Optimal Trading of a Fixed Quantity of Power in an Illiquid Continuous Intraday Market

Gilles Bertrand
CORE/UCLouvain
Louvain-la-Neuve, Belgium
gilles.bertrand@uclouvain.be

Anthony Papavasiliou
CORE/UCLouvain
Louvain-la-Neuve, Belgium
anthony.papavasiliou@uclouvain.be

***Abstract*—The recent integration of renewable resources in electricity markets has increased the need for producers to correct their trading position close to real time in order to avoid volatile real-time prices. The last market to close before delivery is the Continuous Intraday Market. Therefore, this market is an interesting outlet for renewable units that aim at covering their forecast errors. As a starting point for tackling this problem, we characterize an optimal policy for trading a fixed quantity in a simplified market model. We use this analytical solution as a basis for developing an Approximate Dynamic Programming algorithm and an alternative Stochastic Dual Dynamic Programming that can trade under a more realistic set of assumptions.**

***Index Terms*—Stochastic optimal control, continuous intraday market, approximate dynamic programming, stochastic dual dynamic programming**

I. Introduction

The Continuous Intraday Market (CIM) has received increasing attention in recent literature [1]. This can be explained by the growing activity that has been observed in the CIM in recent years. For example, the traded quantity in the German CIM has risen from 10 TWh in 2010 to 41 TWh in 2016 [2]. This is a significant increase, even if this market remains less liquid than the day-ahead market (234 TWh exchanged in Germany in 2016 [3]).

Renewable assets face considerable supply uncertainty, and therefore stand to gain by adjusting their position dynamically in the CIM, as more accurate forecast information arrives for their real-time supply. Moreover, trading later in the day also increases opportunities for profitable trades, since bid-ask spreads in CIMs are empirically observed to decrease as we approach real time. These benefits need to be traded off against the fact that the CIM is typically less liquid than earlier forward (e.g. day-ahead) markets. Therefore, there is a counter-balancing interest for a renewable supplier to sell its power earlier, in order to avoid "pushing" the price against its profits by unloading large quantities of supply in a thin market.

In this work we set the foundation for capturing the latter tradeoff (thin markets), and leave matters associated to the uncertainty of supply and the increasing information that is revealed closer to real time for future work, but set in place the algorithmic framework for this extension. Concretely, we focus on developing an optimal trading strategy for selling a fixed quantity of power in an idealized CIM for which we have a model of the price evolution. Our motivation is to set the basis for approximate dynamic programming (ADP) algorithms that can be used for trading the production of a renewable unit in the real CIM without any assumption on the price evolution model.

Our analytical work draws similarities to early work on optimal control by [4]. In this paper, the authors develop an optimal trading strategy in order to trade a fixed quantity within a certain deadline subject to independent random prices. They prove that the optimal strategy is characterized by a threshold beyond which the producer should trade the required quantity. This work has been extended by [5] and [6], where the authors consider that the trader (i) has the option to store the good for a given holding cost, and (ii) faces a deterministic demand at each time period. They prove that the optimal policy still follows a threshold strategy.

In more recent work in the context of electricity markets [7], the authors derive an optimal strategy for a thermal unit trading in the CIM, while assuming that the price follows an additive Brownian motion. In [8], the authors also present the solution for trading a fixed quantity. The difference with our work is that [7] does not account for any bid-ask spread and [8] only considers the case of a constant bid-ask spread. On the contrary, in our work, we solve the problem using a non constant spread in order to reflect the empirical observation that the CIM remains relatively illiquid, despite its growth.

The results of [4], [5] and [6] have recently been used in [9]. In this paper, the authors use policy function approximation in order to optimize the trading strategies of a storage unit in the CIM. For this purpose, the authors define a parametrization for their policy which relies on the insights of [4], [5] and [6], by specifically developing threshold policies and relying on policy function approximation. By analogy, in the present work we develop a value function approximation for the context of a renewable supplier with uncertain supply.

Our contributions can be summarized as follows: (i) We cast the problem of unloading a fixed quantity of power in a CIM, as a Markov Decision Process (MDP). (ii) We characterize the optimal policy as well as the optimal value function for

This research was funded by the ENGIE Chair on Energy Economics and Energy Risk Management, by an Electrabel grant on "Modeling the Value of Flexibility at Sub-Hourly Operating Time Scales", and by the Belgian National Science Foundation (FSR-FNRS) through a FRIA grant.

this MDP. (iii) We use this optimal value function to develop basis functions for an ADP algorithm. (iv) We use the MDP framework to also develop an SDDP algorithm, that can be used as a benchmark for the ADP algorithm. We validate our algorithms by demonstrating that they both arrive to the optimal analytical solution of a 10-period example.

