Master Degree Thesis

Master's Degree in Automatic Control and Robotics (MUAR)

Efficient Management of Energy Systems including Storage Systems

July 6, 2023

Autor: Ce Xu Zheng

Directors: Vicenç Puig Cayuela

Ramon Costa Castelló

Convocatòria: 06/2023

Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona

Summary

In this master thesis, we are going to work towards a data based control algorithm, that given a complex system and a goal, it can learn how to manipulate the control actions of the system towards achieve that goal. The objective of developing such a control algorithm is to ensure the correct operation of a small smart microgrid with several components that balances the operational costs under optimal conditions and the economical performance of the system.

Particularly, in this thesis, we will describe the components that conforms a standard microgrid including Energy Storage Systems (ESS) considering the efficiency and operational limits. Energy Generation Systems that relies on climatic condition cannot be on demand. Consumption hubs, like a regular household, whose consumption of electrical power depends on the daily habits can change from time to time.

The composition of a microgrid of a "*prosumer*" - generally consumers of energy from the electrical grid with production and storage capabilities - usually relies in several types of ESS with complementary characteristics like a battery with higher storage capacity and a super capacitor with higher power density. The production must have at least one energy source, that is usually a renewable source, but the combination of several sources could increase the reliability of the system.

At this level of management, the system can be considered as a set of ESS whose State of Charge (SOC) can be considered the working space and can be leveled with the control actions. This type of management is traditionally done with an economical criterion, whose optimization requires the definition of an accurate model of the system and the correct parametrization of the goals. We propose a different method that can be used without the previous knowledge of the system dynamics and can simultaneously generate control signals and learn the optimal policy given a parametrized cost function and the sensing of the system states.

The proposed set of methods are called Structured Online Learning methods, and relies in two distinctive parts: the System Identification module, that will learn the dynamics of the system given the collected data as a linear combination of non-linear basis functions of the state; and a Value Function learning that will be updated given the learned model and can generate the control signal that is showed optimal in a long term window.

As those systems require the definition of a differentiable and convex cost over the control effort, we introduced some assumptions over the operational cost that can adapt the economical problem to an equivalent of a Quadratic Regulator. This will also bring us the opportunity to compare our method with some standard solutions like the Ricatti equation for a Linear Quadratic Regulator.

Comparing the standard solutions for the adapted problem that leverages the knowledge of the real model and several versions of the Structured Online Learning algorithm, we can confidently state that the proposed method generates comparable results in a reference tracking problem. But, it can even enhance the economical performance if the generation and consumption profiles has a certain degree of predictability.

Contents

List of Figures

List of Tables

Acronyms

AMI Advanced Metering Infrastructure.

ARE Algebraic Ricatti Equation.

DER Distributed Energy Resources.

EMPC Economic Model Predictive Control.

ESS Energy Storage Systems.

FF Feed Forward.

GD Gradient Descend.

HJB Hamilton Jacobi Bellman equation.

IMP Internal Model Principle.

LQR Linear Quadratic Regulator.

LS Least Squares.

LTI Linear Time Invariant.

MDP Markov Decision Process.

MPC Model Predictive Control.

ODE Ordinary Differential Equation.

PV Photo Voltaic.

RL Reinforcement Learning.

RLS Recursive Least Squares.

SINDy Sparse Identification of Non-linear Dynamics.

SOC State of Charge.

SOL Structured Online Learning-Based Control.

TD Temporal Difference.

VF Value Function.

1 Introduction

1.1 Motivation

We are currently facing a change of paradigm in the energy generation and consumption habits. The existing power units are outdated and approaching their technical end-of-life. This situation coincides with the growing emphasis on decarbonization and ecological transition as part of the agenda from most of the European regulators.

In response, the trend is shifting towards introducing a promoting renewable sources for the energy production, that are generally more extensive than fueled power plants. In this context, there is a general consensus about the prevalence of a Distributed Power Generation (DPG) scenario, where a significant portion of energy is derived from a diverse range of renewable energy sources located near consumption hubs and the introduction of the *"prosumers"*, that acts as productors and consumers of the energy resources simultaneously. This decentralized setup not only allows for local generation and consumption of power but also enhances resilience against disruptions to the main energy grid.

In this scenario, the consumers could also be tempted to introduce energy storage elements that can reduce the costs in the long term by buying and selling to the grid according to the electricity price of the grid. With that comes the optimization problem of managing the states of charge of the household batteries according to the demand and production forecasts, and the price of the grid electricity.

However, this transition to a renewable energy-based mix introduces certain challenges. The generation of electricity becomes unpredictable due to the intermittent nature of renewable sources, and there can be a mismatch between generation and consumption times. For instance, the hours of peak photovoltaic production do not always align with the highest electricity demand from the population. This discrepancy needs to be addressed through the use of Energy Storage Systems, which can absorb excess energy during periods of high production and release it when demand is higher.

Regarding this complex problem, most of the advanced control methods that can assure some degree of robustness and a good performance in the economical sense relies on an accurate modeling of the microgrid and some optimization based control within a time window. Most of the research regarding these methods analyze the heuristics associated to the modeling of the production and consumption profiles along the day [SLP+22], and deals with it with different strategies such as the sensitivity approach, the stochastic approach and the robust optimization [Nas20].

However, there are many novel proposals that follow the new trends of learning a control law based uniquely in sources of data. Among them there are several proposals that are based in Reinforcement Learning techniques, that formulates the problem as a Markov Decision Process (MDP) and maximizes the expected reward only by knowing the states and the reward at each time such as in [AA22], [JWX+19]. And other methods look for analytical solutions that may optimize iteratively the control law given each point of the state space using functions approximators.

The main advantage of those later solutions relies on the "plug and play" implementation of the methods, that are usually more generalizable and, ideally, no prior knowledge of the particular

system must be necessary.

1.2 Project Objectives

The objectives of this thesis will involve developing control algorithms to improve the benefits that energy storage components, as they can move the produced energy in time to the moments when it is needed or even trade with the grid to produce benefits.

It is commonly assumed that there are several levels of hierarchy that contribute to the overall energy management system. From the smaller systems that are considered the lowest level of control, and considering more simple control objectives. Up to the biggest ones, that gathers the lower level systems and tends to consider more varied objectives and in a longer time horizon like the grid level.

