)$$

According to dynamic programming, the optimal control action may be found as

$$\begin{aligned} u^*(x(0), u(-1)) = \\ = \underset{u(0)}{\arg\min}\big(l(y(0), \Delta u(0)) + V(x(1), u(0))\big) \end{aligned} \qquad (6)$$

where $V(x(1), u(0))$ is the value function introduced as the optimal value of the infinite horizon optimization problem:

$$V(x(1), u(0)) = \min_u \sum_{k=1}^{\infty} l(y(k), \Delta u(k)) . \qquad (7)$$

In fact, MPC and Dynamic programming derive current control through minimizing items (3) and (6) respectively, and the value function plays the same role in Equation (6) as the terminal cost plays in Equation (3). Thus, MPC approach may be considered as an implementation of dynamic programming ideas, but with the 'inaccurate' approximation of the value function. Thus a control horizon longer than one must be employed to obtain the satisfactory performance of MPC even though the terminal cost inaccurately estimates the value function of the infinite horizon problem.

If the open-loop control u^*, which is optimal in state x, was known, function $V(x(N), u(N-1))$ could be easily estimated to any reasonable accuracy. However, the optimal control is unknown and a set U of second phase open-loop controls must be used to get a relatively good estimation of the value function as follows:

$$V_K^U(x(N), u(N-1)) = \min_{u \in U} \sum_{k=N}^{K-1} l(x(k), \Delta u(k)) , \qquad (8)$$

here K is the second phase horizon. In particular, the common MPC with different prediction and control horizons uses set U consisting of a single open-loop control which expands the values of the input variables at the control horizon until the end of the prediction horizon.

A set of second phase open-loop controls U containing the optimal control $u^*(x(0), u(-1))$ allows to get the accurate result:

$$V(x(N), u(N-1)) = \lim_{K \to \infty} V_K^U(x(N), u(N-1)) . \qquad (9)$$

Therefore the set of control strategies U must be 'divorce' in a sense that at any point $x(N), u(N-1)$ a close-to-optimal open-loop control can be found in U. On the other hand, the computational complexity of the MPC controller grows for wider sets. Thus, if the set of second phase open-loop controls is chosen as a parametric family of functions, the number of parameters must not be very high in order to avoid the high computational complexity of the method.

Since for any set U consisting of more than one open-loop control the computation of the terminal cost (8) involves an optimization of the system dynamics after the control horizon, this optimization is called 'the second phase' of the proposed MPC.

2.2 Two sets of second phase open-loop controls

Let us consider the following one-parametric set of functions presented in Figure 1:

$$g_\alpha(k) = \alpha \exp(-c_1 k) + (1-\alpha)\exp(-c_2 k) , \qquad (10)$$

where coefficients c_1 and c_2 are fixed. In order to define a open-loop control for the whole MPC, it is needed to define an individual control for each input variable. This can be done in the following way:

$$u_i(N+k-1) = \big(u_i(N-1) - u_i^*\big)g_{\alpha(i)}(k) + u_i^* , \; i = 1,...,l , \quad (11)$$

here $u*$ is the steady state optimal control. Thus the system open-loop control is defined by vector $\alpha = (\alpha_1, \alpha_2, ..., \alpha_l)$, where every element defines the control for the respective manipulated variable.

The proper choice of constants c_1 and c_2 may be a problem if the described above set of open-loop controls is used. In order to avoid it, another set of second phase open-loop controls is introduced by adding a time scale parameter β as follows:

$$g_{\alpha,\beta}(t) = \alpha \exp(-c_1 \beta t) + (1-\alpha)\exp(-c_2 \beta t) . \qquad (12)$$

The open-loop controls of the MPC are constructed similarly to Equation (11):

$$u_i(N+k-1) = \big(u_i(N-1) - u_i^*\big)g_{\alpha(i),\beta}(k) + u_i^* , \; i = 1,...,l \quad (13)$$

Here different parameters α_i are used for different input variables of the MPC, but a single parameter β is used to define the time scale for all input variables. In the present paper, the set of second phase open-loop controls defined by Equation (13) is used for two-phase MPC implementation and testing.
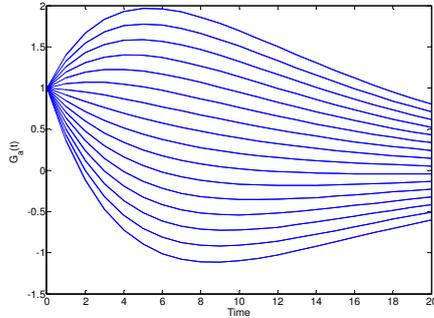

Figure 1. One-parametric set of functions $g_\alpha(k)$ with $c_1 = 0.2$ and $c_2 = 0.1$.

3. DESCRIPTION OF THE GRINDING PLANT

Communition is a huge consumer of electrical power because crushing rocks into powder requires a lot of energy. According to Pomerleau et al. (2000), grinding typically accounts for almost 50% of the costs of a concentrator and, as a result, the optimization of grinding mills is an extremely important research topic. The aim of economic optimization is to maximize the feed rate or to achieve the desired particle size distribution, thus making production more profitable. An overview of the control methods of grinding plants is provided in Hodouin, Jämsä-Jounela et al. (2001).
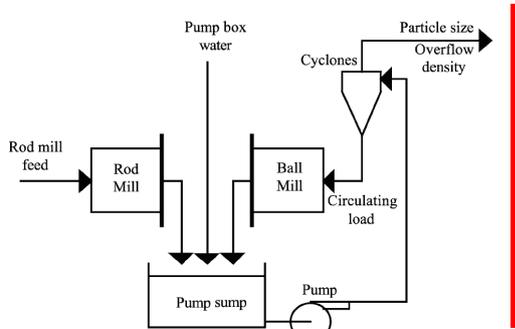
Figure 2. Grinding circuit, Lestage et al. 2002.

The ore is fed to the rod mill and then discharged into the pump sump. The slurry is then fed to a hydrocyclone, where it is separated into the overflow product and a recycled part, which is fed back to the ball-mill (for more details, see Lestage et al. 2002). The whole circuit is presented in Figure 2.

There are two manipulated variables available in the model:

- u_1: the rod-mill feed (t/h)
- u_2: the pump sump water addition (m³/h)

and four output variables:

- y_1: the hydrocyclone overflow density (% solids)
- y_2: the fraction of particles smaller than 325 mesh (47Am) in the product (%)
- y_3: the tonnage through the ball mill (t/h)
- y_4: the pump sump level (%)

Typically the grinding process can be described by means of a transfer function of second order. The transfer function presented in Lestage et al. 2002 is used to test the MPCs:

$$
\begin{array}{ccc}
 & u1 & u2 \\
y1, & G_{11}(s) = \frac{0.0255(1-5600s)e^{-600s}}{(1+5300s)(1+750s)} & G_{12}(s) = \frac{-0.14(1+4050s)}{(1+3200s)(1+60s)} \\
y2, & G_{21}(s) = \frac{-0.2(1-900s)}{(1+5200s)(1+750s)} & G_{22}(s) = \frac{0,012(1+39500s)}{(1+4400s)(1+50s)} \\
y3, & G_{31}(s) = \frac{13.8}{(1+5700s)(1+400s)} & G_{32}(s) = \frac{4.2(1-700s)}{(1+5000s)(1+5s)} \\
y4, & G_{41}(s) = \frac{5,749}{(1+5500s)(1+210s)} & G_{42}(s) = \frac{1,962}{(1+4700s)}
\end{array}
$$

The model is converted into the discrete time state space form (1) with a sample time equals 300s. The system is described with 15 state variables and four controlled variables.

The constraints are defined as product specifications: the hydrocyclone overflow density (y_1) must be above 48% to meet the flotation requirements, and must be below 52% to avoid sedimentation problems. Product specification fineness (y_2) is defined as 47% of the particles smaller than 47 μm. In order not to overload the ball mill, the throughput (y_3) must not exceed 820t/h. The pump sump level (y_4) must remain between 15% and 85%. The constraints are defined in the following order: lower and upper constraints on y_1, lower and upper constraints on y_2, upper constraint on y_3, and lower and upper constraints on y_4. Thus, matrices P and q take the following form:

$$
P = \begin{array}{cccc}
y_1 & y_2 & y_3 & y_4 \\
\end{array}
\begin{bmatrix}
1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & -1
\end{bmatrix},
$$

$q = [52\%, -48\%, 47\%, -47\%, 820\text{t/h}, 85\%, -15\%]^T$.

4. COMPARISON OF THE TWO-PHASE MPC AGAINST THE QP IMPLEMENTATION

Next, the two-phase MPC method is compared against the soft-constrained QP implementation of MPC.

In order to achieve smooth trajectories, the following objective function is used for both controllers:

$$F(x) = \sum_{k=0}^{K}(u(k)-u(k-1))'R(u(k)-u(k-1)) +$$

$$\sum_{k=1}^{K}\sum_{i=1}^{M}Q_i(\max(P_i y(k)-q_i,0))^2 + \qquad , \qquad (14)$$

$$+ \sum_{k=1}^{K}(y(k)-y*)'S(y(k)-y*)$$

where diagonal matrixes R, Q and S are defined as follows:

$$R = \begin{pmatrix} 25 & 0 \\ 0 & 10 \end{pmatrix},$$

$$Q_1 = Q_2 = 100, Q_3 = Q_4 = 25, Q_5 = Q_6 = Q_7 = 1. \qquad (15)$$
$$S_{11} = 5, S_{22} = 0, S_{33} = 0{,}01, S_{44} = 0{,}05$$

To test the controller's ability to follow changing operating conditions, the constraints are varied as shown in Table 1. In every case, the optimal steady states, presented in Table 2, are obtained by maximizing the throughput of the plant.

Table 1. Constraints

Time period (h)	Overflow solids (%)		Particles < than 47μm (%)		Ball mill through t/h	Pump sump level (%)	
	min	max	min	max	max	min	max
0…0.25	48	52	47	47	820	15	85
0.5…5	48	52	49	49	820	15	85
5…10	50	52	49	49	820	15	85
10…15	50	52	47	47	820	15	85

Table 2. Setpoints

Time period (h)	Overflow solids (%)	Particles < than 47μm (%)	Ball mill throughput t/h	Pump sump level (%)
0…0.25	51,83	47,0	820,0	62.06
0.25…5	48,0	49,0	811,95	64.5
5…10	50,0	49,0	739,32	30,87
10…15	51,83	47,0	820	62,06

The prediction horizon K in QP formulation of MPC is always 10 steps longer than the control horizon N. For the two-phase MPC the second phase length is also taken to be 10 steps. QP MPC is tested with control horizons equal to 2 and 10 and two-phase MPC is tested with control horizon equal 2.

The second-phase open-loop control set described by Equation (13) is used for the two-phase MPC implementation. Since there are only two input variables, a second-phase control is defined by three parameters: α_1 and α_2 select the shape of the open-loop controls of the first and the second input respectively, while a parameter β is used for time scaling. Discrete values of α_1 and α_2 are considered with

a step equal 0.4 while β is taken in the following form: $\beta = 0.9^k$ with k having only integer values.

Two controlled variables are compared in Figure 3 and the manipulated variables are presented in Figure 4. In general it is clear that the two-phase MPC with control horizon $N = 2$ demonstrates behaviour, which is closer to QP MPC with longer control horizon $N = 10$ rather than the same control horizon $N = 2$.

The efficiency of the presented methods may be evaluated using the objective function values of every method presented in Table 4. Again, two-phase MPC achieves the same value of the objective function as the QP MPC with the longer prediction horizon $N = 10$ does. QP MPC with the short control horizon $N = 2$ demonstrates 13% bigger objective value.

Table 4. Objective for QP MPC and two-phase MPC

	QP MPC, N = 2	QP MPC, N = 10	MPC 2 phase N = 2
Constraint 1	971	978	993
Constraint 2	1 845	1 358	1 392
Constraint 3	0	0	0
Constraint 4	0	0	0
Setpoint 1	496	446	468
Setpoint 3	1 326	1 106	1 139
Setpoint 4	1 039	889	922
Control 1	827	886	781
Control 2	1 037	1 003	968
Objective	**7 543**	**6 669**	**6 665**
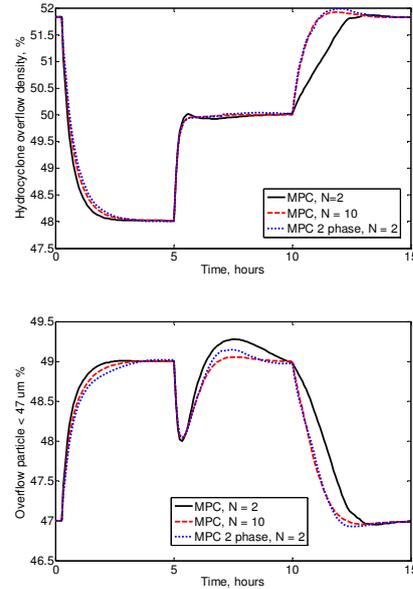
Figure 3. Controlled variables: a – Hydrocyclone overflow density % solids, b – % of particles < 47 μm
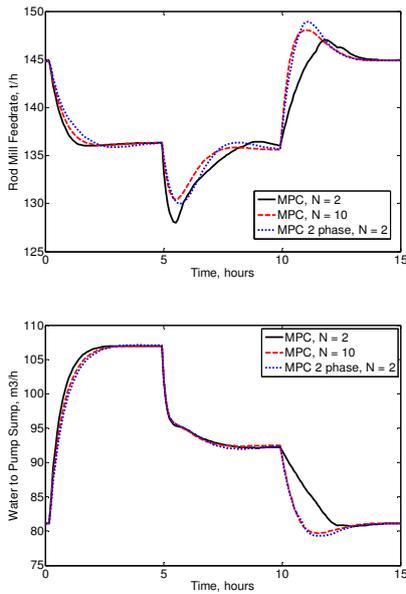
Figure 4. Manipulated variables: a – rod mill feedrate, b – water to pump sump

5. CONCLUSION

A two-phase MPC controller is described in the paper, where the best second phase open-loop control is chosen from a predefined set of open-loop controls. Thus, a more accurate estimation of the value function is used as a terminal cost in comparison with the QP formulation. The numerical comparison against the QP formulation of MPC has shown that similar performance might be achieved with a shorter control horizon. Thereby a progress in the trade-off between the performance and computational demands of MPC is made.

REFERENCES

Fleming W.H., Rishel R.W., (1975), *Deterministic and stochastic control*. New York, Springer.

Richalet J., Rault A., Testud J. L., Papon J., (1978), Model predictive heuristic control: Applications to industrial processes, *Automatica* 14, pp. 413-428.

Prett D.M., Ramarker B.L., Cutler C.R., (1982), Dynamic matrix control method, *United States Patent 4349869.*

Qin S.J., Badgwell T.A. (2003), A survey of industrial model predictive control technology, *Control Engineering Practice*, 11, pp. 733-764.

Bitmead R.R., Gevers M., Wertz V., (1990), *Adaptive optimal control–The thinking man's GPC.* Englewood Cliffs, NJ: Prentice –Hall.

Keerthi S. S., Gilbert E. G., (1988), Optimal, infinite horizon feedback laws for a general class of constrained discrete time systems: Stability and moving-horizon

approximations, *Journal of Optimization Theory and Application,* 57, pp. 265-293.

Mayne D. Q., Michalska H., (1990), Receding horizon control of on-linear systems, *IEEE Transactions on Automatic Control,* 35(5), pp. 814-824.

Mayne D.Q., Rawlings J.B., Rao C.V., Scokaert P.O.M. (2000), Constrained model predictive control: stability and optimality, *Automatica* 36, pp. 789-814.

Magni L.,Sepulchre R., (1997), Stability margins of nonlinear receding-horizon control via inverse optimality, *System & Control letters*, 32, pp. 241-245.

Bemporad A., Morari M., Dua V., Pistikopoulos E.N. (2002), The explicit linear quadratic regulator for constrained systems, *Automatica* 38 pp. 3-20.

Pannocchia G., Rawlings J.B., Wright S.J. (2007), Fast, large-scale model predictive control by partial enumeration, *Automatica*, 43, pp. 852-860.

Rao C.V., Wright S.J., Rawlings J.B. (1998), Application of Interior-Point Methods to Model Predictive Control, *Journal of optimization theory and applications*, 99(3) pp. 723-757.

Di Palma F., Magni L., (2007), *On optimality of nonlinear model predictive control*, System and Control Letters 56, p 58-61.

Sardis G.N, Lee C.G., (1979), An approximation theory of Optimal Control for trainable manipulators, *IEEE Transactions on system, man, and cybernetics*, 9(3), pp. 152-159.

Cannon M., (2004), Efficient nonlinear model predictive control algorithms, *Annual Reviews in Control*, 28, pp. 229-237.

Grune L., Pannek J., (2009), Practical NMPC suboptimality estimates along trajectories, *System & Control Letters*, 58, pp 161-168,

Chmielewski D.J., Manousiouthakis V., (1996), On constrained infinite-time linear–quadratic optimal control, *System and Control Letters*, 29, pp. 121-129.

Sznaier M., Damborg M.J. (1987), Suboptimal control of linear system with state and control inequality constraints, *Proceedings of 26th Conference on Decision and Control*, Los Angeles, CA, December 1987.

Pomerleau A., Hodouin B., Desbiens A., Gagnon E. (2000), A survey of grinding circuit control methods: from decentralized PID controllers to multivariable predictive controllers, *Powder Technology,* 108, pp. 103-115.

Hodouin D., Jämsä-Jounela S.-L., Carvalho M.T., Bergh L. (2001), State of the art and challenges in mineral processing control, *Control Engineering Practice*, 9, pp. 995-1005.

Lestage R., Pomerleau A., Hodouin D. (2002), Constrained real-time optimization of a grinding circuit using steady-state linear programming supervisory control, *Powder Technology,* 124, pp. 254-263.

ARX MPC for people with type 1 diabetes

Dimitri Boiroux, Daniel A. Finan, John Bagterp Jørgensen, Niels Kjølstad Poulsen and Henrik Madsen

Abstract

Type 1 diabetes is a chronic disease characterized by a lack of production of pancreatic insulin, consequently leading to high blood glucose concentrations (hyperglycemia). Hyperglycemia has negative health effects in the long term such as eye, nerve, and kidney disease. Exogenous insulin must be injected to keep the blood glucose in the normoglycemic range (approximately 60 – 140 mg/dL, or 3.3 – 8 mmol/L). However, the dosing of exogenous insulin must be done carefully, because low blood glucose concentrations (hypoglycemia) can have immediate and severe consequences like insulin shock, coma, or even death. Currently, insulin administration is performed by the subject with type 1 diabetes based on infrequent glucose measurements (in the form of finger-sticks), often resulting in an unsatisfactory blood glucose control.

An artificial pancreas is a medical device that injects exogenous insulin automatically in order to regulate the glucose concentration. Blood glucose measurements are obtained from a continuous glucose monitor (CGM). Insulin is administrated either continuously through an insulin pump, or at discrete times using an insulin pen. A control algorithm uses previous glucose measurements and insulin injection information to compute the optimal insulin administration for the current conditions.

We use model predictive control (MPC) to compute the optimal insulin administration for 20 virtual type 1 diabetes subjects. The system (i.e., subject) has one manipulated input (insulin infusion rate), one disturbance input (carbohydrate meals), and one measured output (blood glucose concentration). The subject is represented by a system of nonlinear differential equations describing the dynamic effects of insulin and meals on blood glucose [4]. Twenty parameter sets are used in the study, each representing a different virtual subject.

The model used in the MPC is a low order autoregressive exogenous-input (ARX) model [3]. Due to both the linearity and relative parsimony of the ARX model, there is a significant amount of

subject/model mismatch in the model predictions, reflecting real-world conditions. In general, a simple ARX MPC cannot reject a step disturbance without a resulting offset; thus, the state vector is reformulated using an extended ΔARX description (E-ΔARX) [4], i.e.

$$(1-q^{-1})A(q^{-1})y(t) = (1-q^{-1})B(q^{-1})u(t) + (1-\alpha q^{-1})e(t)$$

in which q^{-1} is the backward shift operator, A and B are polynomials, $e(t)$ is a white noise process, and $0 \leqslant \alpha \leqslant 1$ is a tuning parameter.

The reference signal is time-varying, and is based on the optimal open-loop glucose profile [2]. Insulin-on-board constraints are implemented to avoid overdosing insulin. State estimation is based on a Kalman filter using the noise model described in [1] to simulate a realistic CGM.

We present the MPC results for simulations of the 20 virtual subjects with type 1 diabetes. In particular, we investigate the effects of the prediction horizon length on the control quality of blood glucose and the robustness of the solution.

References

[1] Breton M, Kovatchev B. Analysis, modeling, and simulation of the accuracy of continuous glucose sensors. J Diabetes Sci Technol. 2008;2(5):853-862.

[2] Cobelli C, Dalla Man C, Sparacino G, Magni L, De Nicolao G, Kovatchev B. Diabetes: models, signals, and control. IEEE Rev Biomed Eng. 2009;2:54-96.

[3] Finan DA, Zisser H, Jovanovic L, Bevier WC, Seborg DE. Practical issues in the identification of empirical models from simulated type 1 diabetes data. Diabetes Technol Ther. 2007;9(5):438-450.

[4] Huusom J.K., Poulsen N.K, Jørgensen S.B., Jørgensen J.B. Tuning of methods for offset free MPC based on ARX model representations. 2010 American Control Conference (ACC 2010). *Accepted*.

[5] Kovatchev B., Breton M, Man CD, Cobelli C. In silico preclinical trials: a proof of concept in closed-loop control of type 1 diabetes. J Diabetes Sci Technol. 2009;3(1):44-55.

Tuning of ARX-based Model Predictive Control for Offset-free Tracking.

Jakob Kjøbsted Huusom[a], Niels Kjølstad Poulsen[b],
Sten Bay Jørgensen[a] and John Bagterp Jørgensen[b]

[a]CAPEC, Department of Chemical and Biochemical Engineering
[b]Department of Informatics and Mathematical Modelling
Technical University of Denmark

Keywords: Model Predictive Control, ARX-model, Controller Tuning, Recursive estimation

Model Predictive Control (MPC) is a control technology that uses a model of the system to predict the process output over some future horizon. The controlled input signal is determined by solution of an open-loop optimization problem using the model of the system. The first part of the optimal control sequence is implemented on the process. Feedback is obtained by repeating this procedure as new measurements become available. Advanced control strategies such as Model Predictive Control have gained wide spread interest in many areas in the chemical industries, due to fast algorithms, a well established theory and growing number of successful industrial implementations. The main feature is that the optimal control signal is determined as a constrained optimization which utilizes future predictions of the plant behaviour. Hence the controller has a plant model embedded for state estimation. The achieved closed loop performance is therefore dependent on the quality of the future predictions. The performance of the state estimator is on the other hand dependent on the accuracy of the process and the noise model. In this contribution, we discuss closed loop performance of MPC based on ARX models when applied to systems with unmeasured step disturbances.

The majority of industrial MPC applications are today still based on linear models. ARMAX models (AutoRegresive Moving Average model with eXogenous input) can be identified using standard tools from time series analysis and systems identification. However, for multivariable systems it is difficult to select a structure for the ARMAX model. Furthermore, identification of the parameters in ARMAX models constitutes a non-linear non-convex optimization problem. If the input-output model is simplified to an ARX model (AutoRegresive model with eXogenous input), the estimation problem becomes a convex optimization problem. Furthermore the MIMO system can be handled as easily as SISO systems. The discrete time ARX model structure is given as

$$A(q^{-1})y_k = B(q^{-1})u_k + \varepsilon_k \qquad \varepsilon_k \sim N(0,\sigma^2)$$

where $A(q^{-1})$ and $B(q^{-1})$ are polynomials in the back shift operator q^{-1}.

Unmeasured step disturbances are common in the process industries and appear for instance when a feed source changes. The composition of crude oil in refineries may change significantly when feed is changed from one well to another. Similar unknown steps may occur in a cement plant when raw minerals are changed. Unmeasured steps may also be used to represent the inevitable model-plant mismatch. To reject such disturbances the basic MPC formulation needs to be expanded, such that model errors are introduced in order to include integrators or augment the system with disturbance models. Alternatively the model predictive control algorithm can be combined with a recursive disturbance estimation algorithm. The closed loop performance of the system will depend on the nature of the disturbance and how the disturbance rejection is facilitated by the control algorithm.

This contribution will analyse a range of methods for achieving offset-free tracking in ARX-based model predictive control. Considerations regarding pre-tuning of the free parameters in the controller are also provided based on the infinite horizon, unconstrained model predictive controller, i.e. the LQG control design. The tuning aims at balancing the ability to reject unmeasured disturbances vs. noise sensitivity while keeping both the closed loop input and the output variance small. This type of trade off is illustrated on a small example. A series of closed loop simulations of the MPC has been performed using the third order ARX model as the true system. Three MPC implementations are tested on a fixed scenario. The total simulation horizon is 250 samples. Between time 50 and 100 a step is introduced in the reference and between 150 and 200 an unmeasured step disturbance is acting on the system. The input will be constrained between ±1. The models used by the controller are the ARX model, the ΔARX model and finally the EΔARX model. The Δ in model name indicates that the noise is modelled as integrated white noise. The EΔARX model use an extended model formulation of the noise and has a free parameter α which can vary between 0 and 1 in order to span the range between the two other models.

Deterministic simulations

Noisy simulations

| **ARX-MPC** | **ΔARX-MPC** | **EΔARX-MPC(α=0.7)** |
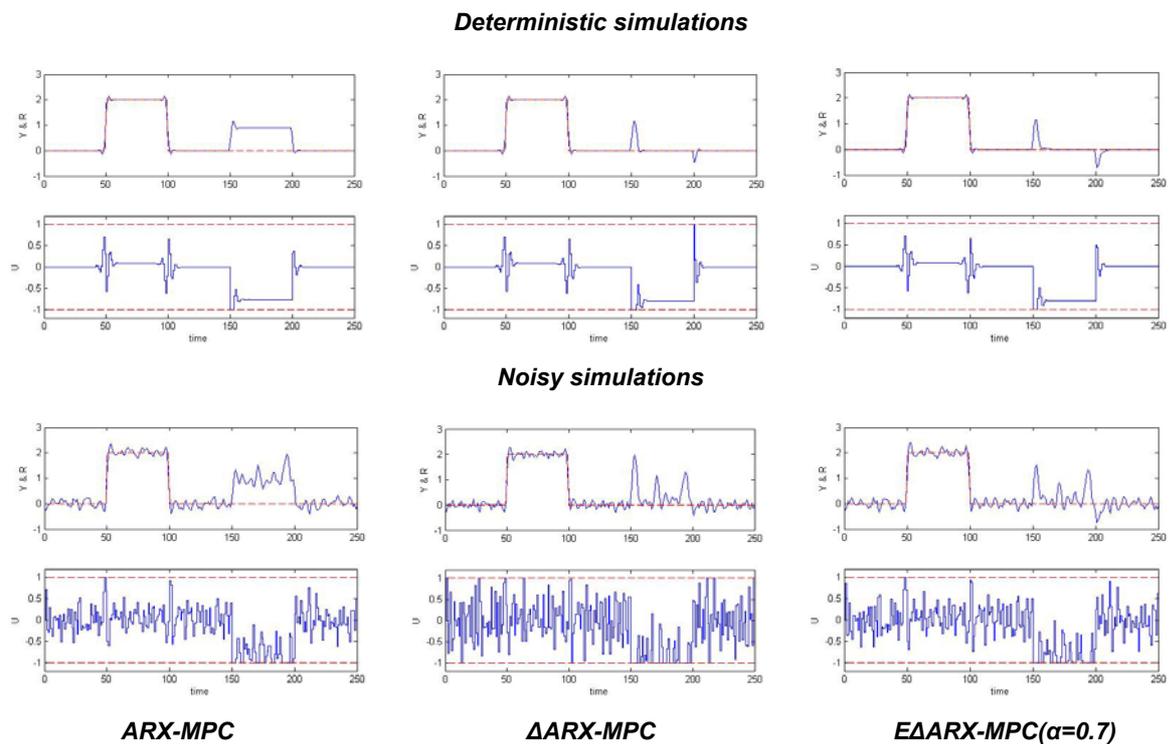
It is clearly seen on the figure that only the ΔARX and the EΔARX models can give offset free tracking. The output trajectories resulting from these two models are very similar but the EΔARX model give a less aggressive control action.

Current research involves extensions and benchmarking of the framework using ARX-based, offset-free MPC algorithms for a series of examples relevant for the process industry involving multivariable and time delay systems.

Comparison of Decentralized Controller and MPC in Control Structure of a CO_2 Capturing Process

Mehdi Panahi, Sigurd Skogestad

Department of Chemical Engineering, Norwegian University of Science and Technology(NTNU), 7491 Trondheim, Norway

Abstract

In a previous study [1], we did the control structure design for a post combustion CO_2 capturing process using Skogestad's method [2]
Based on the designed structure, in this paper we investigate the performance of the control structure with large disturbances where some of the manipulated variables saturate.
To handle this kind of disturbances, a MPC has been designed and implemented on parts of the process and the results show that we achieve good performance. On the other hand, if we implement decentralized PID controllers, one needs to use the reverse pairing of what would be desired for good performance in order to avoid instability when we have saturation.

Keywords: Process control, Plantwide control, Self-optimizing method

Introduction

Absorption/Stripping CO_2 capturing processes are used in post-combustion part of power plants to remove CO_2 from flue gas streams. A simple typical flowsheet of such processes has been shown in Fig.1
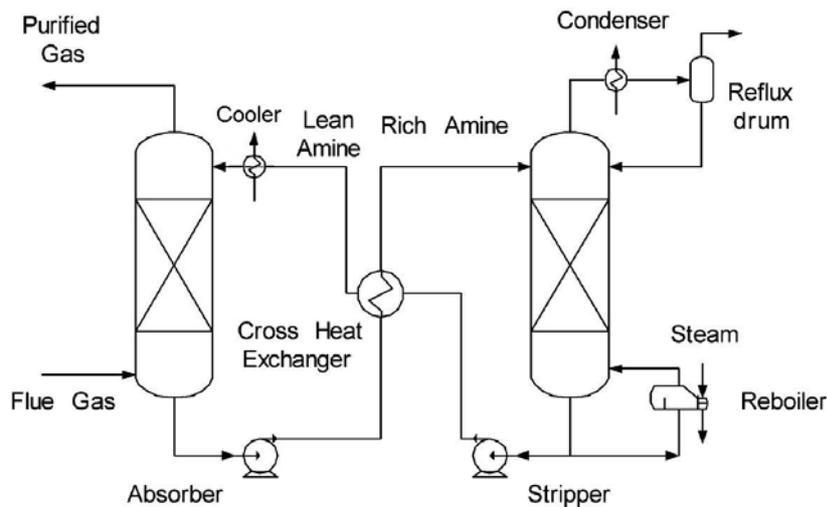
Fig.1- Absorption/Stripping CO_2 capturing process [2]

Energy requirements of CO_2 capturing processes are relatively very high compare to the net generated power in the main plant and operating of this process in the lower operation costs is important to save more energy. To keep this process in the minimum level of energy requirement, a robust control structure is needed.
We have designed a control structure using Skogestad's method to omit the necessity of reoptimization of the process when disturbances happen with an acceptable loss for the energy requirement of the process. Fig.2 shows the proposed control structure. [1]

1

In this structure, recycle amine flowrate is manipulated to remove 90% of the CO_2 content in the feed flue gas and temperature control of tray no.4 (counting from the top) in the stripper (20 trays) has been found as the best self-optimizing variable to be controlled by using the only unconstrained degree of freedom that is reboiler duty of the stripper. [1] The other control loops are level controllers or are the loops to control the variables that have been active during self-optimizing stepwise approach. [3]
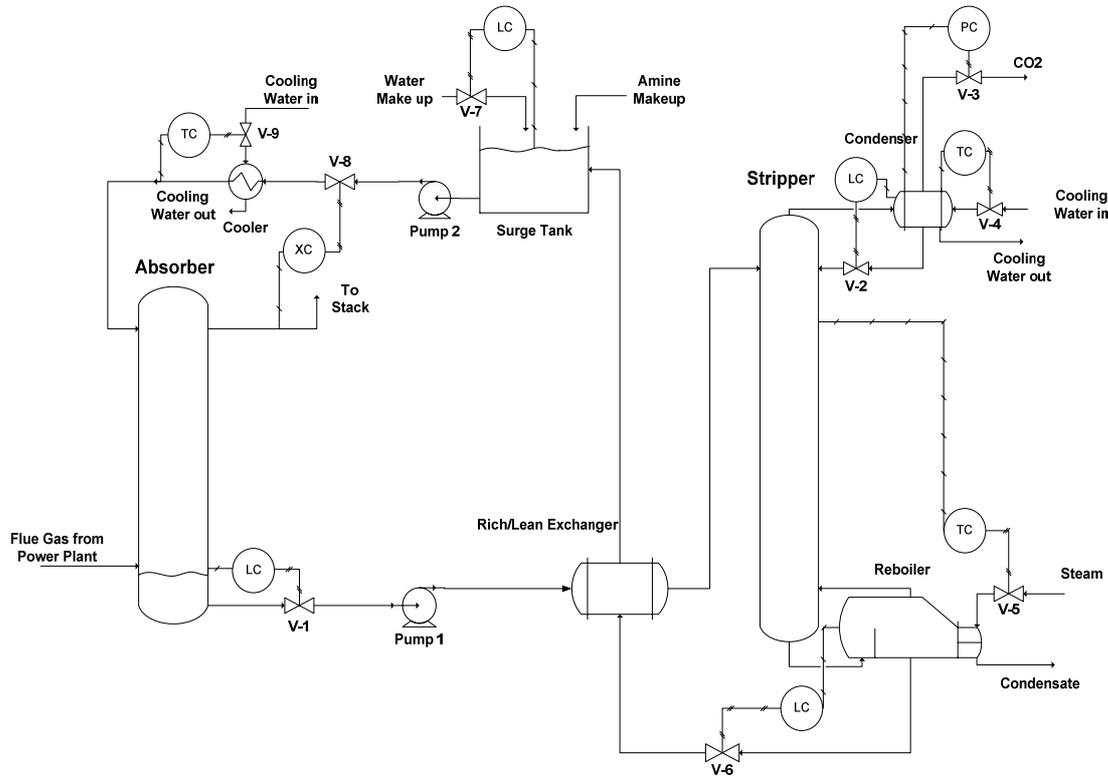
Fig. 2- Process flowsheet with proposed control loops using self-optimizing method [1]

One of the main disturbances in this process is changing of the feed flowrate into the absorber. In this paper we try to investigate and compare the performance of MPC and traditional PID controllers for stability of the proposed control structure. The MPC controller is designed and implemented using Honeywell technology that is named as Robust Model Predictive Control Technology (RMPCT) in UniSim process simulator. [4]

Performance of PID loops and MPC for stabilization of the process

The proposed control structure using decentralized PID controllers gives a stable system as long as the change in flowrate of the flue gas is less than +15% though saturation of reboiler duty happens when flowrate of flue gas reaches to +10%. But, when the accession of flowrate passes +15% then temperature of tray no.4 decreases significantly (Fig.4) and consequently the amount of CO_2 at the bottom of stripper starts increasing and with recycling to the absorber, CO_2 is accumulated in the process and makes the plant unstable. (Fig.5)

2

(a)

(b)

(c)
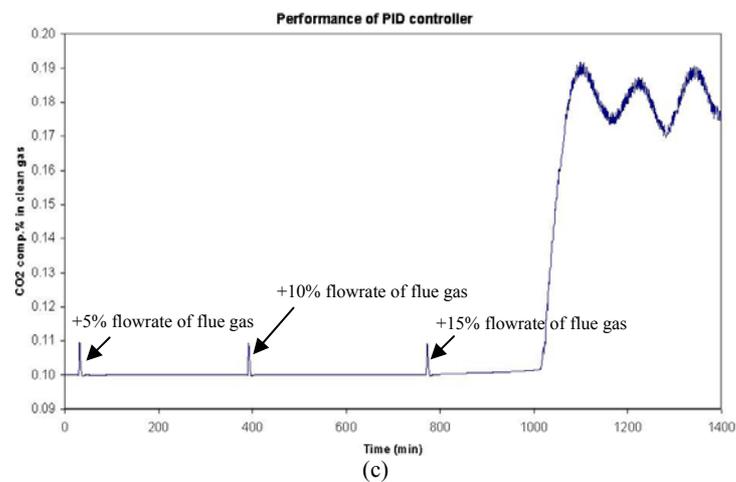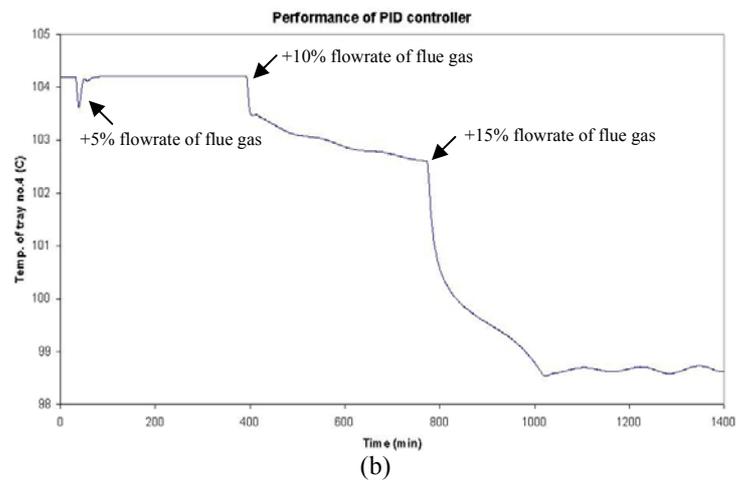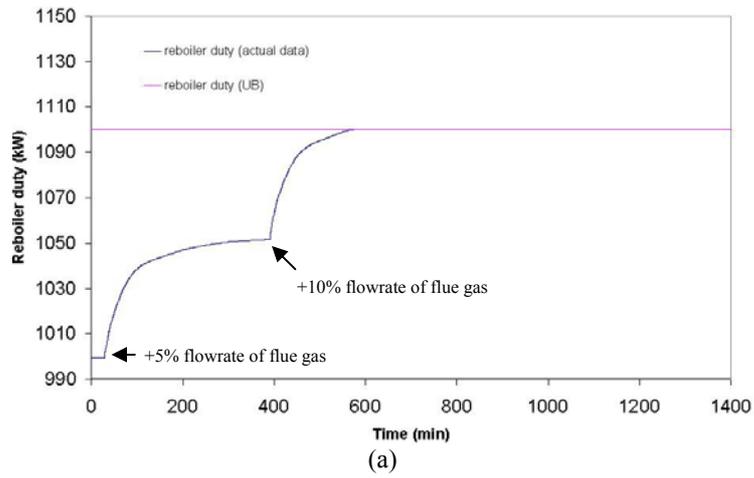
Fig. 3- a: saturation of reboiler duty when the flowrates of flue gas increases to +10%, b: when reboiler duty is saturated, by increasing feed flow to +15%, temperature of tray no.4 decreases significantly, c: instability of the process when flowrate of flue gas increases +15%

To avoid instability in our control structure we have considered two different ways: Reverse pairing and design a MPC.

3

Reverse pairing

We control the temperature of tray no.4 using recycle amine flowrate and give up controlling of CO_2 content in the clean gas when reboiler duty has become saturated. (Fig. 6a and b) Fig. 6b shows that by using recycle amine flowrate we can control tray temperature no. 4 in its setpoint value.

(a)

(b)
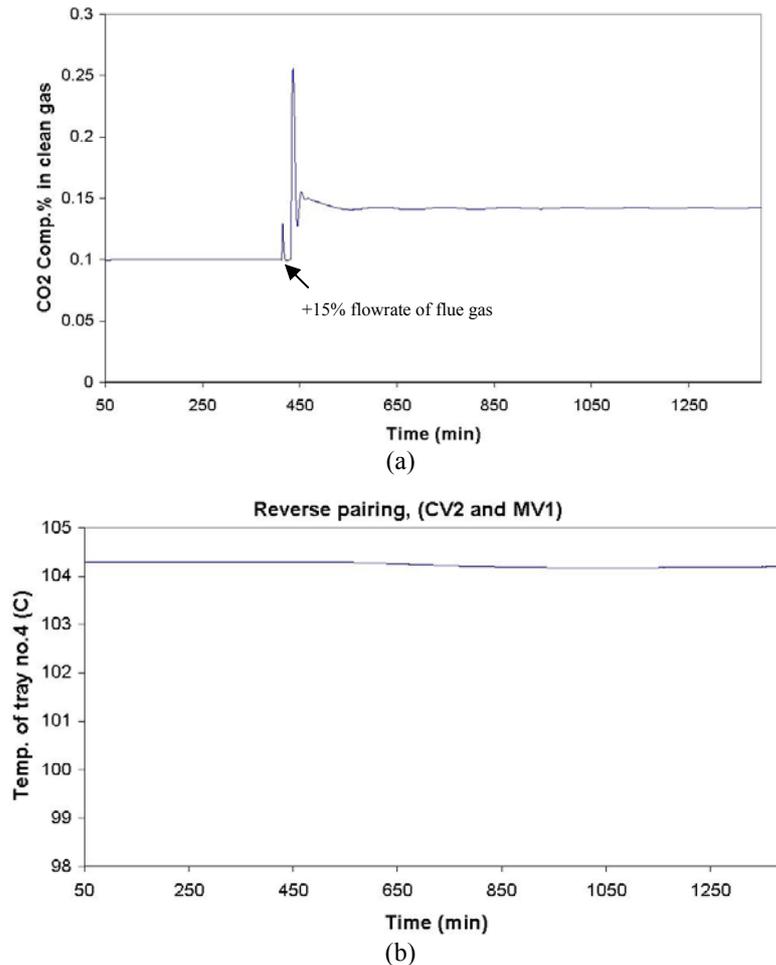
Fig. 6- By reverse pairing and giving up controlling the CO_2 content in the clean gas when reboiler duty is saturated, control structure is stable when flue gas increases +15%

MPC (RMPCT)

A MPC controller is used in order to achieve CO_2 composition control in the clean gas stream and temperature control of tray no.4 in the stripper. In order to keep track of the pressure drop over the stripper, the difference between bottom stage pressure and top stage pressure is included as an auxiliary CV in the identification routine. List of CVs, MVs and DVs for this controller are in table 1.

Table 1. List of CVs, MVs and DVs for design a MPC

CVs	MVs	DVs
CV1- CO_2 content in clean gas CV2- Temperature of tray no.4 in the stripper CV3- pressure drop in stripper	MV1- Recycle amine flowrate to absorber MV2- reboiler duty of the stripper	DV1- change in flowrate of flue gas (feed)

Identification of the model and making the required data files for the MPC are made by Profit Design Studio (PDS) [5] and then the generated files are loaded in UniSim to run the RMPCT.

4

Fig. 7 shows the performance of RMPCT for controlling this process when disturbance with different magnitude happens. As Fig. 6a shows in +15% change of flowrate of flue gas, RMPCT gives up controlling of CO_2 content in the clean gas while it control tray temperature in its setpoint.

(a)

(b)

(c)
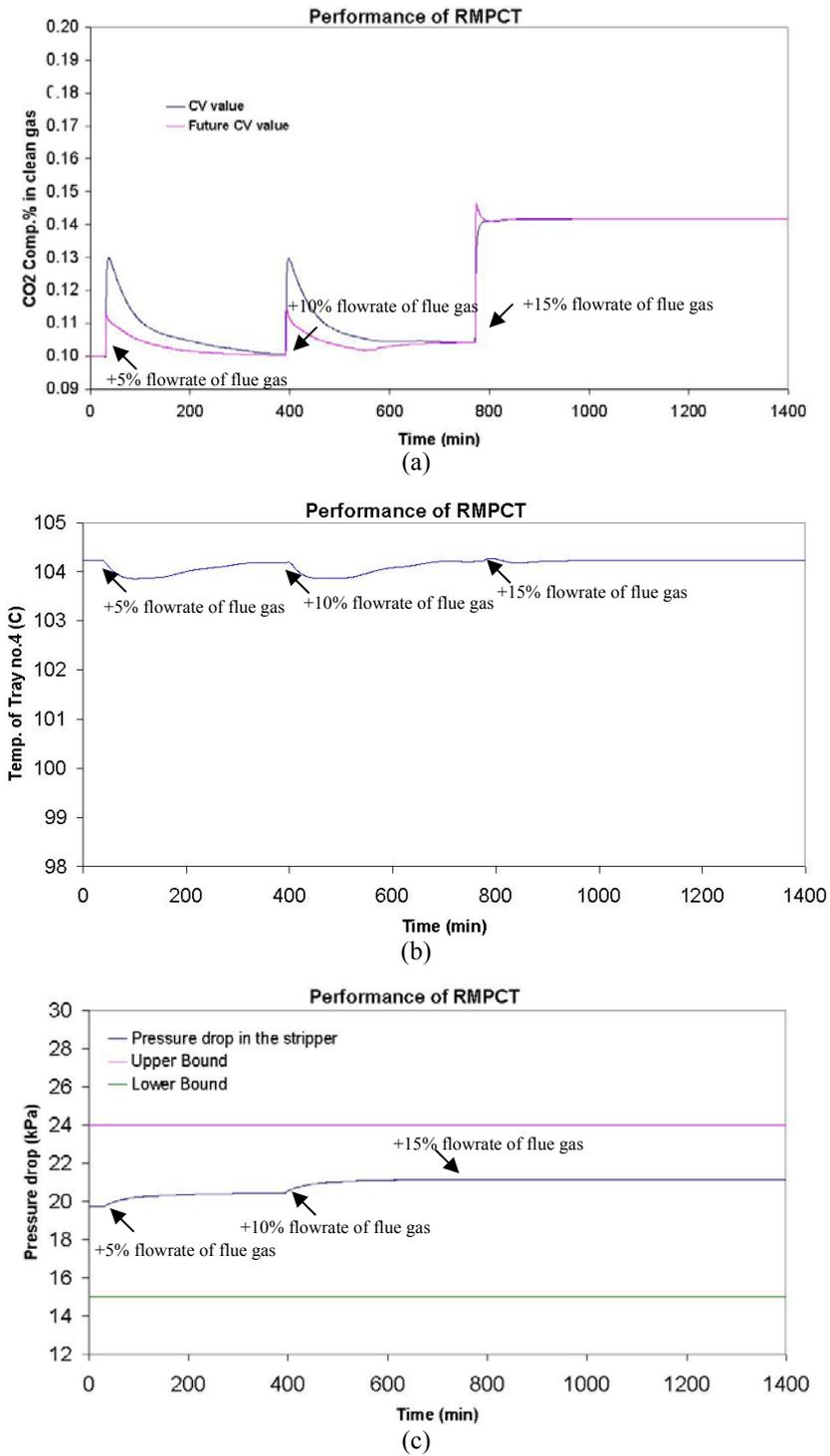
Fig. 7- a: CO_2 content in clean gas stream, RMPCT has given up controlling of CO_2 content in +15% of the disturbance, b: Temperature of tray no.4 in the stripper is always in its setpoint c: pressure drop of stripper

5

Conclusion

Performance of RMPCT and decentralized PIDs for stabilization of a designed control structure by self-optimizing method for a CO_2 capturing process has been considered when large disturbances have happened. Results show that using RMPCT, the control structure will be stable while to keep the process stable using PID loops, reverse pairings of some MVs and CVs has to be done.

Acknowledge

We would like to acknowledge to Bjørn Einar Bjartnes from Honeywell Company for helping in implementing of the RMPCT in UniSim.

References

[1] Panahi, M., M. Karimi, S. Skogestad, M. Hillestad, H. F. Svendsen, 2010, Self-Optimizing and Control Structure Design for a CO_2 Capturing Plant, presented in 2nd gas processing symposium, Doha, Qatar, Jan.12-14.

[2] Skogestad, S., 2004, Control Structure Design for Complete Chemical Plants, Computers and Chemical Engineering, 28, 219-234

[3] Skogestad, S., 2000, Plantwide control: The search for the self-optimizing control structure, Process Control, 10, 487-507

[4] UniSim design R380, Honeywell company

[5] Profit Design Studio (PDS R310) , Honeywell company

6

Potential of Economic Model Predictive Control for Management of Multiple Power Producers and Consumers

Tobias Gybel Hovgaard*, John Bagterp Jørgensen**

*Danfoss A/S, Nordborgvej 81, DK-6430 Nordborg, Denmark
tgh@danfoss.com
**DTU Informatics, Technical University of Denmark, Richard Petersens
Plads, Building 321, DK-2800 Kgs Lyngby, Denmark
jbj@imm.dtu.dk

Keywords : Model Predictive Control, Optimization, Energy Systems, Refrigeration Systems

Economic Model Predictive Control is a receding horizon controller that minimizes an economic objective function rather than a weighted least squares objective function as in Model Predictive Control (MPC). We use Economic MPC to operate a portfolio of power generators and consumers such that the cost of producing the required power is minimized. The power generators are controllable power generators such as combined heat and power generators (CHP), coal and gas fired power generators, as well as a significant share of uncontrollable power generators such as parks of wind turbines. In addition, some of the power consumers are controllable. In this paper, the controllable power consumers are exemplified by large cold rooms or aggregations of super markets with refrigeration systems. We formulate the Economic MPC as a linear program. By simulation, we demonstrate the performance of Economic MPC for a small conceptual example.

We consider a number of dynamically independent systems that have the one thing in common that they all influence a quantity that has to meet a set of constraints. To be more specific we have a portfolio of multiple power generators which delivers a total production for the entire portfolio. On the demand side a cold room consumes power in order to keep the temperature within certain bounds. Furthermore a reference consumption (which is assumed to be predictable) combining all other power consumers and non-controllable power producers like farms of wind turbines is added to the demand side. By adding the cold room to the optimization problem the potential savings gained by controlling flexible loads on the demand side are revealed. The idea is that the thermal capacity in the refrigerated goods can be utilized to store "coldness" such that the refrigeration system can pre-cool when the energy is free (i.e. there is an over production from the generators). Thereby a lower than normally required cooling capacity can be applied later, for a period of time when the energy prices are above zero again. The demands to the temperature in the cold room are not violated at any time since the same total cooling capacity is applied though shifted in a more optimal way. We exploit the special property of the refrigeration system that the dynamics of the temperature are rather slow

while the power consumption can be changed rapidly.

For the power generation and cooling problem we should select the cheapest possible feasible trajectory of inputs. Since the cost is related to producing the power, the problem can be stated as:

$$\min_{\{u_k\}_{k=0}^N} \phi = \sum_{k=0}^N c' u_k \tag{1}$$

with:

$$c = [c_1, c_2, \ldots, c_n, 0]^T \quad u = [u_1, u_2, \ldots, u_n, T_e]^T \tag{2}$$

u_i is the input to power plant i, T_e is the input to the cold room and c_i is the cost related to u_i.

The problem in (1) is subject to:
- system equations:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Ed_k & k = 0, 1, \ldots, N \\ y_k &= Cx_k + Du_k & k = 0, 1, \ldots, N+1 \end{aligned} \tag{3}$$

- limitations on the input:

$$\begin{aligned} u_{min} &\leq u_k \leq u_{max} & k = 0, 1, \ldots, N \\ \Delta u_{min} &\leq \Delta u_k \leq \Delta u_{max} & k = 0, 1, \ldots, N \end{aligned} \tag{4}$$

- cooling capacity must be positive.
- production must be higher than demand at all times.
The last constraint is softened to allow the production to be lower than the demand such that the we do not end up with an infeasible problem that cannot be solved. However the penalty on underproduction is selected sufficiently large such that power demands are met whenever possible.

The cost function and the constraints are formulated as a linear program (LP) on the form:

$$\min_x \phi = g'x \tag{5}$$

s.t.

$$Ax \geq b \tag{6}$$

where A is a block-angular constraint matrix. The LP in (5) and (6) can be solved with standard solvers.

Simulations of a small case study with two power plants (a fast and expensive and a slow but cheaper) and one cold room reveals a potential for significant savings.

The 16th Nordic Process Control Workshop (NPCW'10) 2
Department of Automatic Control and the Process Industrial Centre, Lund University, Lund, Sweden, August 25-27, 2010

38

Fault Detection for the Benfield Process using a Parametric Identification Approach

Johannes P. Maree[1], Fernando R. Camisani-Calzolari

Department of Engineering Cybernetics
Norwegian University of Science and Technology
7491 Trondheim, Norway

The Sasol group of companies specialises in diverse fuel, chemical and related manufacturing and marketing operations. Sasol has interests in oil and gas exploration and production, in crude oil refining and liquid fuels marketing. The efficient and economical recovery of carbon dioxide (CO_2), used in various processes at Sasol, is accomplished by utilizing the Benfield process for CO_2 extraction. The Benfield process is a thermally regenerated cyclical solvent process that uses an activated inhibited Hot Potassium Carbonate solution to remove CO_2, H_2S and other acid gas components.

The current operating philosophy for the Benfield process is to keep it simultaneously hydraulic and loaded with CO_2, as far as possible, to meet optimal profit margins from the gas circuit operations. The hydraulic load is defined as the maximum volume of CO_2 that can be processed. Regeneration efficiency has been identified as one of the major efficiency measures by UOP (Benfield technology licensor). Regeneration efficiency is a measure of how much steam is required per unit volume of CO_2 removed, and gives an indication of the unit cost, the overall pressure drop in the regeneration column, and the solution health. This regeneration is directly dependant on the CO_2 absorption in the wash columns, which in its turn is again affected by the wash flow-rates. Regular foaming and flooding diminishes the efficient CO_2 absorption into the Potassium Carbonate wash solution, resulting in inefficient regeneration. Foaming and flooding is caused by abnormally high differential pressures on either the top, middle or bottom bed. Bed differential pressures increase with increasing liquid and gas loads. High differential pressures that change erratically indicate flooding or foaming. Abnormal high and stable differential pressures indicate partial blockages in packed beds or liquid/gas distributors. Abnormal differential bed pressures are classified as multiplicative faults in abnormal process behaviour, which can be detected by monitoring parameter fluctuations.

This work proposes a parametric model-based approach to Fault Detection (FD). The proposed framework combines a parity space approach (subspace SID), to identify initial process models, with a joint state and parameter estimation method (Extended Kalman filter), to monitor parameter fluctuations used for FD. The motivation for subspace SID methods is that these methods have proven to be computational efficient where no a-priori process knowledge is necessary to estimate a system model. Subspace SID methods thus allow the user to identify black box models, which can be used to monitor processes. The challenge in fault detection with subspace methods comes in how to monitor and evaluate the vast amount of system parameters efficiently and elegantly. Re-identifying the process using subspace methods, necessary to track parameter changes, is also not a feasible solution to fault detection. The reason is due to the vast amount of data samples necessary, which must contain well-excited process

---

dynamics. An elegant solution to the FD problem, using subspace SID methods, would thus be to identify an initial system process model, using a subspace method, where the initial identified parameters of the subspace model are re-cursively updated as new data becomes available. By updating the process parameters, without complete system re-identification, the user is able to track parameter changes, which contributes to the FD and plant-model mismatch problem.

The proposed subspace SID method is based on two consecutive orthogonal decompositions to identify a system in closed-loop. The subspace SID method furthermore guarantees the estimation of stable system matrices by utilizing the shift-invariant property of the extended observability matrix. Extended Kalman filtering is used to recursively update a joint set of initial system states and parameters, using current sampled process data and initial estimated parameters, obtained via the subspace method. To detect a fault in the updated parameter set, it is necessary to compare the updated parameter set, with the initial identified parameter set. This can be accomplished by considering appropriate matrix measures, which will accentuate any fundamental parameter matrix differences. The infinity matrix norm is proposed for detecting discrepancies between the absolute difference of the initial identified, and recently updated parameter set. The infinity matrix norm is used to reduce (averaging) the matrix of differences to one value which is monitored against a threshold value.

Diagnosis of oscillation due to multiples sources using wavelet transforms

Selvanathan Sivalingam and Morten Hovd

Abstract— Control loop oscillations are a common type of plant-wide disturbance and the root-causes can be one or more among poorly tuned controllers, process or actuator non-linearities, presence of model plant mismatch and oscillatory disturbances. The oscillations are the most prominent indications of deteriorated controller performance. This article addresses the detection and diagnosis of oscillations in measurements due to multiple sources under the framework of internal model control. A pattern recognition based approach using cross wavelet transforms is proposed to pinpoint the source(s) of oscillation in the control loop. The phase information in wavelet domain between input and output signals is exploited to diagnose the source of oscillation.

