Integrated Scheduling and Maintenance (GASUEnv)

April 22, 2025

1 Environments in the paper

1.1 Integrated scheduling and maintenance (GASUEnv)

GASUEnv simulates the operation of an air separation unit (ASU) focused on meeting gaseous product demand. Since inventorying gaseous products is impractical, the system allows for external product purchases when demand exceeds production capacity. The ASU consists of ($n =$) three compressors (C), each with a maximum capacity Cap_c. The agent must decide for each compressor whether to operate at a production level $r_c \in [0, 1]$ of its maximum capacity or schedule maintenance, depending on its condition.

Each episode spans 30 days, with the agent determining the appropriate action for each compressor based on daily observations, and deciding whether to place an external purchase order. The objective is to minimize the total operational cost, which includes production and external purchase cost.

Observation Space:
The observation state at time t is represented as a vector,

$$s(t) = (\mathbf{d}_t, \mathbf{e}_t, \mathbf{tslm}_t, \mathbf{tlcm}_t, \mathbf{cdm}_t), \quad t \in [1, 31]$$

For $t \in [1, 31]$, the demand forecast $d_t \in \mathbb{R}_+^{S\times 1}$ represents a vector of predicted demands from day t to $t + S - 1$, expressed in tons, and the electricity price forecast $e_t \in \mathbb{R}_+^{S\times 1}$ provides the corresponding day-ahead electricity prices in KWh. Here, $S = 30$ denotes the length of the lookahead horizon included in the observation. Additionally, $\mathrm{tslm}_t \in \mathbb{Z}_+^{n\times 1}$ denotes the *Time Since Last Maintenance* (a vector of length n, where tslm_{ct} is the value for compressor c), $\mathrm{tlcm}_t \in \mathbb{Z}_+^{n\times 1}$ represents the *Time Left to Complete Maintenance* (a vector of length n, with tlcm_{ct} for compressor c), and $\mathrm{cdm}_t \in \{0, 1\}^n$ is the *Can Do Maintenance* indicator (a binary vector of length n, where $\mathrm{cdm}_{ct} = 1$ means maintenance is allowed for compressor c on day t). The state vector $s(t)$ thus encapsulates both operational and maintenance-related conditions of the ASU.

Action Space:
The action space at each time step t is defined as:

$$a(t) = (a_{\text{maintenance}}(t), a_{\text{production}}(t), a_{\text{purchase}}(t))$$

where $a_{\text{maintenance}}(t) \in \{0,1\}^n$ is a vector indicating whether maintenance is scheduled for each compressor, with $a_{\text{maintenance},c}(t) = 1$ if compressor i is under maintenance, and 0 otherwise. The vector $a_{\text{production}}(t) \in [0,1]^n$ represents the production rate for each compressor, where $a_{\text{production},c}(t) \in [0,1]$ denotes the fraction of the maximum capacity Cap_c of compressor i to be utilized at time t. Additionally, $a_{\text{purchase}}(t) \in [0, 10000]$ represents the amount of external product to be purchased, with $a_{\text{purchase}}(t)$ being a scalar value indicating the external purchase quantity required to meet demand when production is insufficient.

Reward Function:
The reward function represents the cost incurred by the agent for making decisions related to production and external purchases. At each time step t, the reward is defined as:

$$r(t) = -\left(\text{production cost}_t + \text{external purchase cost}_t\right)$$

The production cost at time t is calculated using the production rate, compressor capacity, specific energy consumption, and electricity price. The external purchase cost is incurred when demand exceeds production capacity, calculated by multiplying the purchase amount by the external price.

Penalties:
The agent may incur various penalties based on the decisions it makes at each time step. These penalties represent violations of the system's constraints and help guide the agent towards optimal behavior by discouraging undesirable actions, such as prematurely interrupting maintenance, overloading compressors, or failing to meet demand. The penalties are as follows:

1.1.1 Maintenance Duration Penalty

This penalty is incurred if maintenance is interrupted before it is completed or prolonged, in which case tlcm_{ct} becomes negative. The penalty is defined as:

$$\text{penalty}_{ct,\text{MD}} = \begin{cases} & \text{if } \left(a_{\text{maintenance},c}(t) \neq 1 \text{ and } \text{tlcm}_{ct} > 0\right) \\ -\rho_{\text{MD}} \cdot \exp\left(|\text{tlcm}_{ct}|\right) & \textit{or } \left(a_{\text{maintenance},c}(t) = 1 \text{ and } \text{tlcm}_{ct} = 0 \text{ and } \text{tslm}_{ct} = 0\right) \\ & \textit{or } \left(a_{\text{maintenance},c}(t) = 1 \text{ and } \text{tlcm}_{ct} < 0\right), \\ 0 & \text{otherwise.} \end{cases}$$

