Received: 20 April 2020 | Revised: 16 June 2020 | Accepted: 11 July 2020

DOI: 10.1002/aic.16986

PROCESS SYSTEMS ENGINEERING

AIChE JOURNAL

Nonlinear scheduling with time-variable electricity prices using sensitivity-based truncations of wavelet transforms

Pascal Schäfer[1] | Artur M. Schweidtmann[1] | Alexander Mitsos[1,2,3]

[1]RWTH Aachen University, AVT—Aachener Verfahrenstechnik, Process Systems Engineering, Aachen, Germany

[2]JARA-ENERGY, Aachen, Germany

[3]Forschungszentrum Jülich, Energy Systems Engineering (IEK-10), Jülich, Germany

Correspondence
Alexander Mitsos, AVT Process Systems Engineering, RWTH Aachen University, 52056 Aachen, Germany.
Email: amitsos@alum.mit.edu

Funding information
Bundesministerium für Bildung und Forschung (BMBF) Open access funding enabled and organized by Projekt DEAL.

Abstract

We propose an algorithm for scheduling subject to time-variable electricity prices using nonlinear process models that enables long planning horizons with fine discretizations. The algorithm relies on a reduced-space formulation and enhances our previous work (Schäfer et al., *Comput Chem Eng*, 2020;132:106598) by a sensitivity-based refinement procedure. We therein expose the coefficients of the wavelet transform of the time series of independent process variables to the optimizer. The problem size is reduced by truncating the transform and iteratively adjusted using Lagrangian multipliers. We apply the algorithm to the scheduling of a multi-product air separation unit. The nonlinear power consumption characteristic is replaced by an artificial neural network trained on data from a rigorous model. We demonstrate that the proposed algorithm reduces the number of optimization variables by more than one order of magnitude, whilst furnishing feasible schedules with insignificant losses in objective values compared to solutions considering the full dimensionality.

KEYWORDS

air separation, artificial neural networks, demand side management, reduced-space formulation, wavelet transform

1 | INTRODUCTION

Exploiting time-variable electricity prices induced by an increasing penetration of volatile renewable electricity sources by demand side management (DSM) is nowadays recognized as an important measure to ensure profitability of large industrial consumers.[1] For this purpose, scheduling formulations using discrete-time representations have largely become the method of choice.[2] Here, the overwhelming majority of approaches relies on (mixed-integer) linear programming ([MI]LP) formulations that can be treated efficiently by state-of-the-art solvers.[3-7]

In contrast, many real processes are characterized by strongly nonlinear characteristics. Thus, authors proposed piecewise linearization approaches for capturing the nonlinearities by introducing logical disjunctions represented by binary variables, still leading to MILP formulations.[8,9] At the same time, potentials from using more sophisticated surrogate models for representing highly nonlinear relations in chemical engineering[10,11] are increasingly recognized. Consequently, the utilization of such mostly nonlinear surrogate models to accurately capture the (nonlinear) process characteristics when making scheduling decisions seems promising. However, only few authors tried a direct consideration of nonlinear process models in discrete-time scheduling so far, as this leads to (mixed-integer) nonlinear programs ([MI]NLPs) with potentially multiple suboptimal local optima. Consequently, solution of these scheduling problems either needs to be addressed by confining to local searches[12,13] that involve the risk for suboptimal choices or requires global solution approaches.[14] In the desired latter case, relevant planning horizon lengths with adequate temporal discretizations however lead to large-scale nonlinear optimization problems that are currently prohibitive for deterministic global solution methods.

In order to reduce the dimensionality when focusing on long horizons, time series aggregation has become a widely applied approach in energy systems engineering (cf. the overview given by Teichgraeber and Brandt[15] and the references therein). Essentially, these approaches represent an entire horizon comprising a large number of intervals by only few characteristic intervals, for example, to consider operational decisions during the design optimization of energy systems.[16-18] However, as classic time series aggregation does not consider the chronology of intervals, its application to systems with time-coupling constraints, such as storage or ramping limits, is problematic. In order to overcome this restriction, recent literature proposes an aggregation method relying on the identification of characteristic periods, in which the chronology is preserved,[19] finally enabling the consideration of long-term storage in design decisions.[20]

As an alternative, there are approaches that allow for a low dimensionality of the optimization problem, while explicitly considering all intervals of the original time series and thus guaranteeing constraint satisfaction for all intervals. These approaches make use of a re-assignment of degrees of freedom (DoFs), leading to tailored time grids. More precisely, the dimensionality of the optimization problem is reduced by assigning one DoF to multiple intervals. For instance, both Pineda and Morales[21] as well as Palys and Daoutidis[22] propose to cluster similar consecutive intervals. Likewise, our recent work presents an approach to assign one DoF to multiple similar intervals that do not necessarily need to be consecutive.[23] Considering the application to scheduling problems subject to time-variable electricity prices, we map one DoF to multiple intervals with similar price data, which allows for exploiting repeating price patterns in the course of the planning horizon. Moreover, we apply a wavelet-based adaptation procedure that is largely based on the work of Schlegel et al. concerning suitable control vector parameterizations in dynamic optimization,[24] to iteratively refine the temporal discretization and hence the number of DoFs. Our results indicate that one can thereby identify feasible near-optimal solution points using only a small fraction of DoFs.[23] This makes the algorithm highly advantageous when using reduced-space formulations from previous works addressing both global dynamic[25] and flowsheet optimization.[26] In particular, our recent results demonstrate that the reduction in the number of DoFs, that is, optimization variables, through the grid-adaptation thereby translates into substantial savings in computational times compared to solution approaches considering the full temporal dimensionality.[23]

Despite the successful demonstration that near-optimal solutions can be furnished with only few DoFs actually used, the following issues remain:

1. The refinement procedure needs several restrictions limiting the possibilities for inserting additional grid points in one iteration in order to guarantee that the derivation of the assignment of DoFs to intervals is unambiguous.
2. Furthermore, the algorithm makes use of a systematic yet heuristic procedure for introducing additional DoFs. Consequently, there is no quantitative measure for the expected improvement in the objective value through the refinement step providing a sound basis for decision-making.
3. As a consequence of points (1) and (2), the algorithm's performance can be poor in cases where the optimal schedule is strongly governed by ramping constraints, that is, too many iterations are required to identify promising schedules that involve the introduction of actually negligible DoFs and thus unnecessarily prolong computational times.

Therefore, we herein present a further development of the algorithm resolving issues (1)–(3) in case of scheduling problems comprising only continuous DoFs. For this purpose, we perform optimizations directly in the space of coefficients of the wavelet transform of the time series of DoFs instead of the space of permuted DoFs, making the explicit derivation of a temporal discretization superfluous (issue (1)). Thereby, dimensionality reductions do no longer correspond to assigning one DoF to multiple intervals but rather turn into a truncation of the wavelet transform by setting wavelet coefficients of unused wavelet basis functions to zero, as is widely done in image compression.[27] This procedure allows for calculating the marginal effect of relaxing the constraints truncating the wavelet transform, that is, the Lagrangian multipliers. Hence, it provides the desired quantitative measure for the improvements in the objective value from introducing additional optimization variables and enables decisions based on sensitivity information instead of pure heuristics (issue (2)). Using this for iterative refinement steps finally allows for determining the most important characteristics of the solution and overcomes issue (3). We remark that the proposed refinement strategy using Lagrangian multipliers shows conceptual similarities with approaches for column generation that have recently also gained attention in the context of long-term operational decisions[28-30]; these approaches rely on initially exposing only a restricted number of variables to the optimizer and then iteratively increase this number based on Lagrangian multipliers as well.