I. INTRODUCTION

It is well known that performance degradation in control loops manifest as one or more of the following: (*i*) poor set point (SP) tracking (*ii*) oscillations (*iii*) poor disturbance rejection and (*iv*) excessive final control element variation. Industrial surveys over the last decade indicate that only about one-third of industrial controllers provide acceptable performance and about 30% [1], [2], [3] of the control loops exhibit oscillation. Since oscillations can lead to loss of energy, isolating the root cause for oscillations is important for improving the performance of the oscillating control loops. Also, the presence of the oscillations in a plant increases the variability of the process variables and thus may cause loss of product quality. The oscillations (or vibrations), in general, are a very drastic form of plant performance degradation, which can, in many cases, be induced by the feedback mechanism itself.

Oscillations are attributed to one or more among poor controller tuning, process or actuator non-linearities, presence of model plant mismatch or oscillatory disturbances. A tool to help the engineer should therefore automatically bring oscillatory loops to his or her attention, characterize them and highlight the presence of plant wide oscillations. Several authors have addressed the detection of oscillatory measurements in process data. Early works appear in [4] followed by [5], [6], [7], [8]. [4] proposed a technique to detect oscillating loops "on-line" using the IAE criterion. This method does not assume any particular shape for oscillation and only requires the measurement to deviate significantly from the set point. [4] also proposed a diagnostic procedure for finding the source of oscillation and eliminating it. The diagnostic

Selvanathan Sivalingam and Morten Hovd are with Department of Engineering Cybernetics, Norwegian University of Science and Technology, Norway 7491. Email addresses: selvanathan.sivalingam@itk.ntnu.no and morten.hovd@itk.ntnu.no

procedure is carried out by disconnecting the feedback (*i.e.* switching the controller to manual mode). This approach is simple and efficient and probably the most comprehensive procedure available for diagnosing root cause for oscillations. However, switching the controller to manual mode may not always be allowed, especially if the loop is deemed critical. Further, it will not be possible to apply this approach on thousands of loops in a routine fashion. [5] presented an offline technique for detecting oscillation using a regularity factor. This method requires the user to specify the root-mean-square value of the noise and a thresholds a nontrivial task when applied to hundreds of loops.

[5] and [9] proposed a set of procedures to detect and diagnose oscillating loops using offline data. They combine the techniques of controller performance assessment along with operational signatures (OP-PV plots) and spectral analysis of the controller error for diagnosis. This technique, though not completely automated, can distinguish the cause of oscillation as one of the following: (*i*) poor tuning (*ii*) nonlinearity or (*iii*) external disturbance. However, the downside lies in manually inferring the loop signatures that are based on spectral analysis or on a map of controller output (OP) versus process variable (PV) and isolating the oscillating portion from the entire data. [10] presented a simple, practical approach to distinguish oscillating loops that are caused by external disturbances and static friction. This approach is based on cross-correlation between the controller output (OP) and process output (PV). The cross-correlation technique failed when the data had intermittent oscillations and the set-point was also changing. [11] also proposed a technique to identify stiction using nonlinear filters. The method assumed that information such as mass of stem, diaphragm area, and so on for each valve is readily available. Since in a typical process industry facility there can be thousands of control loops, it may be nearly impossible to build/maintain a knowledge base of control valves, making this technique difficult to implement.

[12] used higher order statistics for detecting nonlinearity in data and have extended the method for diagnosing stiction by fitting an ellipse of the OP-PV plot and inferring the stiction from an assumed stiction model. The success of this approach lies in correctly identifying the oscillation period and its start and end point in the OP-PV data. [8] proposed non-negative matrix factorization for detection and diagnosis of plant-wide oscillations based on source separation techniques. As can be seen, the task of detecting stiction or other nonlinearities in valves from routine operating data is a challenging task. To summarize, data driven techniques
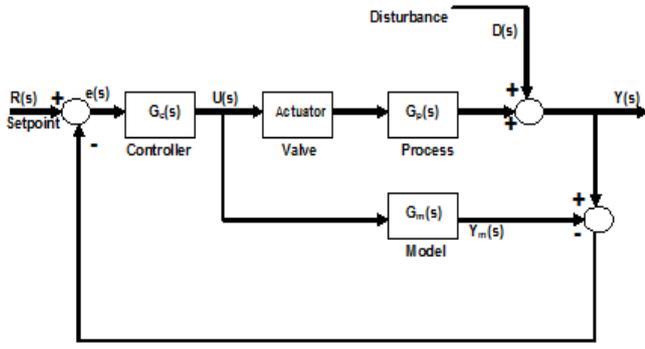
Fig. 1.   Schematic representation of internal model control with actuator

Fig. 2.   Time domain behavior of input and output signals considered for interpretation of wavelet analysis

that are presented in the literature till date are useful in (a) assessing the performance of the controller by calculating a figure of merits given that the cause of poor-performance is only due to either an aggressive or sluggishly tuned controller in pure feedback control, (b) detecting oscillating loops with an user-specified parameter, and (c) limited diagnosis of the cause of oscillation based on cross-correlation, power spectral analysis, or OP-PV plots. The current approaches lack (a) the capability to efficiently diagnose oscillations due to multiple sources, (b) the ability to diagnose the causes of time-varying oscillations and (c) an automated means of oscillation diagnosis.

In this work, we have attempted to address some of the aforementioned drawbacks by using wavelet and cross wavelet transforms.

II. PROBLEM STATEMENT

Oscillations in model based control loops occur due to either one of (*i*) valve stiction (*ii*) model plant mismatch, (*iii*) external oscillatory disturbances or combination of any of these. It becomes vital to diagnose the causes of oscillations in order to take the appropriate remedial action. A procedure based on pattern recognition techniques using cross wavelet transform is devised in this article to diagnose the cause(s) of the oscillation. The problem is setup in the internal model control (IMC) framework (Figure 1).

III. PROPOSED METHODOLOGY

Cross wavelet transform of input and plant and that of input and model output are computed and thereby a specific pattern is sought for root cause diagnosis of oscillation using the direction of wavelet phase difference between the variables.

To illustrate the idea of cross-wavelet transform for an input-output system, an open-loop process with $G_p(s) = \dfrac{1}{10s+1}$ is considered. The process is simulated for a sinusoidal input having two frequencies and the time domain plots of input and output are given in Figure 2. The cross wavelet transform plot between two quantities *u* and *y* is shown in Figure 3.
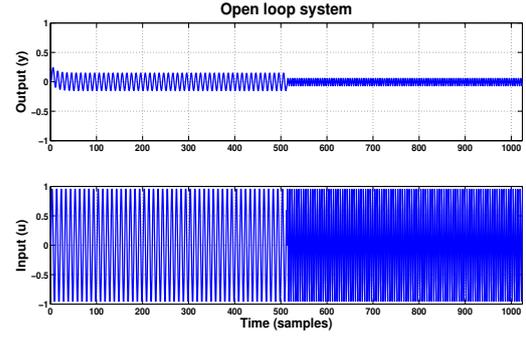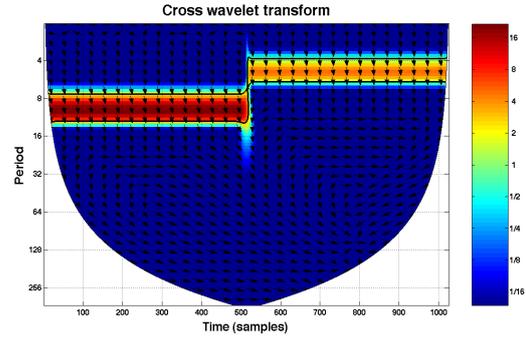
Fig. 3.   Cross wavelet transform between input and output signals

It is known from Figure 3 that the quantities *u* and *y* show high common power at two frequencies between two different time intervals (0.1 Hz, 0-511 and 0.2 Hz, 512-1024) and the arrows indicate the direction of the wavelet phase between *u* and *y* *i.e.,* u leads y by 90° (pointing down). Based on the properties of cross wavelet transform, wavelet phase difference and linear time invariant systems theory, the following methodology is proposed to diagnose the source(s) of oscillation in a control loop.

The quantities controller output (*u*), process output (*y*) and model output (y_m) of an oscillating control loop are obtained either from simulation or from industry. The cross wavelet transforms, $W_{uy}(f, \tau)$ and $W_{uy_m}(f, \tau)$ are computed. By comparing the direction of wavelet phase, the following conclusions can be drawn.

- If the oscillating source is only due to valve stiction, the cross wavelet transform plots should not only exhibit harmonics but also discontinuities.
- If the source is due to gain mismatch, the plots $W_{uy}(f, \tau)$ and $W_{uy_m}(f, \tau)$ of should be identical since the phase spectrum is independent of any changes in gain. The arrows in the plots of $W_{uy}(f, \tau)$ and $W_{uy_m}(f, \tau)$ will be in same direction.
- If the source is due to delay mismatch, the plots of $W_{uy}(f, \tau)$ and $W_{uy_m}(f, \tau)$ will vary in phase direction since the phase spectrum depends on delay changes.

42

IV. SIMULATIONS

A control system consisting of a process characterized by the transfer function $G_p = \dfrac{K_p}{\tau_p s + 1} e^{-d_p s}$ and model $G_m = \dfrac{K_m}{\tau_m s + 1} e^{-d_m s}$ is simulated with IMC controller for a unit step change in the set point. The different case studies analyzed for the diagnosis of oscillation in a control loop are (*i*) oscillation due to valve stiction (*ii*) oscillation due to valve stiction and oscillatory disturbance (*iii*) oscillation due to gain mismatch (*iv*) oscillation due to gain mismatch and oscillatory disturbance and (v) oscillation due to delay mismatch.

A simple yet efficient one parameter model proposed by [4] is used to generate oscillations due to valve stiction. The model is

$$x(t) = \begin{cases} x(t-1) & |u(t) - x(t-1)| \le d \\ u(t) & \text{otherwise} \end{cases} \quad (1)$$

Here $u(t)$ and $x(t-1)$ are present and past valve outputs, $u(t)$ is the present controller output, and d is the valve stiction band. The valve stiction band is expressed in terms of the percentage or fraction of valve movement corresponding to the amount of stiction present in the valve. For instance, if 100 units of force are required to open the valve completely from completely closed position and 10 units of force is required to overcome the amount of static friction in the valve, stiction band is 10% or 0.1. The stiction band of 0.1 is used in the simulation. Model plant mismatch is introduced by changing the values of gain or delay appropriately in the process. The sinusoidal disturbance of frequency 0.01 Hz is considered for the simulation. The time domain plots of controller output (u), plant output (y) and model output (y_m) for different simulation studies is shown in Figures 4, 5, 6, 7 and 8.

The cross wavelet transform computed between controller output and plant output is compared with that computed between controller output and model output. In the case of oscillation due to valve stiction (Figure 4), the plots of cross wavelet transform (Figures 9 & 10) not only show harmonics but also discontinuities which are the characteristics of a sticky valve. Figures 11 and 16 clearly indicate the presence of the valve stiction as one of the sources of oscillation between 800 and 1600 s and the other being the oscillatory component of frequency 0.01 throughout.

If the oscillation is only due to MPM, there will be clearly a single frequency in the cross wavelet transform plot. In the case of gain mismatch (Figure 6), the plots of cross wavelet transform between controller output and plant output and controller output and model output (Figures 13&14) produce identical plots since the phase spectrum is independent of the changes in gain.

The control loop whose time domain trends are characterized by Figure 7 is diagnosed to have gain mismatch as one of the sources of oscillation between 800 and 1600 s and other being the oscillatory component of frequency 0.01 Hz
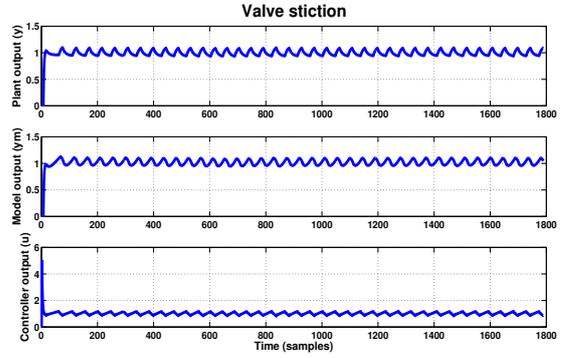
Fig. 4. *Time domain behavior of plant, model and controller outputs for the valve stiction as the source of oscillation.*

(Figures 15 & 16). The presence of oscillatory component can not be due to the presence of delay mismatch since the presence of gain and delay mismatch at a same time will lead to system instability. Hence, the loop can be said to have the external oscillatory disturbance over the entire period and gain mismatch between 800 and 1600 s as the sources of oscillation.

The control loop whose outputs are given in Figure 8 is analyzed for diagnosing the source(s) of oscillations. Figures 17 and 18 indicate the presence of a single frequency component and a directional change in the phase difference. The source of oscillation can be either oscillatory disturbance or delay mismatch.

V. CONCLUSIONS

A pattern recognition technique for the diagnosis of control loop oscillations in internal model control systems due to multiple sources using cross wavelet transform of two quantities has been developed. A diagnostic study of oscillation due to either one of valve stiction, model plant mismatch, oscillatory disturbance or combination of these has been presented. The oscillations due to valve stiction manifest as harmonics as well as discontinuities in the cross wavelet transform plots whereas oscillation due to model plant mismatch leaves distinct signatures in the phase information (arrows). If the oscillations are due to gain mismatch, no change is observed in the phase spectrum computed between controller output and plant output and controller output and model output. On the other hand, oscillation due to delay mismatch or oscillatory disturbance results in a directional change in the phase difference computed between controller output and plant output and controller output and model output.

Further study to distinguish between the oscillatory disturbance and delay mismatch as sources of the oscillation is currently underway. In a parallel study, it is found in our preliminary extensions of this work that the results demonstrated here can be applied to a more generalized problem of the diagnosis of poor control loop performance.
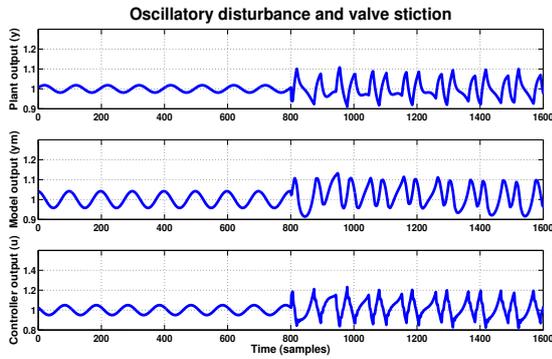
Fig. 5. *Time domain behavior of plant, model and controller outputs for the case oscillatory output and valve stiction as the sources of oscillation*
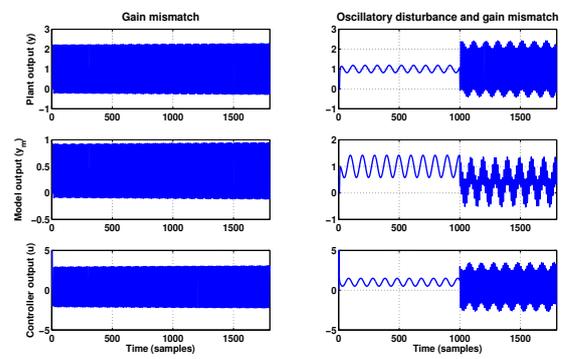

Fig. 7. *Time domain behavior of plant, model and controller outputs for the case of oscillatory disturbance and gain mismatch as the sources of oscillation.*
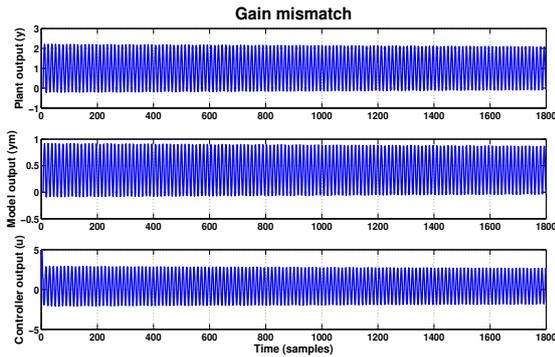
Fig. 6. *Time domain behavior of plant, model and controller outputs for the case of gain mismatch as the source of oscillation.*
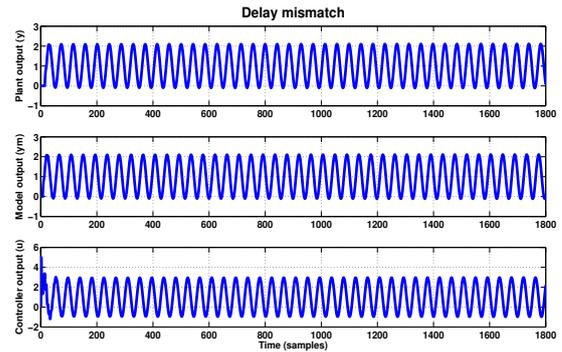
Fig. 8. *Time domain plots of plant, model and controller outputs for the case of delay mismatch as the source of oscillation.*

REFERENCES

[1] W. Bialkowski, "Dreams versus reality: a view from both sides of the gap," *Pulp and Paper Canada*, vol. 94, pp. 19–27, 1993.

[2] L. Desborough and R. Miller, "Increasing customer value of industrial control performance monitoring: Honeywell's experience," *Proc. AIChE Symp. Ser.*, vol. 98, pp. 153–186, 2002.

[3] D. Ender, "Process control performance: not as good as you think," *Control Engineering*, vol. 40, pp. 180–190, 1993.

[4] T. Hägglund, "A control-loop performance monitor," *Control Engineering Practice*, vol. 3, pp. 1543–1551, 1995.

[5] N. Thornhill and T. Hägglund, "Detection and diagnosis of oscillation in control loops," *Control Engineering Practice*, vol. 5, pp. 1343–1354, 1997.

[6] K. Forsman and A. Stattin, "A new criterion for detecting oscillations in control loops," in *CP8-3*. Karlsruhe, Germany.: European control conference, 1999.

[7] R. Rengaswamy, T. Hägglund, and V. Venkatasubramanian, "A qualitative shape analysis formalism for monitoring control loop performance," *Engineering Applications of Artificial Intelligence*, vol. 14, pp. 23–33, 2001.

[8] A. Tangirala, J. Kanodia, and S. Shah, "Non-negative matrix factorization for detection and diagnosis of plant wide oscillations," *Industrial and Engineering Chemistry Research*, vol. 46, pp. 801–817, 2007.

[9] N. Thornhill, B. Huang, and H. Zhang, "Detection of multiple oscillations in control loops," *Journal of Process Control*, vol. 13, pp. 91–100, 2003.

[10] A. Horch, "A simple method for the detection of stiction in control valves," *Control Engineering Practice*, vol. 7, pp. 1221–1231, 1999.

[11] A. Horch and A. Isaksson, "A method for detection of stiction in control valves," in *On-line-fault detection and supervision in the chemical process industry*. Lyon, France: IFAC Workshop, 1998, p. 4B.

[12] M. Choudhury, S. Shah, and N. Thornhill, "Detection and quantification of control valve stiction." Boston, USA: DYCOPS, 2004.

[13] D. Bloomfield, R. McAteer, B. Lites, and P. Judge, "Wavelet phase coherence analysis: Application to a quiet-sun magnetic element," *The Astrophysical Journal*, vol. 617, pp. 623–632, 2004.

[14] C. Torrence and G. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological Society*, vol. 79, pp. 61–78, 1998.
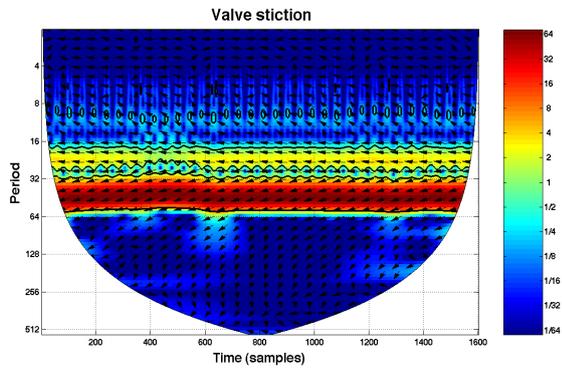
Fig. 9. *Cross wavelet transform plot between u and y_p when the oscillation is only due to valve stiction.*
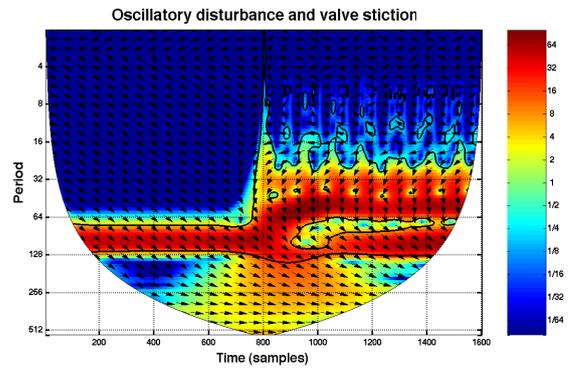
Fig. 12. *Cross wavelet transform plot between u and y_m when the oscillation is due to oscillatory disturbance and valve stiction.*
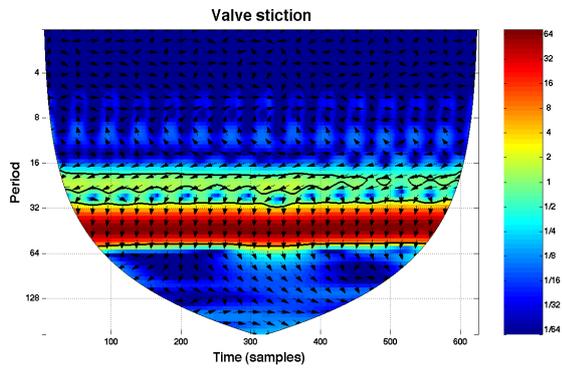
Fig. 10. *Cross wavelet transform plot between u and y_m when the oscillation is only due to valve stiction.*
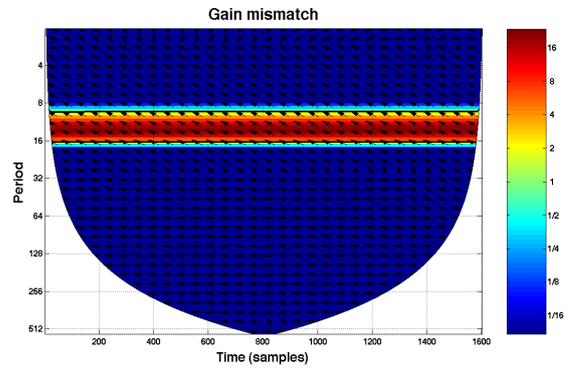
Fig. 13. *Cross wavelet transform plot between u and y_p when the oscillation is gain mismatch*
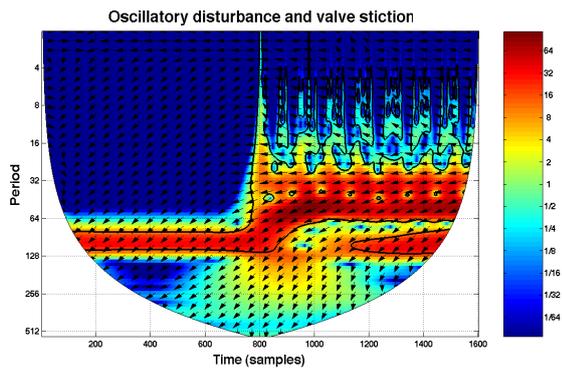
Fig. 11. *Cross wavelet transform plot between u and y_p when the oscillation is due to oscillatory disturbance and valve stiction.*
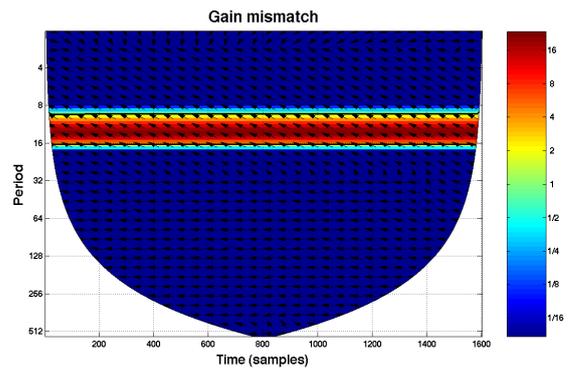
Fig. 14. *Cross wavelet transform plot between u and y_m when the oscillation is gain mismatch*

45
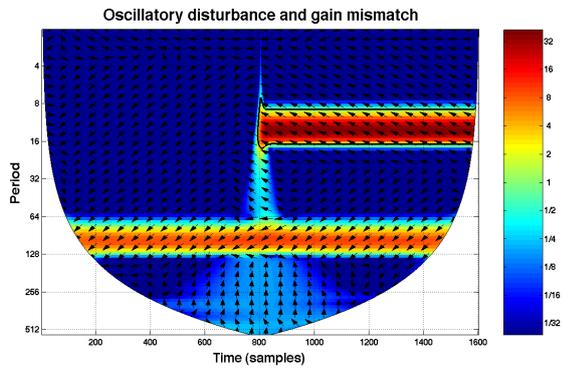
Fig. 15. *Cross wavelet transform plot between* u *and* y_p *when the oscillation is due to oscillatory disturbance and gain mismatch.*
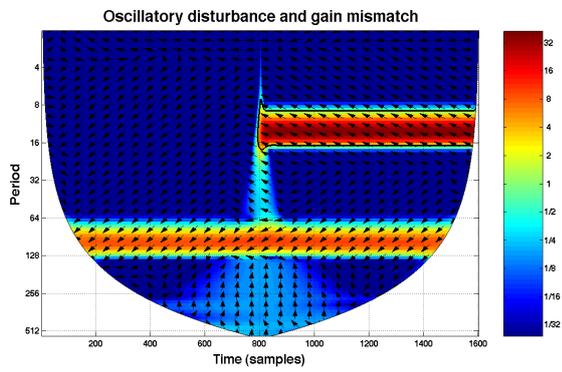
Fig. 16. *Cross wavelet transform plot between* u *and* y_m *when the oscillation is due to oscillatory disturbance and gain mismatch..*
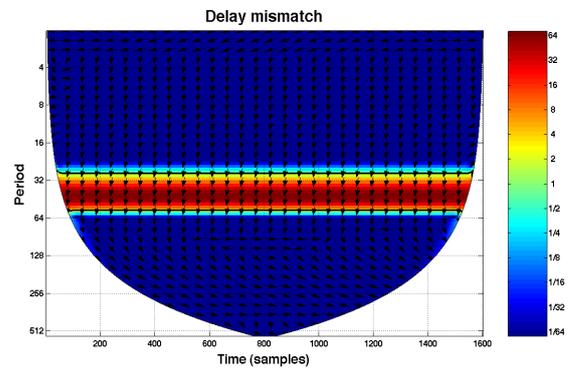
Fig. 18. *Cross wavelet transform plot between* u *and* y_m *when the oscillation is due delay mismatch..*
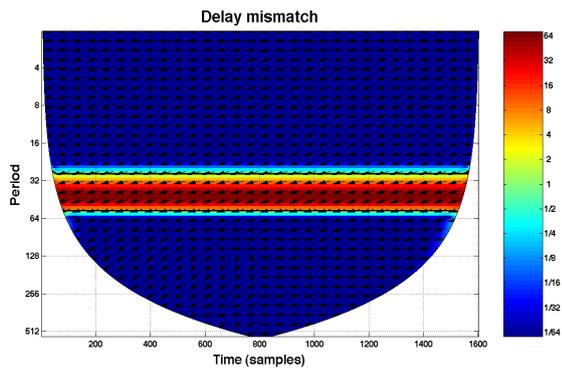
Fig. 17. *Cross wavelet transform plot between* u *and* y_p *when the oscillation is due delay mismatch..*

Availability Estimations for Utilities in the Process Industry

Anna Lindholm
Automatic Control, Lund

Hampus Carlsson
Perstorp AB

Charlotta Johnsson
Automatic Control, Lund

I. INTRODUCTION

An important performance measure of a plant is the plant-availability. The higher availability the better, since a high availability implies a possibility for a large production volume and thereby an increased profit for the company. One way of increasing the plant-availability is by eliminating, or minimizing the effect of disturbances. The cause of a disturbance can be personnel, material or equipment, where material includes both raw materials and utilities.

The aim of this work is to increase the plant-availability by decreasing the effects of plant-wide disturbances caused by utilities. The first step is to determine the set of utilities that can be present at an industrial site, what disturbances these utilities can suffer, and how frequent and safety-critical these disturbances are. A later step will be to determine the effects on the plant-availability, and ways to decrease or eliminate these effects.

The research is performed within the framework of PIC-LU (Process Industrial Centre at Lund University) supported by the Foundation of Strategic Research (SSF).

II. UTILITIES

Utilities, in opposite to raw materials, are materials that are used plant-wide and are crucial for plant operation but are not part of the final product. Common utilities are

- **Steam**: The steam net is commonly used to supply energy for distillation. Other uses are to supply energy for endothermic reactions and to heat a reactor at start-up. There could be several steam nets at the same site, for example one net with high pressure steam and one with low pressure steam.
- **Cooling water**: The cooling water system is used for cooling at exothermic reactions and in the condensing phase of distillation.
- **Electricity**: Electricity is needed in order for the instruments, e.g. pumps, to operate.
- **Water treatment**: A Water treatment utility is used for purification of process water, precipitation and ground water.
- **Combustion of tail gas**: A system for combustion of tail gas, such as a flare, is a safety device needed at unforeseen events.
- **Nitrogen**: Nitrogen is needed to maintain pressure in vessels.
- **Feed water**: Feed water is used in boilers to produce steam.
- **Instrument air**: Instrument air is needed for the pneumatic instruments to work.

- **Vacuum system**: Vacuum is used to lower the boiling point of a liquid to facilitate distillation and to remove gas produced in reactions.

A flowchart can be made for each of the utilities, showing how the utility flows through the areas of the production site. An example is showed in Figure 1.
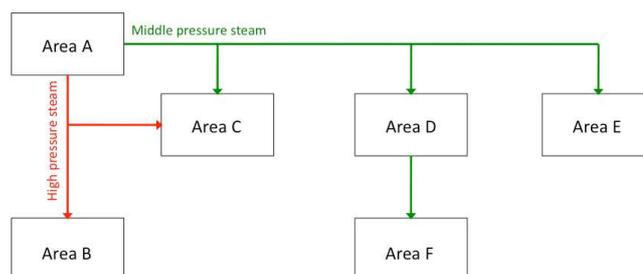
Fig. 1. Example of a utility flowchart for steam. The site contains 6 areas, A-F, and has two steam nets producing high and low pressure steam respectively.

A utility could suffer from different disturbances. For example, a steam net could suffer disturbances such as too high or too low steam pressure. One way of defining when a disturbance on a utility occurs is to set limits, such that if the parameter goes outside this limits, the disturbance will have economical or safety consequences.

III. AVAILABILITY

The availability of a production unit is according to ISO22400, draft 1, the fraction of the main usage time, which is the producing time of the unit, and the planned busy time, which is the time that the production unit is used for the execution of a manufacturing order. The availability of a utility can be estimated by taking the fraction of time when the utility does not suffer a disturbance over the total time. If measurements are available of the key parameters that define the disturbance, the availability can be computed directly from historical data.

IV. FUTURE WORK

When availabilities for the different utilities have been computed, the consequences of each disturbance must be evaluated. When both frequency and severity of all disturbances on utilities are known, focus on handling the disturbance with highest severity×frequency-factor to improve availability of the entire production plant as much as possible.

Detection and Isolation of Oscillations Using the Dynamic Causal Digraph Method

Tikkala, Vesa-Matti*; Zakharov, Alexey; Jämsä-Jounela, Sirkka-Liisa

Aalto University School of Science and Technology

Department of Biotechnology and Chemical Technology

Abstract

This paper proposes a modification to the dynamic causal digraph (DCDG) method in order to address the detection and isolation of oscillations in a process. The proposed detection method takes advantage of the properties of residual signals generated by the DCDG method by studying their zero-crossings. The method is tested in an application to a board making process and the results are presented and discussed.

1 Introduction

Demands to keep industrial processes running efficiently with a high rate of utilization are increasing constantly due to the tightening global competition. Since, the modern industrial processes are complex and large-scale, operator-based monitoring cannot guarantee early enough detection and reliable diagnosis of the faults and abnormalities. Therefore, the detection and diagnosis of different abnormal and faulty conditions in the processes have become increasingly important.

Common problems causing inefficient operation and production losses in the process industry are oscillations. Oscillatory disturbances readily propagate in the process and cause extensive variation in the process variables. The oscillations are usually originated under feedback control, and they may have various causes which have been categorized by Thornhill & Horch (2007) into non-linear and linear causes. Non-linear causes include for example extensive static friction in the control valves, on-off or split range control, sensor faults, process non-linearities and hydrodynamic instabilities. The most common linear causes are poor controller tuning, controller interaction and structural problems involving process recycles (Thornhill & Horch, 2007). According to Choudhury *et al.* (2008), valve stiction is, however, the most common cause of these oscillations in control loops.

The detection and diagnosis of oscillations have been previously addressed by data-based methods which study, for example, the properties of controller error signals (Thornhill & Hagglund, 1997; Forsman & Stattin, 1999), the spectral properties (Thornhill *et al.*, 2003) or the nonlinearity of the measurement signals (Choudhury *et al.*, 2004; Thornhill, 2005). Also, a variety of multivariate methods, such as principal component analysis (Thornhill *et al.*, 2002) and non-negative matrix factorization (Tangirala *et al.*, 2007), have been applied to solve this diagnosis task. A recent trend has been to introduce process information into the diagnosis of plant-wide oscillations. Applications in which the process connectivity information has been integrated into data-based analyses to check hypotheses on the fault origin, have been presented (Yim *et al.*, 2007; Jiang *et al.*, 2009).

This paper aims at further development of the dynamic causal digraph (DCDG) method by addressing the detection and isolation of plant-wide oscillations. A detection algorithm, which is able to deal with oscillatory residuals, is proposed and integrated into the DCDG method. In this paper, the modified DCDG method is used to detect and isolate low-frequency oscillations caused by a valve stiction fault in a board machine process.

The paper is organized as follows. In Section 2, the dynamic causal digraph method and the new detection algorithm are introduced. The process and the test environment are described in Section 3. The results of the testing are presented in Section 4 followed by the conclusions in Section 5.

2 Enhanced DCDG Method for the Detection and Isolation of Oscillations

The dynamic causal digraph method employs the process knowledge formalized as a causal digraph model in order to perform the ordinary fault diagnosis tasks as presented by Cheng *et al.* (2010); Cheng (2009). In the enhanced DCDG, the detection of faults is performed using the proposed method which observes the zero-crossings in the residuals generated by a comparison

---

*Corresponding author: `vesa-matti.tikkala@tkk.fi`, Aalto university, PL 16100, FI-00076 Aalto

of cause-effect models and the process measurements. Next, the isolation is carried out by applying a set of inference rules to the residuals in order to extract the fault propagation path. Finally the arcs in the digraph that explain the faulty behavior are identified. The enhanced DCDG method is described in more detail in the following.

2.1 Fault detection

Fault detection is performed in two steps: residual generation and fault detection in the residuals using the modified cumulative sum (CUSUM) algorithm.

2.1.1 Residual generation with dynamic models

The dynamic causal digraph produces two kinds of residual to be used in fault detection and isolation: global (GR) and local residuals (LR). The global residual is produced from the difference between the measurement and the global propagation value:

$$GR(Y) = Y(k) - \hat{Y}(k), \tag{1}$$

where $Y(k)$ is the measurement and $\hat{Y}(k)$ is the global propagation value obtained by

$$\hat{Y}(k) = f_Y\left(\hat{U}(k-1), \hat{U}(k-2), \ldots\right), \tag{2}$$

where f_Y is a discrete-time model describing the cause-effect relationship from n predecessor nodes U_i to node Y. $\hat{U}(k-\tau) = \{\hat{u}_1(k-\tau), \ldots, \hat{u}_n(k-\tau)\}$ are the lagged global propagation values from the predecessors with time lags $\tau = 1, 2, \ldots$ depending on the system order.

The local residuals are subcategorized into three types: individual local residuals (ILR), multiple local residuals (MLR) and total local residuals (TLR) (Montmain & Gentil, 2000).

The individual local residual is produced by taking the difference between the measurement and the local propagation value with only one measured input, while all the others are propagation values from the parent nodes:

$$ILR_Y^m = Y - \bar{Y}, \tag{3}$$
$$\bar{Y}(k) = f_Y\left(\bar{U}(m, k-1), \bar{U}(m, k-2), \ldots\right),$$

where

$$\bar{U}(m, k-\tau) = \left\{ \bar{u}_i(k-\tau) \,\middle|\, \right.$$
$$\left. \bar{u}_i(k-\tau) = \begin{cases} \hat{u}_i(k-\tau), i \neq m \\ u_i(k-\tau), i = m \end{cases}, 1 \leq i \leq n \right\}, \tag{4}$$

$\hat{u}_i(k)$ is the lagged global propagated value from the predecessors, and $u_i(k-\tau)$ is the measurement for the i-th parent node.

Similarly, the $MLR_Y^{P_Y^l}$ is produced as

$$MLR_Y^{P_Y^l} = Y - \bar{Y}, \tag{5}$$
$$\bar{Y}(k) = f_Y\left(\bar{U}(P_Y^l, k-1), \bar{U}(P_Y^l, k-2), \ldots\right),$$

where

$$\bar{U}(P_Y^l, k-\tau) = \left\{ \bar{u}_i(k-\tau) \,\middle|\, \right.$$
$$\left. \bar{u}_i(k-\tau) = \begin{cases} \hat{u}_i(k-\tau), i \notin P_Y^l \\ u_i(k-\tau), i \in P_Y^l \end{cases}, 1 \leq i \leq n \right\}, \tag{6}$$

P_Y^l is the set of indices of the predecessors which use the measurement as an input. The $TLR(Y)$ is produced with $P_Y^l = P_Y$, where P_Y is the set of indices of all the predecessors of Y.

The residual generation scheme follows the DCDG method developed in (Montmain & Gentil, 2000).

2.1.2 Fault detection using the modified CUSUM method

The proposed detection method utilizes the cumulative sum (CUSUM) method presented by Hinkley (1971), by applying it to the detection of a change in the mean and variance of the zero-crossings in the residual signals. The CUSUM algorithm is defined for a positive change as follows:

$$U_0 = 0$$
$$U_n = \sum_{k=1}^{n} d(k) - \mu_0 - \frac{\beta}{2}, \tag{7}$$
$$m_n = \min_{0 \leq k \leq n} U_k,$$

where β is a user-specified minimum detectable change, $d(k)$ the observed signal with nominal mean value equal to μ_0. Whenever $U_n - m_n > \lambda$, a change is detected, where λ is a design parameter, usually tuned according to the requirements for the false alarm and missed alarm rates.

The signal observed by the CUSUM algorithm, called the detection signal, is defined as follows

$$d(k) = \max\left\{ \bar{\Delta}t(k), \frac{\sigma_{\Delta t}^2(k)}{\bar{\Delta}t(k)} \right\}, \tag{8}$$

where $\max\{\cdot\}$-operator takes the maximum of its arguments, $\bar{\Delta}t(k)$ and $\sigma_{\Delta t}^2(k)$ are the mean and the variance of the time between consecutive zero-crossings in a residual, respectively. Both $\bar{\Delta}t$ and $\sigma_{\Delta t}^2$ are calculated in a moving window of length l: $[e(k-l), e(k)]$, where $e(k)$ is the residual.

In normal operation, when the residuals are assumed to be zero-mean Gaussian noise, $\bar{\Delta}t, \sigma_{\Delta t}^2(k) \approx 2$, since

Table 1: Fault isolation rules of the dynamic causal digraph

$CU(GR(Y))$	$CU(TLR(Y))$	$CU(ILR_Y(m))$	$CU(ILR_Y(i))$	$CU(MLR_Y(P_1))$	$CU(MLR_Y(P_2))$	Decision
0	0	0	0	0	0	No fault
1/-1	0	0*	1/-1*	0*	1/-1*	Fault propagates from the parent node m
1/-1	0	1/-1**	1/-1**	1/-1**	0**	Fault propagates from the nodes with subscript P_2
1/-1	1/-1	1/-1	1/-1	1/-1	1/-1	Local fault on variable Y

* $\forall i \neq m, i \in P_Y, m \in P_1, m \notin P_2, P_Y$ is the set of subscripts of parent nodes of the node Y.

** $\forall i, m, i \in P_Y, m \in P_Y, \forall P_1, P_2 \subseteq P_Y$.

the probability of $e(t)$ having a different sign than $e(t-1)$ is 0.5 for all t. Therefore, the nominal mean value of the observed signal in (7) can be set as $\mu_0 = 2$ and β and λ are then tuned to obtain robust detection with minimal false alarms. The window length l must be selected to be larger than one half of the expected period in the residual.

2.2 Fault isolation

2.2.1 Isolation of the fault propagation path

Fault isolation is performed recursively for the detected nodes by using a set of rules. These isolation rules, developed by Montmain & Gentil (2000), are converted into a table for the convenience of implementation, as shown in Table 1. After the isolation the nature of the fault is determined by using rules in Table 2.

Table 2: Fault nature rules of the dynamic causal digraph

$CU(GR(X))^*$	$CU(TLR(X))$	Fault nature
1/-1	1/-1	Local fault for that child node
1/-1	0	Process fault for the faulty node
0	1/-1	Measurement fault for the faulty node

* X is the subscript of any child node of the node Y.

2.2.2 Isolation of the faulty process component

In the case of a process fault, in addition to locating the fault on the variables (nodes), locating it on the arcs is also desirable. However, the MISO structure of the digraph causes problems by generating multiple possible results as $2^n - 1, n \geq 1$, where n is the number of input arcs of the fault origin node(s).

In order to decrease the number of possible results, an inference mechanism between the arcs proposed in (Cheng *et al.*, 2008) is used. The inference mechanism is based on an inter-arc knowledge matrix $\mathbf{M}$ defined for node U as follows

$$\mathbf{M}_U(i,j) = \begin{cases} 1, \text{ if inconsistency in arc } \langle U, i \rangle \\ \quad \text{causes inconsistency to } \langle U, j \rangle \\ 0, \text{ otherwise,} \end{cases} \quad (9)$$

where i and j refer to the matrix rows and columns, respectively.

Next, each set of suspected arcs is tested in order to determine whether the fault may be caused exactly by the current set of arcs. In order to do it the matrix $\mathbf{M}$ is multiplied with a vector representing the suspected arc set, which is defined as follows

$$\mathbf{sv}(i) = \begin{cases} 1, \text{ if } ARC(\mathbf{M}, i) \in S, \ 1 \leq i \leq N_a \\ 0, \text{ otherwise,} \end{cases} \quad (10)$$

where $ARC(\mathbf{M}, i)$ gives the arc corresponding to the ith row in the matrix $\mathbf{M}$. S is the set of suspected arcs. If the number of non-zero elements of $\mathbf{sv}$ have changed, the current suspected set of arcs must be excluded.

3 Description of the Process and the Valve Stiction Faults

This test focuses on the stock preparation of the board machine at Stora Enso's mills in Imatra, Finland. The simulation tests are run on a board machine simulator model in the APROS simulation environment.

3.1 The Board Machine Process

The board making process begins with the preparation of raw materials in the stock preparation section, as shown in the flowsheet in Figure 1. Different types of pulp are refined and blended according to a specific recipe in order to achieve the desired composition and properties for the board grade to be produced. The consistency of the stock is controlled with dilution water.

The blended stock passes from the stock preparation to the short circulation. First, the stock is diluted in the machine chest to the correct consistency for web formation. The diluted stock is then pumped with a fan pump, which is used to control the basis weight of the board, to cleaning and screening. Next, the stock passes to the head box, from where it is sprayed onto the wire in order to form a solid board web.

The excess water is first drained through the wire and later by pressing the board web between rollers in the
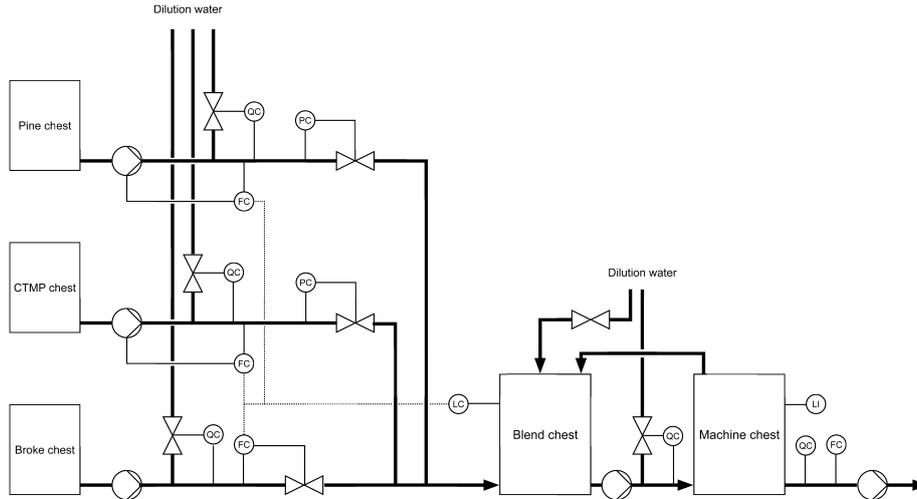
Figure 1: Flowsheet of the stock preparation of the board machine process.

Table 3: Variables of the causal digraph model for the stock preparation of the board machine.

Var.	Description	Type	Unit
vb	valve opening for the broke line	A	-
fb	mass flow of the broke	M	kg/s
vbd	dilution water valve opening for the broke line	A	-
fbd	dilution water flow for the broke line	E	kg/s
cb	broke consistency	M	%
rp	pine pump rotation speed	A	%
fp	mass flow of the pine stock	M	kg/s
vpd	dilution water valve opening for the pine line	A	-
cp	pine consistency	M	%
rc	CTMP pump rotation speed	A	%
fc	mass flow of the CTMP	M	kg/s
vcd	dilution water valve opening for the CTMP line	A	-
cc	CTMP consistency	M	%
$vmcd$	dilution water valve opening for the machine chest	A	-
cmc	consistency before the machine chest	M	%
pp	pressure before the pine valve	M	kg/s
pc	pressure before the CTMP valve	M	kg/s
ct	consistency of the machine chest	M	%

A: Actuator signal, M: Measurement signal

press section. The remaining water is evaporated off in the drying section using steam-heated drying rolls.

The variables used in the causal digraph model of the stock preparation are listed in Table 3.

3.2 Valve Stiction Faults

A control valve is the most common final control element used in the process industry (Choudhury *et al.*, 2008). Therefore, the diagnosis of faults in valves is of great importance. Stiction, short for static friction, is a problem in control valves since it can cause significant disturbances in the process variables. A valve suffering from excessive stiction sticks when the control signal, for example, changes the direction and does not move until the force required to move the valve shaft exceeds

a certain limit. When the valve starts to move, it jumps and then follows the control signal before it sticks again. A sticking valve is likely to cause oscillations when it is involved in a control loop.

The stiction in valves has been modelled and studied e.g. by Stenman *et al.* (2003) and Choudhury *et al.* (2005). This paper considers a stiction fault in a pressure control valve which causes the control loop to oscillate and disturbs the operation of the plant.

4 Testing and Results

4.1 Simulation Environment and Fault Simulation

The Imatra board machine model was developed by Stora Enso and VTT in the APROS environment. It was originally constructed on the basis of modeling and simulation studies carried out during 1998–2002 for Stora Enso's Imatra mills. It has been previously used for grade change simulations and in studies reported by Lappalainen *et al.* (2003).

A valve stiction fault was simulated in the stock preparation part of the board machine using the APROS board machine model. The faulty valve is located in the CTMP line and is used to control the feeding pressure p_c of the blend chest. The two-parameter data-driven valve stiction model proposed by Choudhury *et al.* (2005) was implemented in the APROS simulation software for the simulation. The deadband and slip-jump parameters of the stiction model were set to $S = 0.06$ and $J = 0.06$ respectively. The fault was evoked by a step change to the setpoint of p_c.

The fault occurring at $t = 200$ causes an oscillation
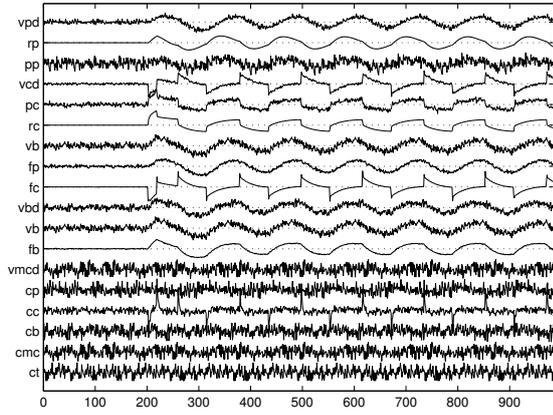
Figure 2: Normalized process variables during the fault simulation.

Figure 3: Global residuals $GR(f_c)$ and $GR(c_c)$, detection signals and the detection results

Figure 4: Local residual $TLR(c_c)$, detection signal and the detection results

with a period of approximately 120 samples, which affects most of the variables in the stock preparation. Figure 2 shows the measured variables during the fault simulation and it demonstrates clearly the effect of the fault in the process.

4.2 Fault Detection and Isolation Results

First, the global residuals for all variables were produced by comparing the measured values of the variables and the estimates generated using the dynamic causal digraph model. Then, the detection signals were produced by calculating the mean and the variance of zero-crossings in the global residuals. The proposed detection method was applied to analyse the residuals in order to detect the faulty nodes. The parameters of the modified CUSUM method were set to the following: $\beta = 10$, $\lambda = 4$ and $l = 200$. The global residuals, detection signals and the detection results for variables f_c and c_c are presented in Figure 3. The fault is detected in both signals $GR(f_c)$ and $GR(c_c)$. The change in the detection signal $d_{f_c}(k)$ was detected for the first time at $k = 203$, three time instants after the fault occurred. However, the detection result is not reliable until $k = 270$. The global residual of c_c is detected later at $k = 315$.

Local residuals were generated in order to carry out the inference to isolate the origin of the fault. The local residual, the detection signal and the detection results of c_c are shown in Figure 4. The detection signal changes slightly after the fault occurrence, but no detection is however made.

The performance of the proposed detection method is satisfactory. The faults are detected with a reasonable delay and no false alarms are generated. Detection in variable c_{mc}, based on the structure of the process and the forecast of the fault propagation, was also expected.
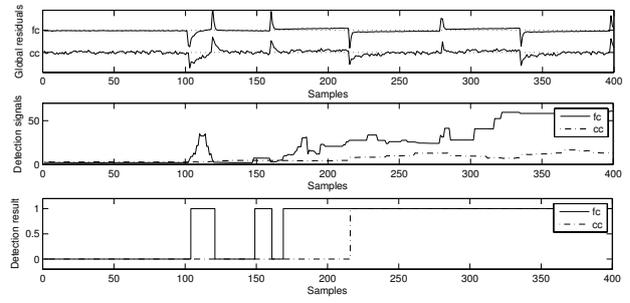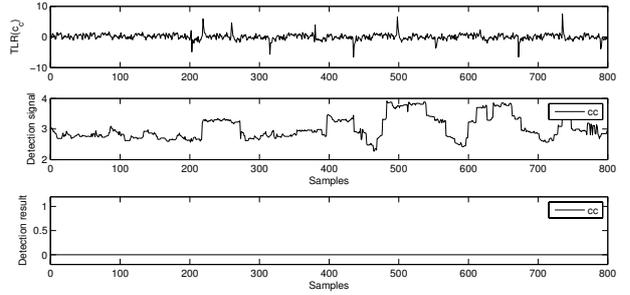
However, based on the simulation studies, it was found out that the effect of the fault attenuates and therefore the change in the global residual of c_{mc} becomes undetectable.

The fault isolation rules presented in Table 1 were applied in order to extract the fault propagation path and the fault origin. The fault origin was located at the node f_c. The nature of the detected fault is diagnosed as a process fault according to the rules presented in Table 2.

Since the fault was a process fault, the structure of the digraph model gave three possible sources for the fault: v_{cd}, p_c and r_c resulting $(3^2 - 1) = 8$ possible sets of arcs explaining the faulty behaviour . The arc sets were analysed using the process knowledge matrix $\mathbf{M}$. However the number of the suspected sets could not be reduced in this case, since the input arcs to the node f_c are independent. If one input arc is faulty, it will not cause inconsistency in other input arcs.

5 Conclusions

A method for detecting oscillatory residual signals was presented in this paper. The method was integrated into the DCDG fault diagnosis method and tested in an application to a board making process.

The proposed method enables the detection and isolation of low-frequency oscillations caused by valve stiction faults in the process by exploiting the statistical properties of the residual signals. The results show that the proposed detection method is able to detect the fault successfully and to provide the information required for fault isolation.

The work presented in this paper represents the first step in addressing the detection and isolation of faults causing oscillatory behaviour in a process using the DCDG method. In future, the aim is to generalize the diagnosis methodology by developing new detection methods that are able to cover a wider range of faults occurring in industrial processes.

References

Cheng, H., 2009. *Causal Digraph Reasoning for Fault Diagnosis in Paper Making Applications.* Ph.D. thesis, Helsinki University of Technology, Helsinki, Finland.

Cheng, H., Nikus, M. & Jämsä-Jounela, S.-L., 2008. Fault diagnosis of the paper machine short circulation process using novel dynamic causal digraph reasoning. *Journal of Process Control,* vol. 18(7–8), pp. 676–691.

Cheng, H., Tikkala, V.-M., Zakharov, A., Myller, T. & Jamsa-Jounela, S.-L., 2010. Application of the enhanced dynamic causal digraph method on a three-layer board machine. *IEEE Transactions on Control System Technology,* (in press).

Choudhury, M. A. A. S., Shah, S. L. & Thornhill, N. F., 2004. Diagnosis of poor control loop performance using higher order statistics. *Automatica,* vol. 40, pp. 1719–1728.

Choudhury, M. A. A. S., Shah, S. L. & Thornhill, N. F., 2008. *Diagnosis of Process Nonlinearities and Valve Stiction.* 1st ed., Springer.

Choudhury, M. A. A. S., Thornhill, N. F. & Shah, S. L., 2005. Modelling valve stiction. *Control Engineering Practice,* vol. 13(5), pp. 641–658.

Forsman, K. & Stattin, A., 1999. A new criterion for detecting oscillations in control loops. In: *European Control Conferenre ECC'99,* p. 878, karlsruhe, Germany, August 31 - September 3, 1999.

Hinkley, D. V., 1971. Inference about the change-point from cumulative sum tests. *Biometrika,* vol. 58, pp. 509–623.

Jiang, H., Patwardhan, R. & Shah, S. L., 2009. Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix. *Journal of Process Control,* vol. 19(8), pp. 1347–1354.

Lappalainen, J., Vehviläinen, O., Juslin, K., Myller, T. & Tuuri, S., 2003. Enhancing grade changes using dynamic simulation. *TAPPI Journal,* vol. 2(12).

Montmain, J. & Gentil, S., 2000. Dynamic causal model diagnostic reasoning for online technical process supervision. *Automatica,* vol. 36, pp. 1137–1152.

Stenman, A., Gustafsson, F. & Forsman, K., 2003. A segmentation-based method for detection of stiction in control valves. *International Journal of Adaptive Control and Signal Processing,* vol. 17(7–9), pp. 625–634.

Tangirala, A., Kanoda, J. & Shah, S. L., 2007. Non-negative matrix factorization for detection and diagnosis of plantwide oscillations. *Industrial and Engineering Chemistry Research,* vol. 46, pp. 801–817.

Thornhill, N. F., 2005. Finding the source of nonlinearity in a process with plant-wide oscillation. *IEEE Transactions on Control Systems Technology,* vol. 13(3), pp. 434–443.

Thornhill, N. F. & Hagglund, T., 1997. Detection and diagnosis of oscillations in control loops. *Control Engineering Practice,* vol. 5(10), pp. 1343–1354.

Thornhill, N. F. & Horch, A., 2007. Advances and new directions in plant-wide disturbance detection and diagnosis. *Control Engineering Practice,* vol. 15, pp. 1196–1206.

Thornhill, N. F., Huang, B. & Zhang, H., 2003. Detection of multiple oscillations in control loops. *Journal of Process Control,* vol. 13, pp. 91–100.

Thornhill, N. F., Shah, S. L., Huang, B. & Vishnubhotla, A., 2002. Spectral principal component analysis of dynamic process data. *Control Engineering Practice,* vol. 10.

Yim, S. Y., Ananthakumar, H. G., Benabbas, L., Horch, A., Drath, R. & Thornhill, N. F., 2007. Using process topology in plant-wide control loop performance assessment. *Computers and Chemical Engineering,* vol. 31, pp. 86–99.