The lower level system is the component level, where the main focus of the problem addresses challenges such as non-linearities, efficient management, safe and reliable behavior of components, and estimation of State of Charge. These issues have been extensively studied by a large community of investigators.

The next level of control hierarchy is the microgrid level. At this state, we can simplify the considerations regarding the individual components, and consider its dynamics solved b their controllers. This will allow to gather several components of varied types with different characteristics and make them work towards a higher objective jointly. There is a growing interest in integrating intermittent renewable energy sources into microgrids. This integration presents significant challenges in terms of reliable operation and control. In the literature, two main families of control approaches can be identified: centralized and decentralized. The centralized approach relies on a central controller, while the decentralized approach facilitates distributed decision-making among various units within the microgrid. This decentralized approach enhances scalability and resilience of control algorithms.

At the grid level, the presence of prosumers (consumers with production and storage capabilities) has sparked extensive research efforts to explore potential evolution of electricity markets and decentralized energy management mechanisms. These mechanisms aim to enable active participation of prosumers in the energy supply. The current energy system is witnessing significant changes, including the increased penetration of small-scale production units, particularly those utilizing renewable energy sources. In this context, distributed methods are employed to decompose the problem, allowing multiple agents to collaboratively reach agreements through iterative processes.

Within those levels of hierarchy, we focused on the microgrid level, where is not needed to model the dynamics of the components, but the goal is to manage the State of Charge (SOC) of the storage elements and the power that will be obtained from the grid, the storage elements and the energy sources.

In order to do it, there are many approaches to solve the management problem. The most popular ones are based on the Economical Model Predictive Control (EMPC) methods, that given a model of the microgrid and a definition of the costs that are involved in the operation of the microgrid, it will minimize numerically the computed costs within a time window. These methods can be very accurate for the operation, and can have intrinsically built in several safety measures like the limits of operation of the system. But they require an accurate model that can

predict the behavior along the future time window and be efficient enough so the solution the optimization problem can be solved in real time.

Our main goal in this thesis is to introduce new methods that can control this type of systems without the need of an accurate model. Most of these methods are not as settled as the previous ones, and they often can not assure an optimal control in the costs terms or satisfy the safety constraints. Regarding those issues, there are some proposals with adaptive models control whose parameters are updated as proposed in [GZ20].

Our study will focus on the Structured Online Learning set of algorithms. They are based on the Value Function basis that describes the optimal cost that can be achieved from each point of the state space. This approach pretends to showcase a method to learn the control law while is performing the control action simultaneously.

1.3 Structure

This project will be structured in four main sections. In the second chapter [2], we will describe the elements that will take part on the microgrid, including the Energy Storage Systems (ESS), the energy sources and the household consumption profile.

In the third cahpter [3], we will describe the theoretical basis of the SOL algorithm, than includes the definition of the Value Function, the Bellman equation and the application of the Hamilton-Jacobi Bellman equation to define the optimal control signal in function of the Value Function. After that, there is a description of some System identification algorithms that can learn the parameters of the model as function of a set of nonlinear functions. These will be leveraged to adjust the Value Function and learn online the optimal control.

The next chapter [4] will be devoted to the statement of the problem that we finally solved. In this section, we identify the components of the microgrid and the configuration that it will take. Then, we describe the hypothesis that allow us to resemble the Economical MPC problem into a tracking problem by adjusting some of the parameters and under some assumptions.

Chapter [5] will present the results that we obtained under different circumstances such as the simple tracking problem, some simplified profiles of production and consumption of energy, and finally, the application on a system with real data. All of this with the corresponding discussion about the obtained results and some Key Performance Indicators (KPIs).

After that, it will only remain some final conclusions about the project and a short dissertation regarding the Economical, Social and Environmental impact.

2 Problem statement

2.1 Smart Grids

In the context of power grids, a smart grid refers to an advanced, modernized infrastructure that incorporates digital communication and information technologies to improve the efficiency, reliability, sustainability and resilience of electricity generation, distribution and consumption. It leverages real-time monitoring, control systems and intelligent automation to optimize grid management and operation [CI12], [Sha20].

Some of the characteristics of a smart grid are:

1. **Advanced Metering Infrastructure (AMI)**: Smart meters are installed at consumers' premises to enable two-way communication between the utility and consumers. This facilitates real-time monitoring of electricity consumption, remote meter reading and demand response programs.

2. **Grid Monitoring and Control**: Enabled by the AMI, smart grids technologies are capable of real-time monitoring of various grid parameters, such as voltage, current, and power flow. This allows early detection of faults, better outage management and improved control of grid operation.

3. **Distributed Energy Resource (DER) Integration** : Smart grids facilitate the integration of renewable energy sources, energy storage systems, electric vehicles and other decentralized energy resources. This allows for better utilization of clean energy and increases grid flexibility.

Within the context of smart grids, our primary focus will be on the energy management challenges faced by households. This involves addressing various aspects, such as defining the electrical components essential to the system, including the Energy Storage System (ESS), Distributed Energy Resources (predominantly utilizing renewable sources), and household consumption.

2.2 Energy Storage Systems

Ideally, a perfect energy storage system (ESS) should have a high energy density [DMBB23]. This allows them to store high amounts of energy in small and light devices. In addition, it must have a high power density, what allows the system to absorb all the power that reaches it and supply as much power as the user needs. This property enables fast charging and discharging of energy. Furthermore, it must be efficient, as there should be no energy losses during the charging and discharging processes. Ideally, it should also have a long service life that makes it reliable for use over a long period of time and cost-effective.

Unfortunately, this type of ESS is not yet available at present and, to meet our storage needs, we will consider two types of ESS: batteries, with a higher and cheaper storage capacity, and supercapactiors (SC), with much higher power restrictions on input and output, but more expensive.

The batteries are ESS that work through chemical reactions inside the material. This allows them to store more energy in smaller and lighter systems leveraging the chemical bounds. But it also comes with the limitation of a slower charge and discharge rate due to the speed at which the component reactions can take place. To model these limitations, Internal Resistance is often

used, which will limit the current the battery can absorb and deliver given the voltage, and effectively limiting the amount of energy it can take in. Moreover, this model can also simulate the amount of energy that is lost during the charging of a battery cell in the form of heat, which will also depend on the power supplied.