1.1.2 Maintenance Failure Penalty

This penalty occurs if the *Time Since Last Maintenance* (tslm_{ct}) exceeds the *Mean Time to Failure* (MTTF_c) of the compressor. The penalty is defined as:

$$\text{penalty}_{ct,\text{MF}} = \begin{cases} -\rho_{\text{MF}} \cdot (\text{tslm}_{ct} - \text{MTTF}_c) & \text{if } (a_{\text{maintenance},c}(t) = 0 \text{ and } \text{tslm}_{ct} > \text{MTTF}_c)\,, \\ -\rho_{\text{MF}} & \text{if } (a_{\text{maintenance},c}(t) = 0 \text{ and } \text{tslm}_{ct} = \text{MTTF}_c)\,, \\ 0 & \text{otherwise.} \end{cases}$$

1.1.3 Early Maintenance Penalty

This penalty is incurred if maintenance is performed on a compressor when it is not yet eligible for maintenance (i.e., when $\text{cdm}_{ct} = 0$, indicating that the compressor has recently undergone maintenance, and $\text{tlcm}_{ct} = 0$). The penalty is proportional to the *Time Since Last Maintenance* (tslm_{ct}) of the compressor.

$$\text{penalty}_{ct,\text{EM}} = \begin{cases} -\rho_{\text{EM}} \cdot \text{tslm}_{ct} & \text{if } (a_{\text{maintenance},c}(t) = 1 \text{ and } \text{cdm}_{ct} = 0 \text{ and } \text{tlcm}_{ct} = 0)\,, \\ 0 & \text{otherwise.} \end{cases}$$

1.1.4 Ramp Penalty

This penalty is incurred when a compressor is ramped up while under maintenance.

$$\text{penalty}_{ct,\text{RP}} = \begin{cases} -\rho_{\text{RP}} \cdot a_{\text{production},c}(t) \cdot \text{Cap}_c & \text{if } (a_{\text{maintenance},c}(t) = 1 \text{ and } a_{\text{prodcution},c}(t) \neq 0)\,, \\ 0 & \text{otherwise.} \end{cases}$$

1.1.5 Demand Penalty

This penalty is incurred if the total supply from production and external purchases does not meet the demand on the current day. The total supply is the sum of the production and external purchase, and if this is less than or greater than the demand, a penalty is imposed proportional to the absolute difference between demand and supply.

$$\text{penalty}_{t,\text{D}} = -\rho_{\text{D}} \cdot |d_t - \text{total\_supply}_t|$$

Therefore, the total penalty at time t can be written as:

$$\text{penalty}_t = \sum_{c \in C} \left(\text{penalty}_{ct,\text{MD}} + \text{penalty}_{ct,\text{MF}} + \text{penalty}_{ct,\text{EM}} + \text{penalty}_{ct,\text{RP}}\right) + \text{penalty}_{t,\text{D}}$$

2 In the Appendix: Environment details

2.1 Plant model and motivation

Energy-intensive chemical processes use Demand Response (DR) to modify their electricity consumption patterns in response to fluctuating electricity prices.

Hence, these industries typically optimize their production schedules on a rolling basis over a certain horizon of available demand and electricity prices. However, optimizing production scheduling alone can be catastrophic, as it fails to consider the inherent condition of the equipment on which production depends. Indeed, recent works have focused on integrating the two, such as integrating condition-based maintenance in the optimization of ASUs [Xenos et al., 2016], as well as for natural gas plants [Huang & Zheng, 2020].

In this we work, we have considered an Air separation unit responsible to meet only aggregated gaseous demand (both gasesous nitrogen, GAN and oxygen, GOX) of its customers. The unit innately depends on its set of three compressors, which have certain capacities which have certain maintenance requirements. Since, demand might exceed the production capacity, possibly becasue of the downtime owing to the maintenance, we consider extnernal product purchase poissble to satisfy the demand. For optimization, we follow a deterministic avaible forecast of 30 days of the demand and electricity prices on rolling basis and consider the episode length of 31 days in the environement. We assume the actions are performed on a daily basis, after whic

2.2 Assumptions

The GASU environment is based on the following assumptions:

- We consider only two operational modes in the plant: "Work" and "Maintenance". Furthermore, in the work mode, we assume a joint production capacity for both GAN and GOX.
- Perfect forecasts are assumed to be available for both demand and electricity prices over the duration of the episode.
- Each compressor has a fixed mean time to failure ($\mathrm{MTTF}_c \in \mathbb{Z}_+$), after which it must undergo maintenance.
- Compressors have a fixed maintenance duration, after which they can be put back into operation. This duration is denoted as the mean time to repair (MTTR_c).
- After maintenance is completed, a compressor must wait for a specified duration before the next maintenance action can be performed. This is referred to as the minimum no-repair duration (MNRD_c). This constraint is important because it allows for the possibility of scheduling early maintenance when a spike in demand is anticipated near the maximum allowable operation time.

2.3 Environment Transition Function

The environment transition function in **GASUEnv** defines how the environment state evolves in response to the agent's actions at each discrete time step $t \in \{1, \ldots, 30\}$. This update is governed by two principal modules:

1. **Plant state update**, which incorporates changes to demand and electricity price signals (perfectly forecasted).

