
Fernán J. Serralunga, Pío A. Aguirre, Miguel C. Mussati
INGAR Instituto de Desarrollo y Diseño (CONICET-UTN), Avellaneda 3657, (S3002GJC) Santa Fe, Argentina
{fernanserralunga, paguir, mmussati}@santafe.conicet.gov.ar

Abstract.
The optimization of heat and power systems operation is a complex task that involves continuous and discrete variables, operating and environmental constraints, uncertain prices and demands and transition constraints for startups or shutdowns. This work proposes a novel methodology for integrating scheduling optimization and real-time optimization (RTO) in order to face and solve such optimization problem. In a first stage, an offline optimization finds a scheduling for the whole horizon under study, which sets the startups and shutdowns of pieces of equipment with long transition times. A second stage solves a multi-period RTO, which corrects the forecasts and adapts the model before optimizing the process. Although the proposed methodology is illustrated through a case study consisting in a heat and power system, it can be generalized to other systems and processes. The obtained results show significant improvements in comparison with applying the results of a single offline scheduling optimization.

Keywords: real-time optimization, scheduling optimization, combined heat and power systems, energy optimization

1 Introduction

The optimization of industrial heat and power systems is often a significant source of savings. It is frequently subjected to changes in steam and power demands along the day, as well as in power prices (which may even change in real-time according to the market power demand). The decision variables can be both continuous (boilers and gas turbine loads, for example) and discrete (startups and shutdowns). The involved constraints can include operating limits of pieces of equipment, operative decisions (control strategies, for example) and environmental regulations.

Scheduling optimization is a possible approach to operate this kind of systems. Mitra et al. [1] have suggested a MILP formulation which accounts for operating modes and transitional behavior of pieces of equipment. That work also includes an excellent review of previous approaches. Other authors present an industrial application of a
A MILP multiperiod formulation which optimizes a refinery utilities system, although it is oriented to mid-term planning [2]. A strategy to find the optimal operation of an utility system for a single period has been developed [3]. It takes into account fixed operating costs and transition costs for startups and shutdowns, as well as the period duration; a successive-MILP approach is used to solve the nonlinearities of energy balances. All these approaches need a forecast of prices and demand, and they solve the schedule assuming no uncertainties in the forecasts and no plant-model mismatch. Therefore, the implementation of the schedule can lead to a suboptimal operation, or it may cause constraint violations.

Real-time optimization (RTO), on the other hand, makes use of process measurements to deal with plant-model mismatch [4], as well as updated prices and demands. A recent work presents an RTO strategy which exploits the structure of energy systems, and included an example of a sugar and ethanol plant utility system [5]. Industrial applications of RTO of energy systems in refineries have been reported [6, 7]. The main drawback of real-time optimization of heat and power systems is that it generally optimizes a single steady state, which does not allow for including transition constraints and accumulated variables, such as tank levels or daily emissions. Furthermore, in general, RTO formulations are of NLP type, which do not account for startups and shutdowns. Some MINLP formulations for single-period RTO including disjunctions have been recently presented by the authors [8, 9].

In other research areas, some desirable properties of RTO have been proposed for scheduling, and vice versa. In the scheduling of batch processes, the uncertainties in forecasts and the plant-model mismatch can be partially overcome by solving iteratively the scheduling problem in a rolling horizon [10]. A multi-period RTO strategy with shrinking (fixed) horizon has been proposed to obtain an optimal blending of gasolines [11], using an NLP model.

This work proposes a novel framework for integrating real-time optimization (RTO) and scheduling optimization (SO). The approach is based on a MINLP formulation and is solved in two stages. Firstly, the optimal schedule is solved for the full period under study (for example, 7 days). This solution provides the startups and shutdowns of pieces of equipment with long transition times (gas turbine and boilers). Secondly, the operation for each period is obtained by a real-time optimization strategy (i.e. adapting the process model and the forecasts), with a fixed horizon at the end of the day. The proposed methodology is illustrated through a case study consisting in a heat and power system, which includes a gas turbine with heat recovery steam generator, two fired boilers, an extraction-condensing turbine that generates electric power and a set of steam turbines that can be replaced by electric motors. A limit to the total NOx production per day is imposed. The results of the strategy are compared with those obtained by the offline optimal schedule.