Regarding battery degradation, the lifetime of the chemical components has improved a lot since the first technologies, but it is still one of the major handicaps of this type of ESS. Gradual degradation of both energy density and power density is inevitable, but has been shown to be faster if operated at temperatures outside the specified range and at higher peak power operations [DMBB23].

In the other hand, the Super Capacitors (SC) are based on a physical properties of the system instead of chemical ones. The voltage difference between two conducting surfaces tends to accumulate ions until reaching the electrical equilibrium. This phenomenon relies on the high capacitance of the systems (coefficient between the voltage difference and the stored energy between the surfaces) and can depend on many factors of the components such as the geometry and the medium in which the elements are immersed.

This provides a limited energy density to this type of ESSs, but it has no intrinsic physical limitations in the power that it can absorb or deliver, what indicates a very high power density. Regarding the efficiency of the SC operation, it might depend on many factors such as temperature and the voltage. But, it usually tends to be more efficient than chemical batteries.

The degradation of this type of systems is not related to any physical phenomenon, but from the chemical reactions that high voltage tension can trigger between several components and the medium. This implies that the lifespan of a SC ESS tends to be much higher and almost independent on the number of charging cycles [Phi22].

2.3 Distributed Energy Resources

The deployment of Distributed Energy Resources (DERs) in regular households commonly includes two widely adopted sources: wind and solar energy. Both sources offer unique advantages, such as clean production and the decreasing costs per Watt-hour over time, as illustrated in Figure 2.1. However, it is important to acknowledge certain challenges associated with these sources. The unpredictability of wind and solar energy production must be taken into account, as it is influenced by meteorological conditions, which are beyond the control of the producer.

Between these two sources, we are going to develop more in depth the solar energy. This decision is favored by many facts, solar energy is safer and easier to install in a common neighborhood due to the lack of moving parts [FOF13]. The cost per Watt-hour is becoming lower in the later years, the lower maintenance that is is usually required and still achieving an optimal energy output. And most importantly for our controlling purpose, the more predictability and regularity that solar energy production can hold [CCR+20] compared with the irregular production of wind turbines [LTS+19],[MVG20].

The availability of energy from the sun varies on a daily basis, influenced by factors such as the position of the sun, and the quantity of solar irradiance. These cyclic patterns determine the amount of energy that can be harnessed. Sunlight tends to be strongest when the sun is at its highest point in the sky, around noon, as the path through the atmosphere is shorter. Additionally, seasonal variations arise from the tilt of the Earth's axis in relation to its orbit

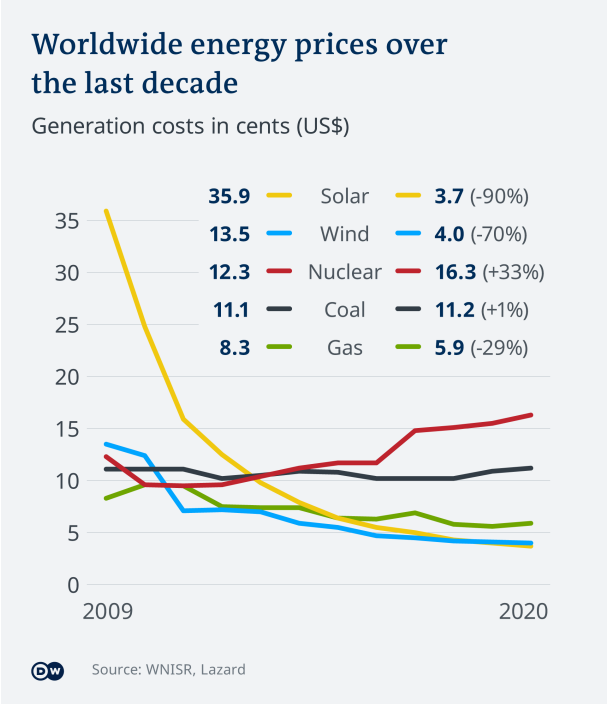


Figure 2.1: Energy costs by sources. Image obtained from [Rue21]

around the sun. As stated in [CCR+20], the solar irradiance can very well modeled in a cloudless sky on the surface of the Earth

$$I_{cs} = bG_{b,\tau}e^{-0.09m(T_L-1)}. \tag{2.1}$$

The factor correction, denoted as b, accounts for adjustments relative to the extraterrestrial irradiance. It helps compensating for variations or discrepancies between the measured irradiance and the idealized extraterrestrial irradiance.

The modified Linke turbidity coefficient, represented as T_L, affects the attenuation of light by the atmosphere. It quantifies the level of atmospheric haze or pollution, with higher values indicating greater attenuation and reduced transmitted light.

The relative optical air mass, denoted as m, measures the amount of air that sunlight must traverse before reaching the Earth's surface. It takes into account factors such as the angle of incidence and the atmospheric conditions, providing an indication of the path length and atmospheric density through which the sunlight passes.

The extraterrestrial irradiance, represented as $G_{b,\tau}$, refers to the incoming solar radiation at the outer boundary of Earth's atmosphere. It depends on various parameters including the zenith angle of sunlight, geographical latitude, angular position between the sun and the local meridian in relation to the equatorial plane, angle between the surface plane and the horizontal plane, angular variation of the local meridian due to the Earth's rotation (approximately 15 degrees per hour), and the angle of incidence.

These parameters collectively influence the quantity of light reaching the Earth's surface, ac-

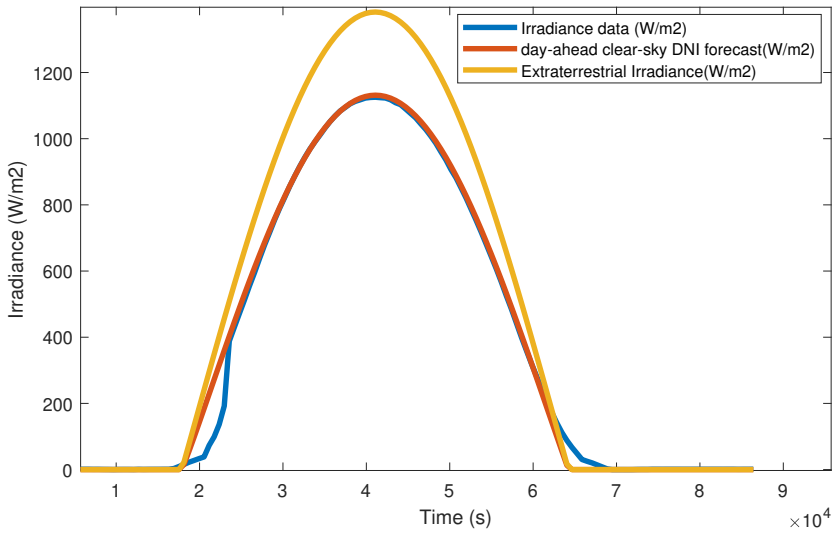


Figure 2.2: Solar irradiance simulation

counting for atmospheric conditions, solar geometry, and the overall path of sunlight through the atmosphere.