The remainder of this article is structured as follows: in the next section, we briefly review key concepts from our previous work, which are required for the proposed algorithm. Afterwards, we introduce both the approach for dimensionality reduction by truncating the wavelet transforms and for sensitivity-based refinements using Lagrangian multipliers. The efficacy of the proposed algorithm is assessed for a case study, which we introduce in the fifth section. Therein, we address the production planning subject to time-variable electricity prices for an industrial-scale air separation unit (ASU), which was designed for increased flexibility in previous work.[31] We utilize high-fidelity power consumption data from steady-state optimizations of a rigorous process model. The nonlinear power consumption characteristic is incorporated into the scheduling problem by embedding artificial neural networks (ANNs) as powerful surrogate models into the optimization.[11] In the sixth section, we finally present the results of our computational study and analyze the convergence behavior of the algorithm with increasing number of optimization variables as well as resulting savings in computational times.

2 | PRELIMINARIES

In the following, we focus on the solution of the generic reduced-space scheduling problem (1a)–(1d). In contrast to our previous work,[23] we confine to problems comprising only continuous optimization variables. Note that this limitation is not as restrictive as it appears: in fact, binary variables often solely stem from piecewise linearizations of nonlinear functions.[9,32] Consequently, we emphasize that a substantial ratio of the scheduling problems considered in the relevant literature could still be formulated equivalently to Equations (1a)–(1d).

$$\min_{d_{i,t}} \ \Phi(d_{i,1},\ldots,d_{i,T}) = \sum_{t=1}^{T} \phi_t(d_{i,t}) \tag{1a}$$

$$\text{s.t.} \ \ 0 \le g_{l,t}(d_{i,1},\ldots,d_{i,t}), \ \ \forall l,t \tag{1b}$$

$$0 = h_m(d_{i,1},\ldots,d_{i,T}), \ \ \forall m \tag{1c}$$

$$d_i^l \le d_{i,t} \le d_i^u, \ \ \forall i,t \tag{1d}$$

In Equations (1a)–(1d), the optimal production schedule is identified by minimizing the objective function $\Phi(\cdot)$, e.g., cost, with respect to the optimization variables, that is, the DoFs $d_{i,\,t}$, which denotes the values of the independent process variable d_i in time interval $t \in \{1, \ldots, T\}$. Furthermore, both inequality and equality constraints denoted by $g_l(\cdot)$ and $h_m(\cdot)$, respectively need to be satisfied. The assumptions concerning the structure of Equations (1a)–(1d) (e.g., objective is a summation of interval-specific, generally multivariate functions, inequalities have to hold in every time step, etc.) are not necessarily required for the application of the algorithm, but rather match common scheduling formulations involving operating limits that have to be obeyed in each interval as well as cumulative production targets. Note that in a reduced-space formulation, $\Phi(\cdot)$, $g_{l,\,t}(\cdot)$, and $h_m(\cdot)$ are explicit functions of the optimization variables $d_{i,\,t}$.

Without loss of generality, we confine to cases with horizons lengths of $T = 2^N$, $N \in \{1, 2, \ldots\}$ for the ease of presentation. Workarounds for different horizons lengths can be found in our previous work.[23] Moreover, we introduce a permutation matrix $\mathbf{P} \in \mathbb{R}^{T \times T}$ that orders the time series of DoFs according to the electricity prices, that is,

$$\mathbf{P}(d_{i,1},\ldots,d_{i,T})^T = \left(\tilde{d}_{i,1},\ldots,\tilde{d}_{i,T}\right)^T \Leftrightarrow (d_{i,1},\ldots,d_{i,T})^T = \mathbf{P}^{-1}\left(\tilde{d}_{i,1},\ldots,\tilde{d}_{i,T}\right)^T \tag{2}$$

with $\tilde{d}_{i,1}$ denoting the value of the independent process variable d_i in the interval with highest electricity price. As a result, the series $\tilde{d}_{i,1},\ldots,\tilde{d}_{i,T}$ is likely characterized by less frequent fluctuations than the original, that is, chronological, time series $d_{i,\,1}, \ldots, d_{i,\,T}$.

Essentially, the permutation procedure above is the key to substantial reductions in the dimensionality of the scheduling problem (1a)–(1d) as it allows for low-dimensional yet accurate approximations of the series of permuted DoFs $\tilde{d}_{i,1},\ldots,\tilde{d}_{i,T}$ by neglecting numerous insignificant fluctuation patterns, that is, differences between intervals with similar prices. For a systematic treatment, we further perform a wavelet transform in the Haar basis,[33] which we denote by $\mathcal{W}_{\psi_{\text{Haar}}}(\cdot)$. Thereby, we represent the entire series $\tilde{d}_{i,1},\ldots,\tilde{d}_{i,T}$ unambiguously by T coefficients $\lambda_{i,\,a,\,b}$ with $a \in \{-1, \ldots, N-1\}$ and $b \in \{0, \ldots, 2^a - 1\}$, which we summarize in the vector $\boldsymbol{\lambda}_i$. The coefficient $\lambda_{i,\,-1,\,0}$ corresponds to the scaled mean of the transformed series, the other $T-1$ coefficients correspond to scaled amplitudes of shifted square waves.

3 | DIMENSIONALITY REDUCTION BY TRUNCATED WAVELET TRANSFORMS

In our previous work,[23] we used the information contained in the coefficients of the wavelet transform to iteratively derive a tailored temporal discretization for the scheduling problem, assigning one DoF to multiple intervals. In contrast, we herein propose an alternative procedure making the derivation of a temporal discretization superfluous. For this purpose, we insert both the inverse Haar-wavelet transform (3) and the permutation (2) into the optimization problem (1a)–(1d).

$$\left(\tilde{d}_{i,1},\ldots,\tilde{d}_{i,T}\right)^T = \mathcal{W}_{\psi_{\text{Haar}}}{}^{-1}(\lambda_i) \tag{3}$$

Following the described procedure allows for an equivalent reduced-space reformulation of (1a)–(1d) exposing the coefficients $\boldsymbol{\lambda}_i$ as only variables to the optimizer (cf. Equations (4a)–(4d)), that is, we formulate the objective (4a) as well as all constraints (4b)–(4c) as explicit functions $\Phi^{\psi}(\cdot)$, $g^{\psi}_{l,t}(\cdot)$, and $h^{\psi}_m(\cdot)$ of the coefficient vector $\boldsymbol{\lambda}_i$.

$$\min_{\lambda_i} \ \Phi^{\psi}(\lambda_i) \tag{4a}$$

$$\text{s.t.} \ \ 0 \le g^{\psi}_{l,t}(\lambda_i), \ \ \forall l,t \tag{4b}$$

$$0 = h^{\psi}_m(\lambda_i), \ \ \forall m \tag{4c}$$

$$\begin{pmatrix} d_i^l \\ \vdots \\ d_i^l \end{pmatrix} \le \mathcal{W}_{\psi_{\text{Haar}}}{}^{-1}(\lambda_i) \le \begin{pmatrix} d_i^u \\ \vdots \\ d_i^u \end{pmatrix}, \ \ \forall i \tag{4d}$$

As the inverse Haar-wavelet transform $\mathcal{W}_{\psi_{\text{Haar}}}{}^{-1}(\cdot)$ corresponds to a matrix multiplication by definition, that is, each value of the series $\tilde{d}_{i,1},\ldots,\tilde{d}_{i,T}$ is a linear combination of 2^N+1 coefficients, we do not expect a substantial deterioration of the computational performance of optimization algorithms if solving Equations (4a)–(4d) instead of Equations (1a)–(1d). Moreover, optimizations in the space of wavelet coefficients enable an exploitation of their physical meaning described above. In particular, summation constraints

on the time series $d_{i,1}, \ldots, d_{i,T}$ that are commonly used to guarantee reaching desired production targets in the scheduling problem can be equivalently reformulated by fixing the coefficient $\lambda_{i,-1,0}$:

$$\sum_{t=1}^{T} d_{i,t} = \gamma \Leftrightarrow \lambda_{i,-1,0} = \frac{\sqrt{2}^{\lfloor \log_2 T \rfloor} \cdot \gamma}{T}.$$

Using the scheduling formulation with the wavelet coefficients as optimization variables (4a)–(4d), dimensionality reductions correspond to truncations of the wavelet transform, essentially taking up a key idea from image compression using wavelet transforms.[27] Consequently, we only allow a subset of all coefficients denoted by U_i to be nonzero. Coefficients $\lambda_{i,a,b}$, $(a, b) \notin U_i$ are set to zero and thereby no longer treated as optimization variables, adding Equation (5) to the scheduling problem (4a)–(4d).

$$\lambda_{i,a,b} = 0, \quad \forall i, (a,b) \notin U_i \tag{5}$$

We finally highlight that for any U_i, any feasible point of the scheduling problem (4a)–(5) becomes a feasible point of the original problem (1a)–(1d) if transformed back using Equations (3) and (2).

