I&EC research
Industrial & Engineering Chemistry Research
ARTICLE
pubs.acs.org/IECR

Active Constraint Regions for Optimal Operation of Chemical Processes

Magnus G. Jacobsen and Sigurd Skogestad*

Department of Chemical Engineering, Norwegian University of Science and Technology, NTNU, N-7491, Trondheim, Norway

ABSTRACT: When designing the control structure of a chemical plant, with optimal operation and control in mind, it is important to know how the active set of constraints changes with disturbances.The generation of optimal active constraints regions using optimization generally is very time-consuming, and this paper discusses how to use process knowledge to minimize the need for numerical calculations—that is, minimize the number of optimization problems that must be solved. We consider the case of two disturbances, as this can be nicely represented graphically. In this paper, we study economically optimal operation of a reactor−separator−recycle process and show how the set of active constraints changes with feed flow rate and energy cost as disturbances. We also identify the economical and physical bottleneck of the process. For this case study process, we find five distinct regions with different active sets. This illustrates that, even for simple cases, the map of active constraint regions can be quite complex.

INTRODUCTION

Motivation. In optimization and control of chemical processes, active constraints play an important role. The set of active constraints influences both plant economy and plant control. If one does not control the active constraints, one will have to accept economical loss.[1] Also, knowing which variables are optimally at their constraint value is crucial for control structure design. If a manipulated variable that is optimally at a constraint is used to control another variable, we cannot keep it at its optimal value at all times. Thus, it is necessary to know how the active constraints vary with disturbances. It is also useful to know which active constraints are important for plant economy and which are not. Say, for example, that a variable is optimally unconstrained in a particular region, and we decide to use this variable for control. Now, if this variable is constrained in a neighboring region, we may have to replace it in our control scheme. However, if backing off from the constraint gives a *small* loss, it might be acceptable to keep the same control structure.

Reactor−separator−recycle systems have been researched actively over the years and is a widely used example in control literature.[2−6] These articles mainly focus on basic regulatory control. Kiss et al.[7] studied state multiplicity, which may complicate optimization of such plants severely. However, little work has gone into identifying economically optimal operating conditions for such systems. Larsson et al.[8] focused on choosing *self-optimizing variables*[9] among the unconstrained variables. Maarleveld and Rijnsdorp[10] offered a study of constraint control on distillation columns and discussed how constraints change with different process variables. Except for ref 10 and the recent work by Jagtap et al.,[11] there is little emphasis on active constraints in the literature.

Contribution and Organization of the Paper. When one seeks to find the (approximate) regions in which each constraint is active, one will usually have to carry out at least a few optimizations. In multiparametric programming, one will typically carry out *many* optimizations, and if the optimization problem itself is difficult, this may be a time-consuming process. In this paper, we address how one can use knowledge about the optimization problem and process model to simplify this work, and we obtain approximate active constraint regions using few optimizations. Part of the purpose is to explain what we can determine by just using process knowledge, and what we must solve for numerically. We suggest a step-by-step method for sketching the active constraint regions for a two-dimensional disturbance space, and we use this method on a reactor−separator−recycle process. The process is to be seen as a "toy example", meaning that some numbers have been defined to illustrate certain points of the paper. Still, we think that the example illustrates how constraint regions may be encountered in a real example. The paper is structured as follows:

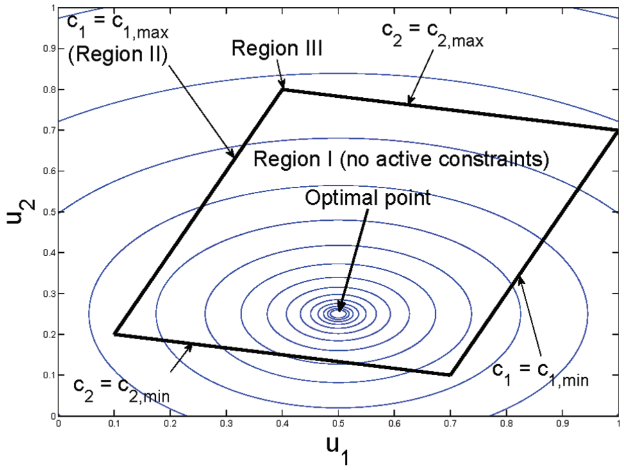
- First, we briefly discuss optimal operation in general and discuss the link between the input space and the disturbance space.
- Some general points about optimization are included to provide background for the rest of the paper.
- A method for finding approximate active constraint regions is outlined.
- We describe the process that we have used for the case study.
- We define optimal operation of the process and provide a degree-of-freedom analysis.
- The results of the case study are shown.
- The results and the efficiency of the method are discussed, and we compare our results to those from Jagtap et al.[11]

OPTIMAL OPERATION AND CONSTRAINT REGIONS

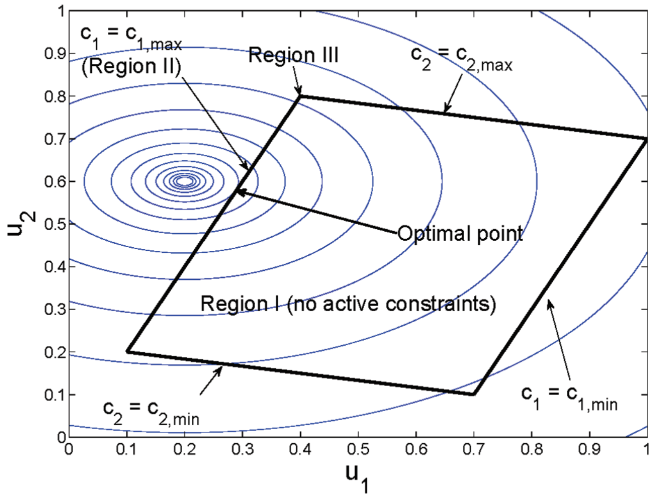
When optimizing the operation of chemical processes, we start by formulating a cost function J that must be minimized, and identifying the degrees of freedom (u) that can be adjusted to minimize J. Since we consider steady-state optimization, causality is not an issue; therefore, it does not matter which variables we

Received: June 7, 2011
Accepted: August 6, 2011
Revised: August 6, 2011
Published: August 07, 2011

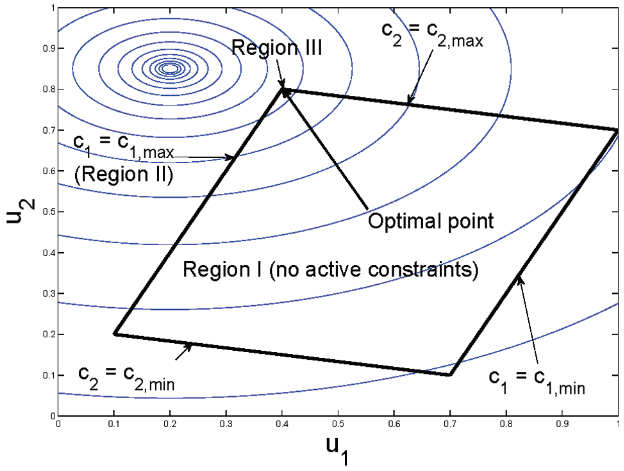
dx.doi.org/10.1021/ie2012196 | *Ind. Eng. Chem. Res.* 2011, 50, 11226–11236



(a) Optimum lies in unconstrained region (I)



(b) Optimum at $c_1 = c_{1,max}$ (region II)



(c) Optimum at $c_1 = c_{1,max}$, $c_2 = c_{2,max}$ (region III)

Figure 1. Constraint lines and objective function as functions of the degrees of freedom (u): (a) optimum lies in the unconstrained region (region I), (b) optimum at $c_1 = c_{1,\max}$ (region II), and (c) optimum at $c_1 = c_{1,\max}$, $c_2 = c_{2,\max}$ (region III).