In Figure 2.2, a comparison is presented between the estimated irradiance and the measured data obtained from a set of panels. The plot clearly demonstrates that the estimations closely align with the corresponding measured values, indicating a remarkably high level of accuracy in the estimation process. This agreement between the estimated and measured irradiance validates the reliability and effectiveness of the estimation method used.

In order to compute an estimation of the energy that can be obtained with a solar panel, we have to evaluate the solar panel efficiency η_p, that will depend employed technologies. The inverter efficiency η_{inv} that accounts for the efficiency of the DC (direct current) to AC (alternating current) conversion process. And the geometrical factors as the panel surface that is exposed to the sunlight S.

$$P_{output} = S\eta_p\eta_{inv}I_{cs}. \tag{2.2}$$

Additionally, in order to have a more accurate scenario, we also have some generation data available obtained from [NSS+21b] that describes a high production household scenario.

2.4 Household Consumption

The daily variability of household consumption is highly pronounced due to the varying behaviors of the individuals residing within it, as mentioned in [SSM+22]. However, by examining larger communities or averaging data from the same household over multiple days, some insights about household consumption can be obtained.

One relevant observation is the presence of a consumption offset, whereby the consumption of idle machines connected to the grid remains constant. Throughout the day, there are multiple

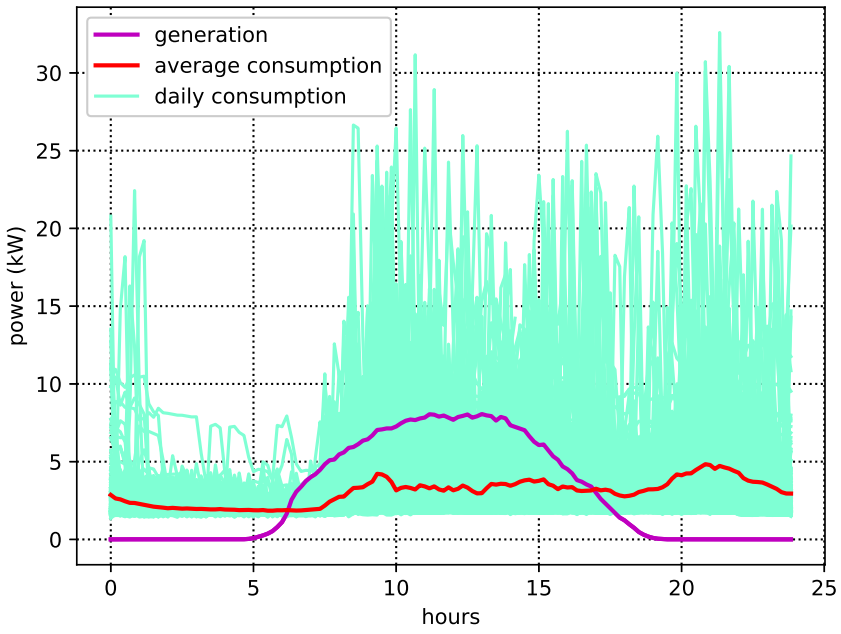


Figure 2.3: Demand and generation of power profiles

consumption peaks. The first peak occurs in the early morning when household members prepare for the day, while another peak is observed at night, characterized by a wider consumption period. It is worth noting that during the daylight hours, the overall consumption remains relatively consistent.

As depicted in Figure 2.3, it is evident that the generation profile of a photovoltaic (PV) array does not align with the consumption profiles, even when the array is appropriately sized to meet the household's needs throughout the day (as discussed in [NSS+21a] and [NSS+21b]). This mismatch implies that without Energy Storage Systems (ESS), a significant portion of the generated power would have to be directed to the grid instead of being utilized for direct consumption. This can result in energy deficiencies and potential losses.

To address this disparity, the installation of ESS can prove beneficial. An ESS would enable the distribution of power from production periods to consumption periods, thereby reducing dependence on the wider electrical grid. By storing excess energy during periods of high generation and utilizing it during times of high consumption, an ESS helps bridge the gap between PV generation and household energy needs, optimizing energy utilization and minimizing reliance on external power sources.

2.5 Microgrids

All the above mentioned components are in essence the components that are needed to conform a microgrid [HPG18].

However, at a microgrid level, we are going to simplify the considerations over those components. The management of the energy resources will not enter into the physical details of the systems

itself, but at higher level in a phenomenological behavior that can affect the grid tension.

Firstly, the voltage of the power line must be constant at any time. This means that the power that is extracted from the power line must be also be supplied by either the power sources, the batteries or the connections to wider grids. Otherwise, the voltage of the power line might drop, and the systems that work under the assumption of a given voltage may malfunction. That control should be instantaneous, and the smaller the microgrid, the more susceptible to drastic changes it is, as the plugging or unplugging of electrical appliances will constitute a significant change. This condition will later lead to the Equilibrium of the Grid restriction.

As for the ESS, the physical models of the systems usually include nonlinearities, and the chemical reactions may need a little time until reaching the full charging speed. For the physical limitations of the ESS, we will consider that it works linearly along all the working space, where the stored energy will change proportionally with respect the input/output power. In order to make those assumptions more correct, the low level controller must usually know the SOC of the ESS and there are usually limitations on the range of SOC that the battery can work safely. Additionally, the longevity concerns of the battery, the inefficiencies and the discourage of particular behaviors can be included as hard and soft limits in an EMPC based control or into the operational costs in a RL based technique.

Regarding the power generators and the connections to the power line, they all must pass through power transformer and inverters so the line keeps its coherence, frequencies and voltage. Despite advancements in the power electronics and the control techniques, the losses can waste from 5% to 15% of power generation depending on the number of back-and-forth conversions[HPG18].

