Dynamic Ramping Constraints for Demand Response of Flat MIMO Processes

Florian Joseph Baader\textsuperscript{a,b}, Philipp Althaus\textsuperscript{a,b}, André Bardow\textsuperscript{a,c}, Manuel Dahmen\textsuperscript{a,*}

\textsuperscript{a} Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research, Energy Systems Engineering (IEK-10), Jülich 52425, Germany
\textsuperscript{b} RWTH Aachen University Aachen 52062, Germany
\textsuperscript{c} ETH Zürich, Energy & Process Systems Engineering, Zürich 8092, Switzerland

\textbf{Abstract:} Volatile electricity prices make demand response (DR) attractive for the process industry. However, the resulting scheduling optimization problems are notoriously challenging to solve, especially in the case of nonlinear process dynamics. Thus, problem reformulations that allow online scheduling are needed. In our previous publication (arXiv:2110.08137), we proposed high-order dynamic ramping constraints with non-constant ramp limits derived rigorously for input-state linearizable single-input single-output (SISO) processes. In the present publication, we generalize dynamic ramping constraints to flat multi-input multi-output (MIMO) processes. Approximating the nonlinear ramping limits by conservative piecewise-affine functions allows to formulate a mixed-integer linear programming (MILP) optimization problem. Moreover, in the MIMO case, dynamic ramping significantly reduces the model dimension. As a case study, dynamic ramping constraints are derived for a heated reactor-separator process with recycle that is subsequently scheduled simultaneously with a multi-energy system. The dynamic ramping formulation bridges the gap between rigorous process models and simplified process representations that allow online scheduling.

\textbf{Keywords:} Demand response, Mixed-integer dynamic optimization, Flatness, Simultaneous scheduling
1 Introduction

Reducing greenhouse gas emissions requires increased renewable electricity production that, however, gives a fluctuating supply. This fluctuating supply can be compensated by consumers that react to time-varying electricity prices by shifting their demand in time in a so-called demand response (DR) (Zhang and Grossmann, 2016). DR can be attractive for energy-intensive production processes with the flexibility to modulate their production rate. To exploit the DR potential, a scheduling optimization is needed, which typically determines operational set-points for a time horizon in the order of one day (Baldea and Harjunkoski, 2014). However, such a scheduling optimization is computationally challenging for nonlinear processes that exhibit scheduling-relevant dynamics. The scheduling optimization becomes even more difficult if processes do not only consume electricity but also heating or cooling as these processes need to be scheduled simultaneously with the local multi-energy supply system (often also referred to as utility system) (Leenders et al., 2019). Local multi-energy supply systems bring integer on/off decisions into the scheduling optimization, especially as they often consist of multiple redundant units (Voll et al., 2013). Thus, the simultaneous scheduling optimization problem usually is a nonlinear mixed-integer dynamic optimization (MIDO) problem (Baader et al., 2021).

Traditionally, such challenging scheduling MIDO problems are simplified by introducing static ramping constraints that define the rate of change of the production rate $\rho$ as degree of freedom $\nu$ and limit $\nu$ with constant ramping limits $\nu_{min}$, $\nu_{max}$, see e.g., Carrion and Arroyo (2006); Mitra et al. (2012); Adamson et al. (2017); Zhou et al. (2017); Hoffmann et al. (2021):

$$\dot{\rho} = \nu \text{ with } \nu_{min} \leq \nu \leq \nu_{max}$$  

(1)

If additionally, the nonlinear energy demand of the production process is approximated as a piecewise-affine (PWA) function, the complete problem can be reformulated as a mixed-integer linear program (MILP) (Schäfer et al., 2020).

Traditional ramping constraints, however, have two shortcomings: They are restricted to (i) first-order dynamics and (ii) constant ramping limits. To tackle both shortcomings, we proposed high-order dynamic ramping constraints in our previous publication (Baader et al., 2021). These high-order dynamic ramping constraints limit the $\delta$-th derivative of the production rate $\rho$ and use limits $\rho_{min}$, $\rho_{max}$ which are functions of the production rate and its time derivatives:

$$\rho^{(\delta)} = \nu \text{ with } \nu_{min}\left(\rho, \dot{\rho}, ..., \rho^{(\delta-1)}\right) \leq \nu \leq \nu_{max}\left(\rho, \dot{\rho}, ..., \rho^{(\delta-1)}\right)$$  

(2)

Dynamic ramping constraints allow to represent energy-intensive processes better than static ramping constraints due to their ability to account for high-order dynamics and non-constant ramping limits that depend on the process state. In Baader et al. (2021), we demonstrated that dynamic ramping constraints can achieve solutions of the scheduling optimization problem that are close to the solutions of the original nonlinear MIDO problem while allowing to formulate MILPs that can be solved sufficiently fast for real-time scheduling. Moreover, we presented a method to derive dynamic ramping constraints rigorously for the special case of single-input single-output (SISO) processes that are exact input-state linearizable. However, this case is quite restrictive, and the question remained how to derive dynamic ramping constraints for more general processes. In particular, it was open how to apply the dynamic ramping method to multi-input multi-output (MIMO) processes that have a variable production rate $\rho$ while, at the same time, other output variables $y$ such as temperatures and concentrations need to be controlled using a set of control inputs $u$ (Figure 1).

The present publication extends the dynamic ramping approach to MIMO processes and presents a method to derive dynamic ramping constraints rigorously for differentially flat MIMO processes. In simple words, a MIMO process with $m$ inputs is differentially flat if an output vector $\xi$ of the same size $m$ exists such that all process states and inputs can be given as a function of the output $\xi$ and its $\beta$ time derivatives $\xi, ..., \xi^{(\beta)}$ (Fließ et al., 1995; Rothfüß, 1997; Rothfüß et al., 1996). We make use of the fact that a flat nonlinear model can be transformed to a linear model (Fließ et al., 1995). However, constraints that are linear for the original model, e.g., bounds on inputs, are nonlinear for the transformed model. In other words, a nonlinear model with linear constraints is transformed into a linear model with nonlinear constraints. Approximating the nonlinear bounds to the safe side with piecewise-affine functions
allows us to formulate a MILP. Note that for SISO processes, flatness is equivalent to exact input-state linearizability (Adamy, 2014), which was the main assumption in our previous work (Baader et al., 2021). For MIMO processes, exact input-state linearizability with static state transformation is a special case of flatness (Adamy, 2014).

The remaining paper is structured as follows: In Section 2, the considered demand response optimization problem and the dynamic ramping constraints are introduced. In Section 3, a method is presented to derive dynamic ramping constraints for flat MIMO processes. In Section 4, a case study is presented. Section 5 provides further discussion and Section 6 concludes the work.

2 Dynamic mixed-integer linear scheduling with ramping constraints

This section briefly introduces the simultaneous scheduling optimization problem of flexible processes and multi-energy supply systems. A more extensive discussion is given in our previous publication (Baader et al., 2021). The optimization problem (P) reads:

$$\min \Phi_{\text{energy}}(t_f) \quad \text{(Pa)}$$

s.t. Dynamic ramping constraints \( (Pb) \)
Process energy demand model \( (Pc) \)
Product storage model \( (Pd) \)
Energy costs \( (Pe) \)
Energy conversion model \( (Pf) \)
Minimum part-load constraints \( (Pg) \)
Energy balances \( (Ph) \)

The objective is to minimize the cumulative energy costs \( \Phi_{\text{energy}} \) at final time \( t_f \). In the following, we focus on the dynamic ramping constraints \( (Pb) \) and the process energy demand model \( (Pc) \). The remaining constraints \( (Pd) \) to \( (Ph) \) are standard constraints (Schäfer et al., 2020; Sass et al., 2020) and further explanation is given in our previous publication (Baader et al., 2021).

In our previous publication (Baader et al., 2021), we only considered a single input. We, therefore, only had to constrain a single time derivative of the production rate \( \rho \) with order \( \delta \), which we defined as the ramping degree of freedom \( \nu = \rho^{(\delta)} \). In the present contribution, we consider multiple inputs \( u \) and can potentially have constraints on all considered derivatives of the production rate \( \rho^{(\gamma)} \) with \( \gamma = 1, \ldots, \delta \) and the integer \( \delta \) being the order of the highest time derivative that is constrained by input bounds. For instance, the bounds of input \( u_1 \) could directly limit the first derivative of the production rate, \( \dot{\rho} \), whereas the bounds on input \( u_2 \) could limit the second derivative of the production rate, \( \dot{\rho}^{(2)} \), directly and only influence the first derivative, \( \dot{\rho} \), through the integration. The generalized dynamic ramping constraints

![Fig. 1. A MIMO process that can vary its production rate \( \rho \) and thus its energy demand while additional control inputs \( u \) are available to control further process outputs variables \( y \).](image-url)
(DRCs) developed in this publication therefore read:

\[ \dot{\rho}^{(\delta)} = \nu \]  \hspace{1cm} (DRCa)

\[ \dot{\rho}^{\text{min}}(\rho) \leq \dot{\rho} \leq \dot{\rho}^{\text{max}}(\rho) \]  \hspace{1cm} (DRCb)

\[ (\rho^{(\gamma)})^{\text{min}}(\rho, \dot{\rho}, ..., \rho^{(\gamma-1)}) \leq \rho^{(\gamma)} \leq (\rho^{(\gamma)})^{\text{max}}(\rho, \dot{\rho}, ..., \rho^{(\gamma-1)}) \]  \hspace{1cm} (DRCc)

\[ \nu^{\text{min}}(\rho, \dot{\rho}, ..., \rho^{(\delta-1)}) \leq \nu \leq \nu^{\text{max}}(\rho, \dot{\rho}, ..., \rho^{(\delta-1)}) \]  \hspace{1cm} (DRCd)

Still, only the highest considered time derivative \( \rho^{\delta} \) is a degree of freedom and thus defined as the ramping degree of freedom \( \nu \).

The limits \((\rho^{(\gamma)})^{\text{min}}, (\rho^{(\gamma)})^{\text{max}}\) as well as \(\nu^{\text{min}}, \nu^{\text{max}}\) are in general nonlinear functions that we derive by coordinate transformation (Section 3). To incorporate the dynamic ramping constraint into a MILP formulation, the true nonlinear limits are approximated by piecewise-affine (PWA) functions for all considered derivatives \(\gamma = 1, ..., \delta\) because PWA functions allow us to formulate a mixed-integer linear scheduling problem. These piecewise-affine functions need to be conservative such that they prohibit constraint violation with respect to the true nonlinear limits to guarantee that the chosen trajectory for \( \rho \) satisfies all bounds on inputs and states. Accordingly, the approximations of \((\rho^{(\gamma)})^{\text{min}}\) and \(\nu^{\text{min}}\) must always be greater than or equal to the true nonlinear limits and the approximations of \((\rho^{(\gamma)})^{\text{max}}\) and \(\nu^{\text{max}}\) must always be lower than or equal to the real nonlinear limits. Choosing the quality of the approximations allows us to explicitly balance the achievable flexibility range against the computational burden. Finally, time is discretized using collocation on finite elements to convert the dynamic ramping constraints (DRC) into a set of piecewise-affine constraints (Biegler, 2010).

As in our previous publication (Baader et al., 2021), the process energy demand \(Q_{\text{dem,e}}(\rho)\) (cf. Equation Pec) for an energy form \(e\) is modeled as a function of the production rate and its time derivatives:

\[ Q_{\text{dem,e}}^{\text{process}}(\rho, \dot{\rho}, ..., \rho^{(\delta-1)}, \nu) \]  \hspace{1cm} (3)

Similar to the DRC, a piecewise-affine function is chosen for \(Q_{\text{dem,e}}^{\text{process}}\) to achieve an MILP formulation.

The problem formulation (P) with DRCs captures a larger share of the dynamic flexibility range of processes than static ramping constraints can do while still allowing to formulate a mixed-integer linear program. However, to choose suitable piecewise-affine ramping limits, the true nonlinear limits of the process need to be derived or approximated. In the following section, we show how these limits can be derived rigorously for MIMO processes that are differentially flat.

3 Deriving dynamic ramping constraints

In Section 3.1, necessary assumptions are stated and then our approach is presented in Section 3.2.