2 Problem Statement

The problem to solve can be stated in a simplified way as follows:
where $k$ (from $k_0$ to $k_f$) is the index corresponding to a period, $u_k$ are the process inputs, $y_k$ are the measured outputs and $\mathbf{D}, \mathbf{z}$ are discrete variables corresponding to disjunctions $d$. The variables $\eta_k = \xi_{Gk}, \eta_k, d \in D$ are performance factors, whose functionality with respect to $u$ is only known in an approximate way by functions $\xi_{G}, p_d(u_k)$. These functions contain all the plant-model mismatch, while the process model $\mathbf{h}_d, d \in D$ includes mass, energy and entropy balances [5], including forecasted demands. Inequality constraints are expressed by functions $\mathbf{g}_d, g_k, d \in D$. Equations $\mathbf{Q}$ include logical constraints. Equations $\Psi$ contain the links between periods. $Q_k$ is a cost function (which includes forecasted prices), and $c_{dk}$ are fixed costs. For the sake of simplicity, transition discrete variables and constraints, as well as transition costs $t_{dk}$, are included implicitly in $\mathbf{Q}$. For a detailed description on possible transition variables, constraints and costs, the reader is directed to reference [1].

3 Integration of Real-Time Optimization and Scheduling Optimization (SO+RTO)

3.1 Proposed Methodology

Let the disjunctions set $D$ be divided in two subsets, $D_1$ and $D_2$. Set $D_1$ includes all discrete decisions (i.e. operating mode selection, startups and shutdowns) with relatively long transition periods. The rest of the discrete decisions are included in set $D_2$.

In an industrial application, the size of a problem such as that of Eq. (1) can be too computationally expensive to be solved with high frequency and applied in real-time. On the other hand, if it is only solved offline and with low frequency, the errors in
modeling (plant-model mismatch) and the deviations from actual and forecasted prices and demands can lead to a suboptimal operation or to constraint violations.

The optimization problem given by Eq. (1) is solved in two stages. Stage 1 solves the optimal scheduling problem for the whole horizon, in order to exploit the higher number of degrees of freedom of the scheduling formulation. As a result of this stage, all the discrete decisions belonging to the set $D_1$ are fixed and applied to the plant.

Stage 2 makes use of the desired feedback properties of real-time optimization, and is solved with a higher frequency. This stage has two main objectives: The first is to update the forecasts of prices and demands. The second is to adapt the model using the available measurements in order to overcome the plant-model mismatch. Having this updated model and information, the optimal operation is recalculated. The degrees of freedom are the decisions belonging to set $D_2$ and the continuous process inputs $u$.

In order to reduce the size of the problem in this stage, the number of periods considered for optimization is reduced, as well as the number of discrete variables.

### 3.2 Stage 1: Full Schedule

The problem given by Eq. (1) is solved from the initial time to the final horizon ($k = 1 \ldots N$) with an initial forecast for demands and prices. The optimal value of the discrete variables belonging to $D_1$, $z_{dk}^\text{opt}$, $d \in D_1$, $k = 1 \ldots N$, is stored and will not be modified by the real-time optimization stage. A new constraint is added to Eq. (1):

$$z_{dk} = z_{dk}^\text{opt}, \quad d \in D_1, \; k = 1 \ldots N$$

(2)

Stage 1 can be performed with a fixed horizon length and repeated when the final time $N$ is reached, but it can also be performed in a rolling horizon strategy, or updated when a significant change in the forecasts or the system conditions is detected. The key characteristic of this stage is that it is performed offline and it sets the startups and shutdowns of pieces of equipment with longer transition periods.

### 3.3 Stage 2: Real-Time Optimization

**Adaptation Strategy.** At each time $k$, online measurements will be used to adapt performance equations from Eq. (1):

$$\eta_{G,i} = p_{G,i}(u) + a_{G,i,k} + b_{G,i,k}^T \cdot v_{G,i} \quad i = 1 \ldots ng$$

$$\eta_d = p_{d}(u) + a_{d,k} + b_{d,k}^T \cdot v_d \quad d \in D$$

(3)

where $a$ and $b$ are constant and gradient modifiers (i.e., correction terms) [12], and $v_{G,i}$ and $v_d$ are subsets or linear combinations of the input set $u$ [5]. $ng$ is the number of “global” performance factors (i.e., the dimension of vector $\eta_G$).