4 | SENSITIVITY-BASED REFINEMENT STRATEGY

In this work, we target the development of an algorithm for the efficient iterative adaptation of the set of used, that is, not-truncated, wavelet coefficients U_i, making the solution of Equations (4a)–(5) converge to the solution of Equations (1a)–(1d) with a substantially smaller number of optimization variables. For this purpose, we herein propose to decide on adding a specific truncated pair (a, b) to the set U_i based on the marginal effect of relaxing the constraints (5) (cf. Figure 1), thus exploiting the relation:

$$\left.\frac{\partial \Phi^{w,*}}{\partial \lambda_{i,a,b}}\right|_{\lambda_{i,a,b}=0} = \mu_{i,a,b}, \quad \forall i, (a,b) \notin U_i \tag{6}$$

with $\Phi^{w,*}$ being at least a local and preferably the global minimizer of Equations (4a)–(5) and $\mu_{i,a,b}$ denoting the Lagrangian multipliers of Equation (5). Following this interpretation, we add those truncated pairs (a, b) to U_i where the (local) sensitivity $|\mu_{i,a,b}|$ of the objective with respect to the value of the coefficient are largest. Thereby, Equation (6) provides a quantitative measure for comparing the benefits in terms of objective value from adding different coefficients to the wavelet representation, thus overcoming an important weakness of our previous work as discussed in the introduction.

As a potential further development of the algorithm in the future, it might be promising to investigate benefits from using higher-order sensitivity information, such as a quadratic approximation of the improvements in the objective function. Along these lines, there would also arise opportunities for an optimization-based selection of truncated pairs (a, b) to add to U_i. That is, having an accurate approximation would allow to construct subproblems for determining the truncated pairs (a, b), whose utilization would lead to the maximum improvement in the objective function, which could improve the convergence behavior.

Considering decisions on deletion of insignificant optimization variables, that is, determination of pairs (a, b) to be truncated, we continue to apply the criterion proposed by Schlegel et al.[24] and used also in our previous work.[23] That is, we continue to decide on deletions by comparing the absolute values of used coefficients to a threshold ε_i^d. Thus, if $|\lambda_{i,a,b}| < \varepsilon_i^d$, the pair (a, b) is removed from U_i. Note that this

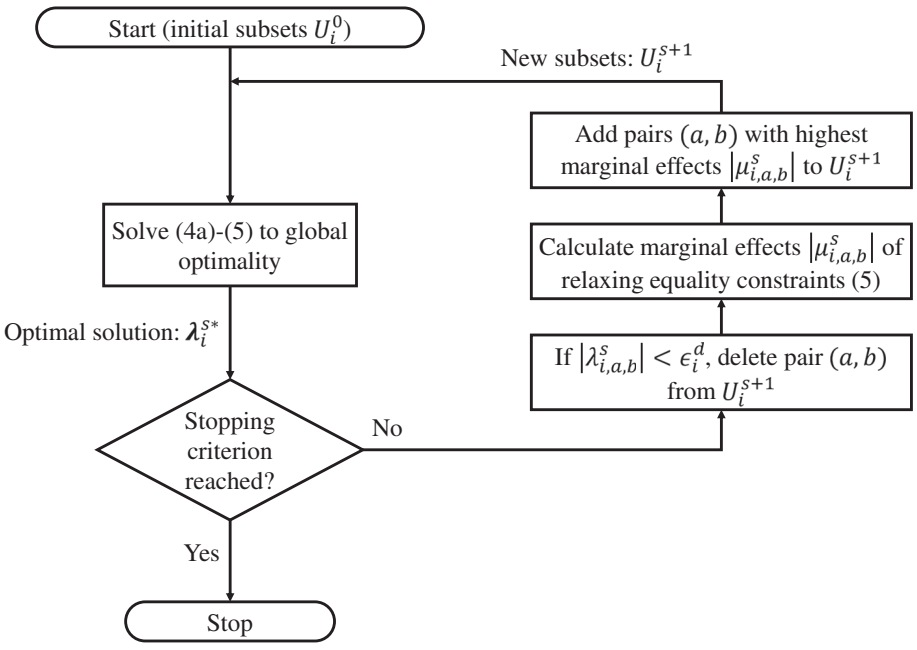

FIGURE 1 Flowchart of the proposed algorithm for sensitivity-based refinements of truncated wavelet transforms. The superscript s in the figure indicates the iteration count

criterion is already based on sound considerations. In particular, using a threshold value for determining pairs (a, b) to be truncated is reasoned by the scaling of wavelet basis functions that ensures a norm equality[34]:

$$\left\| \left(\tilde{d}_{i,1}, \ldots, \tilde{d}_{i,T} \right)^T \right\|_2 = \|\lambda_i\|_2,$$

which implies that truncating small absolute wavelet coefficients has only minor effects on the time series of process variables and thereby will also only insignificantly impair objective values.

We emphasize that in contrast to our previous work,[23] the refinement procedure does not need to obey certain rules; instead, any refinement step is possible at any time. In particular, this enables the use of wavelet coefficients $\lambda_{i,\,a,\,b}$ of a high level of detail, that is, large a, without previously exploring all levels below. Thereby, we are able to circumvent additional iterations of the algorithm and avoid the introduction of insignificant DoFs, that otherwise unnecessarily prolong the computational times for identification of a suitable discretization. Anticipating the presentation of the results below, this allows for an efficient treatment of cases, in which optimal schedules are strongly governed by active ramping constraints.

Finally, the algorithm relies on heuristic stopping criteria for terminating the refinement procedure, such as a threshold on the relative improvement in objective values between consecutive iterations in the previous work[23] or simply a maximum number of used optimization variables. We emphasize that these criteria are required in general to indicate that sufficiently near-optimal solutions have been furnished. Note that this is reasoned by the absence of global lower bounds, that is, from solving relaxations of the problem considering the full temporal dimensionality, in the general case and thus does not involve a certificate of optimality.

5 | AIR SEPARATION CASE STUDY

Due to their high electricity consumption as well as inherent possibilities for liquid product storage, DSM-related scheduling of cryogenic ASUs is an important research field.[2] In this section, we first briefly characterize the considered process configuration. Afterwards, we present the mathematical model as well as the scheduling formulation and briefly discuss them in the context of the relevant literature.

5.1 | Process description

We focus on an industrial-scale ASU producing both high-purity liquid nitrogen (LIN) and oxygen (LOX) that is stored in buffer tanks and distributed to off-site customers. The considered process configuration has been tailored to a high load flexibility and is described in detail our in previous works.[31,35] Most importantly, the process configuration comprises an integrated liquefaction cycle enabling a high ratio of storable liquid products. Moreover, the configuration allows for liquid-assisted operation, where stored cryogenic LIN is fed back to the rectification section. Thereby, the internal column flows and the separation performance can be maintained in the desired operating range, although substantially reducing the power consumption of the process and hence the liquefaction capacity.

The resulting wide operating range makes the considered process configuration on the one hand promising for performing DSM with profitable capabilities for the exploitation of spot electricity market price spreads on short to medium time-scales (e.g., between day and night or between working days and weekends). However, on the other hand, the turbomachinery applied in ASUs is commonly characterized by narrow low-loss operating ranges and significant off-design efficiency losses if leaving these ranges, resulting in nonlinear process characteristics (cf. Figure 2). Consequently, scheduling of the considered process has to be conducted carefully by accounting for its nonlinearities as precisely as possible.