3.1 Assumptions

1. The process degrees of freedom are divided into a control input vector \(u\) and the variable production rate \(\rho\). The process model is a system of ordinary differential equations given by:

\[ \dot{x} = f(x, u, \rho) \]  \hspace{1cm} (4)

with state vector \(x \in \mathbb{R}^n\), and a nonlinear right-hand side function \(f(x, u, \rho)\). We assume that there are no further inputs or disturbances to the process.
2. The control input vector \( \mathbf{u} \), and the production rate \( \rho \) are bounded by minimum and maximum values \( u^{\text{min}}, u^{\text{max}} \), and \( \rho^{\text{min}}, \rho^{\text{max}} \), respectively. Similarly, the states \( \mathbf{x} \) have to be maintained within bounds \( x^{\text{min}}, x^{\text{max}} \).

3. The process \( (4) \) is flat. That is, the process has one or multiple flat output vectors \( \xi \). An output vector \( \xi \) is flat if it satisfies three conditions (Fliess et al., 1995; Rothfuß, 1997; Rothfuß et al., 1996):

(a) The flat output vector can be given as a function \( \phi \) of states \( \mathbf{x} \), inputs \( \mathbf{u} \), production rate \( \rho \), and time derivatives of \( \mathbf{u} \) and \( \rho 

\[
\xi = \phi (\mathbf{x}, \mathbf{u}, \dot{\mathbf{u}}, ..., \mathbf{u}^{(\alpha)}, \rho, \dot{\rho}, ..., \rho^{(\kappa)})
\]

with finite integer numbers \( \alpha, \kappa \). The function \( \phi \) can be seen as a transformation from the original state and input space to the flat output space. Often, it is possible to choose flat outputs that have a physical meaning, e.g., the conversion of a reactor, and that are a function of the states \( \mathbf{x} \) only (Adamy, 2014).

(b) A backtransformation from the flat output and its derivatives to the original states \( \mathbf{x} \) and inputs \( \mathbf{u} \) can be found. Accordingly, the system states \( \mathbf{x} \) and inputs \( \mathbf{u} \) can be given as functions \( \psi_1, \psi_2 \) of the flat outputs \( \xi \), the production rate \( \rho \), and a number of time derivatives of \( \xi \) and \( \rho 

\[
x = \psi_1 (\xi, \dot{\xi}, ..., \xi^{(\beta-1)}, \rho, \dot{\rho}, ..., \rho^{(\zeta-1)})
\]
\[
\mathbf{u} = \psi_2 (\xi, \dot{\xi}, ..., \xi^{(\beta)}, \rho, \dot{\rho}, ..., \rho^{(\zeta)})
\]

with finite integer numbers \( \beta, \zeta \).

(c) The components of \( \xi \) are differentially independent (Rothfuß, 1997). Consequently, they do not fulfill any differential equation:

\[
\mu(\xi, \dot{\xi}, ..., \xi^{(\beta)}) = 0
\]

Condition (5c) is satisfied if condition (5b) is satisfied, \( \dim(\xi) = \dim(\mathbf{u}) = m \), and \( \text{rank}\left(\frac{\partial f(x,u,\rho)}{\partial u}\right) = m \), where \( m \) is the number of inputs (Rothfuß, 1997).

Note: To check conditions (5a) - (5c), a candidate for a flat output vector \( \xi \) is needed. We assume that such a candidate for a flat output vector \( \xi \) can be identified based on engineering intuition.

4. The trajectory of the production rate \( \rho \) is determined by the scheduling optimization. Subsequently, the control input vector \( \mathbf{u} \in \mathbb{R}^m \) is calculated by an underlying process control.

The flatness-based coordinate transformation is visualized in Figure 2. A nonlinear model with linear constraints is transformed into a linear model with nonlinear constraints. The number of control degrees of freedom is maintained: While, in the original model, the \( m \) inputs \( u_k \) \( (k = 1, ..., m) \) are the degrees of freedom, in the linear model, the \( m \) highest time derivatives of the output components \( \xi_k^{(\beta_k)} \) \( (k = 1, ..., m) \) are the degrees of freedom (Adamy, 2014; Fliess et al., 1995). The number of time derivatives \( \beta_k \) can deviate between the flat output components \( \xi_k \) (Adamy, 2014; Fliess et al., 1995). In the linear model, the transformed state vector \( \Xi \) is formed by the outputs and their time derivatives (except for the highest time derivative):

\[
\Xi = \left(\xi_1^{(\beta_1-1)}, \xi_2^{(\beta_2-1)}, ..., \xi_m^{(\beta_m-1)}\right)^T
\]

The dimension of the transformed state vector \( \Xi \) is greater or equal to the dimension of the original state vector \( \mathbf{x} \) (Adamy, 2014; Fliess et al., 1995). Consequently, the original nonlinear process model is converted to a linear model consisting of \( m \) integrator chains. In the linear model, every flat output \( \xi_k \)
3 DERIVING DYNAMIC RAMPING CONSTRAINTS

\[
\begin{align*}
\mathbf{f}_1(\mathbf{x}, \mathbf{u}, \rho) & \rightarrow \mathbf{x}_1 \\
\mathbf{f}_2(\mathbf{x}, \mathbf{u}, \rho) & \rightarrow \mathbf{x}_2 \\
\vdots & \rightarrow \mathbf{x}_n \\
\mathbf{u} & = \begin{bmatrix} u_1 \\
\vdots \\
u_m \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1(\mathbf{x}, \mathbf{u}, \rho) \\
\mathbf{f}_2(\mathbf{x}, \mathbf{u}, \rho) \\
\vdots \\
\mathbf{f}_n(\mathbf{x}, \mathbf{u}, \rho) \end{bmatrix}, \\
\mathbf{u}_{\text{min}} \leq \mathbf{u} \leq \mathbf{u}_{\text{max}} \\
\mathbf{x}_{\text{min}} \leq \mathbf{x} \leq \mathbf{x}_{\text{max}}
\end{align*}
\]

\[
\begin{align*}
\mathbf{f}_1(\mathbf{x}, \mathbf{u}, \rho) & \rightarrow \mathbf{x}_1 \\
\mathbf{f}_2(\mathbf{x}, \mathbf{u}, \rho) & \rightarrow \mathbf{x}_2 \\
\vdots & \rightarrow \mathbf{x}_n \\
\mathbf{u} & = \begin{bmatrix} u_1 \\
\vdots \\
u_m \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1(\mathbf{x}, \mathbf{u}, \rho) \\
\mathbf{f}_2(\mathbf{x}, \mathbf{u}, \rho) \\
\vdots \\
\mathbf{f}_n(\mathbf{x}, \mathbf{u}, \rho) \end{bmatrix}, \\
\mathbf{u}_{\text{min}} \leq \mathbf{u} \leq \mathbf{u}_{\text{max}} \\
\mathbf{x}_{\text{min}} \leq \mathbf{x} \leq \mathbf{x}_{\text{max}}
\end{align*}
\]

Fig. 2. Visualization of flatness-based coordinate transformation showing original nonlinear MIMO process model as introduced in assumption 1 (left), transformations \(\phi, \psi_1, \psi_2\) as introduced in assumption 3 (gray arrows), and linear MIMO process model in transformed coordinate space as introduced in assumption 3 (right).

can be varied with a \(\beta_k\)-th order dynamic independently of the other flat outputs. Note that flatness is a sufficient condition for controllability of a nonlinear process (Adamy, 2014).

To ease notation in the following, we introduce the ramping state vector

\[
\varphi = \left( \rho, \dot{\rho}, \ldots, \rho^{(\delta-2)}, \rho^{(\delta-1)} \right)^T,
\]

and its time derivative

\[
\dot{\varphi} = \left( \dot{\rho}, \rho^{(2)}, \ldots, \rho^{(\delta-1)}, \nu \right)^T.
\]

### 3.2 Approach

The steps to derive dynamic ramping constraints further discussed in the following subsections are summarized in Figure 3: First, a candidate for a flat output vector is selected based on two necessary flatness conditions (Section 3.2.1). Assuming this candidate is, in fact, a flat output, the original nonlinear MIMO process model is transformed into a linear MIMO model (center of Figure 3). In this linear model, the \(m\) components of the flat output vector \(\xi_k\) are decoupled such that the outputs \(\xi_k\) can be controlled independently of each other by manipulating the degrees of freedom \(\xi_k^{(\beta_k)}\) (Figure 2). However, a SISO model is needed for the dynamic ramping constraints where the production rate \(\rho\) is controlled by manipulating the ramping degree of freedom \(\nu\). Thus, second, the flat output components \(\xi_k\) are coupled by choosing an operating strategy \(\pi(\rho)\) that defines every \(\xi_k\) as function of the production rate \(\rho\) (Section 3.2.2). This coupling reduces the number of degrees of freedom from \(m\) to one, leading to a significant model order reduction. Third, backtransformations are found from the ramping SISO model to the original nonlinear MIMO model, and thus flatness is proven (Section 3.2.3).

Based on the backtransformations and the ramping SISO model, the true nonlinear ramping limits are derived from the bounds on inputs \(\mathbf{u}\) and states \(\mathbf{x}\). After approximating the true nonlinear limits with piecewise-affine functions, the dynamic ramping constraints are complete and the problem (P) can be solved as MILP.
3 DERIVING DYNAMIC RAMPING CONSTRAINTS

Fig. 3. Overview of steps to derive dynamic ramping constraints performed in the respective sections. For the nonlinear MIMO model and the linear MIMO model, all variable symbols are omitted for clarity as they are identical to Figure 2. The variables of the ramping SISO model are the production rate \( \rho \), its derivatives \( \dot{\rho}, ..., \rho^{(\delta-1)} \), and the ramping degree of freedom \( \nu \).

3.2.1 Selection of flat output candidate and necessary flatness conditions

First, we propose to apply a necessary condition for flatness from literature based on a graph representation of the process (Schulze and Schenkendorf, 2020) to identify a candidate for the flat output vector \( \xi \) as \( \phi(x, u, ..., u^{(\alpha)}, \dot{\rho}, ..., \dot{\rho}^{(\alpha)}) \) (compare to Equation (5a)). As the second necessary condition for flatness, we propose setting up an equation system that implicitly defines the backtransformation \( \psi \) (compare to Equations (5b1) and (5b2)) and check if this backtransformation equation system is structurally solvable.

For the graph representation, the process model is represented as a directed graph in which all states \( x_i \) and inputs \( u_k \) are represented as vertices \( v_{x_i} \) and \( v_{u_k} \), respectively (Schulze and Schenkendorf, 2020). If \( \frac{\partial f_i(x, u)}{\partial x_j} \) is non-zero, there is an edge from vertex \( v_{x_i} \) to vertex \( v_{x_j} \). In other words: If a state \( x_j \) acts on the derivative of another state \( x_i \), an edge is drawn from \( v_{x_j} \) to \( v_{x_i} \). Similarly, an edge from input \( v_{u_k} \) to state \( v_{x_i} \) exists, if \( \frac{\partial f_i(x, u)}{\partial u_j} \) is non-zero. The necessary condition for a flat output vector \( \xi \) is that it must be possible to match the \( m \) output components \( \xi_i \) to the input components \( u_i \) such that there are \( m \) input-output pairs with pairwise disjoint paths through the graph that cover all process states (Schulze and Schenkendorf, 2020). An illustrative example for this necessary condition is given in the Supplementary Information (SI).

As a starting point, the states which are typically controlled can be tested as flat output candidates. Typical states to be controlled are outlet stream compositions, final effluent temperature, or the process hold-up (Jogwar et al., 2009).