The modifiers are updated using the following procedure:
1. The actual performance factors \( \eta_{G,i,k-1}, \eta_{d,k-1}, d \in D \) are obtained from online measurements and the process mass, energy and entropy balances.

2. The differences \( \beta \) between model and actual values are calculated:

\[
\begin{align*}
\beta_{G,i,k-1} &= \eta_{G,i,k-1}^{\text{real}} - \eta_{G,i,k-1}^d(u_{k-1}) & i = 1...ng \\
\beta_{d,k-1} &= \eta_{d,k-1}^{\text{real}} - \eta_{d,k-1}^d(u_{k-1}) & \forall d \in D / z_{d,k-1} = 1
\end{align*}
\]

The modifiers \( a \) and \( b \) are obtained with a weighted linear regression, using the last \( nr \) periods and a weight \( w \). The number of periods and the weighting strategy can be selected for each implementation, and even for each performance equation. Different criteria for choosing the weights have been proposed [13, 14]. For a disjunction \( d \), if modifiers \( a_d \) and \( b_d \) only appear if \( z_d = \text{true} \), they will only be updated when this disjunction was active in the last RTO cycle:

\[
\begin{align*}
\arg \min_{a,b} & \sum_{j=k-nr+1}^{k-1} w_j \left( \eta_{G,i,j}^{\text{real}} - a - b^T \cdot v_{G,i,j} \right) \\
\text{s.t.} & \quad a_{G,i,j}^L \leq a \leq a_{G,i,j}^U \\
& \quad b_{G,i,j}^L \leq b \leq b_{G,i,j}^U
\end{align*}
\]

\[
\begin{align*}
\arg \min_{a,b} & \sum_{j=k-nr+1}^{k-1} w_j \left( \eta_{d,j}^{\text{real}} - a - b^T \cdot v_{d,j} \right) \\
\text{s.t.} & \quad a_{d,j}^L \leq a \leq a_{d,j}^U \\
& \quad b_{d,j}^L \leq b \leq b_{d,j}^U
\end{align*}
\]

**Forecast Update.** At the same time, prices and demand forecasts are updated. If a new forecast is available, it can be used; otherwise, the original forecasts can be biased using the current error between actual and forecasted values.

**Optimization.** After updating the model and the forecasts, the real-time optimization stage is performed before the beginning of each period \( k \). In order to reduce computational time of the problem, the full scheduling horizon is divided in smaller subsets \( j \):

\[
H_N \Rightarrow \frac{1}{3} k_f_1 \cup \frac{1}{3} k_f_2 \cup \frac{1}{3} k_f_3 + 1...k_f T + 1...N
\]

where \( k_f \ (j = 1...T) \) is the final period of each subset. In practice, a possible size for each subset is of 1 day (with \( k_f \) the last period of the day). It allows dealing with common constraints and costs present in heat and power systems (maximum daily power consumption, total daily natural gas use, total daily emissions, among others).

The optimization problem to be solved is that of Eq. (1), with the addition of Eq. (2). The prediction of the performance factors \( \eta_{G,d}, d \in D \) is modified as proposed in Eq.(3), using the most recent results of Eq. (5). The next period \( k \) is set as the initial period, and the closest future \( k_f \) is selected as the scheduling horizon. If \( k \) is
equal to the current \( k_{f,j} \), a single-period optimization is performed, and the next instance of the optimizer will run with \( k_{f,j+1} \) as horizon.

Figure 1 shows a diagram of the integrated SO+RTO system. In addition to Stages 1 and 2, a data validation step is proposed for correcting the measured inputs and outputs \( u_m^k \) and \( y_m^k \), respectively [15, 16]; and also a validation or filtering step for the optimal results [17, 18].

![Fig. 1. Diagram of the integrated SO+RTO system.](image)

4 Case Study: Application to a Heat and Power System

4.1 Case Description

A heat and power system (shown in Figure 2) was modeled. It consists of three boilers, a gas turbine, a heat recovery steam generator with supplementary firing (afterburning), an extraction-condensing turbine generating power, three steam turbines with spare motors for driving pumps, a steam demand from an industrial process, letdown valves and steam vents. The level of total daily NO\(_x\) emissions is constrained to be lower than 1000 kg/day. The system is studied with an horizon of 7 days; boilers startups and shutdowns are selected in Stage 1 (initial schedule), while on/off decisions for afterburning and steam turbines are selected in real-time (with a fixed horizon at the end of each day). The periods have a length of 6 hours.