5.2 | Scheduling model

We herein consider scheduling of the described ASU subject to time-variable hourly day-ahead spot market prices. We assume that all electricity is purchased at the day-ahead market and that prices for the entire planning horizon are known. Price data provided in the Supporting Information is retrieved from EPEX Spot SE (www.epexspot.com, accessed Apr 2020) and corresponds to German day-ahead prices from end of September/beginning of October 2018. Having fixed cumulative production targets, the objective function to be minimized $\Phi(\cdot)$ corresponds to the cost for electricity purchase, in accordance with the formulation in Equation (1a), given by:

$$\Phi(\mathrm{LIN}_t, \mathrm{LOX}_t) = \sum_{t=1}^{T} \phi_t(\mathrm{LIN}_t, \mathrm{LOX}_t) \quad \text{with} \quad \phi_t(\mathrm{LIN}_t, \mathrm{LOX}_t) = c_t \cdot P(\mathrm{LIN}_t, \mathrm{LOX}_t)$$

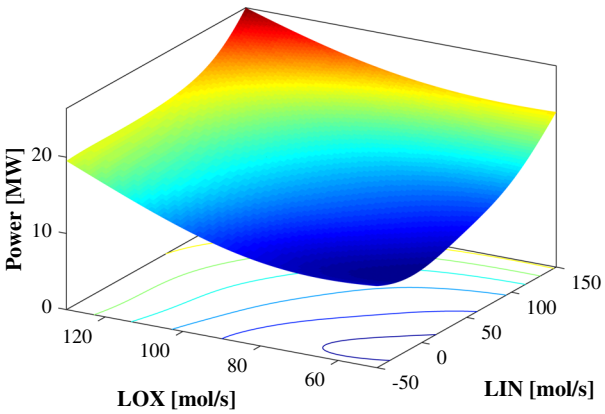

FIGURE 2 Operating range and power consumption characteristics of the considered ASU configuration comprising an integrated liquefaction cycle and allowing for liquid-assisted operation. The power consumption is calculated using an artificial neural network fitted to optimized steady-state operating points using a rigorous process model from literature.[31] ASU, air separation unit [Color figure can be viewed at wileyonlinelibrary.com]

where c_t denotes the day-ahead price for hour t and $P(\cdot)$ the instantaneous power consumption given as a function of the production rates LIN_t and LOX_t. Note that throughout the manuscript, we omit the multiplication with the interval length Δt = 1 hr for readability. Furthermore, we impose operating limits on the two production rates:

$$-50\ \mathrm{mol/s} \le \mathrm{LIN}_t \le 150\ \mathrm{mol/s},\quad \forall t$$
$$50\ \mathrm{mol/s} \le \mathrm{LOX}_t \le 130\ \mathrm{mol/s},\quad \forall t.$$

Note that in the inequalities above, the negative lower bound for the LIN production corresponds to the aforementioned possibility to feedback stored LIN from the tank. In order to ensure reaching the production targets, we require the average production rates to equal the nominal production of $\mathrm{LIN}_{\mathrm{nom}} = \mathrm{LOX}_{\mathrm{nom}} = 120$ mol/s in each case, thus demanding:

$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{LIN}_t = \mathrm{LIN}_{\mathrm{nom}}$$
$$\frac{1}{T}\sum_{t=1}^{T}\mathrm{LOX}_t = \mathrm{LOX}_{\mathrm{nom}}.$$

The power consumption of the process $P_t(\mathrm{LIN}_t, \mathrm{LOX}_t)$ is calculated using an ANN as nonlinear surrogate model resulting in the surface given in Figure 2. Note that this surface is nonconvex, motivating the use of deterministic global optimization for solving the scheduling problems in the following. In particular, one finds univariate functions (fixed LIN and varying LOX and vice versa) with flat plateaus comprising inflection points and partially even with local minima. For generating the training data, we herein used steady-state optimizations of the rigorous dynamic process model from our previous work,[31] extended by the consideration of off-design efficiency losses in the turbomachinery. The ANN is fitted to 184 optimized stationary operating points, which are equally distributed across the entire operating range, using the MATLAB R2019a Deep Learning Toolbox (MathWorks, Inc.) and comprises one hidden layer with six neurons and hyperbolic tangent activation functions. The corresponding MATLAB object including all weights and biases as well as the training data set is provided in the Supporting Information. The ANN fits the training data with a maximum relative error below 2% and a mean percentage error of ~0.5%. The use of a nonlinear surrogate model herein corresponds to an advancement compared to the vast majority of previous literature that either use simplified linear characteristics,[4-6,36] which potentially introduces significant approximation errors, or involve the identification of piecewise-linearized surrogate models,[8,37-39] which potentially requires large numbers of disjunctions for being accurate.

We further assume that the process can be operated in a continuous way within its operating range. In this case, we do not need additional binary variables indicating active modes. In particular, we thereby also omit possibilities for temporary shutdowns of the entire plant, which have been considered in most of the referenced works. However, in case of the herein considered process configuration, opportunities for temporary shut-downs are of limited additional economic potential considering the existing possibilities for reductions of the power consumption by >50% (cf. Figure 2) on the one hand and large durations with off-spec production during shut-down or start-up periods on the other hand.[40]

As ASUs are characterized by large liquid volumes[41] that lead to time constants in a similar order as time scales of spot electricity markets,[13] process dynamics further need to be accounted for in the scheduling.[42] For this purpose, we impose constraints on the change in production rates, as widely done in scheduling practice to ensure that load changes can be tracked without violating product requirements.[6,8,39] Revisiting the results presented in our previous work concerning the process dynamics,[35] the following constraints appear suitable, so that common industrially applied control systems would be able to track all scheduled set-point changes.

$$-15\ \mathrm{mol/s} \le \mathrm{LIN}_t - \mathrm{LIN}_{t-1} \le 15\ \mathrm{mol/s},\quad \forall t$$
$$-15\ \mathrm{mol/s} \le \mathrm{LOX}_t - \mathrm{LOX}_{t-1} \le 15\ \mathrm{mol/s},\quad \forall t.$$

Note that changing the control system to a more advanced one, such as nonlinear model predictive control, could involve substantially loosened ramping limits[43,44] and thereby increase potential savings, which should be thoroughly investigated in the future. Moreover, we acknowledge that numerous works argue to construct dynamic surrogate models[13,45,46] for explicitly capturing the process dynamics and demonstrate their successful application to the scheduling of ASUs.[9,47,48] This discussion is however out of the scope of this manuscript. Nevertheless, we emphasize that the presented approach is conceptually also applicable to scheduling problems involving dynamic process models as will be briefly discussed in the conclusion, but leave this for future work.

Finally, we consider that the storable amount of product is limited. In particular, we restrict the cumulative over-/underproduction, so that the nominal production can be kept up for not more than 12 hr by using the available storage capacities, thus requiring:

$$-6\cdot LIN_{\mathrm{nom}} \le \sum_{\tau=1}^{t}(\mathrm{LIN}_\tau - \mathrm{LIN}_{\mathrm{nom}}) \le 6\cdot\mathrm{LIN}_{\mathrm{nom}},\quad \forall t$$
$$-6\cdot\mathrm{LOX}_{\mathrm{nom}} \le \sum_{\tau=1}^{t}(\mathrm{LOX}_\tau - \mathrm{LOX}_{\mathrm{nom}}) \le 6\cdot\mathrm{LOX}_{\mathrm{nom}},\quad \forall t.$$

The entire scheduling problem can be formulated equivalently to Equations (1a)–(1d) by selecting $d_{1,\,t} = \mathrm{LIN}_t$ and $d_{2,\,t} = \mathrm{LOX}_t$, making the proposed algorithm applicable to this problem. Consequently, we expose the coefficients $\lambda_{1,\,a,\,b}$ and $\lambda_{2,\,a,\,b}$ of the wavelet transforms of the permuted time series of production rates as optimization variables to the solver. The objective and every presented constraint are formulated as explicit functions thereof by using Equations (3) and (2). Note that according to discussion about the physical meaning of the wavelet coefficients, the constraints that ensure reaching the production

target can be reformulated fixing the first wavelet coefficient to $\lambda_{1,-1,0} = \sqrt{2}^{\log_2 T} \cdot T \cdot \mathrm{LIN}_{\mathrm{nom}}$ and $\lambda_{1,-1,0} = \sqrt{2}^{\log_2 T} \cdot T \cdot \mathrm{LOX}_{\mathrm{nom}}$.