Once a flat output vector candidate is identified based on the graph representation as \( \xi = \phi(x, u, \dot{u}, ..., u^{(\alpha)}, \dot{\rho}, ..., \dot{\rho}^{(\alpha)}) \), condition 5a is fulfilled. Further, the rank criterion needed for condition 5c can be checked easily. As a final step to show flatness, existence of the backtransformations \( \psi_1, \psi_2 \) that give states \( x \) and inputs \( u \) as functions of the flat output \( \xi \) and its time derivatives \( \dot{\xi}, ..., \dot{\xi}^{(\beta)} \) needs to be shown (condition 5b). To obtain these backtransformations, a nonlinear system of equations needs to be developed with \( \xi \) and its derivatives \( \xi_i, ..., \xi^{(\beta)} \) on the right-hand side and left-hand side functions including the wanted quantities: states \( x \), inputs \( u \), and potentially derivatives of inputs. To set up the backtransformation, we make use of the fact that \( \xi \) is given by \( \phi(x, u, \dot{u}, ..., u^{(\alpha)}, \dot{\rho}, ..., \dot{\rho}^{(\alpha)}) \) and the
derivatives of $\xi$ can be given as functions of $x$, inputs $u$, and derivatives of inputs by means of the total differential. The equation system for the backtransformation needs to be square such that there are as many equations as there are unknown states, inputs, and derivatives of inputs. We propose to check if this nonlinear equation system is structurally solvable by conducting an analysis similar to the structural index analysis for differential-algebraic equation systems (Unger et al., 1995) commonly used in process systems engineering. An example is given in the SI.

### 3.2.2 Operating strategy

Under the assumption that the flat output candidate identified as discussed in Section 3.2.1 is, in fact, a flat output, the nonlinear model can be transformed into a linear model. This linear model is a MIMO model with $m$ degrees of freedom $\xi^{(m)}$ and $m$ outputs $\xi$, (Figure 2). As discussed above, the linear MIMO is transferred to a SISO ramping model by coupling the flat output components. To this end, we insert an operating strategy that gives the value of every flat output as a function of the production rate $\rho$. This operating strategy differentiates between two possible types of flat output components: First, flat output components might have specifications that should be maintained constant, such as outlet stream compositions, final effluent temperature, and the process hold-up (Jogwar et al., 2009). Accordingly, the operating strategy is to hold such an output $\xi_k$ constant at its nominal value such that $\xi_k = \xi^{\text{nom}}_k$ holds and thus all time derivatives are zero.

Second, flat output components may be unspecified. For instance, in our case study, one flat output component is a concentration for which no specifications are given. As every flat output component $\xi_k$ corresponds to one control input, i.e., degree of freedom, $u_k$, if specifications are given for $l$ outputs, and $l$ is smaller than $m$, $m-l$ flat output components remain as degrees of freedom in steady-state. Thus, the $m-l$ free flat output component $\xi_k$ can, in principle, be chosen to have any value in steady-state as long as no variable bounds are violated. To have the optimal steady-state operating points, we use a steady-state optimization to determine the optimal values $\xi_k$ in advance as a function of the production rate $\rho$ such that $\xi_k = \pi_k(\rho)$. For instance, the objective can be to find the steady-state operating points that minimize energy consumption. In our case study, we choose the flat output component such that the overall heat demand is minimal for steady-state points (Section 4).

The operating strategy $\pi_k(\rho)$ can be any nonlinear function. The only requirement is that the function $\pi_k(\rho)$ must be differentiable with respect to $\rho$ sufficiently often so that all derivatives of $\xi_k$ which are part of the backtransformation discussed in the previous section are defined by the total differential, e.g.,

$$\dot{\xi}_k = \frac{\partial \pi_k}{\partial \rho} \rho, \quad \xi^{(2)}_k = \frac{\partial^2 \pi_k}{\partial \rho^2} \rho^2 + \frac{\partial \pi_k}{\partial \rho} \rho.$$

When all output components with specifications are maintained at their nominal values and a function $\pi_k(\rho)$ is chosen for all other output components, the operating strategy can be summarized as $\xi = \pi(\rho)$. Consequently, the flat output vector $\xi$ and all relevant time derivatives $\dot{\xi}, \ddot{\xi}, ..., \xi^{(3)}$ (compare to Equations (5b1) and (5b2)) are defined as function of the production rate $\rho$ and a number of its time derivatives. The highest time derivative $\rho^{(d)}$ that occurs defines the order of the dynamic ramping constraint (DRC) and the ramping degree of freedom $\nu = \rho^{(d)}$. Consequently, the backtransformations $\psi_1, \psi_2$ discussed in the following only depend on the ramping state vector $\varphi$ and the ramping degree of freedom $\nu$.

### 3.2.3 Backtransformation and reformulation to dynamic ramping constraints

First, the flat output vector $\xi$ and its time derivatives $\dot{\xi}, \ddot{\xi}, ..., \xi^{(3)}$ are replaced in the nonlinear equation system for the backtransformation derived in Section 3.2.1. While $\xi$ is replaced by $\pi(\rho)$, the derivatives are replaced by building the total differential of $\pi(\rho)$. Next, we solve the equation system to get $x = \psi_1(\varphi)$ and $u = \psi_2(\varphi, \nu)$. It is favorable to solve the equation system analytically. Still, it is not necessary to derive the functions $\psi_1, \psi_2$ analytically as they can also be evaluated numerically as long as their solution is unique. In this paper, we use the computer algebra package SymPy (Meurer et al., 2017) to obtain analytic functions $\psi_1, \psi_2$. Note that it might not be possible to solve the system of equations because the graphical and structural criteria proposed in Section 3.2.1 are only necessary flatness conditions. In case the nonlinear system of equations cannot be solved, one can test other flat output candidates. If the
functions $\psi_1, \psi_2$ are found, condition 5b is fulfilled, and flatness is shown at least locally on the subspace defined by the operating strategy.

This flat system constitutes one integrator chain with the degree of freedom $\nu$ that can be chosen arbitrarily at any point in time. Thus, mathematically, the production rate can be changed infinitely fast by choosing infinitely high values for $\nu$. However, the real process control inputs $u$ are bounded by maximum and minimum values (assumption 3), and therefore, dynamic ramping constraints (DRCs) are needed to ensure that the real process inputs are maintained within bounds.

To get the dynamic ramping constraints (DRCs), we consider the input bounds row by row. For the $k$-th row, we get

$$u_k^{\min} \leq \psi_{2,k}(\rho, \dot{\rho}, ..., \rho^{(\gamma)}) \leq u_k^{\max},$$

where $\rho^\gamma$ is the highest time derivative needed to compute $u_k$. Equation (9) implicitly limits $\rho^\gamma$ for given $\rho, \dot{\rho}, ..., \rho^{(\gamma-1)}$. These limits can be given as an explicit function if it is possible to symbolically invert $\psi_{2,k}$ for the derivative $\rho^{(\gamma)}_k$ and thus derive an analytic function $\theta$ of the production rate $\rho$, derivatives $\dot{\rho}, ..., \rho^{(\gamma-1)}$, and the input $u_k$:

$$\rho^{(\gamma)}_k = \theta (\rho, \dot{\rho}, ..., \rho^{(\gamma-1)}, u_k)$$

Inserting $u_k^{\min}, u_k^{\max}$ into $\theta$ gives $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$. Alternatively, the limits can be derived numerically for given $\rho, \dot{\rho}, ..., \rho^{(\gamma-1)}$ by sampling different values for $\rho^{(\gamma)}_k$ and evaluating if the resulting $u_k$ is within the allowed range.

Note: Here, we discuss the case of one allowed region for $\rho^{(\gamma)}_k$ given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$. In principle, there could be two (or even more) non-connected allowed regions given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$ and $\rho^{(\gamma)}_k^{\min_2}, \rho^{(\gamma)}_k^{\max_2}$, with $\rho^{(\gamma)}_k^{\max_1} < \rho^{(\gamma)}_k^{\min_2}$. However, the real process inputs are bounded by maximum and minimum values, and therefore, dynamic ramping constraints (DRCs) are needed to ensure that the real process inputs are maintained within bounds. Note: Here, we discuss the case of one allowed region for $\rho^{(\gamma)}_k$ given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$. In principle, there could be two (or even more) non-connected allowed regions given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$ and $\rho^{(\gamma)}_k^{\min_2}, \rho^{(\gamma)}_k^{\max_2}$, with $\rho^{(\gamma)}_k^{\max_1} < \rho^{(\gamma)}_k^{\min_2}$. However, the real process inputs are bounded by maximum and minimum values, and therefore, dynamic ramping constraints (DRCs) are needed to ensure that the real process inputs are maintained within bounds. Note: Here, we discuss the case of one allowed region for $\rho^{(\gamma)}_k$ given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$. In principle, there could be two (or even more) non-connected allowed regions given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$ and $\rho^{(\gamma)}_k^{\min_2}, \rho^{(\gamma)}_k^{\max_2}$, with $\rho^{(\gamma)}_k^{\max_1} < \rho^{(\gamma)}_k^{\min_2}$. However, the real process inputs are bounded by maximum and minimum values, and therefore, dynamic ramping constraints (DRCs) are needed to ensure that the real process inputs are maintained within bounds. Note: Here, we discuss the case of one allowed region for $\rho^{(\gamma)}_k$ given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$. In principle, there could be two (or even more) non-connected allowed regions given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$ and $\rho^{(\gamma)}_k^{\min_2}, \rho^{(\gamma)}_k^{\max_2}$, with $\rho^{(\gamma)}_k^{\max_1} < \rho^{(\gamma)}_k^{\min_2}$. However, the real process inputs are bounded by maximum and minimum values, and therefore, dynamic ramping constraints (DRCs) are needed to ensure that the real process inputs are maintained within bounds. Note: Here, we discuss the case of one allowed region for $\rho^{(\gamma)}_k$ given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$. In principle, there could be two (or even more) non-connected allowed regions given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$ and $\rho^{(\gamma)}_k^{\min_2}, \rho^{(\gamma)}_k^{\max_2}$, with $\rho^{(\gamma)}_k^{\max_1} < \rho^{(\gamma)}_k^{\min_2}$. However, the real process inputs are bounded by maximum and minimum values, and therefore, dynamic ramping constraints (DRCs) are needed to ensure that the real process inputs are maintained within bounds. Note: Here, we discuss the case of one allowed region for $\rho^{(\gamma)}_k$ given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$. In principle, there could be two (or even more) non-connected allowed regions given by limits $\rho^{(\gamma)}_k^{\min_1}, \rho^{(\gamma)}_k^{\max_1}$ and $\rho^{(\gamma)}_k^{\min_2}, \rho^{(\gamma)}_k^{\max_2}$, with $\rho^{(\gamma)}_k^{\max_1} < \rho^{(\gamma)}_k^{\min_2}$.
Fig. 4. Case study of reactor-separator process with recycle: States x are the concentrations of component A and B, c_A, and c_B, respectively, and the temperature T in the reactor (1) and in the flash (2). Manipulated control inputs u are the bottom stream F_B, the purge stream F_P, the heat input to the reactor Q_1, and the heat input to the flash Q_2. The scheduling degree of freedom is the production rate ρ. All other material flow rates are given as functions of F_B, F_P, and ρ, e.g., the reactant stream is equal to the purge F_P plus the production rate ρ as no accumulation of material occurs.

specifications for the outlet stream are given. As fourth output component, we choose the concentration c_A1 based on the graph representation in Figure 5. The four input-output pairs F_B − c_A2, F_P − c_B2, Q_2 − T_2, Q_1 − c_A1 fulfill the necessary condition for a flat output (Figure 5).

Fig. 5. Graph representation for the reactor-separator process with recycle (compare to Figure 4). The output ξ = (c_A2, c_B2, T_2, c_A1)^T fulfills the necessary condition for a flat output.

By differentiating the components of the output ξ = (c_A2, c_B2, T_2, c_A1) up to three times, we receive a structurally solvable system of equations (Table S3 in the SI).

4.2 Operating strategy

For the operating strategy, we assume that the composition and temperature of the product stream ρ must be maintained constant. Accordingly, ξ_1 = c_A2, ξ_2 = c_B2, ξ_3 = T_2 must be maintained at their nominal values ξ_1^{nom} = 0.4539, ξ_2^{nom} = 0.4610, ξ_3^{nom} = 455K. Thereby, the considered operating region is already significantly reduced.