II. Overview of European Continuous Intraday Markets

In this section, we describe the CIM operations of a typical EU market, using the German market as a prototypical example. For a presentation of the positioning of the CIM in German electricity market operations, the reader can refer to [9]. In the German CIM, buy and sell bids can arrive at any moment and can be 'locked in' by market participants who find the bids favourable. Each of these bids is associated with a delivery hour, a type (buy/sell), a price [in Eur/MWh] a quantity [in MWh]. This means that, at any moment of the CIM, a producer observes a collection of bids. This collection of bids is called an order book. This order book can be further split into 24 "hourly" order books[1], one for each delivery hour. We present such an order book in Fig. 1. The order book consists of two parts: (i) the buy side, which contains all the bids of traders which want to buy power from us (in blue), and (ii) the sell side, which contains all the bids of traders which want to sell power to us (in red). In the next section, we explain how we simplify the representation of the order book in order to derive an analytical solution to the optimal trading problem.

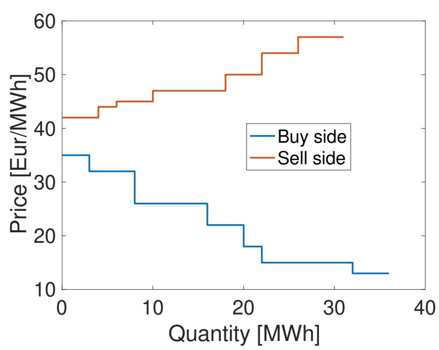

Fig. 1. Example of an order book.

III. Continuous Intraday Market Model

In order to derive an optimal trading strategy, we decompose the order book into three components that are presented in Fig. 2: (i) a bid-ask spread, which is the price difference between the most favorable sell and buy bids that have yet to be matched (illustrated on the left panel); (ii) the center of the bid-ask spread, which is the average price between the best sell bid and the best buy bid (illustrated on the left panel); and (iii) a linear price impact (illustrated on the right panel). We exploit the linearity in the derivation of the optimal trading policy.

[1]In the remainder of the paper, we only consider "hourly" order books.

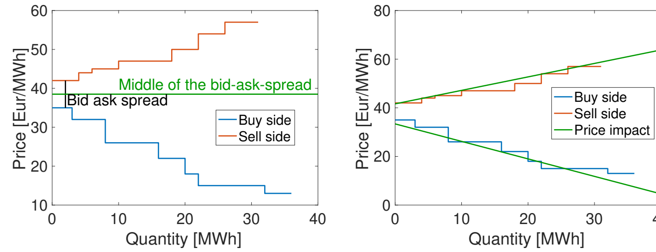

Fig. 2. Illustration of the different components of the orderbook.

This decomposition of the order book at time step t can also be expressed mathematically as:

$$p_t^s(q_t) = p_t - \Delta_t - 2rq_t \quad (1)$$

where (i) $p_t^s(q_t)$ is the marginal price at which we sell quantity q_t; (ii) p_t is the center of the bid ask spread; (iii) Δ_t is half the bid-ask spread; (iv) $2 \cdot r$ is the slope of the linear impact; and q_t is the quantity that we sell. The term $2 \cdot r \cdot q_t$ therefore represents the impact of the producer on the price.

We assume the following about the parameters in Eq. (1):

- p_t follows a stochastic evolution, according to the following model:

$$p_{t+1} = p_t + \epsilon_t \quad (2)$$

  where ϵ_t can follow any distribution respecting $\mathbb{E}[\epsilon_t] = 0$, where the expectation is conditional on the information available in time t.
- The parameter r is assumed to be deterministic. In order to estimate it, we use confidential data from the German CIM for $2015 - 2016$ which has been sourced from the European Power Exchange (EPEX). We use the following strategy, which is initially proposed in [8]. We record the state of the market in different instances. For each of these instances i, we compute the marginal price $p^{s\,i}(q^j)$ that we would obtain if we were selling different quantities of power q^j. For each of these quantities, we compute the average price obtained from the different instances:

$$\bar{p^s}(q^j) = \sum_{i=1}^{I} \frac{p^{s\,i}(q^j)}{I},$$

  where I is the number of instances. These averages are presented on the left panel of Fig. 3. Finally, we use a linear regression to obtain the slope of the red line which is equal to the price impact coefficient $2r$. Estimating the value against the 200 first days of 2015 yields a value of $r = 0.0095$.
- Δ_t is assumed to be deterministic. As for r, we record the state of the market in different instances, although we now separate the instances in different batches, depending on how much time before delivery the instance has been captured. For each of these batches, we compute the average spread. The results of this computation are presented on the right panel of Fig. 3. One important insight from this graph is that, as we arrive closer to

delivery, the bid-ask-spread decreases. This indicates that it can be favorable to wait for market closure to trade power.

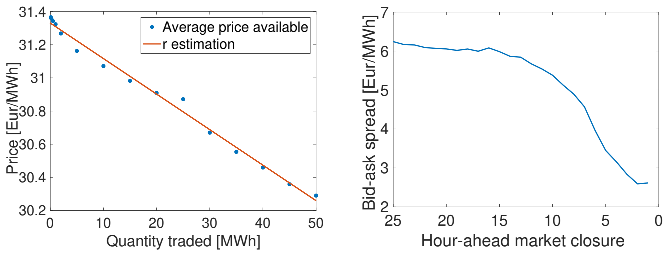

Fig. 3. Estimation of r (left), and expected bid-ask-spread (right).

IV. Analytical Solutions

A. Assumptions

In the analytical derivations of this section, we consider trading a fixed quantity of power. We assume the following:

- We discretize time. This appproach is similar to the one proposed in [9] and [10].
- From one time step to the next one, the order book evolves as described in section III. Notice that we do not make any assumption on the precise distribution of the center of the bid-ask spread. Indeed, as shown in Eq. (7), we only need to characterize the expectation of ϵ_t.
- In order to simplify the analysis, we assume that the closer we are to market closure, the smaller the bid-ask spread is. This is consistent with the results that we obtain from the right panel of Fig. 3.
- We cannot be in imbalance at the end of the CIM. This assumption originates from the German regulation which discourages resources from being in imbalance on purpose [11].
- By considering every delivery period independently, we ignore time coupling effects. These have been considered in previous work by the authors in the context of storage units [9].

B. Modelling the problem as an MDP

We model our problem using the MDP framework. To this aim, we need to define the state space, the action space, the reward function, as well as the state transition function.

1) State space: Our state space contains two variables: (i) s_t, the quantity that still needs to be traded at time step t (for a trader in the CIM, it would correspond to the difference between the quantity that it had to trade initially and the quantity that it has already traded), and (ii) p_t, the center of the bid-ask-spread at time step t.

2) Action space: Our action space consists of q_t, the quantity that we sell at time step t.

3) Reward: The reward at time step t is the earnings obtained from selling q_t in the order book $p_t^s(q_t)$. It is expressed as:

$$
\begin{aligned}
rev(q_t) &= \int_0^{q_t} p_t^s(z)dz \\
&= \int_0^{q_t} (p_t - \Delta_t - 2rz)dz \\
&= p_t q_t - \Delta_t q_t - r q_t^2
\end{aligned}
$$

4) State transition function: The transition function links the state variables at time step $t+1$ with the one at time step t:

$$s_{t+1} = s_t - q_t \quad (3)$$

$$p_{t+1} = p_t + \epsilon_t \quad (4)$$

Eq. (3) describes the evolution of the quantity that still needs to be traded. Eq. (4) corresponds to the price evolution model of Eq. (2).

C. Optimal trading policy

In this section, we derive the optimal decision at each time step, as well as the optimal value function for the case in which we have a positive[2] quantity to sell $s_t \geq 0$. We prove by induction, starting at the last time step T, that the value function and the optimal decision at time step t are characterized by the formula in Table I, where:

- $C_i \doteq \frac{\sum_{j=i+1}^{T} \Delta_j - (T-i)\Delta_i}{2(T-i+1)r}$,
- $X_t \doteq [\frac{(T-t)\Delta_t - \sum_{i=t+1}^{T} \Delta_i}{2r}, \frac{(T-t+1)\Delta_{t-1} - \sum_{i=t}^{T} \Delta_i}{2r}[$.