All those simplifications make a feasible EMPC formulation that can be solved under a reasonable amount of time. An MPC is a multivariable control strategy that uses ideally an accurate state-space model, some possible constraints on the process variables, and an objective function to solve optimization problems. The predictive control solves an optimization problem using a moving time horizon window. MPC is not only able to predict in advance the next control, but it can also select the optimal control actions.

Classically, the MPC approach was formulated for tracking purposes, where the costs are formulated in a quadratic form with respect to the desirable set-points. It penalizes deviations of the states and control inputs from their reference trajectories while explicitly enforcing the constraints.

The constraints can be determined by the system characteristics like the working limits of the of the system states or the control signal limitations $h(x_k, u_k) \leq 0$. Additionally, the states of the system mus suffice the modeled dynamics at each time step $x_{k+1} = f(x_k, u_k) \quad \forall k$. This will transform the control problem into a optimization problem with several restrictions.

$$\begin{aligned} \min_{x_k} \quad & \sum_{k=1,\dots,H} (x_k - x_k^{ref})^T Q (x_k - x_k^{ref}) + u_k^T R u_k \\ s.t. \quad & x_{k+1} = f(x_k, u_k) && \forall k \ \textit{Predictive model constraints} \\ & h(x_k, u_k) \leq 0 && \forall k \ \textit{System constraints} \end{aligned} \tag{2.3}$$

where x_k and u_k are the states and control signals at each time step of the prediction horizon.

The optimization function is added up along all the horizon window, H represents the size of the horizon window and x_{ref} the set point for the states. f is the predictive model, that must be accurate in order to obtain precise solutions, but computationally efficient, and h represent the conditions that the states and control signals must satisfy.

In the economical MPC, the main difference will reside in the formulation of the optimization function since will not require a reference trajectory or set point, and will uniquely based on the operational costs that the control signal may generate. There is not an specific shape of those costs, so it might take any form depending on the objectives that we want to optimize in each case. The solution of this problem will simultaneously generate the trajectory of the states and the control signal sequence to reach it

$$\begin{aligned} \min_{x_k,u_k} \quad & \sum_{k=1,\dots,H} L(x_k,u_k) \\ s.t. \quad & x_{k+1} = f(x_k,u_k) \quad \forall k \ \textit{Predictive model constraints} \\ & h(x_k,u_k) \leq 0 \qquad\qquad \forall k \ \textit{System constraints.} \end{aligned} \tag{2.4}$$

One of the most common formulations of the economic MPC related to the management of a smart-grid can be formulated with the simplified subsystems [NBP20]. The control signals that can be considered in the model will be the power flow of each one of the subsystems including all the power sources such as the grid, the load on the storage systems and even the power generators.

As system constraints, it may limit the power flow of each one of the subsystems

$$u^{min} \leq u_k \leq u^{max}, \tag{2.5}$$

where the limits are set by the particular installation in the smart-grid. Some particular power flows may have additional temporal upper limits u_k^+, like the maximum generation that some renewable power sources may have. Those will be instantaneous and depend on the climate conditions, like the wind velocity for wind turbines or the available solar irradiance for photovoltaic panels.

$$u_k \leq u_k^+ \tag{2.6}$$

Regarding the storage elements, we can consider that State of Charge of each one of them will conform the working state space of the system. In general, they will be instantaneous, and will depend on the previous State of Charge and the input and output power

$$SOC_{k+1} = SOC_k + \zeta_c u_k^{in} + \zeta_d u_k^{out} \tag{2.7}$$

where ζ_c and ζ_d are the charging and discharging efficiency factors that will be given by each of the storage elements separately.

Additionally, the SOC of each storage system has its own capacity and operational limits, so they must also satisfy those system conditions

$$SOC^{min} \leq SOC_k \leq SOC^{max} \quad \forall k. \tag{2.8}$$

As we stated before, in order to ensure the stability of the system, it must be satisfied the equilibrium condition at each node of the microgrid, so the input and output power flows must add up at any time instant

$$\sum_i u_{i,k}^{in} = \sum_i u_{i,k}^{out}. \tag{2.9}$$

Regarding the economical costs of the EMPC formulation, most of the cost associated to electrical power production are related to the purchase and maintenance of generators, as well as their accessories. Additionally, legal canons (taxes) and electricity costs can also be included in the associated economic costs

$$f_k^E = (\alpha_1 + \alpha_{2,k})^T u_k \Delta t, \tag{2.10}$$

where Δt is the sampling time, α_1 is the time independent part of the control costs and $\alpha_{2,k}$ represents the time dependent part, such as the price of the grid electricity.

Other common costs are those associated to a smooth operation, that will penalize sudden changes on the operational behavior

$$f_k^{\Delta u} = \Delta u_k^T \Delta u_k, \tag{2.11}$$

where $\Delta u_k = u_k - u_{k-1}$ is the control input variation between two consecutive time steps.

And the costs associated to a safety measures, that will introduce a penalization when the SOC of some ESS components goes below some safety threshold δ

$$\delta_i - \epsilon_{i,k} \leq SOC_{i,k}, \quad \forall i, k \tag{2.12}$$

where the ϵ is a vector of slack variables that should be minimized by including an extra term in the MPC cost function

$$f_k^S = \epsilon_k^T \epsilon_k. \tag{2.13}$$

Then, the function that a regular EMPC would try to minimize is composed by a linear combination of all those costs along all he prediction horizon

$$L(\mathbf{x}, \mathbf{u}) = \lambda^E f_k^E + \lambda^S f_k^S + \lambda^{\Delta u} f_k^{\Delta u} \tag{2.14}$$

3 Reinforcement Learning based Control

In this chapter, our focus will be on exploring concepts related to the management of system with the objective of minimizing operational costs. As mentioned earlier in this document, our project aims to develop algorithms that are independent of pre-existing models and instead rely on data-driven approaches for control.

In this particular context, numerous approaches have been proposed. Among the most widely recognized ones, it is noteworthy to mention the model-free approaches. These approaches determine the optimal control based on the changes in states and the anticipated costs. Such formulation is commonly associated with Reinforcement Learning (RL), which requires knowing the states at each moment and the scalar cost of operation (also called *stage cost*). Typically, this cost is assumed to be a function denoted as $L(x, u)$, where x represents the current state and u denotes the control action.