As there are 4 control inputs, we can maintain the first three flat output components at their nominal values and still have one degree of freedom left. Consequently, an operating strategy for the fourth flat output ξ_4 = c_A1 can be chosen freely. In a steady-state optimization, we search for the steady-state operating points that minimize the total heating Q_1 + Q_2 and fix ξ_4 to be a function of the production rate ξ_4 = π_4(ρ). Further details on this steady-state optimization are given in the SI.
4 CASE STUDY: REACTOR-SEPARATOR PROCESS WITH RECYCLE

Tab. 1. Structural dependency resulting from nonlinear transformation and operating strategy of states and inputs on the production rate $\rho$ and its time derivatives $\dot{\rho}$ and $\nu$ ($= \rho^{(2)}$). States which are held constant are marked with a star (*).

| states: | $\rho$ | $\dot{\rho}$ | $\nu$ |
|---------|--------|-------------|-------|
| $c_{A1}$ | x      |             |       |
| $c_{B1}$ | x      |             |       |
| $T_1$   | x      | x           |       |
| $c_{A2}^*$ | x   |             |       |
| $c_{B2}^*$ | x   |             |       |
| $T_2^*$ | x      |             |       |

| inputs: |        |           |
|---------|--------|-----------|
| $F_B$  | x      |           |
| $F_p$  | x      | x         |
| $Q_1$  | x      | x         | x      |
| $Q_2$  | x      |           |

4.3 Reformulation to dynamic ramping constraints

After inserting the operating strategy defined in Section 4.2, we solve the nonlinear backtransformation equation system. To this end, we use the computer algebra package SymPy (Meurer et al., 2017) to find explicit algebraic expressions for all states and inputs, except for the temperature $T_1$. The temperature $T_1$ has to be determined numerically because the two reaction terms in the differential Equation (S9) (compare to the SI) lead to an equation of the form

$$0 = b_0 + b_1 e^{-\frac{E_1}{RT_1}} + b_2 e^{-\frac{E_2}{RT_1}},$$

with parameters $b_0, b_1, b_2$. The implicit Equation (11) has a unique solution for the temperature, as the exponential functions are monotonic. The numerical solution of Equation (11) is found using the python package SciPy (Virtanen et al., 2020).

Table 1 provides the structural dependency of the states and inputs on the production rate $\rho$ and its time derivatives $\dot{\rho}$ and $\nu$ ($= \rho^{(2)}$). The highest time derivative of the production rate that appears is two. Therefore, the ramping degree of freedom $\nu$ is equal to $\rho^{(2)}$. More detailed information about the reformulations is given in the SI.

For the dynamic ramping constraints, it has to be analyzed which state and input bounds limit the time derivatives of the production rate $\rho$. States and inputs that are either held constant or exclusively depend on the production rate $\rho$ do not need to be checked because it is already known from the steady-state optimization that these variables take feasible values for all considered production rates. Consequently, the variables $c_{A1}, c_{B1}, c_{A2}, c_{B2}, T_2, F_B$ do not influence the dynamic ramping constraints (compare to Table 1).

The variables $T_1, F_p, Q_2$ depend on $\rho$ and $\dot{\rho}$ but not on the second time derivative $\nu$. Accordingly, the bounds of these variables limit the first time derivative $\dot{\rho}$. In Figure 6, the limits on $\dot{\rho}$ resulting from variable bounds are shown over the production rate $\rho$. The calculation of these limits is explained in the SI. While the lower limit on $\dot{\rho}$ results from the bound $F_p^{\min}$, the upper limit is given by $F_p^{\max}$ for small production rates and by $Q_2^{\min}$ for large production rates. Graphically, we choose conservative linear functions for the limits $\dot{\rho}^{\min}(\rho), \dot{\rho}^{\max}(\rho)$ (compare to (DRCb)).

The heating input $Q_1$ is the only variable that depends on the ramping degree of freedom $\nu$. We choose linear limits $\nu^{\min}(\rho, \dot{\rho}), \nu^{\max}(\rho, \dot{\rho})$ (compare to DRCd) which cover 80 % of the feasible area. Further details are given in the SI.

With the limits on $\nu$, the second-order dynamic ramping constraints are completely parameterized and have the form:

$$\rho^{(2)} = \nu$$

$$\dot{\rho}^{\min}(\rho) \leq \dot{\rho} \leq \dot{\rho}^{\max}(\rho)$$

$$\nu^{\min}(\rho, \dot{\rho}) \leq \nu \leq \nu^{\max}(\rho, \dot{\rho})$$

(12)

(13)

(14)
4 CASE STUDY: REACTOR-SEPARATOR PROCESS WITH RECYCLE

Fig. 6. Limits on the scaled time derivative of the production rate $\dot{\rho}$ over the scaled production rate $\rho$ resulting from the minimum (min) and maximum (max) values of the state $T_1$ and the inputs $F_p, Q_2$.

4.4 Investigation 1: Ramp optimizations

To illustrate the ramping behavior with the derived dynamic ramping constraints, we perform two as-fast-as-possible ramp optimizations shown in Figure 7. The ramp-up from minimum production rate to maximum production rate takes 1.3 hours, and the corresponding ramp-down takes 3.1 hours. The ramp-up is first constrained by the acceleration, i.e., the bounds on $\nu$, and then by the speed, i.e., the bounds on $\dot{\rho}$. In contrary, the ramp-down is always constrained by the bounds on $\nu$.

To visualize such an optimized ramp on the full-order process model, we simulate the ramp-up by using the optimized production rate trajectory (left part of Figure 7) as input to a simulation and calculate the control inputs $u$ using the backtransformation function $u = \psi_2(\rho, \dot{\rho}, \nu)$ derived above. While the first three flat output components $c_{A2}, c_{B2}, T_2$ are maintained at their nominal values, the fourth output component $c_{A1}$ follows the function of the production rate $c_{A1} = \pi_4(\rho)$ specified in the operating strategy (Figure 8). All other states and inputs are within their respective bounds. Moreover, in the first half hour, when the ramp-up is limited by the limit on the acceleration $\nu_{\max}$ (compare to Figure 7, left), $Q_1$ is close to its maximum value (Figure 8) because the upper limit of $\nu_{\max}$ is derived from $Q_{1\max}$. However, $Q_1$ does not reach its maximum value due to the conservative linear approximation. During the second half hour, the ramp-up is limited by the maximum speed $\dot{\rho}_{\max}$ (compare to Figure 7, left) and thus the control input $Q_2$, which limits the speed for high production rates $\rho$ (compare to Figure 6), is close to its bounds (Figure 8). Here, $Q_2$ comes very close to its bound as the conservative approximation of the ramping limit on $\dot{\rho}$ is very close to the true nonlinear limit (Figure 6).

Ramp optimization and corresponding simulation show that even though the dynamic ramping constraints are formed by purely linear equations, they can capture dynamics that are significantly more complex than traditional static first-order ramps. Moreover, the original process model with 6 states and 5 degrees of freedom is reduced to a dynamic ramping constraint with only 2 states, i.e., $\rho, \dot{\rho}$, and one degree of freedom $\nu$. Accordingly, coupling the flat outputs to the production rate reduces the model size and thus simplifies optimization.

4.5 Investigation 2: Demand response application

To demonstrate the dynamic ramping constraints in a DR application, the flexible process is considered together with a multi-energy system and additional non-flexible heat and electricity demands (Figure 9). Similar to our previous publication (Baader et al., 2021), we consider an energy system based on (Sass et al., 2020). However, instead of one combined heat and power plant (CHP) and one boiler (B), we extended the system to 4 CHPs and 2 boilers to study a larger energy system with more discrete on/off-decisions. Additionally, electricity can be bought from and sold to the grid for the day-ahead price that may change hourly. More detailed information about the multi-energy system and the resulting
First, the energy system operation is optimized without accounting for the energy demand of the flexible process to obtain the energy costs of the inflexible demands only. Second, the operation of the energy system is optimized with the flexible process operating in steady-state such that the demand of the flexible process is constant (compare to Figure 10, right). Third, the DR optimization is performed using dynamic ramping constraints (compare to Figure 10, left).

The DR operation of the process reduces the total energy costs by 4.7 % compared to steady-state operation. Considering only the energy costs associated with the flexible process, the cost reduction through demand response is 10.2 %.

In the resulting operation, two types of periods can be distinguished: In times of low electricity prices, heat is preferably produced by the boilers, and electricity is bought from the grid (hours 14 - 18). In times of high electricity prices, heat is preferable provided by the CHPs, and excess electricity is sold to the grid (hours 1-13, 19-24). The DR case reduces costs due to two reasons: First, the boilers are operated less. Instead of 12 hours, the boilers are only active in 9 hours (Figure 10). The amount of heat provided by the boilers is reduced by 8 %. Second, the heat demand of the flexible process and, therefore, the electricity production of the CHPs is shifted from hours of low prices to hours with higher prices (Figures 10). For instance, the heat demand is lower in hour 15 and higher in hour 5. Consequently, the derived dynamic ramping constraints allow to reduce costs substantially compared to steady-state process operation.

The optimization problem terminates after a maximum runtime of 5 minutes that we set following Harjunkoski et al. (2014). The remaining optimality gap is 4.2 %. Thus, even if the on/off-status of 6 energy system units has to be optimized simultaneously with the process operation, the optimization provides a schedule achieving substantial cost reductions within this maximum optimization runtime of 5 minutes.
5 DISCUSSION

In this section, we discuss possible adaptations and limitations of our approach.

Throughout this paper, we assume that the production rate $\rho$ is a degree of freedom (compare to assumptions in Section 3.1). However, the method can be adapted to cases where the production rate is not a degree of freedom but a component of the flat output vector $\xi_k$. For instance, if the flow rate of the product stream would be hydraulically driven by a filling level $h_{\text{fill}}$, the production rate would be given as part of the flat output by a function of $h_{\text{fill}}$. In such cases, the production rate cannot be controlled directly but only through manipulating the corresponding input $u_k$. For instance, the filling level $h_{\text{fill}}$ could be controlled by manipulating the feedflow of the corresponding unit.

In our operating strategy discussed in Section 3.2.2, all flat output components without given specifica-
Fig. 10. Resulting schedule for the flexible process performing demand response (left) and the flexible process being operated in steady-state (right). Top: electricity price. Second row: heat supplied by the 4 combined heat and power plants (CHP1 - CHP4) and the two boilers (B1, B2). The inflexible heat demand is shown as thin black line and the total heat demand as bold black line. Third row: Electricity generated by CHP1 - CHP4, sold electricity to the grid, and electricity bought from the grid. The inflexible electricity demand is shown as black line. Bottom: Production rate $\rho$.

Sections are coupled to the production rate. This coupling reduces the dimensionality of the model. Instead of $m$ flat outputs and their derivatives, only the production rate and its derivatives are variables in the optimization problem. Consequently, there are $\delta$ states and one degree of freedom. This dimensionality reduction strongly reduces the computational complexity of the optimization problem. Still, coupling all flat outputs with the production rate might be unfavorable in cases where some flat outputs can only be changed much slower than the production rate. In such cases, the outputs without specifications could be kept as independent integrator chains in the ramping model (right part in Figure 3). In our case study, the concentration $c_{A1}$, which is the fourth flat output component, could be uncoupled such that there are two integrator chains in the ramping model: one for the production rate $\rho$ and one for the concentration $c_{A1}$. Consequently, dynamic ramping constraints need to be derived for both integrator chains. While this uncoupled version makes the ramping model computationally more challenging, it is also more flexible and thus might enable higher profits in some cases. An extreme example is electrolyzers that can often adapt their production rate rapidly but have slow temperature dynamics (Simkoff and Baldea, 2020; Flamm et al., 2021). Thus, for electrolyzers, it is not favorable to couple the temperature with the production rate. Instead, we have shown recently in a conference paper that it is favorable to keep both
production rate and temperature as degrees of freedom and formulate dynamic ramping constraints on the temperature (Baader et al., 2022a).

The present work focuses on demand response applications where the production rate \( \rho \) is the main scheduling-relevant variable. However, it is straightforward to adapt the approach to cases where a different variable is scheduling-relevant. For example, in multi-product processes, the concentration may be varied to yield different products. Thus, the scheduling needs to account for the dynamics of the concentration during product transitions (Flores-Tlacuahuac and Grossmann, 2006; Baader et al., 2022b). Our approach can be transferred to such a multi-product process if the production rate \( \rho \) is replaced by the concentration.