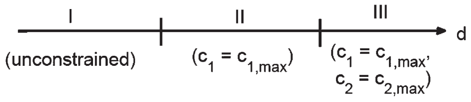


Figure 2. Constraint regions in one-dimensional disturbance space (one disturbance).

select as our degrees of freedom, as long as the remaining variables of the model form an independent set. For example, when optimizing a distillation column, the degrees of freedom could be flows (e.g., reflux L and vapor boilup V, so $u = [L\ V]$) or compositions ($u = [x_B\, x_D]$). Finally, we must identify the constraints (c) on operation (for example $L,V > 0$, $V \leq V_{\max}$), and the most important disturbances (d).

The optimal solution will often be found at the intersection of constraints. In Figure 1, we illustrate how constraints and objectives may change when disturbances change, for a simple example with constraints

$$\begin{aligned} c_1 &\geq c_{1,\min} \\ c_1 &\leq c_{1,\max} \\ c_2 &\geq c_{2,\min} \\ c_2 &\leq c_{1,\max} \end{aligned} \tag{1}$$

and two degrees of freedom (u_1 and u_2). The constraints define a two-dimensional region in the input space, and the optimal solution may lay in the interior of this region (Figure 1a), on one of its edges (Figure 1b), or at one of its corners (Figure 1c). Operation outside this region is infeasible. In this example, the feasible set of u is shown as being unchanged as d changes. This will typically be the case if the disturbance that changes is a price. If it is a process disturbance, the constraint lines also are likely to change.

In this paper, we focus on how the set of active constraints depends on the *disturbances*. The different points indicated in Figures 1a−1c would belong in different *constraint regions* in the disturbance space. In the case of one disturbance, each region corresponds to a line segment, as shown in Figure 2. It is worth noticing that "region III" in Figures 1a−1c refers to just one point, whereas in the disturbance space, "region III" includes all disturbances d, which makes that particular point optimal.

In the case of *two* disturbances, each region is a subset of the now two-dimensional disturbance space. Later in this paper, we use a two-dimensional example to illustrate the suggested method for identifying the active constraint regions. Thus, Figure 4a (presented later in this paper) is a two-dimensional equivalent to Figure 2.

■ OPTIMIZATION THEORY

General Form of the Optimization Problem. Optimization of chemical processes is typically a nonlinear problem of the form

$$\min_{u}\ J(x, u, d) \tag{2}$$

subject to

$$\begin{aligned} f(x, u, d) &= 0 \\ c(x, u, d) &\leq 0 \end{aligned}$$

where J is the economical objective, $f(x,u,d)$ the process model equations, and $c(x,u,d)$ the process constraints. The process model equations may be included in the optimization problem, or solved separately. The latter gives rise to different (and more complex) shapes of J and c, but eliminates x from the equations.

The choice of which approach to use depends on which is more robust, i.e., which approach is most certain to return a result. Especially in cases where recycles are involved, it may be beneficial (or even necessary) to let the optimization solver also solve the flowsheet model (see ref 12 for a discussion of this topic).

A solution of such a nonlinear optimization problem is characterized by the Karusch−Kuhn−Tucker conditions.[13] These conditions are described as follows (with X including both x and u, but not the disturbance d):

$$\begin{aligned} &\nabla_X \mathcal{L}(X^*, \lambda^*) = 0 \\ &c_i(X^*) = 0 \quad \text{for } i \in \mathcal{E} \\ &c_i(X^*) \leq 0 \quad \text{for } i \in \mathcal{I} \\ &\lambda_i^* \geq 0 \quad \text{for } i \in \mathcal{I} \\ &\lambda_i^* c_i(X^*) = 0 \quad \text{for } i \in \mathcal{E} \cup \mathcal{I} \end{aligned} \tag{3}$$

The optimal solution (X^*,λ^*) is parametrized by the disturbance d. We seek to find the disturbance value for which a constraint switches from active to inactive—let us call this value d_{active}. Here, we take advantage of the fact that for any constraint c_i, either the constraint value itself or its corresponding Lagrange multiplier λ_i (or both) is zero, and that the sum of the two must be monotonous at least in a small area around d_{active}. At d_{active}, we have that $c_i + \lambda_i = 0$. Thus, finding d_{active} is equivalent to solving the equation

$$s_i(d) = c_i(d) + \lambda_i(d) = 0 \tag{4}$$

for d. Since the optimal solution $X^*(d)$ is the result of an iterative process, one cannot use an analytic method to solve for $s_i = 0$; thus, an interpolation method is the simplest option. When using MATLAB, this will typically be the **fzero.m** solver, which has been used in this work. When using this solver, one must give two initial points between which the solver should search for the solution. The sign of $c_i + \lambda_i$ must be different at the two initial points in order for the solver to work.

Significance of Lagrange Multipliers for Active Constraints. Knowing the active constraint regions allows us to design a control structure based on the expected disturbances. However, although we may find an optimal control structure for each region, we may be interested in simplifying it, for example, by using the same control structure in several regions, even if it is not optimal. If using a control structure different from the optimal one, we may end up having to back off from a constraint that is optimally active. We may also have to back off from active constraints because of dynamic reasons (for example, controller overshoot) (see Figure 3). This gives rise to a loss, and in a small region, the magnitude of the loss ($|\Delta J|$) is related to the Lagrange multiplier as follows:[13]

$$|\Delta J| = \lambda_i |\Delta c_i| \tag{5}$$

where $|\Delta c_i|$ is the distance from the active constraint c_i (corresponding to "back off" in Figure 3) and λ_i is the corresponding Lagrange multiplier. What this means is that if we back off from the active constraint by a small margin Δc_i, we will have a loss that is locally proportional to the corresponding Lagrange multiplier. Thus, the Lagrange multiplier tells us how hard we get punished by backing off from a constraint. Obviously, we get punished harder if we back off from a constraint when we are far from a region where it is inactive, than if we back off from the same constraint at a point where it becomes active. In economics, $|\Delta J|$ is called a *shadow price*;[14] in optimal control theory, we have the related concept of *costate equations*.[15]

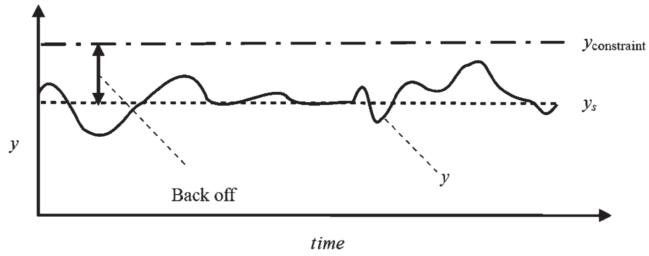


Figure 3. "Back off" from constraint due to imperfect control. (Illustration taken from ref 1. Copyright 2009, NTNU.)