Using this definition, we can characterize the action policy $\pi(x)$ as the function responsible for determining the control actions at every state point, along with the corresponding long-term expected cost for each state and action policy function $J(x, \pi)$. When considering continuous time and an infinite prediction horizon, this can be expressed as follows:

$$J(x_0, \pi) = \int_{t_0}^{\infty} L(x, \pi(x))dt, \tag{3.1}$$

where the states will evolve according to the applied control action $\pi(x)$ and the system and $x_0 = x(t_0)$ represents the initial point of the trajectory.

A commonly held assumption is the existence of an optimal Value Function, denoted as V, which assigns the optimal long-term expected cost to each state point

$$V(x) = \inf_{\pi} J(x, \pi) \quad \forall x. \tag{3.2}$$

This optimal cost can only be achieved by applying the optimal control signals at each time step, thereby defining the optimal policy function as well

$$\pi^* = \operatorname*{arginf}_{\pi} J(x, \pi) \quad \forall x, \tag{3.3}$$

where π^* denotes the optimal control policy that would minimize the long term cost $J(x, \pi)$ from any point of the state space.

Many of the prevailing learning algorithms depend on the acquisition of data regarding state changes before and after executing actions, as well as the associated costs incurred at each state [MKS+13]. Utilizing this gathered information, one can attempt to approximate the actual Value Function and Action Value Function.

For each lapse of time, we can deduce that the definition of the long term cost must be consistent along the time:

$$\begin{aligned} J(x_0, \pi) &= \int_{t_0}^{\infty} L(x, \pi(x))dt \\ &= \int_{t_0}^{t_0+\tau} L(x, \pi(x))dt + \int_{t_0+\tau}^{\infty} L(x, \pi(x))dt \\ &= \int_{t_0}^{t_0+\tau} L(x, \pi(x))dt + J(x_\tau, \pi), \end{aligned} \tag{3.4}$$

where the notation $x_0 = x(t_0)$ represents the initial state and $x_\tau = x(t_0 + \tau)$ represents the state of the system at time $t_0 + \tau$, which is obtained by applying the given policy π starting from the initial time t_0 for a duration of τ. If we can measure the states and compute the associated costs, it becomes possible to compute the integral of the cost over a specific time period. By doing so, we can estimate the values of the long-term cost for any state by comparing the previous long-term cost function with the estimation provided by the right-hand side of equation (3.4).

The difference between any candidate representing the long-term cost for a given problem $J(x_0, \pi)$ and the approximation obtained from the experience is commonly known as the Temporal Difference (TD) equation:

$$J(x_t, \pi) + \int_{t}^{t+\tau} L(x, \pi(x))dt - J(x_{t+\tau}, \pi), \tag{3.5}$$

where x_t represents the state at time t for any particular realization. Given an initial estimate of the long-term cost function, we can iteratively refine it using the collected data and the Temporal Difference formula (3.5). The goal is to minimize the discrepancy until equivalence is achieved for all data points using the assignation in the Equation (3.6).

$$J(x(t), \pi) \leftarrow J(x(t), \pi) + \alpha\Big(\int_{t}^{t+\tau} L(x(s), \pi(x))ds + J(x(t+\tau), \pi) - J(x(t), \pi)\Big), \tag{3.6}$$

where α is the learning rate that determines how fast the Value Function can be learned. After establishing the fundamental concepts of Reinforcement Learning, one can delve into the specific methods of data collection for learning and the techniques for minimizing the Temporal Difference. These explorations lead to the development of various algorithms in the class of Temporal Difference-based Reinforcement Learning.

3.1 The Ricatti analogy

These methods can also be applied to more classical problems like the Linear Quadratic Regulator, where the system follows a Linear Time-Invariant(LTI) law:

$$\dot{x} = Ax + Bu, \qquad x(0) = x_0, \tag{3.7}$$

where $x \in \mathbb{R}^n$ is the state and $u \in \mathbb{R}^m$ is the control input. Additionally, the cost of performing a certain policy is given by the quadratic cost:

$$J(x,\pi) \triangleq \int_0^\infty [x^T(s)Qx(s) + u^T(s)Ru(s)]\ ds, \tag{3.8}$$

where the stage cost depends quadratically on the state and the control cost $L(x,u) = x^T(t)Qx(t) + u^T(t)Ru(t)$, and both $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ are symmetric and positive definite matrices. As usually denoted as $Q \succ 0$ and $R \succ 0$.

With the definition of the Value Function in Equation (3.2) and the conditions of the Temporal Difference in Equation (3.4) for any policy, we can establish Bellman's principle of optimality [Bel52]. This principle states that for every state x and any time interval τ, the Value Function must satisfy:

$$V(x) = \inf_\pi \Big\{ \int_t^{t+\tau} x^T(s)Qx(s) + \pi(x)^T R\pi(x) ds + V(x(t+\tau)) \Big\}. \tag{3.9}$$

From this condition, it can be elaborated by subtracting the left hand side of the equation

$$0 = \inf_\pi \Big\{ \int_t^{t+\tau} x^T(s)Qx(s) + \pi(x)^T R\pi(x) ds + V(x(t+\tau)) - V(x) \Big\}, \tag{3.10}$$

and dividing both sides by the interval of time τ we can obtain

$$0 = \inf_\pi \Big\{ \frac{1}{\tau} \int_t^{t+\tau} x^T(s)Qx(s) + \pi(x)^T R\pi(x) ds + \frac{V(x(t+\tau)) - V(x)}{\tau} \Big\}. \tag{3.11}$$

As this equality must hold for any time interval, we can take the limit as the interval approaches zero, $\lim_{\tau\to 0}$. In doing so, we can recover the definition of derivatives:

$$0 = \inf_\pi \Big\{ x^T(t)Qx(t) + \pi(x)^T R\pi(x) + \frac{dV(x)}{dt} \Big\}. \tag{3.12}$$

The last term can be developed a little bit further applying the chain rule leaving us with the following necessary condition of the Value Function:

$$\begin{aligned} 0 &= \inf_\pi \Big\{ x^T(t)Qx(t) + \pi(x)^T R\pi(x) + \frac{\partial V(x)}{\partial x}\dot{x} \Big\} \\ &= \inf_\pi \Big\{ x^T(t)Qx(t) + \pi(x)^T R\pi(x) + \frac{\partial V(x)}{\partial x}(Ax + B\pi(x)) \Big\}. \end{aligned} \tag{3.13}$$

The last equality is renowned the **Hamilton-Jacobi-Bellman equation**(HJB equation). It has been proved in the Section 2.2.3 of the reference [FL23] that this HJB equation also provides a sufficient condition for the optimal Value Function as it can be shown that a function that satisfy this equation is unique.