A limitation of this work is that solving the equation system to derive the backtransformation requires a lot of manual effort, even though a computer algebra system is used as support. It is an open question to what extent our approach can be automated to enable the analysis of large-scale processes.

6 Conclusion

Dynamic ramping constraints simplify the simultaneous demand response scheduling optimization of production processes and their energy systems compared to an optimization considering the full-order process model. Still, dynamic ramping constraints can capture more flexibility than traditional static ramping constraints as they allow high-order dynamics and non-constant ramp limits. In this paper, we extend our method to rigorously derive dynamic ramping constraints from input-state linearizable single-input single-output (SISO) processes to flat multi-input multi-output (MIMO) processes. In the MIMO case, dynamic ramping constraints reduce the problem dimensionality by coupling all flat outputs with the production rate. In our case study, a system with 6 states and 5 degrees of freedom is reduced to 2 states and one degree of freedom. Additionally, we demonstrate that an operational strategy can be chosen for flat outputs such that the steady-state production points are optimal, e.g., with respect to energy consumption.

Our case study demonstrates that dynamic ramping constraints allow for finding DR schedules for flexible processes and multi-energy systems that substantially reduce energy costs compared to a steady-state operation. Even though discrete on/off-decisions in the multi-energy system add to the computational complexity, the problem can be solved within the time limit for online scheduling.

Overall, dynamic ramping constraints allow bridging the gap between nonlinear process models and simplified process representations suitable for real-time scheduling optimization.

Author contributions

Florian J. Baader: Conceptualization, Methodology, Software, Investigation, Validation, Visualization, Writing - original draft. Philipp Althaus: Conceptualization, Writing – review & editing. André Bardow: Funding acquisition, Conceptualization, Supervision, Writing – review & editing. Manuel Dahmen: Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

We have no conflict of interest.

Acknowledgements

This work was supported by the Helmholtz Association under the Joint Initiative “Energy System Integration”.
6 CONCLUSION

Nomenclature

Abbreviations
- CHP: combined heat and power plant
- CSTR: continuous stirred tank reactor
- DR: demand response
- DRC: dynamic ramping constraint
- MIDO: mixed-integer dynamic optimization
- MILP: mixed-integer linear program
- MIMO: multi-input multi-output
- MINLP: mixed-integer nonlinear program
- SISO: single-input single-output

Greek symbols
- $\alpha$: integer number (compare to assumption 3a)
- $\alpha_K$: relative volatility of component K
- $\beta$: integer number (compare to assumption 3b)
- $\gamma$: integer number (compare to assumption 3b)
- $\delta$: order of dynamic ramping constraint
- $\zeta$: integer number
- $\kappa$: integer number (compare to assumption 3a)
- $\nu$: ramping degree of freedom
- $\Xi$: transformed state vector
- $\xi$: flat output
- $\pi$: operating strategy
- $\rho$: production rate
- $\varphi_F$: density
- $\Phi$: objective
- $\phi$: transformation to flat output
- $\varphi$: ramping state vector
- $\psi$: backtransformation from flat output

Latin symbols
- $a$: coefficient
- $b$: parameter
- $C$: coverage
- $C_p$: heat capacity
- $c$: concentration
- $E$: activation energy
- $F$: flow rate
- $f$: nonlinear function
- $H$: enthalpy
- $k$: reaction constant
- $m$: number of control inputs
- $n$: number of states
- $P$: power
- $Q$: heat flow
- $R$: gas constant
- $T$: temperature
- $t$: time
- $u$: control input
- $V$: volume
- $v$: vertex
- $x$: state
- $z$: binary variable

Subscripts
- 0: feed
- 1: reactor
- 2: flash
- A: component A
- B: component B or bottom
- C: component C
- dem: demand
- el: electric
- f: final time
- p: purge
- s: steady-state
- V: vaporization

Superscripts
- max: maximum
- min: minimum
- nom: nominal
Bibliography

Adamson, R., Hobbs, M., Silcock, A., and Willis, M. J. (2017). Integrated real-time production scheduling of a multiple cryogenic air separation unit and compressor plant. *Computers & Chemical Engineering*, 104:25–37.

Adamy, J. (2014). *Nichtlineare Systeme und Regelungen*. Springer Vieweg, Berlin, Germany.

Baader, F. J., Althaus, P., Bardow, A., and Dahmen, M. (2021). Dynamic ramping for demand response of processes and energy systems based on exact linearization. *arXiv:2110.08137*.

Baader, F. J., Bardow, A., and Dahmen, M. (2022a). MILP formulation for dynamic demand response of electrolyzers. *International Symposium on Process Systems Engineering – PSE 2021+ (accepted)*.

Baader, F. J., Bardow, A., and Dahmen, M. (2022b). Simultaneous mixed-integer dynamic scheduling of processes and their energy systems. *AIChE Journal*, 2022:e17741.

Baldea, M. and Harjunkoski, I. (2014). Integrated production scheduling and process control: A systematic review. *Computers & Chemical Engineering*, 71:377–390.

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

Carrion, M. and Arroyo, J. M. (2006). A computationally efficient mixed-integer linear formulation for the thermal unit commitment problem. *IEEE Transactions on Power Systems*, 21(3):1371–1378.

Christofides, P. D., Liu, J., and La Muñoz de Peña, D. (2011). *Networked and Distributed Predictive Control*. Springer London, London, UK.

Flamm, B., Peter, C., Büchi, F. N., and Lygeros, J. (2021). Electrolyzer modeling and real-time control for optimized production of hydrogen gas. *Applied Energy*, 281:116031.

Fliess, M., Lévine, J., Martin, P., and Rouchon, P. (1995). Flatness and defect of non-linear systems: Introductory theory and examples. *International Journal of Control*, 61:13–27.

Flores-Tlacuahuac, A. and Grossmann, I. E. (2006). Simultaneous cyclic scheduling and control of a multiproduct CSTR. *Industrial & Engineering Chemistry Research*, 45(20):6698–6712.

Harjunkoski, I., Maravelias, C. T., Bongers, P., Castro, P. M., Engell, S., Grossmann, I. E., Hooker, J., Méndez, C., Sand, G., and Wassink, J. (2014). Scope for industrial applications of production scheduling models and solution methods. *Computers & Chemical Engineering*, 62:161–193.

Hoffmann, C., Hübner, J., Klaucke, F., Milojević, N., Müller, R., Neumann, M., Weigert, J., Esche, E., Hofmann, M., Repke, J.-U., Schomäcker, R., Strasser, P., and Tsatsaronis, G. (2021). Assessing the realizable flexibility potential of electrochemical processes. *Industrial & Engineering Chemistry Research*, 60(37):13637–13660.

Jogwar, S. S., Baldea, M., and Daoutidis, P. (2009). Dynamics and control of process networks with large energy recycle. *Industrial & Engineering Chemistry Research*, 48(13):6087–6097.

Leenders, L., Bahl, B., Hennen, M., and Bardow, A. (2019). Coordinating scheduling of production and utility system using a stackelberg game. *Energy*, 175:1283–1295.

Meurer, A., Smith, C. P., Paprocki, M., Čertík, O., Kirpichev, S. B., Rocklin, M., Kumar, A., Ivanov, S., Moore, J. K., Singh, S., Rathnayake, T., Vig, S., Granger, B. E., Muller, R. P., Bonazzi, F., Gupta, H., Vats, S., Johansson, F., Pedregosa, F., Curry, M. J., Terrel, A. R., Roučka, Š., Saboo, A., Fernando, I., Kulul, S., Cimrman, R., and Scopatz, A. (2017). SymPy: Symbolic computing in Python. *PeerJ Comput. Sci.*, 3:e103.
Mitra, S., Grossmann, I. E., Pinto, J. M., and Arora, N. (2012). Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Computers & Chemical Engineering*, 38:171–184.

Rothfuß, R. (1997). *Anwendung der flachheitsbasierten Analyse und Regelung nichtlinearer Mehrgrößen-systeme*. VDI-Verl., Düsseldorf, Germany.

Rothfuß, R., Rudolph, J., and Zeitz, M. (1996). Flatness based control of a nonlinear chemical reactor model. *Automatica*, 32(10):1433–1439.

Sass, S., Faulwasser, T., Hollermann, D. E., Kappatou, C. D., Sauer, D., Schütz, T., Shu, D. Y., Bardow, A., Gröll, L., Hagenmeyer, V., Müller, D., and Mitsos, A. (2020). Model compendium, data, and optimization benchmarks for sector-coupled energy systems. *Computers & Chemical Engineering*, 135:106760.

Schulze, M. and Schenkendorf, R. (2020). Robust model selection: Flatness-based optimal experimental design for a biocatalytic reaction. *Processes*, 8(2).

Schäfer, P., Daun, T. M., and Mitsos, A. (2020). Do investments in flexibility enhance sustainability? A simulative study considering the german electricity sector. *AIChE Journal*, 66(11):e17010.

Simkoff, J. M. and Baldea, M. (2020). Stochastic scheduling and control using data-driven nonlinear dynamic models: application to demand response operation of a chlor-alkali plant. *Industrial & Engineering Chemistry Research*.

Unger, J., Kröner, A., and Marquardt, W. (1995). Structural analysis of differential-algebraic equation systems—theory and applications. *Computers & Chemical Engineering*, 19(8):867–882.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, 17:261–272.

Voll, P., Klaffke, C., Hennen, M., and Bardow, A. (2013). Automated superstructure-based synthesis and optimization of distributed energy supply systems. *Energy*, 50:374–388.

Zhang, Q. and Grossmann, I. E. (2016). Planning and scheduling for industrial demand side management: Advances and challenges. In Martin, M. M., editor, *Alternative Energy Sources and Technologies*, Engineering, pages 383–414. Springer, Cham, Switzerland.

Zhou, D., Zhou, K., Zhu, L., Zhao, J., Xu, Z., Shao, Z., and Chen, X. (2017). Optimal scheduling of multiple sets of air separation units with frequent load-change operation. *Separation and Purification Technology*, 172:178–191.
1 Exemplary model for necessary flatness conditions

In this section, the two necessary conditions for flatness (compare to Section 3.2.1 in the main paper) are illustrated using an example (E) with three states and two inputs:

\[
\begin{align*}
\dot{x}_1 &= -x_1 - u_1 \quad (Ea) \\
\dot{x}_2 &= -x_2 + x_1 + u_2 \quad (Eb) \\
\dot{x}_3 &= -x_3 + 2x_1 - u_2 \quad (Ec)
\end{align*}
\]

In the example E, the output vector \( \xi = (x_1, x_2)^T \) does not fulfill the necessary graphical condition for flat outputs because the state \( x_3 \) is not covered (Figure S1, center). However, the vector \( \xi = (x_3, x_2)^T \) is a candidate for a flat output as all states are covered (Figure S1, right).

![Graph representation of example process (E)](image)

**Fig. S1.** Graph representation of example process (E) (left). While the candidate \( \xi = (x_1, x_2)^T \) does not fulfill the necessary condition for a flat output vector (center), the candidate \( \xi = (x_3, x_2)^T \) does (right).

To receive the equation system for the backtransformations, the flat output candidate \( \xi = \phi(x) = \begin{pmatrix} x_3 \\ x_2 \end{pmatrix} \) has to be differentiated two times to obtain a square backtransformation equation system:

\[
\begin{align*}
\xi &= \phi(x) = \begin{pmatrix} x_3 \\ x_2 \end{pmatrix} \\
\dot{\xi} &= \frac{\partial \phi(x)}{\partial x} \dot{x} = \frac{\partial}{\partial x} \begin{pmatrix} x_3 \\ x_2 \end{pmatrix} f(x, u) \\
\xi^{(2)} &= \frac{\partial^2 \phi(x)}{\partial x^2} \ddot{x} + \frac{\partial \dot{\phi(x)}}{\partial u} \dot{u}
\end{align*}
\]

Structurally, the backtransformation equation system (S1) - (S3) is solvable for the three states \( x_1, x_2, x_3 \), the two inputs \( u_1, u_2 \), and the input derivative \( \dot{u}_2 \) (Table S1). The input derivative \( \dot{u}_2 \) is part of the equation system as both second derivatives \( \xi^{(2)}_1, \xi^{(2)}_2 \) are functions of \( \dot{u}_2 \).