■ METHOD FOR FINDING ACTIVE CONSTRAINT REGIONS AS A FUNCTION OF DISTURBANCE

In this section, we consider the two-dimensional case (with two disturbances) unless otherwise stated. We start this section by defining the terms *constraint curve* and *region*:

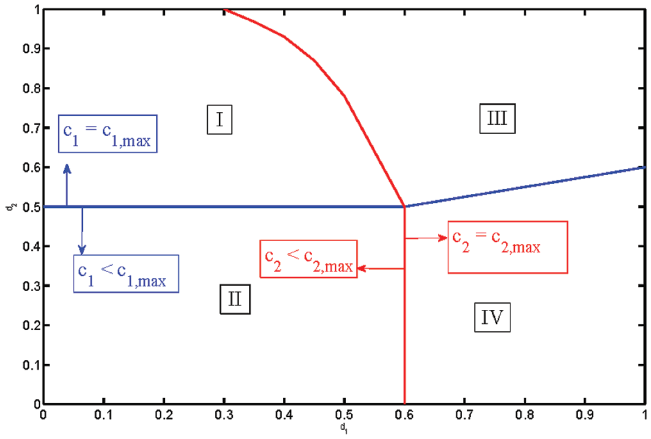
Definition 1: The *constraint curve* corresponding to a constraint c is the line separating the regions where c is optimally active from the regions where c is optimally inactive. [In some cases, a constraint curve will be straight and can be referred to as a constraint line. In an N-dimensional disturbance space, we will instead have $(N - 1)$-dimensional constraint surfaces.]

Definition 2: In this paper, a *region* refers to a part of the disturbance space, bounded by constraint curves, and is described by which constraints are active within that region.

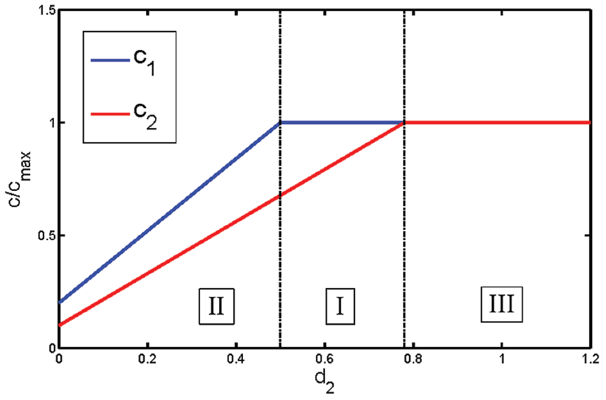
With these definitions in mind, we can make some general points about how the constraint curves and regions will behave:

- Constraint curves may cross each other, so one constraint curve may span the border between more than two regions. Where two constraint curves cross each other, four regions will meet in a point.
- When one constraint curve crosses another, it will generally change slope. This is because the nature of the optimization problem will change when the set of active constraints changes.
- When there are N constraints, which all may be either active or inactive, there may be as many as 2^N *active constraint sets*. In simple cases, where constraint curves do not cross each other more than once, this means we also have a maximum of 2^N regions (as illustrated by Figure 4a, where we have two constraints and four regions). In more-complex cases, where two constraint curves may cross each other more than once, we may have more than 2^N regions. In that case, some regions will share the same active set. However, the number of regions is usually smaller than 2^N, since some constraint combinations may not occur (for example, because they are physically impossible). For example, we often have both maximum and minimum constraints on the same variable—these will obviously never be active at the same time.
- In the one-dimensional case (i.e., one disturbance), each region corresponds to a line segment on the disturbance axis (x-axis). This is illustrated in Figure 4b. Here, the y-axis is used to plot the optimal value of the constraint functions as functions of a single disturbance.

We now want to outline a method for identifying active constraint regions without having to optimize at a large number of points across the entire disturbance space. As an illustrative example, let us consider a hypothetical problem



(a) Two-dimensional case (two disturbances)



(b) One-dimensional (one single disturbance)

Figure 4. Example figure illustrating how constraint curves divide the disturbance space into regions, and how the regions translate to line segments in the one-dimensional (1D) case.

with two constraints:

$$\begin{bmatrix} c_1 & c_2 \end{bmatrix} < \begin{bmatrix} c_{1,\max} & c_{2,\max} \end{bmatrix} \tag{6}$$

We have two disturbances, d_1 and d_2. Both c_1 and c_2 are continuous functions of the disturbances, and we have four regions:

(1) Both constraints are inactive
(2) Only c_1 is active
(3) Only c_2 is active
(4) Both constraints are active

The regions are shown in Figure 4a. Constraint c_1 is active ($c_1 = c_{1,\max}$) above the blue constraint curve in the figure, and c_2 is active ($c_2 = c_{2,\max}$) to the right of the red constraint curve in the figure. We also show the one-dimensional case where $d_1 = 0.5$ and d_2 is on the x-axis (see Figure 4b). Using this example as an illustration, we now outline a method for finding active constraint regions. First, it is reasonable to make the following assumptions:

(1) In a two-dimensional plot with (d_1,d_2) along the axes, two neighboring regions will only differ by one active constraint, except when two constraint curves cross each other (in Figure 4a, this happens at the point (0.6,0.5)). For the opposite to be true, two constraint curves would have to follow each other exactly (at least along a segment).
(2) We assume that the same set of active constraints does not appear in two separate regions of the disturbance space (thus, two constraint curves will not cross each other twice).
(3) We assume that there will always be a maximum feed rate ("bottleneck") above which we cannot satisfy all constraints.

The generality of the assumptions is discussed later in the paper.

Based on the above definitions, considerations and assumptions, we suggest using the following procedure to find active constraint regions:

(1) Using knowledge about the process model and the optimization problem, determine if any constraints will be active (or inactive) for *all* values of the disturbances, thus reducing the number of possible regions by a factor of 2 for each constraint that is always active.
(2) Also use the same insight to predict whether some region borders (part of constraint curves) will be independent of one of the disturbances. In a two-dimensional (2D) graph, these borders will correspond to vertical or horizontal lines. In the example (Figure 4a), the constraint curve separating regions I and III from regions II and IV is horizontal (independent of d_1). The line segment that separates regions II and IV is vertical. (Notice that this segment is only a part of a constraint curve!)
(3) Locate the region borders that are found to be vertical or horizontal, by solving for the disturbance value at which the corresponding constraint changes between inactive and active. In Figure 4a, this corresponds to finding the value of d_2 for which c_1 becomes active.
(4) Similarly, determine the value of d_1 for which we go from region II to region IV (that is, for which c_2 becomes active, with d_2 lower than the value we found in the previous step), thus locating the vertical part of the constraint curve for c_2.
(5) Find at least one more point along the line separating regions I and III. This means we need to do find a value of (d_1,d_2) which makes constraint c_2 switch from active to inactive. By fixing d_2 at a higher value than the one we found in step 3, and solving for the value of d_1 which makes $s_1 = 0$, we find a point on this line.
(6) In the same manner, find at least one more point along the line separating regions III and IV (with d_1 fixed at a value *higher* than found in step 4).
(7) If we are confident that the two last region borders are straight lines, or are satisfied with it as an approximation, we do not need more points. However, the true borders will often be curved (as the border between regions I and III in Figure 4a), and if we want to describe them more accurately, we need to find more pairs (d_1,d_2).