Optimal Control of the Oil Reservoir Water-flooding process

Eka Suwartadi, NTNU; Stein Krogstad, SINTEF ICT; Bjarne Foss, NTNU Norway

Abstract

In the second phase of oil recovery water flooding is a common way to sweep remaining oil in the reservoir. The process can be posed as a nonlinear optimization problem. This talk will address optimization of water flooding in the context of gradient based optimization. The gradient is computed using the adjoint method. In the optimization problem there will be constraints on both control inputs and state variables. The latter constraints are notoriously difficult to handle since they affect the efficiency of the adjoint method adversely. We propose some methods to mitigate this. Further, we present a second order adjoint method to avoid numerical problems which may arise in quasi Newton methods like the BFGS method.

State-Constrained Control Based on Linearization of the Hamilton-Jacobi-Bellman Equation

Torsten Wik, Per Rutqvist and Claes Breitholtz

Abstract— **For continuous time state constrained stochastic control problems a method based on optimization is presented. The method applies to systems where the control signal and the disturbance both enters affinely, and it has one main tuning paramater, which determines the control activity. If the disturbance covariance is unknown, it can also be used as a tuning parameter (matrix) to adjust the control directions in an intuitive way. Optimal control problems for this type of systems result in Hamilton Jacobi Bellman (HJB) equations that are problematic to solve because of nonlinearity and infinite boundary conditions. However, by applying a logarithmic transformation we show how and when the HJB equation can be transformed into a linear eigenvalue problem for which there are sometimes analytical solutions and if not, it can readily be solved with standard numerical methods. Sufficient and necessary conditions for when the method can be applied are derived, and their physical interpretation is discussed. A MIMO buffer control problem is used as an illustration.**

I. INTRODUCTION

Consider a dynamical system described by

$$\dot{x} = f(x) + B(x)u + G(x)w, \tag{1}$$

where $x \in \Omega \subset \mathbb{R}^n$ is the system state, $u \in \mathbb{R}^m$ is the control signal, $f : \Omega \to \mathbb{R}^n$, $B : \Omega \to \mathbb{R}^{n \times m}$ and $G : \Omega \to \mathbb{R}^{n \times n_w}$ are functions that describe the system dynamics, and $w = \dot{\omega} \in \mathbb{R}^{n_w}$ is a Gaussian white noise (where ω is a Wiener process) having the covariance matrix W. The boundary $\partial\Omega$ of the state space Ω defines the state constraints, i.e. the system must be controlled such that x never leaves Ω.

The idea now is to formulate the control problem as an optimal control problem for which a control policy $u(t, x)$ can be determined analytically or by straight forward application of numerical methods. In general, the optimal control problem for a system on the form (1) requires the solution of a nonlinear partial differential equation (PDE), the so-called Hamilton-Jacobi Bellman equation, that is both nonlinear and will have infinite boundary conditions that are difficult to handle numerically. However, by applying a logarithmic transformation we show how and when the PDE can be transformed into a linear eigenvalue problem for which there are sometimes analytical solutions and if not, it can readily be solved with standard numerical methods.

Rutquist et al. [7] used this logarithmic transformation to linearize the HJB equation in the scalar one dimensional case. Concurrently, Itami [2] also studied the one-dimensional case and used the transformation to make

T. Wik and C. Breitholtz are with Department of Signals and Systems, Chalmers University of Technology, SE 412 96 Göteborg, Sweden `tw@chalmers.se`, `claesbr@chalmers.se`

P. Rutquist is with Tomlab Optimization AB, Västerås, Sweden `per.rutquist@gmail.com`

a coupling between quantum mechanics (the Schrödinger equation) and the Hamilton equation, which was earlier pointed out also by Rosenbrock [4].

In a later study [5] the use of the transformation was further developed to several dimensions for the special case when the disturbance enters the system in the same way as the control input, i.e. $G = B$. Here, we generalize the method to the case when $G \neq B$, give necessary and sufficient conditions for when this linearization can be applied, and analyze the implications of these conditions. When linearization can be applied, it gives an intuitive control method for state constrained systems with only one main tuning parameter that determines the control aggression.

Two examples where the results are applied are presented here. The first one shows how to deal with disturbances that are not purely white. The other is a buffer example treating the pumping of wastewater to a wastewater treatment plant in Göteborg, Sweden.

II. THE OPTIMAL CONTROL PROBLEM

We define the problem as to find a feedback control policy $u(t, x)$ that minimizes

$$V(x(t), t) =$$
$$\mathrm{E}\left\{ \int_t^{t_f} \left(l(x(\tau)) + u^T(\tau)Qu(\tau) \right) \, d\tau + V_f(x(t_f))|x(t) \right\}, \tag{2}$$

where $t \in \mathbb{R}$ is the current time, $t_f > t$ is the final time, $l : \Omega \to \mathbb{R}^+$ describes the (time independent) cost (non-singular on Ω) associated with the state, $V_f : \Omega \to \mathbb{R}$ is the final cost, and $u^T Q u$ defines the cost of the control signal.

The stochastic HJB equation for the minimization of (2) can be formulated as

$$-\frac{\partial V}{\partial t} = \min_u \{ l + (\nabla V)(f + Bu) + u^T Q u$$
$$+ \frac{1}{2}\mathrm{tr}\left[(\nabla^T \nabla V)GWG^T \right] \}, \tag{3}$$

where $V : [t, t_f] \times \Omega \to \mathbb{R}$ is the so-called cost-to-go function (see for instance [1] for details on the derivation of this equation) and the gradient ∇V is defined as a row vector.

Minimization of this quadratic expression with respect to u gives the optimal control input

$$u = -\frac{1}{2}Q^{-1}B^T(\nabla V)^T. \tag{4}$$

Inserting the optimal u into (3) gives

$$-\frac{\partial V}{\partial t} = l - \frac{1}{4}(\nabla V)BQ^{-1}B^T(\nabla V)^T + (\nabla V)f \\ + \frac{1}{2}\text{tr}\big[(\nabla^T\nabla V)GWG^T\big] \tag{5}$$

with infinite boundary conditions on $\partial\Omega$ due to the state constraints. The solution to this equation inserted into (4) gives the optimal control policy. The problem though, is that this PDE is not readily solved because it is nonlinear in V and the infinite boundary conditions are difficult to handle numerically.

III. Linearization

Now, applying the transformation

$$V = -2\kappa\log Z, \tag{6}$$

where κ is an arbitrary real constant, gives

$$\nabla V = -\frac{2\kappa}{Z}\nabla Z$$

$$\nabla^T\nabla V = \frac{2\kappa}{Z^2}(\nabla Z)^T\nabla Z - \frac{2\kappa}{Z}(\nabla^T\nabla Z).$$

Equation (5) is then transformed into

$$2\kappa\frac{\partial Z}{\partial t} = zl - \frac{\kappa^2}{Z}(\nabla Z)BQ^{-1}B^T(\nabla Z)^T - 2\kappa(\nabla Z)f \\ + \frac{\kappa}{Z}\text{tr}\big[(\nabla Z)^T(\nabla Z)GWG^T\big] \\ - 2\kappa\text{tr}\big[(\nabla^T\nabla Z)GWG^T\big]. \tag{7}$$

If two matrices A and B have matching dimensions $\text{tr}[AB] = \text{tr}[BA] = \text{tr}\big[A^TB^T\big]$, and the trace of a scalar equals the scalar itself. Therefore we have that

$$(\nabla Z)BQ^{-1}B^T(\nabla Z)^T = \text{tr}\big[(\nabla Z)^T(\nabla Z)BQ^{-1}B^T\big]$$
$$\text{tr}\big[(\nabla Z)^T(\nabla Z)GWG^T\big] = (\nabla Z)GWG^T(\nabla Z)^T,$$

which gives

$$\frac{\partial Z}{\partial t} = \frac{l}{2\kappa}Z + \frac{1}{2Z}(\nabla Z)(GWG^T - \kappa BQ^{-1}B^T)(\nabla Z)^T \\ - (\nabla Z)f - \text{tr}\big[(\nabla^T\nabla Z)GWG^T\big]. \tag{8}$$

Now, if

$$GWG^T = \kappa BQ^{-1}B^T \tag{9}$$

the transformed HJB equation becomes a linear PDE:

$$\frac{\partial Z}{\partial t} = \frac{l}{2\kappa} - (\nabla Z)f - \text{tr}\big[(\nabla^T\nabla Z)GWG^T\big] \tag{10}$$

with boundary conditions

$$Z = 0, \quad x \in \partial\Omega \tag{11}$$

and

$$Z_f \triangleq Z(t_f, \mathbf{x}) = \exp\left(\frac{-V_f}{2\kappa}\right). \tag{12}$$

For the finite time case this PDE can be solved using variable separation, i.e. $Z(t,x) = T(t)\phi(x)$ [6]. The time dependent part $T(t)$ has an analytical solution and the state dependent part $\phi(x)$ becomes a (partial) linear differential equation which may have an analytical solution or be readily solved numerically using an eigenvalue solver.

For the stationary problem $t_f \to \infty$ and the integral in (2) will not converge. However, since the process is ergodic the expected cost per unit time (λ) will eventually be the same everywhere in Ω, i.e.

$$\lambda = -\frac{\partial V}{\partial t} = \frac{2\kappa}{Z}\frac{\partial Z}{\partial t}, \tag{13}$$

which gives us the *linear eigenvalue problem*

$$\lambda Z = lZ - 2\kappa(\nabla Z)f - \kappa\text{tr}\big[(\nabla^T\nabla)GWG^T\big] \\ Z = 0 \qquad \text{on } \partial\Omega, \tag{14}$$

where the solution with the least λ is sought since it corresponds to the lowest cost.

Clearly, the problem of determining the optimal solution is greatly simplified by the transformation (6) if (9) also holds so that the problem becomes linear as well. In the following we will state the necessary and sufficient conditions for the linearization to be applicable, and an analysis of the implications of (9) on the optimal control.

Preliminaries

A singular value decomposition (SVD) of B gives

$$B = U_1\Lambda_B U_2^T.$$

where U_1 ($n \times n$) and U_2 ($m \times m$) are orthogonal matrices, and the $r_B = \text{rank}(B)$ singular values on the upper left diagonal of Λ_B are the square roots of the nonzero eigenvalues of both BB^T and B^TB. Then the columns of U_1 and U_2 give the orthonormal bases for all four fundamental subspaces [8]:

$$E_{R_B} = \text{the first } r_B \text{ columns of } U_1,$$
$$E_{N_B^T} = \text{the last } n - r_B \text{ columns of } U_1,$$
$$E_{R_B^T} = \text{the first } r_B \text{ columns of } U_2,$$
$$E_{N_B} = \text{the last } m - r_B \text{ columns of } U_2,$$

where E_{R_B} has columns that are the base vectors for the column space $\mathcal{R}(B)$, $E_{N_B^T}$ for the left nullspace $\mathcal{N}(B^T)$, $E_{R_B^T}$ for the row space $\mathcal{R}(B^T)$, and E_{N_B} for the nullspace $\mathcal{N}(B)$ of B.

For a symmetric positive definite matrix, such as the covariance matrix W, SVD is identical to diagonalization with orthogonal eigenvectors, i.e.

$$W = U_W\Lambda_W U_W^T,$$

where Λ_W is a ($n_w \times n_w$) diagonal matrix with the real positive eigenvalues $\lambda_{W,i}$ of W on the diagonal. We may then define the square root of such a matrix as

$$W^{1/2} = U_W\Lambda_W^{1/2}U_W^T,$$

where $\Lambda_W^{1/2} = \text{diag}(\sqrt{\lambda_{W,1}}, \ldots \sqrt{\lambda_{W,n_w}})$. Clearly, $W^{1/2}$ is also symmetric and positive definite, and $W^{1/2}W^{1/2} = W$. Congruously,

$$W^{-1/2} = U_W\Lambda_W^{-1/2}U_W^T,$$

where $\Lambda_W^{-1/2} = \text{diag}(1/\sqrt{\lambda_{W,1}}, \ldots 1/\sqrt{\lambda_{W,n_w}})$.

Theorem

The HJB equation (3) can be linearized if and only if $\mathcal{R}(G) \subseteq \mathcal{R}(B)$, which is true if and only if

$$E_{N_B^T}^T G = 0, \tag{15}$$

The optimal control policy is then given by the linearized HJB (14) and

$$u = B^+ G W G^T (B^+)^T B^T \frac{1}{Z} (\nabla Z)^T, \tag{16}$$

where B^+ denotes the pseudoinverse of B. The corresponding cost matrix Q is implicitly given by

$$Q^{-1} = \frac{1}{\kappa} \left(B^+ G W G^T (B^+)^T + E_{N_B} \Theta \Theta^T E_{N_B}^T \right. \\ \left. + \epsilon E_\Delta E_\Delta^T \right), \tag{17}$$

where $\kappa > 0$ is an arbitrary scalar, Θ is any full rank matrix, and

$$E_\Delta = (I_m - \Phi(\Phi^T \Phi)^{-1} \Phi^T) E_{R_B^T},$$

where $\Phi = B^+ G W^{1/2}$, and $\epsilon \to 0$.

Remarks

1) The test (15) corresponds to checking that all directions of Gw can be directly counteracted by Bu. It is noteworthy that this sufficient and necessary condition for HJB linearization has this clear physical interpretation.

2) Q determines the cost of control action through $u^T Q u$ in (2). In the same way as for W the SVD of Q is

$$Q = U_Q \Lambda_Q U_Q^T,$$

where $\Lambda_Q = \text{diag}(\lambda_{Q,1}, \dots \lambda_{Q,m})$, determines the cost in the directions given by U_Q. Since the directions are preserved in the SVD of Q^{-1}:

$$Q^{-1} = U_Q \Lambda_Q^{-1} U_Q^T, \tag{18}$$

where $\Lambda_Q^{-1} = \text{diag}(1/\lambda_{Q,1}, \dots 1/\lambda_{Q,m})$, Q^{-1} determines the "preferred" control action directions.

- The second term in (17) is in the null space of B and therefore does not contribute to Bu.
- The first term, on the other hand, corresponds to a complete matching of the directions of Gw (see the proof).
- The last term, which corresponds to the possible directions of the control input that cannot directly counteract the directions of the disturbance, is only needed for Q to be formally invertible. However, by setting ϵ to a small number and invert Q^{-1} it becomes clear from large elements in Q what directions (combinations of input signals) that should/will be avoided.

3) κ is the main tuning parameter that determines the trade-off between cost on the states and cost of control activity. The smaller κ the more agressive the control.

4) If the covariance matrix W has not been determined from data, it is next to κ the remaining parameter. If it is evident, when applying the control policy, that we are never close to a boundary in one direction, the covariance in the corresponding disturbance direction can be decreased. If the constraints in one direction are violated, the corresponding covariance should be increased.

5) For the case when we have input noise on all inputs, (17) reduces to $Q^{-1} = \kappa^{-1} W$, which is the case studied in [5].

Proof

We need to show that if the required conditions are fulfilled, (17) implies that the equality (9) holds when $\epsilon \to 0$.

Presuming Q is positive definite and symmetric, Q^{-1} is also positive definite and symmetric according to (18), and we may define the square root in the same way as shown for W. Equation (9) can then be written

$$G W^{1/2} (G W^{1/2})^T = (\sqrt{\kappa} B Q^{-1/2})(\sqrt{\kappa} B Q^{-1/2})^T.$$

Clearly, the HJB equation is linearized if and only if

$$G W^{1/2} = \sqrt{\kappa} B Q^{-1/2}. \tag{19}$$

Because $Q^{-1/2}$ has full rank, the columns on the right hand side can take any, and no other, vector values than those spanned by the columns of B, i.e. $\mathcal{R}(B)$. So, if and only if the columns of $G W^{1/2}$ are entirely in $\mathcal{R}(B)$ will (19) have a solution. According to the fundamental theorem of linear algebra [8] the left nullspace $\mathcal{N}(B^T)$ is the orthogonal complement of $\mathcal{R}(B)$. Hence, if the columns of $G W^{1/2}$ are entirely in $\mathcal{R}(B)$ all columns should be orthogonal to all base vectors for $\mathcal{N}(B^T)$, which is exactly the test (15).

Now, regard the columns of $Q^{-1/2}$ as vectors (in $\mathbb{R}^m$). Each of these vectors can be written as a sum of one vector in the rowspace $\mathcal{R}(B^T)$ of B and another vector in the nullspace $\mathcal{N}(B)$ of B, since these two spaces are orthogonal and span the entire $\mathbb{R}^m$. We may therefore split $Q^{-1/2}$ such that

$$Q^{-1/2} = Q_{R^T}^{-1/2} + Q_N^{-1/2},$$

where $Q_{R^T}^{-1/2}$ have columns entirely in $\mathcal{R}(B^T)$ and $Q_N^{-1/2}$ have columns entirely in $\mathcal{N}(B)$. Then

$$\begin{aligned} Q^{-1} &= Q^{-1/2}(Q^{-1/2})^T \\ &= Q_{R^T}^{-1/2}(Q_{R^T}^{-1/2})^T + Q_N^{-1/2}(Q_N^{-1/2})^T \\ &= Q_{R^T}^{-1} + Q_N^{-1} \end{aligned}$$

since $(Q_{R^T}^{-1/2})^T Q_N^{-1/2} = 0$ because of the orthogonality.

The component of the columns of $G W^{1/2}$ in $\mathcal{R}(B^T)$ are

$$Q_{G,R^T}^{-1/2} = \frac{1}{\sqrt{\kappa}} B^+ G W^{1/2}.$$

Thus, $Q^{-1} = Q_{G,R^T}^{-1} = Q_{G,R^T}^{-1/2}(Q_{G,R^T}^{-1/2})^T$ is enough to satisfy (9) since $B Q_N^{-1/2} = 0$ by definition. However, if B or G does not have full rank $Q^{-1} = Q_{R^T}^{-1}$ will not be invertible and we have to add symmetric matrices to give Q^{-1} full rank. If B is not full rank we add the basis for the nullspace. From the SVD of B we have that the columns of

E_{N_B} are an orthonormal base for $\mathcal{N}(B)$. Hence, we can fill the space using

$$Q_N^{-1} = E_{N_B}\Theta(E_{N_B}\Theta)^T = E_{N_B}\Theta\Theta^T E_{N_B}^T,$$

where Θ is an arbitrary full rank matrix.

Since

$$BQ_N^{-1}B = (BE_{N_B})(\Theta\Theta^T E_{N_B}^T)B^T = 0$$

Q_N^{-1} will not contribute to (9). Further,

$$Q_N^{-1}B^T(\nabla V)^T = ((\nabla V)BQ_N^{-1})^T = 0$$

because Q_N^{-1} is symmetric. Hence, Q_N^{-1} will neither have an effect on the control signal (4) and nor on the cost (2). This is important because it means that no control signals are wasted in the null space of B, and nor are any control signals generated that would not contribute to the cost function.

If the test (15) is fulfilled

$$Q^{-1} = Q_{G,R^T}^{-1} + Q_N^{-1}$$

will still not have full rank if $GW^{1/2}$ does not fill Q_{R^T}, i.e. there are missing dimensions in $\mathcal{R}(B^T)$. These dimensions are obtained if we project $\mathcal{R}(B^T)$ on $Q_{G,R^T}^{-1/2}$:

$$E_\Delta = (I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)E_{R_B^T},$$

where $\Phi = B^+GW^{1/2}$. The columns of E_Δ corresponds to the directions of the control input not needed to directly counteract the disturbances. If we add

$$Q_{B-G,R^T}^{-1} = E_\Delta(E_\Delta)^T$$

to $Q_{G,R^T}^{-1} + Q_N^{-1}$, Q^{-1} becomes invertible but (9) will no longer hold. However, since the inversion of Q is not explicitly needed for the calculation of the optimal u, we may add $\epsilon Q_{B-G,R^T}^{-1}$ to get (17), i.e.,

$$Q^{-1} = Q_{G,R^T}^{-1} + Q_N^{-1} + \epsilon Q_{B-G,R^T}^{-1}$$

and then let $\epsilon \to 0$.

IV. EXAMPLES

First we illustrate with a very simple example how the vector spaces affect Q. Then two applications of the results are presented. The first one is the case when the disturbance enters as an addition to the control input, but is not purely white. The other application is a buffer example treating the pumping of wastewater to a wastewater treatment plant in Göteborg, Sweden.

A. Illustration of the components of Q^{-1}

Let

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

Clearly B has a null space since the last row are zeros. Further, we see that $\mathcal{R}(G) \subseteq \mathcal{R}(B)$ since G equals twice the first column of B. Hence, we may linearize the HJB and

all three terms of (17) are non-zero. In fact, there will be one dimension in each term:

$$E_{N_B} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad E_\Delta = \begin{bmatrix} 0 & 0 \\ 0.61 & 0.35 \\ 0.61 & 0.35 \end{bmatrix}.$$

Also, because G equals twice the first column of B we realize that only u_1 should need to be used. Calculation of Q from (17) with $\kappa = 1$ and $\epsilon = 10^{-4}$ gives

$$Q = \begin{bmatrix} 0.46 & 0.27 & -0.27 \\ 0.27 & 500 & 500 \\ -0.27 & 500 & 500 \end{bmatrix}$$

which confirms that neither u_2 nor u_3 will be used in the optimal control.

B. Coloured disturbance

For physical reasons almost all disturbances are more or less of low pass character. However, if the system dynamics are slower than the time constants of the disturbances the assumption of white noise is often an acceptable approximation. In some cases though, there may be slow components that also need to be considered. Consider the following system with input disturbance

$$\begin{aligned} \dot{x} &= Ax + B(u + v) \\ y &= Cx + Du + e, \end{aligned} \tag{20}$$

where the disturbance v contains slow dynamics. From physical modeling, or spectral analysis, a model of the disturbance dynamics has been derived:

$$\begin{aligned} \dot{x}_w &= A_w x_w + B_w w \\ v &= C_w x_w + D_w w, \end{aligned} \tag{21}$$

where w is white Gaussian noise with covariance W. Inserting this disturbance model into (20) gives

$$\begin{aligned} \dot{x} &= Ax + Bu + BC_w x_w + BD_w w \\ &= Ax + B\tilde{u} + Gw \end{aligned} \tag{22}$$

where $\tilde{u} = u + C_w x_w$ and $G = BD_w$.

The disturbance state variables x_w are not controllable but a successful observer (for example a Kalman filter) giving estimates $\hat{x}$ and $\hat{x}_w$ close to the real states is assumed to be at hand. Assuming fully known states, x and x_w, the system (22) fits into (1) and because $G = BD_w$ we have that $\mathcal{R}(G) \subseteq \mathcal{R}(B)$, which implies that the condition (15) always holds. Hence, we may apply the linearization and solve the linearized HJB equation for (22) to obtain the control policy $\tilde{u}(x)$. The actually applied control signal is then

$$u = \tilde{u} - C_w \hat{x}_w.$$

The cost matrix Q is now acting on $\tilde{u}$ through

$$\tilde{u}^T Q\tilde{u} = u^T Qu + 2u^T QC_w x_w + x_w^T C_w QC x_w.$$

The first term is the same as before and has the same interpretation as before, and the last term has no effect on the optimal control policy since x_w is not controllable and does not depend on u. The second term is likely to have only a
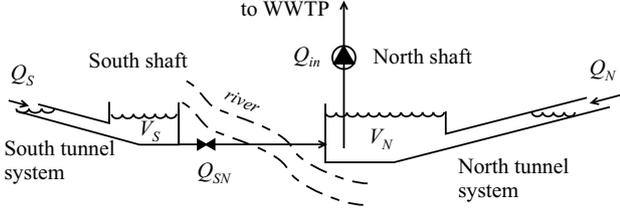
Fig. 1. The Rya WWTPs buffer system for their influent pumping station (24 m below ground level).

limited influence on u because $E\{x_w\} = 0$. However, using SVD the added cost can be written as

$$u^T Q C_w x_w = u^T U_Q \Lambda_Q U_Q^T C_w x_w,$$

where $\Lambda = \text{diag}(\lambda_{Q,1}, \ldots, \lambda_{Q,2})$ defines the costs in the directions of U_Q. It can therefore be seen as an additional cost for the influence of the slow dynamic disturbances ($C_w x_w$) in the directions of u already undesired because of the direct term.

If the disturbance does not enter as an addition to the input as in (20), i.e.

$$\dot{x} = Ax + Bu + \tilde{G}v$$

the treatment is only slightly changed if there exists a matrix M such that $\tilde{G}C_w = BM$. Then we get $u = \tilde{u} - M\hat{x}_w$ and $G = \tilde{G}D_w$. The interpretation of the linearization requirement remains the same, i.e. the direct terms from the white noise must be possible to directly counteract by the control signal.

C. A wastewater treatment buffert system

The Rya wastewater treatment plant (WWTP) treats the wastewater from the Göteborg region. Wastewater is transported to the Rya WWTP through a large tunnel system that can be considered as two separated systems of about the same size, separated by the river Göta älv (see Fig. 1).

Simplified, and with the notation

$$x = \begin{bmatrix} V_N \\ V_S \end{bmatrix}, \quad u = \begin{bmatrix} Q_{in} \\ Q_{SN} \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} Q_N \\ Q_S \end{bmatrix},$$

where all variables are deviations from the operating point, the system can be described by

$$\frac{dx}{dt} = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

w_1 and w_2 are considered (within the time scale considered) as white with a fairly strong correlation, since they are both the results of figuratively the same human habits and rain run off. A covariance matrix

$$W = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \tag{23}$$

is assumed.

Pumping is one of the major costs for this plant and many other WWTPs. For this plant the average flow (including run-off and infiltration) is about 4 m^3/s and hence, every saved

meter of elevation corresponds to more than 340 MWh/year. Remembering that we deal with deviations from an operating point, the variable cost for pumping is assumed to be proportional to $h_{1,\max} - h_1(V_N(t))$, where h_1 depends on the horizontal cross sectional area of the north shaft. However the levels in the north and south shafts should always be above a minimum level and below a maximum level to avoid overflows and the pumps from running dry.

The stationary control problem can now be formulated as

$$\min_u E\left\{h_{1,\max} - h(x_1) + u^T Q u\right\}$$
$$\text{s.t.} \quad 0 \leq x_1 \leq x_{1,\max} \tag{24}$$
$$0 \leq x_2 \leq x_{2,\max}$$

where x_1 and x_2 are the volumes above minimum level. We may assume that the relations between height and volume of the north shaft are roughly given by [3]

$$V_N = 10^4 \cdot h_1^2$$
$$V_S = 10^4 \cdot (0.11 h_2^2 + 0.07 h_2^3).$$

Thus, $h_1 = 0.01\sqrt{x_1}$, and for $h_{1,\max} = 4$ m and $h_{2,\max} = 6$ m we get $x_{1,\max} = 1.6 \cdot 10^5$ m^3 and $x_{2,\max} = 1.9 \cdot 10^5$ m^3.

For this example we have that both B and G have full rank, so

$$Q^{-1} = \frac{1}{\kappa} B^+ G W G^T (B^+)^T,$$

which gives

$$Q = \begin{bmatrix} 2.8 & -5 \\ -5 & 10 \end{bmatrix}$$

for $\kappa = 1$. Because the volume in the north shaft (x_1), is affected by both disturbance flows (through u_2) the cost for u_1 is less than the cost for u_2.

Solving the linearized HJB-equation (14) for $\kappa = 1$ and calculating u gives the result presented in Fig. 2. As can be seen, the control policy produces a control vector that aims away from the constraints, and guarantees that we do not violate them. However, the cost for control activity is not sufficiently low to keep the level high in the north shaft, which was desired in order to reduce the pumping costs. If $\kappa = 0.01$ we allow more control activity and then it should be possible to stay closer to the maximum level in the north shaft. This is confirmed in Fig. 3. The average relative height in the north shaft was increased from 3.0 m ($0.89 \cdot 10^5$ m^3) to 3.40 m ($1.15 \cdot 10^5$ m^3). Note that the full buffer capacity in the south shaft is used in both cases.

Now, assume we have misjudged the disturbance so that the variance of the disturbance in the North shaft is in fact only a tenth what we thought, i.e.

$$W = \begin{bmatrix} 0.1 & 0.25 \\ 0.25 & 1 \end{bmatrix} \tag{25}$$

if the correlation coefficient is unchanged. In Fig. 4 a simulation of that situation is shown ($\kappa = 0.01$). Because of the overestimated variance the controller keeps the level in the north shaft unnecessarily low. To push the level in the north shaft a little higher we decrease the variance of the
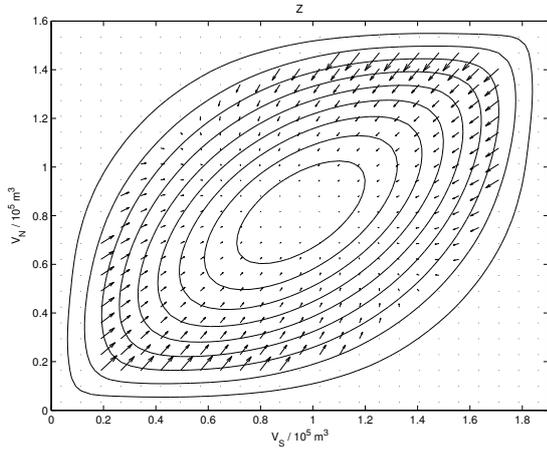
Fig. 2. Contour of the solution Z and direction of the input Bu when $\kappa = 1$. The stationary point where the system has its maximum probability to be is inside the inmost contour.
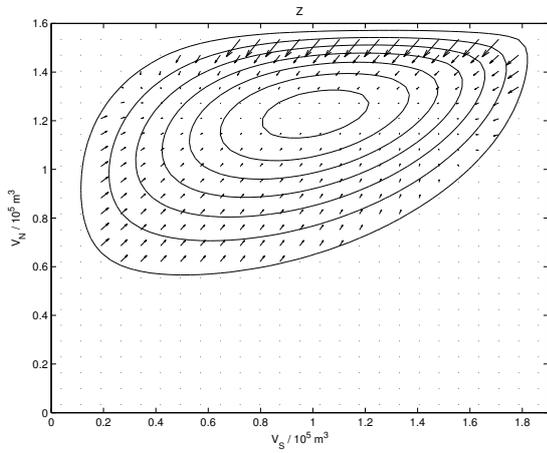

Fig. 3. Contour of the solution Z and direction of the input Bu when $\kappa = 0.01$. The stationary point where the system has its maximum probability to be is inside the inmost contour.

influent flow to the north shaft (keeping the cross correlation coefficient). As can be seen in Fig. 5, redoing the design with the smaller and correct W means that we can stay closer to the upper limit.

V. CONCLUSIONS

A method for state constrained control based on stochastic optimal control has been presented. The method is a result of an exact linearization of the HJB-equations, which has implications on how the cost for control is defined. However, the implications have been thoroughly investigated and found to be in agreement with intuitive reasoning about suitable cost matrices. There is one main tuning parameter, which determines the control activity, and then the covariance matrix of the white noise disturbance can also be used for tuning the preferred (or undesired) directions of the control. The method assumes white noise but it has been shown how it can also be applied in cases when the disturbance
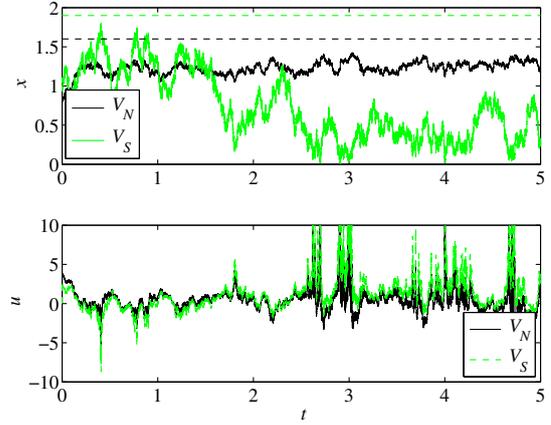
Fig. 4. Simulation of the controlled system with the policy calculated for W given by (23) and simulated for the true W given by (25).
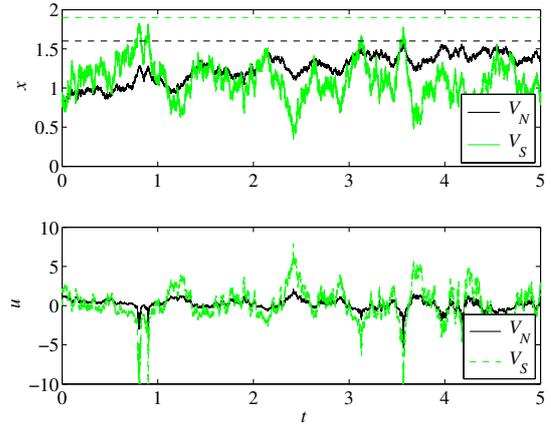
Fig. 5. Simulation of the controlled system with the policy calculated for W given by (25) and simulated for the same (true) W.

is coloured. A buffer example, pumping of wastewater to a wastewater treatment plant from two different shafts in a sewer system, has illustrated the use of the method.

REFERENCES

[1] P. Dorato, T. C. Abdallah, and V. Cerone. *Linear Quadratic Control: An Introduction*. Prentice-Hall, 1995.
[2] T. Itami. Nonlinear optimal control as quantum mechanical eigenvalue problems. *Automatica*, 41:1617–1622, 2005.
[3] J. Lindqvist, T. Wik, D. Lumley, and G. Äijälä. Influent load prediction using low order adaptive modeling. In *2nd IWA Conference on Instrumentation, Control and Automation*, Busan, South Korea, 2005.
[4] H.H. Rosenbrock. *Physics letters*, 110A:343–346.
[5] P. Rutquist, C. Breitholtz, and T. Wik. On the infinite time solution to state-constrained stochastic optimal control problems. *Automatica*, 44:1800–1805, 2008.
[6] P. Rutquist, C. Breitholtz, and T. Wik. Finite-time state-constrained optimal control for input affine systems with actuator noise. *Automatica*, Submitted.
[7] Per Rutquist, Claes Breitholtz, and Torsten Wik. An eigenvalue approach to infinite-horizon optimal control. In *Proc. 16th IFAC World Congress*, Prague, Czech Republic, jul 2005.
[8] G. Strang. *Linear Algebra and its Applications*. Harcourt Brace and Company, Florida, 3 edition, 1988.

Application of Optimal Control Theory to a Batch Crystallizer using Orbital Flatness

Steffen Hofmann[1] and Jörg Raisch[1,2]

[1]Technische Universität Berlin, Einsteinufer 17, 10857 Berlin, Germany
[2]Max-Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr.1, 39106 Magdeburg, Germany

Abstract

In this contribution we suggest an efficient application of the Pontryagin Minimum Principle (PMP) to the solution of an optimal control problem for a standard moment model of a batch crystallizer. The application is based on a time scaling that turns the moment model differentially flat, and on a simplifying assumption. We present efficient solutions and also consider uniqueness of the solutions. We demonstrate ways to assess and partially correct the error due to simplification. Finally, we use a case study to demonstrate the optimal trajectories, numerical effort and the magnitude of the error.

1 Introduction

Optimal control of batch crystallizers has been an active topic for a long time. Mullin and Nyvlt [1] recognized in 1971 that the final crystal size can be increased by using a "programmed" crystallization temperature trajectory rather than natural (exponential) cooling of the solution. In the time following, specific optimal control problems were formulated and solved using optimal control theory as well as numerical methods [2], [3], [4], [5], [6]. Requirements on the final crystalline product are commonly expressed in terms of the crystal size distribution (CSD) or of its moments. The field remains active, with recent publications like [7]. Also, a new focus on feedback control has emerged. Schemes such as online-optimization or model predictive control [8], [9], [10] demand for computationally fast solutions or other means of efficient control algorithms.

By recognizing the orbital flatness property of a standard moment model describing the crystallization of a single substance out of solution, with crystallization temperature T, or jacket temperature T_j, being the control input, a way was opened up in [11] to approach some related control problems in an analytical way. In particular, it was shown that the model can be inverted analytically in order to achieve a desired final CSD stemming from nucleation. Based on this, an optimization scheme for a common setup based on a parametrization of this final CSD was suggested. Here, we show that the results in [11] allow the application of the Minimum Principle (PMP) for optimal control of the batch crystallizer in an efficient way if a suitable assumption is made, requiring very little numerical effort.

2 Model and optimization task

We consider a standard moment model that describes the population dynamics for the crystallization of a single substance out of solution:

$$\dot{\mu}_i = iG\left(\cdot\right)\mu_{i-1}, \ i = 1, 2, \ldots$$
$$\dot{\mu}_0 = B\left(\cdot\right), \tag{1}$$

where $G\left(\cdot\right)$ and $B\left(\cdot\right)$ are the (crystal size independent) growth and nucleation rate, respectively, which both depend on other variables in the system. In fact, a model including only the first four moments, μ_0 to μ_3, can be closed via the mass balance equation

$$m_l\left(t\right) = m_{l,0} - \rho k_V\left(\mu_3\left(t\right) - \mu_3\left(0\right)\right), \tag{2}$$

where m_l denotes the mass of dissolved substance, ρ is the density of the crystals, k_V is a volume shape factor, and $m_{l,0} = m_l(0)$. $G\left(\cdot\right)$ and $B\left(\cdot\right)$ are defined as in [11]:

$$G\left(\cdot\right) = k_g\left(S - 1\right)^g \tag{3a}$$
$$B\left(\cdot\right) = k_b\left(S - 1\right)^b \mu_3\left(t\right) \tag{3b}$$

Only growth and secondary nucleation are assumed to take place, and k_g, k_b, g and b are positive constants that depend on the substances used and on design variables, like stirrer speed. Furthermore, it is assumed that $b > g$. This assumption is justified for many practically important substances. The supersaturation S is given by

$$S = \frac{c}{c_{sat}}, \quad c = \frac{m_l}{m_W}, \tag{4}$$

with m_W being the mass of the solvent. The *solubility* relation is approximated as a polynomial

$$c_{sat}\left(T\right) = a_0 + a_1 T + a_2 T^2, \tag{5}$$

with positive coefficients a_0, a_1 and a_2, where c_{sat} is the saturation concentration, and T is the temperature of the slurry, or crystallization temperature, and is assumed to be spatially constant.

2.1 Orbital flatness and separation of growth and nucleation

In [11] it is shown that a standard moment model of a batch cooling crystallizer is orbitally flat, i.e. flat in an appropriate, scaled time domain. A new, scaled time τ is introduced, together with the transformation

$$d\tau = G(t)\,dt \qquad (6)$$

This time scaling turns the moment model into a chain of integrators, where growth and nucleation rate enter only at the input of the first integrator. Here, we apply the transformation to a model describing separately the evoultion of the moments $\mu_{i,s}$ of the CSD of growing seed crystals and the moments $\mu_{i,n}$ of the CSD of newly nucleated (and growing) crystals. This split model (in normal time) is also used e.g. in [12], [7]:

$$\frac{d}{d\tau}\mu_{i,s} = i\mu_{i-1,s},\ i = 1\ldots 3 \qquad (7a)$$

$$\frac{d}{d\tau}\mu_{0,s} = 0 \qquad (7b)$$

$$\frac{d}{d\tau}\mu_{i,n} = i\mu_{i-1,n},\ i = 1\ldots 3 \qquad (7c)$$

$$\frac{d}{d\tau}\mu_{0,n} = \frac{B(\mu_{3,s} + \mu_{3,n}, T(\tau))}{G(\mu_{3,s} + \mu_{3,n}, T(\tau))} \qquad (7d)$$

The crystallization temperature, or slurry temperature, T is commonly used as a manipulated variable. This is signified in the equations by explicitly denoting the dependency on τ. It is assumed here that it can be adjusted precisely and arbitrarily fast. In practice, it can be regulated using a low-level thermostat controller. An additional equation is added to keep track of real time t, which is also an important process variable. Storing this variable allows the transformation of computed trajectories from the τ- into the t-domain:

$$\frac{d}{d\tau}t = \frac{1}{G(\mu_{3,s} + \mu_{3,n}, T(\tau))} \qquad (7e)$$

The initial conditions of the system (7) are:

$$\mu_{i,s}(0) = \mu_{i0,s}\ ;\ \ \mu_{i,n}(0) = 0\ ;\ \ t(0) = 0 \qquad (8)$$

where $\mu_{i0,s}$, $i = 0\ldots 3$ are the (positive) moments of the CSD of seed crystals. Note that, as indicated in (7d) and (7e), we want to emphasize that $B(\cdot)$ and $G(\cdot)$ are functions only of T and the total third moment, $\mu_3 = \mu_{3,s} + \mu_{3,n}$, which are obtained by combining (2) to (5). From (7a) to (7b) it can be seen that the evolution of the moments of the grown seeds CSD is decoupled from that of the moments of the nucleated CSD and depends only on initial conditions and on τ. The third moment of the grown seeds CSD is given by:

$$\mu_{3,s}(\tau) = \mu_{00,s}\tau^3 + 3\mu_{10,s}\tau^2 + 3\mu_{20,s}\tau + \mu_{30,s} \qquad (9)$$

In the following discussion, we consider $\mu_{i0,s}$ as constants and refer to $\mu_{3,s}(\tau)$ as a known function of τ. Note that the case of monodisperse seed crystals (seed CSD thought to be concentrated at one length L_0), which is often used in literature (e.g. [2]), leads to a special case of (9).

2.2 Optimal control problem

When the crystallization temperature T is considered as the manipulated variable, or control input, one wants to find an optimal input $T^*(t)$. Such an optimal control problem is usually formulated using a cost function, constraints that must hold while the process evolves and final time constraints. The basic requirement is to produce a large amount of crystalline product in a short time, which could be formulated in different configurations of cost function and constraints. But, at least for the growth and nucleation models used here, this problem setting leads to trivial solutions, like keeping temperature always as low as possible (or allowed).

Other problems have also been treated. For example, in [2] the final length of growing seed crystals is maximized. In [5], [11], the final ratio of crystalline masses stemming from nucleation and growing seeds,

$$\mathrm{imp}_f := \frac{\mu_{3f,n}}{\mu_{3f,s}}, \qquad (10)$$

where $\mu_{3f,n} = \mu_{3,n}(\tau_f)$ and $\mu_{3f,s} = \mu_{3,s}(\tau_f)$, is minimized subject to a constraint on the final total crystallized mass, $m_f := \rho k_V(\mu_{3f,s} + \mu_{3f,n})$. In [13] the difference $\mu_{3f,n} - \mu_{3f,s}$ is minimized. In [7] a number of related optimization problems are considered.

These approaches are mostly motivated by the idea that crystalline mass stemming from nucleation is undesirable, e.g. because these crystals do not have a lot of time to grow and are too small (product requirements, filtration etc.) at the end of the batch. Especially, if one wants to achieve a very narrow final CSD consisting of the shifted CSD of the seed crystals, any nucleation is disturbing. In the following, we develop a new method for the solution of the following optimal control problem, similar to the problem in [11]:

$$\min_{u(\tau)}\ \mu_{3,n}(\tau_f)\ \text{ s.t. }\ \begin{array}{rcl} \mu_{3,s}(\tau_f) & = & \mu_{3f,s,c} \\ t(\tau_f) & \leq & t_{f,c} \\ u(\tau) & \in & U\ \forall \tau \in [0, \tau_f] \end{array} \qquad (11)$$

where τ_f is the final scaled time and U is the compact set of allowed input values. Note that, for simplicity, (11) uses an equality constraint on $\mu_{3,s}(\tau_f)$, rather than the economically meaningful inequality constraint $\mu_{3,s}(\tau_f) \geq \mu_{3f,s,c}$. However, numerical results indicate that increasing the constraint $\mu_{3f,s,c}$ will increase the minimal cost when, at the same time, the constraint on t_f is maintained. Note also that, when the final $\mu_{3f,s} = \mu_{3,s}(\tau_f)$ is fixed, minimization of $\mu_{3f,n}$ is equivalent to minimization of imp_f. One could also think of variations of the problem (11) that, for example, involve constraints on $\mu_{3,n}(\tau_f)$. The general goal is to keep $\mu_{3,n}(\tau_f)$ low compared to $\mu_{3,s}(\tau_f)$ while producing "enough" of the desired $\mu_{3,s}(\tau_f)$.

With T as the control input, i.e. $u(\tau) \triangleq T(\tau)$, U could be defined as $U = [T_{\min}, T_{\max}]$. The bounds on T can be determined by the capabilities of the thermostat used for regulation, etc. Here, we define a different control input, namely the inverse of the growth rate,

in which case $u(\tau) \triangleq 1/G(\tau)$, and define fixed bounds on this variable:

$$U = \left[\left(\tfrac{1}{G} \right)_{\min}, \left(\tfrac{1}{G} \right)_{\max} \right] \qquad (12)$$

With the growth law (3a), $\left(\tfrac{1}{G} \right)_{\min}$ and $\left(\tfrac{1}{G} \right)_{\max}$ can be seen as upper and lower limits on supersaturation S, respectively. These bounds could make sense in that the process is operated in a metastable region, where the model equations can be considered to be sufficiently accurate. We will allow $\left(\tfrac{1}{G} \right)_{\max}$ to be chosen arbitrarily large, so that it is never hit. It will be shown in the sequel that, under normal conditions, the crystallization temperature T that is needed to realize $1/G$ can be uniquely computed.

3 Application of the PMP

For solving the proposed optimization problems, the necessary conditions of optimality according to the Pontryagin Minimum Principle [14], [15] (PMP) have been stated several times in the respective literature, but analytical solutions have been found to be difficult to obtain. Numerical methods have then been employed, for example control vector iteration is used in [2]. In other contributions, a simplification is suggested to make solutions easier. For derivation of a constant growth rate trajectory in [1], and for optimization in [3], and subsequently in [4], it is suggested to neglect nucleation as it is supposed to be very low when the process is operated within the metastable region.

In contrast to the latter publications, we do consider nucleation as it is part of the problem (11). However, since the goal is to suppress nucleation, we neglect the feedback that it has on the crystallization process. This feedback would normally be due to the mass balance (feedback of $\mu_{3,n}$ on concentration and thus supersaturation) and to secondary nucleation. The effects of this simplification will be assessed later on. The resulting optimization problem has a particularly simple structure when it is written in the transformed time (τ)-domain.

In preparation for doing so, we relabel the state variables. Remember that the moments $\mu_{i,s}$ of the CSD of growing seed crystals can be expressed as functions of τ, especially $\mu_{3,s}(\tau)$ is given by (9). Therefore, it is not necessary to regard these moments as state variables. Hence, we choose the state according to

$$x_1 \triangleq \mu_{3,n}; \;\; x_2 \triangleq \mu_{2,n}; \;\; x_3 \triangleq \mu_{1,n}; \;\; x_4 \triangleq \mu_{0,n}; \;\; x_5 \triangleq t; \qquad (13)$$

We arrive at a system of differential equations for a model with five states and explicit time (τ)-depedency:

$$\tfrac{d}{d\tau} x_1 = 3x_2, \quad \tfrac{d}{d\tau} x_2 = 2x_3, \quad \tfrac{d}{d\tau} x_3 = x_4 \qquad (14a)$$

$$\tfrac{d}{d\tau} x_4 = \tfrac{B}{G}\left(\mu_{3,s}(\tau) + x_1, u(\tau) \right) \qquad (14b)$$

$$\tfrac{d}{d\tau} x_5 = u(\tau) \qquad (14c)$$

where $u(\tau) \triangleq \tfrac{1}{G}(\tau)$ is the control input, and $\tfrac{B}{G}(\cdot)$ shall be understood as:

$$\tfrac{B}{G}\left(\mu_{3,s}(\tau) + x_1, u(\tau) \right) := B\left(\mu_{3,s}(\tau) + x_1, \overline{T}(\tau) \right) u(\tau) \qquad (15)$$

where, for every instant of τ, $\overline{T}(\tau)$ is the (positive) solution of

$$G\left(\mu_{3,s}(\tau) + x_1, \overline{T}(\tau) \right) = (u(\tau))^{-1} \qquad (16)$$

For the simple growth and nucleation laws (3a) and (3b), the temperature $\overline{T}$ need not be computed explicitly. The required supersaturation $\overline{S}$ can be obtained from (3a) and then directly be inserted in (3b). The resulting expression for $\tfrac{B}{G}$ is

$$\tfrac{B}{G}\left(\mu_{3,s}(\tau) + x_1, u(\tau) \right)$$
$$= k_b k_g^{-\frac{b}{g}} u(\tau)^{\frac{g-b}{g}} \left(\mu_{3,s}(\tau) + x_1 \right) \qquad (17)$$

However, as $\overline{T}$ is the manipulated variable, it has to be computed. It can be seen from (4), (5) that, as $a_0, a_1, a_2 > 0$, a real positive (inc. zero) solution exists for $\overline{T}$ if and only if $\overline{c}_{sat}(\tau) = c(\tau)/\overline{S}(\tau) \geq a_0$.

We now propose an idealization of the model that will substantially simplify the solution of the optimal control problem discussed in the following. We negelect any feedback of the nucleated part of the CSD on the crystallization process, i.e., we assume that

$$\mu_3 = \mu_{3,s} + \mu_{3,n} \approx \mu_{3,s} \qquad (18)$$

As a consequence, we replace $\mu_3 = \mu_{3,s} + x_1$ by $\mu_{3,s}$. With this simplification, the total third moment that affects the system does only depend on the value of τ, see (9). The system can then be written

$$\tfrac{d}{d\tau} x_1 = 3x_2, \quad \tfrac{d}{d\tau} x_2 = 2x_3, \quad \tfrac{d}{d\tau} x_3 = x_4 \qquad (19a)$$

$$\tfrac{d}{d\tau} x_4 = \tfrac{B}{G}\left(\mu_{3,s}(\tau), u(\tau) \right) \qquad (19b)$$

$$\tfrac{d}{d\tau} x_5 = u(\tau), \qquad (19c)$$

where $\mu_{3,s}(\tau)$ is given by (9). Nevertheless, it is important to track the third moment from nucleation, $\mu_{3,n} \triangleq x_1$, as one process requirement is to keep it as low as possible. Only its feedback on the system is neglected.

The optimization is a Mayer type problem, because the cost function in (11) is expressed in terms of the final state and does not include an integral term, with the Hamiltonian $H(\mathbf{x}, u, \boldsymbol{\psi}, \tau) = \boldsymbol{\psi}^T \tfrac{d}{d\tau} \mathbf{x}$:

$$H(\mathbf{x}, u, \boldsymbol{\psi}, \tau) = 3\psi_1 x_2 + 2\psi_2 x_3 + \psi_3 x_4$$
$$+ \psi_4 \tfrac{B}{G}\left(\mu_{3,s}(\tau), u \right) + \psi_5 u \qquad (20)$$

Next, we state the necessary conditions (NC) for an optimal trajectory. The differential equations of the adjoint system $\boldsymbol{\psi}$ are, in general, $\tfrac{d}{d\tau} \boldsymbol{\psi} = -\tfrac{\partial}{\partial \mathbf{x}} H(\cdot)$. Hence,

$$\tfrac{d}{d\tau} \psi_1 = 0 \qquad (21a)$$

$$\tfrac{d}{d\tau} \psi_2 = -3\psi_1, \quad \tfrac{d}{d\tau} \psi_3 = -2\psi_2, \quad \tfrac{d}{d\tau} \psi_4 = -\psi_3 \qquad (21b)$$

$$\tfrac{d}{d\tau} \psi_5 = 0 \qquad (21c)$$

At each instant of τ, as long as the control is not singular, the optimal control input $u^*(\tau)$ must satisfy

$$u^*(\tau) = \underset{u \in U}{\arg\min} \, H(\mathbf{x}, u, \boldsymbol{\psi}, \tau)$$
$$= \underset{u \in U}{\arg\min} \, \psi_4 \tfrac{B}{G}\left(\mu_{3,s}(\tau), u \right) + \psi_5 u \qquad (21d)$$

where U is the compact set of allowed input values (12). Like said before, the upper limit $(\frac{1}{G})_{max}$ is only auxiliary and shall be chosen as large as necessary so that it is never hit. The boundary conditions for the states are

$$\mathbf{x}(0) = \mathbf{0}; \quad x_5(\tau_f) \leq t_{f,c} \quad (21e)$$

From (11) follows the cost function ϕ, and, as x_{1f} is not part of a constraint, the terminal condition for ψ_1:

$$\phi(\mathbf{x_f}, \tau_f) = x_{1f}; \quad \frac{\partial\phi(\cdot)}{\partial x_{1f}} = 1 \;\Rightarrow\; \psi_1(\tau_f) = 1 \quad (21f)$$

The final values of the moments $\mu_{0f,n}$, $\mu_{1f,n}$ and $\mu_{2f,n}$, i.e. of $x_4(\tau_f)$, $x_3(\tau_f)$ and $x_2(\tau_f)$, are neither part of the cost function nor of constraints, so that

$$\psi_2(\tau_f) = \psi_3(\tau_f) = \psi_4(\tau_f) = 0 \quad (21g)$$

From (21a) through (21c) and (21g) it can be seen that, with knowledge of τ_f, the evolution of all adjoint states can be readily computed (uniquely) up to two multiplicative constants, k_1 and k_2:

$$\psi_1(\tau) = k_1 \quad (22a)$$
$$\psi_2(\tau) = 3k_1(\tau_f - \tau) \quad (22b)$$
$$\psi_3(\tau) = 3k_1(\tau_f - \tau)^2 \quad (22c)$$
$$\psi_4(\tau) = k_1(\tau_f - \tau)^3 \quad (22d)$$
$$\psi_5(\tau) = k_2 \quad (22e)$$

In order to find the optimal solution, the two constants k_1 and k_2 have to be determined. Additionally, (21d) has to be solved for u^* at each τ between 0 and τ_f. The final "time" τ_f is given by the equality constraint $\mu_{3f,s,c}$ and can be computed beforehand by solving (9) (where all $\mu_{i0,s} > 0$). Because of (21f), $k_1 = 1$.

With (17), and the simplification, (21d) becomes:

$$u^*(\tau) = \operatorname*{argmin}_{u \in U} \psi_4 k_b k_g^{-\frac{b}{g}} u^{\frac{g-b}{g}} \mu_{3,s}(\tau) + \psi_5 u \quad (23)$$

A candidate for an unconstrained minimizer, u^o, if it exists, is obtained by setting the derivative of H with respect to u equal to zero:

$$\left.\frac{\partial H(\cdot)}{\partial u}\right|_{u^o} = -\psi_4 \frac{b-g}{g} k_b k_g^{-\frac{b}{g}} (u^o)^{-\frac{b}{g}} \mu_{3,s}(\tau) + \psi_5 \overset{!}{=} 0$$

$$u^o = \frac{1}{kg}\left(\frac{\psi_4}{\psi_5}\frac{b-g}{g} k_b \mu_{3,s}(\tau)\right)^{\frac{g}{b}} \quad (24)$$

For a real valued solution to exist, the term inside the parantheses must be positive. Because of (22d), with $k_1=1$, it follows that $\psi_4(\tau) > 0$, $\forall \tau \in [0, \tau_f)$ and $\psi_4(\tau_f) = 0$. Note that $(g-b)/g$ in (23) is negative because of the assumptions that $g > 0$, $b > 0$ and $b > g$. Also, $k_b > 0$, $k_g > 0$ and $\mu_{3,s}(\tau) > 0$. Hence, $H(\cdot)$ is strictly monotonically increasing in u when $\psi_4 = 0$ and $\psi_5 > 0$ and it is strictly monotonically decreasing in u when $\psi_4 > 0$ and $\psi_5 \leq 0$ or when $\psi_4 = 0$ and $\psi_5 < 0$. The second derivative with respect to u is

$$\frac{\partial^2 H(\cdot)}{\partial u^2} = \psi_4 \frac{b-g}{g}\frac{b}{g} k_b k_g^{-\frac{b}{g}} u^{\frac{-b-g}{g}} \mu_{3,s}(\tau) \quad (25)$$

and its sign is always equal to the sign of ψ_4 ($u > 0$). Thus, $H(\cdot)$ is convex in u when ψ_4 is positive and linear in u when ψ_4 is zero. The limits for $u \to 0$ and $u \to \infty$, of the part of H given in (23), named $H^{(u)}$, are

$$\lim_{u \to 0} H^{(u)}(\cdot) = \qquad \lim_{u \to \infty} H^{(u)}(\cdot) =$$
$$\begin{cases} 0 & \psi_4 = 0 \\ \infty & \psi_4 > 0 \end{cases} \qquad \begin{cases} -\infty & \psi_5 < 0 \\ 0 & \psi_5 = 0 \\ \infty & \psi_5 > 0 \end{cases} \quad (26)$$

So, as $\psi_4 \geq 0$, and as long as not both ψ_4 and ψ_5 are zero, which by now we conclude is only possible at $\tau = \tau_f$, (23) has always a unique solution which is either $(\frac{1}{G})_{min}$, (24) or $(\frac{1}{G})_{max}$. Also, if a real valued solution exists for u^o, it is really an unconstrained minimizer (no maximizer or saddle point). Only when both ψ_4 and ψ_5 are zero, $H(\cdot)$ is independent of u. The fact that this is not possible for an extended period in (τ)-"time" prohibits the existence of solutions with singular control intervals.