6 | PERFORMANCE ASSESSMENT

The reduced-space scheduling problem is implemented and solved using our in-house open-source software for deterministic global optimization MAiNGO[49] based on McCormick relaxations.[25,50,51] The reduced-space implementation substantially reduces the number of optimization variables when using ANNs, leading to significant improvements in the computational performance.[11] The ANNs functions have been made accessible using the MeLOn—Machine Learning models for Optimization toolbox (`git.rwth-aachen.de/avt.svt/public/MeLOn`).

We apply a relative optimality tolerance of 2% in accordance with the maximum regression error of the ANN. Furthermore, we impose a CPU time limit of 100,000 s. In order to get the values of the Lagrangian multipliers of the constraints (5), we perform an additional a posteriori local search starting from the best feasible point using IPOPT.[52] All calculation are conducted on one core of an Intel Xeon CPU E5-2630 v3 at 2.40 GHz with 192 GB RAM.

Moreover, we apply threshold values $\varepsilon_1^d = \lambda_{1,-1,0} \cdot 10^{-3}$ and $\varepsilon_2^d = \lambda_{2,-1,0} \cdot 10^{-3}$ respectively for truncating a coefficient of the wavelet transform in the optimization problem in the next iteration. Afterwards, we add those pairs (a, b) of truncated wavelet coefficient to the set U_i for the next iteration, where the marginal effect of relaxing Equation (5), that is, the absolute of the Lagrangian multiplier, is highest. This is done until the total number of optimization variables between two iteration is increased by four. Note that this procedure is primarily chosen for illustration purposes as it allows for displaying the effects of adding additional optimization variables in a high resolution. With regard to a practical application of the algorithm, other criteria are possible as well as discussed previously.

6.1 | Convergence of the algorithm

In this section, we analyze the course of objective and optimal variable values over the iterations of the algorithm. Note that global deterministic solutions when considering the full dimensionality, that is, using an untruncated wavelet transform, are computationally intractable. Thus, a stochastic global approach relying on multi-start local searches is applied to obtain benchmark values for using the untruncated transforms. This choice is justified by the good performance of local searches with IPOPT performed during pre-processing in MAiNGO, that is, no improvements of the upper bounds have been observed within the CPU time limit in any case. Consequently, we herein scale all objective values for comparison purpose between the benchmark from multi-start local optimizations using the untruncated transform (0%) and the objective value of a constant production using only the fixed coefficients $\lambda_{1,-1,0}$ and $\lambda_{2,-1,0}$ (100%). Consequently, the scaled objective values indicate which percentage of the maximum achievable cost savings are discarded by truncating the wavelet transform.

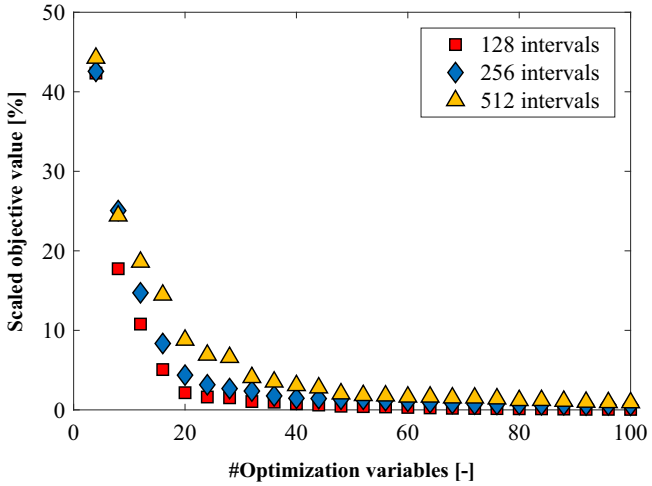

FIGURE 3 Optimal objective value in the course of the algorithm as a function of the number of variables exposed to the solver in the respective iterations. Values are scaled between a constant operation (100%) and the benchmark from multi-start local optimization when using the untruncated wavelet transform (0%). Symbols indicate varying horizon lengths. Squares (□): horizon length of 128 hourly intervals, diamonds (◇): 256 hourly intervals, triangles (△): 512 hourly intervals [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Comparison of the minimum number of required optimization variables for reaching a desired scaled objective value in case of varying horizon lengths

	#Opt. variables for desired objective value (–)			
#Intervals	10%	5%	2%	1%
128	13	17	22	35
256	15	20	35	62
512	20	31	49	92

In Figure 3, we show the course of the scaled objective value over the iterations of the algorithm for the different horizon lengths of T = 128, 256, 512 hourly intervals. The algorithm furnishes a series of feasible solutions with objective values that decrease monotonically and converge towards the benchmark values from multi-start local optimizations, that is, the presumable global optima using the untruncated wavelet transform, for all considered horizon lengths. Moreover, it can be seen that high percentages of the maximum achievable cost savings can be reached even when using substantially truncated wavelet transforms.

Table 1 further gives a comparison of the minimum numbers of optimization variables that are required for reaching desired percentages of the maximum savings. Thereby, we illustrate how the algorithm circumvents a linear scaling in the number of optimization variables with the number of considered scheduling intervals by confining to low-dimensional yet accurate approximations that suffice for exploiting large proportions of the saving potentials. For instance, if

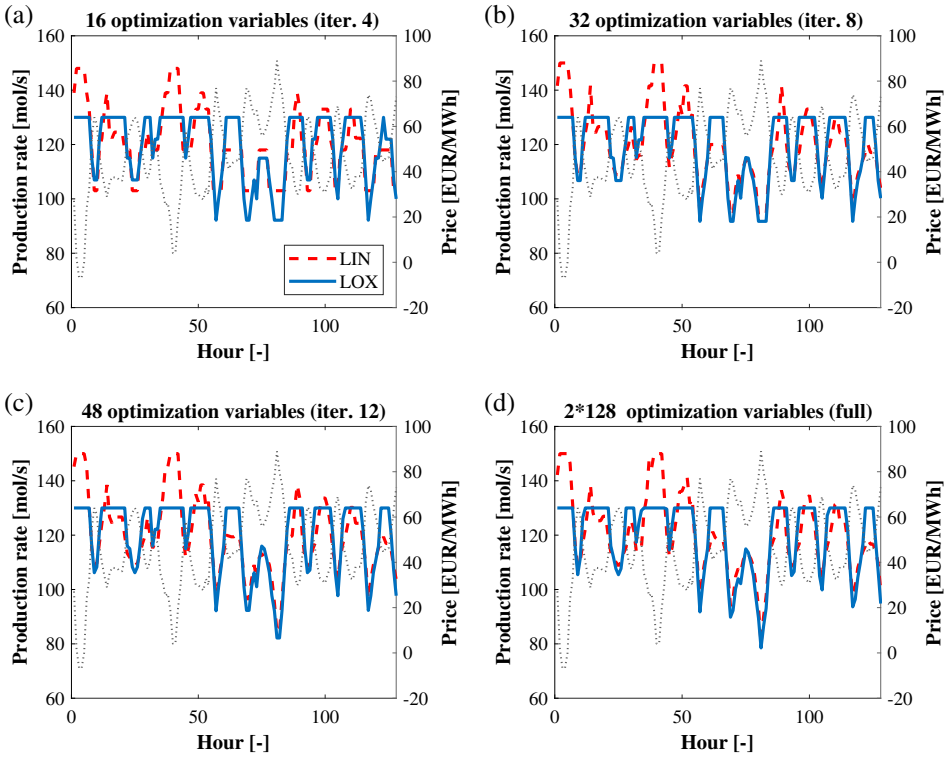

FIGURE 4 Optimized production schedules after selected iterations of the algorithm. (a): after the fourth iteration using 16 optimization variables, (b): after the eighth iteration using 32 optimization variables, (c): after the twelfth iteration using 48 optimization variables, and (d): from multi-start local solutions considering the full time series, that is, using an untruncated wavelet transform with 256 optimization variables. Dashed lines depict the LIN production, solid lines correspond to LOX. The thin dotted lines give the electricity price for comparison. LIN, liquid nitrogen; LOX, liquid oxygen [Color figure can be viewed at wileyonlinelibrary.com]

comparing the 128 hr scheduling horizon to the 512 hr, that is, four times as many intervals, we find that only ∼50% more optimization variables are required if confining to 90% of the maximum savings. Likewise, ∼80% more optimization variables are required for 95% of the maximum savings and 2.2 times as many for 98%. Even if demanding the exploitation 99% of the saving potentials, 2.6 and thus still substantially less than four times as many optimization variables are required.