When choosing starting points for each interpolation search, one may be taken at a spot where the active constraints are already known (for example, the nominal optimum). The other may be taken close to the feasibility limits, which can be determined by carrying out a few sample optimizations.

The interpolation itself has been carried out using the **fzero** solver in MATLAB, which uses a combination of bisection, secant, and inverse quadratic interpolation.

We have used this approach successfully in a case study that we will discuss in the following sections.

■ DESCRIPTION OF THE EXAMPLE PROCESS

In this paper, we consider the simple reactor—separator—recycle process shown in Figure 5. We stress that the main intention of

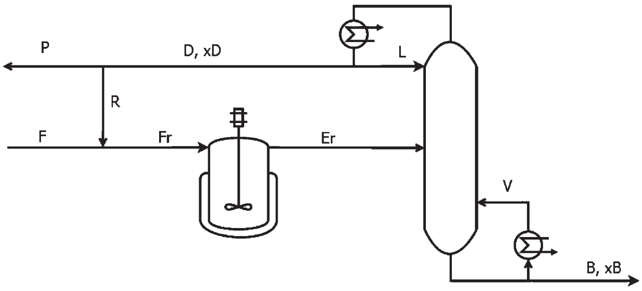


Figure 5. Flowsheet of the reactor−separator−recycle system.

the example is to illustrate the method, and the optimal process operation and constraint regions are likely to be quite different for other (and possibly more-realistic) examples.

Fresh feed of A (F) is mixed with recycled distillate from the column. The mixture (F_r) goes to the reactor, which is a continuously stirred tank reactor (CSTR) with two reactions (see eqs 7aa and 7ab). The first yields B, which is the desired product, and the second yields C, which is an undesired byproduct. A real-life example of such a process could be, for example, a hydrocarbon isomerization process, where cracking occurs as a side reaction (this is quite common in hydrocarbon refining processes).

$$A \rightarrow B \tag{7a}$$

$$A \rightarrow 2C \tag{7b}$$

In general, the steady-state mass balance of a CSTR with one feed stream and one product stream can be written as follows:

$$0 = \frac{dn_r}{dt} = F_r x_{F,r} + M_r \nu r - E_r x_r [\text{mol/s}] \tag{8}$$

where n_r is the reactor holdup vector (given in units of mol), F_r the flow rate entering the reactor (expressed in units of mol/s), x_{FR} the reactor feed composition, E_r the reactor exit flow rate (also expressed in units of mol/s), ν a matrix of stoichiometric coefficients, r a vector of reaction rates ($r = [r_1\ r_2]^T [s^{-1}]$), and x_r the composition inside the reactor (expressed in units of mole fraction) (in a CSTR, the exit stream is assumed to have the same composition as the reactor contents). M_r is the total reactor holdup (given in units of mol).

We will, for later use, define the reactant conversion X for the reactant and the product yield Y_i for each product i:

$$X\ (\%) = \frac{\dot{n}_{\text{reactant,in}} - \dot{n}_{\text{reactant,out}}}{\dot{n}_{\text{reactant,in}}} \times 100 \tag{9}$$

$$Y_i\ (\%) = \frac{\dot{n}_{i,\text{out}} - \dot{n}_{i,\text{in}}}{\dot{n}_{\text{reactant,in}}} \times 100 \tag{10}$$

where $\dot{n}$ indicates molar flow rate of a given component.

The reaction model uses first-order kinetics; the reaction rate of reaction i is given by

$$r_i = k_i x_A \tag{11}$$

where x_a is the mole fraction of A in the reactor and

$$k_i = A_i \exp\left(-\frac{E_{a,i}}{\mathscr{R}T}\right) \tag{12}$$

Table 1. Reaction Kinetics Parameters

reaction	A (s^{-1})	E_a (J/mol)
7a	1×10^5	6×10^4
7b	5×10^6	8×10^4

Table 2. Distillation Column Parameters

parameter	value
α_{AC}	0.70
α_{BC}	0.60
number of stages	30
feed stage location	15
feed liquid fraction	1
V_{max}	30 mol/s

where $E_{a,i}$ is the activation energy of reaction i (in J/mol), T the temperature (given in Kelvin), and $\mathscr{R}$ the gas constant ($\mathscr{R}$ = 8.3145 J/(mol K). Kinetic data are shown in Table 1. In this example, reaction 7a has a lower rate constant ($A_1 < A_2$), but also a lower activation energy; thus, it is the favored reaction at lower temperatures.

The reactor product E_r is separated in a column. The distillate D is rich in byproduct (C) and unreacted A, whereas the bottom product ($\mathscr{B}$) is rich in the desired product B. For the distillation column, we have used a simple column model, using the following assumptions: constant relative volatilities, constant molar overflow, constant pressure over the entire column, equilibrium at every stage, and negligible vapor holdups. Francis' weir equation is used to calculate liquid flow rates. The column data are shown in Table 2. The column model is based on the "Column A" model described in the work of Skogestad and Morari.[16]

A fraction of the distillate is recycled (R); the remaining distillate leaves the system as a purge stream (P).

DEFINING OPTIMAL OPERATION OF THE REACTOR−SEPARATOR−RECYCLE PROCESS

Steady-State Operational Economy Objective. The objective J should cover all economical aspects that are influenced by the steady-state operation. In general, these include cost of raw materials, energy, and utilities (such as cooling water), and the value of products. The cost of a product stream may be positive (if the product is sold, or processed to form valuable products later) or negative (if it is waste which must be disposed of). We may generalize to write

$$J = \sum p_{F,i} F_{F,i} + \sum p_{U,j} F_{U,j} + \sum p_{P,k} F_{P,k} \tag{13}$$

where $F_{F,i}$, $F_{U,j}$, and $F_{P,k}$ are the flow rates of feeds, utility streams, and product streams, respectively. (All of these terms are expressed in terms of mol/s, except for energy usage, which is given in \$/kJ.) The terms $p_{F,i}$, $p_{U,j}$, and $p_{P,k}$ are the prices of the respective streams. In this example, we use the following objective function:

$$J = p_F F + p_V V - p_B B - p_P P \tag{14}$$

where F, V, B, and P refer to Figure 5, and the prices are given in Table 3. (The value given for p_V is listed as variable, it is used as a disturbance later in the process.)

Table 3. Prices Used in Optimization[a]

component	variable	value
feed	p_F	1 \$/mol
product	p_B	2 \$/mol
purge	p_P	0.5 \$/mol
energy	p_V	variable

[a] The value given for p_V is the nominal value.