1) Time step T: As explained in section IV-A, at the last time step we have to cover our position. Therefore, the decision is $q_T = s_T$, and the associated value function is given by:

$$V_T^*(s_T, p_T) = p_T s_T - \Delta_T s_T - r s_T^2 \quad (5)$$

2) First step of the induction: We first derive the optimal solution if we are one time step before delivery. The value function is given by:

$$
\begin{aligned}
& V_{T-1}^*(s_{T-1}, p_{T-1}) = \max_{0 \leq q_{T-1} \leq s_{T-1}} p_{T-1} q_{T-1} - \Delta_{T-1} q_{T-1} \\
& - r(q_{T-1})^2 + \int_{-\infty}^{\infty} V_T^*(s_T, p_T) f(p_T) dp_T \\
& = \max_{0 \leq q_{T-1} \leq s_{T-1}} p_{T-1} q_{T-1} - \Delta_{T-1} q_{T-1} - r(q_{T-1})^2 \\
& + \int_{-\infty}^{\infty} \left(p_T s_T - \Delta_T s_T - r s_T^2\right) f(p_T) dp_T \quad (6) \\
& = \max_{0 \leq q_{T-1} \leq s_{T-1}} p_{T-1} q_{T-1} - \Delta_{T-1} q_{T-1} - r(q_{T-1})^2 \\
& + \mathbb{E}[p_T] s_T - \Delta_T s_T - r s_T^2 \\
& = \max_{0 \leq q_{T-1} \leq s_{T-1}} p_{T-1} q_{T-1} - \Delta_{T-1} q_{T-1} - r(q_{T-1})^2 \\
& + (p_{T-1} + \mathbb{E}[\epsilon_{T-1}])(s_{T-1} - q_{T-1}) - \Delta_T (s_{T-1} - q_{T-1}) \\
& - r(q_{T-1})^2 + 2r s_{T-1} q_{T-1} - r s_{T-1}^2 \quad (7) \\
& = \max_{0 \leq q_{T-1} \leq s_{T-1}} \Delta_T (q_{T-1} - s_{T-1}) - \Delta_{T-1} q_{T-1} \\
& - 2r(q_{T-1})^2 + p_{T-1} s_{T-1} + 2r s_{T-1} q_{T-1} - r s_{T-1}^2
\end{aligned}
$$

[2] The case with $s_t < 0$ can be computed similarly.

TABLE I
SUMMARY OF THE VALUE FUNCTION AND OPTIMAL DECISION FOR STEP t.

Range of quantity s_t to be traded	X_T	$\cdots$	X_{t+1}	$R^+\backslash(X_T \cup \cdots \cup X_{t+1})$
Optimal quantity to trade q_t^*	0	$\cdots$	0	$\frac{s_t}{T-t+1} + \frac{\sum_{i=t+1}^{T} \Delta_i - (T-t)\Delta_t}{2(T-t+1)r}$
Value function $V_t^*(s_t, p_t)$	$V_T^*(s_t, p_t)$	$\cdots$	$V_{T-t+1}^*(s_t, p_t)$	$p_t s_t - \left(\sum_{i=t}^{T} \Delta_i\right) \frac{s_t}{T-t+1} - \frac{r s_t^2}{(T-t+1)} + \sum_{i=t}^{T-1} \frac{(T-i+1)r}{T-i} C_i^2$

For Eq. (6), we use the definition of the value function (Eq. (5)). In Eq. (7), we use the fact that $\mathbb{E}[p_T] = p_{T-1} + \mathbb{E}[\epsilon_{T-1}]$ and the transition function of s_t defined in Eq. (3). As this objective function is quadratic, we can compute the maximum if we would be ignoring the constraints. To this aim, we compute the point at which the gradient vanishes:

$$\begin{aligned} 0 &= \Delta_T - \Delta_{T-1} - 4rq_{T-1}^* + 2rs_{T-1} \\ 4rq_{T-1}^* &= \Delta_T - \Delta_{T-1} + 2rs_{T-1} \\ q_{T-1}^* &= \frac{s_{T-1}}{2} + \frac{\Delta_T - \Delta_{T-1}}{4r} \end{aligned}$$

There are three cases to consider for this maximum:

- **The maximum is feasible,** $0 \leq q_{T-1}^* \leq s_{T-1}$**:** This condition is equivalent to $\frac{\Delta_{T-1} - \Delta_T}{2r} \leq s_{T-1}$. In this case, we have the optimal decision. The associated value function is given by the following expression, in which we fix $C_{T-1} = \frac{\Delta_T - \Delta_{T-1}}{4r}$:

$$\begin{aligned} &V_{T-1}^*(s_{T-1}, p_{T-1}) = \Delta_T(-\frac{s_{T-1}}{2} + C_{T-1}) \\ &- \Delta_{T-1}(\frac{s_{T-1}}{2} + C_{T-1}) - 2r(\frac{s_{T-1}}{2} + C_{T-1})^2 \\ &+ p_{T-1}s_{T-1} + 2rs_{T-1}(\frac{s_{T-1}}{2} + C_{T-1}) - rs_{T-1}^2 \\ &= \Delta_T(-\frac{s_{T-1}}{2} + C_{T-1}) - \Delta_{T-1}(\frac{s_{T-1}}{2} + C_{T-1}) \\ &- \frac{r}{2}s_{T-1}^2 - 2rC_{T-1}^2 + p_{T-1}s_{T-1} \\ &= -\frac{\Delta_T}{2}s_{T-1} - \frac{\Delta_{T-1}}{2}s_{T-1} - \frac{r}{2}s_{T-1}^2 + p_{T-1}s_{T-1} \\ &+ C_{T-1}(\Delta_T - \Delta_{T-1} - 2rC_{T-1}) \\ &= -\frac{\Delta_T}{2}s_{T-1} - \frac{\Delta_{T-1}}{2}s_{T-1} - \frac{r}{2}s_{T-1}^2 \\ &+ p_{T-1}s_{T-1} + 2rC_{T-1}^2 \end{aligned}$$

- $q_{T-1}^* < 0$**:** This condition is equivalent to $s_{T-1} \leq \frac{\Delta_{T-1} - \Delta_T}{2r}$. In this case, we observe that the derivative is always negative, because (i) $\Delta_T - \Delta_{T-1} + 2rs_{T-1}$ is negative, and (ii) $-4rq_{T-1}$ is negative for $q_{T-1} \geq 0$. Therefore, it is optimal to have $q_{T-1}^* = 0$. This case represents the situation in which the spread at the last time step is significantly smaller than the one at time step $T-1$ ($\Delta_{T-1} >> \Delta_T$). Therefore, it is optimal to trade the entire remaining quantity at the last time step. The associated value function is given by:

$$V_{T-1}^*(s_{T-1}, p_{T-1}) = p_{T-1}s_{T-1} - \Delta_T s_{T-1} - rs_{T-1}^2$$

- $q_{T-1} > s_{T-1}$**:** This solution can never occur because, as assumed in section IV-A, $\Delta_T \leq \Delta_{T-1}$.

In conclusion, we have a different value function formula and optimal decision depending of the value that still needs to be traded (for $s_{T-1} \leq \frac{\Delta_{T-1} - \Delta_T}{2r}$ and for $s_{T-1} > \frac{\Delta_{T-1} - \Delta_T}{2r}$). This is summarized in Table II.

3) General step of the induction: In order to conclude the proof, we need to prove that, if the value function and the optimal solution verify the format of Table I at time step t, it will also be the case at time step $t-1$. The proof for this argument is provided in the electronic supplement[3]. In the electronic supplement, we also present an example of the analytical solution using the true parameters estimated in section III.

D. Insights from the analytical solution

In this section, we present the insights that we gain from the analytical solution.

a) Optimal quantity to trade: The optimal quantity to trade at time step t (there are still $T-t+1$ chances to trade) is given by:

$$\frac{s_t}{T-t+1} + \frac{\sum_{i=t+1}^{T} \Delta_i - (T-t)\Delta_t}{2(T-t+1)r}.$$

If we would consider only the first term, it would imply that we trade $\frac{1}{T-t+1}$ of the capacity available. The second term is a correction for the difference in spread between the different time steps. This term is negative, which means that we always trade at most $\frac{s_t}{T-t+1}$. We can analyse two extreme cases (i) $r \to 0$**:** This means that the second term is very negative. In this situation, the optimal decision is to always trade 0 until the last time step. This results from the fact that the price impact is negligible, and therefore we can trade all the power when the spread is the lowest. (ii) *The spread is constant:* This means that the second term is equal to 0. In this case, it is optimal to trade the same quantity at each time step. This can be interpreted as follows: As the spread is constant, there is no reason to prefer one time step over another and therefore we simply aim at minimizing our impact on the price.