In addition to the discussion of objective values, the courses of the optimal variable values over the iterations of the algorithm, that is, the optimized production schedules, are given in Figure 4. It can be seen that the algorithm furnishes feasible solutions that converge to the benchmark time series considering the full dimensionality, that is, an optimal untruncated wavelet transform, even in cases where the optimum is strongly governed by active ramping constraints. Furthermore, we notice that only a very limited number of nonzero wavelet coefficients, that is, optimization variables, suffices for a detailed reproduction of the most important fluctuations of the optimum when considering the full dimensionality. In particular, the iterations in Figure 4 correspond to reductions in the number of optimization variables compared to an untruncated wavelet transform (d) of 93.75% (a), 87.5% (b), and 81.25% (c), respectively, demonstrating the ability of accurately reproducing large and strongly fluctuating time series from only few important wavelet coefficients.

Finally, in Figure 5, we compare the convergence behavior of our proposed sensitivity-based refinement strategy to the systematic yet heuristic procedure from our previous work[23] for a horizon

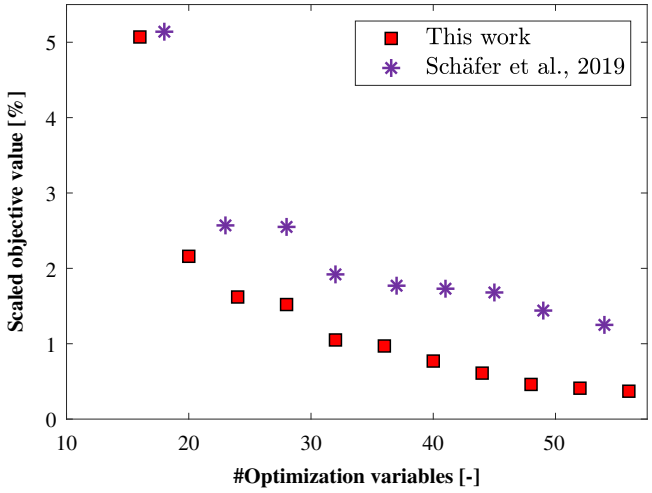

FIGURE 5 Comparison of the course of the optimal objective value for a horizon length of 128 using the sensitivity-based refinement strategy presented in this work (squares □) vs. the algorithm from our previous work[23] using a heuristic criterion for refinements (asterisks *) [Color figure can be viewed at wileyonlinelibrary.com]

length of T = 128 hourly intervals. As can be seen, the ability to identify suitable low-dimensional representations tailored to the characteristics of the original problem, such as active ramping constraints, is substantially enhanced by the sensitivity-based refinement procedure proposed herein. Thus, if we instead apply the refinement procedure from the previous work, the performance is poorer, so that

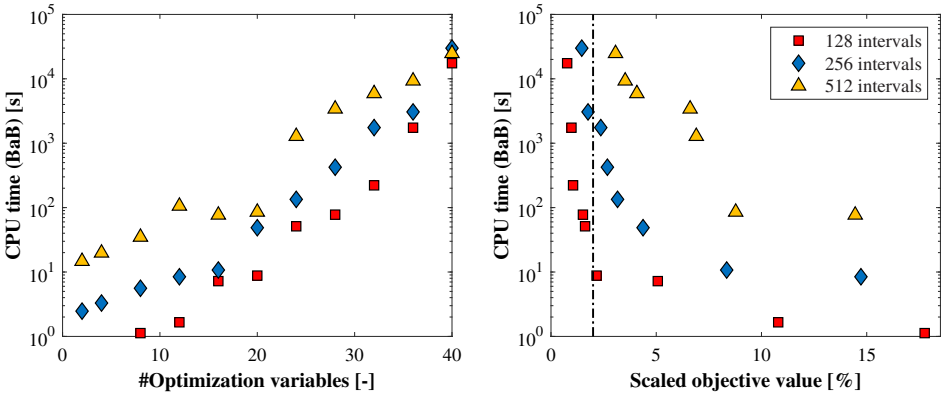

FIGURE 6 CPU time spent in MAiNGO's branch and bound (BaB) for converging the objective value to within the optimality tolerance. (a) Given as a function of the number of optimization variables; (b) plotted over the corresponding objective value. Squares (□): horizon length of 128 hourly intervals, diamonds (◇): 256 hourly intervals, triangles (△): 512 hourly intervals. The vertical dashed-dotted line further depicts the applied relative optimality tolerance for MAiNGO [Color figure can be viewed at wileyonlinelibrary.com]

similarly close approximations of the optima when considering the full dimensionality in terms of objective and variable values require more optimization variables. Note that due to a worst-case exponential scaling of computational times when performing deterministic global optimizations (see, below) in both cases, the achieved reductions in the number of optimization variables through the sensitivity-based refinement strategy correspond to highly promising times savings. Likewise, iterations using a similar number of optimization variables and thus resulting in similar computational times furnish improved approximations through the sensitivity-based refinement strategy.

6.2 | Computational performance

In this section, we discuss the savings in computational time enabled by the use of the proposed algorithm in comparison to solution approaches considering the full temporal dimensionality. In particular, we observe beneficial effects of truncating the wavelet transform for converging the lower bound in branch and bound algorithms to deterministic global optimization, which allows for obtaining converged solutions within the time limit. For illustration, we summarize the CPU times spent in MAiNGO's branch and bound procedure in Figure 6 for all problems, where a converged solution could have been obtained within the time limit. As can be seen, using the proposed refinement strategy successfully circumvents an unfavorable, worst-case exponential, scaling of computational times with the horizon length, that would be observed in solution approaches considering the full dimensionality. One rather finds that the CPU times only scale exponentially with the number of optimization variables (given a fixed horizon length), whereas they scale linearly with the horizon length (given a fixed number of optimization variables). Thus, by sensibly truncating the number of optimization variables, solutions are enabled in reasonable time that leave only small percentages of the maximum achievable savings unexploited that are in a similar order of magnitude as the relative optimality tolerance of the solver. More precisely, the 2% gap in savings is undercut after a CPU time of $\mathcal{O}\left(10^1\ \mathrm{s}\right)$ in case of T = 128 hourly scheduling intervals. In case of T = 256, this can still be achieved within 10^3...10^4 s. Even for T = 512, we expect this to be computationally tractable, considering ready-to-use opportunities for speed-ups by using parallel computing capabilities on local machines (∼10 cores).

Beside the opportunities for reducing the computational time for obtaining converged global optima, we observe a nearly linear scaling of computational times in the pre-processing with the number of optimization variables. Consequently, the algorithm appears also promising for reducing the computational burden for obtaining stationary feasible points.