Degrees of Freedom. For example by using the method outlined in ref 17, we find that the process shown in Figure 5 has got five steady-state degrees of freedom (not including the feed flow rate, which is considered a disturbance in this paper). We may also find this number by examining the model, determining that it has 103 variables (again excluding the feed flow rate) and 98 independent equations. When using the method obtained from ref 17, the five degrees of freedom may be related to

(1) recycle/purge split, P/D
(2) reactor holdup, M_r
(3) reactor temperature, T_r (since the CSTR has a cooling jacket as shown in Figure 5, we may adjust this)
(4) column reboiler duty, Q_R
(5) column condenser duty, Q_C

However, in simulation, we may specify *any* five variables as long as the resulting 98 × 98 system is not structurally or numerically singular. When initializing the process model, the following set of specifications was used:

- Reactor temperature, T_r
- Mole fraction of component B in product stream $\mathcal{B}$, $x_{\mathcal{B},\mathrm{B}}$
- Reactor holdup, M_r
- Column reflux, L
- Purge/distillate ratio, P/D

Constraints. When optimizing a chemical process plant, we will encounter the following types of constraints:

- There are always capacity constraints; maximum levels in tanks and liquid-phase reactors, maximum available amount of utilities such as steam and cooling water, and maximum feed and product rates.
- There will usually be requirements on product quality, often in terms of the maximum content of impurities.
- In addition, there are typically constraints on pressures (due to limits in piping strength) and temperatures (for example, to limit catalyst degradation).

In this work, we have included maximum limits on reactor temperature T_r, reactor holdup M_r, and column boilup V, as well as a minimum limit on the fraction of component B in stream

$\mathcal{B}$

. In addition, we require that all flow rates are ≥0, and that the purge flow rate is <5 mol/s. (See Table 4.)

Disturbances. In a chemical process, there are many possible disturbances. The most important ones are usually related to feed conditions (flow rate, composition, and pressure), because these often are dependent on the operation of an upstream process. In addition, the prices of feeds, products, and utilities are often changing on a daily basis.

We may also have changes in process parameters. For example, in a catalytic process, the catalyst activity may degrade, so that the value of A in the expression for the reaction rate constant k decreases. Deposits inside piping may increase the pressure drop and decrease heat transfer. Finally, downstream processes may demand changes in product flow rates and compositions.

Table 4. Constraint Values Used in Optimization

parameter	value
$x_{\mathcal{B},\mathrm{B},\min}$	0.90
$T_{r,\max}$	390 K
$M_{r,\max}$	11000 mol
$V_{\max}$	30 mol/s
$P_{\max}$	5 mol/s
$R_{\min}$	0 mol/s
$P_{\min}$	0 mol/s
$B_{\min}$	0 mol/s

In this work, we consider two disturbances:

- Feed flow rate, F (nominal value: 1.1 mol/s). This is the flow rate that is most likely to be given by another process unit.
- Energy cost, p_V in the column (nominal value: 0.01 \$/mol) (either of the four price parameters could be used; what is really important is the energy price, relative to the difference in value between feed and products).

Since the feed flow rate is used as a disturbance, we only have *five* degrees of freedom in the optimization problem, compared to six degrees of freedom for initialization.

Feed Rate as Disturbance or as a Degree of Freedom. Previously in this paper, we stated that the feed flow rate is a disturbance in operation. This is typically the case when we consider a part of a bigger plant, and we just must process the feed that we get. If the considered unit is a stand-alone unit, however, the feed rate will rather be a degree of freedom. An exception to this is if we are bound, by a contract, to process a given amount of feed anyway.

When the feed rate is a disturbance, the goal is to process this feed at a minimum cost, while satisfying the process constraints c. At some point, the feed cannot be increased anymore without breaking the given constraints: this is when the process reaches its true bottleneck.

When the feed rate is a degree of freedom, however, we also may have an *economic* bottleneck. At a given feed rate, we may not be able to produce any more of the valuable product; thus, it will not be optimal to increase the feed rate. Below this economical bottleneck, the active constraint regions will be the same. When the feed rate is considered a degree of freedom, one can see the maximum available feed rate $F_{\max}$ as a disturbance.

Later in this paper, we address this economic bottleneck.

■ RESULTS

Initialization. In order to have a feasible starting point for subsequent optimization, we specified six variables as described above, and used **fmincon** in the MATLAB software to solve the model equations. To do this, the model equations and specifications were included as equality constraints in the optimization problem (corresponding to $f(x,u,d)$ in eq 2), and a dummy objective function with a constant value was used. This approach is used with success in ref 18. An alternative approach could be to use an equation-solving method minimizing the sum of squares of the residuals of the equalities. Table 5 shows the values used for initialization, plus the resulting values of other chosen variables (reactor holdups ($n_{r,i}$), conversion (X_A), and yield (Y_i), as well as

Table 5. Initial Data for the Reactor−Separator−Recycle System, Used as a Starting Point in Optimization

variable	value
F	1.1 mol/s
$x_{\mathcal{B},B}$	0.901
T_r	355 K
M_r	11000 mol
L	25 mol/s
P/D	0.2
R	1.512 mol/s
$n_{r,A}$	5064 mol
$n_{r,B}$	4159 mol
$n_{r,C}$	1777 mol
X_A	67.33%
Y_1	61.82%
Y_1	5.51%
B	0.7649 mol/s
$x_{A,B}$	0.099
$x_{\mathcal{B},B}$	0.901
$x_{C,B}$	0.000
D	1.8901 mol/s
$x_{A,D}$	0.607
$x_{\mathcal{B},D}$	0.167
$x_{C,D}$	0.227

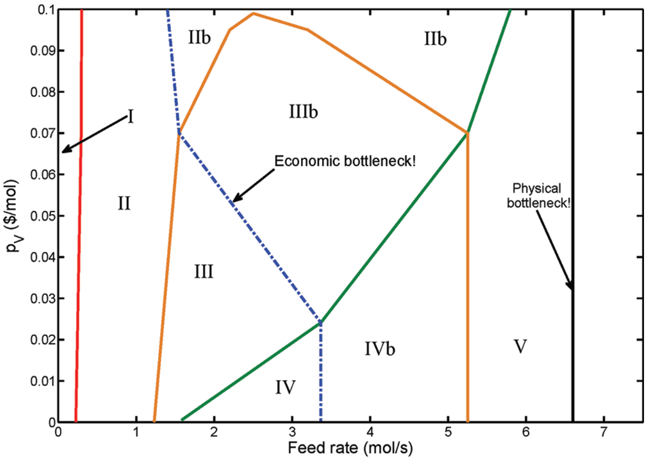


Figure 6. Active constraint regions for the reactor−separator−recycle system, for F values up to 6 mol/s and p_V values up to 0.10 $/mol.