b) Intuition about the value function: The optimal value function is presented in Table I. From this table, we observe that for different ranges of power to sell, s_t, we have different expressions for the value function. These different ranges represent different time steps over which we are required to trade. For instance, (i) X_T represents the case for which we have a small quantity of power to sell. In this case, we can ignore the price impact and trade all the power at the last time step where the bid-ask-spread is the lowest. (ii) X_{T-1}

[3]The electronic supplement is available at the following link: https://sites.google.com/site/gillesbertrandresearch/publications/app-powertech-2021

TABLE II
SUMMARY OF THE VALUE FUNCTION AND OPTIMAL DECISION FOR STEP $T-1$.

Range of quantity s_{T-1} to be traded	$[0, \frac{\Delta_{T-1}-\Delta_T}{2r}[$	$[\frac{\Delta_{T-1}-\Delta_T}{2r}, \infty[$
Optimal quantity to trade q^*_{T-1}	0	$\frac{s_{T-1}}{2} + \frac{\Delta_T-\Delta_{T-1}}{4r}$
Value function $V^*_{T-1}(s_{T-1}, p_{T-1})$	$(p_{T-1} - \Delta_T - rs_{T-1})s_{T-1}$	$\left(p_{T-1} - \frac{\Delta_T}{2} - \frac{\Delta_{T-1}}{2} - \frac{r}{2}s_{T-1}\right)s_{T-1} + 2rC^2_{T-1}$

represents the case for which we have more power to sell. In this situation, we would be affected more significantly by the price impact, and we therefore split sell between the two last time steps.

c) Mathematical format of the value function: As explained before, the value function is a piecewise function. For each of these pieces, the value function can be decomposed as (i) a quadratic function in s_t, and (ii) a bilinear term in $p_t s_t$. We exploit this observation in the algorithmic section.

V. ALGORITHMIC APPROACHES

We now use the modeling setup and the insights of the analytical solution as a basis for an ADP and SDDP algorithm.

A. Approximate dynamic programming

Finding the optimal solution to an MDP is equivalent to finding the action-value function which verifies the Bellman optimality equation [12]:

$$q^*(s,a) = \mathbb{E}[R_{t+1} + \max_{a'} q^*(S_{t+1}, a')|S_t = s, A_t = a]$$

where (i) $\mathbb{E}[R_{t+1}]$ is the expected reward obtained after applying action a in state s, and (ii) $\max_{a'} q^*(S_{t+1}, a')$ is the value function associated to the best possible action in state S_{t+1}.

In our problem, we face continuous state and action spaces. In order to make the problem tractable for ADP, we discretize the action space. This discretization of the action space is needed because the ADP algorithm requires finding the action associated with the best action-value function.

In order to develop an ADP algorithm, we parametrize the action-value function $q^*(s,a)$ as $\hat{q}(s,a;w)$, where w is a set of parameters that we need to optimize.

In order to determine basis functions for the value function, we exploit the insights from the analytical solution. By definition of the action-value function $Q_t(s_t, p_t, q_t)$ [12], we know that:

$$Q_t(s_t, p_t, q_t) \doteq p_t q_t - \Delta_t q_t - r(q_t)^2 + V^*_{t+1}(s_t - q_t, p_t)$$

where $V^*_{t+1}(s_t - q_t, p_t)$ is the optimal value function described in section IV-C. The parametrization of this action-value function can therefore be split into two parts:

$$\hat{q}((s_t, p_t), q_t; w) = f_0(s_t, p_t, q_t; w_0) + \sum_{j=1}^{T} f_j(s_t - q_t, p_t; w_j) \cdot \mathbf{1}_j$$

In this expression:

- f_0 contains the parameters for the payoff obtained at time step t which can be parametrized as:
$$f_0(s_t, p_t, q_t; w_0) = w_{0,0} \cdot (p_t - \Delta_t) \cdot q_t + w_{0,1} \cdot q_t^2.$$
- f_j is the basis function that represents $V^*_j(s_t - q_t, p_t)$. As we know that this function is piecewise quadratic in $s_t - q_t$ and has a bilinear term in $p_t(s_t - q_t)$, we can parametrize it as:
$$f_j(s_t - q_t, p_t; w_j) = w_{j,0} + w_{j,1} \cdot (s_t - q_t) + w_{j,2} \cdot (s_t - q_t)^2 + w_{j,3} \cdot p_t(s_t - q_t).$$
- $\mathbf{1}_j$ is an indicator function which evaluates to 1 if f_j should be used and 0 otherwise. Based on Table I, there are two cases in which f_j should be used (i) $t < j$ and $s_t - q_t \in X_j$, and (ii) $t = j$ and $s_t - q_t \in R^+ \backslash (X_T \cup \cdots \cup X_{j+1})$.

Having defined basis functions, the intuition of the algorithm is to optimize the w parameter in order to minimize the error between the prediction from our value function $\hat{q}(S_t, A_t; w_t)$ and the obtained outcome from the episode G_t. This error can be written as:

$$L = (G_t - \hat{q}(S_t, A_t; w_t))^2 .$$

We minimize this loss function using a stochastic gradient algorithm. Concretely, we run episodes of our trading problem following an $\epsilon-$greedy policy derived from our action-value function $\hat{q}(S_t, A_t; w)$. When the episode is finished, we update w using the following update:

$$w_{t+1} = w_t + \alpha_t \left[G_t - \hat{q}(S_t, A_t; w_t)\right] \nabla \hat{q}(S_t, A_t; w_t).$$

As we use a parametrization derived from the optimal value function, we know that a particular set of weights w gives the optimal value function (computed analytically). Moreover, as the parametrization is linear in w, we are guaranteed to obtain the best possible parametrized policy [13]. This means that this algorithm is guaranteed to converge to the optimal value function when applied on the simplified market model. Nevertheless, the problem of trading in the real-market will typically not respect the model of section 3 and, therefore, this guarantee will no longer hold.

B. Stochastic Dual Dynamic Programming

SDDP is a method for solving a specific class of multi-stage stochastic convex programs. For a complete description of the algorithm, the reader can refer to [14]. In order to use SDDP, we need to define the subproblem faced at every stage. This subproblem, referred to as the Nested L-Shaped Decomposition Subproblem (NLDS), is defined for every time step t and node k of the uncertainty model. In our case, the NLDS is expressed as:

$$\max_{q_t^s \geq 0, q_t^b \geq 0} (p_{t,k} - \Delta_t - rq_t^s)q_t^s - (p_{t,k} + \Delta_t + rq_t^b)q_t^b$$
$$s_t = s_{t-1} - q_t^s + q_t^b$$

There are two specificities compared to the classical problems solved using SDDP. (i) Our objective function is quadratic in q_t^s and q_t^b. These terms can be linearized using a first-order Taylor approximation [15]. (ii) The uncertain parameter $p_{t,k}$ appears in the objective function and follows an auto-regressive process, based on our model for the price evolution (Eq. (2)). This is not suitable for classical implementations of SDDP [16] because the value function is convex in s_t and concave in p_t Therefore, we use the modified version of the algorithm that is presented in [16] and implemented in toolbox [17]. As explained in [16], this algorithm is guaranteed to converge almost surely to the optimal solution.

VI. Case study

In this section, we compare the numerical results obtained by the analytical solution, ADP and SDDP for a problem with 10 time steps. We present, on the left panel of Fig. 4, the error in the value function computed by ADP and SDDP compared to the value function from the analytical solution. We observe that these differences are very small (around 1 euro) compared to the magnitude of the value function (around 1000 euros). We also present on the right panel of Fig. 4, the evolution of the quantity s_t that still needs to be traded for each time time step. We observe that the decisions that are reached by SDDP and ADP are very close to the ones obtained by the analytical solution. This confirms that both methods are able to solve this simplified problem and can therefore be considered for the real CIM trading problem.

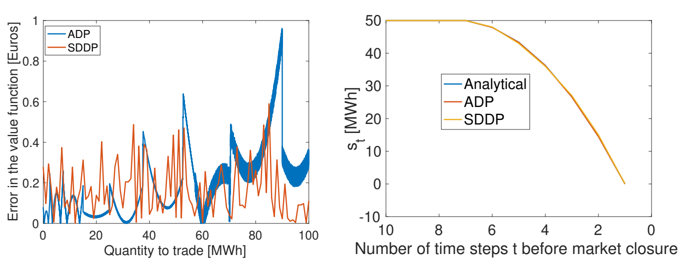

Fig. 4. Left: difference of ADP and SDDP value functions with the analytical solution. Right: quantity that still needs to be traded, s_t, with an initial quantity of 50MWh.