7 | CONCLUSION

We present an algorithm for identifying low-dimensional yet highly accurate approximations of nonlinear scheduling problems subject to time-variables electricity prices. We first apply a wavelet transform of the time series of independent process variables, allowing to formulate a reduced-space optimization problem exposing only the wavelet coefficients to the optimizer. Thereby, we enable reductions in the dimensionality of the optimization problem by truncating the wavelet transform. Decisions concerning the use of specific wavelet coefficients are iteratively made based on sensitivity information from the Lagrangian multipliers. The algorithm is applied to the scheduling of an industrial-scale ASU, where we use an ANN as a surrogate model representing the nonlinear power consumption characteristics of the plant obtained from optimizations using a rigorous process model.

We demonstrate that substantially truncated wavelet transforms are still able to reproduce time series that allow for the exploitation of high percentages of the maximum achievable savings. Moreover, we show that the reductions in the dimensionality of the optimization problems lead to significant savings in computational times when performing deterministic global optimizations. This circumvents a worst-

case exponential scaling of solution times with the horizon length, which characterizes solution approaches considering the full dimensionality. Consequently, we hold the proposed algorithm to be highly useful if targeting the solution of strongly nonconvex optimization problems involving multiple local solutions.

Furthermore, the identified savings in computational times for local searches make the algorithm promising for applications where solution times are of crucial importance, for example, to ensure real-time capability. Here, we highlight in particular that single-shooting approaches to dynamic optimization are formulated in a similar way as the reduced-space quasi-stationary scheduling addressed in this work with both the objective as well as the path-constrained variables being explicit functions of the parameterized, commonly piecewise-constant input signals while hiding the actual model, that is, the state integration in case of dynamic optimization, from the optimizer.[53] Consequently, such a formulation would also allow to expose the coefficients of the wavelet transforms of the input signals as only variables to the optimizer, making the proposed refinement strategy applicable. Thus, promising future fields of application will certainly also include moving horizon applications involving nonlinear dynamic optimizations, which aim at integrating both the scheduling and the control perspective. The application of the presented algorithm for scheduling with low-dimensional dynamic surrogate models[47,54] and economic optimizations directly in the control layer using a detailed dynamic model[35,55,56] is thus left for future work.

An important research perspective lies in the extension of the algorithm towards the handling of scheduling problems that comprise logic disjunctions or conditional statements, that is, to those problems where binary variables in state-of-the-art formulations do not solely stem from piecewise linearizations and could thus not be eliminated by using the original (smooth) nonlinear function. Here, it should be thoroughly investigated if alternative formulations exist that describe the relationships but forgo binary variables. In particular, progress mostly made in the area of dynamic optimization regarding nonsmooth optimization[57-60] as well as the handling of complementarity constraints[61-64] already bear a potential for eliminating the binary variables in many conventional problem formulations, which would in these cases allow for the direct application of the proposed sensitivity-based refinement strategy using Lagrangian multipliers. In addition, we are still interested in a tailored treatment for binary DoFs similar to that of continuous ones, that is, using truncatable transforms and providing quantitative measures for decisions on refinements. Along these lines, future research should also investigate if the underlying concepts of our work (in particular, the reduced-space optimization and the truncated wavelet transforms) are computationally beneficial as well in case of mixed-integer linear scheduling problems, which have been extensively studied in literature.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Kopernikus project SynErgie by the Federal Ministry of Education and Research (BMBF) and the project supervision by the project management organization Projektträger Jülich. The authors further thank Adrian Caspari for providing the case study and Dominik Bongartz as well as Jaromil Najman for helpful advice during the use of MAiNGO. Open access funding enabled and organized by Projekt DEAL.

ORCID

Pascal Schäfer https://orcid.org/0000-0002-3268-8976
Artur M. Schweidtmann https://orcid.org/0000-0001-8885-6847
Alexander Mitsos https://orcid.org/0000-0003-0335-6566

REFERENCES

1. Mitsos A, Asprion N, Floudas CA, et al. Challenges in process optimization for new feedstocks and energy sources. *Comput Chem Eng*. 2018;113:209-221.
2. Zhang Q, Grossmann IE. Enterprise-wide optimization for industrial demand side management: fundamentals, advances, and perspectives. *Chem Eng Res Des*. 2016;116:114-131.
3. Daryanian B, Bohn RE, Tabors RD. Optimal demand-side response to electricity spot prices for storage-type customers. *IEEE Trans Power Syst*. 1989;4:897-903.
4. Ierapetritou MG, Wu D, Vin J, Sweeney P, Chigirinskiy M. Cost minimization in an energy-intensive plant using mathematical programming approaches. *Ind Eng Chem Res*. 2002;41:5262-5277.
5. Karwan MH, Keblis MF. Operations planning with real time pricing of a primary input. *Comput Oper Res*. 2007;34:848-867.
6. Mitra S, Grossmann IE, Pinto JM, Arora N. Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes. *Comput Chem Eng*. 2012;38:171-184.
7. Mitra S, Sun L, Grossmann IE. Optimal scheduling of industrial combined heat and power plants under time-sensitive electricity prices. *Energy*. 2013;54:194-211.
8. Zhang Q, Grossmann IE, Heuberger CF, Sundaramoorthy A, Pinto JM. Air separation with cryogenic energy storage: optimal scheduling considering electric energy and reserve markets. *AIChE J*. 2015;61:1547-1558.
9. Kelley MT, Pattison RC, Baldick R, Baldea M. An MILP framework for optimizing demand response operation of air separation units. *Appl Energy*. 2018;222:951-966.
10. Cozad A, Sahinidis NV, Miler DC. Learning surrogate models for simulationbased optimization. *AIChE J*. 2015;60:2211-2227.
11. Schweidtmann AM, Mitsos A. Deterministic global optimization with artificial neural networks embedded. *J Optim Theory App*. 2019;180:925-948.
12. Lizarraga-Garcia E, Ghobeity A, Totten M, Mitsos A. Optimal operation of a solar-thermal power plant with energy storage and electricity buy-back from grid. *Energy*. 2013;51:61-70.
13. Pattison RC, Touretzky CR, Johansson T, Harjunkoski I, Baldea M. Optimal process operations in fast-changing electricity markets: framework for scheduling with low-order dynamic models and an air separation application. *Ind Eng Chem Res*. 2016;55:4562-4584.
14. Ghobeity A, Mitsos A. Optimal time-dependent operation of seawater reverse osmosis. *Desalination*. 2010;263:76-88.
15. Teichgraeber H, Brandt AR. Clustering methods to find representative periods for the optimization of energy systems: an initial framework and comparison. *Appl Energy*. 2019;239:1283-1293.
16. Lythcke-Jorgensen CE, Münster M, Ensinas AV, Haglind F. A method for aggregating external operating conditions in multi-generation system optimization models. *Appl Energy*. 2016;166:59-75.
17. Poncelet K, Höschle H, Delarue E, Virag A, Dhaeseleer W. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Trans Power Syst*. 2017;32:1936-1948.
18. Bahl B, Lützow J, Shu D, et al. Rigorous synthesis of energy systems by decomposition via time-series aggregation. *Comput Chem Eng*. 2018;112:70-81.