If there exists a solution V and an optimal control policy π^* that fulfills this equation, we can guarantee the stability of the closed-loop system. This is derived from the definition of the stage cost $L(x, u)$, which is always positive definite, implying that the Value Function is also positive definite. Additionally, the derivative of the Value Function under the optimal policy control is provided in the HJB Equation (3.12):

$$x^T(t)Qx(t) + \pi^*(x)^T R\pi^*(x) + \frac{dV(x)}{dt} \;\Rightarrow\; \frac{dV(x)}{dt} = -\Big(x^T(t)Qx(t) + \pi^*(x)^T R\pi^*(x)\Big). \quad (3.14)$$

The negative definiteness of the derivative of the Value Function implies that we have identified a potential Lyapunov function candidate. Thus, the system is deemed stable, at least in the Lyapunov sense.

So far, our analysis has been applicable to a broader scenario, considering both the stage cost of operation denoted as $L(x, u)$ and the system dynamics represented by $\dot{x}$. However, given the specific problem at hand, we can now establish a necessary condition for the policy function as well. As the control action that minimizes the value in the bracketed equation 3.13 must be a critical point, the partial derivative of the control policy with respect to the given function must be zero. This results in

$$2R\pi * + B^T \frac{\partial V(x)}{\partial x}^T = 0 \Rightarrow \pi = -\frac{1}{2}R^{-1}B^T(\frac{\partial V(x)}{\partial x})^T. \quad (3.15)$$

If we replace the expression of the optimal control signal into the HJB Equation (3.13),

$$\begin{aligned} 0 &= x^T(t)Qx(t) + \frac{1}{4}\frac{\partial V(x)}{\partial x}BR^{-1}B^T\frac{\partial V(x)}{\partial x}^T + \frac{\partial V(x)}{\partial x}Ax - \frac{1}{2}\frac{\partial V(x)}{\partial x}BR^{-1}B^T(\frac{\partial V(x)}{\partial x})^T \\ &= x^T(t)Qx(t) + \frac{\partial V(x)}{\partial x}Ax - \frac{1}{4}\frac{\partial V(x)}{\partial x}BR^{-1}B^T\frac{\partial V(x)}{\partial x}^T, \end{aligned} \quad (3.16)$$

we obtain a differential equation that the Value Function must satisfy. The linear constraints on the partial derivatives of the Value Function in relation to the states suggest that a possible choice for the Value Function could be quadratic in form.

$$V(x) = x^T P(t)x, \quad (3.17)$$

where P is assumed to be symmetric and positive definite matrix. Then, the HJB Equation becomes:

$$0 = x^T Qx + 2PAx - x^T PBR^{-1}B^T Px, \quad (3.18)$$

that can be rewritten into:

$$0 = x^t \left(Q + PA + A^T P^T + PBR^{-1}B^T P \right) x. \tag{3.19}$$

As this must be satisfied for all the states, we can then establish the condition that the P matrix must satisfy:

$$0 = Q + PA + A^T P^T + PBR^{-1}B^T P, \tag{3.20}$$

and finally, the control action policy must take the form of

$$\pi(x) = -R^{-1}B^T \left(\frac{\partial V(x)}{\partial x} \right)^T = -R^{-1}B^T Px. \tag{3.21}$$

With this approach, we have reached the general solution for the Linear Quadratic Regulator (LQR). To obtain the actual control signal, we only need to solve the matrix differential equation that is the well known *Algebraic Ricatti Equation* (ARE).

3.2 A Structured Approximate Optimal Control

To derive an analytical solution for generating a control signal to control nonlinear systems, we can begin by making a few assumptions about the problem. Firstly, we assume that the nonlinear system is control-affine.

When we say that a nonlinear system is control-affine, it means that the system dynamics depend linearly on the control signal, but the coefficients and independent terms can be nonlinear. In this context, let define the state variable as $x \in D \subseteq \mathbb{R}^N$ and the control input as $u \in \Omega \subseteq \mathbb{R}^m$, where D represents the space of all possible states and Ω represents the space of all possible control inputs.

Furthermore, we have two nonlinear continuous functions: $f : D \to \mathbb{R}^n$, which describes the state dynamics, and $g : D \to \mathbb{R}^{n \times m}$, which relates the control input to the state dynamics. These functions help define the system behavior.

$$\dot{x} = F(x, u) = f(x) + g(x)u. \tag{3.22}$$

In the general case, we cannot assume that the system will reach the desired equilibrium point or be able to maintain it with a null control action. This situation may lead to an infinite valued $J(x_0, \pi)$ for any initial condition. In order to address this problem, we will consider a time-dependent stage cost that gradually diminishes the importance of the current stage cost over time. This can be achieved by incorporating an exponential decay factor into the stage cost function

$$J(t, x, u) = \lim_{T \to \infty} \int_t^T e^{-\gamma s} \Big(x(s)^T Q x(s) + u^T R u \Big) ds, \tag{3.23}$$

where $Q \in \mathbb{R}^{n\times n}$ is positive semi-definite, $\gamma \leq 0$ is the discount factor, and $R \in \mathbb{R}^{m\times m}$ positive definite. In this case, we can follow similar definitions of the optimal control policy and the Value Function as outlined in Equations (3.2) and (3.3):

$$V(t,x) = J(t,x,\pi^*) = \min_{\pi} J(t,x,\pi). \tag{3.24}$$

Indeed, in this case, the Value Function will explicitly depend on time due to the discount factor, which introduces a time-dependency in the stage cost. However, we can still follow a similar reasoning as in the time-independent formulation and derive the equivalent of Equation (3.11).

$$0 = \inf_{\pi} \left\{ \frac{1}{\tau} \int_t^{t+\tau} e^{-\gamma s} \big(x^T(s)Qx(s) + \pi(x)^T R\pi(x)\big) ds + \frac{V(t+\tau, x(t+\tau)) - V(t,x)}{\tau} \right\}. \tag{3.25}$$