The remaining task is to find the value of $k_2 = \psi_5(\tau)$. The monotonicity considerations show that, as $k_1 > 0$, and in case $k_2 \leq 0$, u^* would have to be $(\frac{1}{G})_{max}$ at least for $\tau < \tau_f$. Looking at (19c), it is clear that for any value of $t_{f,c}$, $(\frac{1}{G})_{max}$ can be chosen large enough so that the terminal constraint $t_f \leq t_{f,c}$ is violated. We conclude that $k_2 > 0$. Also, the fact that $k_2 \neq 0$ shows that the constraint on t_f cannot be removed, i.e. it has to be active (otherwise the state x_5 could have been removed from the problem, resulting in (23) without the ψ_5-term, i.e. $\psi_5 = 0$). The constraint $x_{5f} \leq t_{f,c}$ can be replaced by the new boundary condition

$$x_{5f} = t_{f,c} \quad (27)$$

When the value of k_2 that makes the NC hold is unique, then the whole solution is unique. A solution is, however, not guaranteed. The minimum time t_f possible to produce the desired $\mu_{3f,s,c}$ is obtained by setting $u(\tau) = (\frac{1}{G})_{min}$ $\forall \tau \in [0, \tau_f]$, leading to the maximum possible $\mu_{3f,n}$. If this time is greater than $t_{f,c}$, then no solution is feasible. Also, if the increase in μ_3 is too large, i.e., concentration is reduced too much, then the control $u^*(\tau)$ can be computed, but it cannot be realized by controlling T (see end of this section). We assume that the process is operated in a region where this is excluded.

In order to compute a solution to the NC, the only remaining task is to find one value, k_2, known to be greater than zero. If an equation error can be defined which is strictly monotonic in this value, then one way to numerically find an arbitrarily close approximation is to use bisection with a preceding extrapolation step. In fact, when $k_2 = \psi_5(\tau)$ gets smaller, then the unconstrained minimizer (24) gets larger for any instant of τ. Then, $\forall \tau \in [0, \tau_f]$, the constrained minimizer $u^*(\tau)$ will increase or stay the same. Near τ_f, when $\psi_4(\tau) \to 0$, there will always be an interval where $u^*(\tau) < (\frac{1}{G})_{max}$. When the case $u^*(\tau) = (\frac{1}{G})_{min}$ $\forall \tau \in [0, \tau_f]$ has already been excluded (this is only the optimal solution if the resulting t_f is

exactly equal to $t_{f,c}$), it is clear that at least for some time u^* will increase (when the constraints $(\frac{1}{G})_{\min}$ and $(\frac{1}{G})_{\max}$ are not active). This means that the final time t_f, which is the integral over u^*, will grow strictly monotonically with decreasing k_2.

This also shows that the value of k_2 that makes the NC hold is unique. Together with the unique solution for u^* at every τ (including τ_f as $k_2 = \psi_5(\tau) > 0$), a unique solution to the necessary conditions is found (if it exists), which must then be truly optimal.

The solution of the optimal control problem yields not only $u^*(\tau)$, but also the state trajectory $\mathbf{x}^*(\tau)$ and, amongst others, $\mu_{3,s}^*(\tau) = \mu_{3,s}(\tau)$ and supersaturation $S^*(\tau)$. Finally, to be able to realize the trajectory $u^*(\tau) = (\frac{1}{G})^*(\tau)$ by manipulating $T(t)$ in real time t, two steps must be taken

1. $T^*(\tau)$ has to be computed from $(\frac{1}{G})^*(\tau)$ by using (2), (3a), (4) and (5), $\tau \in [0, \tau_f]$.

2. $T^*(\tau)$ has to be converted to $T^*(t)$. The trajectory $t^*(\tau) = x_5^*(\tau)$ can be used.

For the first step it has to be checked whether $c_{sat}^*(\tau) = c(\tau)/S^*(\tau) \geq a_0 \; \forall \tau \in [0, \tau_f]$. As the concentration is smallest and the supersaturation is at its upper limit at $\tau = \tau_f$, it is enough to consider this instant.

3.1 Efficient solutions

For the given growth and nucleation laws, there exists an alternative, and more efficient way to determine the parameter k_2. Recognizing from (24) that the unconstrained minimizer of the Hamiltonian will always tend to zero towards the end of the batch, when $\psi_4(\tau)$ goes to zero, it is clear that the inequality constraint on the input will be active when τ approaches τ_f. It is then assumed that the control consists of two arcs, one unconstrained (sensitivity seeking) arc at the beginning and one constrained arc at the end. We denote the switching time between these arcs τ_s. Equation (24) can also be written

$$u^o(\tau) = \psi_5^{-\frac{g}{b}} \frac{1}{kg} \left(\psi_4(\tau) \frac{b-g}{g} k_b \mu_{3,s}(\tau) \right)^{\frac{g}{b}} \quad (28)$$

Note that no differential equations need to be integrated to compute $u^o(\tau)$. At the switching instant the unconstrained minimizer u^o will be exactly equal to the constrained minimizer $(\frac{1}{G})_{\min}$. This gives a relation between the switching instant τ_s and the constant $k_2 = \psi_5$ required for the switching to happen at this moment:

$$\psi_5(\tau_s) = (\tfrac{1}{G})_{\min}^{-\frac{b}{g}} k_g^{-\frac{b}{g}} \psi_4(\tau_s) \frac{b-g}{g} k_b \mu_{3,s}(\tau_s) \quad (29)$$

By integrating (28), the evolution of real time t dependent on τ can be computed for the unconstrained arc, up to $\psi_5 = k_2$, resulting in the real time t_s corresponding to τ_s:

$$t_s(\tau_s) =$$

$$\psi_5(\tau_s)^{-\frac{g}{b}} \int_0^{\tau_s} \frac{1}{kg} \left(\psi_4(\tau) \frac{b-g}{g} k_b \mu_{3,s}(\tau) \right)^{\frac{g}{b}} d\tau \quad (30)$$

Because the term with $\psi_5(\tau_s)$ has been put outside the integral, the integration can be done numerically in a cumulative way, $\forall \tau_s \in [0, \tau_f]$. It is not necessary to integrate from zero to τ_s, for every value of τ_s. Knowing τ_f and the required final $t_f = t_{f,c}$, the real time can also be integrated backwards from τ_f to τ_s. In this case, the constrained control input has to be used:

$$t_s(\tau_s) = t_{f,c} + \int_{\tau_f}^{\tau_s} (\tfrac{1}{G})_{\min} d\tau = t_{f,c} + (\tfrac{1}{G})_{\min}(\tau_s - \tau_f) \quad (31)$$

Because (31) and (30) (where (29) is substituded for $\psi_5(\tau_s)$) have to be consistent, the switching instant can be computed as the intersection of the corresponding curves. This is equivalent to detecting the zero crossing of $(30)-(31)$. Finally, ψ_5 can be obtained by again using (29).

4 Analysis of the error

As mentioned earlier, because we consider T as the manipulated variable and the real time is t, $T^*(t)$ has to be computed from $u^*(\tau) = (\frac{1}{G})^*(\tau)$. However, there exists a mismatch between the simplified model employed for finding the optimal control and the non-simplified, *detailed* model, (1) to (5), where the third moment from nucleation, $\mu_{3,n}$, has feedback on the system in two ways:

1. via (2), (3), (4): feedback on supersaturation via the mass balance

2. via secondary nucleation, (3b), where really $\mu_3 = \mu_{3,s} + \mu_{3,n}$

This mismatch will result in an error when the detailed model is simulated with $T^*(t)$, which has been computed using the simplified model, as the control. By computing a new control, $T'(t)$, we now propose a way to correct this error in the sense that the constraints $t_{f,c}$, $\mu_{3f,s,c}$ are maintained. Therefore, the non-simplified model in τ-domain, (14), (17), is simulated with $u'(\tau)$ as the control, where we set $u'(\tau) = u^*(\tau)$, resulting in trajectories $\mathbf{x}'(\tau) = [\mu_{3,n}', \mu_{2,n}', \mu_{1,n}', \mu_{0,n}', t']^T(\tau)$, $\mu_{3,s}'(\tau)$.

Of course, $\mu_{3,s}'(\tau) = \mu_{3,s}^*(\tau) = \mu_{3,s}(\tau)$, $\forall \tau \in [0, \tau_f]$. From (14c) it is also clear that $t'(\tau) = t^*(\tau)$, $\forall \tau \in [0, \tau_f]$, whereas the positive feedback of x_1 in (14a), (14b), (17) causes $\mu_{3,n}'(\tau) > \mu_{3,n}^*(\tau)$, $\forall \tau \in (0, \tau_f]$.

Like before, the control $T'(\tau)$ is obtained using (2), (3a), (4) and (5), this time setting $\mu_3 = \mu_3' = \mu_{3,s} + \mu_{3,n}'$. Knowing $t'(\tau) = t^*(\tau)$, $T'(\tau)$ can be converted to $T'(t)$. In τ-domain, $t^*(\tau)$ is a state trajectory. Changing the independent variable from τ to t, one gets a corresponding state trajectory, $\tau^*(t)$. Because $t'(\tau) = t^*(\tau)$ in τ-domain, $\tau'(t) = \tau^*(t)$ in t-domain. Thus $\mu_{3,s}'(t) = \mu_{3,s}^*(t)$. From there it is clear that both constraints will be maintained when the detailed model is simulated in normal t-time with $T'(t)$ as the control.

However, $T'(t)$ is not, in general, the optimal control for the detailed model, i.e., by solving the optimal control problem for the detailed model, one should achieve a result $\mu_{3,n}^{**}(\tau_f)$, which is less than $\mu_{3,n}'(\tau_f)$, whilst maintaining the constraints $t_{f,c}$ and $\mu_{3f,s,c}$. We use the sub-optimal result $\mu_{3f,n}' = \mu_{3,n}'(\tau_f)$ as an upper bound for $\mu_{3,n}^{**}(\tau_f)$. For finding a lower bound, we consider (14a), (14b) and (17). In the detailed model in τ-domain, the feedback of $\mu_{3,n} \overset{\triangle}{=} x_1$ is only via (17), which can be factorized:

$$\frac{B}{G}\left(\mu_{3,s}(\tau) + x_1, u(\tau)\right)$$
$$= k_b k_g^{-\frac{b}{g}} u(\tau)^{\frac{g-b}{g}} \mu_{3,s}(\tau) + k_b k_g^{-\frac{b}{g}} u(\tau)^{\frac{g-b}{g}} x_1 \quad (32)$$

The second term, which must be zero or positive, is only considered in the detailed model. Looking at the linear chain of integrators, (14a), (14b), it is clear that this additional nucleation term, as soon as it is greater than zero, will cause an inevitable positive contribution to $\mu_{3,n}(\tau)$. We conclude that optimal control of the detailed model cannot yield less $\mu_{3,n}(\tau_f)$ than optimal control of the simplified model can do. Consequently,

$$\mu_{3,n}^*(\tau_f) \le \mu_{3,n}^{**}(\tau_f) \le \mu_{3,n}'(\tau_f) \quad (33)$$

We define $\varepsilon_{\max} := \mu_{3,n}'(\tau_f) - \mu_{3,n}^*(\tau_f)$ as the maximum amount of $\mu_{3,n}$ that could be avoided by computing the optimal control for the detailed model, over computing the optimal control for the simplified model and applying the proposed error correction scheme.

5 Case study

The parameters and initial conditions for the case study are adapted from [5], [11] and are given in Table 1. The final time constraint is set to $t_{f,c} = 100$ min and the control constraint is set to $\left(\frac{1}{G}\right)_{\min} = 5 \frac{\min}{mm}$. In order to show the impact of the error caused by simplification, different values were identified for the constraint $\mu_{3f,s,c}$, that let specific final ratios $\text{imp}_f' = \mu_{3f,n}'/\mu_{3f,s,c}$ result from optimal control of the simplified model and application of the proposed error correction scheme. We were able to find these values for $\mu_{3f,s,c}$ by repeating the optimization a number of times and using a bisection scheme.

Figures 1 to 4 show optimal trajectories for the simplified model, as well as the results of the application of the error correction scheme, for $\text{imp}_f' = 0.33$. Furthermore, results are compared to those obtained with constant supersaturation control, denoted by S_{cst}. The

Table 1: Case study

Process parameters (KNO$_3$, water):		
$k_b = 3.47 \cdot 10^4 \frac{1}{\text{mm}^3\text{min}}$; $b = 1.78$; $k_g = 6.97\frac{\text{mm}}{\text{min}}$; $g = 1.32$		
$a_0 = 0.1286$; $a_1 = 5.88 \cdot 10^{-3} \frac{1}{\circ C}$; $a_2 = 1.721 \cdot 10^{-4} \frac{1}{\circ C^2}$		
$\rho = 2.11\frac{\text{g}}{\text{cm}^3}$; $k_V = 1$		
Initial conditions:		
$f_s(\ell, 0) = N_0 \cdot \delta(\ell - \ell_0)$; $\ell_0 = 200\mu\text{m}$; $\mu_{i0,s} = N_0\ell_0^i$		
$m_{s,0} = \rho_c k_v \mu_{30,s} = 0.5$ g $\to \mu_{30,s} = 237.0$ mm^3		
$m_W = 1650.1$ g; $c_0 = 0.493 \to m_{l,0} = 813.5$ g;		

Table 2: Results obtained by optimization and subsequent application of the error correction scheme ($'$) and with constant supersaturation control (S_{cst}), with error bounds $\varepsilon_{\max}$, for different ratios imp_f'

imp_f'	[%]	1	5	10	20	33
$\mu_{3f,s,c}$	[mm^3]	1344	2665	3768	5471	7251
$\mu_{3f,n}'$	[mm^3]	13.4	133.2	376.8	1094.1	2392.8
$\mu_{3f,n}^{S_{\text{cst}}}$	[mm^3]	15.3	148.3	415.7	1196.8	2603.6
$\frac{\mu_{3f,n}^{S_{\text{cst}}} - \mu_{3f,n}'}{\mu_{3f,n}'}$	[%]	13.7	11.3	10.3	9.4	8.8
$\frac{\varepsilon_{\max}}{\mu_{3f,n}'}$	[%]	0.05	0.33	0.73	1.58	2.73
$\frac{\varepsilon_{\max}}{\mu_{3f,n}^{S_{\text{cst}}} - \mu_{3f,n}'}$	[%]	0.38	2.92	7.08	16.87	30.99

constant S can be chosen so as to meet the constraints $t_{f,c}$ and $\mu_{3f,s,c}$. The control $T^{S_{\text{cst}}}(\tau)$, respectively $T^{S_{\text{cst}}}(t)$, which realizes the constant S, can then be determined with the help of the detailed model.

Table 2 contains the results for various imp_f'. The relative amount of $\mu_{3,n}$ is given that can be avoided by using the approximately optimal control $T'(\tau)$, respectively $T'(t)$, over using constant supersaturation control. The improvement is rather small, around 10 percent. The error bound $\varepsilon_{\max} = \mu_{3f,n}' - \mu_{3f,n}^*$ is given relative to $\mu_{3f,n}'$ as well as relative to the improvement over a constant supersaturation trajectory.
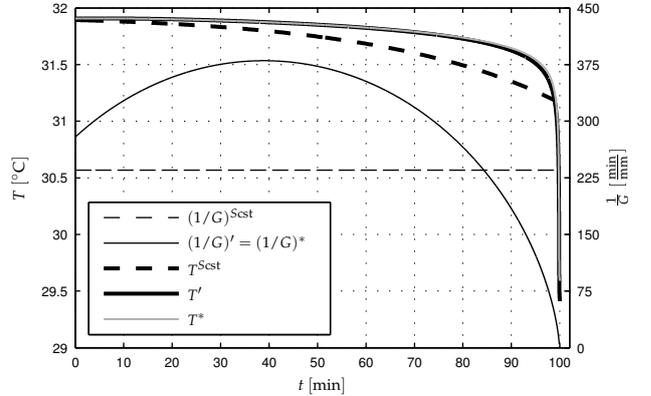

Figure 1: Control in t-domain
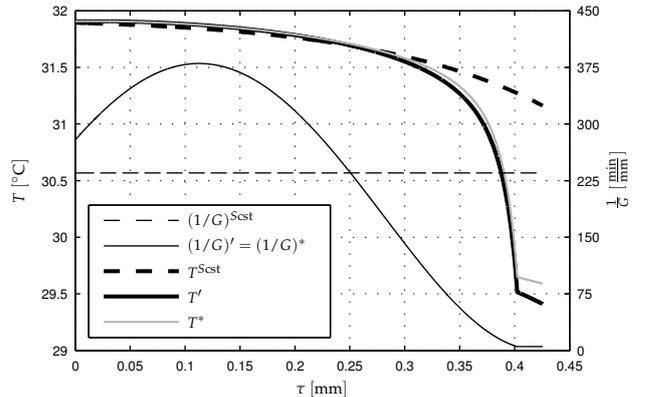
Figure 2: Control in τ-domain

66
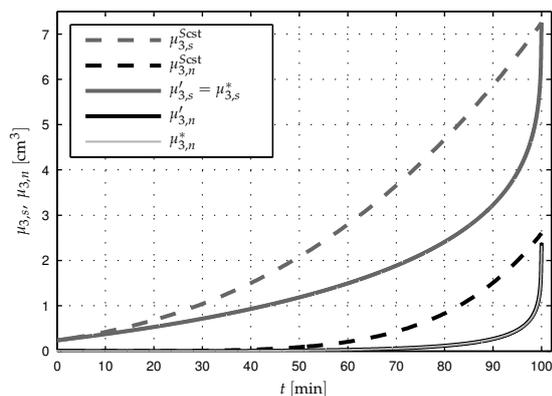
Figure 3: Trajectories of the third moments
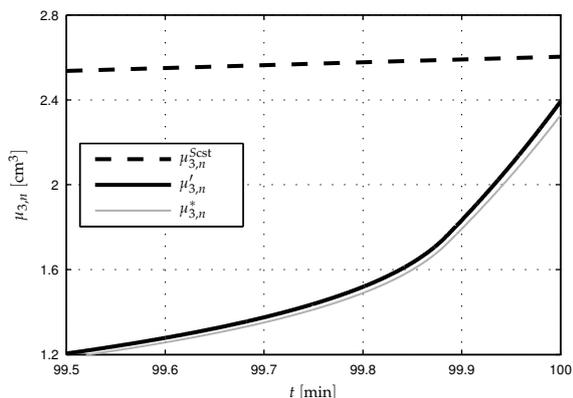
Figure 4: Magnification of Figure 3

These results confirm that, the smaller the relative amount of the third moment of the nucleated CSD gets, the better the simplified model is suited for finding an approximate solution to the optimal control problem. Other tests have shown that the relative errors stay very similar for a wide range of final time constraints $t_{f,c}$.

As in previous results, the computed optimal trajectories exhibit a steep drop in temperature at the end, corresponding to a surge in supersaturation. This effect can also be seen on the curves in τ-domain, which are "stretched" near τ_f, compared to the curves in t-domain. Crystals that nucleate as τ approaches τ_f do not attain a high final length (only $\tau_f - \tau$). Therefore, nucleation and growth can be greatly increased toward τ_f. Also, in τ-domain, the interval with constrained input $u^* = \left(\frac{1}{G}\right)_{\min}$ is clearly visible.

The implementation was done in Matlab^{TM}. It uses equidistant discretization of the interval $[0, \tau_f]$ into ten thousand steps and trapezoidal integration. The computing time for a single solution of the optimal control problem for the simplified model was about 0.04 s on one core of an AMD Athlon^{TM} 64 X2 Dual Core Processor 3800+ running at 1 Ghz. The application of the error correction scheme took about 1 s when using ode45 for the numerical integration of the detailed model.

6 Conclusion

We have presented a new, mostly analytic, solution for the optimal control of a batch crystallizer. The solution is based on a time transformation that turns a moment model of the crystallizer flat, on the Pontryagin Minimum Principle, and on an appropriate, simplifying assumption. It has been argued and shown in a case study, that the error coming from simplification is sufficiently small when the goal is to suppress nucleation. Current work is done on solving variations of the optimal control problem as well as incorporating constraints on temperature rather than on growth rate or supersaturation.

References

[1] J. W. Mullin and J. Nyvlt, *Programmed cooling of batch crystallizers.* Chem. Eng. Sci. 26:3 (1971), pp. 369-377.

[2] A. G. Jones, *Optimal Operation of a Batch Cooling Crystallizer.* Chem. Eng. Sci. 29 (1974), pp. 1075-1087.

[3] M. B. Ajinkya and W. H. Ray, *On the Optimal Operation of Crystallization Processes.* Chem. Eng. Commun. 1:4 (1974), pp. 181-186.

[4] M. Morari, *Some Comments on the Optimal Operation of Batch Crystallizers.* Chem. Eng. Commun. 4:1-3 (1980), pp. 167-171.

[5] S. M. Miller and J. B. Rawlings, *Model identification and control strategies for batch cooling crystallizers.* AIChE J. 40:8 (1994), pp. 1312-1327.

[6] Y. Lang, A. M. Cervantes and L. T. Biegler, *Dynamic Optimization of a Batch Cooling Crystallization Process.* Ind. Eng. Chem. Res. 38:4 (1999), pp. 1469-1477.

[7] J. P. Corriou and S. Rohani, *A new look at optimal control of a batch crystallizer.* AIChE J. 54:12 (2008), pp. 3188-3206.

[8] G. P. Zhang and S. Rohani, *On-line optimal control of a seeded batch cooling crystallizer.* Chem. Eng. Sci. 58:9 (2003), pp. 1887-1896.

[9] Z. K. Nagy and R. D. Braatz, *Robust nonlinear model predictive control of batch processes.* AIChE J. 49:7 (2003), pp. 1176-1786.

[10] E. Visser, B. Srinivasan, S. Palanki and D. Bonvin, *A Feedback-Based Implementation Scheme for Batch Process Optimization.* J. of Proc. Cont. 10:5 (2000), pp. 399-410.

[11] U. Vollmer and J. Raisch, *Control of batch cooling crystallisers based on orbital flatness.* Int. J. of Cont. 76:16 (2003), pp. 1635-1643.

[12] Q. Hu, S. Rohani, D. X. Wang and A. Jutan, *Optimal Control of a Batch Cooling Seeded Crystallizer.* Pow. Tech. 156:2-3 (2005), pp. 170-176.

[13] Q. Hu, S. Rohani and A. Jutan, *Modelling and optimization of seeded batch crystallizers.* Comp. and Chem. Eng. 29:4 (2005), pp. 911-918.

[14] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes.* John Wiley & Sons (1962).

[15] D. E. Kirk, *Optimal Control Theory.* Prentice-Hall (1970).

The setpoint overshoot method: A super-fast approach to PI tuning

Mohammad Shamsuzzoha[a],*Sigurd Skogestad[a], **Ivar J. Halvorsen[b]**

[a]Norwegian University of Science and Technology,
*Trondheim, Norway, (shamsuzz@chemeng.ntnu.no), (*skoge@ntnu.no)*
[b]SINTEF ICT, Applied Cybernetics, N-7465 Trondheim, Norway

Abstract: A simple method has been developed for PI controller tuning of an unidentified process using closed-loop experiments. The proposed method is similar to the Ziegler-Nichols (1942) tuning method, but it is faster to use and does not require the system to approach instability with sustained oscillations. The method requires one closed-loop step setpoint response experiment using a proportional only controller with gain K_{c0}. From the setpoint response one observes the overshoot, the corresponding time to reach the peak (t_p) and the steady-state change ($b=y(\infty)/y_s$). Based on simulations for a range of first-order with delay processes, simple correlations have been derived to give PI controller settings similar to those of the SIMC tuning rules (Skogestad, 2003). The controller gain (K_c/K_{c0}) is only a function of the overshoot observed in the setpoint experiment whereas the controller integral time (τ_I) is mainly a function of the time to reach the peak (t_p). Importantly, the method includes a detuning factor F that allows the user to adjust the final closed-loop response time and robustness. The proposed tuning method, originally derived for first-order with delay processes, has been tested on a wide range of other processes typical for process control applications and the results are comparable with the SIMC tunings using the open-loop model.

Keywords: PI controller, step test, closed-loop response, IMC, overshoot

1. INTRODUCTION

The proportional integral (PI) controller is widely used in the process industries due to its simplicity, robustness and wide ranges of applicability in the regulatory control layer. On the basis of a survey of more than 11 000 controllers in process industries, Desborough and Miller (2002) have reported that more than 97% of regulatory controllers utilise the PID algorithm. A recent survey (Kano and Ogawa; 2009) from Japan shows that the ratio of applications of PID control, conventional advanced control (feedforward, ratio, valve position control, etc.) and model predictive control is about 100:10:1. In addition, the vast majority of the PID controllers do not use derivative action. Even though the PI controller only has two adjustable parameters, they are often poorly tuned. One reason is that quite tedious plant tests may be needed to obtain improved controller setting. The objective of this paper is to derive a method which is simpler to use than the present ones.

To obtain the information required for tuning the controller one may use open-loop or closed-loop plant tests. Most tuning approaches are based on open-loop plant information; typically the plant's gain (k), time constant (τ) and time delay (θ). One popular approach is direct synthesis (Seborg et al., 2004) which includes the IMC-PID tuning method of Rivera et al. (1986). The original direct synthesis approaches give very good performance for setpoint changes but give sluggish responses to input (load) disturbances for lag-dominant (including integrating) processes with τ/θ larger than about 10. To improve load disturbance rejection, Skogestad (2003) proposed the modified SIMC method where the integral time is reduced for processes with a large value of the time constant τ. The SIMC rule has one tuning parameter, the closed-loop time constant τ_c, and for "fast and robust" control

is recommended to choose $\tau_c = \theta$, where θ is the (effective) time delay. However, these approaches require that one first obtains an open-loop model of the process. There are two problems here. First, an open-loop experiment, for example a step test, is normally needed to get the required process data. This may be time consuming and may upset the process and even lead to process runaway. Second, approximations are involved in obtaining the process parameters (e.g., k, τ and θ) from the data.

The main alternative is to use closed-loop experiments. One approach is the classical method of Ziegler-Nichols (1942) which requires very little information about the process. However, there are several disadvantages. First, the system needs to be brought its limit of instability and a number of trials may be needed to bring the system to this point. To avoid this problem one may induce sustained oscillation with an on-off controller using the relay method of Åström and Hägglund, (1984). However, this requires that the feature of switching to on/off-control has been installed in the system. Another disadvantage is that the Ziegler-Nichols (1942) tunings do not work well on all processes. It is well known that the recommended settings are quite aggressive for lag-dominant (integrating) processes (Tyreus and Luyben, 1992) and quite slow for delay-dominant process (Skogestad, 2003). A third disadvantage is of the Ziegler-Nichols (1942) method is that it can only be used on processes for which the phase lag exceeds -180 degrees at high frequencies. For example, it does not work on a simple second-order process.

Therefore, there is need of an alternative closed-loop approach for plant testing and controller tuning which avoids the instability concern during the closed-loop experiment, reduces the number of trails, and works for a wide range of processes. The proposed new method satisfies the above

concerns: In summary, the proposed method is simpler in use than existing approaches and allows the process to be kept under closed-loop control.

An obvious alternative to the proposed method is a two-step procedure where one first identifies an open-loop model from the closed-loop setpoint experiment, and then obtains the PI or PID controller using standard tuning rules (e.g., the SIMC rules of Skogestad, 2003). This approach was used by Yuwana and Seborg (1982). We found that this two-step approach gives result comparable or slightly inferior (Shamsuzzoha and Skogestad, 2010) to the more direct approach proposed in this paper by using the SIMC method. In addition, the proposed approach avoids the extra step of obtaining the process parameters (k, τ, θ) and is therefore simpler to use.

2. SIMC PI TUNING RULES

In process control, a first-order process with time delay is a common representation of the process dynamics:

$$g(s) = \frac{ke^{-\theta s}}{\tau s + 1} \tag{1}$$

Here k is the process gain, τ the dominant lag time constant and θ the effective time delay. Most processes in the chemical industries can be satisfactorily controlled using a PI controller:

$$c(s) = K_c\left(1 + \frac{1}{\tau_I s}\right) \tag{2}$$

The PI controller has two adjustable parameters, the proportional gain K_c and the integral time τ_I. The ratio $K_I = K_c/\tau_I$ is known as the integral gain.

The SIMC tuning rule is widely used in the process industry and for the process in Eq. (1) is given as:

$$K_c = \frac{\tau}{k(\tau_c + \theta)} \tag{3}$$

$$\tau_I = \min\{\tau, 4(\tau_c + \theta)\} \tag{4}$$

Note that the original IMC tuning rule (Rivera et al., 1986) always uses $\tau_I = \tau$, but the SIMC rule increases the integral contribution for close-to integrating processes (with τ large) to avoid poor performance (slow settling) to load disturbance. There is one adjustable tuning parameter, the closed-loop time constant (τ_c), which is selected to give the desired trade-off between performance and robustness. Initially, this study is based on the "fast and robust" setting $\tau_c = \theta$, which gives a good trade-off between performance and robustness. In terms of robustness, this choice gives a gain margin is about 3 and a sensitivity peak (M_s-value) of about 1.6. On dimensionless form, the SIMC tuning rules with $\tau_c = \theta$ become

$$K_c' = kK_c = 0.5\frac{\tau}{\theta} \tag{5}$$

$$K_I' = \frac{kK_c}{\tau_I/\theta} = \max\left(0.5, \frac{1}{16}\frac{\tau}{\theta}\right) \tag{6}$$

The dimensionless gains K_c' and K_I' are plotted as a function of τ/θ in Fig. 1. We note that the integral term (K_I') is most important for delay dominant processes ($\tau/\theta < 1$), while the proportional term K_c' is most significant for processes with a smaller time delay. These insights are useful for the next step when we want to derive tuning rules based on the closed-loop setpoint response.

3. CLOSED-LOOP EXPERIMENT

As mentioned earlier, the objective is to base the controller tuning on closed-loop data. The simplest closed-loop experiment is probably a setpoint step response (Fig. 2) where one maintains full control of the process, including the change in the output variable. The simplest to observe is the time t_p to reach the (first) overshoot and its magnitude, and this information is therefore the basis for the proposed method.

We propose the following procedure:

1. Switch the controller to P-only mode (for example, increase the integral time τ_I to its maximum value or set the integral gain K_I to zero). In an industrial system, with bumpless transfer, the switch should not upset the process.

2. Make a setpoint change that gives an overshoot between 0.10 (10%) and 0.60 (60%); about 0.30 (30%) is a good value. Record the controller gain K_{c0} used in the experiment. Most likely, unless the original controller was quite tightly tuned, one will need to increase the controller gain to get a sufficiently large overshoot.

Note that small overshoots (less than 0.10) are not considered because it is difficult in practice to obtain from experimental data accurate values of the overshoot and peak time if the overshoot is too small. Also, large overshoots (larger than about 0.6) give a long settling time and require more excessive input changes. For these reasons we recommend using an "intermediate" overshoot of about 0.3 (30%) for the closed-loop setpoint experiment.

3. From the closed-loop setpoint response experiment, obtain the following values (see Fig. 2):
   - Fractional overshoot, $(\Delta y_p - \Delta y_\infty)/\Delta y_\infty$
   - Time from setpoint change to reach peak output (overshoot), t_p
   - Relative steady state output change, $b = \Delta y_\infty/\Delta y_s$.

Here the output variable changes are:

Δy_s: Setpoint change
Δy_p: Peak output change (at time t_p)
Δy_∞: Steady-state output change after setpoint step test

To find Δy_∞ one needs to wait for the response to settle, which may take some time if the overshoot is relatively large (typically, 0.3 or larger). In such cases, one may stop the experiment when the setpoint response reaches its first minimum and record the corresponding output, Δy_u.

$$\Delta y_\infty = 0.45(\Delta y_p + \Delta y_u) \tag{7}$$

4. CORRELATION BETWEEN SETPOINT RESPONSE AND SIMC-SETTINGS

The objective of this paper is to provide a more direct approach similar to the Ziegler-Nichols (1942) closed-loop method. Thus, the goal is to derive a correlation, preferably as simple as possible, between the setpoint response data (Fig. 2) and the SIMC PI-settings in Eq. (3) and (4), initially with the choice $\tau_c = \theta$. For this purpose, we considered 15 first-order with delay models $g(s) = ke^{-\theta s}/(\tau s + 1)$ that cover a wide range of processes; from delay-dominant to lag-dominant (integrating):

τ/θ=0.1,0.2,0.4,0.8,1.0,1.5,2.0,2.5,3.0,7.5,10.0,20.0,50.0,100

Since we can always scale time with respect to the time delay (θ) and since the closed-loop response depends on the product of the process and controller gains (kK_c) we have without loss of generality used in all simulations k=1 and θ=1.
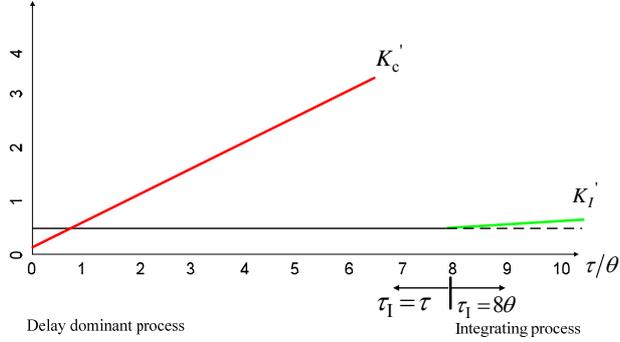

Fig. 1. Scaled proportional and integral gain for SIMC tuning rule.
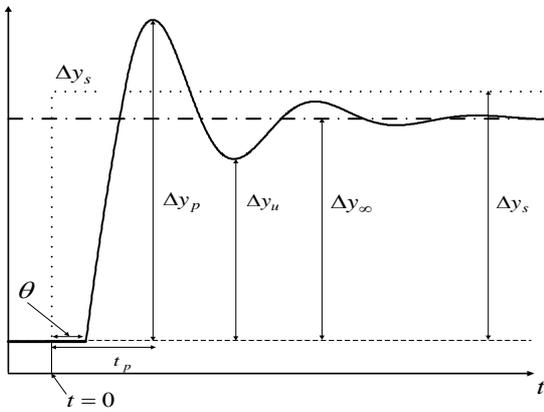
Fig. 2. Closed-loop step setpoint response with P-only control.

For each of the 15 process models (values of τ/θ), we obtained the SIMC PI-settings (K_c and τ_I) using Eqs. (3) and (4) with the choice $\tau_c=\theta$. Furthermore, for each of the 15 processes we generated 6 closed-loop step setpoint responses using P-controllers that give different fractional overshoots.

Overshoot= 0.10, 0.20, 0.30, 0.40, 0.50 and 0.60

In total, we then have 90 setpoint responses, and for each of these we record four data: the P-controller gain K_{c0} used in the experiment, the fractional overshoot, the time to reach the overshoot (t_p), and the relative steady-state change, b = $\Delta y_\infty/\Delta y_s$.

Controller gain (K_c). We first seek a relationship between the above four data and the corresponding SIMC-controller gain K_c. Indeed, as illustrated in Fig. 3, where we plot kK_c (SIMC PI-controller for the corresponding process) as a function of kK_{c0} for the 90 setpoint experiments, the ratio K_c/K_{c0} is approximately constant for a fixed value of the overshoot, independent of the value of τ/θ. Thus, we can write

$$\frac{K_c}{K_{c0}}=A \qquad (8)$$

where the ratio A is a function of the overshoot only. In Fig. 4 we plot the value of A, which is obtained as the best fit of the slopes of the lines in Fig. 3, as a function of the overshoot. The following equation (solid line in Fig. 4) fits the data in Fig. 3 well,

$$A=[1.152(overshoot)^2 -1.607(overshoot)+1.0] \qquad (9)$$
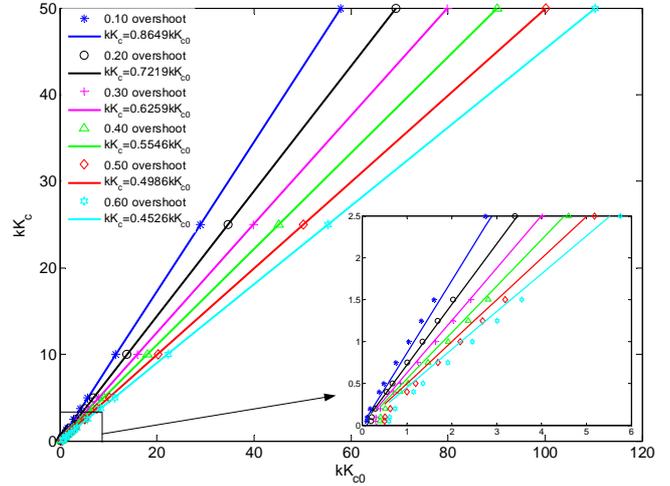
Fig. 3. Relationship between P-controller gain kK_{c0} used in setpoint experiment and corresponding SIMC controller gain kK_c.

Actually, a closer look at Fig. 3 reveals that a constant slope, use of Eq.(8) and (9), only fits the data well for $K_c'=kK_c$ greater than about 0.5. Fortunately, a good fit of the controller gain K_c is not so important for delay-dominant processes ($\tau/\theta<1$) where $K_c'<0.5$, because we recall from the discussion of the SIMC rules (Fig. 1) that the integral gain K_I is more important for such processes. This is discussed in more detail below.

Integral time (τ_I). Next, we want to find a simple correlation for the integral time. Since the SIMC tuning formula in Eq. (4) uses the minimum of two values, it seems reasonable to look for a similar relationship, that is, to find one value ($\tau_{I1}=\tau$) for processes with a relatively large delay, and another value ($\tau_{I2}=8\theta$) for processes with a relatively small delay including integrating processes.

(1) Process with relatively large delay. For processes with a relatively large delay $\tau/\theta<8$ or $\theta>\tau/8$, the SIMC-rule is to use $\tau_I = \tau$. Inserting $\tau = \tau_I$ into the SIMC rule for K_c in Eq. (5) and solving for τ_I gives:

$$\tau_{I1}=2kK_c\theta \qquad (10)$$

As just mentioned, for processes with a relatively large delay it is the integral gain $K_I=K_c/\tau_I$ that matters most (Fig. 1) and to avoid that any error in K_c originating from our correlation Eq.(8) propagates into K_I, we should in Eq. (10) use $K_c = K_{c0}A$, where A is given as a function of the overshoot in Eq. (9). In (10), we also need the value of the process gain k, and to this effect write

$$kK_c= kK_{c0}.K_c/ K_{c0} \qquad (11)$$

Here, the value of the loop gain kK_{c0} for the P-control setpoint experiment is given from the value of b:

$$kK_{c0}=\left|\frac{b}{(1-b)}\right| \tag{12}$$
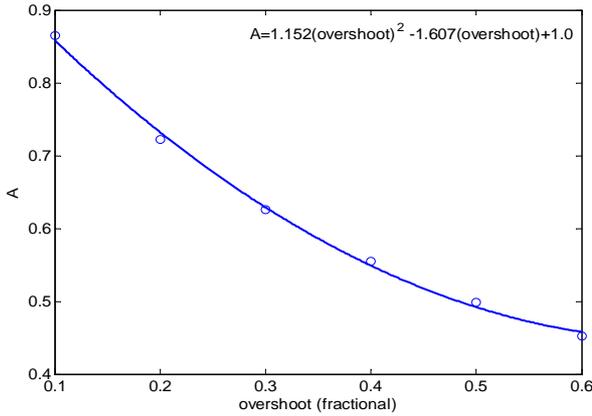
Fig. 4. Variation of A with overshoot using data (slopes) from Fig. 3.

To prove this, the closed-loop setpoint response is $\Delta y/\Delta y_s = gc/(1+gc)$ and with a P-controller with gain K_{c0}, the steady-state value is $\Delta y_\infty/\Delta y_s = kK_{c0}/(1+kK_{c0})=b$ and we derive Eq.(12). The absolute value is included to avoid problems if $b>1$, as may occur for an unstable process or because of inaccurate data.

In summary, we have derived following expression for τ_I for a delay-dominant process:

$$\tau_{I1}=2A\left|\frac{b}{(1-b)}\right|\theta \tag{13}$$

One could obtain the **effective time delay** θ directly from the closed-loop setpoint response, but this is generally not easy. Fortunately, as shown in Fig. (5), there is a reasonably good correlation between θ and the setpoint peak time t_p which is easier to observe. For processes with a relatively large time delay ($\tau/\theta<8$), the ratio θ/t_p varies between 0.27 (for $\tau/\theta=8$ with overshoot=0.1) and 0.5 (for $\tau/\theta=0.1$ with all overshoots). For the intermediate overshoot of 0.3, the ratio θ/t_p varies between 0.32 and 0.50. A conservative choice would be to use $\theta=0.5t_p$ because a large value increases the integral time. However, to improve performance for processes with smaller time delays, we propose to use $\theta=0.43t_p$ which is only 14% lower than 0.50 (the worst case).

In summary, we have for process with a relatively large time delay:

$$\tau_{I1}=0.86\ A\left|\frac{b}{(1-b)}\right|t_p \tag{14}$$

(2) Process with relatively small delay. For a lag-dominant (including integrating) process with $\tau/\theta>8$ the SIMC rule gives

$$\tau_{I2}=8\theta \tag{15}$$

For $\tau/\theta>8$ we see from Fig. (5) that the ratio θ/t_p varies between 0.25 (for $\tau/\theta=100$ with overshoot=0.1) and 0.36 (for $\tau/\theta=8$ with overshoot 0.6). We select to use the average value $\theta=0.305t_p$ which is only 15% lower than 0.36 (the worst case). Also note that for the intermediate overshoot of 0.3, the ratio θ/t_p varies between 0.30 and 0.32. In summary, we have for a lag-dominant process

$$\tau_{I2}=2.44t_p \tag{16}$$

Conclusion. The integral time τ_I is obtained as the minimum of the above two values:

$$\tau_I=\min\left(0.86A\left|\frac{b}{(1-b)}\right|t_p,\ 2.44t_p\right) \tag{17}$$
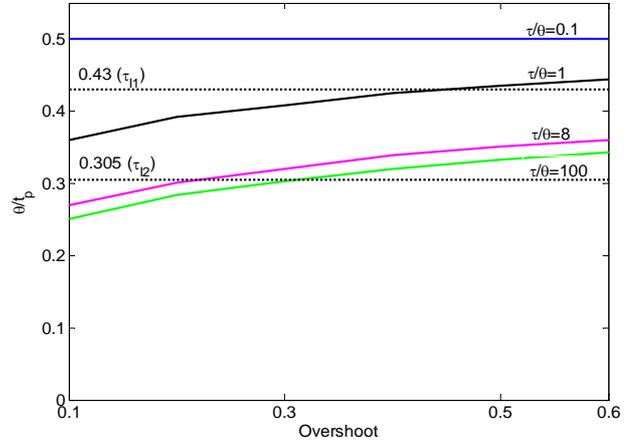
Fig. 5. Ratio between delay and setpoint overshoot peak time (θ/t_p) for P-only control of first-order with delay processes (solid lines); Dotted lines: values used in final correlations.

5. ANALYSIS AND SIMULATION

Closed-loop simulations have been conducted for 7 different processes and the proposed tuning procedure provides in all cases acceptable controller settings with respect to both performance and robustness. For each process, PI-settings are obtained based on step response experiments with three different overshoot (around 0.1, 0.3 and 0.6) and are compared with the SIMC settings.

The closed-loop performance is evaluated by introducing a unit step change in both the set-point and load disturbance i.e, ($y_s=1$ and $d=1$). To evaluate the robustness, the maximum sensitivity, M_s, defined as $M_s=\max\left|1/[1+gc(i\omega)]\right|$, is used. Since M_s is the inverse of the shortest distance from the Nyquist curve of the loop transfer function to the critical point $(-1, 0)$, a small M_s-value indicates that the control system has a large stability margin.

The results for the 7 example processes, which include the different types of the process mainly stable, integrating and unstable plant dynamics, are listed in Table 1.

All results are without detuning ($F=1$). The complete simulation results with additional examples are available in a technical report (Shamsuzzoha and Skogestad, 2010). As expected, when the method is tested on first-order plus delay processes, similar to those used to develop the method, the responses are similar to the SIMC-responses, independent of the value of the overshoot. Typical cases are E1, E2 (pure time delay) and E3 (integrating with delay); see Figs. 6-8.

For models that are *not* first-order plus delay (typical cases are E4, E5 and E6, see Fig. 9 for E6 only), the agreement with the SIMC-method is best for the intermediate overshoot (around 0.3). A small overshoot (around 0.1) typically give

"slower" and more robust PI-settings, whereas a large overshoot (around 0.6) gives more aggressive PI-settings. In some sense this is good, because it means that a more "careful" step response results in more "careful" tunings. Also note that the user always has the option to use the detuning factor F to correct the final tunings. Case E7 (Fig. 10) illustrates that the method works well for a simple unstable process with delay.
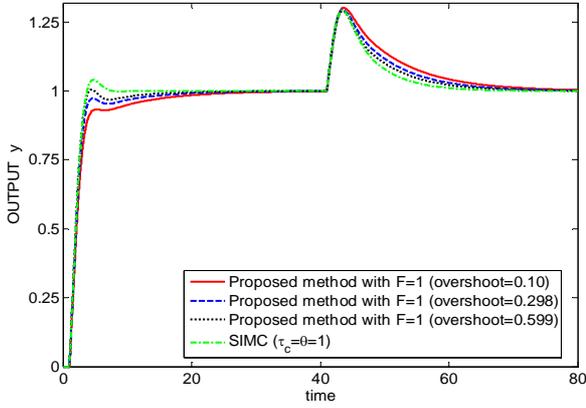
Fig. 6. Responses for PI-control of $g = e^{-s}/(5s+1)$ (E1).
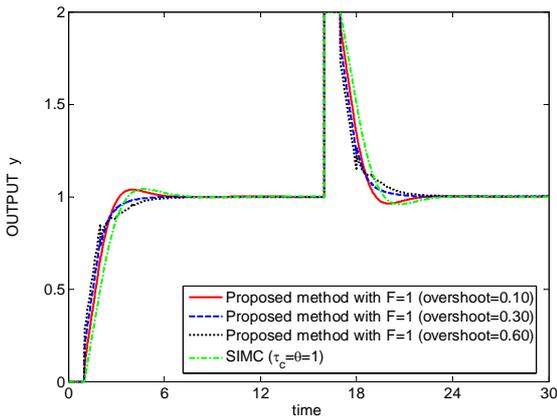
Fig. 7 Responses for PI-control of $g = e^{-s}$ (E2).
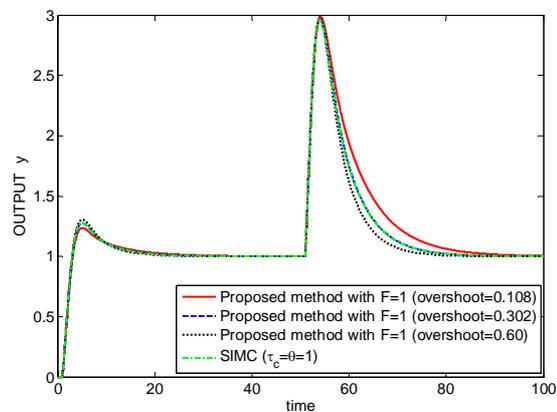
Fig. 8. Responses for PI-control of $g = e^{-s}/s$ (E3).
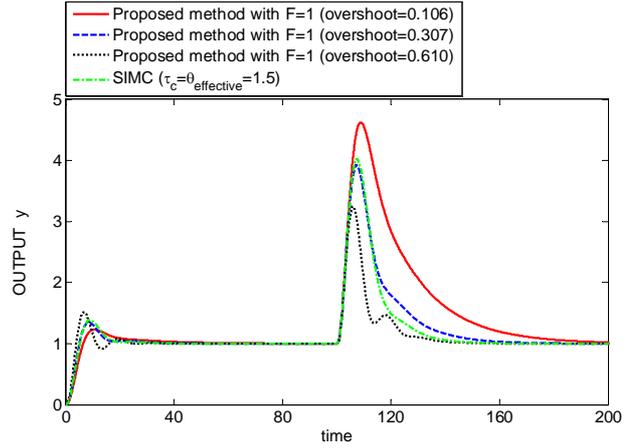
Fig. 9. Responses for PI-control of $g = 1/\left[s(s+1)^2 \right]$ (E6).
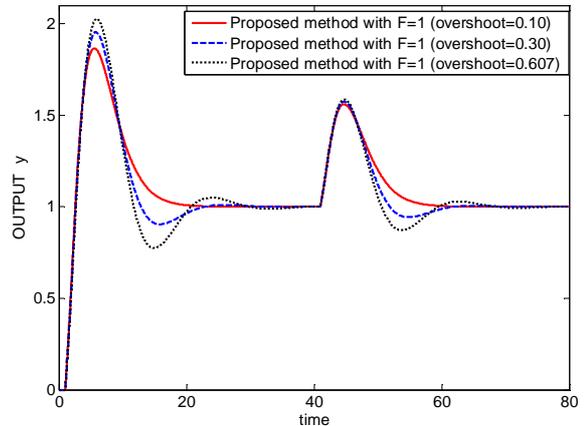
Fig. 10. Responses for PI-control of $g = e^{-s}/(5s-1)$ (E7).

6. CONCLUSION

A simple and new approach for PI controller tuning has been developed. It is based on a single closed-loop setpoint step experiment using a P-controller with gain K_{c0}. The PI-controller settings are then obtained directly from following three data from the setpoint experiment:

- Overshoot, $(\Delta y_p - \Delta y_\infty)/\Delta y_\infty$
- Time to reach overshoot (first peak), t_p
- Relative steady state output change, $b = \Delta y_\infty / \Delta y_s$.

If one does not want to wait for the system to reach steady state, one can use the estimate $\Delta y_\infty = 0.45(\Delta y_p + \Delta y_u)$.

The proposed tuning formulas for the proposed "**Setpoint Overshoot Method**" method are:

$$K_c = K_{c0}A/F$$

$$\tau_I = \min\left(0.86A \left| \frac{b}{(1-b)} \right| t_p, \ 2.44 t_p F \right)$$

where, $A = \left[1.152(\text{overshoot})^2 - 1.607(\text{overshoot}) + 1.0 \right]$

The factor F is a tuning parameter and F=1 gives the "fast and robust" SIMC settings corresponding to $\tau_c = \theta$. To detune the response and get more robustness one selects F>1, but in special cases one may select F<1 to speed up the closed-loop response.

The Setpoint Overshoot Method works well for a wide variety of the processes typical for process control, including the standard first-order plus delay processes as well as

integrating, high-order, inverse response, unstable and oscillating process.

We believe that the proposed method is the simplest and easiest approach for PI controller tuning available and should be well suited for use in process industries.

Rivera, D. E., Morari, M., Skogestad, S. (1986). Internal model control. 4. PID controller design, *Ind. Eng. Chem. Res.,* 25 (1) 252–265.

Seborg, D. E., Edgar, T. F., Mellichamp, D. A., (2004). Process Dynamics and Control, 2nd ed., John Wiley & Sons, New York, U.S.A.

Table 1: PI controller setting for proposed method and comparison with SIMC method ($\tau_c=\theta_{effective}$)

Case	Process model	k_{c0}	overshoot	t_p	b	k_c	τ_I	M_s
E1	$\dfrac{e^{-s}}{(5s+1)}$	2.75	0.10	3.60	0.733	2.338	7.240	1.50
		4.0	0.298	3.049	0.80	2.494	6.538	1.56
		5.75	0.599	2.705	0.852	2.592	6.030	1.60
		SIMC	-	-	-	2.50	5.0	1.59
E2	e^{-s}	0.10	0.10	2.0*	0.091	0.085	0.146	1.60
		0.30	0.30	2.0	0.231	0.187	0.321	1.53
		0.60	0.60	2.0	0.270	0.465	0.375	1.59
		SIMC	-	-	-	$k_c/\tau_I=0.50$		1.59
E3	$\dfrac{e^{-s}}{s}$	0.59	0.108	3.976	1.0	0.495	9.702	1.67
		0.80	0.302	3.282	1.0	0.496	8.008	1.70
		1.10	0.60	2.909	1.0	0.496	7.098	1.72
		SIMC	-	-	-	0.50	8.0	1.70
E4	$\dfrac{(s^2+2s+9)(-2s+1)(s+1)e^{-2s}}{(s^2+0.5s+1)(5s+1)^2}$	0.07	0.112	18.132	0.387	0.058	8.198	1.46
		0.12	0.301	15.043	0.519	0.074	8.667	1.61
		0.18	0.583	13.71	0.618	0.082	8.684	1.70
		SIMC	-	-	-	-	-	-
E5	$\dfrac{1}{(s+1)(0.2s+1)}$	5.0	0.127	0.710	0.833	4.074	1.732	1.33
		15.0	0.322	0.393	0.937	9.031	0.958	1.74
		40.0	0.508	0.230	0.976	19.23	0.561	2.62
		SIMC	-	-	-	5.5	0.80	1.56
E6	$\dfrac{1}{s(s+1)^2}$	0.32	0.106	8.985	1.0	0.270	21.923	1.51
		0.58	0.307	6.188	1.0	0.357	15.10	1.75
		1.15	0.610	4.492	1.0	0.516	10.961	2.30
		SIMC	-	-	-	0.330	12.0	1.76
E7	$\dfrac{e^{-s}}{(5s-1)}$	3.10	0.10	4.647	1.476	2.636	10.54	2.12
		4.0	0.30	3.671	1.333	2.487	7.852	2.33
		5.30	0.607	3.164	1.233	2.379	6.475	2.67
		SIMC	-	-	-	-	-	-

* For the pure time delay case (E4) use the end time of the peak (or add a small time constant to get t_p in simulation).

REFERENCES

Åström, K. J., Hägglund, T. (1984). Automatic tuning of simple regulators with specifications on phase and amplitude margins, *Automatica,* (20), 645–651.

Desborough, L. D., Miller, R. M. (2002). Increasing customer value of industrial control performance monitoring—Honeywell's experience. Chemical Process Control–VI (Tuscon, Arizona, Jan. 2001), AIChE Symposium Series No. 326. Volume 98, USA.

Kano, M., Ogawa, M. (2009). The state of art in advanced process control in Japan, IFAC symposium ADCHEM 2009, Istanbul, Turkey.

Shamsuzzoha, M., Skogestad. S. (2010). Report on the setpoint overshoot method (extended version) http://www.nt.ntnu.no/users/skoge/.

Skogestad, S., (2003). Simple analytic rules for model reduction and PID controller tuning, *Journal of Process Control,* 13, 291–309.

Tyreus, B.D., Luyben, W.L. (1992). Tuning PI controllers for integrator/dead time processes, *Ind. Eng. Chem. Res.* 2628–2631.

Yuwana, M., Seborg, D. E., (1982). A new method for on-line controller tuning, *AIChE Journal* 28 (3) 434-440.

Ziegler, J. G., Nichols, N. B. (1942). Optimum settings for automatic controllers. Trans. *ASME*, 64, 759-768.

Comparing PI tuning methods in a real benchmark temperature control system

Finn Haugen[*]

May 27, 2010

Abstract

This paper demonstrates a number of PI controller tuning methods being used to tune a temperature controller for a real air heater. Indexes expressing setpoint tracking and disturbance compensation, and stability margin (robustness) are calculated. From these indexes and a personal impression about how quick a method is to deliver the tuning result and how simple the method is to use, a winning method is identified.

1 Introduction

The PI (proportional plus integral) controller function is the most frequently used controller function in practical applications. The PI controller stems from a PID controller with the D-term (derivative) deactived. The D-term is often deactivated because it amplifies random (high-frequent) measurement noise, causing abrupt variations in the control signal. This paper assumes PI control (not PID).