2 Illustrative example for multiple feasible regions

As discussed in Section 3.4 of the main manuscript, multiple distinct feasible regions can occur for the dynamic ramping constraints. To illustrate this phenomenon, we consider an illustrative model (IM) with
Tab. S1. Analysis of structural solvability of the backtransformation equation system to calculate states \( x \), and inputs \( u \) with flat output candidate \( \xi = (x_3, x_2)^T \) for example process (E). All states and inputs that appear in an equation are marked with a cross. A necessary condition for solvability is that it must be possible to circle exactly one cross in every row and exactly one cross in every line. In other words: Each equation should be solved for one variable.

| states | inputs and derivatives |
|--------|------------------------|
| \( x_1 \) | \( x_2 \) | \( x_3 \) | \( u_1 \) | \( u_2 \) | \( \dot{u}_2 \) |
| \( \xi_1 \) | \( \xi_2 \) | \( \checkmark \) | \( x \) | \( x \) | \( \checkmark \) |
| \( \xi_1 \) | \( \xi_2 \) | \( x \) | \( x \) | \( \checkmark \) |
| \( \xi_1^{(2)} \) | \( \checkmark \) | \( x \) | \( x \) | \( x \) | \( \checkmark \) |
| \( \xi_2^{(2)} \) | \( x \) | \( x \) | \( \checkmark \) |

three states \( x \), two inputs \( u \), and a production rate \( \rho \):

\[
\begin{align*}
\dot{x}_1 &= x_2 + \rho & \text{(IMa)} \\
\dot{x}_2 &= -x_2 + u_1 & \text{(IMb)} \\
\dot{x}_3 &= x_2^2 + u_2 & \text{(IMc)}
\end{align*}
\]

As shown in the following, this process has a flat output \( \xi = \phi(x) = \left( x_1 \ x_3 \right) \). We assume an operating strategy \( \xi = \pi(\rho) = \left( \frac{a\rho}{b\rho} \right) \) with two constants \( a, b \).

For \( x_1 \) and \( x_3 \), the back transformation is given by the function \( \pi \). For \( x_2 \), Equation (IMa) is set equal to \( \dot{x}_1 = a\dot{\rho} \) which gives

\[
\dot{x}_2 = a\dot{\rho} - \rho \quad \text{(S4)}
\]

Similarly, inserting \( \dot{x}_2 = a\rho^{(2)} - \dot{\rho} \) into Equation (IMb) gives

\[
u_1 = a\dot{\rho} - \rho + a\rho^{(2)} - \dot{\rho} \quad \text{(S5)}
\]

Finally, inserting \( \dot{x}_2 = \dot{x}_3 = b\dot{\rho} \) into Equation (IMc) gives

\[
u_2 = b\dot{\rho} - \left( \frac{a\rho - \rho}{x_2} \right)^2 \quad \text{(S6)}
\]

As all states and inputs can be given as functions of production rate \( \rho \) and derivatives \( \dot{\rho}, \rho^{(2)} \), the process is de facto flat for the given operating strategy.

Two feasible regions for the dynamic ramping constraints can result from the second input \( u_2 \) being a quadratic function of the first derivative of the production rate \( \dot{\rho} \) (Equation (S6)) (Figure S2). One could either restrict the operating range to one of the two regions or would have to introduce a binary variable that indicates which region is active.

3 PWA approximation of ramping limits

As mentioned in the main manuscript and discussed in our previous paper Baader et al. (2021), the nonlinear ramping limits need to be approximated by piecewise-affine (PWA) functions. This approximation
3 PWA APPROXIMATION OF RAMPING LIMITS

**Fig. S2.** Sketch for the input $u_2$ as function of the first derivative of the production rate $\dot{\rho}$ (compare to Equation (S6)). Potentially, there can be two feasible regions for $\dot{\rho}$, FR1 and FR2, respectively.

**Fig. S3.** Constraints for ramping degree of freedom $\nu$ as function of production rate $\rho$ for an illustrative case with first-order ramping constraints and two limiting inputs $u_1$, $u_2$. For first-order ramping constraints, the ramping state vector $\varphi$ is of dimension one and equal to the production rate $\rho$. Consequently, the limits on the ramping degree of freedom $\nu$ only depend on $\rho$. The true nonlinear limits caused by the minimum and maximum values of the two inputs $u_1$, $u_2$ are compared to static limits (dotted), linear limits (dashed) and piecewise-affine (PWA) limits (dashed-dotted).

is shown in Figure S3 for an illustrative case with first-order dynamics and two limiting inputs $u_1$, $u_2$. The upper ramping limit is only determined by the maximum input $u_1^{\text{max}}$ as the upper limit in the flat system resulting from $u_2^{\text{min}}$ is always above the limit from $u_1^{\text{max}}$. In contrast, the lower ramping limit is given by an intersection between the lower limits from $u_1^{\text{min}}$ and $u_2^{\text{max}}$, respectively. If static ramping limits are chosen, a large amount of the feasible region needs to be cut off for the illustrative case in Figure S3 (horizontal dotted lines). In contrast, linear (dashed lines in Figure S3) and piecewise-affine (dashed-dotted lines in Figure S3) limits allow to come closer to the true nonlinear limits and thus realize a larger flexibility range. For the lower ramping limit, piecewise-affine limits can be realized without the addition of binary variables as the feasible region is convex. However, for the upper limit, the feasible region is non-convex and binary variables are needed, making the optimization more computationally
challenging.

As, in the general case, the limit functions \((\rho(\gamma))^\min, (\rho(\gamma))^\max\) are multivariate functions, multivariate regression methods, e.g., hinging hyperplanes (Breiman, 1993; Adeniran and Ferik, 2017; Kämper et al., 2021), convex region surrogates (Zhang et al., 2016; Schweidtmann et al., 2021), or artificial neural networks with ReLU activation functions (Grimstad and Andersson, 2019; Lueg et al., 2021), can be used to find piece-wise affine approximations.

4 Supplementary information to case study

4.1 Model equations

In this section, we present the model equations of the case study (compare to Figure 4 in the main paper). All symbols other than states \(x = (c_{A1}, c_{B1}, T_1, c_{A2}, c_{B2}, T_2)^T\), control inputs \(u = (F_B, F_p, Q_1, Q_2)^T\), and production rate \(\rho\) are parameters marked in gray with values listed in Table S2. The index 0 indicates the feed quantities.

\[
\begin{align*}
\dot{c}_{A1} &= \frac{\rho + F_B}{V_1} (c_{A0} - c_{A1}) + \frac{F_B - F_p}{V_1} (c_{A2} - c_{A1}) - k_1 c_{A1} e^{-\frac{c_{A1}}{T_1}} \tag{S7} \\
\dot{c}_{B1} &= \frac{\rho + F_B}{V_1} (c_{B0} - c_{B1}) + \frac{F_B - F_p}{V_1} (c_{B2} - c_{B1}) + k_1 c_{A1} e^{-\frac{c_{A1}}{T_1}} - k_2 c_{B1} e^{-\frac{c_{B1}}{T_2}} \tag{S8} \\
\dot{T}_1 &= \frac{\rho + F_B}{V_1} (T_0 - T_1) + \frac{F_B - F_p}{V_1} (T_2 - T_1) - \frac{\Delta H_1}{C_p} k_1 c_{A1} e^{-\frac{c_{A1}}{T_1}} \tag{S9} + \frac{Q_1}{\varrho F C_p V_1} \\
\dot{c}_{A2} &= \frac{\rho + F_B}{V_2} (c_{A1} - c_{A2}) - \frac{\rho}{V_2} \left( \frac{\alpha_A c_{A2}}{\alpha_A c_{A2} + \alpha_B c_{B2} + \alpha_C (1 - c_{A2} - c_{B2})} - c_{A2} \right) \tag{S10} \\
\dot{c}_{B2} &= \frac{\rho + F_B}{V_2} (c_{B1} - c_{B2}) - \frac{\rho}{V_2} \left( \frac{\alpha_B c_{B2}}{\alpha_A c_{A2} + \alpha_B c_{B2} + \alpha_C (1 - c_{A2} - c_{B2})} - c_{B2} \right) \tag{S11} \\
\dot{T}_2 &= \frac{\rho + F_B}{V_2} (T_1 - T_2) - \frac{\Delta H_1}{C_p} \frac{\rho}{\varrho F C_p V_2} + \frac{Q_2}{\varrho F C_p V_2} \tag{S12}
\end{align*}
\]

4.2 Operating strategy

As discussed in the main paper, the fourth flat output \(\xi_4\) can be chosen freely. Here, we calculate the optimal steady-state operating points for \(\xi_4\) and fix \(\xi_4\) to be a function of the production rate \(\xi_4 = \pi_4(\rho)\). Accordingly, we minimize the total heat input to the process \(Q_1 + Q_2\) as the objective function to obtain the steady-state operating points. First, we determine the optimal steady-state operating points. To this end, a steady-state optimization problem \((P_s)\) is formulated by sampling the production range...
Tab. S2. Parameters of the reactor-separator process with recycle.

| symbol               | value                  |
|----------------------|------------------------|
| volume CSTR          | $V_1$ = 30 m$^3$       |
| feed concentration A | $c_{A0}$ = 1           |
| reaction constant reaction 1 | $k_1$ = $3.21 \times 10^6$  |
| activation energy reaction 1 | $E_1$ = $5 \times 10^4$  |
| gas constant         | $R = 8.314$ kJ mol$^{-1}$ K$^{-1}$ |
| feed concentration B | $c_{B0}$ = 0           |
| reaction constant reaction 2 | $k_2$ = $3.21 \times 10^6$  |
| activation energy reaction 2 | $E_2$ = $6 \times 10^4$  |
| feed temperature     | $T_0$ = 300 K          |
| reaction enthalpy 1   | $\Delta H_1$ = -261 kJ mol$^{-1}$ |
| heat capacity        | $C_p = 4.2$ kJ kg$^{-1}$ K$^{-1}$ |
| reaction enthalpy 2   | $\Delta H_2$ = -304 kJ mol$^{-1}$ |
| density              | $\varrho_F = 1000$ kg m$^{-3}$ |
| volume flash          | $V_2$ = 30 m$^3$       |
| relative volatility of A | $\alpha_A$ = 0.5  |
| relative volatility of B | $\alpha_B$ = 0.25  |
| relative volatility of C | $\alpha_C$ = 1  |
| enthalpy of vaporization | $\Delta H_V$ = $7.2 \times 10^4$ kJ mol$^{-1}$ |

$\rho_{\text{min}} \leq \rho \leq \rho_{\text{max}}$ with 21 equally distributed points $\rho_j$. The optimization problem ($P_s$) reads:

$$
\min_{\mathbf{x}_j, \mathbf{u}_j} \sum_{j=1}^{21} Q_{1,j} + Q_{2,j} \quad (P_s.a)
$$

s.t. 0 = $f(\mathbf{x}_j, \mathbf{u}_j, \rho_j)$ \quad $\forall j$ \quad ($P_s.b$)

$$
\phi_{k,j}(\mathbf{x}) = \xi_k^{\text{nom}} \quad \forall k = 1, 2, 3 \quad \forall j \quad (P_s.c)
$$

$\mathbf{u}_{\text{min}} \leq \mathbf{u}_j \leq \mathbf{u}_{\text{max}}$ \quad $\forall j$ \quad ($P_s.d$)

$\mathbf{x}_{\text{min}} \leq \mathbf{x}_j \leq \mathbf{x}_{\text{max}}$ \quad $\forall j$ \quad ($P_s.e$)

For every point $j$, the process needs to be in steady state (Equation ($P_s.b$)), the first three flat output components $\xi_k = \phi_k(\mathbf{x})$ need to be at their nominal values (Equation ($P_s.c$)), and inputs and states needs to be within bounds (Constraints ($P_s.d$), ($P_s.e$)). All bounds are given in Table S4. We implement the optimization problem $P_s$ using pyomo (Hart et al., 2017) and solve it using BARON version 20.10.16 (Khajavirad and Sahinidis, 2018) in heuristic mode.