All mass, energy and entropy balances were modeled rigorously. For performance factors (boilers efficiencies, gas turbine heat rate and maximum power, turbines efficiencies), as well as for NO\(_x\) emissions in boilers, gas turbine and afterburning, an approximate model is has been built. NO\(_x\) emissions have been modeled using emission factors [19].

For the purpose of evaluating the proposed strategy, a different model of the performance factors was built. It is assumed to represent exactly the process under study.
This model (called hereafter real plant) is used to simulate the real process, evaluate the actual cost achieved after applying optimization solutions and provide the factors $\eta_G^{\text{real}}$ required for Eq.(4). The main (structural) differences between the approximate model and the real plant are summarized in Table 1. It can be noticed that the NOx factor for the afterburning changes with time (i.e. with the cycle $k$). The rest of the differences (performance of turbines, turbogenerator and gas turbine) can be obtained from reference [8].

A forecast for power price, power demand, steam demand and ambient temperature is used for scheduling and RTO. For the purpose of evaluation of the approach, a different set of demands, power price and temperature represents the actual values of these properties along the scheduling horizon. These actual values are used for simulation (together with the real plant model) and for updating the forecast for the real-time optimization model (by biasing the current and future values using the error between actual and forecasted value at current cycle $k$). Forecasted and actual values are shown in Figure 3.

**Table 1.** Differences between the approximated model and the real plant. $F$: Boiler steam flow (t/h); $Q$: Boiler/Gas Turbine/Afterburning fuel use (Gcal/h); $k$: Period number

<table>
<thead>
<tr>
<th>Performance index</th>
<th>Approximate model</th>
<th>Real plant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency Boilers (%)</td>
<td>92</td>
<td>$75 + 0.48F -0.0035-F^2$</td>
</tr>
<tr>
<td>NOx Boiler 1 (kg/Gcal)</td>
<td>0.08</td>
<td>$0.07 + 2Q^{-1}$</td>
</tr>
<tr>
<td>NOx Boiler 2 (kg/Gcal)</td>
<td>0.08</td>
<td>$0.08 + 3Q^{-1}$</td>
</tr>
<tr>
<td>NOx Boiler 3 (kg/Gcal)</td>
<td>0.08</td>
<td>$0.06 + 4Q^{-1}$</td>
</tr>
<tr>
<td>NOx Gas Turbine (kg/Gcal)</td>
<td>0.055</td>
<td>$0.03 + 10Q^{-1}$</td>
</tr>
<tr>
<td>NOx Afterburning (kg/Gcal)</td>
<td>0.07</td>
<td>$0.05 + 5Q^{-1} + 0.001k$</td>
</tr>
</tbody>
</table>
Fig. 2. Heat and power system diagram.

The objective function to minimize is the operating cost:

$$
\sum_{k=0}^{k_f} \left( 7.2 \cdot \left( Q_{TG,k} + Q_{AB,k} + \sum_{bl=1}^{3} Q_{bl,k} \right) + 0.5 \cdot FW_k + C_{pow,k} \cdot P_k \right) + 20 \cdot \sum_{i=1}^{3} \left( \zeta_{off-on,k}^{on-off} + \zeta_{on-off,k}^{off-on} \right) + 50 \sum_{j}^{} EX_{j}^{NOx} 
$$

(6)

where $Q_{TG,k}$, $Q_{AB,k}$ and $Q_{bl,k}$ are the fuel consumptions in the gas turbine, the afterburning and a boiler, respectively, in Gcal/h); $FW_k$ is the boiler feedwater use in t/h, $C_{pow,k}$ is the power purchase cost ($/MWh$) and $P_k$ is the net power import (MW); and $\zeta_{on-off,k}^{on-off}$ and $\zeta_{off-on,k}^{off-on}$ are binary variables indicating that, a turbine $t$ has been turned off or on, respectively, in cycle $k$. After each day $j$, the excess NOx produced over the constraint $EX_{j}^{NOx}$ (i.e. all NOx production higher than 1000 kg/h) is penalized with a cost.