The continuous-time PI controller function is as follows:

$$u(t) = K_c e(t) + \frac{K_c}{T_i} \int_0^t e(\tau) \, d\tau \quad (1)$$

where u is the control signal (the controller output), $e = r - y$ is the control error, where r is the reference or setpoint and y is the

process output variable (process measurement), K_c is the controller gain, and T_i is the integral time. K_c and T_i are the controller parameters which are to be tuned. In most practical applications the continuous-time PI controller is implemented as a corresponding discrete-time algorithm based on a numerical approximation of the integral term. Typically, the sampling time of the discrete-time controller is so small – compared to the dynamics (response-time or time-constant) of the control system – that there is no significant difference between the behaviour of the continuous-time PI controller and the discrete-time PI controller. Concequently, in this paper the sampling time is not regarded as a tuning parameter.

This paper compares a number of methods for tuning PI controllers using the following measures:

1. Performance related to setpoint tracking and disturbance compensation

2. Robustness against parameter changes in the control loop

3. How quick and simple the method is to use

Numerous studies about simulated control systems exist (O'Dwyer, 2003) (Seborg et.al, 2004). However, in this paper only experiments on a physical system will be used as the basis of the comparison of the tuning methods. The system is a laboratory scale air heater, cf. Section 2. I think it is particularly valuable to see various methods being applied to a physical system because a

[*]Telemark University College, Kjolnes ring 56, 3918 Porsgrunn, Norway. E-mail: finn.haugen@hit.no

1

physical system will always differ – more or less – from the underlying model or assumptions of the controller tuning method. So, applying a method to a physical system is *real* testing. Of course, it would be nice to accomplish such real tests with several different real processes, but that may be the topic of future paper.

This paper contains the following subsequent sections:

- Section 2: The experimental setup

- Section 3: The methods to be compared
  Section 4: Measures to compare the tuning methods

- Section 5: Control tunings and results

- Section 6: Summary and discussion

- Section 7: Conclusions

2   The experimental setup

The physical system used in the experiments is the air heater laboratory station shown in Figure 1. The temperature of the air outlet is controlled by adjusting the control signal to the heater.[1] The fan speed can be adjusted manually with a potensiometer. Changes of the fan speed is used as process disturbance. The voltage drop across the potensiometer is used to represent this disturbance.[2]

Figure 2 shows a block diagram of the temperature control system.

The nominal operating point of the system is temperature at 34 $^{\text{o}}$C and fan speed potensiometer position at 2.4 V (corresponding

Figure 1: The air heater lab station with NI USB-6008 analog I/O device
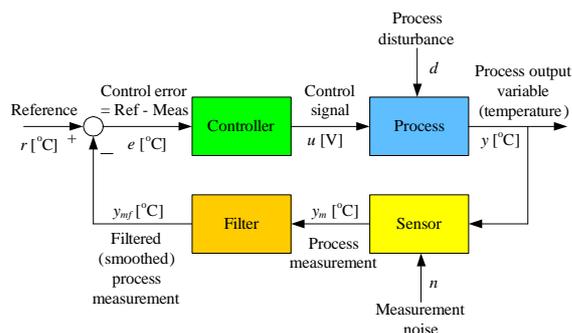
Figure 2: Block diagram of the temperature control system

to a relatively low speed). The measurement filter is a time-constant filter with time-constant 0.5 s. To demonstrate the setpoint tracking the setpoint is changed from 34 to 35 $^{\text{o}}$C, and – thereafter – to demonstrate the disturbance compensation, the fan speed (air flow through the pipe) is changed from minimum (i.e. indicating voltage of 2.4 V) to maximum (5.0 V).

The temperature control system is implemented with National Instruments LabVIEW running on a PC.

---

[1] The supplied power is controlled by an external voltage signal in the range [0 V, 5 V] applied to a Pulse Width Modulator (PWM) which connects/disconnects the mains voltage (220 VAC) to the heater. The temperature is measured with a Pt100 element which in the end provides a voltage measurement signal. The National Instruments USB-6008 is used as analog I/O device. Additional information about the air heater is available at http://home.hit.no/~finnh/air_heater.

[2] The potensiometer voltage is roughly in range 2.4 – 5.0 V, with 2.4 V representing minimum speed.

3 The methods to be compared

In general, both experimental (model-free) and model-based tuning methods are available. In this presentation methods of both these classes will be tested, but among the model-based methods only those methods which that can be applied without automatic system identification are compared. This is because it is my view that system identification tools should not be used unless the user has knowledge about the basic theoretical foundation of such methods and are able to evaluate different estimated models, and few control engineers have such knowledge. In other words: The mathematical model to be used in the tuning method must simple and easy to estimate from experiments, e.g. time-constant with time-delay models.

The methods which will be compared are the following:

Open-loop methods, which are methods based on experiments on the open-loop system (i.e. on the process itself, independent of the controller, which may be present or not):

- Skogestad's Model-based method (or: the SIMC method – Simple Internal Model Control) (Skogestad, 2003)

- Ziegler-Nichols' Process Reaction Curve method (or the Ziegler-Nichols' Open-Loop method) (Ziegler and Nichols, 1942)

- Hägglund and Åstrøm's Robust tuning method (Hägglund and Åstrøm, 2002)

Closed-loop methods, which are methods based on experiments on the already established closed-loop system (i.e. the feedback control system):

- Ziegler-Nichols' Ultimate Gain method (or the Ziegler-Nichols' Closed-Loop method) (Ziegler and Nichols, 1942)

- Tyreus-Luyben's method (which is based on the Ziegler-Nichols' method, but with more conservative tuning), (Luyben and Luyben, 1997)

- Relay method (using a relay function to obtain the sustained oscillations as in the Ziegler-Nichols' method) (Åstrøm and Hägglund, 1995)

- Sham's Setpoint method (based on Skogestad's SIMC method) (Shamsuzzoha et. al., 2010)

- Good Gain method (Haugen, 2010)

Each of these methods are described in their respective subsections of Section 5 of this paper.

The above list of tuning methods contains well-known methods (i.e. often refered to in literature), and also some methods which I personally find interesting.

4 Measures for comparing the tuning methods

The measures for comparing the different methods of PI controller tuning are as follows:

1. **Performance** related to setpoint tracking and disturbance compensation:

   (a) *Setpoint tracking*: The setpoint is changed as a step of amplitude 1, from 34 to 35 $^\circ$C. The IAE (Integral of Absolute Error) index, which is frequently used in the litterature to compare different control functions, is calculated over an interval of 100 sec. The IAE is

   $$IAE = \int_{t_i}^{t_f} |e| \, dt \qquad (2)$$

   where t_i is the initial (or start) time and t_f is the final time. $t_f - t_i = 100$ sec. This IAE index is denoted IAE_s. The less IAE_s value, the better control (assuming that the behaviour of the control signal has no weight).

(b) *Disturbance compensation*: After the temperature has settled at the new setpoint, a disturbance change is applied by adjusting the fan speed voltage from 2.4 (min speed) to 5 V (max speed). Again the IAE index is calculated over an interval of 100 sec. This IAE index is denoted $\mathrm{IAE_d}$.

2. **Robustness** against parameter changes in the control loop is in terms of stability robustness against parameter variations in the control loop. An adjustable gain, K_L, is inserted into the loop (between the controller and the process, in the LabVIEW program). Nominally, $K_L = 1$. For each of the tuning methods, the K_L that brings the control system to the stability limit (i.e. the responses are sustained oscillations) is found experimentally. This K_L value is then the gain margin, ΔK, of the control loop.

It might be interesting also to insert an adjustable time-delay, T_{delay}, into the loop (between the controller and the process, in the LabVIEW program) and find experimentally the time-delay increase in the loop which brings the control system to the stability limit. (This is closely related to finding the phase margin of the control loop in a frequency response analysis.) However, it is assumed the gain margin is suffient to express the stability robustness of the loops in our case, and to simplify the analysis, only the gain margin is considered.

3. **How quick and simple** a given method is to use. It is necessary for a tuning method to be attractive to a user that it gives good results, but also that it is not too complicated to use (i.e. requires lots of calculations) and that the experiments does not take too long time to accomplish. Both the quickness and the simplicity of each of the methods are evaluated with a number ranging from 10 (best) to 0.

5 Controller tunings and results

The subsequent sections describes the controller tuning principle and the actual tuning and results for each of selected tuning methods. The results are summarized in Table 3.

5.1 Skogestad's method

Skogestad's PID tuning method (Skogestad, 2003) (or: the SIMC method – Simple Internal Model Control) is a model-based tuning method where the controller parameters are expressed as functions of the process model parameters. The process model is some continuous-time transfer function. The control system tracking transfer function $T(s)$, which is the transfer function from the setpoint to the (filtered) process measurement, is specified as a first order transfer function with time delay:

$$T(s) = \frac{y_{mf}(s)}{y_{SP}(s)} = \frac{1}{T_C s + 1} e^{-\tau s} \qquad (3)$$

where T_C is the time-constant of the control system which the user must specify, and τ is the process time delay which is given by the process model (the method can however be used for processes without time delay, too). Figure 3 shows the response in y_{mf} after a step in the setpoint y_{SP} for (3).
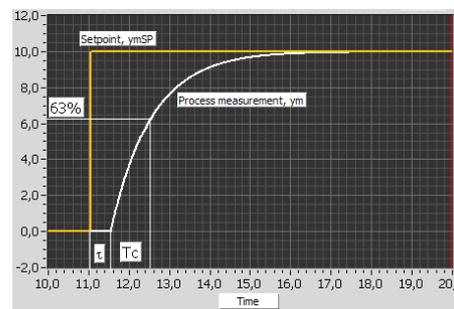

Figure 3: Step response of the specified tracking transfer function (3) in Skogestad's PID tuning method

By equating tracking transfer function given by (3) and the tracking transfer function for the given process, and making some simplifying approximations to the time-delay term, the controller becomes a PID controller or a PI controller for the process transfer function assumed.

The process transfer functions for which Skogestad has calculated tuning formulas includes time-constant with time-delay:

$$H_{psf}(s) = \frac{K}{Ts+1}e^{-\tau s} \qquad (4)$$

which is the model we will use. (The other process models are given in (Skogestad, 2003).) For this process a PI-controller is tuned as follows:[3]

$$K_c = \frac{T}{K(T_C + \tau)} \qquad (5)$$

$$\underline{T_i} = \min\left[T, c(T_C + \tau)\right] \qquad (6)$$

Originally, Skogestad sets the factor c to

$$c = 4 \qquad (7)$$

This gives good setpoint tracking. But the disturbance compensation may become quite sluggish (e.g. in integrator with time-delay processes). In most control loops the disturbance compensation is the most important task for the controller. To obtain faster disturbance compensation, you can try e.g. $c = 2$. The drawback of such a reduction of c is that there will be somewhat more overshoot in the setpoint step response, and that the stability of the control loop will be somewhat reduced (the stability margins will be reduced). Both values of c (4 and 2) will be tried in this paper.

Skogestad suggests setting the closed-loop system time-constant to

$$T_C = \tau \qquad (8)$$

Application to the air heater

To find a proper transfer function model, the process was excited by a step change from 1.5
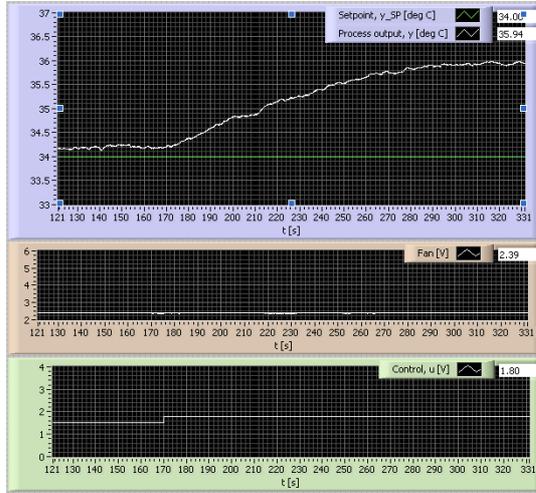
Figure 4: Skogestad's method: Process step response

to 1.8 V, see Figure 4. The response indicates that a proper model is time-constant with time-delay:

$$H_{psf}(s) = \frac{K}{Ts+1}e^{-\tau s} \qquad (9)$$

From the step response I found[4]

$$K = 5.7 \text{ °C/V}; \ T = 60 \text{ s}; \ \tau = 4.0 \text{ s} \qquad (10)$$

For the controller tuning I use (8):

$$T_C = \tau = 4.0 \text{ s} \qquad (11)$$

The PI controller parameters are

$$\underline{K_c} = \frac{T}{K(T_C + \tau)} = \frac{60}{5.7 \cdot (4+4)} = \underline{1.3} \qquad (12)$$

$$\begin{aligned} \underline{T_i} &= \min\left[T, c(T_C + \tau)\right] & (13) \\ &= \min\left[60, 4(4+4)\right] = \underline{32.0 \text{ s}} & (14) \end{aligned}$$

Figure 5 shows control system responses with the above PI settings.

---

[3] "min" means the minimum value (of the two alternative values).

[4] An exact value of the time-delay is not so easy to determine from the response, but other experiments indicate 4 sec or a somewhat less, so I set 4.0.
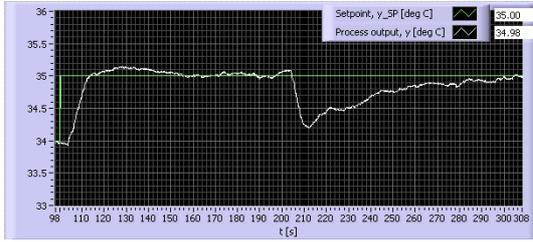
Figure 5: Skogestad's method: Closed-loop responses

The IAE indexes and the gain margin was

$$IAE_s = 12.5; \ IAE_d = 27.2; \ \Delta K = 2.4 = 7.6 \text{ dB} \tag{15}$$

Figure 6 shows the responses with this gain increase.
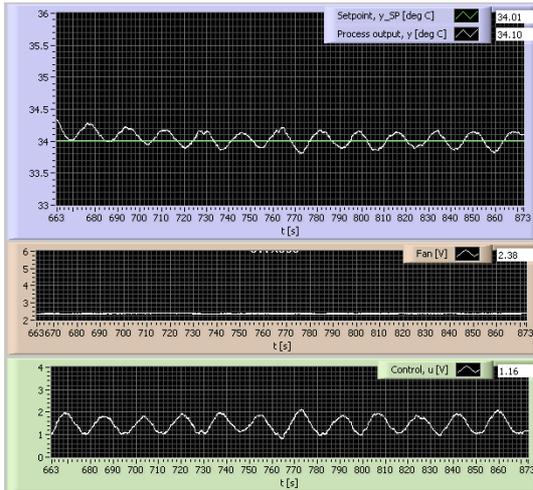
Figure 6: Skogestad's method: Responses with loop gain increase of 2.4

One interesting observation is that the actual time-constant (63% response time) as seen from Figure 5 is approximately 5 sec, which corresponds well with the specified time-constant of 4 sec.

Finally, to try to obtain faster disturbance compensation, let's set

$$c = 2 \tag{16}$$

Now we get

$$
\begin{aligned}
\underline{T_i} &= \min\left[T, \, c\left(T_C + \tau\right)\right] & (17) \\
&= \min\left[60, \, 2\left(4 + 4\right)\right] = \underline{16.0 \text{ s}} & (18)
\end{aligned}
$$

The controller gain is as before:

$$\underline{K_c = 1.3} \tag{19}$$

Figure 7 shows control system responses with the above PI settings.
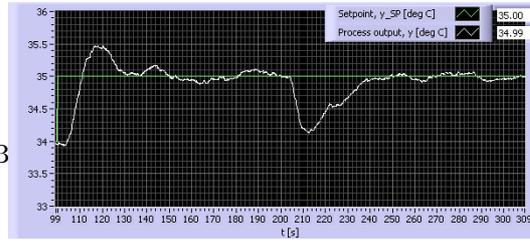
Figure 7: Skogestad's method: Closed-loop responses with $c = 2$

The IAE indexes and the gain margin was

$$IAE_s = 18.1; \ IAE_d = 18.3; \ \Delta K = 2.2 = 6.8 \text{ dB} \tag{20}$$

Thus, by setting $c = 2$ in stead of 4, the setpoint tracking is worse, but the disturbance compensation is better. The gain margin is only a little worse with $c = 2$. In most control loops the disturbance compensation is the most important task for the controller.

5.2 Ziegler-Nichols' Process Reaction Curve method

Ziegler and Nichols (1942) designed two tuning rules – known as the Ultimate Gain method and the Process Reaction Curve method – to give fast control but with acceptable stability. They used the following definition of acceptable stability: The ratio of the amplitudes of subsequent peaks in the same direction (due to a step change of the disturbance or a step change of the setpoint in the control loop) is approximately 1/4.

The Ziegler-Nichols' Process Reaction Curve method ([9]) is based on characteristics of the

step response of the process to be controlled (i.e. the open-loop system). The PID parameters are calculated from the response in the (filtered) process measurement y_{mf} after a step with height U in the control variable u. From the step response in y_{mf}, read off the equivalent dead-time or lag L and the rate or slope R, see Figure 8.
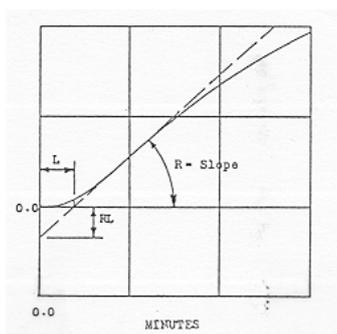
Figure 8: Ziegler-Nichols' open loop method: The equivalent dead-time L and rate R read off from the process step response. (The figure is a reprint from [9] with permission.)

After this step response test, the controller parameters are calculated with the formulas in Table 1.

	K_p	T_i	T_d
P controller	$\frac{1}{LR/U}$	∞	0
PI controller	$\frac{0.9}{LR/U}$	$3.3L$	0
PID controller	$\frac{1.2}{LR/U}$	$2L$	$0.5L = \frac{T_i}{4}$

Table 1: Ziegler-Nichols' open loop method: Formulas for the controller parameters.

Application to the air heater:

To tune the PI controller, I use the data from the open-loop experiment recorded for Skogestad's method, cf. Section 5.1. The time-delay is

$$L = \tau = 4.0 \text{ s} \qquad (21)$$

The slope R can be calculated as the initial slope of the step response. For a first order system,

$$R = \frac{KU}{T} \qquad (22)$$

where we have

$$K = 5.7 \text{ }^{\circ}\text{C/V}; \ T = 60 \text{ s} \qquad (23)$$

The PI settings becomes

$$\underline{K_c} = \frac{0.9}{LR/U} = \frac{0.9}{L\frac{KU}{TU}} = \frac{0.9}{LK/T} = \underline{2.4} \quad (24)$$

$$\underline{T_i} = 3.3 \cdot L = 3.3 \cdot 4 \text{ s} = \underline{13.2 \text{ s}} \qquad (25)$$

(Reading off R more directly from Figure 4 gives $R = 0.025 \text{ }^{\circ}\text{C/s}$, and $K_c = 2.7$.)

Figure 9 shows control system responses with the above PI settings.
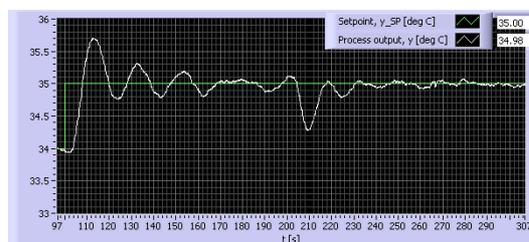
Figure 9: Ziegler-Nichols' Process Reaction Curve method: Responses in the control system

The setpoint response indicates that the stability is very poor. However, the disturbance response indicated that the stability is ok. The latter is due to the fact the increased fan speed (increased air flow) reduces the process gain and the process time-delay – thereby improving the stability of the control loop.

The IAE indexes and the gain margin was

$$\text{IAE}_s = 19.5; \ \text{IAE}_d = 8.6; \ \Delta K = 1.2 = 1.6 \text{ dB} \qquad (26)$$

5.3 Hägglund-Åstrøm's robust tuning

Hägglund and Åstrøm (2002) have derived PI controller tuning rules for "integrator with time-delay" processes and "time-constant with time-delay" processes giving maximum

performance given a requirement on robustness. The air heater looks like a "time-constant with time-delay" process. Assuming the process model is

$$H_{psf}(s) = \frac{K}{Ts+1}e^{-\tau s} \qquad (27)$$

the PI controller settings according to Hägglund and Åstrøm are as follows:

$$K_c = \frac{1}{K}\left(0.14 + 0.28\frac{T}{\tau}\right) \qquad (28)$$

$$T_i = \tau\left(0.33 + \frac{6.8T}{10\tau + T}\right) \qquad (29)$$

Application to the air heater:

To tune the PI controller, I use the data from the open-loop experiment recorded for Skogestad's method, cf. Section 5.1:

$$K = 5.7 \ ^{\mathrm{o}}\mathrm{C/V}; \ T = 60 \ \mathrm{s}; \ \tau = 4.0 \ \mathrm{s} \qquad (30)$$

The PI settings become

$$\underline{K_c = 0.76} \qquad (31)$$

$$\underline{T_i = 17.6 \ \mathrm{s}} \qquad (32)$$

Figure 10 shows control system responses with the above PI settings.
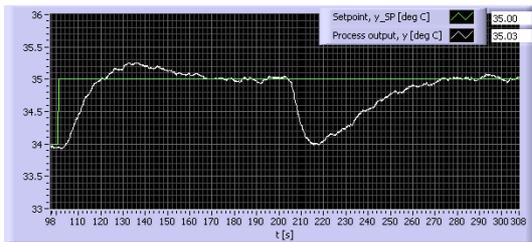
Figure 10: Hägglund-Åstrøm's Robust tuning method: Closed-loop responses

The IAE indexes and the gain margin was

$$\mathrm{IAE_s} = 17.5; \ \mathrm{IAE_d} = 32.8; \ \Delta K = 3.6 = 11.1 \ \mathrm{dB} \qquad (33)$$

5.4 Ziegler-Nichols' Ultimate Gain method (Closed-loop method)

The Ziegler-Nichols' Ultimate Gain method is based on experiments executed on an established control loop (a real system or a simulated system): The ultimate proportional gain K_{cu} of a P-controller (which is the gain which causes sustained oscillations in the signals in the control system without the control signal reaching the maximum or minimum limits) must be found, and the ultimate (or critical) period P_u of the sustained oscillations is measured. Then, the controller is tuned using K_{cu} and P_u in the formulas shown in Table 2.

	K_c	T_i	T_d
P controller	$0.5K_{cu}$	∞	0
PI controller	$0.45K_{cu}$	$\frac{P_u}{1.2}$	0
PID controller	$0.6K_{cu}$	$\frac{P_u}{2}$	$\frac{P_u}{8} = \frac{T_i}{4}$

Table 2: Formulas for the controller parameters in the Ziegler-Nichols' closed loop method.

Application to the air heater

Figure 11 shows the oscillations in the temperature response with the ulitmate gain
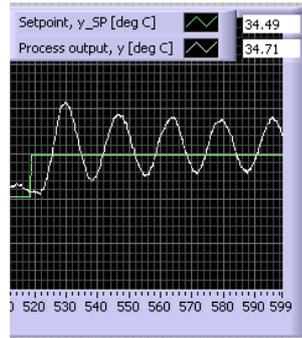
Figure 11: Ziegler-Nichols' Ultimate Gain method: Response with ultimate gain

$$K_{cu} = 3.4 \qquad (34)$$

The period of the oscillations is

$$P_u = 15 \ \mathrm{s} \qquad (35)$$

The PI parameter values become

$$\underline{K_c} = 0.45 K_{cu} = 0.45 \cdot 3.4 = \underline{1.5} \qquad (36)$$

$$\underline{T_i} = \frac{P_u}{1.2} = \frac{15 \text{ s}}{1.2} = \underline{12.5 \text{ s}} \qquad (37)$$

Figure 12 shows control system responses with the above PI settings.
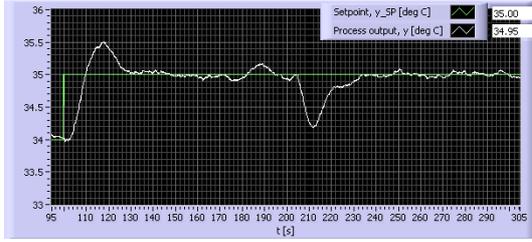
Figure 12: Ziegler-Nichols' Ultimate Gain method: Responses in the control system

The IAE indexes and the gain margin was

$$\text{IAE}_s = 13.8; \text{ IAE}_d = 11.7; \Delta K = 1.8 = 5.1 \text{ dB} \qquad (38)$$

5.5 Tyreus-Luyben's tuning method

The Tyreus and Luyben's tuning method (Luyben and Luyben, 1997) is based on oscillations as in the Ziegler-Nichols' method, but with modified formulas for the controller parameters to obtain better stability in the control loop compared with the Ziegler-Nichols' method. For a PI controller they suggest

$$K_c = 0.31 K_{cu} \qquad (39)$$

$$T_i = 2.2 P_u \qquad (40)$$

Application to the air heater

Applying the same data as for the Ziegler-Nichols' Ultimate Gain method, cf. Section 5.4, we get

$$\underline{K_c} = 0.31 K_{cu} = 0.31 \cdot 3.4 = \underline{1.1} \qquad (41)$$

$$\underline{T_i} = 2.2 P_u = 2.2 \cdot 15 = \underline{33 \text{ s}} \qquad (42)$$
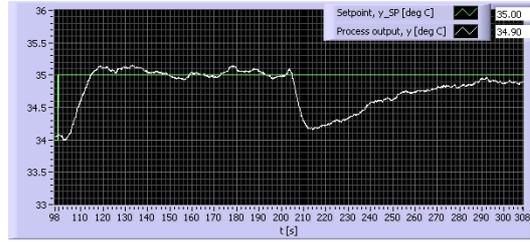

Figure 13: Tyreus-Luyben's method: Responses in the control system

Figure 13 shows control system responses with the above PI settings.

The IAE indexes and the gain margin was

$$\text{IAE}_s = 14.2; \text{ IAE}_d = 35.7; \Delta K = 3.1 = 9.8 \text{ dB} \qquad (43)$$

5.6 Relay-based tuning method

Åstrøm-Hägglund's relay-based method (Åstrøm and Hägglund, 1995) can be regarded as a practical implementation of the Ziegler-Nichols' Ultimate Gain method. In the Ziegler-Nichols' method it may be time-consuming to find the ultimate gain K_{cu}. This problem is eliminated with the relay-method of Åstrøm-Hägglund. The method is based on using a relay controller – or on/off controller — in the place of the PID controller to be tuned during the tuning. Due to the relay controller the sustained oscillations in control loop will come *automatically*. These oscillations will have approximately the same period as if the Ziegler-Nichols' closed loop method were used, and the ultimate gain K_{cu} can be easily calculated, as explained below.

The parameters of the relay controller are the "high" (or "on") and the low (or "off") control values, U_{high} and U_{low}, respectively. Once they are set, the amplitude A of the relay controller is

$$A = \frac{U_{\text{high}} - U_{\text{low}}}{2} \qquad (44)$$

If "large" oscillation amplitude is allowed,

you can set (assuming that the control signal is scaled in percent)

$$U_{\text{high}} = U_{\text{max}} = 100\% \text{ (typically)} \quad (45)$$

and

$$U_{\text{low}} = U_{\text{min}} = 0\% \text{ (typically)} \quad (46)$$

But there may be no relay controller in the control system! You can turn the PID controller into a relay controller with the following settings:

$$K_c = \text{very large, e.g. } 1000; \; T_i = \infty; \; T_d = 0 \quad (47)$$

With the relay controller in the loop, sustained oscillations comes automatically. The ultimate gain of the relay controller can be calculated as:

$$K_{cu} = \frac{\text{Amplitude of relay output}}{\text{Amplitude of relay input}} = \frac{A_u}{A_e} \approx \frac{\frac{4A}{\pi}}{E} \quad (48)$$

where $A_e = E$ is the amplitude of the oscillatory control error signal, and $A_u = 4A/\pi$ is the amplitude of the first harmonic of an Fourier series expansion of the square pulse train on the output of the relay controller.

So, after the relay controller is set into action, you read off the ultimate period P_u from any signal in the loop, and also calculate the ultimate gain K_{cu} with (48). Finally, the controller parameters can be calculated using the Ziegler-Nichols' formulas given in Table 2.[5]

Application to the air heater

The high and low control signals are, according to their physical limits:

$$U_{\text{high}} = U_{\text{max}} = 5 \text{ V} \quad (49)$$

[5]I have experienced (at least with the PID Advanced controller in LabVIEW) that the period of the oscillations are smaller than expected when the PID controller is turned into a Relay controller by setting $K_c$very large, e.g. 1000, and T_i also very large. Probably this problem is due to the anti-windup function combined with the P control action of the controller. In the experiments (see below) accomplished for this paper, I deactivated the anti-windup function by setting the max and min control signal limits to very high values: 1000 and –1000, respectively. Doing this I got the same amlitude and period of the osciolltions as with an ideal relay function.

and

$$U_{\text{low}} = U_{\text{min}} = 0 \text{ V} \quad (50)$$

Hence, the relay amplitude is

$$A = 2.5 \text{ V} \quad (51)$$

Figure 14 shows the oscillations in the tuning phase. From Figure 14 we find the ultimate
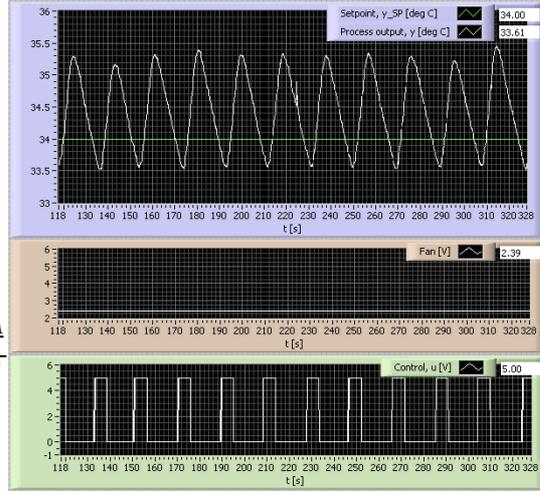
Figure 14: Relay-tuning method: Responses in the control system with relay controller

period

$$P_u = 18.0 \text{ s} \quad (52)$$

(which is quite equal to the period found with the ultimate gain in Ziegler-Nichols' method). The amplitude of the control error is appoximately

$$A_e = 0.9 \text{ °C} \quad (53)$$

The ultimate gain becomes, cf. (48),

$$K_{cu} = \frac{4A}{\pi A_e} = \frac{4 \cdot 2.5 \text{ V}}{\pi \cdot 0.9 \text{ °C}} = 3.54 \text{ V/°C} \quad (54)$$

The PI parameter values become

$$\underline{K_c} = 0.45 K_{cu} = 0.45 \cdot 3.54 = \underline{1.6} \quad (55)$$

$$\underline{T_i} = \frac{P_u}{1.2} = \frac{18 \text{ s}}{1.2} = \underline{15 \text{ s}} \quad (56)$$

Figure 15 shows control system responses with the above PI settings.
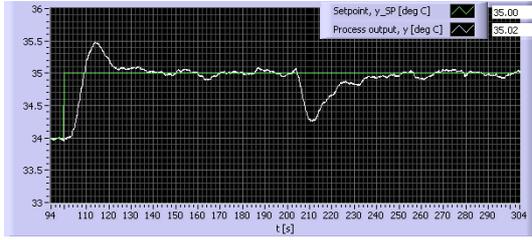
Figure 15: Relay tuning method: Responses in the control system

The IAE indexes and the gain margin was

$$\text{IAE}_s = 13.4; \text{IAE}_d = 12.9; \Delta K = 2.0 = 6.0 \text{ dB} \tag{57}$$

5.7 Sham's Setpoint method

Sham's Setpoint method (Shamsuzzoha et. al., 2010) is based on Skogestad's SIMC method. According to Skogestad himself, "It is simpler to use because it requires only one setpoint experiment. It uses P-control – a bit similar to Ziegler-Nichols – but is much quicker because one does not need to crank up the gain to get persistent oscillations."

The method is as follows: Start by using a P-controller with gain K_{c0}, and apply a setpoint change of amplitude Δy_{SP}. K_{c0} should be selected so that you get a proper overshoot in the setpoint response (in the process output). A typical value is claimed to be 0.3. From the setpoint response you read off the maximum response, $y_{\max}$, and the steady-state response, $y(\infty)$, and the time to reach the peak, t_p. Assume that the process output has value y_0 before the setpoint change. From these quantities you calculate the actual overshoot:

$$S = \frac{y_{\max} - y(\infty)}{y(\infty) - y_0} \tag{58}$$

Also calculate the relative steady-state change of the process output:

$$b = \frac{y(\infty) - y_0}{\Delta y_{SP}} \tag{59}$$

Define the following parameters:

$$F = 1 \tag{60}$$

($F = 1$ for "fast robust control" corresponding to $T_C = \tau$ in Skogestad's SIMC method, but use $F > 1$ to detune), and

$$A = 1.152 \cdot S^2 - 1.607 \cdot S + 1.0 \tag{61}$$

The PI parameter settings are

$$K_c = K_{c0} \frac{A}{F} \tag{62}$$

$$T_i = \min\left[(0.86At_p \frac{b}{1-b}, 2.44t_pF)\right] \tag{63}$$

Application to the air heater

Figure 16 shows the closed-loop response to a setpoint step change of amplitude $\Delta y_{SP} = 1.0$ °C with a P-controller with gain

$$K_{c0} = 1.8 \tag{64}$$
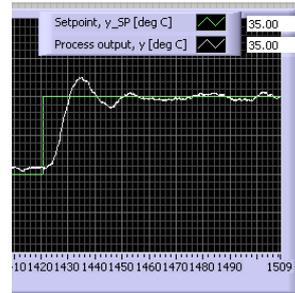
Figure 16: Sham's Setpoint method: The responses with a P-controller with gain $K_{c0} = 1.8$

which gives a stable response and a reasonable overshoot.From the responses we find the actual overshoot as

$$S = \frac{y_{\max} - y(\infty)}{y(\infty) - y_0} = \frac{35.25 - 35.0}{35.0 - 34.1} = 0.28 \tag{65}$$

The relative steady-state change of the process output is

$$b = \frac{y(\infty) - y_0}{\Delta y_{SP}} = \frac{35.0 - 34.1}{1.0} = 0.9 \tag{66}$$

We read off the peak time as

$$t_p = 14 \text{ sec} \qquad (67)$$

The PI parameter settings becomes

$$\underline{K_c} = K_{c0}\frac{A}{F} = 1.8\frac{0.64}{1} = \underline{1.2} \qquad (68)$$

$$
\begin{aligned}
\underline{T_i} &= \min\left[\left(0.86At_p\frac{b}{1-b}, 2.44t_pF\right)\right] &(69)\\
&= \min\left[(69.4, 34.2)\right] = \underline{34.2 \text{ s}} &(70)
\end{aligned}
$$

Figure 17 shows control system responses with the above PI settings.
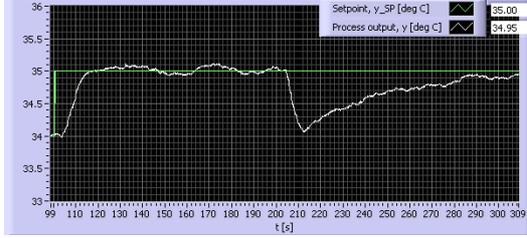
Figure 17: Sham's Setpoint method: The responses in the control system

The IAE indexes and the gain margin was

$$\text{IAE}_s = 12.2; \text{IAE}_d = 36.1; \Delta K = 2.7 = 8.6 \text{ dB} \qquad (71)$$

5.8 Good Gain method

The Good Gain method[6] (Haugen, 2010) is a simple method based on experiments with a P-controller, like in Sham's method and Ziegler-Nichols' Ultimate Gain method. As in Sham's method, the system is not brought into marginal stabililty during the tuning, which is benefical. The procedure described below assumes a PI-controller.

First, the process should be brought close to the specified operation point with the controller in manual mode. Then, ensure that the controller is a P-controller with $K_c = 0$

---
[6]I am responsible for this name.

(set $T_i = \infty$ and $T_d = 0$). Switch the controller to automatic mode. Find a good gain, K_{cGG}, by trial-and-error which gives the control loop good stability as seen in the response in the measurement signal due to a step in the setpoint. I assume that a response with a small overshoot and a barely observable undershoot (or the opposite, if the setpoint step is negative) represents good stability. A proper value of the integral time T_i is (hopefully)

$$T_i = 1.5T_{ou} \qquad (72)$$

where T_{ou} is the time between the first overshoot and the first undershoot of the step response (a step in the setpoint) with the P-controller, see Figure 18.
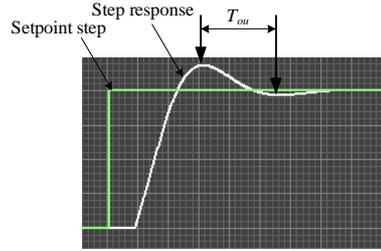
Figure 18: Reading off the time between the first overshoot and the first undershoot of the step response with P controller

Due to the inclusion of the integral term, the control loop will get somewhat reduced stability than with the P-controller only. This can be compensated for by reducing K_c to e.g. 80% of the original value.

Here is the background of this method: Assume that the control loop with the P-controller behaves approximately as an underdamped second order system with the following transfer function model from setpoint to process output:

$$\frac{y(s)}{y_{SP}(s)} = \frac{K\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2} \qquad (73)$$

It can be shown that with $\zeta = 0.6$ the step response is damped oscillations with an overshoot of about 10% and a barely observable

undershoot, as in the Good Gain tuning, and that the period of the damped oscillations is

$$P_d = \frac{2\pi}{\sqrt{1-\zeta^2}\omega_0} = \frac{2\pi}{\sqrt{1-0.6^2}\omega_0} \quad (74)$$

$$= \frac{2\pi}{0.8\omega_0} = P_{GG} = 2T_{ou} \quad (75)$$

If the oscillations are undamped, as with the Ziegler-Nichols' Ultimate Gain method, the period of the oscillations is

$$P_{ZN} = \frac{2\pi}{\omega_0} \quad (76)$$

Hence, the relation between the period of the damped oscillations of the Good Gain method and the undamped oscillations of the Ziegler-Nichols' method is approximately

$$P_{ZN} = 0.8P_{GG} = 1.6T_{ou} \quad (77)$$

In the Ziegler-Nichols' method we set

$$T_i = \frac{P_{ZN}}{1.2} = \frac{1.6T_{ou}}{1.2} = 1.33T_{ou} \quad (78)$$

If we make the T_i setting somewhat more relaxed (to obtain better stability and better robustness), we can increase T_i to

$$\underline{T_i = 1.5T_{ou}} \quad (79)$$

In the Ziegler-Nichols' method the controller gain K_c of a PI-controller is 90% of the gain of the P-controller. To compensate for the inclusion of the integral term we can reduce the original controller gain of the Good Gain method to 90%, but to relax the setting even more, let's set

$$\underline{K_c = 0.8K_{cGG}} \quad (80)$$

Application to the air heater

Figure 19 shows the closed-loop response to a setpoint step change with a P-controller with gain

$$K_{cGG} = 1.5 \quad (81)$$

The half-period is

$$T_{ou} = 12 \text{ s} \quad (82)$$
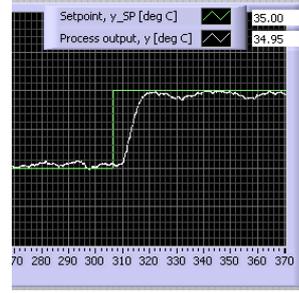
Figure 19: Good Gain method: Response with P-controller with gain $K_{cGG} = 1.4$

The PI parameter values becomes

$$\underline{K_c} = 0.8 \cdot K_{cGG} = 0.8 \cdot 1.5 = \underline{1.2} \quad (83)$$

$$\underline{T_i} = 1.5 \cdot T_{ou} = 1.5 \cdot 12 = \underline{18 \text{ s}} \quad (84)$$

Figure 20 shows control system responses with the above PI settings.
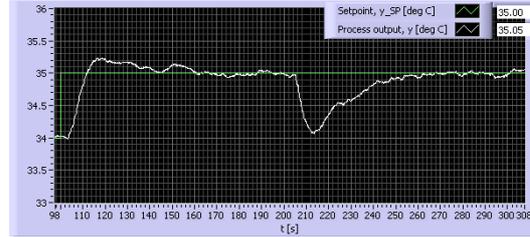
Figure 20: Good Gain method: The responses in the control system

The IAE indexes and the gain margin was

$$\text{IAE}_s = 14.3; \text{IAE}_d = 21.5; \Delta K = 2.4 = 7.6 \text{ dB} \quad (85)$$

6 Summary and discussion

Table 3 summarizes the results with the different tuning methods. Both the quickness and the simplicity of each of the methods are evaluated with a number ranging from 10 (best) to 0.

	K_c	T_i	IAE_s	IAE_d	ΔK	$_Q$	$_S$
S 1	1.3	32.0	12.5	27.2	2.4	10	9
S 2	1.3	16.0	18.1	18.4	2.2	10	9
Z N - P	2.4	13.2	19.5	8.6	1.2	10	8
H Å	0.76	17.6	17.5	32.8	3.6	10	9
Z N - U	1.5	12.5	13.8	11.7	1.8	5	6
T L	1.1	33.0	14.2	35.7	3.1	5	6
R	1.6	15.0	13.4	12.9	2.0	8	6
S S	1.2	34.2	12.2	36.1	2.7	5	6
G G	1.2	18.0	14.3	21.5	2.4	7	10

Table 3: Results for different PI controller tunings. (S1 = Skogestad original, with $c = 4$. S2 = Skogestad with $c = 2$. ZN-P = Ziegler-Nichols' Process Reaction Curve method. HÅ = Hägglund-Åstrøm's method. ZN-U = Ziegler-Nichols' Ultimate Gain method. TL = Tyreus-Luyben's method. R = Relay method. SS = Sham's Setpoint method. GG = Good Gain method.

Methods which results in gain margin ΔK less than 2.0 are here regarded as poor. So, the Ziegler-Nichols's methods are in trouble. Also, a method should not give poor disturbance compensation, i.e. not too large IAE index compared with other methods. In this light, Tyreus-Luyben's method, Sham's Setpoint method, and Hägglund-Åstrøm's method are in trouble, although the latter gives a large stability margin. Also Skogestad's method with the original value $c = 4$ is doubtful. The Relay method works well, but it is on the limit regarding robustness, and it may be a little difficult to implement in a practical system because the relay (on/off) function is not immediately available. The simple-to-use Good Gain method performs ok, although it gives a little slow disturbance compensation.

However, we want a winner, and *Skogestad's method with $c = 2$ is the lucky one*! It gives good results and is simple to use.[7] The model that the method requires can be obtained

from a simple step-response test.

7 Conclusions

This paper has demonstrated a number of PI controller tuning methods being used to tune a temperature controller for a real air heater. Indexes expressing setpoint tracking and disturbance compensation, and stability margins (robustness) were calculated. From these indexes and a personal impression about how quick a method is to deliver the tuning results and how simple the method is to use, a winning method has been identified (from the tests reported in this paper), namely *Skogestad's method* (with a modified integral time tuning).

References

[1] Åstrøm, K. J, Hägglund, T., *PID Controllers: Theory, Design and Tuning*, ISA, 1995

[2] Haugen, F., *Basic Dynamics and Control*, ISBN 978-82-91748-13-9, pp. 127-129, TechTeach, 2010

[3] Hägglund, T., Åstrøm, K. J, *Revisiting the Ziegler-Nichols' Tuning Rules for PI Control*, Asian J. Of Control, Vol. 4, 364, 2002

[4] Luyben, W. L., Luyben, M. L., *Essentials of Process Control*, McGraw-Hill, New York, 1997

[5] O'Dwyer, A., *Handbook of Controller Tuning Rules*, Imperial College Press, London, 2003

[6] Seborg, D. E., Edgar, Th. F., Mellichamp, D. A., *Process Dynamics and Control,* Ch. 12, John Wiley and Sons, 2004

[7] Shamsuzzoha, M., Skogestad, S., Halvorsen, I. J., *On-Line PI Controller Tuning Using Closed-Loop Setpoint*

---

[7]Additional benefits of Skogestad's method are application to processes without time-delay where stability-based methods fail (as in level control), and easy continuous adaptation of controller parameters to possibly varying process model parameters.

Responses for Stable and Integrating Processes, to be presented at IFAC conference on dynamics and control of process systems processes (DYCOPS), Belgium, July 2010

[8] Skogestad, S., *Simple analytic rules for model reduction and PID controller tuning*, Journal of Process Control, Vol. 13, pp. 291-309, 2003

[9] Ziegler, J. G. and N. B. Nichols: *Optimum Settings for Automatic Controllers*, Trans. ASME, Vol. 64, page 759-768, 1942

Tentative Dependence Analysis of Process Variables in a Circulating Fluidized Bed Boiler

Lohiniva, Laura & Leppäkoski, Kimmo
Systems Engineering Laboratory, Department of Process and Environmental Engineering
University of Oulu, Finland
laura.lohiniva@oulu.fi, kimmo.leppakoski@oulu.fi

1. INTRODUCTION

Ash-induced problems in boilers include deposit formation, accelerated corrosion and erosion of surfaces and decrease in energy efficiency due to formation of insulating layers. Co-combustion of fossil fuels and biomasses of various quality may even worsen the problems in unfavourable process conditions and fuel mixes. There are various different approaches to investigating slagging and fouling of heat exchanger surfaces. There have been various attempts to model, predict and monitor deposit formation (for example Henderson et al. 2006, Ma et al. 2008, Räsänen et al. 2006). If conventional process measurement variables could be used in monitoring, slagging and fouling could be investigated during normal operation of the plant.

Preliminary data analysis, preceding a more accurate analysis and later control design, is used to study properties of data and dependence among variables, but also in detecting errors. Methods can be quantitative or qualitative. As computing power and data storage capacity have increased, effective utilization of soft computing methods has become possible. The Self-Organizing Map (SOM) is one of the most popular methods used in data mining; it performs a nonlinear dimensionality reduction on the data by using competitive unsupervised learning and then produces a visualization of the results, usually on a 2-dimensional regular lattice (Kohonen 2001). In addition to visualization and preliminary analysis, the SOM is also an efficient tool for modelling and data preprocessing purposes as well as for reducing noise and computational load in clustering.

2. DATA AND METHODS

Ordinary process measurement data from a circulating fluidized bed boiler (CFB) was retrieved from the process automation system of a power plant located in northern Finland. Main fuels used in the plant include peat (up to 80%), wood, bark and other wood-based materials. CFB boilers tolerate fuels of fluctuating quality and properties, which makes them highly suitable for co-combustion. Data from the collection period, between September 2st and December 31st 2008, was extracted from the data storage of the process automation system by an Excel application. Missing values in data were replaced with mean approximations during collection. The plant normally operates throughout the year. During the four collection months there were two shutdown periods (September 29th – October 10th, October 23rd – 29th) that were excluded from the analysis. The limited data consisted of 138 days, and since the data consists of minute averages, the number of sample rows was 198720. Two methods were used in data analysis. Regular correlation coefficient calculation can be used to obtain numerical results. The Self-Organizing Map (SOM) provides visual presentation of the results. Before analysis, data was further preprocessed in several ways. Flue gas emissions were reduced to standard 6 % O_2 content in order to obtain comparable and usable values. Data was mean-centered and normalized to unit variance for SOM.

3. RESULTS AND DISCUSSION

Correlation coefficients (covariance of two variables divided by the product of their standard deviations, in the range of [-1,1] and indication of linear dependency) were calculated for a group of variables. The results could be compared to previous results (Lohiniva & Leppäkoski 2010) from bubbling fluidized bed boiler (BFB) data. The differences in BFB and CFB processes, most importantly in the recirculation of bed material and unburnt fuel, and the location of measurement points account for some of the differences.

The obtained coefficients showed that SO_2 and NO_X contents in flue gas had medium correlation (0.4647). In previous studies, the two variables have been considered quite independent in the case of CFB boilers. SO_2 content had medium correlation with flue gas temperatures in different parts of the flue. Flue gas oxygen contents

at economizer and before stack had surprisingly low correlation (0.5469). Flue gas pressure differences over three parts of the flue gas duct had strong or almost linear correlation with each other. Flue gas temperatures did not correlate with each other as strongly as expected. Temperatures and pressure differences had only medium correlation (table 1). Most of the results differed from the results obtained from BFB data.

Table 1. Correlation coefficients between flue gas temperatures and pressure differences

	T after superheaters	T after economizer	T after air preheaters
p_d at superheaters	0.5695	0.6197	0.6256
p_d at economizer	0.8735	0.8644	0.8705
p_d at air preheaters	0.7401	0.7192	0.7345

The SOM results (map size [69,32], quantization error 1.590, topographical error 0.115) supported the results from correlation calculation. Component planes effectively visualize, for example, the independence of SO_2 content from other variables. The CO-O_2-relation can be clearly seen. Steam flow, pressure difference and flue gas temperature maps have similar variation, although the temperature planes are not alike. These differences could be further investigated.
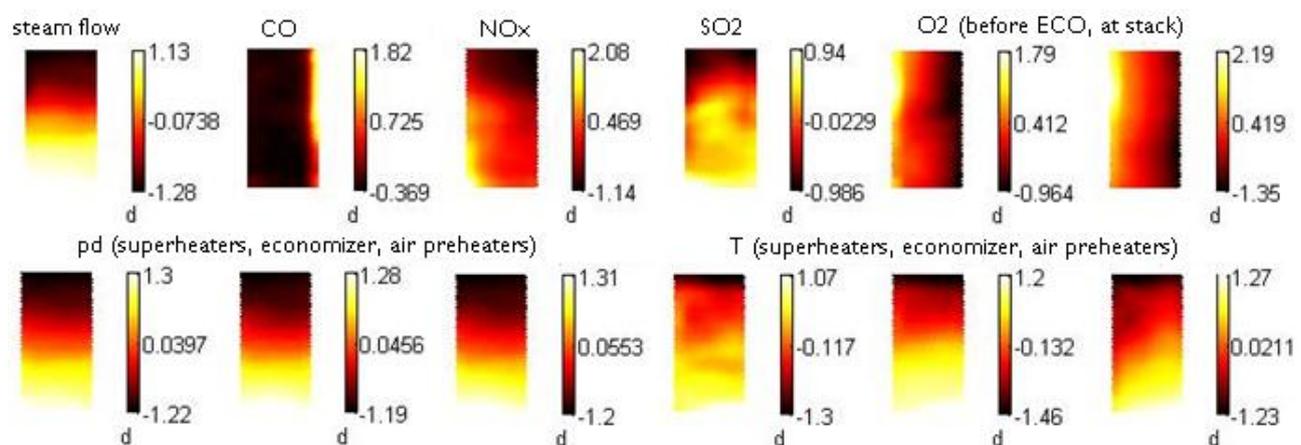
Figure 1. SOM component planes

Both methods provided fast and illustrative preliminary presentation of the data. The methods used are inexpensive to compute and could be used in online monitoring.

REFERENCES

Henderson, P., Szakálos, P., Pettersson R, Andersson, C., and Högberg, J. (2006). Reducing superheater corrosion in wood-fired boilers. *Materials and Corrosion*, 57 (2), pp. 128-134.

Kohonen, T. (2001). *Self-Organising Maps*, 3. p.. Springer Verlag, Berlin, Heidelberg, Germany.

Lohiniva, L. & Leppäkoski, K. (2010). Preliminary Dependence Analysis of Process Variable: A Case Study of a Bubbling Fluidized Bed Boiler. IFAC Conference on Control Methodologies and Technology for Energy Efficiency, Vilamoura, Portugal, March 29-31 2010. (CD)

Ma, Z., Iman F., Lu P., Sears R., Kong L., Rokanuzzaman A.S., McCollor D.P. and Benson S.A. (2008). A comprehensive slagging and fouling prediction tool for coal-fired boilers and its validation/application. *Fuel Processing Technology,* 88 (11-12), pp. 1035-1043.

Räsänen, T., Kettunen, A., Niemitalo, E., and Hiltunen, Y. (2006). Self-refreshing SOM for dynamic process state monitoring in a circulating fluidized bed energy plant. *3nd International IEEE Conference on Intelligent Systems*. September 2006, pp. 344-349.

Automated Controller Design using Linear Quantitative Feedback Theory for Nonlinear systems

Roozbeh Kianfar and Torsten Wik*

Abstract— **A method to design simple linear controllers for mildly nonlinear systems is presented. In order to design the desired controller we approximate the behavior of the nonlinear system with a set of linear systems which are derived through linearizations. Classical local linearization is carried out around stationary points but in order to have a better approximation of the nonlinear system selected non-stationary points are taken into account as well. This set of linear models are considered as an uncertainty description for a nominal plant. Qunatitative Feedback theory (QFT) may be used to guarantee specification to be fulfilled for all linear models in such an uncertainty set. Traditionally QFT design is carried out in a Nichols diagram by loop shaping of the nominal linear plant. This task highly depends on the experience of the designer and is difficult for unstable systems. In order to facilitate this task an optimization algorithm based on Genetic algorithm is used to automatically synthesize a fixed structure controller. For illustration and evaluation the method is succesfully applied to a Wiener system and a nonlinear Bioreactor benchmark problem.**

KEYWORDS: **Nonlinear, QFT, loop shaping, linearization, non-stationary point, genetic algorithm.**

I. INTRODUCTION

In the process industry ease of implementation is without doubt one of the most important aspects of automatic control. Provided the performance is acceptable, fixed structure and linear controllers, such as PID controllers, are therefore advantageous even though the process itself may be nonlinear. In line with this, the aim of the work presented here is a semi-automized method for determination of fixed structure low order linear controllers for mildly nonlinear single input single output (SISO) processes.

Depending on the character of a nonlinear process there are many methods for designing nonlinear controllers, such as Feedback linearization, Sliding control, Adaptive control and Model predictive control (c.f. [9]). However, in many control systems there is little support for these methods and operators are untrained in their use. As a consequence, most mildly nonlinear plants are controlled by linear controllers, mainly PID controllers, which are either tuned experimentally or synthesized for a specific operating point. Because of the nonlinearities the system will have deteriorating properties and may become unstable when operated too far away from the design point. The idea here is to find a controller parameterization that gives a robust system in the sense that it has an acceptable performance in a large operating region.

Since there is an abundance of efficient methods for synthesis of linear controllers from linear models, the use of a linear process model is in many cases motivated. One way is then to use a linear model and treat the nonlinearities as model uncertainties. Schweickhardt and Allgöwer [14], [15], [16] use this to define the best linear model as the one with the smallest gain of the uncertainty. In [15] they pursue by determining the linear controller such that the Small gain theorem can be used to guarantee stability. The drawbacks are difficulties in the computation of the nonlinearity measure (uncertainty gain), that the process needs to be stable and that the use of the Small gain theorem introduces conservatism in the resulting solution. Basically, the problem of conservatism and its connection to the nonlinearity measure originates from the fact that the gain is considered for signals that the controllers might neither use nor apply. To some degree this can be taken into account by bounding the input amplitude [16].

Olesen *et al.* [13] also used the idea of treating the nonlinearities as uncertainties and disturbances to show that only a few linear controllers are needed in gain scheduling control of the temperature in an exothermic tank reactor. They use model linearizations to generate a set of transfer functions that can be interpreted as a model uncertainty description. This is then followed by a controller design using Quantitative Feedback Theory (QFT) to guarantee robustness specifications for all transfer functions in the set. By adding non-stationary linearization points the robustness and operating window for each controller could be made significantly larger. The use of off-equilibrium linearization has also been shown to improve performance of gain scheduling control when the controller parameters are interpolated [8].

QFT was originally developed for linear systems with uncertainties (c.f. [7]), but there are also extensions to nonlinear systems with uncertainties, which are based on finding an equivalent linear model for the nonlinear system (see [1] and references therein). However, this requires the knowledge of what specific input signal that generates the desired output, which currently limits its use.

Basically, standard linear QFT design is based on loop shaping the nominal loop transfer function such that for each frequency considered it does not violate frequency dependent Horowitz Sidi (HS) bounds. A drawback of standard QFT for linear systems is that the manual loop-shaping in the Nichols chart highly depends on the experience of the designer. During the last decade solutions to how to automate this step have therefore been proposed. Basically, they rely on optimization where the bounds constrain the search space. The optimization problem, however, is in general non-convex. Chait *et al.* [1] convexifies the HS-bounds and

*Department of Signals and Systems, Chalmers University of Technology, SE-412 96 Göteborg, Sweden, E-mail: roozbeh@student.chalmers.se, tw@chalmers.se

solve the problem using linear programming. However, the use is limited because the method requires that the closed loop poles are known beforehand. Another approach is to use a global optimization routine. Nataraj *et al.* [12], [11] propose an interval analysis, and Chen *et al.* [2] use genetic algorithm. To improve the accuracy for a given numerical effort Fransson *et al.* [4] use a combination of a global (DIRECT method) and a local optimization routine. It should be noted that optimized control of uncertain linear systems can also be determined using the structured singular value for the constraints, as in [5] and [17]. However, for SISO systems with less than 8 parameters to optimize the use of HS-bounds can in general be recommended [17].