19. Baumgärtner N, Bahl B, Hennen M, Bardow A. RiSES3:rigorous synthesis of energy supply and storage systems via time-series relaxation and aggregation. *Comput Chem Eng* 2019;127:127-139.
20. Baumgärtner N, Shu D, Bahl B, Hennen M, Hollermann DE, Bardow A. DeLoop: decomposition-based long-term operational optimization of energy systems with time-coupling constraints. *Energy*. 2020;198:117272.
21. Pineda S, Morales JM. Chronological time-period clustering for optimal capacity expansion planning with storage. *IEEE Trans Power Syst*. 2018;33:7162-7170.
22. Palys MJ, Daoutidis P. Using hydrogen and ammonia for renewable energy storage: a geographically comprehensive techno-economic study. *Comput Chem Eng*. 2020;136:106785.
23. Schäfer P, Schweidtmann AM, Lenz PHA, Markgraf HMC, Mitsos A. Wavelet-based grid-adaptation for nonlinear scheduling subject to time-variable electricity prices. *Comput Chem Eng*. 2020;132:106598.
24. Schlegel M, Stockmann K, Binder T, Marquardt W. Dynamic optimization using adaptive control vector parameterization. *Comput Chem Eng*. 2005;29:1731-1751.
25. Mitsos A, Chachuat B, Barton P. McCormick-based relaxations of algorithms. *SIAM J Optim*. 2009;20:573-601.
26. Bongartz D, Mitsos A. Deterministic global optimization of process flowsheets in a reduced space using McCormick relaxations. *J Global Optim*. 2017;69:761-796.
27. Antonini M, Barlaud M, Mathieu P, Daubechies I. Image coding using wavelet transform. *IEEE Trans Image Process*. 1992;1:205-220.
28. Flores-Quiroz A, Palma-Behnke R, Zakeri G, Moreno R. A column generation approach for solving generation expansion planning problems with high renewable energy penetration. *Electr Pow Syst Res*. 2016;136:232-241.
29. Lara CL, Mallapragada DS, Papageorgiou DJ, Venkatesh A, Grossmann IE. Deterministic electric power infrastructure planning: mixed-integer programming model and nested decomposition algorithm. *Eur J Oper Res*. 2018;271:1037-1054.
30. Flores-Quiroz A, Pinto JM, Zhang Q. A column generation approach to multiscale capacity planning for power-intensive process networks. *Optim Eng*. 2019;20:1001-1027.
31. Caspari A, Offermanns C, Schäfer P, Mhamdi A, Mitsos A. A flexible air separation process: 1. Design and steady-state optimizations. *AIChE J*. 2019;65:e16705.
32. Schäfer P, Westerholt H, Schweidtmann AM, Ilieva S, Mitsos A. Model-based bidding strategies on the primary balancing market for energy-intense processes. *Comput Chem Eng*. 2019;120:4-14.
33. Hubbard BB. *The world according to wavelets: the story of a mathematical technique in the making*. Wellesley, MA: AK Peters, Ltd; 1996.
34. Mallat SG. Multiresolution approximations and wavelet orthonormal bases of L2(R). *Trans Am Math Soc*. 1989;315:69-87.
35. Caspari A, Offermanns C, Schäfer P, Mhamdi A, Mitsos A. A flexible air separation process: 2. Optimal operation using economic model predictive control. *AIChE J*. 2019;65:e16721.
36. Mitra S, Pinto JM, Grossmann IE. Optimal multi-scale capacity planning for power-intensive continuous processes under time-sensitive electricity prices and demand uncertainty. Part I: modeling. *Comput Chem Eng*. 2014;65:89-101.
37. Zhang Q, Sundaramoorthy A, Grossmann IE, Pinto JM. A discrete-time scheduling model for continuous power-intensive process networks with various power contracts. *Comput Chem Eng*. 2016;84:382-393.
38. Zhang Q, Morari MF, Grossmann IE, Sundaramoorthy A, Pinto JM. An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load. *Comput Chem Eng*. 2016;86:106-119.
39. Obermeier A, Windmeier C, Esche E, Repke J-U. A discrete-time scheduling model for power-intensive processes taking fatigue of equipment into consideration. *Comput Chem Eng*. 2019;195:904-920.
40. Kender R, Wunderlich B, Thomas I, Peschel A, Rehfeldt S, Klein H. Pressure-driven dynamic simulation of start up and shutdown procedures of distillation columns in air separation units. *Chem Eng Res Des*. 2019;147:98-112.
41. Miller J, Luyben WL, Belanger P, Blouin S, Megan L. Improving agility of cryogenic air separation plants. *Ind Eng Chem Res*. 2008;47:394-404.
42. Baldea M, Harjunkoski I. Integrated production scheduling and process control: a systematic review. *Comput Chem Eng*. 2014;71:377-390.
43. Cao Y, Swartz CLE, Flores-Cerrillo J. Preemptive dynamic operation of cryogenic air separation units. *AIChE J*. 2017;63:3845-3859.
44. Schäfer P, Bering LF, Caspari A, Mhamdi A, Mitsos A. Nonlinear dynamic optimization for improved load-shifting agility of cryogenic air separation plants. In: Eden MR, Ierapetritou MG, Towler GP, eds. *13th International Symposium on Process Systems Engineering (PSE 2018). Volume 44 of Computer Aided Chemical Engineering*. Amsterdam: Elsevier; 2018:547-552.
45. Du J, Park J, Harjunkoski I, Baldea M. A time scale-bridging approach for integrating production scheduling and process control. *Comput Chem Eng*. 2015;79:59-69.
46. Tsay C, Baldea M. Integrating production scheduling and process control using latent variable dynamic models. *Control Eng Pract*. 2020;94:104201.
47. Pattison RC, Touretzky CR, Harjunkoski I, Baldea M. Moving horizon closed-loop production scheduling using dynamic process models. *AIChE J*. 2017;63:639-651.
48. Tsay C, Kumar A, Flores-Cerrillo J, Baldea M. Optimal demand response scheduling of an industrial air separation unit using data-driven dynamic models. *Comput Chem Eng*. 2019;126:22-34.
49. Bongartz D, Najman J, Sass S, Mitsos A. *MAiNGO—McCormick-based Algorithm for Mixed-integer Nonlinear Global Optimization*. 2018. http://permalink.avt.rwth-aachen.de/?id=729717.
50. McCormick GP. Computability of global solutions to factorable nonconvex programs: part I—convex underestimating problems. *Math Program*. 1976;10:147-175.
51. Tsoukalas A, Mitsos A. Multivariate McCormick relaxations. *J Global Optim*. 2014;59:633-662.
52. Wächter A, Biegler LT. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math Program*. 2006;106:25-57.
53. Biegler LT. *Nonlinear programming*. Philadelphia: Society for Industrial and Applied Mathematics; 2010.
54. Otashu JI, Baldea M. Demand response-oriented dynamic modeling and operational optimization of membrane-based chlor-alkali plants. *Comput Chem Eng*. 2019;121:396-408.
55. Cao Y, Swartz CLE, Flores-Cerrillo J. Optimal dynamic operation of a high-purity air separation plant under varying market conditions. *Ind Eng Chem Res*. 2016;55:9956-9970.
56. Schäfer P, Caspari A, Mhamdi A, Mitsos A. Economic nonlinear model predictive control using hybrid mechanistic data-driven models for optimal operation in real-time electricity markets: in-silico application to air separation processes. *J Process Control*. 2019;84:171-181.
57. Sahlodin AM, Watson HAJ, Barton PI. Nonsmooth model for dynamic simulation of phase changes. *AIChE J*. 2016;62:3334-3351.
58. Khan KA, Barton PI. Generalized derivatives for hybrid systems. *IEEE Trans Autom Control*. 2017;62:3193-3208.
59. Stechlinski P, Patrascu M, Barton PI. Nonsmooth differential-algebraic equations in chemical engineering. *Comput Chem Eng*. 2018;114:52-68.
60. Barton PI, Khan KA, Stechlinski P, Watson HAJ. Computationally relevant generalized derivatives: theory, evaluation and applications. *Optim Methods Softw*. 2018;33:1030-1072.
61. Raghunathan AU, Biegler LT. Mathematical programs with equilibrium constraints (MPECs) in process engineering. *Comput Chem Eng*. 2003;27:1381-1392.
62. Raghunathan AU, Soledad Diaz M, Biegler LT. An MPEC formulation for dynamic optimization of distillation operations. *Comput Chem Eng*. 2004;28:2037-2052.
63. Baumrucker BT, Renfro JG, Biegler LT. MPEC problem formulations and solution strategies with chemical engineering applications. *Comput Chem Eng*. 2008;32:2903-2913.

64. Caspari A, Lüken L, Schäfer P, et al. Dynamic optimization with complementarity constraints: smoothing for direct shooting. *Comput Chem Eng*. 2020;139:106891.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Schäfer P, Schweidtmann AM, Mitsos A. Nonlinear scheduling with time-variable electricity prices using sensitivity-based truncations of wavelet transforms. *AIChE J*. 2020;66:e16986. https://doi.org/10.1002/aic.16986