Table 6. Constrained Variables in Each Region in Figure 6

region	constrained variables
I	$x_{\mathcal{B},B}$, M_r, R
II	$x_{\mathcal{B},B}$, M_r, R, T_r, (F)
IIb	$x_{\mathcal{B},B}$, M_r, R, T_r
III	$x_{\mathcal{B},B}$, M_r, T_r, (F)
IIIb	$x_{\mathcal{B},B}$, M_r, T_r
IV	$x_{\mathcal{B},B}$, M_r, T_r, V, (F)
IVb	$x_{\mathcal{B},B}$, M_r, T_r, V
V	$x_{\mathcal{B},B}$, M_r, T_r, V, R

product and distillate flow rates and compositions). Notice that the specification on $x_{\mathcal{B},B}$ was set slightly above the minimum value; this was to give an initial solution with a little margin to the most important constraint. On the other hand, the reactor holdup was initialized at its maximum value. [The reason for having T_r = 355 K, despite the maximum, which is shown as $T_{r,\max}$ = 390 K in Table 4, is that we increased the maximum value in order to make the region where the constraint on T_r was inactive more visible. Originally, the $T_{r,\max}$ value was 355 K, but this made the region very small.]

Active Constraint Regions. According to the method outlined earlier in the paper, we start by checking if any constraints will be active for all (F,p_V). Indeed, there are two, namely, the constraints on product purity $x_{\mathcal{B},B}$ and reactor holdup M_r. The first one follows from the "Avoid Product Giveaway rule", which states that *when the product prices are constant, the minimum purity constraint in the valuable product stream is always active*. The latter is also easy to explain: Reducing the holdup leads to a lower single-pass conversion, yielding more A in the column feed, without improving selectivity in favor of the desired product. Thus, the column feed will contain more A and less B, leading to a higher boilup rate. Thus, keeping the holdup at maximum saves energy.

We can also assume that one constraint will never be active, namely, $P \geq 0$. This is because as long as C is produced, we need to provide a way for it to exit the system. Since C is the most volatile component, we will always have some of it in the distillate; therefore, we must purge some of the distillate to avoid accumulation of C within the system.

This means we must determine when the following variables are at their constraint values: T_r, R, V, D, and B. Using the solvers **fmincon.m** for optimization and **fzero.m** for interpolation, we come up with the regions shown in Figure 6. Each region is referenced by a number, and Table 6 lists which variables are at their constraint value in each respective region. The constraint curves that define the regions are as follows:

- The red constraint curve in Figure 6 indicates where $T_{r,\max}$ becomes active.
- The orange constraint curve in Figure 6 indicates where $R_{\min}$ becomes active. (It is active *below* this curve.)
- The green constraint curve in Figure 6 indicates where $V_{\max}$ becomes active.
- The dashed blue line in Figure 6 could be seen as a "quasi"-constraint curve; it indicates where increasing F is no longer economically optimal, so it shows the *economic bottleneck* of the process.

In the Discussion section, we explain further the shape of the regions. In Table 7, we show the optimal value of selected model variables at four points in the disturbance space. Variables that are at a constraint are shown in bold font.

■ DISCUSSION

Shape of Active Constraint Regions. In the following, we will explain the presence of each active constraint region, at least for those that are not obvious.

- The dashed blue line in Figure 6 indicates where it is no longer optimal to increase the feed rate. That is, if F was a degree of freedom and $F_{\max}$ was specified instead, the blue line shows where the constraint $F < F_{\max}$ would no longer

Table 7. Optimal Values of Selected Variables for Different Values of F and p_V[a]

variable	Value region I	region II	region IV	region IIIb
F[mol/s]	0.1	1.0	3.0	3.0
p_V[\$/mol]	0.02	0.02	0.01	0.07
F_r[mol/s]	0.1000	1.0006	5.4219	3.3829
E_r[mol/s]	0.1024	1.0871	5.6584	3.6079
T_r[K]	328.8	**390.0**	**390.0**	**390.0**
$x_{A,r}$	0.2356	0.0815	0.2229	0.2121
$x_{B,r}$	0.7168	0.7589	0.4604	0.6225
$x_{C,r}$	0.0476	0.1596	0.3167	0.1654
Y_1[%]	73.43	82.50	61.11	68.44
Y_2[%]	2.44	8.64	6.30	7.17
B[mol/s]	0.0781	0.9108	2.3696	2.0511
$x_{A,B}$	0.1000	0.0900	0.0996	0.0996
$x_{\mathscr{B},B}$	**0.9000**	**0.9000**	**0.9000**	**0.9000**
$x_{C,B}$	0.0000	0.0100	0.0004	0.0004
D[mol/s]	0.0243	0.1762	3.2888	1.5569
$x_{A,D}$	0.6707	0.0375	0.3117	0.3603
$x_{\mathscr{B},D}$	0.1290	0.0300	0.1436	0.2569
$x_{C,D}$	0.2003	0.9325	0.5447	0.3828
L[mol/s]	1.6231	2.1365	26.7112	14.0009
V[mol/s]	1.6475	2.3128	**30.0000**	15.5578
P/D	1.0000	1.0000	0.2636	0.7541
R [mol/s]	**0.0000**	**0.0000**	2.4219	0.3829
J [\$/s]	−0.035	−0.863	−1.873	−1.378

[a] Numbers in bold font indicate active constraints.

be active. We illustrate this further with Figure 7a, where the value of the objective function J at the optimal solution is given as a function of F at three different energy prices.

- We notice that the right part of the orange constraint curve in Figure 6, separating regions IVb and V, is vertical. This is easily explained; V_{max} is active here, and once the optimal value of V reaches V_{max}, the next region boundary must be independent of p_V. Therefore, it is vertical. The same applies to the other constraint lines in the right-hand part of Figure 6.
- We also see that for very low F, where no capacity constraints are active, we have a region (region I in Figure 6) where the maximum reactor temperature ($T_{r,max}$) is not an active constraint. This is because, when the overall conversion is very high (as it is at low flow rates), we benefit from increasing the reaction *selectivity* in favor of the desired reaction (by reducing the temperature). This compensates for the lower overall conversion, which also results from lower temperature. (As mentioned in the process description, reaction 7a has a lower activation energy, and thus will be favored at low temperatures.)
- R_{min} is active ($R = 0$) at low F: If F is sufficiently low, the reactor exit stream contains very little unreacted A. Thus, there is no benefit from recycling, because we would only be recycling byproduct C. If the value of the purge stream was zero; however, we would recycle as long as there is any A in the distillate at all.
- R_{min} is active at high F: As F increases, the reactor conversion decrease (see Figure 8a) and the product stream will contain more unreacted A. This must be compensated by increasing V or decreasing L in the column. If $V = V_{max}$, we cannot increase it further and our only option is to reduce L, meaning the distillate flow rate increases and the bottoms flow rate decreases (see Figure 8b. In this situation, we have nothing to earn from recycling more, so the entire increase in distillate flow rate goes into the purge stream P, leading to a higher purge ratio P/D. For higher p_V, we may find the same, even when $V < V_{max}$, because an increase in V costs more than it gives. We discuss this further below.