The method we present here for nonlinear processes is based on the manual method used by Olesen *et al.* [13], combined with an optimization using genetic algorithm, which has the advantage that no initial guess is required - a valuable property from an automation point of view. Based on systematically selected simulations of the nonlinear system new linearization point are added to the set of transfer functions until performance and robustness no longer is improved. This method is then applied to a Wiener system [15] and a nonlinear benchmark problem, an unstable bioreactor [10]. Some solutions have been proposed for this problem such as [3], though it appears as if no linear controller for the process has been evaluated earlier. The PID controller derived with the method presented here performs well over the operating window and also compare well to the sliding mode controller by Mehmed *et al.* [3].

II. METHOD

A. Quantitative Feedback Theory for design and analysis

QFT is a method for design and analysis of feedback controlled uncertain system, originally developed by I. Horowitz [7]. The uncertainties and closed loop specifications should be translated to frequency domain and one arbitrary transfer function P_{nom} will be considered as the nominal one. Instead of simultaneous design for all the loop transfer functions defined by the uncertainties, the design can then be carried out only for the nominal loop transfer function, $L_{nom}(j\omega) = P_{nom}(j\omega)G(j\omega)$, where $G(j\omega)$ is the controller (see Fig. 1). The uncertainties are taken into account in a translation of the specifications into frequency dependent boundaries (the Horowitz-Sidi bounds) which the nominal loop transfer function should not violate.
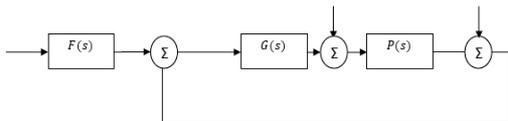
Fig. 1. Two degree of freedom controller.

The design of the feedback compensator, is carried out in the following steps:

- Define the uncertainties in the process by a set of transfer function $P_j(s)$. One of these plant transfer functions is selected to be the nominal one. Then we calculate the so called *templates* for relevant frequencies, i.e, the values of all $P_j(j\omega_i)$ for chosen relevant frequencies.
- Formulate closed loop specifications, such as servo specification and sensitivity specifications.
- Use the templates to calculate the corresponding Horowitz-Sidi bounds for the specifications.
- Given the nominal plant and the Horowitz-Sidi bounds exploit the methods in classical control and loop shaping techniques to shape the nominal loop transfer function such that it satisfies all the Horowitz-Sidi bounds.
- Check the stability of the closed loop system for all plants with the Nyquist criterion.
- If the system response is not within the acceptable servo specification envelope, a prefilter $F(s)$ is needed prior the loop.

III. AUTOMATED CONTROLLER SYNTHESIS USING GENETIC ALGORITHM

In practice QFT loop-shaping is carried out in a Nichols diagram for a finite number of frequencies, $\Omega = \{\omega_i\}$. The idea in this work is to translate this problem into an optimization problem. If we assume that the controller has a fixed structure

$$G(s) = \frac{\theta_m s^m + \theta_{m-1} s^{m-1} + ... + \theta_0}{s^n + \theta_{m+n} s^{n-1} + ... + \theta_{m+1}} \quad (1)$$

where θ is the parameter vector to be determined by optimization. The Horowitz-Sidi bounds at each frequency ω_i are denoted $B_i(\angle L_0(j\omega_i, \theta), \omega_i)$. These bounds have different shapes and may be single-valued or multiple valued, depending on the specifications.

The objective here is to synthesize a controller such that:
- The Horowitz-Sidi bounds at each frequency ω_i are not violated.
- The nominal loop-transfer function is stable.
- The controller has low complexity.

In most QFT literature the aim is to minimize the high frequency gain of controller. In this work we follow that tradition by choosing a cost function $J(\theta)$ that is the high frequency gain of the controller, However low frequency disturbance rejection, as in [4], could equally well have been used. The Horowitz-Sidi bounds are translated to nonlinear constraint inequalities as below:

$$ub_i(\theta) = B_i(\angle L_0(j\omega_i, \theta), \omega_i) - |L_0(j\omega_i, \theta)| \leq 0 \quad (2)$$
$$lb_i(\theta) = |L_0(j\omega_i, \theta)| - B_i(\angle L_0(j\omega_i, \theta), \omega_i) \leq 0 \quad (3)$$

where ub_i and lb_i are upper and lower single-valued bound constraints. Multiple valued bounds are split into one upper and one lower bound. There are in general no analytical functions for these bounds and in this work we derive them numerically using the QSYN toolbox for Matlab [6]. The nominal closed loop transfer function stability imposes one more constraint to the problem: the roots λ of $1 + L_0 = 0$

should be in the LHP. Hence, the problem can now be formulated as

$$\min_{\theta} J(\theta)$$

$$\begin{aligned}
\textbf{subject to:} \quad ub_i(\theta) &\leq 0 \ \forall i \\
lb_i(\theta) &\leq 0 \ \forall i \\
Re[\lambda(1 + L_0)] &\leq 0
\end{aligned}$$

This problem is classified (to be) in the global optimization category with nonlinear constraints, a class that classical gradient based optimization methods are generally not suited for. However, Genetic algorithm, which is a powerful evolutionary method with the ability to handle the nonlinear constraints, is a good candidate to solve this problem.

Advantages of this method are: (1) the order of the controller can be assigned beforehand. (2) There is no need for an initial guess. (3) There is no need to determine optimization variable search space in advance. In order to illustrate the efficiency of this automated synthesis, a manual design example from the QSYN-manual is compared with the solution derived by optimization.

Example 1

An uncertain plant, with parametric uncertainty is given:

$$P(s) = K.\frac{s + a}{1 + 2\zeta s/\omega_n + s^2/w_n^2} \tag{4}$$

where $K \in [2, 5]$, $a \in [1, 3]$, $\zeta \in [0.1, 0.6]$ and $\omega_n \in [4, 8]$, and the structure of $G(s)$ is

$$G(s) = \frac{\theta_3 s^3 + \theta_2 s^2 + \theta_1 s + \theta_0}{s^4 + \theta_6^3 + \theta_5 s^2 + \theta_4} \tag{5}$$

i.e. the same as the controller presented in the manual. The design specifications are:

$$M_T \leq 0.1 \tag{6}$$

$$T_s \leq 10s \tag{7}$$

$$\|S(j\omega)\| = \frac{1}{|1 + G(j\omega)P(j\omega)|} \leq 6dB \tag{8}$$

where M_T is the maximum overshoot of the closed loop step response, T_s is the settling time and $S(j\omega)$ is the sensitivity function. First of all, the design specifications are translated to frequency domain. In QSYN this task is carried out through an approximation of the closed loop system by a low order system, from which the correspondence between time and frequency domain is used.

The objective function to be minimized is the high frequency gain of controller, i.e. $J(\theta) = \theta_3$, and the optimization variables are the coefficients of $G(s)$, i.e. $\theta = [\theta_0, \theta_1, ... \theta_6]^\top$. As can be seen from Fig. 2, the automatically synthesized controller satisfies all specifications, that is, the nominal loop transfer function for selected frequencies is outside the sensitivity HS bounds and above the bounds for the servo specifications. The nominal loop transfer function $L_0(j\omega)$ is stable and the high frequency gain of the controller
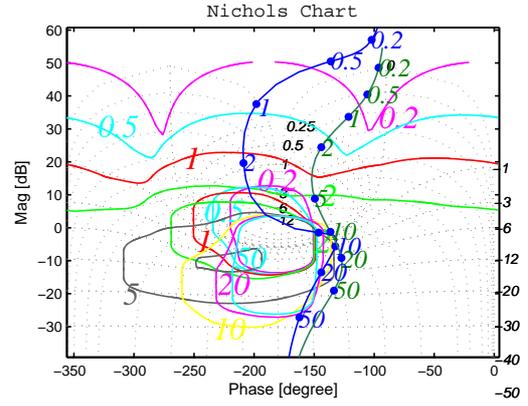

Fig. 2. The green curve to the right is the nominal loop transfer function (manual design). The blue curve to the left is derived from automated loop-shaping.

in the automated design is less than the solution given in the manual.

The foregoing example showed how Genetic algorithm can be used to solve the loop-shaping problem for an uncertain linear system. The Matlab Toolbox for genetic algorithm was used to solve the problem with a population size of 60. The time of optimization highly depends on the number of optimization variables and population size, for this specific example it is around 30 min.

The ultimate goal is to present a controller that not only works in a small region around the equilibrium points but also is robust to rather large deviations in initial points from equilibria. Quantitative feedback theory is a useful method for design and analysis of uncertain linear systems. There are solutions to design QFT controller for nonlinear systems. However, a so-called equivalent linear model needs to be found then. To find this equivalent linear model it is required to have a good knowledge about which input signal will yield the desired output, and for an unstable system this is not a trivial task (cf. the issues in finding the best linear model [15]). Different methods of finding this equivalent linear model have been discussed in [1]. In general it can be divided into global and local approach. Because of some difficulties to deal with the global approach, the local approach is our interest here in this work.

For a nonlinear system:

$$\dot{x} = f(x, u) \tag{9}$$

$$y = h(x, u) \tag{10}$$

local linearization around $(\bar{x}, \bar{u})$ gives:

$$\begin{aligned}
\Delta \dot{x} &= A(\bar{x}, \bar{u})\Delta x + B(\bar{x}, \bar{u})\Delta u \\
&\quad + R(\bar{x}, \bar{u}) \tag{11}
\end{aligned}$$

$$\text{where} \quad A = \frac{\partial f}{\partial x}(\bar{x}, \bar{u}), \quad B = \frac{\partial f}{\partial u}(\bar{x}, \bar{u}) \tag{12}$$

and Δx and Δu are the deviations from $\bar{x}$ and $\bar{u}$ respectively.

In classic linearization this task is carried out only for stationary points which are derived from:

$$f(\bar{x}, \bar{u}) = 0 \ and \ \bar{y} = h(\bar{x}, \bar{u}) \tag{13}$$

which implies $R(\bar{x}, \bar{u}) = 0$. Linearization around non-stationary points introduces a constant $R(\bar{x}, \bar{u})$ and due to this term, properties such as stability are meaningless, because these points are reached only during transients. Anyway (11) approximates the possibly transient dynamics of the nonlinear system when the trajectory is close to $R(\bar{x}, \bar{u})$. Under the assumption that $R(\bar{x}, \bar{u})$ is small enough, R can also be considered as a process disturbance. The nonlinear system is approximated by a family of LTI systems and a set of process disturbances.

In this work the design is carried out through the following steps:

1) Define specifications and cost function.
2) Determine equilibria and linearize around them to get the initial set $\{P_{0i}(j\omega_k)\}$.
3) Determine the relevant non-stationary points in the desired operating window and linearize around them to get the new templates $\{P_i(j\omega_k)\}$.
4) Translate the specifications into Horowitz-Sidi bounds in the Nichols chart.
5) Decide the structure of controller.
6) Run optimization algorithm using GA.
7) Simulate the system with initial conditions in the desired operating window.
8) If the response becomes unstable go back to step 3 and repeat the algorithm again.

Looking at the simulation result and direction of transient response (for the case of second order system such as our problem looking at the phase plane) is informative and might help to decide about new non-stationary points. The design procedure will be explained in detail for a benchmark problem in next section.

Example 2

In [15] an example is presented which compares the capability of two controllers in rejecting disturbances. One of the controllers is designed for the *best linear model* and another one is designed for a linear model derived thorough classic linearization. Here, we applied the proposed method, based on QFT, on the same nonlinear system and compare the result with the two previous ones.

The nonlinear model is a simple Wiener system. The Wiener system is given by the series connection of the linear system $G(s) = \frac{1}{(s+1)^3}$ followed by a static nonlinearity $f(x) = x + x^3$. The control signal is limited to be $u \in [-2, 2]$.

First of all, the nonlinear system is linearized around stationary and relevant non-stationary points. The effect of linearizing at different points is only in the DC gain of the system. It can be interpreted that we can replace the nonlinear system with an uncertain linear system with uncertainty only in the DC gain. Then, in the next step the servo and disturbance rejection specifications are translated to frequency domain. Finally, Genetic algorithm is used to synthesize a PID controller for this system such that the high frequency gain of the controller is minimized.

The two controllers presented in [15], are the controller designed for the best linear model

$$u = (0.2 + \frac{0.12}{s})e \tag{14}$$

and the controller derived from local linearization:

$$u = (1 + \frac{0.6}{s})e \tag{15}$$

The controller designed using QFT and linearization at stationary and non-stationary points (to minimize the high frequency gain of the controller are) is

$$u = (\frac{0.397s^2 + 0.562s + 0.334}{s})e \tag{16}$$

As can be seen in Fig. 3 and 4 adding non-stationary linearization points improves the robustness and performance considerably. The derived PID controller also compares well to the controller based on the best linear model [15]. We also used the following criterion

$$J_{LF} = \|S(s)\|_\infty \tag{17}$$

where $S(s)$ is the sensitivity function. The foregoing criterion is a measure of the system's ability to compensate LF output disturbances (see Fig. 4). The controller derived for this criterion is:

$$u = (\frac{0.588s^2 + 0.614s + 0.336}{s})e \tag{18}$$
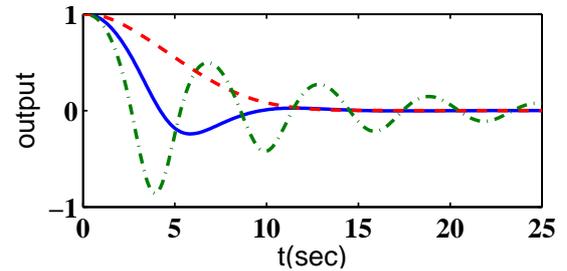
Fig. 3. Closed loop responses to a step with height one as output disturbance. The solid curve (blue) corresponds to the QFT controller, the dashed curve (red) is the response corresponding to the best linear model, and the dashed-dot curve (green) is the closed loop response with the controller based on local linearization.

IV. BENCHMARK PROBLEM

A. MODELING

In order to further evaluate the performance and characteristics of our method, we have selected a Bioreactor Benchmark problem as the plant to be controlled due to its interesting characteristics [10]. Although this process is rather simple and only has two state variables it is difficult to control due to strong nonlinearity. The bioreactor is a continuous flow stirred tank reactor (CSTR) with water and cells (e.g., yeast or bacteria) which consumes nutrients ('substrate') and produce products (both desired and undesired) and more cells. The stated control problem is tracking a desired amount of cell mass.
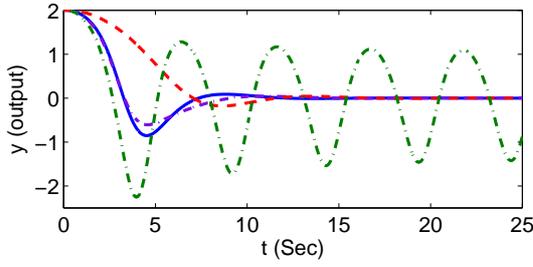
Fig. 4. Closed loop responses to the step with height two as the output disturbance. The solid curve (blue) corresponds to the QFT controller minimized respect to HF gain and the purple curve is the QFT controller minimzed respect to LF disturbance rejection, the dashed curve (red) is the response corresponding to the best linear model and the dashed-dot curve (green) is the closed loop response with the controller based on local linearization.
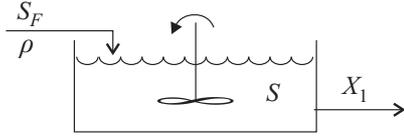

Fig. 5. Bioreactor with ρ as input and x_1 as output

The state space equations of the plant are:

$$\dot{X}_1 = -X_1\rho + X_1(1-X_2)e^{X_2/\gamma} \qquad (19)$$

$$\dot{X}_2 = -X_2\rho + X_1(1-X_2)e^{X_2/\gamma}\frac{1+\rho}{1+\rho-X_2} \qquad (20)$$

where X_1 is dimensionless cell mass and X_2 is nutrient conversion, defined as $(S_F - S)/S_F$, where S_F is the concentration of nutrient in the feed to the reactor and S is the concentration (of nutrient) in the reactor. The constraints on the state variables are, $X_1 \geq 0$ and $X_2 \leq 1$. ρ is the control signal, which is the flow rate through the reactor $(0 \leq \rho \leq 2)$. The constants β and γ determine the rate of cell growth and nutrient consumption. From the equations we may also deduce that cell growth in moderate nutrient concentrations is faster than at very high or low conversion.

This system is a challenging benchmark because it is highly nonlinear and for some values of ρ limit cycle is unavoidable, see Fig. 6. The system is also unstable, as can be seen in the phase portrait in Fig. 7. It can be noted that the system has one stable and one unstable eigenvalue in this area so the equilibrium points are saddle points. The system response is very sensitive to parameter variation. It means that a small error in the model can cause a large change in the control problem.

B. CONTROL DESIGN

According to our design procedure we begin by linearizing the system around its stationary curve. To obtain stationary points we need to solve the (19) and (20) at a steady state, which gives

$$\dot{x}_1 = 0 \Rightarrow \rho_{ss} = (1-x_2)^{x_2/\rho} \qquad (21)$$

$$\dot{x}_2 = 0 \Rightarrow x_2 = 0 \text{ or } x_2 = x_1\frac{1+\beta}{1+\beta-X_2} \qquad (22)$$
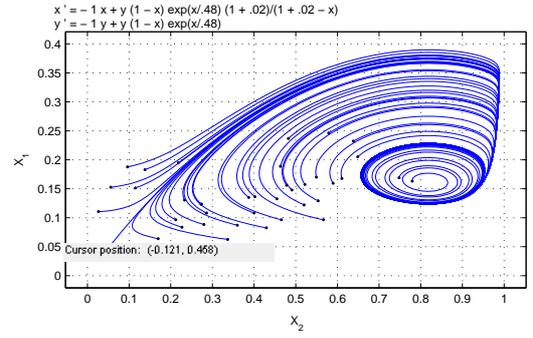
Fig. 6. Limit cycle for $\rho = 1$
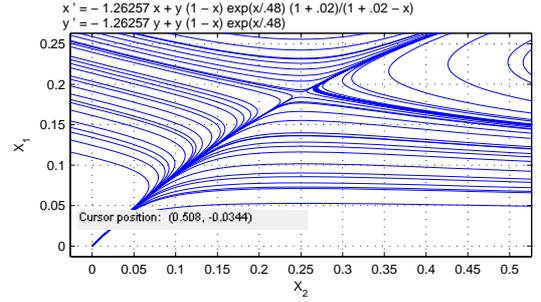
Fig. 7. saddle point equilibrium for $\rho = 1.26$

From the plot in Fig. 8. we observe that in a steady state we cannot achieve any value larger than $\frac{1+\beta}{4} = 0.255$ for cell mass. As mentioned, the main goal is to track a desired cell mass (X_1). We limit our design to $0 \leq X_1 \leq 0.255$ and $0 \leq X_2 \leq 0.51$. The nonlinear plant is linearized around stationary points in this region and one of these plant is selected as the nominal one. Then, the servo specifications are translated from time domain into frequency domain. Output disturbance rejection constraint is also applied to the system in the form of a constraint on the sensitivity function $\|S\| \leq 3$. The Matlab toolbox Qsyn [6] is used to calculate the corresponding Horowitz-Sidi bounds for these specifications. In Fig. 9 the nominal plant together with the Horowitz-Sidi bounds is portrayed for different frequencies. Clearly, the nominal plant is unstable and violates all the Horowitz-Sidi bounds. We also observe that the gain uncertainty of the template, especially for frequencies smaller
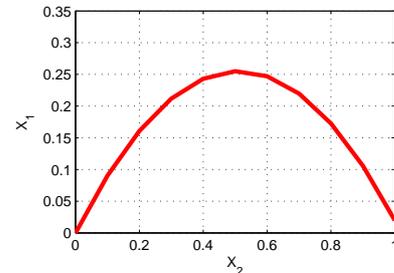
Fig. 8. Stationary points for bioreactor
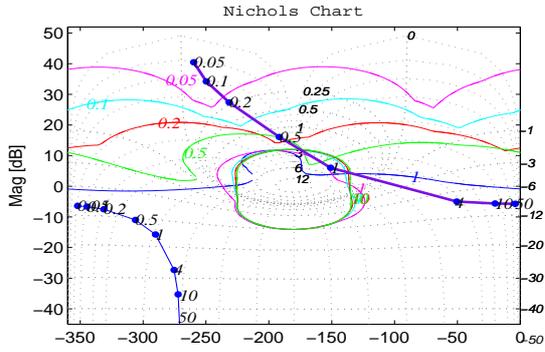
Fig. 9. The blue curve to the left is nominal loop transfer function before design and the purple thick curve to the right is the loop transfer function after design for only stationary points

Fig. 11. The blue curve is nominal loop transfer function after designing the controller

the system response is satisfactory but for the larger perturbations in the initial condition from equilibrium points the system becomes unstable. These new non-stationary points are then added to the set of linearization points. This imposes tougher boundaries on the nominal loop transfer function $L_0(j\omega_i)$ in the Nichols chart (see Fig. 11). The problem is then solved with genetic algorithm once more. The simulation results for this new controller, the former one and a sliding mode solution is presented in the next section. In Fig. 12 the gain extent of the closed loop system together with the uncertainty in the template is portrayed. We see that after designing the feedback the uncertainty is reduced to an acceptable level. We also conclude that there is no need to design a prefilter $F(s)$.

Fig. 10. The blue curve is nominal plant. Small circles, show the uncertainties defined by the template. The red envelope is servo specification.

than the bandwidth frequency, is larger than the tolerance specification (see Fig. 10).Feedback control is needed to reduce the uncertainty within the acceptable envelope. The idea here is to automatically design a PID controller such that the closed loop system becomes stable and fulfill the specifications for all frequencies. The controller has the the following (ideal) transfer function:

$$G(s) = \frac{K_D s^2 + K_P s + K_I}{s} \qquad (23)$$

The optimization variables are $\theta = [K_D, K_P, K_I]^\top$, and the objective function to be minimized in this case is K_D (high frequency gain of controller) subject to the following specifications:

- Servo specification

$$a(\omega) \le \left| \frac{F(j\omega)G(j\omega)P(j\omega)}{1 + G(j\omega)P(j\omega)} \right| \le b(\omega) \qquad (24)$$

- Sensitivity specification

$$\left| \frac{1}{1 + G(j\omega)P(j\omega)} \right| \le 3 \qquad (25)$$

Fig. 9. shows the nominal loop transfer function after design of the PID controller. From the plot we can see that the system becomes stable and the nominal loop transfer function satisfies the specifications for all frequencies. However, we cannot claim that it has the desired performance on the original nonlinear system unless we test our design through simulation. When we simulate the system from an initial condition in a region close enough to the stationary points,

Fig. 12. The blue curve is nominal plant, the small circles show the uncertainty in the template, and the red envelope is the servo specification.

V. SIMULATION RESULTS

Simulations were carried out in Simulink for different initial values and different square waves as reference signal. First of all, the two controllers, one derived from linearization around only stationary points and another one from linearization around both stationary and non-stationary points, are simulated for two different initial values. For an initial condition close enough to the stationary curve both controllers work, but as can be seen in Fig. 13 for an initial condition $x_1(0) = 0.15$ and $x_2(0) = 0.3$ the controller designed from linearization around stationary points only results in large overshoots. In Fig.14 we perturbed the system harder by giving an initial condition equal to $x_1(0) = 0.09$ and $x_2(0) = 0.4$. For this rather large deviation from the

96

stationary curve the first controller gives an unstable response but the second one is more robust and shows a satisfactory response. In [3] a sliding mode controller is designed for this
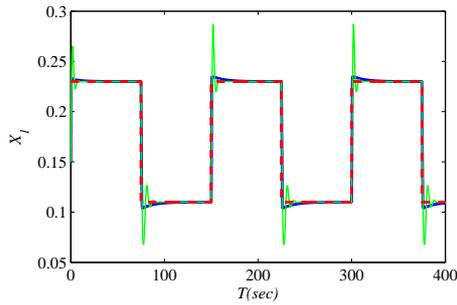
Fig. 13. Red dashed line is reference signal, blue curve is for controller designed using non-stationary points and the green response is the response from controller designed for stationary points only.
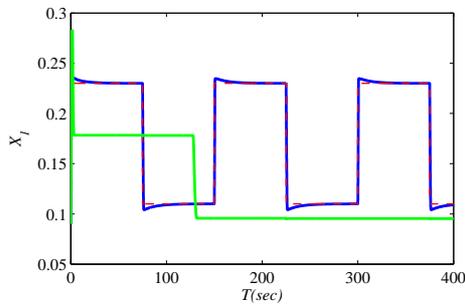
Fig. 14. Red dashed is reference signal, blue curve is for the controller designed with QFT for non-stationary points and the green response is the response for the controller designed for stationary points only.

system. In Fig. 15 that sliding mode controller is compared to the PID controller designed with the QFT method. The minimum and maximum values of the square wave are close to the maximum values that the system can reach. The systems are simulated for an initial values that causes large initial error in the control. As can be seen in Fig. 15 the PID controller response has an acceptable response though a significant overshoot. However, from an implementation point of view the PID controller is clearly preferable.

VI. CONCLUSION

In this paper a method based on QFT is used to design simple linear controllers for mildly nonlinear systems. The design is based on local linearization of the nonlinear system. In addition to classical linearizations around only equilibrium points non-equilibruim points are taken into account as well. Simulation results demonstrates that this method improve both transient response and robustness of the controller. In order to facilitate the design procedure loop-shaping is carried out with using an optimization algorithm based on Genetic Algorithm.
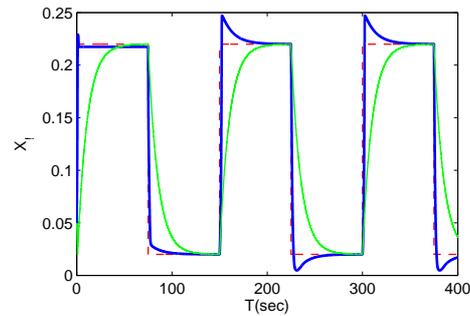
Fig. 15. Red dash line is the reference signal, the blue one is the QFT controller and the green one is the sliding mode controller response. $x_1(0) = 0.05, x_2(0) = 0.3$

REFERENCES

[1] A. Bafios, O. Yaniv, and F.J. Montoya. Nonlinear QFT based on local linearization. *International Journal of Control*, 2003.
[2] W. Chen, D.J. Ballance, and Y. Li. Automatic loop-shaping in QFT using genetic algrithm. *3rd Asia-Pacific Conference on Control and Measurement, pages 175-179, China, Dunhuang*, Sep 1998.
[3] Mehmet Onder Efe. Model refernce sliding mode control for bioreactor benchmark problem. *International Workshop on Variable Structure Systems*, 2008.
[4] C. Fransson, T. Wik, B. Lennartson, M. Saunders, and P. Gutman. Non-conservative robust control: Optimized and constrained sensitivity functions. *IEEE Transactions on Control Systems Technology, 17:298-308*, May 2009.
[5] C. M. Fransson, B. Lennartson, T. Wik, and K. Holmström. Multi criteria controller design for uncertain MIMO systems using nonconvex global optimization. In *40th IEEE Conference on Decision and Control*, volume 1-5, pages 3976–3981, Orlando, FL, USA, Dec 2001.
[6] P.O. Gutman. *Manual for the toolbox for Robust Control System Design for use with Matlab*, May 1996.
[7] I. Horwitz. *Quantitative Feedback Design Theory*. QFT Publications., 1993.
[8] T. A. Johansen, K. J. Hunt, P. J. Gawthrop, and H. Fritz. Off-equilibrium linearisation and design of gain-scheduled control with application to vehicle speed control. pages: 167-180. *Control enginnering practice*, 1998.
[9] H. Khalil. *Nonlinear systems*. Prentice Hall., 2003.
[10] W. T. Miller, R. S. Sutton, and P. J. Werbos. *Neural Networks for Control*. MIT Press, pp.387-402, 1990.
[11] P. S. V. Nataraj and M. M. Deshpande. Automated synthesis of fixed structure QFT controller using interval constraint satisfaction techniques. In *Proceedings of the 17th IFAC World Congress*, pages 4976–4981, Seoul, South Corea, July 6-11 2008.
[12] P. S. V. Nataraj and S. Tharewal. An interval analaysis algorithm for automated controller synthesis in QFT design. *Journal of Dynamic Systems, Measurement, and Control, Vol. 129. pages:311-321*, 2007.
[13] V. Olesen, C. Breitholtz, and T. Wik. Tank reactor temperature control using quantitative feedback theory. In *Proceedings of the 17th IFAC World Congress*, pages 4970–4975, Seoul, South Corea, July 6-11 2008.
[14] T. Schweickhardt and F. Algöwer. Linear control of nonlinear systems based on nonlinearity measures. *J. Process Control*, 17:273–284, 2007.
[15] T. Schweickhardt and F. Allgöwer. A robustness approach to linear control of mildly nonlinear processes. *International Journal of Robust Control*, 17:1163–1182, 2007.
[16] T. Schweickhardt and F. Allgöwer. On system gains, nonlinearity measures, and linear models for nonlinear systems. *IEEE transactions on automatic control*, 54(1):62–78, 2009.
[17] T. Wik, C. M. Fransson, and B. Lennartson. Feedforward feedback controller design for uncertain systems. In *42th IEEE Conference on Decision and Control*, pages 5328–5334, Maui, Hawaii, USA, Dec 2003.

OPTIMAL CONTROLLED VARIABLE SELECTION FOR INDIVIDUAL PROCESS UNITS IN SELF OPTIMIZING CONTROL WITH MIQP FORMULATIONS

Ramprasad Yelchuru[a], Sigurd Skogestad[b]

[a,b] Department of Chemical Engineering

Norwegian Science and Technological University, Trondheim, 7032

Corresponding Author's E-mail: skoge@chemeng.ntnu.no

Keywords: Optimal operation; selection of controlled variables; measurement combination; plantwide control; Mixed Integer Quadratic Programming

EXTENDED ABSTRACT

Optimal operation aids in improved productivity and profitability of the process plants. To facilitate the optimal operation in the presence of disturbances, the optimal control structure selection is important. The decision on which variables should be controlled, which variables should be measured, which input variables should be manipulated and which links should be made between them is called control structure selection. Usually, control structure decisions are based on the intuition of process engineers or on heuristic methods. This does not guarantee optimality and makes it difficult to analyze and improve the control structure selection proposals.

This paper considers the selection of controlled variables (CVs) associated with the unconstrained degrees of freedom. We assume that the CVs c s are selected as a subset or combination of all available measurements y. This may be written as

$c=Hy$ where $ny \geq nc$;

ny: number of measurements, nc: number of CVs = number of unconstrained DOFs

where the objective is to find a good choice for the matrix H. In general, we also include inputs (MVs) in the available measurement set y.

Skogestad and coworkers have proposed to use the steady state process model to find "self-optimizing" controlled variable as combinations of measurments. The objective is to find 'H' such that when the CVs are kept at constant set points, the operation gives acceptable steady state loss from the optimal operation even in the presence of disturbances.

The theory for self-optimizing control (SOC) is well developed for quadratic optimization problems with linear models. This may seem restrictive, but any unconstrained optimization problem may locally be approximated by this. The "exact local method" of Halvorsen et al. (2003) handles both disturbances and measurement noise. The problems of finding CVs as optimal variable combinations (c=Hy, where H is a full matrix) are found to be difficult to solve numerically (Halvorsen, 2003), but recently it has been shown that it may be reformulated as a quadratic optimization problem with linear constraints (Alstad et al., 2009).

We consider three interesting problems related to finding 'H':

1) Selection of CVs as combination of best subset of n measurements. Where $n \in \{nu : ny\}$
2) Selection of CVs as combination of disjoint measurement subsets using all measurements
3) Selection of CVs as combination of disjoint measurement subsets using n measurement subset. Where $n \in \{nu : ny\}$

We consider the solution of these problems when applied to the exact local method formulation of Halvorsen et al. (2003). Problem 1 has been solved by Kariwala and Cao (2009) and MIQP based approaches by Yelchuru et al. (2010).

Problem 2 is more appealing for practical usage of the SOC concepts than Problem 1, as using MVs of a process unit to control CVs as the measurement combinations (i.e. all/subsets of measurements) of the same process unit in a process flow sheet. This improves the process controllability significantly as the MVs and CVs are local than that of using CVs as combinations of all the measurements of the process flow sheet. Even though Problem 2 makes intuitive sense it is still an open problem. Following the QP formulation proposed by Alstad et al., 2009, we propose to solve Problem 2 by solving a MIQP problem for a given measurement subset to find the CVs as combinations of disjoint measurement subsets.

In this paper we propose a method to solve Problem 2 by reformulating the exact local method problem formulation for a given measurement subset as MIQP problem. Further we extend the formulation to solve Problem 3. The developed methods are evaluated on a toy problem and on a distillation column case study with 41 trays. The developed MIQP based methods for solving Problem 2 and Problem 3 in SOC are generic and can easily be evaluated for any system.

References

1. S. Skogestad, Plantwide control: the search for the self-optimizing control structure. Journal of Process Control, 10(5), 487 - 507.
2. J. Halvorsen, S. Skogestad, J.C. Morud, and V. Alstad., Optimal selection of controlled variables. Industrial Engineering and Chemistry Research, 42, 14, 3273 – 3284, 2003
3. V. Kariwala and Y.Cao., Bidirectional branch and bound for controlled variable selection. Part II: Exact local method for self-optimizing control, Computers and Chemical Engineering, 33, 8, 1402 – 1414, 2009
4. S. Skogestad and I. Postlethwaite., Multivariable Feedback Control: Analysis and Design. John Wiley & Sons, Chichester, UK, 2nd edition, 2005.
5. V. Alstad, S. Skogestad, Eduardo S. Hori, Optimal measurement combinations as controlled variables, Journal of Process Control, 19, 138 – 148, 2009.
6. Y. Cao and V. Kariwala, Bidirectional branch and bound for controlled variable selection Part I. Principles and minimum singular value criterion, Computers and Chemical Engineering, 32, 2306 – 2319, 2008.
7. S. Skogestad. Dynamics and control of distillation columns – A tutorial introduction. Trans. IChemE Part A, 75:539-562, 1997
8. R. Yelchuru, S. Skogestad and H. Manum, MIQP formulation for Controlled Variable Selection in Self Optimizing Control, DYCOPS 2010, accepted for oral presentation.

MIQP formulation for Controlled Variable Selection in Self Optimizing Control

Ramprasad Yelchuru*, Sigurd Skogestad*, Henrik Manum*

*Department of Chemical Engineering
Norwegian Science and Technological University, Trondheim 7032.

Abstract In order to facilitate the optimal operation in the presence of process disturbances, the optimal selection of controlled variables plays a vital role. In this paper, we present a Mixed Integer Quadratic Programming methodology to select controlled variables $c=Hy$ as the optimal combinations of fewer/all measurements of the process. The proposed method is evaluated on a toy test problem and on a binary distillation column case study with 41 trays.

Key words: Optimal operation, selection of controlled variables, measurement combination, plantwide control, Mixed Integer Quadratic Programming

1. INTRODUCTION

To facilitate the optimal operation in the presence of disturbances, the optimal control structure selection is important. The decision on which variables should be controlled, which variables should be measured, which input variables should be manipulated and which links should be made between them is called control structure selection. Usually, control structure decisions are based on the intuition of process engineers or on heuristic methods. This does not guarantee optimality and makes it difficult to analyze and improve the proposals.

This paper considers the selection of controlled variables (CVs) associated with the unconstrained degrees of freedom. We assume that the CVs c s is selected as individual measurements or combinations of subset or all available measurements y. This may be written as

$c=Hy$ where $ny \geq nc$;

ny : number of measurements;

nc: number of CVs = number of unconstrained DOFs = nu ;

where the objective is to find a good choice for the matrix H. In general, we also include inputs (MVs) in the available measurements set y.

Skogestad and coworkers have proposed to use the steady state process model to find "self-optimizing" controlled variable as combinations of measurements. The objective is to find H such that when the CVs are kept at constant set points, the operation gives acceptable steady state loss from the optimal operation in the presence of disturbances.

The theory for self-optimizing control (SOC) is well developed for quadratic optimization problems with linear models. This may seem restrictive, but any unconstrained optimization problem may locally be approximated suitably by this method. The "exact local method" of Halvorsen et al. (2003) handles both disturbances and measurement noise. The problems of finding CVs as

optimal variable combinations ($c=Hy$, where H is a full matrix) are originally believed to be difficult to solve numerically (Halvorsen, 2003), but recently it has been shown that SOC problem may be reformulated as a quadratic optimization problem with linear constraints (Alstad et al., 2009). The problem of selecting individual measurements as controlled variables (so H contains nc number of columns with a single 1 and rest of the columns are zero, mathematically $HH^T = I$) is more difficult. The maximum gain rule (Halvorsen et al., 2003) may be useful for prescreening but it is not exact. Even though these methods simplify the loss evaluation for a single alternative, it requires evaluation of every feasible alternative to find the optimal solution. As the number of alternatives increase rapidly with the process dimensions, resorting to exhaustive search methods to find the optimal solution is computationally intractable. Kariwala and Cao (2009) have derived effective branch and bound methods for the exact local method. These branch and bound methods require monotonicity property in the objective function. Furthermore, branch and bound methods are quite complex and they require derivation of good upper and lower bounds. This motivates the need to develop simple and efficient methods to find the optimal solution.

We consider three interesting problems related to finding H:

1. Selection of best individual measurements as CVs (select $n = nc$ measurements)

2. Selection of CVs as combination of all (ny) measurements.

3. Selection of CVs as combination of best subset of n measurements. Where $n \in \{nu, ny\}$

We consider the solution of these problems when applied to the exact local method formulation of Halvorsen et al. (2003). Problem 2 is the easiest one, Problems 1 and 3 involve structural decisions (discrete variables) and are therefore more difficult to solve. Nevertheless, from a practical point of view Problems 1 and 3 are important as

it is not wise to use more measurements than necessary to get an acceptable loss.

To solve Problem 1, Cao and Kariwala (2008) has developed bidirectional branch and bound methods to find the best individual measurements as CVs using minimum singular value criterion. To solve Problem 2, Alstad et al. (2009) has reformulated the self optimizing control problem as a constrained quadratic optimization problem. To solve Problem 3, Kariwala and Cao (2009) developed partial bidirectional branch and bound (PB³) methods to find best subset of measurements. The methods developed by Kariwala and Cao (2009) exploit the monotonic property of objective function in SOC problem and these methods are of limited/no use if the objective functions are not monotonic.

In this paper we propose a different method to solve Problems 1 and 3 by reformulating the exact local method problem formulation as a Mixed Integer Quadratic Programming (MIQP) problem. The MIQP formulation is simple and can easily be extended to other cost functions. The developed methods are evaluated on a toy problem and on a binary distillation column with 41 trays. The developed MIQP methods for SOC are generic and can easily be evaluated for any system.

2. EXACT LOCAL METHOD FORMULATION

We here review the "exact local method" formulation from Halvorsen et al. (2003) and its optimal solution from Alstad et al. (2009). We want to operate the plant close to optimal steady state operation, by using available degrees of freedom $u_{all} = \{u_{ac}\} \bigcup \{u\}$. The steady state cost function $J(u_{all}, d)$ is minimized for any given disturbance d. The possible process parameter variations are also included as disturbances. Few of the available degrees of freedom u_{ac} are used to implement optimally "active constraints", so that u contains only the remaining unconstrained steady state degrees of freedom.

The "reduced space" unconstrained optimization problem then becomes

$$\min_u J(u,d) \tag{1}$$

In this work we want to find a set of nc = nu controlled variables c, or more specifically optimal measurement combinations

$$c = Hy \tag{2}$$

such that a constant set point policy (where u is adjusted to keep c constant) yields optimal operation (equation (1)), at least locally. With a given d, solving equation (1) for u gives $J_{opt}(d)$, $u_{opt}(d)$ and $y_{opt}(d)$. In practice, presence of implementations errors and changing disturbances makes it impossible to have $u = u_{opt}(d)$ and results in deviation from optimal operation and this deviation is quantified as loss. The resulting loss (L) is defined as the difference

between the cost J, when using a non-optimal input u, and $J_{opt}(d)$ as in Skogestad and Postlethwaite (2005):

$$L = J(u,d) - J_{opt}(d) \tag{3}$$

The local second-order accurate Taylor series expansion of the cost function around the nominal point $(u*; d*)$ can be written as

$$J(u,d) = J(u^*,d^*) + [J_u \; J_d]^T \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix} \tag{4}$$
$$+ \frac{1}{2} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}^T \begin{bmatrix} J_{uu} & J_{ud} \\ J_{ud}^T & J_{dd} \end{bmatrix} \begin{bmatrix} \Delta u \\ \Delta d \end{bmatrix}$$

where $\Delta u = (u - u^*)$ and $\Delta d = (d - d^*)$. nu and nd are sizes of Δu and Δd. For a given disturbance ($\Delta d = 0$), the second-order accurate expansion of the loss function around the optimum ($J_u = 0$) becomes

$$L = \frac{1}{2}(u - u^{opt})^T J_{uu}(u - u^{opt}) \tag{5}$$
$$= \frac{1}{2} z^T z = \frac{1}{2}\|z\|^2$$

where $z \triangleq J_{uu}^{1/2}(u - u^{opt})$

In this paper, we consider a constant set point policy for the controlled variables which are chosen as linear combinations of the measurements as in equation (2).

The constant set point policy implies that u is adjusted to give $c_s = c + n^c$ where n^c is the implementation error for c. Here we assume implementation error is caused by the measurement error i.e. $n^c = H*n^y$. Now we want to express the loss variables z in terms of d and ny when we use a constant set point policy.

The linearized (local) model in terms of the deviation variables is written as

$$\Delta y = G^y \Delta u + G_d^y \Delta d \tag{6}$$
$$\Delta c = G\Delta u + G_d \Delta d \tag{7}$$

where $G = HG^y$ and $G_d = HG_d^y$

For a constant set point policy ($\Delta c_s = 0$) (Halvorsen et. al. 2003)

$$\Delta u^{opt} = -J_{uu}^{-1} J_{ud}\Delta d$$
$$\Delta y^{opt} = -(G^y J_{uu}^{-1} J_{ud} - G_d^y)\Delta d = F\Delta d \tag{8}$$

The F in equation (8) is the disturbance sensitivity matrix from disturbances d to measurements y at optimal operating points. And this F can be evaluated directly from optimal process operating data. For illustration, select the process operating data close to optimal operation for the possible process disturbances Δd and for these disturbances Δy^{opt} are known and disturbance sensitivity matrix F can be calculated directly. And this obviates the need to calculate G^y, G_d^y and J_{uu}, J_{ud}. The magnitudes of the disturbances d and measurement error ny are quantified by the diagonal scaling matrices W_d and W_n^y respectively. And we write

$$\Delta d = W_d d' \tag{9}$$

$$n^y = W_{n^y} n^{y'} \tag{10}$$

and by introducing the magnitudes of Δd and ny, the loss variables z in equation (3) can be written as

$$z = M_d d' + M_{n^y} n^{y'} \tag{11}$$

where
$$M_d = -J_{uu}^{1/2}(HG^y)^{-1}HFW_d \tag{12}$$

$$M_n = -J_{uu}^{1/2}(HG^y)^{-1}HW_{n^y} \tag{13}$$

$$Y = [(G^y J_{uu}^{-1} J_{ud} - G_d^y)W_d \quad W_n]_{ny \times (ny+nd)} \tag{14}$$

Using the equations (12), (13), (14) and (5) the loss can be rewritten as

$$L = \frac{1}{2}\left\| (J_{uu}^{1/2}(HG^y)^{-1}HY) \begin{bmatrix} d' \\ n^{y'} \end{bmatrix} \right\|^2 \tag{15}$$

The loss in equation (15) can be minimized with H as the decision variable. Similar to Halvorsen et.al. 2003 the norm of d', $n^{y'}$ is chosen to be constrained by $\left\| \begin{bmatrix} d' \\ n^{y'} \end{bmatrix} \right\| \leq 1$, and the opitmization problem is formulated to minimize the worst case loss and average loss as in Kariwala et al. (2008).

$$\min_H \frac{1}{2}\bar{\sigma}(J_{uu}^{1/2}(HG^y)^{-1}HY)^2 \tag{16}$$

$$\min_H \frac{1}{6(ny+nd)}\left\| (J_{uu}^{1/2}(HG^y)^{-1}HY) \right\|_F^2 \tag{17}$$

For these SOC problems Kariwala et.al. (2008) proved that the combination matrix H that minimizes the average loss in equation (17) is super optimal and in the sense that the same H minimizes the worst case loss in equation (16). Hence solving the optimization problem in equation (17) is considered in the rest of the paper. The scaling factor $\frac{1}{6(ny+nd)}$ does not have any effect on the solution of the equation (17) and hence it is omitted in the problem formulation.

Lemma 1: The problem in equation (17) may seem non-convex, but it can be reformulated as a constrained quadratic programming problem (Alstad et al., 2009).

$$\min_H \quad \|HY\|_F^2$$
$$st. \quad HG^y = J_{uu}^{1/2} \tag{18}$$

Proof: From the original problem in equation (17) the optimal solution H is non-unique. If H is a solution then $H_1 = DH$ is also a solution as $(J_{uu}^{1/2}(HG_y)^{-1})(HY) = (J_{uu}^{1/2}(H_1G_y)^{-1})(H_1Y)$ for any non-singular matrix D of nu x nu size. This means the objective function is unaffected by the choice of D. One implication is that HG_y can be chosen freely. We can thus make H unique by adding a constraint, for example $HG^y = J_{uu}^{1/2}$. More importantly this simplifies the optimization problem in equation (17) to optimization problem shown in equation (18). **End proof**

The problem in equation (18) is a constrained quadratic programming problem in measurement combination matrix H. We further reformulate the problem in (18) by vectorizing the decision matrix H to a vector x as described in Alstad et al., (2009).

First X is introduced as $X \triangleq H^T$. The matrices X and $J_{uu}^{1/2}$ are split into vectors as $X = [x_1 x_2 \cdots x_{nu}]; J_{uu}^{1/2} = [J_1 J_2 \cdots J_{nu}];$ and we further introduce the long vectors as below

$$X_\delta = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{nu} \end{bmatrix}_{(nu*ny)\times1} \quad ; J_\delta = \begin{bmatrix} J_1 \\ J_2 \\ \vdots \\ J_{nu} \end{bmatrix}_{(nu*ny)\times1}$$

and the large matrices

$$G_\delta^T = \begin{bmatrix} G^{y^T} & 0 & 0 & \cdots \\ 0 & G^{y^T} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & \cdots & G^{y^T} \end{bmatrix}_{(nu*nu)\times(ny*nu)} \quad ; Y_\delta = \begin{bmatrix} Y & 0 & 0 & \cdots \\ 0 & Y & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & \cdots & Y \end{bmatrix}_{(nu*ny)\times(nu*(nu+nd))}$$

As

$$\|HY\|_F^2 = \left[(HY)^T(HY)\right] = \left\| \begin{bmatrix} x_1^T Y \\ x_2^T Y \\ \vdots \\ x_{nu}^T Y \end{bmatrix} \right\|_F^2 = \left\| x_1^T Y \ x_2^T Y \cdots x_{nu}^T Y \right\|_F^2$$

$$= \left\| X_\delta^T Y_\delta \right\|_F^2 = \left\| Y_\delta^T X_\delta \right\|_F^2 = X_\delta^T Y_\delta Y_\delta^T X_\delta$$

and as J_{uu} is symmetric positive definite matrix, $J_{uu}^{1/2}$ is also symmetric positive definite $HG_y = G_y^T H^T = G_y^T X = J_{uu}^{1/2}$ and as

$$G^{y^T} X = \left[G^{y^T} x_1 \ G^{y^T} x_2 \cdots G^{y^T} x_n \right] = [J_1 J_2 \cdots J_{nu}]$$

the constraint can be written as

$$\begin{bmatrix} G^{y^T} x_1 \\ G^{y^T} x_2 \\ \vdots \\ G^{y^T} x_{nu} \end{bmatrix}_{(nu*nu)\times1} = \begin{bmatrix} J_1 \\ J_2 \\ \vdots \\ J_n \end{bmatrix}_{(nu*nu)\times1} \rightarrow G_\delta^T X_\delta = J_\delta .$$

In summary, the optimization problem (18) for finding the optimal H can be written as a constrained quadratic programming problem in the variables X_δ as follows.

$$\min_{XN} \quad X_\delta^T Y_\delta Y_\delta^T X_\delta \tag{19}$$

$$st. \quad G_\delta^T X_\delta = J_\delta$$

Note here that X_δ is a stacked vector of all the columns in X or H^T.

3. MIQP FORMULATION

The mixed integer quadratic programming (MIQP) approach provides a different method to solve Problems 1 and 3 described in introduction. Note here that Problem 1 and Problem 2 may be considered as special cases of Problem 3. The main advantages with the MIQP formulation are that these are simple, easily extendable and exact.

We start from the formulation given in (19) to find the optimal loss for the exact local method. Then we address this best measurement subset selection problem by formulating the problem in equation (19) as a Mixed Integer Quadratic Programming (MIQP) problem as described below. Let $\sigma_1, \sigma_2, \cdots \sigma_{ny} \in \{0,1\}$ be binary variables and let rest of the variables be the same as in equation (19). For the chosen measurement subset in the ny measurements, the decision variables associated to that binary variables are chosen to be bounded in a range of $-M$ to M. And these bounds are formulated as big-M constraints. Thus the MIQP problem with big-M constraints can be written as in equation (20).

$$\min_{x_{aug}} \quad \left(x_{aug}^T F x_{aug} \right)$$

$$st. \quad G_{new}^{y^T} x_{aug} = JN \tag{20}$$

$$P x_{aug} = n$$

$$\begin{bmatrix} -M & 0 & 0 & \cdots \\ 0 & -M & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ 0 & 0 & \cdots & -M \end{bmatrix} \sigma_i \le \begin{bmatrix} x_i \\ x_{ny+i} \\ \vdots \\ x_{(nu-1)*ny+i} \end{bmatrix} \le \begin{bmatrix} M & 0 & 0 & \cdots \\ 0 & M & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ 0 & 0 & \cdots & M \end{bmatrix} \sigma_i$$

$$\text{for } i = 1, 2, \cdots, ny$$

$$\sigma_i \in \{0,1\}$$

where $x_{aug} = \begin{bmatrix} X_\delta \\ \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_{ny} \end{bmatrix}_{(nu*ny+ny)\times 1}$ and

$F = [Y_\delta Y_\delta^T \ zeros(ny,ny)]; G_{new}^{y^T} = [G_\delta^T \ zeros(nu*ny,ny)];$
$P = [zeros(1, nu*ny) \ ones(1,ny)]$
and n is the measurement subset size.

In MIQP formulations, selections of a higher value for M in big-M constraints guarantee optimal solution, when bounds on decision variables are unknown. Note that each binary variable σ in inequality constraints in equation (20) provides bounds on nu number of elements in X_δ vector.

But higher M requires increased computational time in finding the optimal solution. Hence to find the suitable M value in finding optimal solution in an acceptable computational time, the constrained QP problem in (19) with ny measurements is solved. Based on the solution of equation (19) M is chosen as 2 times the maximum absolute value of the solution. Then MIQP problem in equation (20) is solved for different values of n from nu to ny. Later, the optimal measurement subset size n can be selected for the concerned process.

Lemma 2: The best individual measurements in exact local method (Problem 1) can be obtained from the MIQP problem formulation (equation (20)) solution for measurement subset size equal to nc.

Proof: As mentioned in the proof of Lemma 1, If H is a solution then $H_l = DH$ is also a solution for any non-singular matrix D of size $nuxnu$ as $(J_{uu}^{-1/2}(HG_y)^{-1})(HY) = (J_{uu}^{-1/2}(H_1 G_y)^{-1})(H_1 Y)$. Hence the objective function is unaffected by the choice of D.

Let H_{nc}. be the optimal solution to this MIQP problem (equation 20) for best nc measurements combination matrix. Now by choosing $D = H_{nc}^{-1}$ and we find the best indiviual measurements H_{im}.(Solution to Problem 1) **End proof**

Application to toy test problem. To illustrate the problem formulation, consider the toy problem of Halvorsen et a.l. (2003) which has two inputs $u = (u_1 \, u_2)^T$, one disturbance d and two output measurements $x = (x_1 \, x_2)^T$. The cost function is

$$J = (x_1 - x_2)^2 + (x_1 - d)^2$$

where the outputs depended linearly on u, d as

$$x = G^x u + G_d^x d \text{ with } G^x = \begin{bmatrix} 11 & 10 \\ 10 & 9 \end{bmatrix}; G_d^x = \begin{bmatrix} 10 \\ 10 \end{bmatrix};$$

At the optimal point we have $x_1 = x_2 = d$ and $J_{opt}(d) = 0$. Both the inputs and outputs are included in the candidate set of measurements y. For the example, the steady gain matrix from y to u (G^y), steady disturbance gain matrix from y to d (G_d^y), hessian of cost function with u, d J_{uu}, J_{ud} and disturbance, noise weight matrices W_d, W_n used are

$$C = \begin{bmatrix} y_1 \\ y_2 \\ u_1 \\ u_2 \end{bmatrix}; G^y = \begin{bmatrix} 11 & 10 \\ 10 & 9 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}; G_d^y = \begin{bmatrix} 10 \\ 10 \\ 0 \\ 0 \end{bmatrix};$$

$$J_{uu} = \begin{bmatrix} 244 & 222 \\ 222 & 202 \end{bmatrix}; J_{ud} = \begin{bmatrix} 198 \\ 180 \end{bmatrix}; W_d = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; W_n = 0.01 * \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The resulting optimal sensitivity matrix is computed as follows

$$Y = [(G^y J_{uu}^{-1} J_{ud} - G_d^y)W_d \quad W_n]_{ny \times (ny+nd)}$$

These matrices are used to get the stacked vector X_δ, J_δ, G_δ^T and Y_δ and the associated matrices in MIQP formulation in equation (20) are

$$F = \begin{bmatrix} 2 & 2 & -18 & 18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & -18 & 18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -18 & -18 & 162 & -162 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 18 & 18 & -162 & 162 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 2 & -18 & 18 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 2 & -18 & 18 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -18 & -18 & 162 & -162 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 18 & 18 & -162 & 162 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{12 \times 12}$$

$$G_{new}^T = \begin{bmatrix} 11 & 10 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 10 & 9 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 11 & 10 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 9 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}_{4 \times 12}$$

$$J_n = \begin{bmatrix} 11.6 \\ 10.5 \\ 10.5 \\ 9.6 \end{bmatrix}_{4 \times 1}$$

$$M = 2$$

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}_{1 \times 12}$$

4. RESULTS

4.1 Toy problem

The minimized loss function with the number of measurements used as CVs (i.e. the measurement combinations) is shown in Figure 1. From Figure 1, the loss is minimized as we use more number of measurements to find the CVs as the combinations of measurements. And the reduction in loss is very small when we increase the measurement subset size from 3 to 4.
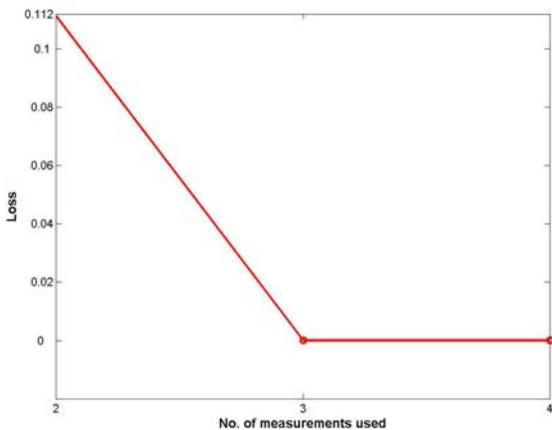

Figure 1. Optimal average loss with best measurement combinations vs no. of measurements used.

Based on the Figure 1, we can conclude that using CVs as combinations of 3 measurement subset is optimal for this toy problem.