2. **Compressor physical condition update**, which tracks the evolution of maintenance status and compressor readiness based on operational decisions.

Together, these modules determine the updated observation vector $s(t+1)$ from the action vector $a(t)$, following the dynamics detailed below.

2.3.1 Action Interpretation

At each time step t, the agent executes an action $a(t) \in \mathcal{A}$ defined as:

$$a(t) = (a_{\text{maintenance}}(t), a_{\text{production}}(t), a_{\text{purchase}}(t))$$

- $a_{\text{maintenance}}(t) \in \{0,1\}^n$: Maintenance decision per compressor.
- $a_{\text{production}}(t) \in [0,1]^n$: Fractional production rates relative to maximum capacity Cap_c.
- $a_{\text{purchase}}(t) \in [0, 10000]$: Scalar value denoting external product purchase to meet excess demand.

2.3.2 Observation State Update

Let $s(t) = (d_t, e_t, \mathbf{tslm}_t, \mathbf{tlcm}_t, \mathbf{cdm}_t)$ be the observation vector at time t. The transition to $s(t+1)$ is governed by the following procedures:

Update of Exogenous Forecasted Variables From the perfect-forecast arrays:

$$d_{t+1} \leftarrow \texttt{demand\_array}[t+1], \quad e_{t+1} \leftarrow \texttt{price\_array}[t+1]$$

Compressor Physical Condition Transition For each compressor $c \in C$, update:

- **Time Left to Complete Maintenance (TLCM)**: $\text{tlcm}_{ct} \in \mathbb{Z}_+$
- **Time Since Last Maintenance (TSLM)**: $\text{tslm}_{ct} \in \mathbb{Z}_+$
- **Can Do Maintenance (CDM)**: $\text{cdm}_{ct} \in \{0,1\}$

Let MTTR_c be the fixed maintenance duration, and MNRD_c the minimum no-repair duration. The updates are:

$$\text{tlcm}_{c,t+1} = \begin{cases} \text{MTTR}_c - 1, & \text{if } a_{\text{maintenance},c}(t) = 1 \wedge \text{cdm}_{ct} = 1 \\ \text{tlcm}_{ct} - 1, & \text{if } a_{\text{maintenance},c}(t) = 1 \wedge \text{cdm}_{ct} = 0 \\ \text{tlcm}_{ct}, & \text{otherwise} \end{cases}$$

$$\text{tslm}_{c,t+1} = \begin{cases} 0, & \text{if } a_{\text{maintenance},c}(t) = 1 \\ \text{tslm}_{ct} + 1, & \text{otherwise} \end{cases}$$

$$\text{cdm}_{c,t+1} = \begin{cases} 1, & \text{if } \text{tslm}_{c,t+1} \geq \text{MNRD}_c \\ 0, & \text{otherwise} \end{cases}$$

These transitions reflect the maintenance eligibility and reset mechanisms based on countdown timers and repair cycles.

2.3.3 Reward and Cost Computation

The immediate cost incurred at time t is:

$$\text{cost}_t = \texttt{production\_and\_external\_purchase\_cost}(a(t))$$

The reward is defined as:

$$r_t = -\text{cost}_t, \quad R = \sum_{t=1}^{30} r_t$$

2.3.4 Transition Summary

The complete transition from state $s(t)$ to $s(t+1)$ includes:

1. Receiving action $a(t)$
2. Computing operational cost
3. Advancing time $t \leftarrow t + 1$
4. Updating d_{t+1}, e_{t+1}
5. Updating compressor state variables $\text{tslm}_{ct}, \text{tlcm}_{ct}, \text{cdm}_{ct}$
6. Returning $s(t+1)$, reward r_t, and auxiliary info

This transition logic forms the backbone of the environment's dynamics and supports reinforcement learning over industrial scheduling with embedded maintenance constraints.

2.4 MILP counterpart

Define all the variables, parameters, and the optimization model to derive the optimal action.

2.5 Representativeness and Generalizability

The model can be used for independent production scheduling by relaxing the maintenance constraints, allowing for simpler operation-focused planning. Additionally, the model is highly extensible; incorporating additional compressors requires only minor modifications to the compressor configuration in the input JSON file, making the framework scalable to more complex plant setups.

The environment's capability can be readily enhanced to account for uncertainty in demand and electricity price forecasts. This extension would enable the development and benchmarking of robust or stochastic reinforcement learning policies, and constitutes a key direction for future work.

2.6 Limitations

The current environment could be made more realistic and robust by incorporating condition-based maintenance strategies, where the timing of maintenance is informed by compressor health indicators rather than fixed schedules.

Moreover, the current assumption of external gas purchases during supply shortfalls could be replaced by a more detailed process-based mechanism such as the activation of an internal vaporizer, which vaporizes stored cryogenic liquids (liquid oxygen or nitrogen) to meet demand. This would improve the physical fidelity of the environment and allow for more nuanced control strategies.