The physical bottleneck indicated by the black line in Figure 6 is reached when the optimal value of the purge flow rate P reaches its maximum value. The full set of active constraints at the physical quasi-bottleneck is

$$x_{\mathscr{B},B} = x_{\mathscr{B},B,min}$$

$$M_r = M_{r,max}$$

$$T_r = T_{r,max}$$

$$V = V_{max}$$

$$P = P_{max}$$

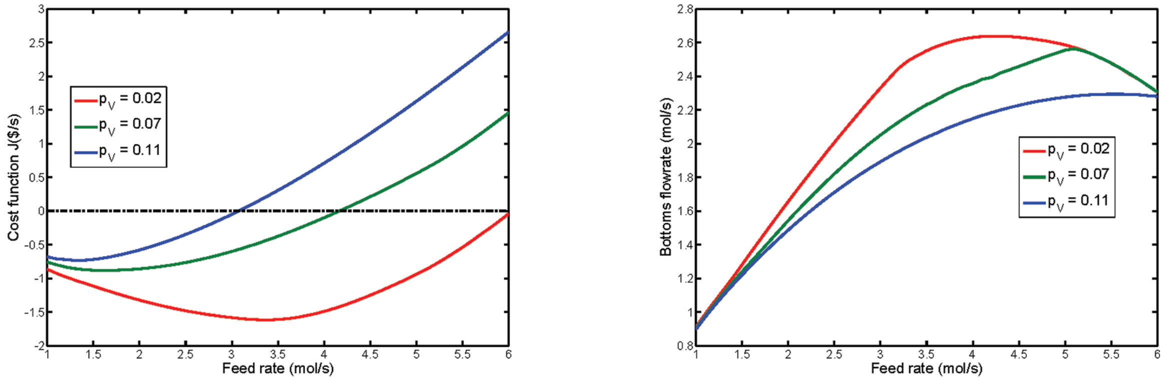
$$R = 0$$

At this point, the plant cannot process any more feed without breaking the purity constraint on the product.

More on the Effect of Recycling. As mentioned above, we found that for low and high F, it was optimal to purge all the distillate from the column, rather than recycling some of it. In Figure 9a, we show how the optimal value of R (the recycle flow rate) varies with F, and Figure 9b shows the same for vapor boilup in the column. As we can see, at the lowest p_V value, the maximum recycle flow rate is reached just as V reaches V_{max}. However, for the two higher p_V values, the maximum recycle flow rate is reached *before* V reaches V_{max}. To understand this, consider the following: Let the fresh feed F be given, and start at a p_V value where V_{max} is inactive. Now consider that we reduce the purge fraction, thereby recycling more. This means production of B goes up, but column boilup V must increase accordingly. When the energy price becomes sufficiently high, the increased column boilup costs *more* than we earn from the increased bottoms product flow rate. In other words, we *lose* money by recycling more. If the p_V value is sufficiently high, it will be optimal to have $R = 0$ for *all* F; in Figure 6, we see that this happens at $p_V \approx 0.1$ \$/mol.

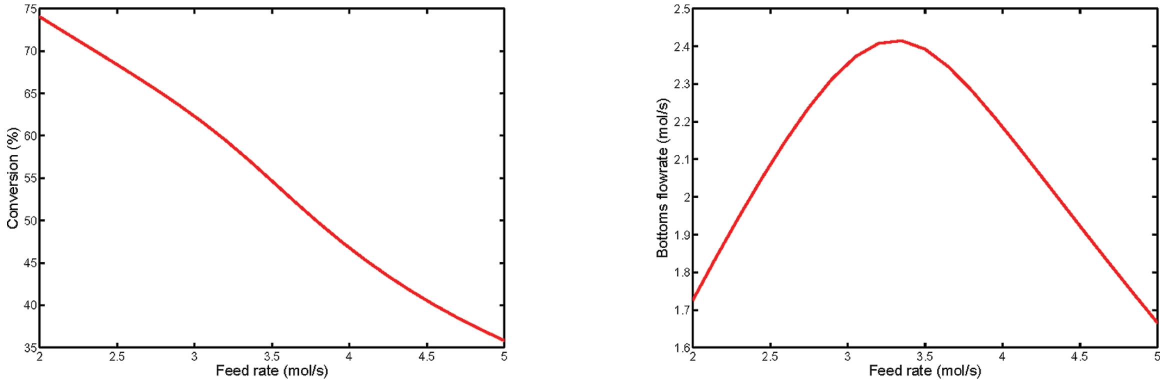
It could be argued that when we operate in the region where the purge stream is rich in the reactant (A), it would be more economic to simply bypass the plant, since we are not processing additional A to products anyway.

Efficiency and Applicability of Method. All optimizations were carried out without using analytical Jacobians, so the optimization solver would typically need 50−100 iterations to



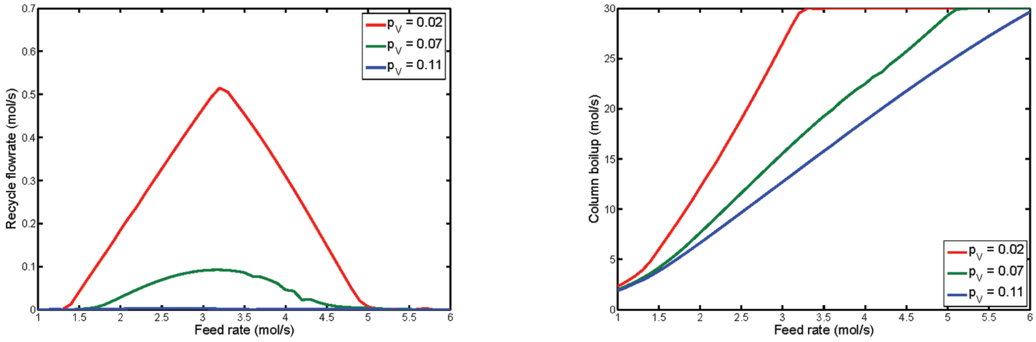
(a) Objective function J (b) Column bottoms flow rate B

Figure 7. Optimal objective function value J and column bottom product flow rate B, each as a function of the feed flow rate F at three values of p_V.



(a) Conversion as function of F (b) Bottoms flowrate as function of F

Figure 8. Reactor conversion X and bottoms flow rate B, each as a function of F when M_r, $x_{\mathcal{B},B}$, T_r, V, and P/D are fixed.



(a) Optimal value of recycle flow rate R as function of F (b) Optimal value of boilup flow rate V as function of F

Figure 9. Selected variables as a function of feed flow rate F for three different values of p_V.

converge to a solution. However, each iteration was quick, since calculation of the objective and constraint function values was easy (because they involve no further equation solving). An interpolation search would typically require ∼10 function evaluations (optimizations). If one should find the active constraint regions by just mapping the active constraints at a number of points evenly distributed over the disturbance space, it would be necessary to carry out many more optimizations, even if one kept the assumption that every region boundary was straight.

The assumptions stated earlier in the paper are only *necessary* for the method to a certain extent. However, they help simplify the thinking involved. We can treat one constraint curve at a time, and the regions will be found this way. This is simpler than starting by completely defining one region, and then trying to explore each neighboring region (if this is at all possible for highly nonlinear cases). If any single constraint function behaves highly nonlinear in parts of the disturbance space, it may be difficult to apply the method at all.

As stated in the Introduction, part of the purpose of the paper is to explore just how much can be done using process insight and how much must be determined by calculations. Our experience from this paper is that when one has less *a priori* knowledge of the effect of disturbances on the process, the amount of time spent on calculations increases, but not necessarily drastically. For example, if one does not know whether a constraint line segment will be vertical/horizontal or not, one must find just one more point more than if one knows this. (That is, as long as one is satisfied with a straight line as an approximation for the true constraint curve.)