Next, a function $\pi_4(\rho)$ needs to be chosen that does not increase the objective function ($P_s.a$) too much. In other words, the steady-state operation points reached with the additional constraint

$$
\xi_{4,j} = \pi_4(\rho_j) \quad \forall j \quad (P_s.f)
$$

should give an objective function close to the objective function reached without the additional constraint ($P_s.f$). As a first candidate, we investigate the simplest possible operating strategy, that is to hold $\xi_4$ constant, i.e., $\pi_4(\rho) = \xi_4^{\text{nom}}$. This optimization leads to a feasible solution. Accordingly, it is feasible to always hold the fourth flat output component at a constant value. However, the objective function worsens by 2.9 % compared to the first optimization without constraint ($P_s.f$) because it is not possible to choose a single constant value $\xi_4^{\text{nom}}$ that is optimal for all operating points. As a second candidate, we therefore investigate the linear strategy

$$
\xi_4 = \pi_4(\rho) = a_0^{\xi_4} + a_1^{\xi_4}\rho, \quad (S13)
$$
Tab. S3. Analysis of structural solvability with flat output candidate \( \xi = (c_{A2}, c_{B2}, T_2, c_{A1})^T \) for the reactor-separator cycle (compare to Table S1).

| states | control inputs and derivatives |
|--------|-----------------------------|
| \(\xi_1\) | \(c_{A1}\) | \(c_{B1}\) | \(T_1\) | \(c_{A2}\) | \(c_{B2}\) | \(F_B\) | \(F_p\) | \(Q_1\) | \(Q_2\) | \(\dot{F}_B\) | \(\dot{F}_p\) | \(\dot{F}^{(2)}_p\) |
| \(\xi_2\) | \(\times\) | \(\times\) | \(\times\) | \(\times\) |
| \(\xi_3\) | \(\times\) | \(\times\) | \(\times\) | \(\times\) |
| \(\xi_4\) | \(\times\) | \(\times\) | \(\times\) | \(\times\) |

\(\xi_1^{(2)}\) \(\xi_2^{(2)}\) \(\xi_3^{(2)}\) \(\xi_4^{(2)}\)

\(\xi_1^{(3)}\) \(\xi_2^{(3)}\)

Tab. S4. State and input bounds

| variable | lower bound | upper bound |
|----------|-------------|-------------|
| \(c_{A1}\) | 0 | 1 |
| \(c_{B1}\) | 0 | 1 |
| \(T_1\) | 410 K | 460 K |
| \(F_B\) | \(0 \frac{m^3}{h}\) | \(20 \frac{m^3}{h}\) |
| \(F_p\) | \(0 \frac{m^3}{h}\) | \(8 \frac{m^3}{h}\) |
| \(Q_1\) | \(0 \frac{kJ}{h}\) | \(10 \times 10^6 \frac{kJ}{h}\) |
| \(Q_2\) | \(0 \frac{kJ}{h}\) | \(4 \times 10^6 \frac{kJ}{h}\) |
| \(\rho\) | \(4.2 \frac{m^3}{h}\) | \(6.3 \frac{m^3}{h}\) |

with optimization degrees of freedom \(a_{\xi_1}^{\xi_2}, a_{\xi_1}^{\xi_3}\). The optimization finds values \(a_{\xi_1}^{\xi_2}, a_{\xi_1}^{\xi_3}\) such that the objective value worsens only 0.05% compared to the first optimization without the coupling constraint \((P_s f)\). As this linear strategy is very close to the theoretical optimum, we do not consider other functions, e.g., higher-order polynomials, and continue with the linear operating strategy \((S13)\).

4.3 Derivation of flat process reformulation

In this section, the nonlinear coordinate transformation is derived based on the nonlinear system of equations and the operating strategy. The aim is to express the states \(x = (c_{A1}, c_{B1}, T_1, c_{A2}, c_{B2}, T_2)^T\) and control inputs \(u = (F_B, F_p, Q_1, Q_2)^T\) as functions of production rate \(\rho\), and its first two derivatives \(\dot{\rho}\), and \(\nu\).

For the reformulation, first, the strategy for the 4 flat output components \(\xi = (c_{A2}, c_{B2}, T_2, c_{A1})^T\) discussed in Section 4.2 of the main manuscript is inserted:

\[ c_{A2} = c_{A2}^{\text{nom}} \] (S14)
\[ c_{B2} = c_{B2}^{\text{nom}} \] (S15)
\[ T_2 = T_2^{\text{nom}} \] (S16)
\[ c_{A1}(\rho) = a_{\xi_1}^{\xi_2} + a_{\xi_1}^{\xi_3} \rho \quad \text{(Equation (S13))} \] (S17)
While the first three outputs are chosen independent of the production rate, the fourth output \( c_{A1} \) is a function of the production rate \( \rho \).

Next, we consider the equation for the derivative of the first flat output, i.e., \( \dot{c}_1 = \dot{c}_{A2} \). As \( c_{A2} \) is held constant, the derivative \( \dot{c}_{A2} \) is zero and we solve Equation (S10) for the bottom flow \( F_B \) that is a function of the production rate \( \rho \), i.e.,

\[
F_B(\rho) = \rho \left( \frac{c_{A1}^{\text{nom}} - c_{A2}^{\text{nom}}}{c_{A1}(\rho) - c_{A2}^{\text{nom}}} - 1 \right),
\]

with \( c_{A1}^{\text{nom}} = \frac{\alpha_A c_{A2}^{\text{nom}}}{\alpha_A c_{A2}^{\text{nom}} + \alpha_B c_{B2}^{\text{nom}} + \alpha_C (1 - c_{A2}^{\text{nom}} - c_{B2}^{\text{nom}})} \),

and \( c_{A1}(\rho) \) as defined in Equation (S17).

The second output is maintained constant, too. Accordingly, we receive the condition \( \dot{c}_2 = \dot{c}_{B2} = 0 \). After inserting \( F_B(\rho) \) from Equation (S18) into Equation (S11), we get the equation:

\[
0 = \rho \left( (c_{A1}(\rho) - c_{A2}^{\text{nom}}) \left( c_{B2}^{\text{nom}} - c_{Bv}^{\text{nom}} \right) + (-c_{A2}^{\text{nom}} + c_{A1}(\rho)) (c_{B1} - c_{B2}^{\text{nom}}) \right)
\]

\[
\frac{\dot{F}_B(\rho) - F_B(\rho)}{c_{B1} - c_{B2}^{\text{nom}}}
\]

As \( \rho \) is always nonzero, we get \( c_{B1} \) as:

\[
c_{B1}(\rho) = \frac{c_{A1}(\rho) c_{B2}^{\text{nom}} - c_{A1}(\rho) c_{Bv}^{\text{nom}} + c_{A2}^{\text{nom}} c_{Bv}^{\text{nom}} - c_{A1}(\rho) c_{B2}^{\text{nom}}}{c_{A2}^{\text{nom}} - c_{A1}(\rho) c_{Bv}^{\text{nom}}}
\]

The third output component, we get the condition \( \dot{c}_3 = \dot{T}_2 = 0 \) and thus solve Equation (S12) for \( Q_2 \):

\[
Q_2(\rho, T_1) = -\theta_F C_p (\rho + F_B(\rho))(T_1 - T_2^{\text{nom}}) + H_{1\rho} \rho.
\]

The input \( Q_2 \) still depends on the temperature \( T_1 \) for which no expression has been derived up to this point.

From the fourth flat output component, we get \( \dot{c}_4 = \dot{c}_{A1} = \alpha_1^{\xi_1} \dot{\rho} \) (compare to Equation (S17)) and thus Equation (S7) changes to:

\[
\alpha_1^{\xi_1} \dot{\rho} = -c_{A1}(\rho) k_1 e^{-\frac{c_{A1}(\rho)}{k_1}} + \frac{(F_B(\rho) - F_p) (-c_{A1}(\rho) + c_{A1}(\rho))}{V_1}
\]

Equation (S22) is solved numerically for \( T_1 \). However, first, an expression for the input \( F_p \) is needed. This expression is derived from the condition that according to our operating strategy the second derivative of the second output is zero, i.e., \( c_{A1}^{(2)} = c_{B2}^{(2)} = 0 \). From this condition, we receive

\[
F_p(\rho, T_1) = \Psi_{F_p}(\rho, T_1),
\]

with a nonlinear function \( \Psi_{F_p}(\rho, T_1) \), which is not stated here to preserve readability. After inserting Equation (S23) into Equation (S22), Equation (S22) implicitly defines \( T_1 \) as a function of \( \rho \) and \( \dot{\rho} \). As discussed in the main paper, we cannot solve for \( T_1 \) analytically, but the solution received numerically is unique. With \( T_1(\rho, \dot{\rho}) \), we can calculate \( Q_2 \) as \( Q_2(\rho, \dot{\rho}) \) (compare to Equation (S21)) and \( F_p \) as \( F_p(\rho, \dot{\rho}) \) (compare to Equation (S23)).

The only variable that is missing is the reactor heating \( Q_1 \). To get an expression for \( Q_1 \), Equation (S22) is differentiated with respect to time. Before doing so, we replace \( F_p \) with \( \Psi_{F_p}(\rho, T_1) \) (compare to Equation (S23)) and insert the functions for \( F_B \) and \( c_{A1} \), which depend on \( \rho \). After bringing everything to the left hand side, Equation (S22) changes to

\[
0 = \Psi(\rho, \dot{\rho}, T_1).
\]
Building the total differential with respect to time gives

\[
0 = \frac{\partial \Psi(\rho, \dot{\rho}, T_1)}{\partial \rho} \dot{\rho} + \frac{\partial \Psi(\rho, \dot{\rho}, T_1)}{\partial \dot{\rho}} \ddot{\rho} + \frac{\partial \Psi(\rho, \dot{\rho}, T_1)}{\partial T_1} \frac{\rho + F_p}{V_1} (T_0 - T_1) + \frac{F_B - F_p}{V_1} (T_2 - T_1) - \frac{\Delta H_1}{C_p} k_1 c_{A1} e^{-\frac{A}{T_1}} - \frac{\Delta H_2}{C_p} e^{\frac{C}{C_B} 1 e^{-\frac{E}{T_1}}} + \frac{Q_1}{\varrho_F C_p V_1} \]  
(S25)

\[
\Psi_0(\rho, \dot{\rho}, T_1) + \Psi_\nu(\rho, \dot{\rho}, T_1) \nu + \Psi_1(\rho, \dot{\rho}, T_1) Q_1
\]
with \( \Psi_\nu(\rho, \dot{\rho}, T_1) = \frac{\partial \Psi(\rho, \dot{\rho}, T_1)}{\partial \dot{\rho}} \), \( \Psi_1(\rho, \dot{\rho}, T_1) = \frac{\partial \Psi(\rho, \dot{\rho}, T_1)}{\partial T_1} \frac{1}{\varrho_F C_p V_1} \),
and \( \Psi_0(\rho, \dot{\rho}, T_1) \) as the remaining part.

As \( Q_1 \) enters linearly, we can easily solve Equation (S25) for \( Q_1 \). Moreover, the ramping degree of freedom \( \nu \) enters linearly as well, which allows to solve Equation (S25) for \( \nu \) and calculate the bounds of \( \nu \) from the bounds of \( Q_1 \) (compare to Figure S4).

At this point, all states and inputs can be calculated as functions of the production rate \( \rho \) and its first two derivatives \( \dot{\rho} \) and \( \nu \).

Note that we do not derive expressions for the input derivatives \( \dot{F}_B, \dot{F}_B^{(2)} \), and \( \dot{F}_p \) that are shown in Table S3. Instead, we replace \( F_B \) with \( F_B(\rho) \) (Equation (S18)), and \( F_p \) with \( F_p(\rho, T_1) \) (Equation (S23)) in the equations where they appear and subsequently differentiate the complete equations with respect to time.