VII. Conclusion and perspectives

In this paper, we model the problem of selling a fixed quantity of power in a simplified model of the CIM using the MDP framework. We derive the optimal trading strategy for this problem through backward induction. We use the optimal value function to develop basis functions for an ADP algorithm. We then develop ADP and SDDP algorithms for testing against our idealized CIM model and demonstrate that they both arrive to the optimal policy on a 10-step example. In future work, we aim at enriching this algorithmic framework with renewable supply uncertainty, in order to also tackle the risk-information trade-off faced by a renewable supplier trading in the CIM. We are also interested in relaxing the assumption that we have access to a model for the price evolution.

References

[1] P. Shinde and M. Amelin, "*A Literature Review of Intraday Electricity Markets and Prices,*" 2019 IEEE Milan PowerTech, Milan, Italy, 2019, pp. 1-6

[2] H. Martin and S. Otterson, "*German Intraday Electricity Market Analysis and Modeling Based on the Limit Order Book,*" 15th International Conference on the European Energy Market (EEM), Lodz, 2018.

[3] EPEXSPOT, "*Trading on EPEX SPOT 2020,*" [Online]. Available: https://www.epexspot.com/sites/default/files/download_center_files/20-01-24_TradingBrochure.pdf

[4] W. Morris, "*Some analysis of purchasing policy,*" in Management Science, vol. 5, no.4, pp. 443-452, Jul. 1959.

[5] B. Kingsman, "*Commodity purchasing,*" in Journal of the Operational Research Society, vol. 20, no. 1, pp. 59-79, Mar. 1969.

[6] K. Golabi. "*Optimal inventory policies when ordering prices are random,*" in Operations Research, vol. 33, no. 3, pp. 575-588, 1985.

[7] R. Aid, P. Gruet and H. Pham, "*An optimal trading problem in intraday electricity markets,*" in Mathematics and Financial Economics, vol. 10, no. 1, pp. 49-85, Jan. 2016.

[8] A. Cartea, S. Jaimungal and J. Penalva, "*Algorithmic and high-frequency trading,*" in Cambridge University Press, 2015.

[9] G. Bertrand and A. Papavasiliou, "*Adaptive Trading in Continuous Intraday Electricity Markets for a Storage Unit,*" in IEEE Transaction on Power Systems, vol. 35, no. 3, pp. 2339-2350, Dec. 2019.

[10] A. Skajaa, K. Edlund and J. Morales, "*Intraday Trading of Wind Energy,*" in IEEE Transactions on Power Systems, vol. 30, no. 6, pp. 3181-3189, Nov. 2015.

[11] TENNET [Online]. Available: https://www.tennet.eu/fileadmin/user_upload/The_Electricity_Market/German_Market/Grid_customers/contracts/BNetzA-BKC_englisch.pdf

[12] R. Sutton and A. Barto, "*Reinforcement Learning: An Introduction,*" in MIT press, 2018.

[13] J. Tsitsiklis and B. Van Roy, "*An analysis of temporal-difference learning with function approximation,*" in IEEE Transactions on Automatic Control, vol. 42, no. 5, pp. 674-690, May 1997.

[14] M. Pereira and L. Pinto, "*Multi-stage stochastic optimization applied to energy planning,*" in Mathematical Programming, vol. 52, pp. 359-375, 1991.

[15] D. Bertsimas et al. "*Adaptive robust optimization for the security constrained unit commitment problem,*" in IEEE Transactions on Power Systems, vol. 28, no. 1, pp. 52-63, 2012.

[16] A. Downward, O. Dowson and R. Baucke, "*Stochastic dual dynamic programming with stagewise dependent objective uncertainty,*", in Operations Research Letters, vol. 48, no. 1, pp. 33-39 , Jan. 2020.

[17] O. Dowson and L. Kapelevich, "*SDDP.jl: a Julia package for stochastic dual dynamic programming,*", in INFORMS Journal on Computing, 2020.