4.2 Binary distillation column Problem

The binary distillation column and the associated data are taken from Skogestad (1997). The distillation column in LV-configuration with 41 stages is used. The 41 stage temperatures are taken as candidate measurements. Note that we do not include the inputs in the candidate measurements for this case study. The economic objective J for the indirect composition control problem is

$$\sqrt{J} = \Delta X = \sqrt{\left(\left(\frac{x_{top}^H - x_{top,s}^H}{x_{top,s}^H} \right)^2 + \left(\frac{x_{btm}^L - x_{btm,s}^L}{x_{btm,s}^L} \right)^2 \right)}$$

where ΔX is the root mean square of the relative steady stae compostion deviation. $x_{top}^H, x_{btm}^L, x_{top,s}^H, x_{btm,s}^L$, L and H denote the heavy component composition in top tray, light component composition in bottom tray, specification of heavy component composition in top tray, specification of light component composition in bottom tray, light and heavy key components respectively. We formulated the MIQP problem for the distillation column with 41 trays to find the 2 CVs as the combinations of 41 tray temperatures. An MIQP is set up for this distillation column with an M value of 1 in big-M constraints in equation (20). We solved the MIQP to find the CVs as the combinations of best measurement subset size from 2 to 41. The CPLX solver in Tomlab environment is used to solve the MIQP problem. The same problem is solved by downwards branch and bound, partial bidirectional branch bound methods of Kariwala and Cao (2009). The computational times (CPU time) taken by MIQP, Downward BAB, PB[3] method and exhaustiv|e search method are shown in Figure 3. Note that exhaustive search is not performed and an estimate of CPU time assuming 0.01 s for each evaluation is plotted. From Figure 3, it can be seen that the MIQP finds optimal solution in 6 orders faster than exhaustive search methods in computational (CPU) time. MIQP method runs relatively quickly for measurement subset size from 25 to 41, but it took fairly longer time for subset sizes from 10 to 19. As these subset sizes (10 to 19) have very high number of possibilities ($41C_{10}$ to $41C_{19}$), the longer time taken by MIQP method is justifiable. But on an average basis MIQP methods are slower by 1 order to PB[3] and 0.5 orders slower than Downwards BAB methods. In conclusion, even though the MIQP methods are not computationally attractive to that of Downwards BAB and PB[3] methods; the variation in the computational time by order of 1 is acceptable as these optimal CVs selection problems are performed offline. Despite these, MIQP method is valuable as the method is simple and can easily be extended to any quadratic cost functions to find optimal CVs in SOC framework. The minimized loss function with the number of measurements
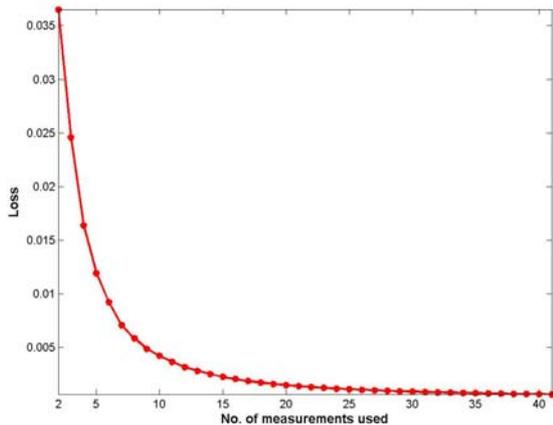
Figure 2. Optimal average loss using MIQP method with best measurement combinations vs no. of measurements used.
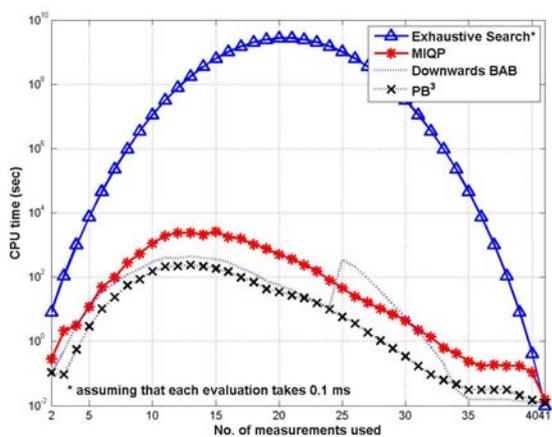

Figure 3. Comparsion of computation times

used for CVs (i.e. the measurement combinations) is shown in Figure 2. From Figure 2, it can be seen that the loss decreases rapidly when the number of measurements increased from 2 to 14, and from 14 very slowly. Based on the Figure 2, we can conclude that using CVs as combinations of 14 measurements subset is optimal for this 41 stage binary distillation column problem. MIQP formulations are easy than the BAB methods and structural constraints such as selection of certain number of measurements from top section, selection of certain number of measurement from bottom section can be done easily.

5. CONCLUSIONS

Optimal CV selection as measurement combinations to minimize the loss from the optimal operation is solved. The CV selection problem in self optimizing control framework is reformulated as a QP and CVs selection as combinations of measurement subsets is formulated as an MIQP problem. The developed MIQP based methods are easier compared to the bidirectional branch and bound methods reported in literature to find the CVs as

combinations of measurement subsets. And MIQP methods cover wider spectrum of quadratic based objective functions whereas bidirectional branch and bound methods are limited to objective functions with monotonic properties. MIQP based methods takes longer time than bidirectional branch and bound methods, but this is acceptable as the optimal CV selection problem is done offline. MIQP problem formulations are easily extendable for optimal measurement subset selection for systems with few structural constraints.

REFERENCES

V. Alstad, S. Skogestad, Eduardo S. Hori, Optimal measurement combinations as controlled variables, Journal of Process Control, 19, 138 – 148, 2009.

Y. Cao and V. Kariwala, Bidirectional branch and bound for controlled variable selection Part I. Principles and minimum singular value criterion, Computers and Chemical Engineering, 32, 2306 – 2319, 2008.

I. J. Halvorsen, S. Skogestad, J.C. Morud, and V. Alstad., Optimal selection of controlled variables. Industrial Engineering and Chemistry Research, 42, 14, 3273 – 3284, 2003

V. Kariwala and Y.Cao., Bidirectional branch and bound for controlled variable selection. Part II: Exact local method for self-optimizing control, Computers and Chemical Engineering, 33, 8, 1402 – 1414, 2009

V. Kariwala, Y.Cao and S. Janardhan., Local self-optimizing control with average loss minimization, Industrial Engineering and Chemistry Research, 47, 1150-1158, 2008.

S. Skogestad, Plantwide control: the search for the self-optimizing control structure. Journal of Process Control, 10(5), 487 - 507.

S. Skogestad and I. Postlethwaite., Multivariable Feedback Control: Analysis and Design. John Wiley & Sons, Chichester, UK, 2nd edition, 2005.

S. Skogestad. Dynamics and control of distillation columns – A tutorial introduction. Trans. IChemE Part A, 75:539-562, 1997

TOMLAB v7.1 - The *TOMLAB* Optimization Environment in Matlab (1999)

Dynamic Characteristics of Counter-Current Flow Processes

Jennifer Puschke[a], Heinz A Preisig[b]

[a]*RWTH Aachen, Templergraben 55, 52062 Aachen, Germany,*
Jennifer.Puschke@rwth-aachen.de
[b]*Chemical Engineering,, NTNU, N – 7491 Trondheim, Norway,*
heinz.preisig@chemeng.ntnu.no (corresponding author)

Abstract

In industry counter-current flow processes are common. Although these processes have been widely studied in literature, relatively little has been published on their dynamic behaviour. Two very common counter-current flow processes are heat exchangers and distillation columns. Ma's study based on dynamic models of heat exchanger's dynamic behaviour [1] reports an internal resonance effect, also earlier reported by Profos in 1943 [4] and Friedly in 1972 [3]. Here the study is repeated with lumped models, first for heat exchangers and thereafter for simple distillation columns. Not unexpectedly, the dynamic properties change gradually as the number of lumps increases towards the distributed systems and for high frequencies similar internal resonance effects evolve and the envelopes show a very low-order behaviour, which though somewhat surprisingly is independent of the number of lumps. Finally we show that the eigenvalues of the normed system matrix lie on a circle in the complex plane.

Keywords: Modeling, distributed/lumped model, Resonance effect, Frequency analysis

1. Introduction

An industrial process consists of a series of material transformations. Examples of such operations are batch reactors, compressors, heat exchangers and distillations. Many of these operations are based on two phases exchanging material and/or energy in the form of heat. The two phases are passing each other either in co-current or counter-current fashion often arranged in stages in each of which one drives the system towards equilibrium.

Although counter-current flow processes have been widely studied in literature, little of it reports on their fundamental dynamic behaviour. Common dynamic models for heat exchangers are simple first-order plus dead-time models. Exception are Profos (1943) [4] reporting the internal resonance effects, Friedly (1972) [3] derived reduced-order models and X H Ma [1] who derived a new set of high-fidelity low-order models also confirming the internal resonance effect, which years later has been show to exist in an experimental study by Grimm [2].

In a distributed model the temperatures on the inner and the outer tube of the heat exchanger are considered as continuous functions of time and spatial coordinates yielding a set of partial differential equations (PDE's). Ma's distributed model shows the presence of the internal resonance effect in the high frequencies domain. She splits the transfer function into a resonance and a non resonance part assuming a linear underlaying behaviour. This procedure yields high-fidelity analytical low-order models being the envelopes of the oscillating transfer function. For distillation columns however, no such behaviour has been reported. Since standard tray columns are better described as counter-current staged processes, Ma's study was repeated with lumped models. The model is constructed as a network of communicating capacities for each of which a mass and an energy balance is constructed. The energy balances are being

Figure 1: Model of the counter current double pipe heat exchanger with n equal three stages

transformed into the alternative state space of the intensive quantity temperature all of which forms a set of ordinary differential equations (ODE's). It is expected that, as the number of lumps approaches infinity, the solution approximates the solution of the distributed model hoping that the resonance effect shows also for low-order models. In a second step this is applied to a kind of a mass transport distillation column.

2. Lumped Model of the Heat Exchanger

Ma's double pipe heat exchanger is analyzed approximating it as series of heat-exchanging, paired lumps on one side representing the hot stream and on the other the cold one. Since the result shall be compared with distillation models only the counter-current flow pattern is being considered. Ma's work discusses different cases, which are based on different sets of assumptions. Here we focus on Ma's case I. We start with a 3-stage process as shown in Figure 1. The assumptions are:

- The total volume of stream A (consisting of all equal-sized lumps a_i) is the same as the total volume of stream B and both are constant: $V^A = V^B = V = const$

- The volume of each lump has the same size: $V_{a1} = \cdots = V_{an} = V_{b1} \cdots = V_{bn} = \dfrac{V}{n}$

- The heat transfer area O_i with $i = [1,2,\cdots, n]$ between two lumps with the same index i is constant and equal:
  $O_i = \dfrac{O_{overall}}{n} \forall i$, with $O_{overall}$ as the overall heat transfer area between stream A and stream B.

- Heat is only transferred between two lumps with the same index.

2.1. State Space Model Equations for n Stages

The energy balances is drawn up for each lump and solved for the temperature T resulting in the state space representation: $\dot{x} = Ax + Bu$ with the two matrices *A* and *B* being

$$
A =
\begin{pmatrix}
-\tau_A - d_A & \tau_A & & & & d_A & & \\
& \ddots & \ddots & & & & \ddots & \\
& & \ddots & \tau_A & & & & \ddots \\
& & & -\tau_A - d_A & & & & d_A \\
& d_B & & & -\tau_B - d_B & & & \\
& & \ddots & & & \tau_B & & \ddots \\
& & \ddots & & & & \ddots & \ddots \\
& & d_B & & & & \tau_B & -\tau_B - d_B
\end{pmatrix}
; \quad
B =
\begin{pmatrix}
0 & 0 \\
\vdots & \vdots \\
\tau_A & 0 \\
0 & \tau_B \\
\vdots & \vdots \\
0 & 0
\end{pmatrix}
\quad (3)
$$

And the $y = Cx$ with the matrix $C = \begin{pmatrix} 1 & \cdots & 0 \\ 0 & \cdots & 1 \end{pmatrix}$

With $\tau_m = \dfrac{\hat{V}^m}{V/n}$; $d_m = \dfrac{k_m O_i}{\rho_m c_{pm} V}$, $m \in [A,B]$

The quantities are: k_m :: heat transfer coefficient of stream m, O_i :: heat transfer area between two lumps, ρ_m :: density of the stream m, c_{pm} :: specific heat of stream m, V/n :: individual lump volume.

The state is $x = \begin{pmatrix} T_{a1} & \cdots & T_{an} & T_{b1} & \cdots & T_{bn} \end{pmatrix}^T$, the input is $u = \begin{pmatrix} T_\alpha & T_\beta \end{pmatrix}^T$ and the output is $y = \begin{pmatrix} T_\gamma & T_\delta \end{pmatrix}^T$.

2.2. Bode Plots

The dynamic behaviour of the models is depicted in Bode Plots of the model transfer functions. The transfer functions are derived by transforming the state space model into the frequency domain solving for the output y = x in dependence of u. The transfer function matrix is then simply: $G = C(sI - A)^{-1} B$ (4)

The transfer functions of input α to the output γ shows similar behavior as the one from the input β to the output δ. Only one of the two down stream responses, namely G_{11} from the input α to the output γ is shown in Figure 2. The same applies to the cross stream transfer functions, where only the transfer function G_{12} from the input α to the output δ is plotted. The behavior of all transfer functions approaches Ma's distributed model as the number of stages *n* approaches infinity which is also shown in Figure 2 as a reference.

Figure 2: Bode plots of the down stream transfer function G_{11} (left) and of the cross stream transfer function G_{12} (right) with different numbers of stages n. The parameters are chosen to be $d_A=d_B=0.01$; $\tau_A=1$; $\tau_B=1$.

Down Stream Response: The behavior of the transfer functions varies with the number of stages *n*. In the amplitude plot with an increasing number of stages *n* the slope of the amplitude decreases. As the number of stages approaches infinity the slope approaches zero. The latter implies that there exists only one gain. In the phase plot an increasing number of lumps increase the negative phase shift. For an infinite number of stages the phase lags go to minus infinity, which indicates the existence of a dead time. But there is not a resonance effect in the amplitude or the phase.

Cross Stream Response: The Bode plot shows the resonance effect in amplitude and phase. Furthermore one observes that the curves show a first corner at the frequency of ω=1Hz for the chosen set of parameters. Above this corner frequency the slope in the amplitude plot of the resonance part is in average minus one. And in the phase plot the resonance part average is -90 degree. The transfer functions with the number of stages being small than infinite show a decaying resonance part with increasing frequency, which finally disappears. The apparent length of the resonance part depends on the number of stages: With an increasing number of stages, the resonance part grows longer until the infinite case, where the resonance part does not decay anymore. Also the models with the number of stages being less than infinity, the final slope in the amplitude plot is -2 and the final phase lag is -180 degree. Hence this transfer functions show a second corner frequency under which the resonance part decays. Both corner frequencies depend on the number of stages.

Figure 3: The normed eigenvalues of the system matrix A in the complex plane with n=100.

2.3. Detailed Analysis of the Cross Stream Response

To get more information about the second-order behavior of the cross stream response, one needs the pole excess of the transfer function. Due to the structure of the matrices B and C only four entries of the matrix $(Is-A)^{-1}$ are relevant for the transfer functions matrix and only two of these entries for the cross-stream transfer functions. The zeros of the transfer functions are the zeros of the adjoint matrix adj(Is-A). For the number of stages *n = 3 or 4* it is easy to show that the respective adjoints have 2n-2 zero. Since the poles are the eigenvalues of A, their number is 2n. So the pole excess is

2n-(2n-2)=2, which explains the observed second-order behavior. In addition, by closer examination of the poles, one finds that the normed eigenvalues of the matrix A form a circle with radius one and the center at (-1,0) as shown in Figure 3.

3. Lumped Model of a Distillation Column

Figure 4 depicts a distillation column and an abstraction there off which underlays the construction of the model equations [5].

3.1. The State Space Model

The mass balances drawn for each lump, assuming a linear transfer law making the mass transfer proportional to the composition differentiate and solved for the concentration c of the lumps yields again a linear state space model (A,B,C,D), with the state x , input u and the output y being

$$x = \begin{pmatrix} c_{a1} & \cdots & c_{an} & c_{b1} & \cdots & c_{bn} \end{pmatrix}^T ; \quad u = c_\alpha \quad \text{and}$$

$$y = \begin{pmatrix} c_\beta & c_\gamma \end{pmatrix}^T \quad \text{with:}$$

Figure 4: Process of the distillation column (left) and model of the distillation column (right)

$$A = \begin{pmatrix}
-\tau_{a1}-\tau_\beta & & & & & & \tau_{a1} & & & & & \\
\tau_{a2} & -\tau_{a1}-d & & & & & & d & & & & \\
& \ddots & \ddots & & & & & & \ddots & & & \\
& & \tau_{an-1} & -\tau_{an-1}-d & & & & & & d & & \\
& & & -\tau_{an} & -\tau_{an}-\tau_\gamma & & & & & & 0 \\
0 & & & & & -\tau_{b1} & & & & \\
& d & & & & -\tau_{b2}-d & \tau_{b2} & & \\
& & \ddots & & & & \ddots & \ddots & \\
& & & d & & & & -\tau_{bn-1}-d & \tau_{bn-1} \\
& & & & \tau_{bn} & & & & -\tau_{bn}
\end{pmatrix}$$

;

$$B = \begin{pmatrix} 0 & \cdots & 0 & \tau_\alpha & 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}^T ; \quad C = \begin{pmatrix} 1 & 0 & \cdots & & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} ; \quad D = \begin{pmatrix} 0 & 0 \end{pmatrix}^T \quad \text{with} \quad \tau_m = \frac{\hat{V}^m}{V}$$

and $\quad \tau_{i2} = \tau_{i3} \ldots = \tau_{in-1}$, $i \in \{a, b\}$

These sparse system matrices for the distillation column show a similar structure as the system matrices of the heat exchanger.

3.2. Bode Plots

The transfer functions in the following Bode plots are obtained in the same way as this was done in the analysis of the heat exchanger.

The curves in the magnitude plot of the transfer function from the input α to the output β in the Bode plots of Figure 5 shows a resonance effect. This resonance part decays before the curves reach a multiple corner frequency, which depends on the number of stages n. With a larger number of lumps the length of the resonance part is longer. If the number of stages go to infinity one could assume, that there is only a steady state gain with resonance, whereas in the phase shift plot the resonance part does not appear. The general behavior of the curves with the number of stages n going to infinity suggests the existence of a dead time. The magnitude plot of the transfer function from the input α to the output γ (see Figure 5) shows a comparable response behavior as the transfer function from the input α to the output β. The differences are in the amplitude of the resonance part and the corner frequency. But the phase plot of the of the transfer function from the input α to the output γ is a resonance part.

By closer examination of the poles in the complex plane, one finds again that the standardized eigenvalues of *A* form a circle with radius one and the center at (-1,0) as Figure 3 shows.


Figure 5: Bode plots of the transfer function $G_{\alpha\beta}$ (left) from the input α to the output β and of the transfer function $G_{\alpha\gamma}$ (right) from the input α to the output γ with different numbers of stages n.

4. Conclusion

The dynamic characteristics of two counter-current processes are compared: a single tube heat exchanger and a staged distillation column. For both simple linear transfer models are assumed yielding linear systems that are of very similar structure. If normed, both show the same behaviour with respect to the system eigenvalues: they lay on a shifted unit circle in the complex domain. Both systems show resonance effects for some parts. Heat exchangers show it for cross-stream transfer functions, but not for down-stream transfer functions, whilst in distillation one finds the resonance also in the down stream transfer function, at least in the amplitude. In both cases, the magnitude of the resonance effect is a function of the number of lumps or stages.

In case of the heat exchanger the pole excess is 2, but the second corner frequency approaches infinity as the number of lumps approaches infinity. Thus for the distributed system the pole excess is only 1. This behaviour is also detected in the phase plot with a max phase shift of -180 degrees for finite number of lumps and -90 degrees for the distributed system.

The cross transfer functions for the distillation column behaves like a dead time for high frequencies, though the position of the multiple zeros shifts to higher and higher frequencies as the number of stages increases.

References

[1] Ma, X H. Dynamic Modelling, Simulation and Control of Heat Exchanger. PhD thesis, School of Chemical Engineering and Industrial Chemistry, University of New South Wales, Kensington, Sydney, Australia,1993.

[2] Grimm, R. Low-Order Modelling of the Dynamic Bahaviour of Heat Exchangers:Theory and Experimental Verification. Diploma thesis, 1999.

[3] Friedly, J C. Dynamic Behaviour of Processes. Prentice-Hall, Englewood Cliffs, NJ, 1972.

[4] Profos, P. Die Behandlung von Regelproblemen vermittels des Frequenzganges des Regelkreises. PhD thesis, ETH, Zuerich, 1943.

[5] Dones, Ivan and Preisig, Heinz, Graph tehory and model simplifcation. Case study: distillation column. Comp & Chem Eng. Accepted for publication, 2009.

Observer design for the activated sludge process

Marcus Hedegärd and Torsten Wik
Automatic Control, Department of Signals and Systems,
Chalmers University of Technology
SE-412 96 Göteborg, Sweden

The activated sludge process (ASP) is the most common process in biological wastewater treatment. However, they are very costly to operate, with energy for aeration being their largest cost. To optimize the process, a model is needed. The most widely used model for modeling of ASP:s is the Activated Sludge Model NO.1 (ASM1) [1]. It is physically based and a good compromise between accuracy and simplicity, and from this comes its popularity. Unfortunately, the model contains several concentrations that cannot reliably be measured online. Some of these are degradable dissolved and particulate organic matter, and biomass concentration. Totally Suspended Solids (TSS) measurements give indication of biomass concentration though. Online substrate analyzers have been available for many years but have historically been considered unreliable. Under the assumption that all relevant concentrations in the ASM1 were available online, the operation of ASP:s could be optimized to an extent that is not possible today. Some of the inputs to be optimized are the different input flowrates (water), TSS in the sludge recycle flow, the aerobic volume, external carbon addition and the air flowrates in the aerobic compartments. Motivated by this, an observer has been designed for the process based on the ASM1.

Observers based on the ASM1 have earlier been formulated by [3] and [4]. Both of these are for one aerobic reactor in the benchmark model [2] and the biomass concentration is assumed to be known and constant. A reduced order observer model is used in both cases. In [3] an altered version of the ASM1 which only includes one kind of substrate is used as the model for a nonlinear observer. In [4] an extended Kalman filter approach is taken. Different sets of measurements are considered in these cases but in both it is concluded that all relevant variables cannot be estimated at the same time. When lab analysis of substrate in the input is fed to the observers they are convergent though.

In this work we try to estimate all concentrations including unknown inputs in a reduced order model of two aerobic reactors with an extended Kalman filter. The success to estimate all unkwnown inputs compared to [3] and [4] relies on the following features:

- Contrary to the work mentioned above, the process considered here is *predenitrifying with postnitrification in trickling filters*. This means that nitrification stands for a very small portion of the aerobic reactions in the ASP, which in turn leads to that less

concentrations in ASM1 need to be considered. The concentrations in the reduced order model are: S_O (oxygen), S_S (readily biodegradable substrate), X_S (slowly biodegradable substrate) and X_{BH} (heterotrophic biomass). Among these, only oxygen is measured.

- By including two reactors, one additional measurement of oxygen is gained and better coupling between the states is achieved. The process model holds totally 11 states.

- The assumption that the biomass concentration is constant can for most plants in reality only be assumed on a very short time basis (hours). This can namely vary fast with varitions in the ratio between the input flowrates and with TSS in the return sludge flow. On the other hand, the composition of the sludge can be assumed to be slowly time varying. Instead of estimating the unknown input of biomass directly the observer estimates a parameter $\gamma_{X_{BH}}$: the ratio of the return sludge being heterotrophic biomass. The sludge concentration in the reactors can be simulated from TSS measurements in the recycle flow. This approach allows for larger variations in biomass concentration.

- The unknown input concentrations: S_S, X_S and the parameter $\gamma_{X_{BH}}$ are all modeled as random walk processes.

Initially, it was considered to estimate $\gamma_{X_{BH}}$ in the upstream anoxic compartments by using that some of the Monod expressions in the ASM1 model then can be assumed to be saturated. The developed EKF gives close to unbiased estimates of all states. The transient of the filter for three of the estimates of simulated variables are shown in Figure (1). The parameter $\gamma_{X_{BH}}$ is shown for a longer time period because of its slower variation.

(a) Readily biodegradable substrate at the inlet of the first reactor.

(b) Slowly biodegradable substrate at the inlet to the first reactor.

(c) The parameter $\gamma_{X_{BH}}$
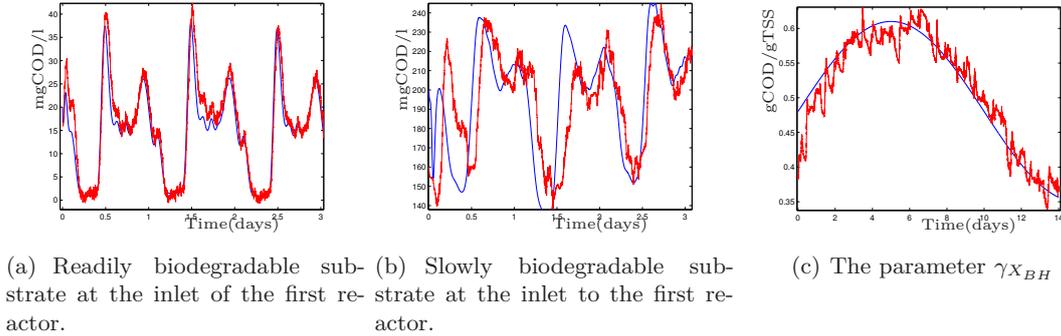
Figure 1: Estimated (noisy) and true variables (smooth).

Unfortunately, the system can be very sensitive to model errors, especially for errors in the $K_L a$ (Oxygen transfer) function, which is known to be time varying. To make the filter useful in an application it is most probably necessary to estimate the $K_L a$ function on a continuous basis. Most methods for estimating this function needs excitation of the air flowrate. The purpose of the observer is to save money and excitation is costly. However a $K_L a$ estimation method where the air flowrate is excited with small amplitude has been described in [5].

References

[1] M. Henze, C.P.L. Grady, W. Jr., Gujer, G.V.R. Marais, T. Matsuo, *Activated sludge model no. 1, IAWQ Scientific and Technical Report No. 1*, London, UK, 1987.

[2] B. Boulkroune a, M. Darouach, M. Zasadzinski, S. Gillé, D. Fiorelli, *A nonlinear observer design for an activated sludge wastewater treatment process*, Journal of Process Control 19 (2009) 1558–1565

[3] F. Benazzi, K.V. Gernaey, U. Jeppsson and R. Katebi, *On-line estimation and detection of abnormal substrate concentrations in WWTP:S using a software sensor: a benchmark study*, Environmental Technology, Vol. 28. pp 871-882

[4] J. Alex, J.F. Beteau, J.B. Copp, C. Hellinga, U. Jeppsson, S. Marsili-Libelli, M.N. Pons, H. Spanjers, H. Vanhooren, *Benchmark for evaluating control strategies in wastewater treatment plants*, in: European Control Conference 1999, ECC'99, Karlsruhe, Germany, August 31–September 3 1999.

[5] G. Olsson, B. Newell, *Waste water treatment systems: modeling, diagnosis and control*, Sweden and Australia, 1999.

Greenhouse Illumination Control

Anna-Maria Carstensen and Torsten Wik

Automatic Control, department of Signals and Systems

Chalmers University of Technology

SE-412 96 Göteborg, Sweden

Illumination of European greenhouses consumes about as much electricity as the whole of Sweden. Today, the greenhouse industry almost exclusively uses HPS lamps (High Pressure Sodium lamps). The spectrum of the HPS lamp fits poorly with the action spectrum of the photosynthesis, and it is estimated that more than 75% of this illumination is wasted. By combining High Brightness LEDs of different wavelengths, a spectrum closer to the action spectrum of the photosynthesis can be achieved. Experiments carried out with the Heliospectra HB LED lamp on the crop basil showed 40% improved efficiency compared to HPS lamps.

Further improvements are possible, since the LED lamp allows the intensity of each waveband to be adjusted to the varying needs of the plants. A plant's need, and its efficiency in using light, can be measured through plant fluorescence and reflectance. The aim of this project is to derive mathematical models that correlate plant growth with reflectance and illumination, and fluorescence with physiological state of the plant and plant stress. These models will then be used for a control system for the LED lamp such that the plants can modulate their own light environment. Expected results include decreased energy use, predictable and earlier harvest, and higher crop quality.

Reflectance and growth rate

In general, reflectance from plants are related to chlorophyll content in leafs, leaf area index and plant height. Estimates of green biomass has previously been done from remote sensing of reflectance from satellites as well as on ground level through the use of reflectance indices such as the Normalized Difference Vegetation Index (NDVI). These are based on relations between the reflectance of the wavelengths that are the least and the most sensitive to chlorophyll content.

Since plant reflectance is sensitive to chlorophyll content it is reasonable to assume that it also can be correlated to growth rate. We hope to experimentally find a correlation between growth rate and plant reflectance. In our experiments we combine measurements of dry matter content in plants, with remote (at a distance of about one meter) and on-leaf measurements of plant reflectance. The relationship between growth rate and reflectance will be used for non-destructively determining biomass content and predicting the time for harvest, as well as controlling the timing of the crop.

Fluorescence and plant stress

Fluorescence measurements contains information about the physiological state of plants. The physiological state of a plant is affected by access to specific wavelengths of light (especially 680 and 700 nm), CO_2, water and nutrients. Unbalanced access to these growth factors blocks the transfer of absorbed light energy and causes de-excitation of chlorophyll molecules, which gives rise to fluorescence. Hence, fluorescence signals reveal the dynamics of a plants physiology and can be used to measure if the plants are stressed due to lack of any of these parameters.

From fluorescence measurements one can determine whether the plants are exposed to unnecessary high or damaging light intensities. High light intensities, such as mid-day light, causes chlorophyll and other central parts in the photosynthetic apparatus to break down, leading to so-called photoinhibition. Within our project we will build a mathematical model of plant fluorescence and use the model to determine, e. g., when maximum photosynthesis is achieved. This information will be used to control the light intensity from the lamp, and prevent damaging light intensities as well as waste of energy. Furthermore, certain repair mechanisms of photoinhibited plants are trigged by specific frequencies. When photoinhibition is detected from fluorescence signals, this information could be used to enable light with repairing functions.

Model predictive control for plant-wide control of a reactor-separator-recycle system

Dawid Jan Białas[1], Jakob Kjøbsted Huusom[1], John Bagterp Jørgensen[2], Gürkan Sin[1]

[1] Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

[2] Department of Informatics and Mathematical Modelling, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

ABSTRACT

Ethylene glycol is produced by reaction of ethylene oxide and water in continuous-stirred tank reactor. To increase the overall conversion, the unreacted substrates are separated from the product in a distillation column and recycled to the reactor. This process is an example of a reactor-separator-recycle system that is widely used for manufacture of chemicals in industrial scale. We use a mathematical dynamical model of this process to investigate Model Predictive Control structures for plant-wide control of reactor-separator-recycle system.

Previously, different regulatory control structures and strategies for plant-wide control of reactor-separator-recycle system have been investigated. This work we investigate linear Model Predictive Control (MPC) for plant-wide control of the ethylene glycol reactor-separator-recycle system. The MPC gives setpoints to the regulatory controllers. The regulatory control structure is determined using a plant-wide control methodology such that suitable trade-off between sensitivity to disturbances and agile tracking of setpoints is achieved. The MPC coordinates the operation of each regulatory controller such that disturbances are rejected faster with less upset of plant operation. Consequently, the MPC facilitates operations aimed to maximize the throughput, minimize operating costs and ensure product quality.

In one plant-wide control strategy, the following controlled variables and manipulated variables were paired: reactor feed composition vs. fresh water flow-rate, water fraction in the bottom vs. vapor flow-rate, glycol composition in the top vs. recycle flow-rate. The essence of MPC is to optimize the predicted process behavior over a future horizon by manipulating the inputs subject to a number of constraints (process and actuators). A nonlinear process model was developed to simulate the ethylene glycol reactor-separator-recycle system. The filtering and predictions in the MPC are based on various linear models (state-space and reduced input-output models). The controller performances are investigated for scenarios with different disturbances entering the process. The key contribution of this paper is a demonstration of linear MPC for plant-wide control of a reactor-separator-recycle system.

Fuel quality soft-sensor for control strategy improvement of the Biopower 5 CHP plant

Jukka Kortela, Sirkka-Liisa Jämsä-Jounela

Abstract— This paper aims to present an enhanced method for estimating fuel quality on a BioGrate combustion process and its use in a control strategy improvement. The dynamic model based method (DMBM) utilizes combustion power which is calculated using the oxygen consumption of the furnace and the energy balance of the boiler. The proposed method is tested with data from the industrial scale Biopower 5 CHP plant and compared with the method currently used in industry, and finally the results are analyzed and discussed.

I. Introduction

Biomass fired power plants are usually controlled by means of conventional feedback control strategies where the main measurements are obtained from steam generation. However, these strategies face challenges due to great delays in control schemes associated to fuel feed and air supply [4]. Disturbances in combustion should thus be detected as early as possible before they have a significant effect on the process and its operation.

The main disturbances to the process are caused by fuel quality variations. Even for the same type of bio fuels, their chemical properties may differ greatly; for example – on account of harvesting, storing and transporting conditions [11]. In order to control the combustion process in an optimal way, it is essential to compensate variations in the fuel quality.

Referring to the theoretical studies and practical tests by Kortela & Lautala [3] for a coal power plant, the fuel combustion power in the furnace can be estimated on the basis of the measured oxygen consumption. The on-line measurement of the oxygen consumed is used when a new cascade compensation loop was built to optimally control the fuel flow. In that control strategy, the set point of the feed mainly depends on the output of the master controller, which obtained the main reference signal from the drum pressure. Taking into account the oxygen consumption and total air feed, the combustion power was calculated. The fuel feed setpoint was also modified as a direct result of making use of this variable. It is reported that the amplitude and the settling time of the response of the generator power decreased to about one third of the original when this cascade compensation loop was added to the present system. Combustion power control (CPC) was implemented also in peat power plants [6], where it was able to stabilize the furnace. The control actions of the burning air flow decreased when variations in the oxygen consumption were eliminated. The control strategy could thus reduce the standard deviation

Corresponding author: `jukka.kortela@tkk.fi`, Aalto University, PL 16100, FI-00076 Aalto

of the flue gas oxygen content and the air flow could be lowered close to the optimal flow. As a consequence, the flue gas losses were reduced. Furthermore, the stabilized steam temperatures reduced thermal stress on superheaters and associated pipes. In addition an integrated optimisation and control system has been applied for minimization of (NO_x), (SO_2) and (CO) emissions in the bubbling fluidized bed boiler in [5], where the use of the combustion power control algorithm made possible to stabilise the burning conditions in co-combustion. This leads to the better control of flue gas emissions when using the CPC together with the expert system. Hence, the combustion power control was reported cutting steam pressure deviation by 50 %.

However, there are still some challenges and objectives in the combustion power method. For example, variations in the moisture of fuel should be taken into account in order to correct any estimation errors in the combustion power. It is reported that temperature measurement in the furnace (or in the boiler) and calculation of the steam enthalpy could be used to estimate the fuel quality. However, temperature measurements are sensitive to process and measuring disturbances and the system dynamics introduce a delay in the method using enthalpy measurements [3].

The model-based predictive control was used by Havlena & Findejs [2] to enable tight dynamical coordination between air and fuel to take into account variations in power levels. The results showed that this approach could be used to increase boiler efficiency while considerably reducing the production of (NO_x) emissions. Similar results are also reported on an application of a local model networks (LMN) based multivariable long-range predictive control (LRPC) strategy for a simulation of 200 MW oil-fired drum-boiler thermal plant [10].

This paper sets forth to introduce an enhanced method for fuel quality estimation and its use in a control strategy. The paper is organized as follows. In Section 2, the biopower plant process is presented. Section 3 presents an enhanced control strategy, combustion power calculation for a BioGrate process and dynamic models of a boiler. The process experiments with varying fuel quality and the diagnosis results are given in Section 4 in order to demostrate the applicability of the method, followed by the conclusions in Section 5.

II. Description of the process and its control strategy

In the Biopower 5 CHP plant, the heat used for steam generation is obtained by burning solid biomass fuel: bark, sawdust and pellets, which are fed to the steam boiler

together with combustion air. As a result combustion heat and flue gases are generated. The heat is then used in the steam-water circulatory process.

Fig. 1 shows the boiler part of the Biopower 5 CHP plant. The essential components of the water-steam circuit are an economizer, a drum, an evaporator and superheaters. Feed water is pumped from a feed water tank to the boiler. First the water is led to the economizer (4) that is heated by flue gases. The temperature of flue gases is decreased by the economizer and the efficiency of the boiler is improved and thus further optimized.
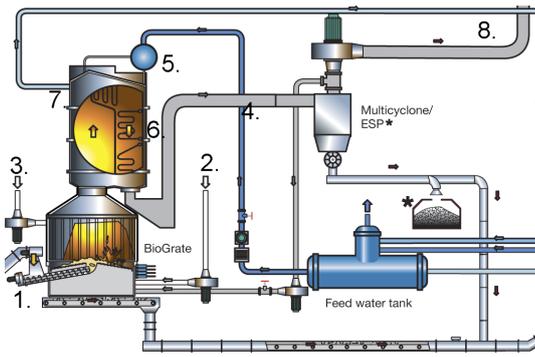

Fig. 1. 1. Fuel, 2. Primary air, 3. Secondary air, 4. Economizer, 5. Drum, 6. Evaporator, 7. Superheaters, 8. Superheated steam

From the economizer, heated feed water is led to the drum (5) and along downcomers into the bottom of the evaporator (6) tubes that surround the boiler. From the evaporator tubes the heated water and steam return back to the steam drum, where steam and water are separated. Steam rises to the top of the steam drum and flows to the superheaters (7). Steam heats up furthermore so it superheats. The superheated high-pressure steam (8) is led to a steam turbine, where electricity is generated.

A. Fuel composition and fuel quality

The composition and the quality of fuel have big effect on its heat value. Thus fuel quality is playing a key role when designing a control strategy of a biopower plant and guaranteeing its optimal operation. Common elements to all biomass fuels are carbon (C), hydrogen (H), oxygen (O) and nitrogen (N). In addition, biomass fuels contain substances from soil, such as water, minerals, rock materials and sulphur (S).

The actual volatile components of fuels are carbon, hydrogen and sulphur. Sulphur is an unwanted component, because it forms harmful sulphur dioxide, when it is burned. Nitrogen is also harmful; part of nitrogen reacts with oxygen and forms nitrogen oxides.

Water in fuel requires heat for its evaporation. Because of this, moisture decreases the heat value of fuel. Table I lists the composition and typical moisture content of wood fuels burned in the Biopower 5 CHP plant. The heat value of fuel can be determined by the equation that has been derived from heat values between combustible components

TABLE I
THE COMPOSITION OF WOOD FUELS BURNED IN THE BIOPOWER 5 CHP PLANT

Fuel	C	H	S	O	N	Ash	M
Wood	50.4	6.2	-	42.5	0.5	0.4	50
Pine	54.5	5.9	-	37.6	0.3	1.7	60
Spruce	50.6	5.9	-	40.2	0.5	2.8	60

and oxygen [1]. For solid fuel the following equation can be used.

$$q_f = 34.8 \cdot m_C + 93.8 \cdot m_H + 10.5 \cdot m_S$$
$$+6.3 \cdot m_N - 10.8 \cdot m_O - l_{25} \cdot m_{H_2O}[MJ/kg] \tag{1}$$

where m is the mass percent of a component and the sub index C is carbon, H hydrogen, S sulphur, N nitrogen, O oxygen and H_2O the water content of fuel. In order to use Equation 1, the composition of fuel has to be known.

B. Control strategy of the biograte process

The main aim of the control strategy of the biograte process is to produce desired amount of energy by keeping the drum pressure constant. At the highest level of the control strategy, the drum pressure control defines the power of the boiler. At a lower level, the needed boiler power is produced by controlling the amount of combustion air and fuel.

The primary air flow is controlled by the set point that comes from the pressure control. The fuel feed is controlled to track the primary air flow measurement. The needed amount of primary air and secondary air for diverse fuel and power levels are specified by air curves that have been calculated in a boiler design phase. The flue gas oxygen controller acts as a master controller while the set point of the secondary air controller is adjusted to provide the desired amount of excess air for the combustion.

III. ENHANCED CONTROL STRATEGY FOR BIOPOWER 5 CHP PLANT

In the enhanced control strategy, (taking into account the oxygen consumption and total air feed), the amount of fuel burned is estimated and the fuel feed setpoint is modified accordingly. The integrator in Fig. 2 removes steady-state offset in the control loop. Furhermore, the enhanced control strategy uses the dynamic model of the boiler to take into account variations in the moisture content of fuel. Therefore, it is possible to control the combustion process dynamically preventing steam temperature and pressure oscillations.

The biograte process is characterized by large time constants and long time delays. Thus the drum pressure control has to be tuned slow to maintain stability. The disturbances in fuel quality and fuel feed have a strong effect and proportionate direct correlation on steam pressure. These pressure and temperature disturbances settle slowly. To overcome this limitation, a dynamic model has been developed using the combustion power and energy balance of the boiler which are described next.
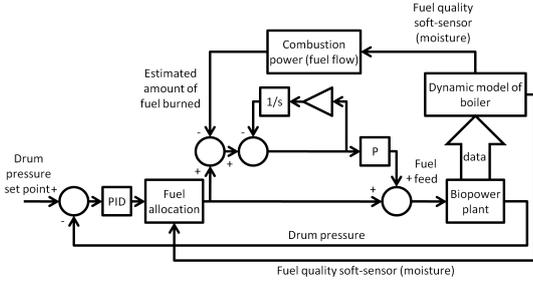
Fig. 2. Enhanced strategy based on the oxygen consumption and the energy balance of the boiler

A. Estimation of the combustion power

The combustion reaction in fossil fuel power plants occurs mainly between carbon and oxygen. Therefore, a good measure of heat generation in the furnace is the oxygen consumption [4].

When the fuel composition and combustion reactions are known, the combustion air and the composition of flue gases can be calculated. This information can then be used to conclude the completeness of the combustion and the correctness of the fuel-air ratio. Table II gives moles per unit of fuel from mass fractions of the fuel. The amount

TABLE II
MOLES OF THE COMPONENTS OF THE FUEL PER UNIT MASS

Component	Mass w_i ()	(g/mol)	(mol/kg)
C	$w_c(1-w/100)$	12.011	$w_C \cdot 10/n_C$
H	$w_h(1-w/100)$	2.0158	$w_H \cdot 10/n_H$
S	$w_s(1-w/100)$	32.06	$w_S \cdot 10/n_S$
O	$w_o(1-w/100)$	31.9988	$w_O \cdot 10/n_O$
N	$w_n(1-w/100)$	28.01348	$w_N \cdot 10/n_N$
Water	w	18.0152	$10/n_W$

of oxygen needed for fuel combustion can be determined from the reaction equations. By summing up the oxygen needed for different components and subtracting the amount of oxygen in the fuel, the theoretical amount of oxygen needed to burn completely one kilogram of the fuel is:

$$N_{O_2}^g = n_C + 0.5 \cdot n_{H_2} + n_S - n_{O_2} [mol/kg] \quad (2)$$

Air contains mainly oxygen and nitrogen. Argon is often included in nitrogen portion, so there is 21% oxygen and 79% nitrogen in the air. Theoretical amount of dry air needed is then:

$$N_{Air} = N_{O_2}^g \cdot \frac{1}{0.21} = N_{O_2}^g \cdot 4.76 [mol/kg] \quad (3)$$

In addition to combustion products, nitrogen N that comes with the air, is included in flue gases. There is 3.76 times more nitrogen compared with needed oxygen in flue gases. Incombustible components for example water are included as such. Flue gas flow for one kilogram of fuel is

$$N_{fg} = n_C + n_{H_2} + n_S + 3.76 \cdot N_{O_2}^g + n_{N_2} + n_{H_2O} [mol/kg] \quad (4)$$

Similarly, the flue gas losses per kilogram of fuel are

$$q_{fg}^g = (n_C C_{CO_2} + n_S C_{SO_2} + (n_{H_2O} + n_{H_2})C_{H_2O}$$
$$+(3.76 \cdot N_{O_2}^g + n_{N_2})C_N + (F_{Air}/(22.41 \cdot 10^{-3} \cdot m_f)$$
$$-4.76 \cdot N_{O_2}^g)C_{Air}) \cdot (t_{fg} - t_0)[J/kg] \quad (5)$$

where C_i is specific heat capacity i $(J/molT)$, F_{Air} is total air flow (m^3/s), m_f is fuel flow (kg/s), t_{fg} is temperature of the flue gas $(°C)$, and t_0 reference temperature $(°C)$.

The combustion power of the BioGrate boiler is estimated using Equations 6 - 10.

The total oxygen consumption is

$$N_{O_2}^{tot} = 0.21 \cdot n_{Air} - \frac{X_{O_2}}{100} \cdot n_{fg} [mol/s] \quad (6)$$

where $N_{O_2}^{tot}$ is total oxygen consumption (mol/s), n_{Air} is total air flow (mol/s), $X_{O_2}(t+\tau)$ is oxygen content of flue gas (%), and n_{fg} flue gas flow (mol/s).

The flue gas flow is

$$n_{fg} = m_f \cdot N_{fg} + n_{Air} + 4.76 \cdot m_f \cdot N_{O_2}^g [mol/s] \quad (7)$$

On the other hand, the oxygen consumption can be presented in the form:

$$N_{O_2}^{tot} = m_f \cdot N_{O_2}^g [mol/s] \quad (8)$$

and thus the fuel flow is

$$m_f = \frac{(0.21 - \frac{X_{O_2}}{100})n_{Air}}{N_{O_2}^g + \frac{X_{O_2}}{100}(N_{fg} - 4.76 \cdot N_{O_2}^g)} [kg/s] \quad (9)$$

Finally, the net combustion power for a given fuel flow is

$$P = (q_f - q_{fg}^g - q_{cr}) \cdot m_f [MW] \quad (10)$$

where q_{cr} is convection and radiation losses (MJ/kg).

B. Dynamic models of the boiler and estimation of moisture

Much of the behaviour of the boiler is captured by global mass and energy balances [12]. The heat released by the combustion of fuel is transferred to the water and steam of the boiler where each section can be considered as a thermal system. Therefore, the model is developed using the combustion power and the energy balances of the boiler to detect fluctuations in fuel quality.

The global mass balance is

$$\frac{d}{dt}(\varrho_s V_{st} + \varrho_w V_{wt}) = m_f - m_s [kg/s] \quad (11)$$

where ϱ_s is specific density of steam (kg/m^3), V_{st} is volume of steam (m^3), ϱ_w is specific density of water (kg/m^3), V_{wt} is volume of water (m^3), m_f is feed water flow (kg/s), and m_s steam flow rate (kg/s).

The global energy balance is

$$\frac{d}{dt}(\varrho_s u_s V_{st} + \varrho_w u_w V_{wt} + m_t C_p t_m)$$
$$= Q + m_f h_f - m_s h_s [kJ/s] \quad (12)$$

where u_s is specific internal energy of steam (kJ/kg), u_w is specific internal energy of water (kJ/kg), m_t is total mass

of the metal tubes and the drum (kg), C_p is specific heat of the metal (kJ/kgK), t_m is temperature of the metal (K) and Q is heat transfer from metal walls to steam/water (kJ/s).

Since the internal energy is $u = h - p/\varrho$, the global energy balance is

$$\frac{d}{dt}(\varrho_s h_s V_{st} + \varrho_w h_w V_{wt} + pV_t + m_t C_p t_m)$$
$$= Q + m_f h_f - m_s h_s [kJ/s] \tag{13}$$

where h_s is specific enthalpy of steam (kJ/kg), h_w is specific enthalpy of water (kJ/kg), and h_f specific enthalpy of water (kJ/kg).

and

$$V_t = V_{st} + V_{wt} [m^3] \tag{14}$$

Multiplying Equation 11 by h_w and subtracting the result from Equation 13 gives

$$(h_s - h_w)\frac{d}{dt}(\varrho_s V_{st}) + \varrho_s V_{st}\frac{h_s}{dt} + \varrho_w V_{wt}\frac{h_w}{dt}$$
$$-V_t\frac{dp}{dt} + m_t C_p\frac{dt_s}{dt}$$
$$= Q - m_f(h_w - h_f) - m_s(h_s - h_w)[kJ/s] \tag{15}$$

Equation 15 is used for economizer, evaporator, and superheaters subsections of the boiler and it is modified as necessary. If the drum level is controlled well, the variations in the steam volume are small and energy balance for a subsection is [8],[9]

$$\frac{dh_2}{dt} = \frac{1}{\varrho V}(Q + m_f h_f - m_s h_s)[kJ/(s \cdot kg)] \tag{16}$$

Energy balance for tube walls is

$$\frac{dt_m}{dt} = \frac{1}{m_t C_p}(P - Q)[K/s] \tag{17}$$

Heat transfer from metal walls to steam/water for convection heat transfer (superheaters) is

$$Q = \alpha m_s^{0.8}(t_m - T)[kJ/s] \tag{18}$$

where α is conversion heat transfer coefficient and for boiling heat transfer (water wall)

$$Q = \alpha(t_m - T)^3[kJ/s] \tag{19}$$

$$T = (T_1 + T_2)/2[°C] \tag{20}$$

where T_1 is input temperature $(°C)$, and T_2 output temperature $(°C)$.

The temperature of a secondary superheater is kept constant at specific temperature. De-superheating spray is used to achieve mixing between the superheated steam at the outlet of the preceding component. Because the attemperator has a relatively small volume, the mass storage inside it is negligible. The steady state energy balance yields

$$m_{in}h_{in} + m_{ds}h_{ds} = m_{out}h_{out}[kJ/s] \tag{21}$$

In normal operation, the steam flow m_{out} in the secondary superheater is imposed by the load controller, the enthalpy of primary superheater h_{in} is determined by the upstream superheater and enthalpy of de-superheating spray h_{ds} is nearly constant.

The value for the moisture parameter w is obtained by minimizing

$$\min J(w) = \sum_{i=0}^{N}(h - \hat{h}) \tag{22}$$

where N is prediction horizon, h is measured output enthalpy of the boiler (kJ/kg), and $\hat{h}$ estimated output enthalpy of the boiler (kJ/kg).

IV. TEST RESULTS OF THE FUEL QUALITY SOFT-SENSOR AND THE ENHANCED CONTROL STRATEGY

First, the performance of the fuel quality soft-sensor is tested with real data obtained from the Biopower 5 CHP plant. Next, the three different control strategies (the current control strategy, the combustion power control strategy and the enhanced control strategy proposed in this paper) are evaluated on the Biopower 5 CHP plant simulator using MATLAB simulation environment.

A. Performance of the fuel quality soft-sensor

Process tests were carried out in the Biopower 5 CHP plant and ten hours of data were logged with 1 second as the sampling time (benchmark interval). Two different fuels were varied during the time period. In order to show the performance of the fuel quality soft-sensor, the monitoring of the most important variables is investigated to study the effect of the fuel quality: drum pressure, the temperature of the furnace, the temperature over the secondary superheater and temperature of the flue gases. The value of the fuel moisture parameter w in Fig. 3 agrees with the power of the boiler. Although the fuel flow in Fig. 4 increases, the power of the boiler drops. Therefore, the value of the fuel moisture w captures the fuel quality well. The combustion power method estimates the combustion power of the fuel flow well when the fuel quality does not change significantly. Since the fuel type was varying greatly, there is an error in the estimation of the power in Fig. 5. The temperature of the furnace in Fig. 5 drops when the value of the fuel moisture w increases. However, the change in fuel quality is shown 20 minutes later if compared with the fuel moisture parameter. Moreover, the temperature of the furnace is disturbed by air flows that cool the furnace. Fig. 6 shows the performance of drum pressure control when the fuel type changes. Since the controller is not tuned to handle variations in fuel quality it saturates and the power of the boiler is not controlled. Moreover, the drum pressure is characterized with a large time delay. The temperature difference between the secondary superheater in Fig. 7 correlates with the power of the boiler. However, the temperature is greatly disturbed. On the other hand, the flue gas temperature in Fig. 7 increases when moisture content in fuel increases. Since more fuel and air are needed to achieve the same amount of energy, more heat flows through
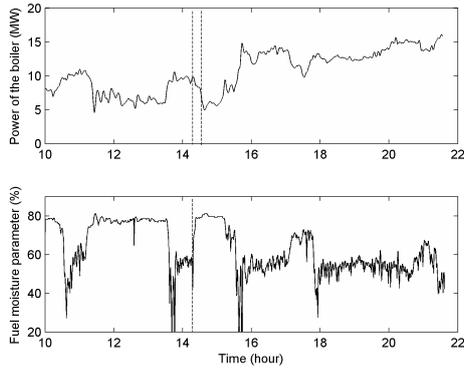
Fig. 3. The power of the boiler and the value of the fuel moisture parameter
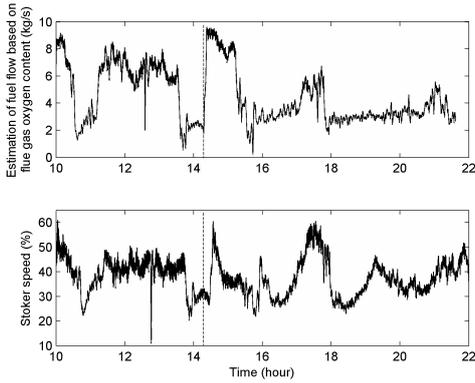

Fig. 4. The estimated fuel flow and the stoker speed

Fig. 5. Combustion power and temperature of furnace

Fig. 6. Drum pressure control

Fig. 7. The temperature difference between the secondary superheater and flue gas temperature

B. Comparision of control strategies.

Fig. 8 and 9 present disturbance situations in the fuel feed. The power demand of the boiler is kept on 13.5 MW while the moisture content of fuel is varied from 55 % to 65 %. The settling time in the response of the current control strategy is about 2 hours, whereas it is about 10 minutes when using the enhanced control strategy. With the combustion power control, there are minor oscillations. Fig. 10 and 11

Fig. 8. Disturbances in power and pressure.

the gases, and therefore altering its temperature. Also the measurement is greatly filtered. Therefore, it cannot be used as a measure of fuel quality. The validity of the fuel quality soft-sensor was tested by calculating the cross correlation between different values in Table III. The positive delay 378 s means that the moisture value shows the change in fuel quality earlier than any other value. Also value -0.74 means good correlation with the power of the boiler.

According to these tests, the fuel quality soft-sensor is a promising way to measure the changing fuel quality and could be used in a control strategy.

present boiler load disturbances. With the control strategy

TABLE III
CORRELATIONS OF DIFFERENT FUEL QUALITY INDICATORS

|	Corr	Delay (s)
Moisture content, Boiler power	-0.74	378
Furnace temperature, Boiler power	0.92	-666
Drum pressure, Boiler power	0.81	-424
Temperature over superheater, Boiler power	0.64	298
Flue gas temperature, Boiler power	-0.67	-1464
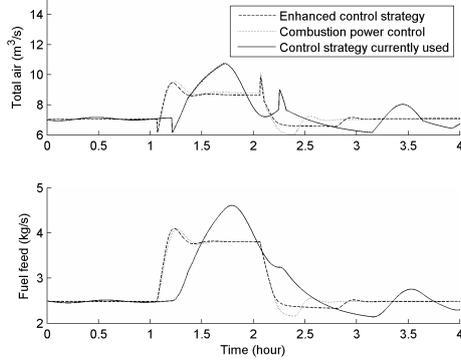
Fig. 9.   Disturbances in air and fuel.

used currently, load disturbances cause strong oscillations. Using the enhanced method no oscillations occur. Also it has half of the settling time compared with the method based only in oxygen consumption, because the fluctuating fuel quality is taken into account already in combustion power calculation.
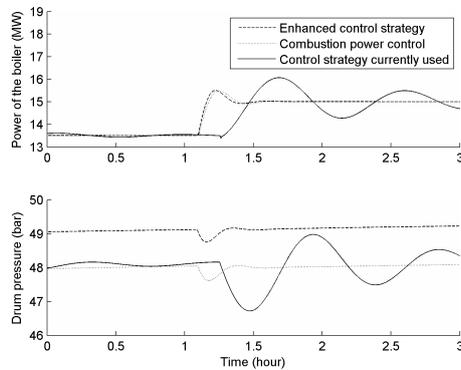

Fig. 10.   Load disturbances in power and pressure

The process simulation tests proved that the enhanced control strategy is able to efficiently stabilize the combustion process. The control strategy managed to keep the pressure level by using air flow, oxygen content and fuel quality soft-sensor for the estimation of the fuel flow and the allocation of fuel.