Other constraints include the need for keeping always one boiler in operation, and that each backpressure turbine can only be turned on or off once every 24 hours. The turbogenerator cannot be stopped.

4.2 Results and Discussion

The case study was implemented in GAMS and solved using DICOPT. Stage 1 (full schedule) has 2541 equations and 2745 variables (196 discrete). It is solved in a computer with an Intel Core i7-2670QM (2.2 GHz) processor and 8 GB of RAM memory RAM. The CPU time required for the solution was of 8.7 s. Each of the real-time stages requires a CPU usage < 1.9 s, and it has a maximum size of 362 equations and 418 variables (16 discrete).

The results obtained by applying only Stage 1 are also calculated for comparison. The actual cost and constraints are calculated for each period using the real model.

The total operating cost for the comparison case is $204792, while the proposed strategy leads to a cost of $161182. The relative reduction in cost is 21.3%.

Figure 4 shows the cost evolution for the proposed strategy and the comparison case. It can be observed that the model and forecast updates provided by the real-time optimization strategy impact on a reduction of total cost in all periods. The NOx penalty cost cannot be related to each period but for each day, but for the purpose of illustrating the results it is distributed equally over the periods of the corresponding day (25% of the total daily NOx penalty is added to the cost of each period of the day).

Figure 5 shows the evolution of NOx total production. Again, the feedback properties of real-time optimization reduce the number and the size of violations of the daily
NOx production constraint. The plant-model mismatch and the uncertainty in the forecasted demands and prices lead to errors in the predicted fuel use and NOx production in the boilers and the gas turbine, which is (at least partially) corrected by real-time optimization. In this example, the model does not predict any violations of the NOx constraint for any of the days. However, the application of the results of the comparison case to the real plant causes a NOx production higher than the allowed limit. In this case, the limit is modeled as a “soft” constraint, which can be slightly violated but with a high penalization in the objective function. In other situations, a limit like this could be a hard constraint, and therefore the proposed schedule would be infeasible.

Fig. 3. Forecasted and actual values for ambient temperature, power price and steam and power demand.
**Fig. 4.** Evolution of the objective function for the proposed SO+RTO strategy and the comparison case (offline scheduling optimization)

**Fig. 5.** Daily accumulated NOx production for offline scheduling optimization and SO+RTO strategies.

Figure 6a shows the steam vents for each period, with significant improvements when using the proposed methodology. This is directly related with the real-time management of backpressure turbines, shown in Figure 6b. Again, the real-time corrections of the model and the current and forecasted conditions lead to a reduction of the cost.

**Fig. 6.** (a) Steam vent and (b) Number of turbines operating for offline scheduling optimization and SO+RTO strategies.
5 Conclusions

A novel framework for integrating optimal scheduling and real-time optimization of continuous processes has been presented. It can be particularly useful for continuous improvement of the operation of combined heat and power systems. However, it can be generalized to other systems or processes.

This framework makes use of the higher number of degrees of freedom of an optimal scheduling formulation, as well as of the feedback properties and the lower computational cost of real-time optimization. Decision variables with long transition periods are fixed in the offline scheduling stage, while the remaining degrees of freedom are set in a multi-period RTO stage. The scheduling optimization can be performed offline and with low frequency, while RTO is performed using online measurements, which allows correcting the forecasted conditions and the model parameters. Except for the corrected parameters, the mixed-integer nonlinear models used in the scheduling and the RTO stage are the same.

A case study that optimizes 7-days a week operation of a combined heat and power system was used to illustrate the proposed methodology. A combined scheduling optimization+RTO strategy with a fixed horizon for the offline scheduling and a shrinking horizon for the RTO stage was implemented. The results show a significant improvement in operating cost reduction and constraint satisfaction, in comparison with the direct implementation of the offline scheduling results.

The MINLP problem solved in the real-time stage may present convergence problems that lead to an infeasible solution or to not obtaining a solution in the required (limited) time. A multiperiod MINLP real-time optimization system must have a strategy to recover from these problems, which will be the subject of future work.

Acknowledgements: The authors gratefully acknowledge CONICET and Soteica Latinoamérica S.A. for the financial support.

References