The fact that the minimum constraint on R is active for both high and low F, but not for intermediate, gave rise to an additional challenge. For the **fzero.m** solver from MATLAB to work, it requires two end points where the function for which we seek a zero has different signs. Since the constraint was active both at high and low F, choosing values near 0 and $F_{\max}$ would yield two end points where the function value was positive. Thus, we had to carry out a few extra optimizations at intermediate F to find a point where the optimal $R \neq 0$. In general, however, finding initial points for interpolation was easy. Generally, it worked well to use the nominal point, combined with a point near the borders of the feasible region.

Comparison with Jagtap et al.[11]. Jagtap et al.[11] studied a process where two reactants A and B react according to the following reaction scheme:

$$\begin{aligned} &A + B \rightarrow C \\ &B + C \rightarrow D \end{aligned} \tag{15}$$

with C being the desired product and D being an unwanted byproduct. Instead of a purge stream, they use *two* distillation columns, and the desired product C is taken out in the distillate stream of the second column (however, the bottoms stream of that column may be seen as a replacement for a purge stream, because this stream takes care of most of the byproduct D). They use the "Avoid Product Giveaway" rule to fix *three* specifications, all of which are directly related to exit stream compositions in the distillation columns. However, they use more indirect specifications, namely, the ratio of B to C in the bottom stream from the first column, and the *loss* of C and D in the bottom and distillate streams, respectively, in column 2. These seem to be chosen more for the sake of easy steady-state convergence of the process model.

Similar to that observed in this work, they find that when fresh feed is a degree of freedom, there is a feed rate at which the plant profit reaches a maximum, and any further increase in feed rate leads to a *decrease* in profit. This is because the increased feed rate does not lead to a sufficiently large increase in the amount of valuable product. An increased feed flow rate means a lower conversion percentage, so one gets more of the impurities in the reactor product. Then more needs to be recycled, and the increase in production rate is not high enough to "pay" for the increase in feed consumption rate.

A notable difference is that the region where the reactor temperature constraint is inactive is much larger, and another constraint becomes active before it ($V_{\max}$ in the first column). This is probably because the reaction parameters are such that the temperature has a much stronger influence on reaction selectivity.

Implications for Control. For the case where the feed is a disturbance, we have five degrees of freedom for control. In the self-optimizing framework, these should be selected as follows:

(1) Active constraints should be controlled. In this case, this means that we should control $x_{B,B}$ and M_r in all operating regions. In addition, T_r should be controlled in the majority of regions.

(2) The remaining unconstrained degrees of freedom should be used to control variables whose optimal values are insensitive to disturbances. Such an analysis is outside the scope of this paper, but should be addressed in future work.

The above implies that M_r should be controlled at its maximum value ($M_{r,\max}$). However, Larsson et al.[8] have argued that if energy is inexpensive (i.e., p_V is small) and $V_{\max}$ is large, it could be beneficial, for control purposes, to let the holdup vary, giving better disturbance rejection.

CONCLUSIONS

In this paper, we have outlined a method for finding active constraint regions for a chemical process. We have applied the method to a simple reactor–distillation–recycle process, with feed flow rate and energy cost as disturbances.

For the example process, we find five distinct active constraint regions. There is a maximum feed flow rate (the physical bottleneck) above which we cannot operate without breaking constraints. If the feed flow rate is a degree of freedom, rather than a disturbance, then the active constraints regions are identical; however, in addition, we find an economic bottleneck that occurs at lower feed flow rates. Above this economic bottleneck, increasing the feed rate leads to an economic loss.

AUTHOR INFORMATION

Corresponding Author

*E-mail: sigurd.skogestad@chemeng.ntnu.no.

REFERENCES

(1) Aske, E. Design of plantwide control systems with focus on maximizing throughput. Ph.D. Thesis, NTNU, Trondheim, Norway, 2009.

(2) Gilliland, E.; Gould, L.; Boyle, T. Dynamic effects of material recycle. *Prepr. JACC, Stanford, CA* **1964**, 140–146.

(3) Wua, K.; Yu, C.; Luyben, W.; Skogestad, S. Reactor/separator processes with recycles—2. Design for composition control. *Comput. Chem. Eng.* **2001**, *27*, 401–421.

(4) Wu, K.-L.; Yu, C.-C. Reactor/separator processes with recycle—1. Candidate control structure for operability. *Comput. Chem. Eng.* **1996**, *20*, 1291–1316.

(5) Luyben, M.; Fluodas, C. Analyzing the interaction of design and control—2. Reactor–separator recycle system. *Comput. Chem. Eng.* **1994**, *18*, 971–993.

(6) Luyben, W. L. Dynamics and control of recycle systems. 1. Simple open-loop and closed-loop systems. *Ind. Eng. Chem. Res.* **1993**, *32*, 466–475.

(7) Kiss, A.; Bildea, C.; Dimian, A.; Iedema, P. In *European Symposium on Computer-Aided Process Engineering-14*, 37th European Symposium of the Working Party on Computer-Aided Process Engineering; Barbosa-Póvoa, A., Matos, H., Eds.; Computer Aided Chemical Engineering; Elsevier: Amsterdam, 2004; Vol. 18, pp 223−228.

(8) Larsson, T.; Govatsmark, M.; Skogestad, S.; Yu, C. Control structure selection for reactor, separator, and recycle processes. *Ind. Eng. Chem. Res.* **2003**, *42*, 1225–1234.

(9) Skogestad, S. Plantwide control: The search for the self-optimizing control structure. *J. Process Control* **2000**, *10*, 487.

(10) Maarleveld, A.; Rijnsdorp, J. Constraint control on distillation columns. *Automatica* **1970**, *6*, 51–58.

(11) Jagtap, R.; Kaistha, N.; Skogestad, S. Submitted to *Ind. Eng. Chem. Res.* (DOI: 10.1021/ie2024358).

(12) Biegler, L. T. *Nonlinear Programming: Concepts, Algorithms and Applications to Chemical Processes*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, 2010.

(13) Nocedal, J.; Wright, S. *Numerical Optimization*; Springer−Verlag, 1999.

(14) Kanbur, R. In *The New Palgrave Dictionary of Economics*; Durlauf, S. N., Blume, L. E., Eds.; Palgrave Macmillan: Basingstoke, Hampshire, U.K., 2008.

(15) Naidu, D. *Optimal Control Systems*; CRC Press: Boca Raton, FL, 2003.

(16) Skogestad, S.; Morari, M. Understanding the dynamic behavior of distillation columns. *Ind. Eng. Chem. Res.* **1988**, *27*, 1848–1862.

(17) de Araújo, A.; Govatsmark, M.; Skogestad, S. Application of plantwide control to the HAD process. I—Steady-state optimization and self-optimizing control. *Control Eng. Pract.* **2007**, *15*, 1222–1237.

(18) Lid, T. Data reconciliation and optimal operation. Ph.D. Thesis, NTNU, Trondheim, Norway, 2007.