V. CONCLUSIONS

An enhanced method for estimating fuel quality on a BioGrate combustion process, along with its use in a control strategy improvement environment at the Biopower 5 CHP
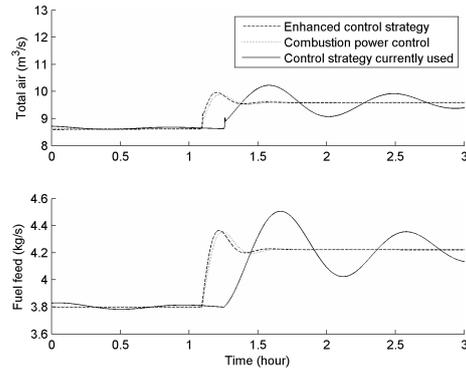

Fig. 11.   Load disturbances in air and fuel

plant was presented in this paper. Using the enhanced control strategy, it is possible to control the combustion process dynamically preventing steam temperature and pressure oscillations.

The fuel quality estimation method was tested with real data. The enhanced control strategy was tested in the controlled simulation environment. The results of the tests dramatically demonstrate that the enhanced control strategy efficiently stabilizes the combustion process.

REFERENCES

[1] Effenberger, H. (2000) *Dampferzeuger*. Germany: Springer.
[2] Havlena, V & Findejs, J. (2005) *Application of model predictive control to advanced combustion control Control*. Control Engineering Practice, Vol 13, Issue 6, June 2005, pp. 671-680.
[3] Kortela, U. & Lautala, P. (1981) A New Control Concept for a Coal Power Plant. *Proceedings of the 8th IFAC World Congress, Kyoto, Japan 1981*.
[4] Kortela, U. & Marttinen, A. (1985). Modelling, Identification and Control of a Grate Boiler. *American Control Conference, 1985, Boston, MA, USA 19-21 June 1985*. pp. 544-549
[5] Kortela, U., Ikonen, I., Kotajärvi H. & Heikkinen, P. (1994) *Modelling, Simulation and Control of Fluidized Bed Combustion Process*. International Journal of Power & Energy Systems, Vol 14, no. 3, pp. 92-97.
[6] Lehtomäki, K., Kortela, U., Wahlström, F. & Luukkanen, J. (1982) New Control Methods for Combustion Stabilization in Peat Power Plants. *The Joint Soviet-Finnish Symposium on Automation in Process Industries, Espoo, Finland 13-16 December 1982*.
[7] Leppäkoski, K., Mononen, J. & Kovács, J. (2000) Integrated Optimisation and Control System to Reduce Flue Gas Emissions. *Power Plants and Power Systems Control 2000: A Proceedings Colume from the IFAC Symposium Brussels, Belgium, 26-29 April, 2000*. pp. 113-118
[8] Lu, S. (1999) *Dynamic modelling and simulation of power plant systems*. Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy, Vol 213, Number 1/1999, pp. 7-22.
[9] Lu, S. Hogg, W. (2000) *Dynamic nonlinear modelling of power plant by physical principles and neural networks*. International Journal of Electrical Power & Energy Systems, Vol 22, Issue 1, January 2000, pp. 67-78.
[10] Prasad, G., Swidenbank, E. & Hogg, B. W. (1998) *A Local Model Networks Based Multivariable Long-Range Predictive Control Strategy for Thermal Power Plants*. Automatica, Vol 34, Issue 10, October 1998, pp. 1185-1204.
[11] Yin, C., Rosendahl, L. & Kær, S. (2008) *Grate-firing of biomass for heat and power production*. Progress in Energy and Combustion Science, Vol 34, Issue 6, December 2008, pp. 725-754.
[12] Åström, K. Bell, R. (1999) *Drum-boiler dynamics*. Automatica, Vol 36, Issue 3, March 2000, pp. 363-378.

Convex approximation of the static output feedback problem with application to MIMO-PID

Henrik Manum and Sigurd Skogestad*
Department of Chemical Engineering
Norwegian University of Science and Technology
N-7491 Trondheim

1 Introduction

In this contribution we derive convex approximations to the static output feedback (SOF) problem. This may seem like a restrictive formulation, but the problem of finding an optimal multiple input − multiple output (MIMO) proportional integral deriative (PID) controller can be posed on this form. In the literature it is proved that a problem closely related to SOF belongs to a class of problems that cannot be solved by polynomial time algorithms (NP-hard problems), and it is further conjectured that also the SOF problem is an NP-hard problem [Blondel and Tsitsiklis, 1997]. For an overview of different approaches to this problem the reader is referred to the survey paper by Syrmos et al. [1997].

2 Problem formulation

Consider a linear process on discrete form

$$
\begin{aligned}
x_{k+1} &= Ax_k + Bu_k + d_k, \\
y_k &= Cx_k + n_k,
\end{aligned}
\tag{1}
$$

where $x_k \in \mathbb{R}^{n_x}$ are the states, $u_k \in \mathbb{R}^{n_u}$ are the inputs and $y_k \in \mathbb{R}^{n_y}$ are the measurements. The $\mathcal{H}_2$ infinite horizon static output feedback (IH-SOF) problem is to choose an output feedback $u_k = -K^y y_k$ that minimizes the objective function

$$
J_{\text{IH-SOF}} = \sum_{i=0}^{\infty} x_i' Q x_i + u_i' R u_i
\tag{2}
$$

for all disturbances $d_k \in \mathbb{R}^{n_x}$ and noise-terms $n_k \in \mathbb{R}^{n_y}$. Throughout the paper we assume that $Q = Q' \geq 0$ and $R = R' > 0$. A related problem is the finite horizon static output feedback problem (FH-SOF), which is the same problem as IH-SOF, except that the objective function

*Corresponding author: skoge@chemeng.ntnu.no

to be minimized is finite:

$$
J_{\text{FH-SOF}} = x_N' P x_N + \sum_{i=0}^{N-1} \left(x_i' Q x_i + u_i' R u_i \right).
\tag{3}
$$

The problems posed above are *feedback* problems and as already mentioned they are conjectured to be NP-hard. To derive new convex approximations, we do not directly minimize the cost (3), but rather we minimize the (worst-case) *loss* to an optimal open loop controller, which in this case is the finite-horizon linear quadratic controller (FH-LQR). More precisely, the loss we want to minimize is

$$
L(d) = J_{K^y}(d) - J_{\text{FH-LQR}}(d),
\tag{4}
$$

where $J_{K^y}(d)$ is the cost of a particular (static output) feedback implementation and $J_{\text{FH-LQR}}(d)$ is the cost of a FH-LQR problem.

In the previous Nordic Process Control Workshop (Porsgrunn, 2009) we proposed a convex approximation based on implementing the first move of a sequence of open loop moves $\begin{bmatrix} u_0' & u_1' & \dots & u_{N-1}' \end{bmatrix}' = K y_0$. In this paper we present a new approximation, which seems to get closer to the global optimum, but at the expense of some more computational effort. The resulting problem is a quite large quadratic program (QP).

3 Solution methods

By using a result from Alstad et al. [2009] we pose the following problem as a convex relaxation of the SOF problem:

$$
\min_H \|HF\|_F
$$
$$
\text{s.t. } H \text{ on the form } \begin{bmatrix} H^y & H^u \end{bmatrix} = \begin{bmatrix} \text{diag}(K^y) & I \end{bmatrix}
\tag{5}
$$

Here $F = -(G^y J_{uu}^{-1} J_{ud} - G_d^y)$ is the optimal sensitivity matrix from the disturbance $d = x_0$ to the "measurements" $y = (y_0, y_1, \dots, y_{N-1}, u_0, u_1, \dots, u_{N-1})$ and J_{uu},

J_{ud}, G^y and G_d^y are derived from the objective function and the linear model as shown below.

Derivation of J_{uu} and J_{ud} Consider the FH-LQR problem, which is an *open loop* problem in the inputs $u = (u_0, u_1, \ldots, u_{N-1})$ defined by first writing the linear model $x_{k+1} = Ax_k + Bu_k$ as

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} A \\ A^2 \\ \vdots \\ A^N \end{bmatrix}}_{G_{x_0}^x} x_0 + \underbrace{\begin{bmatrix} B & & & \\ AB & B & & \\ \vdots & & \ddots & \\ A^{N-1}B & \cdots & \cdots & B \end{bmatrix}}_{G_u^x} u \quad (6)$$

By further defining $\bar{Q} = \text{diag}(Q, \cdots, Q, P)$ and $\bar{R} = \text{diag}(R, \cdots, R)$ and let the *disturbance $d = x_0$* we can write the FH-LQR problem as

$$J_{\text{LQR}}^*(d) = \min_u \begin{bmatrix} u \\ d \end{bmatrix}' \begin{bmatrix} J_{uu} & J_{ud} \\ J_{ud}' & J_{dd} \end{bmatrix} \begin{bmatrix} u \\ d \end{bmatrix} \quad (7)$$

with

$$J_{uu} = G_u^{x\prime} \bar{Q} G_u^x + \bar{R}, \quad (8)$$
$$J_{ud} = G_u^{x\prime} \bar{Q} G_{x_0}^x, \quad (9)$$
$$J_{dd} = G_{x_0}^{x}{}' \bar{Q} G_{x_0}^x. \quad (10)$$

Derivation of G^y and G_d^y Using the linear model $x_{k+1} = Ax_k + Bu_k$, $y_k = Cx_k + Du_k$, we have:

$$y_k = CA^k x_0 + \sum_{j=0}^{k-1} CA^j B u_{k-1-j} + Du_k, \quad (11)$$

which on matrix form can be written as

$$y_m = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix} = \underbrace{\begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{N-1} \end{bmatrix}}_{G_{x_0}^{ym}} x_0 + \ldots$$

$$+ \underbrace{\left(\begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ CB & 0 & \ldots & 0 & 0 \\ CAB & AB & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ CA^{N-2}B & CA^{N-3}B & \ldots & CB & 0 \end{bmatrix} + \text{diag}(D) \right)}_{G_u^{ym}} u. \quad (12)$$

We now have that

$$y = \begin{bmatrix} y_m \\ u \end{bmatrix} = \underbrace{\begin{bmatrix} G_u^{ym} \\ I \end{bmatrix}}_{G^y} u + \underbrace{\begin{bmatrix} G_{x_0}^{ym} \\ 0 \end{bmatrix}}_{G_d^y} \underbrace{x_0}_{d}. \quad (13)$$

4 Application to MIMO-PID

In order to use static output feedback synthesis for this problem we augment the plant output with the integrated output and derivative. The augmented plant can be written as

$$\begin{bmatrix} x_{k+1} \\ \sigma_{k+1} \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & I \end{bmatrix} \begin{bmatrix} x_k \\ \sigma_k \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u_k$$
$$\begin{bmatrix} y_k^P \\ y_k^I \\ y_k^D \end{bmatrix} = \begin{bmatrix} C & 0 \\ 0 & I \\ \frac{1}{T_s}C(A-I) & 0 \end{bmatrix} \begin{bmatrix} x_k \\ \sigma_k \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ B \end{bmatrix} u_k. \quad (14)$$

We can now use the method above to calculate a static output feedback K^y for the augmented output vector.

Example: Distillation We have calculated MIMO-PI and -PID controllers for "column A" in [Skogestad, 1997] and we found that in terms of closed-loop norms did the convex approximations come quite close to the true optimal SOF controllers (found by nonlinear search starting from the convex approximations). In addition we have tight bound in terms of the worst case loss from the LQR controller, and we found that the loss was less than 3.5% for both cases. This will be further illustrated in the presentation.

5 Conclusions

A new convex approximation to the static output feedback problem has been given, and we have shown that the approximation can be used to find MIMO-PID controllers for interesting chemical engineering cases, such as distillation control.

References

V. Alstad, S. Skogestad, and E.S. Hori. Optimal measurement combinations as controlled variables. *Journal of Process Control*, 19(1):138 – 148, 2009.

V. Blondel and J.N. Tsitsikilis. NP-Hardness of some linear control design problems. *Siam Journal of Control and Optimization*, 35(6):2118–2127, November 1997.

S. Skogestad. Dynamics and control of distillation columns - a tutorial introduction. *Trans IChemE, Part A*, 75:539–562, September 1997.

V.L. Syrmos, C.T. Abdallah, P. Dorato, and K. Grigoriadis. Static Output Feedback − A Survey. *Automatica*, 33(2):125–137, 1997.

Modeling and Optimization of Grade Changes for Multistage Polyethylene Reactors

Per-Ola Larsson, Johan Åkesson, Staffan Haugwitz, Niklas Andersson

I. INTRODUCTION

Polyethylene reactors are today able to produce different grades by manipulating inflows of raw material. It is imperative for polyethylene manufacturers to change product grades to increase their profitability as market demands change, but also due to market competition and raw material pricing. During grade transitions it is of importance that production of off-specification material, raw material and time is minimized. We present an optimization procedure for grade change of a Borstar® polyethylene plant.
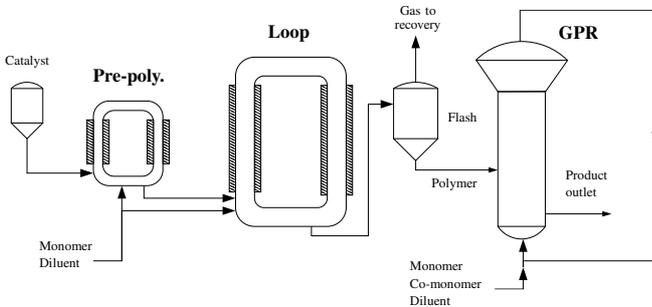

Fig. 1. Reactor chain of a Borstar® process: Pre-polymerization, Loop, and Gas phase reactor (GPR).

II. PLANT MODELING

The Borstar® polyethylene plant at Borealis AB incorporates two slurry reactors, pre-polymerisation and loop reactor, and a gas phase reactor, see Figure 1. The model of the plant includes both first principles, semi-empirical, and empirical relations. A total of 12 inputs flows, denoted $\mathbf{u}$, are available at optimization and outputs such as masses of both fluid and solid components, reaction rates, instantaneous and bed averaged component concentrations, split factor, catalyst and polymer properties, denoted $\mathbf{y}$, can be used. Together with algebraic variables $\mathbf{w}$, the model can be written in the general non-linear index 1 differential algebraic equation form

$$
\begin{aligned}
\mathbf{0} &= F(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{w}, \mathbf{u}) \\
\mathbf{y} &= g(\mathbf{x}, \mathbf{w}, \mathbf{u}).
\end{aligned} \tag{1}
$$

and contains approximately 70 differentiated variables, 180 algebraic variables and 250 equations.

P. Larsson and J. Åkesson are with the Department of Automatic Control, Lund University, Lund, Sweden, {perola|jakesson}@control.lth.se.

S. Haugwitz is with Borealis AB, Stenungsund, Sweden, Staffan.Haugwitz@borealisgroup.com.

N. Andersson is with the Department of Chmical Engineering, Lund University, Lund, Sweden. niklas.andersson@chemeng.lth.se.

Modelica, a high level language for encoding of complex physical systems, is used for plant modeling. The Optimica extension, see [1], gives constructs for cost functions, constraints and mechanisms to select inputs and parameters to optimize. Using JModelica.org, an open source project targeted at dynamic optimization, see [2], the optimization problem is translated into a non-linear programming problem using collocation on finite elements and solved using the large-scale NLP solver IPOPT [3].

III. OPTIMAL GRADE TRANSITION

The grade transition example will change conditions in all three reactors and corresponds to two grades currently produced at Borealis AB. The main objectives are to change raw material concentrations and concentration ratios, split factor, and production rates. At transition start and end time, i.e., t_1 and t_2, the plant fulfills the static non-linear equations

$$
\begin{aligned}
\mathbf{0} &= F(\mathbf{0}, \mathbf{x}^{\circ}, \mathbf{w}^{\circ}, \mathbf{u}^{\circ}) \\
\mathbf{y}^{\circ} &= g(\mathbf{x}^{\circ}, \mathbf{w}^{\circ}, \mathbf{u}^{\circ}),
\end{aligned} \tag{2}
$$

which corresponds to Eq. (1) when all derivatives equal 0 and superscript $^{\circ}$ indicate constant value. Initial and end conditions of the transition for states, inflows and algebraic variables are given by solving the non-linear equations in Eq. (2), i.e., a DAE initialization problem is posed and contains approximately 280 equality constraints and 290 variables, of which 180 are algebraic and 230 have both upper and lower limits. Solving the NLP takes less than 10 seconds per grade.

A quadratic cost function that includes deviations from the grade to be are used, giving the possibility to emphasize the importance of different variables. Also the deviation from inflows yielding the new grade in stationarity will be used, removing too large over- and undershoots. Introducing the deviation vectors

$$
\Delta \mathbf{u} = \mathbf{u} - \mathbf{u}_2 \qquad \Delta \mathbf{y} = \mathbf{y} - \mathbf{y}_2,
$$

where $\mathbf{u}_2$ and $\mathbf{y}_2$ are inputs and outputs defining the new grade solved for in the DAE initialization problem, the dynamic grade transition optimization problem can be formulated as

$$
\min_{\mathbf{u}} \int_{t_1}^{t_2} \begin{bmatrix} \Delta\mathbf{y} \\ \Delta\mathbf{u} \\ \dot{\mathbf{u}} \end{bmatrix}^T \begin{bmatrix} \mathbf{Q}_{\Delta\mathbf{y}} & 0 & 0 \\ 0 & \mathbf{Q}_{\Delta\mathbf{u}} & 0 \\ 0 & 0 & \mathbf{Q}_{\dot{\mathbf{u}}} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{y} \\ \Delta\mathbf{u} \\ \dot{\mathbf{u}} \end{bmatrix} dt \tag{3}
$$

$$
\text{subj. to} \quad \mathbf{0} = F(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{w}, \mathbf{u}), \quad \mathbf{y} = g(\mathbf{x}, \mathbf{w}, \mathbf{u})
$$

$$
\mathbf{y}_{\min} \le \mathbf{y} \le \mathbf{y}_{\max}, \qquad \mathbf{u}_{\min} \le \mathbf{u} \le \mathbf{u}_{\max}
$$

$$
\mathbf{w}_{\min} \le \mathbf{w} \le \mathbf{w}_{\max}, \qquad \dot{\mathbf{u}}_{\min} \le \dot{\mathbf{u}} \le \dot{\mathbf{u}}_{\max},
$$
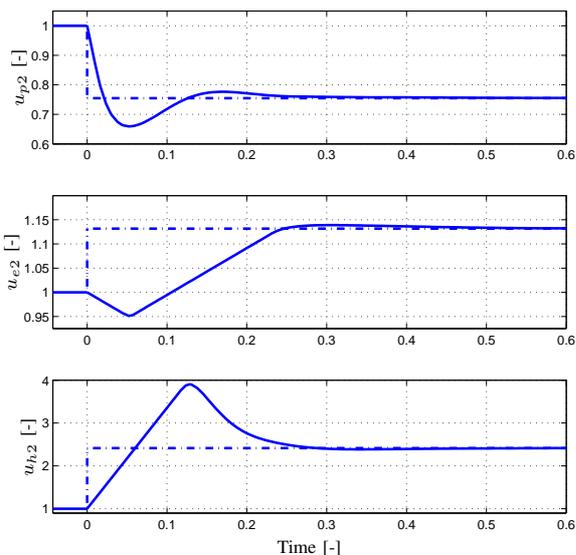
Fig. 2. Scaled optimal inflows to loop reactor at grade transition – propane u_{p2}, ethylene u_{e2} and hydrogen u_{h2}.

where also a cost of inflow derivatives is added such that smoothness of inflows can be controlled. The weights $\mathbf{Q}_{\Delta\mathbf{y}}$, $\mathbf{Q}_{\Delta\mathbf{u}}$ and $\mathbf{Q}_{\dot{\mathbf{u}}}$ are chosen diagonal for simplicity and the initial state of the plant is defined by the solution of the DAE initialization problem.

Over- and undershoots are accepted up to a certain limit for the instantaneous concentrations and ratios. However, for the bed average concentrations and ratios and the split S, no over- or undershoots are accepted in the grade change. The constraints on the algebraic variables $\mathbf{w}$ are for instance limits on volumes, component masses, and pressure, while constraints on inflows, both magnitudes and rates of changes, concern physical limits such as e.g., pump capacities.

After discretization, the NLP problem contains about 20.000–200.000 variables depending on number of elements and collocation points. Initial trajectories can be generated in JModelica.org via simulation using SUNDIALS, see [4], with inflows ramping from initial to end values found in the DAE intialization problem. With an Intel® Core™2 Duo CPU@3.00GHz, a solution is obtained in 5-90 minutes depending on number of variables and initial values.

Figures 2–3 show the resulting optimal inflows, component concentrations, and production rate of the loop reactor and the split factor between the loop reactor and GPR. Note the scaling, i.e., the transition is 1 time unit and all variables have initial value 1.

Since the production rate Q_2 is to be increased, the inflow of ethylene is increased in total and at the same time inflow of the diluent propane is decreased as shown in Figure 2. This results in a longer hold up time of the polymer and thus also a larger mass of polymer in the loop. The concentrations of ethylene and hydrogen in the loop are higher in the new grade and the decrease of diluent is not enough for the hydrogen specification to be met. Thus, the inflow of hydrogen is increased and to reach the specification of the hydrogen-ethylene ratio rapidly, the inflow of ethylene is initially
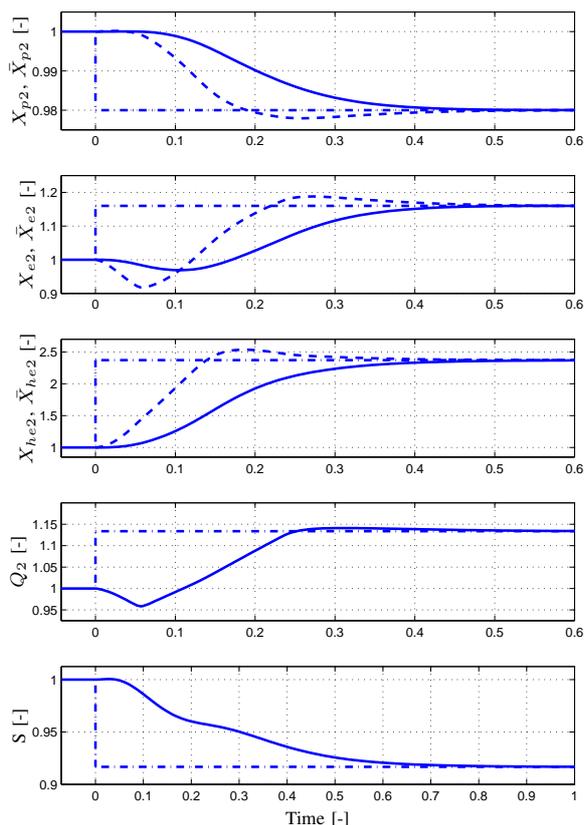

Fig. 3. Scaled key parameters for loop reactor and split factor at grade transition. Bed averaged (solid) and instantaneous (dashed) ethylene conc. $\bar{X}_{e2}$, X_{e2}, propane conc. $\bar{X}_{p2}$, X_{p2}, hydrogen-ethylene conc. ratios $\bar{X}_{he2}$, X_{he2}, production rate Q_2, split factor S.

decrased. Note that both the inflow of ethylene and hydrogen have their derivative constraints active in the beginning, seen by the linear decrease and increase. From Figure 3 it is seen that the over- or undershoot constraints on the averaged concentrations and ratios are obeyed and the instantaneous measures have over- or undershoots. The split, see Figure 3, which indirectly depends on the production rates in both loop and gas phase reactor, is decreased by lowering production rate in the GPR, i.e., decreasing the ethylene inflow to the gas phase reactor. The transition in loop reactor is completed after 0.5 time units. Similar trajectories for key parameters and inflows are available for the pre-polymerization and gas phase reactor.

REFERENCES

[1] J. Åkesson, "Optimica—An Extension of Modelica Supporting Dynamic Optimization," in *In 6th International Modelica Conference 2008.* Modelica Association, Mar. 2008.

[2] J. Åkesson, K.-E. Årzén, M. Gäfvert, T. Bergdahl, and H. Tummescheit, "Modeling and optimization with optimica and jmodelica.org-languages and tools for solving large-scale dynamic optimization problem," *Computers and Chemical Engineering*, Jan. 2010, doi:10.1016/j.compchemeng.2009.11.011.

[3] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming." *Mathematical Programming*, vol. 106, no. 1, pp. 25–58, 2006.

[4] C. f. A. S. C. Lawrence Livermore National Laboratory, "SUNDIALS (SUite of Nonlinear and DIfferential/ALgebraic equation Solvers)," 2009, https: //computation.llnl.gov/casc/sundials/main.html.

Magnus Glosli Jacobsen:
Challenges in optimization of operation of LNG plants

Abstract for the Nordic Process Control Workshop 2010, Lund, Sweden

Keywords: LNG, optimal operation, self-optimizing control, simulation

As pointed out by Jensen [1], most of the open research on LNG plants focuses on process design. A typical study will seek to maximize profit over the life span of a plant, by chosing the best configuration. This includes how many pressure levels the process should have, whether to use single component or multi component refrigerants, what kind of drives to use for the compressors and so on. Often such optimizations take the form of mixed-integer nonlinear problems (MINLP).

The final result depends on factors like expected gas prices, distance between gas field and liquefaction facility, and climate at the plant location. For example, if temperatures vary significantly between summer and winter, one must have a plant design which is robust to changes in ambient temperature.

Once the process design is done, however, one should obviously seek to run the process as close to optimal as possible. This means that disturbances need to be handled – for example by using MPC controllers or by controlling variables whose optimal values are not very sensitive to disturbances. The latter approach is called *self-optimizing control* [2] – i.e. one seeks to minimize loss caused by disturbances, by choosing the best variables to control at constant setpoints.

In order to do the latter, one needs to perform off-line optimization of the plant with given plant data. One will inevitably have fewer degrees of freedom for optimization in this case, since equipment size, driver configurations, coolant compositions and so on are fixed. Typically one can vary pressures and flowrates in certain streams. One must optimize the process for nominal conditions and for different disturbance scenarios.

The problem does usually not contain any binary/integer decision variables, all variables are continuous. This means we are left with a constrained non-linear problem, which may be solved using a sequential quadratic programming method. However, the problem is not always easy. Since we operate with small temperature differences (especially in the cold part of the plant – often as low as 1°C) we encounter problems in the models used to calculate the objective function and constraints.

Since each calculation of constraints and objectives requires convergence of the steady-state process model, it is critical that the model is robust enough to handle the steps taken by the optimization method. If, for example, the independent variables are temperatures in both ends of a heat exchanger, it is easy to specify a step in those variables which is physically infeasible. When you combine flows that go in closed loops with small margins to constraints, you are bound to have a difficult optimization problem. Examples of difficulties are:

- A small change in one flow might result in temperature crossover unless the model is good enough
- A small change in one temperature might lead to another temperature becoming infeasible – for example, the location of the smallest temperature difference might move to the other end of the heat exchanger, giving a very different solution
- The active set (of constraints) changes, especially if disturbances occur

It is necessary to know the process well in order to handle, or possibly avoid, these problems.

The example process that the work is focused on is the Air Products C3-MR process [3], which is the most widely used process for liquefaction of natural gas to date. The main approach to optimization has been to model the process in Honeywell's Unisim simulation software [4], and using Matlab's Optimization Toolbox [5] to carry out optimization. The two are linked using the **actxserver** function, which makes Unisim a COM server for Matlab and allows Matlab to specify variables in the Unisim model.

In this work we have investigated possibilities for simplification of the optimization problem – this includes changing the set of specifications, reformulating the models and finally splitting up the flowsheets and optimizing each part with respect to appropriately changed objective functions. We have also sought to identify the parts of the process which are most likely to produce problems for optimization.

References:

[1] Jensen, J.B., Skogestad. S: "Optimal operation of a simple LNG process", International Symposium on Advanced Control of Chemical Processes, Gramado, Brazil, 2006.

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

[3] Newton, C. L.; Kinard, G. E.; Liu, Y. N.: "C3-MR Processes for baseload liquefied natural gas". Liquefied Natural Gas VIII Volume 1, Sessions I & II, June 15-19 1986, Los Angeles, California.

[4] http://hpsweb.honeywell.com/Cultures/en-US/Products/ControlApplications/simulation/UniSimDesign/default.htm

[5] http://www.mathworks.com/products/optimization/?BB=1

Production Optimization for Two-Phase Flow in an Oil Reservoir

Carsten Völcker, John Bagterp Jørgensen, Per Grove Thomsen

Department of Informatics and Mathematical Modeling
Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Erling H. Stenby

Department of Chemical and Biochemical Engineering
Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Petroleum reservoirs are subsurface formations of porous rocks with hydrocarbons (oil and/or gas) trapped in the pores. Initially a reservoir may be under sufficient pressure to push the fluids to the surface. However, as the fluids are produced the pressure declines and production reduces over time. When natural drive becomes insufficient, then the pressure can be maintained artificially by injection of water. Conventional technologies for recovery leaves more than 50 % of the oil in the reservoir. Wells with adjustable downhole flow control devices coupled with modern control technology offer the potential to increase the oil recovery significantly.

The objective is to maximize production by manipulating the well rates and bottom hole pressures of injection and production wells. Optimal control settings of injection and production wells are computed by solution of a large scale constrained optimal control problem. We present a two-phase immiscible flow model and describe a gradient based method to compute the optimal control strategy. An explicit singly diagonally implicit Runge-Kutta (ESDIRK) method with adaptive stepsize control is used for computationally effecient solution of the model. The gradients are computed by the adjoint method. The adjoint equations associated with the ESDIRK method are solved by integrating backwards in time. The necessary information for the adjoint computation is calculated and stored during the forward solution of the model. The backward adjoint computation then only requires the assembly of this information to compute the gradients.

We demonstrate the optimal control strategy on a simple waterflooding example using one injector and one producer, which are divided into several individually controllable inflow valves.

Acknowledgement

Convex optimization for the crystal shape manipulation

Naim Bajcinca[1], Ricardo Perl[1], Jörg Raisch[1,2], Christian Borchert[1], and Kai Sundmacher[1,3]

[1]Max-Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr.1, 39106 Magdeburg, Germany
[2]Technische Universität Berlin, Einsteinufer 17, 10587 Berlin, Germany
[3]Otto-von-Guericke Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany

Crystal shape manipulation is a critical engineering venture in numerous industries. It has been shown that many properties of dispersed phase products are strongly linked to their shape, see [6]. For instance, dissolution rate or catalytic activity may depend on the ratio of the face area of a crystal particle. From the engineering point of view, manipulation of the crystal morphology is therefore essential, see [3]. In contrast to the traditional techniques that utilize chemical additives for blocking or promoting of certain crystal faces despite the undesired chemical impurities, [5], here, shape manipulation by means of temperature control only is considered.

The growth (*i.e.* dissolution) dynamics of a single particle of dimension n is defined by a set of n ODEs corresponding to each ℓ_i-coordinate ($i = 1, 2, \cdots, n$), which are coupled by a feedback mass-balance equation. Diverse solution approaches to optimal control for crystal shape manipulation utilizing the minimum principle and efficient numerical algorithms have been recently proposed in [1]. Optimal switching trajectories with a number of subsequent growth and dissolution phases, see also [4], have been shown to be indispensable in order to achieve morphologies that do not result from a pure growth process, see Figure 1. This is due to the inherent constraints in the particle growth (or dissolution) vector, which is confined to lie within a cone in the model state-space. As a consequence, instantaneous switching between subsequent phases leads to optimal piecewise-continuous temperature profiles.
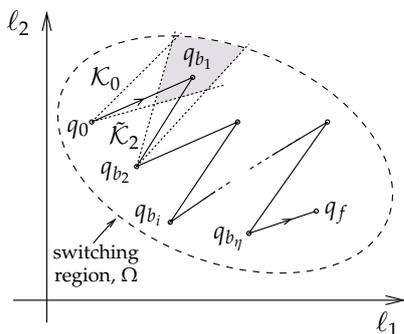

Figure 1: Optimal switching path

The main contribution in this paper is a convenient reformulation of the minimum-time problem as a convex optimization problem. While in [1] the *switching manifolds* in state-space are prespecified (*e.g.* referring to constant particle volume), in this paper, we introduce a *switching region*, where the switching points are required to lie within. Note that switching regions do not put any restriction to the number of trajectory sections, as opposed to the optimization scenario with constraints imposed by the switching manifolds, see [1].

In our new approach, we exploit two useful properties of the optimal solutions as proven in [1]. In particular, the minimum-time solutions are composed of constant supersaturation (*i.e.* $\sigma =$ const) sections. Hence, the optimal path in state-space consists of a number of connected straight line sections. Furthermore, we utilize the invertibility property of the particle growth (or dissolution) dynamics, which reduces the dynamic optimization problem into a static NLP problem. As a result, the geometric path of the particle motion in the state-space is to be optimized instead. The required optimal temperature profile is then uniquely determined, see [1]. In the sequel, we describe the main aspects of our approach.

Without loss of generality, we consider a two-dimensional particle. Let q_0 and q_f denote the initial and final particle shape in the state-space (ℓ_1, ℓ_2), see Figure 1. The points q_{b_i}, $i = 1, 2, \cdots, \eta$ in the figure denote the switching points, where jumps in the temperature (*i.e.* supersaturation) profile are exerted. In our optimization setup we assume that the number of switching points η, as well as, the sequence of the growth and dissolution phases is fixed. Thereby, sequences with two or more subsequent growth (or dissolution) cycles are legal, too. The optimization problem, see Figure 1, may be now stated as

$$\min_{q_b} \ t = \sum_{i=1}^{\eta+1} t_i(q_{b_{i-1}}, q_{b_i}), \ \ \text{s. t. } q_{b_i} \in \Omega \cap \mathcal{K}_{i-1} \cap \tilde{\mathcal{K}}_{i+1}, \ i = 1, 2, \cdots, \eta. \tag{1}$$

Here q_b is a vector variable that collects all switching points q_{b_i}, $i = 1, 2, \cdots, \eta$. The switching region Ω is per construction a convex set. The cone $\mathcal{K}_i$ includes the set of all reachable points by forward integration starting at $q_{b_{i-1}}$, while $\tilde{\mathcal{K}}_i$ includes the reachable points by backwards integration starting from q_{b_i}, see Figure 1. For convenience, we assume $q_{b_0} = q_0$ and $q_{b_{\eta+1}} = q_f$. Therefore, $\Omega \cap \mathcal{K}_{i-1} \cap \tilde{\mathcal{K}}_{i+1}$ is a convex set in the q_b-space. Note that, the constraints in (1) imply an upper and lower level for the supersaturation σ, that is, the temperature constraints as introduced in [1] are dropped here.

Since the condition $\sigma = \text{const}$ uniquely determines the slope of a straight line section, the net time required to move from q_0 up to q_f can be expressed in terms of the decision variables $q_{b_i} = (q_{b_{i,1}}, q_{b_{i,2}})$. It turns out that

$$t = \sum_{i=1}^{\eta+1} K_i \cdot \left(q_{b_i,1} - q_{b_{i-1},1} \right)^{\frac{g_{i2}}{g_{i2}-g_{i1}}} \cdot \left(q_{b_i,2} - q_{b_{i-1},2} \right)^{\frac{-g_{i1}}{g_{i2}-g_{i1}}} \tag{2}$$

The exponential parameters g_{i1} and g_{i2}, corresponding to the coordinates ℓ_1 and ℓ_2 of the crystal particle, as well as the K_i's in (2) are positive constants. Note that every growth (or, dissolution) phase shares the same values for K_i, g_{i1} and g_{i2}.

To examine the convexity of the objective function $t = t(q_b)$ in (2), it is instructive to consider first any summand term as they all obviously share the same form. Indeed, it can be shown that the Hessian of $t_i(q_{b_{i-1}}, q_{b_i})$ w.r.t. to the decision variables $q_{b_{i-1}}$ and q_{b_i} is a 4×4 symmetric matrix, with three eigenvalues being identically zero and a fourth one $\lambda_{4,i}$ of the form

$$\lambda_{4,i} = C_{gi} \frac{g_{i1} g_{i2}}{(g_{i2}-g_{i1})^2} \left(\left(q_{b_i,2} - q_{b_{i-1},2}\right)^{\frac{-g_{i1}}{g_{i2}-g_{i1}}} \left(q_{b_i,1} - q_{b_{i-1},1}\right)^{\frac{2g_{i1}-g_{i2}}{g_{i2}-g_{i1}}} + \left(q_{b_i,2} - q_{b_{i-1},2}\right)^{\frac{g_{i1}-2g_{i2}}{g_{i2}-g_{i1}}} \left(q_{b_i,1} - q_{b_{i-1},1}\right)^{\frac{g_{i2}}{g_{i2}-g_{i1}}} \right).$$

The parameters C_{gi} are again positive constants determined by the growth (i.e. dissolution) kinetics. It is easy to see, that $\lambda_{4,i} > 0$, i.e the function $t_i(q_{b_{i-1}}, q_{b_i})$ is convex for all $i = 1, 2, \cdots, \eta$. Notice that for $i = 1$ and $i = \eta$, the Hessian matrix is 2×2. Yet, the positive semi-definiteness of the Hessian is in both cases maintained. In other words, the objective function $t = t(q_b)$ in (2) is itself convex, and the optimal problem as defined in equation (1) is convex, too. This is an important outcome, since the optimal solution is global and a variety of efficient methods and tools for its solution are available, see [2].

The extension of the optimization algorithm for crystal particles of a dimension $n > 2$, as shown in [1], can always be recast as a two-dimensional problem. This is due to the fact that the supersaturation σ is uniquely determined by the projection of the motion into a two-dimensional plane. Note that the convexity and simplification of the introduced optimization algorithm, as compared to those proposed in [1], are achieved at the price of a double number of decision variables. Figure 2 illustrates the above algorithm for the case with one switching, i.e. $\eta = 1$. Notice that the shortest path is not necessarily the fastest one.
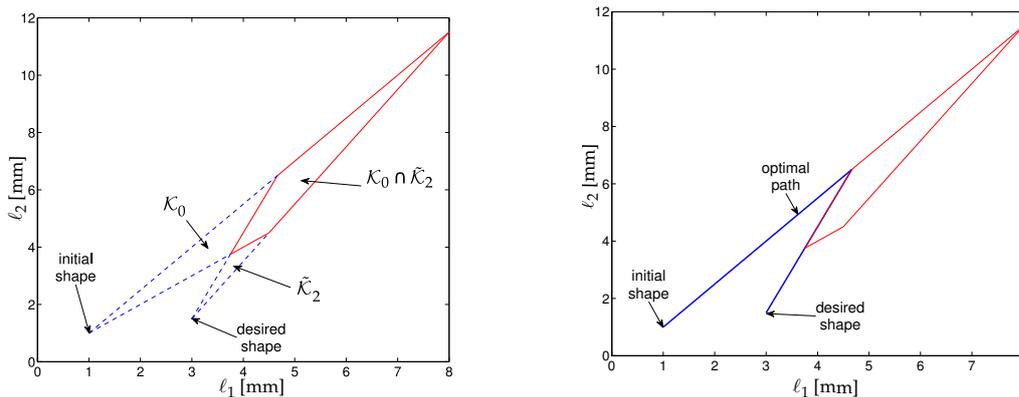

Figure 2: An illustration for the KDP-crystal. The problem data are taken from [1] and the references therein.

References

[1] N. Bajcinca, V. de Oliveira, C. Borchert, J. Raisch, and K. Sundmacher. Optimal control solutions for crystal shape manipulation. *20-th European Symposium on Computer Aided Process Engineering – ESCAPE20*, 2010.

[2] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[3] Daniel B. Patience and J. B. Rawlings. Particle-shape monitoring and control in crystallization processes. *AIChE Journal*, 47(9):2125–2130, 2001.

[4] R. C. Snyder, S. Studener, and M. F. Doherty. Manipulation of crystal shape by cycles of growth and dissolution. *AIChE Journal*, 53(6):1510–1517, June 2007.

[5] I. Weissbuch, R. Popovitzbiro, M. Lahav, and L. Leiserowitz. Understanding and control of nucleation, growth, habit, dissolution and structure of 2-dimensional and 3-dimensional crystals using tailor-made auxiliaries. *Acta Crystallographica Secton B-Structural Science*, 51(Part 2):115–148, APR 1 1995.

[6] Hua Gui Yang, Cheng Hua Sun, Shi Zhang Qiao, Jin Zou, Gang Liu, Sean Campbell Smith, Hui Ming Cheng, and Gao Qing Lu. Anatase TiO2 single crystals with a large percentage of reactive facets. *NATURE*, 453(7195):638–U4, MAY 29 2008.

Comparison of two main approaches for operating Kaibel distillation columns

Maryam Ghadrdan, Ivar J. Halvorsen*, Sigurd Skogestad,

NTNU, Deparment of Chemical Engineering
*) SINTEF ICT, Applied Cybernetics

The divided-wall distillation column (DWC) realizes the fully thermally coupled Petlyuk column, which is to separate the feed in a prefractionator-sidestream arrangement with a direct coupling of vapor and liquid streams between prefractionator and main column, into a single shell. This arrangement for separating the feed to 4 products is called Kaibel arrangement. This tight integration makes it challenging to design and control the column, compared to the conventional sequence of simple columns. The design challenges have been mostly solved, but operation and control remains largely an open issue.

The objective of this paper is to study the Kaibel distillation column from operability point of view. Also a qualitative study is done for finding proper self-optimizing control variables. Two different objectives, namely minimizing energy consumption at fixed product purities and maximizing product purities with a fixed boilup are considered. It is usually assumed that the objective is to make products of given purity using the minimum energy. However, in practical operation this is often not the issue, but rather to make the purest possible products with a given energy.

The idea behind self-optimizing control is to find a variable which characterize operation at the optimum, and the value of this variable at the optimum should be less sensitive to variations in disturbances than the optimal value of the remaining degrees of freedom. Thus if we close a feedback loop with this candidate variable controlled to a setpoint, we should expect that the operation will be kept closer to optimum when a disturbance occur. Self-optimizing control is when we can achieve an acceptable loss L with constant setpoint values c, for the controlled variables (Skogestad 2000).

The steady-state model used for this purpose is developed in UNISIM. The model in UNISIM is optimized and purturbed around the optimum from MATLAB. This is because of the more powerful optimization tool provided in MATLAB. The feed stream contains the first four simple alcohols (Methanol, Ethanol, 1-Propanol, 1-butanol) as the components. The model has six degrees of freedom: boilup rate (V), reflux (L), side stream flows ($S1$, $S2$), liquid split (Rl) and vapour split (Rv). Boilup rate is set as a constraint in the first approach.This constraint should be met as the operating conditions change. After finding the optimal nominal case, we have visualized the objective functions around the optimum. This can be used to get insight in column behavior and as a basis for a systematic control structure design. Another issue is that how the stage design affects this comparison. This is also considered in this study.

References

1. S. Skogestad, *J. Process Control*, 10 (2000) 487-507

PRODUCTION OF DISTRICT HEATING AT SÖDRA CELL MÖRRUM

Karin Axelsson*, Veronica Olesen**

*Department of Chemical Engineering, Lund University
**Solvina AB, Gruvgatan 37, Västra Frölunda, Sweden

1 INTRODUCTION

Södra Cell Mörrum, SCM pulp mill produces steam in the chemical recovery process. Some of the steam is used to make hot water for district heating. The hot water for district heating is transported to the local energy company Karlshamn Energi AB, KEAB.

A varying flow of cold water is sent back from KEAB to SCM. The water is heated up and sent back at a, by KEAB, predefined temperature. However, there have been oscillations in the process leading to variations in outgoing temperature to KEAB. In a project at Solvina the source of these oscillations was identified, a dynamic model of the process was made and measures of how to eliminate the oscillations were suggested.

2 PROCESS DESCIPTION

Surplus steam from the recovery process is fed through a condenser. The heat from the condensing steam is used for heating up water. A pair of control valves on the condensate pipe regulates the condensate flow from the condenser. Changing the positions of these valves, changes the condensate level and hence the area available for condensation and thus incoming steam flow to the condenser. A controller working with the steam flow as its set point regulates the condensate valves. The steam flow set point is given by the total steam demand in the process and is in this application an uncontrollable variable.

To compensate for unpredictable changes in steam supply a heat buffer, an accumulator, is used to keep the outgoing temperature at its set point. The accumulator is a hot water reserve, storing hot water in the top and cold water in the bottom. Flow can go in both directions through the accumulator, depending on whether it is being charged or discharged with hot water. See Figure 1.

As seen in the figure, there is also a possibility to bypass cold water, to keep the outgoing temperature down.
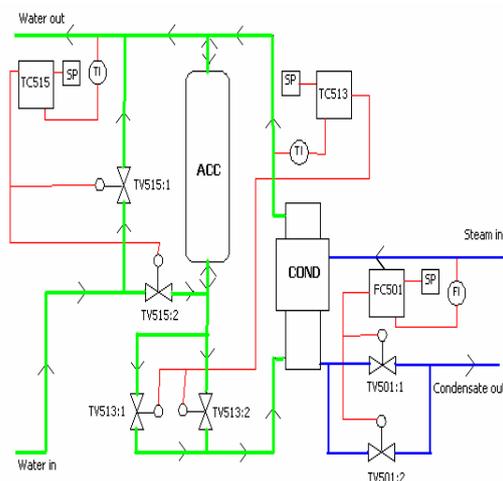

Figure 1: Sketch of the system, control loops are marked in red

3 THE MODEL

The simulation of the system was performed using Extend. This modelling tool provides a platform where all components can be modelled as blocks containing first-principle models. All blocks can be connected, forming a sequence to be solved in each time step.

The resulting model was validated using data from SCM, collected during half a year. During this period of time the need for district heating varied from very low in the summer to high in the winter. The resulting model was able to capture the large scale dynamics of the system. However, an ideal model of the system did not capture the oscillations.

4 PROBLEM IDENTIFICATION

The oscillations observed were found to originate in the condenser and were especially prominent in the following measured variables: incoming steam flow, outgoing water temperature from the condenser, condensate level and output from the steam flow controller. Closer analysis of the graphs showing steam flow and output from the steam flow controller revealed that they followed patterns typical for systems with valve friction. Hence it was concluded that the oscillations in all probability were caused by friction in the large condensate valve. See Figure 2 for details.
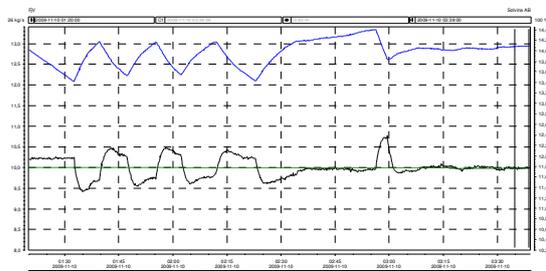
Figure 2. Data for the large condensate valve. Green line is the steam flow set point. Blue curve shows the control output. Black curve is steam flow measured value.
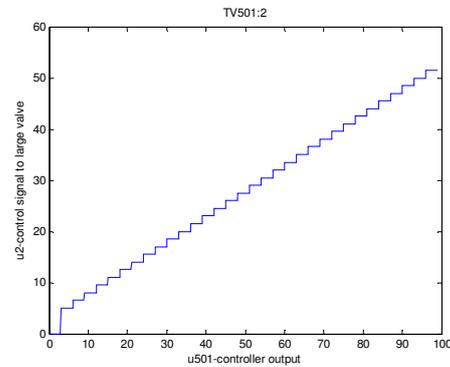
Figure 3: Control signal to large valve as a function of controller output

5   IMPROVED CONTROL STRATEGY

The steam flow to the condenser is regulated by two condensate valves. The output from the flow controller is converted into two control signals, one for each valve. Prior to this project, split range control was used to divide the control signal to the valve pair. The small valve was used for small demands on steam flow. At higher demands, the small valve was fully open and the large valve was used for control of the steam flow. This strategy relies on both valves being able to move continuously and hence fails when friction is present in the large valve.

A new strategy was proposed where both control valves work simultaneously in the entire range of controller outputs. The control signal to the large valve as a function of controller output was suggested as a discrete stepwise function. The small valve could then be used for fine tuning at each discrete level. This strategy has two apparent advantages; it does not rely on the large valve being able to move continuously and it allows the small valve to operate in its optimal working span, i.e. near 50 % opened.

6   RESULTS

The control strategy was implemented in the control system. The oscillations in outgoing temperature were reduced and the system showed significant improvement in its ability to follow the steam flow set point using the new strategy.

As a positive side effect of this project, all control loops could be better tuned as the cause of oscillations was had been found and adjusted for.

CONTROL OF AN HMR PRE-COMBUSTION GAS POWER CYCLE

Lei Zhao[a], Finn A. Michelsen[b], Bjarne Foss[a]

[a]*NTNU, Department of Engineering Cybernetics,NO-7491 Trondheim, Norway*
[b] *SINTEF ICT, Applied Cybernetics, NO-7491 Trondheim, Norway*

Abstract: CO_2 capture and storage is becoming an increasingly important part of any discussion on clean coal and natural gas based power production. Statoil has recently developed and patented a pre-combustion gas power cycle based on a hydrogen membrane reformer (HMR). This is a promising option for capturing CO_2 in natural gas based power generation plants.

The plant consists of a pre-reformer, an HMR reactor, a medium temperature and a low temperature conventional water gas shift (WGS) stage, gas and steam turbines, a heat recovery system, a CO_2 separation unit, and several heat exchangers, separation and mixing units. Steam methane reforming (SMR) is among the most common technologies for converting hydrocarbons (methane) to hydrogen. A mix of natural gas and steam is fed to one of the sides of the HMR and undergoes steam reforming. The produced gas is a hydrogen rich syngas. Compressed air drawn from the gas turbine compressor is supplied to the other side of the HMR reactor. Permeated hydrogen is combusted, consuming approximately all oxygen in the air stream. This gives "CO_2 free" heat for the endothermic SMR reactions. Syngas with high concentrations of H_2, CO_2 and CO is fed via several heat exchangers to the WGS stages converting CO to H_2. The outlet gas from the permeate side contains mainly H_2O and N_2 and is used to dilute the hydrogen fuel recovered in the CO_2 removal process. CO_2 removal may be performed by using a conventional absorption unit. This process has shown higher efficiency than other pre-combustion processes, and it has a potential for cost reduction compared to other pre-combustion processes.

For this type of reforming, high operability and robustness is required. This is partly achieved through an understanding of the system dynamics and robust control structure design. The paper identifies important dynamic features of the plant. Based on this analysis, the paper explores various options for conventional control strategies for this plant, and suggests a reasonable control strategy based on realistic disturbance scenarios.

Control of industrial chromatography steps

Jan Peter Axelsson, Karolinska Institute,
Department of Biosciences and Nutrition, 141 57 Stockholm, Sweden.

There is an increased interest to decrease process variation and to reduce production costs in the biopharmaceutical industry today. This trend is facilitated by the so-called PAT-initiative from FDA a few years ago. In this talk I will describe two different ideas of control of chromatography steps with industrial relevance. Both examples address the fact that the incoming material may vary in quality and how to adapt the operation of a chromatography step to these variations. In the first example the batch is divided into several parts and after processing of the first sub-batch product quality is measured and the information is used for adjustments of the operating conditions. Regression of historical data is used to obtain a simple linear control law. In the second example the focus is to extract the long-term trends of performance despite considerable variation in the incoming material. A simple control law is obtained. Data from commercial production illustrate the benefits.
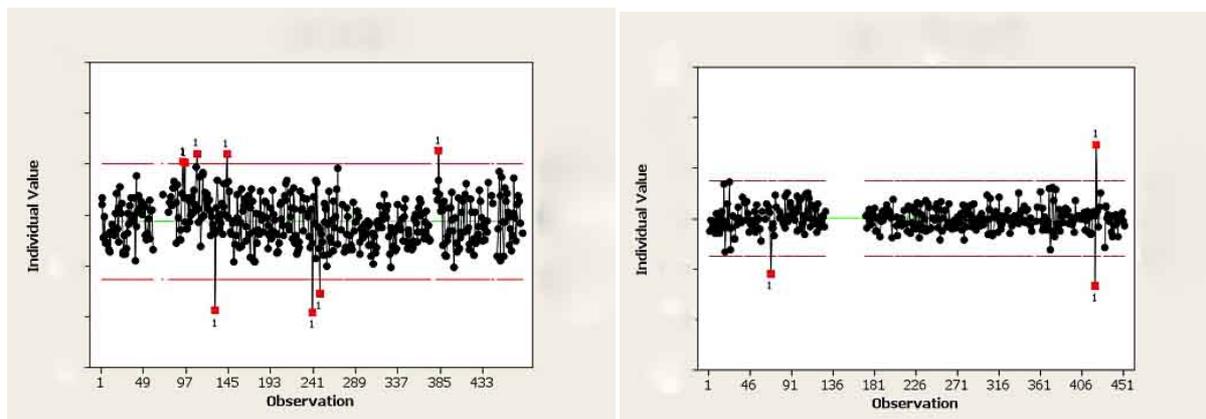
Figure 1 Production data from the first example illustrating the possible improvement. In the left diagram variation during manual control is shown. The right diagram shows the prediction error that set the limit of the control performance. During the gap in data (observation 135-180) the production was operated differently and the prediction model was not relevant. Scales are the same in the two figures.

Basic control of complex distillation columns

Deeptanshu Dwivedi, Ivar J. Halvorsen[1], Maryam Ghadrdan, Mohammad Shamsuzzoha and Sigurd Skogestad

Norwegian University of Science and Technology, Department of Chemical Engineering, Trondheim, Norway
1) SINTEF ICT, Applied Cybernetics, Trondheim Norway
Email: IvarJ.Halvorsen@sintef.no

Introduction

Dividing wall distillation columns (DWC) have received considerable attention in the last decades. Although the patent of DWC was submitted by Wright in 1946, and the basic theory that outlined potential energy savings by fully thermally coupled columns was presented by Petlyuk in 1965 the industry were reluctant. The breakthrough came with the work of Kaibel 1987 and several DWCs were realised within BASF through the last decade of 1900. Theoretical expressions for minimum energy for 3-component ideal zeotropic mixtures were presented by Fidkowski in 86. Extension to any number of components, sharp and non-sharp splits, and the general extended Petlyuk arrangement was presented by Halvorsen (2001). Several papers based on model realized in rigorous simulation tools have been presented in the last decade. Typical savings are reported in the range from 15% to 35% compared to conventional sequences.

However, not many papers have been published on operation. Wolff and Skogestad (1995) showed that the setting of liquid and vapour splits are very important. Triantafyllou and Smith investigated selection of manipulated variables both from simulations and by a pilot plant column (1992). Halvorsen explored the steady state properties of the Petlyuk arrangement. A recent paper by Lyuben (2009) discusses composition and minimum energy control.

The objective is to explore the behaviour of a complex column in real operation and propose how it should be controlled to obtain the potential energy savings in industrial practice.
The Approach in this paper is based on both simulations and on results from a 4-product Kaibel-type laboratory column at NTNU. Several issues are studied:
What happens when the feed composition is changed? What happens if the required liquid- and vapour splits as not properly set? What happens if the product draws are not properly set? How should a complex column best be started up and stabilized? How should we ensure real minimum energy operation in practice? And how does suboptimal operation affect the column profile and product purities, that is, how can we identify non-optimal operation? How do the control loops interact?

Many of these questions arose when we were operating the laboratory column, and to answer them we have mainly used dynamic simulation. Note that the dynamic model was adjusted to match the experimental column.

References

Wright, R.O., "Fractional Apparatus", US Patent 2471134, May 1946
Petlyuk, F.B. "Thermodynamically optimal method for separating multi-component mixtures", Int.Chem.Eng. Vol 5, No 3, pp 555-561, 1965.
Fidkowski, Z and Krolikowski, "Thermally Coupled system of distillation columns. Optimization procedure.", AIChE Journal Vol 32, No 4 , 1986
Kaibel, G. "Distillation columns with vertical partitions", Chem. Eng. Tech. 10 (1987) 92-98.
Wolff, E and Skogestad, S. "Operation of integrated 3-product (Petlyuk) distillation columns" Ind. Eng. Chem. Res. 1995, 34, 2094-2103.
Triantafyllou, C. and Smith, R. "The design and operation of fully thermally coupled distillation columns", Trans. IChemE, 7-(Part A), 118-132, 1992.
Halvorsen, I.J. "Minimum Energy Requirements in Complex Distillation Arrangements", PhD Thesis , NTNU 2001:43