4.4 Calculating the limits for \( \dot{\rho} \)

As discussed in the previous section, the limits for \( \nu \) can be calculated in a straightforward way from Equation (S25). In this section, we calculate the limits for \( \dot{\rho} \) resulting from the bounds of \( T_1, Q_2, F_p \) (Figure 6 in the main paper). First, the bounds are calculated as a function of \( T_1 \) by dividing Equation (S22) by \( \partial \varphi^\xi \), giving \( \dot{\rho} \) as

\[
\dot{\rho} = \theta_{T_1}(\rho, T_1),
\]  
(S26)

with a nonlinear function \( \theta_{T_1} \). Note that while Equation (S22) cannot be solved for \( T_1 \) analytically, Equation (S22) can be solved analytically for \( \dot{\rho} \). Accordingly, inserting the bounds of \( T_1 \) gives the limits of \( \dot{\rho} \) shown in Figure 6 in the main paper.

In the case of \( Q_2 \), which is calculated as a function of \( Q_2(\rho, T_1) \) from Equation (S21), Equation (S21) cannot be solved for \( \dot{\rho} \) as no analytic expression was derived for \( T_1 \) that depends on \( \rho \) and \( \dot{\rho} \). Instead, Equation (S21) is solved for \( T_1 \) as \( T_1(Q_2, \rho) \), which is straightforward as \( T_1 \) enters Equation (S21) linearly. Inserting \( T_1(Q_2, \rho) \) into Equation (S26) then gives

\[
\dot{\rho} = \theta_{T_1}(\rho, T_1(Q_2, \rho)),
\]  
(S27)

which allows to calculate the limits of \( \dot{\rho} \) from the bounds of \( Q_2 \) as shown in Figure 6 in the main paper.

For the purge stream \( F_p \), we follow the same strategy to first calculate \( T_1 \) and then \( \dot{\rho} \). However, Equation (S23) cannot be solved for \( T_1 \) as function of \( F_p \) and \( \rho \) analytically due to the exponential functions (see discussion around Equation (11) in main paper). Consequently, Equation (S23) is solved numerically for \( T_1 \) as function of the bounds on \( F_p \). This value of \( T_1 \) is then used to calculate the limits of \( \dot{\rho} \) shown in Figure 6 in the main paper.

4.5 Ramping limits on \( \nu \)

In this section, the derivation of the ramping limits on the ramping degree of freedom \( \nu \) is given in detail. First, the space given by the limits \( \nu_{\text{min}}^\text{true}(\rho, \dot{\rho}), \nu_{\text{max}}^\text{true}(\rho, \dot{\rho}) \) is sampled using \( 51 \times 51 \) points. For every point, the true limits \( \nu_{\text{min}}^\text{true}(\rho, \dot{\rho}), \nu_{\text{max}}^\text{true}(\rho, \dot{\rho}) \) are calculated from \( Q_{\text{min}}, Q_{\text{max}} \) (compare to Equation (S25)) (Figure S4). For a conservative approximation of these limits, we test linear functions
4.6 Multi-energy system and optimization problem for demand response application

In this section, we describe the multi-energy system and the optimization problem in the second study, which considers a demand response application. Our aim for this case study is to construct an illustrative example that demonstrates the application and capabilities of our new method. The considered multi-energy system is based on the third day of the benchmark problem given by Sass et al. (2020). Note that the data are publicly available at Helmholtz Energy Computing Initiative (2021). The given heat and electricity demands are multiplied by 3 as we increase the number of units from 2 to 6.

Instead of one combined heat and power plant (CHP) and one boiler (B) as in Sass et al. (2020), we use 4 CHPs and 2 boilers. All 4 CHPs have a nominal thermal output power of 450 kW, and the 2 boilers have 530 kW. The nominal efficiencies are given in Table S5.

\begin{align*}
\nu_{\text{min}}(\rho, \dot{\rho}) &= a_{0}^{\text{min}} + a_{\rho}^{\text{min}} \rho + a_{\dot{\rho}}^{\text{min}} \dot{\rho}, \\
\nu_{\text{max}}(\rho, \dot{\rho}) &= a_{0}^{\text{max}} + a_{\rho}^{\text{max}} \rho + a_{\dot{\rho}}^{\text{max}} \dot{\rho},
\end{align*}

with parameters \(a_{0}^{\text{min}}, a_{\rho}^{\text{min}}, a_{\dot{\rho}}^{\text{min}}, a_{0}^{\text{max}}, a_{\rho}^{\text{max}}, a_{\dot{\rho}}^{\text{max}}\). The parameters are calculated in an optimization, which minimizes the difference to the true bounds while ensuring conservativeness such that \(\nu_{\text{true}}(\rho, \dot{\rho}) \leq \nu_{\text{min}}(\rho, \dot{\rho}) < \nu_{\text{max}}(\rho, \dot{\rho}) \leq \nu_{\text{true}}(\rho, \dot{\rho})\) holds for every point \((\rho, \dot{\rho})\).

To evaluate the quality of the bounds, we calculate the coverage \(C(\rho, \dot{\rho})\):

\[
C(\rho, \dot{\rho}) = \frac{\nu_{\text{max}}(\rho, \dot{\rho}) - \nu_{\text{min}}(\rho, \dot{\rho})}{\nu_{\text{true}}(\rho, \dot{\rho}) - \nu_{\text{true}}(\rho, \dot{\rho})},
\]

which is the distance between the chosen bounds in relation to the distance between the true bounds. The arithmetic mean of the coverage is 80%. Accordingly, the chosen linear limits capture approximately 80% of the feasible region while being computationally cheap.
The minimum part-load fraction is 50% for the CHPs and 20% for the boilers (Sass et al., 2020). Following Voll (2014), the part-load efficiency curves are discretized with one affine element to achieve an accurate discretization.

The electricity price used is the German day-ahead price from April 2nd, 2021 (SMARD Marktdaten, 2021). This price curve is chosen because it features times with low prices that make the operation of the boiler favorable and times of high prices which make the operation of the CHPs favorable.

For the optimization problem, which schedules the flexible operation of the reactor-separator process, a process energy demand model is needed (Pc). As the dynamic ramping constraint is second order, the heat demand of the process is modeled as a function of \( \rho, \dot{\rho}, \) and \( \nu \), i.e.,

\[
Q_{\text{dem, heat}}(\rho, \dot{\rho}, \nu) = \quad (S31)
\]

To fit the heat demand model, we sample the operating region given by the limits \( \rho_{\min}, \rho_{\max}, \dot{\rho}_{\min}(\rho), \dot{\rho}_{\max}(\rho), \) and \( \nu_{\min}(\rho, \dot{\rho}), \nu_{\max}(\rho, \dot{\rho}) \) using \( 11 \times 11 \times 11 \) points. First, a purely linear function is used for the heat demand and the parameters are fitted using linear regression. Because the average absolute approximation error is 10% of the nominal heat demand, we try a piecewise-affine (PWA) model to increase the accuracy. For this new PWA model, the operating region is split into segments in which affine functions are used. Here, the operating region is split into four segments by making two distinctions: If \( \dot{\rho} \) is greater or smaller than zero and if \( \nu \) is greater or smaller than zero. To implement the PWA model in the optimization, two binary variables are needed: \( z_{\dot{\rho}} \) indicates whether \( \dot{\rho} \) is positive and \( z_{\nu} \) indicates whether \( \nu \) is positive. No distinction is made for \( \rho \) as we find that the heat demand is roughly linear with \( \rho \). The PWA model reduces the mean approximation error to 5% of the nominal heat demand.

The optimization problem is discretized using orthogonal collocation with 2 elements per hour and 3 collocation points per element. The binary variables are discretized with one-hour timestep to match the hourly changing prices and demands. There are 8 binary variables per timestep: 4 on/off binaries for the CHPs, 2 on/off binaries for the boilers, and 2 binaries for the PWA heat demand model. Consequently, the optimization problem has \( 8 \times 24 = 192 \) binary variables for the 24 h time horizon. The ramping degree of freedom \( \nu \) is discretized to be piecewise linear for every hour.

The optimization problem is formulated using pyomo (Hart et al., 2017, 2011) and pyomo.dae (Nicholson et al., 2018) for discretization. The optimization problem is solved using gurobi version 8.1.0 (Gurobi Optimization, LLC, 2021). As Harjunkoski et al. (2014) state that the maximum acceptable optimization runtime for scheduling problems is typically between 5 and 20 minutes, we set the maximum optimization runtime to 5 minutes. All calculations are performed on a Windows 10 machine with an Intel(R) Core(TM) i5-8250U CPU and 24 GB RAM.

### Bibliography

Adeniran, A. A. and Ferik, S. E. (2017). Modeling and identification of nonlinear systems: A review of the multimodel approach—part 1. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(7):1149–1159.

Baader, F. J., Althaus, P., Bardow, A., and Dahmen, M. (2021). Dynamic ramping for demand response of processes and energy systems based on exact linearization. *arXiv:2110.08137*. 
Breiman, L. (1993). Hinging hyperplanes for regression, classification, and function approximation. *IEEE Transactions on Information Theory*, 39(3):999–1013.

Grimstad, B. and Andersson, H. (2019). ReLU networks as surrogate models in mixed-integer linear programs. *Computers & Chemical Engineering*, 131:106560.

Gurobi Optimization, LLC (2021). Gurobi optimizer reference manual. [http://www.gurobi.com](http://www.gurobi.com) (accessed 09 February 2021).

Harjunkoski, I., Maravelias, C. T., Bongers, P., Castro, P. M., Engell, S., Grossmann, I. E., Hooker, J., Méndez, C., Sand, G., and Wassick, J. (2014). Scope for industrial applications of production scheduling models and solution methods. *Computers & Chemical Engineering*, 62:161–193.

Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hackebeil, G. A., Nicholson, B. L., and Sirola, J. D. (2017). *Pyomo—Optimization modeling in Python*, volume 67. Springer Science & Business Media, second edition.

Hart, W. E., Watson, J.-P., and Woodruff, D. L. (2011). Pyomo: Modeling and solving mathematical programs in python. *Mathematical Programming Computation*, 3(3):219–260.

Helmholtz Energy Computing Initiative (2021). [https://jugit.fz-juelich.de/heci/energy-benchmark](https://jugit.fz-juelich.de/heci/energy-benchmark) (accessed 01 October 2021).

Kämper, A., Holtwerth, A., Leenders, L., and Bardow, A. (2021). AutoMoG 3D: Automated data-driven model generation of multi-energy systems using hinging hyperplanes. *Frontiers in Energy Research*, 9:430.

Khajavirad, A. and Sahinidis, N. V. (2018). A hybrid LP/NLP paradigm for global optimization relaxations. *Mathematical Programming Computation*, 10(3):383–421.

Lueg, L., Grimstad, B., Mitsos, A., and Schweidtmann, A. M. (2021). reluMIP: Open source tool for milp optimization of ReLU neural networks.

Nicholson, B., Siirola, J. D., Watson, J.-P., Zavala, V. M., and Biegler, L. T. (2018). Pyomo.dae: A modeling and automatic discretization framework for optimization with differential and algebraic equations. *Mathematical Programming Computation*, 10(2):187–223.

Sass, S., Faulwasser, T., Hollermann, D. E., Kappatou, C. D., Sauer, D., Schütz, T., Shu, D. Y., Bardow, A., Gröll, L., Hagemeyer, V., Müller, D., and Mitsos, A. (2020). Model compendium, data, and optimization benchmarks for sector-coupled energy systems. *Computers & Chemical Engineering*, 135:106760.

Schweidtmann, A. M., Weber, J. M., Wende, C., Netze, L., and Mitsos, A. (2021). Obey validity limits of data-driven models through topological data analysis and one-class classification. *Optimization and Engineering*, 23:855–876.

SMARD Marktdaten (2021). [https://www.smard.de/home/marktdaten](https://www.smard.de/home/marktdaten) (accessed 01 April 2021).

Voll, P. (2014). *Automated optimization-based synthesis of distributed energy supply systems*. PhD thesis, RWTH Aachen University.

Zhang, Q., Grossmann, I. E., Sundaramoorthy, A., and Pinto, J. M. (2016). Data-driven construction of convex region surrogate models. *Optimization and Engineering*, 17(2):289–332.