And performing the limit of $\tau \to 0$ we recover the differential formulation

$$\begin{aligned} 0 &= \inf_{\pi} \left\{ e^{-\gamma t} \big(x^T(t)Qx(t) + \pi(x)^T R\pi(x)\big) + \frac{dV(t,x)}{dt} \right\}. \\ 0 &= \inf_{\pi} \left\{ e^{-\gamma t} \big(x^T(t)Qx(t) + \pi(x)^T R\pi(x)\big) + \frac{\partial V(t,x)}{\partial t} + \frac{\partial V(t,x)}{\partial x}\dot{x} \right\}. \end{aligned} \tag{3.26}$$

Since the partial time derivative of the Value Function does not depend on the control policy, we can take it out of the brackets and isolate it on the other side of the equation.

$$\frac{\partial V(t,x)}{\partial t} = \inf_{\pi} \left\{ e^{-\gamma t} \big(x^T(t)Qx(t) + \pi(x)^T R\pi(x)\big) + \frac{\partial V(t,x)}{\partial x} F(x,\pi) \right\}. \tag{3.27}$$

This equation is commonly known as the **Hamilton-Jacobi-Bellman equation** (HJB equation). It is worth noting that the uniqueness of the Value Function and the optimality of any control policy have been demonstrated in Section 2.2.3 of the book [FL23]. To simplify the formulation, we can define the Hamiltonian operator H

$$H\Big(x,u,\rho\Big) := -L(x,u) + \rho^T F(x,u). \tag{3.28}$$

If we evaluate this on $\rho = -\frac{\partial V}{\partial x}^T$, we can recover the expression of the right hand side of the HJB equation.

$$H\Big(x,u,-\frac{\partial V}{\partial x}^T\Big) = -L(x,u) - \frac{\partial V}{\partial x} F(x,u). \tag{3.29}$$

And the HJB equation can be simplified to

$$\frac{\partial V}{\partial t} = \sup_{\pi} H\Big(x, u, -\frac{\partial V}{\partial x}^T\Big), \tag{3.30}$$

what should hold the property of having a null derivative for the optimal control policy, as it must be a critical point

$$\frac{\partial H}{\partial u}\Big(x, \pi^*, -\frac{\partial V}{\partial x}^T\Big) = 0. \tag{3.31}$$

Moving forward, we will make an important assumption that each component of f and g can be expressed or effectively approximated within the domain of interest by a linear combination of a set of p basis functions $\phi_i \in \mathcal{C}^1 : D \to \mathbb{R}$ for $i = 1, 2, ..., p$. Consequently, the system dynamics can be rewritten as follows:

$$\dot{x} = W\Phi(x) + \sum_{j=1}^{m} W_j \Phi(x) u_j, \tag{3.32}$$

where $W \in \mathbb{R}^{n\times p}$ are the parameters that multiplies the basis functions, $W_j \in \mathbb{R}^{n\times p}$ are the parameters that multiplies the basis functions for each control action, and $\Phi(x) = [\phi_1(x), ..., \phi_p(x)]^T$ are the basis functions itself. In order to be able to represent the costs with the same set of basis functions, we are going to assume that we will always have a constant term and a set of linear terms in the first places of basis, and the other non-linear ones are placed last $\Phi(x) = [1, x_1, ..., x_n, \phi_{n+2}(x), ..., \phi_p(x)]^T$.

With this formulation, the policy cost can be rewritten in terms of these basis functions.

$$J(x_0, u) = \lim_{T\to\infty} \int_0^T e^{-\gamma t}\Big(\Phi(x)^T \bar{Q} \Phi(x) + u^T R u\Big) dt, \tag{3.33}$$

In the rewritten policy cost expression, $\bar{Q} = \text{diag}(0, Q, \mathbf{0}_{(p-n-1)\times(p-n-1)})$ represents a block diagonal matrix with all zeros except for the block corresponding to the linear part for the states x.

Now, let assume that the Value Function has an exponential time dependency, similar to the stage cost, and a quadratic dependence on the basis functions. We can represent the Value Function as follows:

$$V(t, x) = e^{-\gamma t}\hat{V}(x) = e^{-\gamma t}\Big(\Phi(x)^T P \Phi(x)\Big), \tag{3.34}$$

where $P \in \mathbb{R}^{p\times p}$ is a symmetric matrix. These conditions enables us to effectively describe the various functions associated with the controller. These functions encompass the system dynamics, the stage cost of operation, the optimal value function, and consequently, the control policy. We can achieve this by employing a sufficient number of basis functions that span the state space.

From the Equation (3.28), taking the Value Function as (3.34) and the parameterized dynamics of (3.32), we can start developing the Hamiltonian in order to simplify the terms:

$$H = -e^{-\gamma t}\big(\Phi(x)^T \bar{Q}\Phi(x) + u^T R u\big) - e^{-\gamma t}\frac{\partial\big(\Phi(x)^T P\Phi(x)\big)}{\partial x}\Big(W\Phi(x) + \sum_{j=1}^{m} W_j\Phi(x)u_j\Big). \quad (3.35)$$

Performing the partial derivative over the Value Function, and applying the distributive rule, we can obtain

$$\begin{aligned} H = &-e^{-\gamma t}\big(\Phi(x)^T \bar{Q}\Phi(x) + u^T R u\big) \\ &- e^{-\gamma t}\Phi(x)^T P\frac{\partial\Phi(x)}{\partial x}\Big(W\Phi(x) + \sum_{j=1}^{m} W_j\Phi(x)u_j\Big) \\ &- e^{-\gamma t}\Big(W\Phi(x) + \sum_{j=1}^{m} W_j\Phi(x)u_j\Big)^T\frac{\partial\Phi(x)}{\partial x}^T P\Phi(x). \end{aligned} \quad (3.36)$$

Following that, the terms within the brackets can be separated using the distributive property. In the work by [FL23], the Structured Approximate Optimal Control algorithm was developed under the assumption that the R matrix was diagonal. As a result, the cost associated with the control action can be divided into a sum of various components of the control signal u_j. Later on, in Section [4], we will work in a generalization of the solutions with a generic R matrix.

$$\begin{aligned} H = -e^{-\gamma t}\Bigg(&\Phi(x)^T \bar{Q}\Phi(x) + \sum_{j=1}^{m} u_j(x)^2 \cdot r_{j,j} + \Phi(x)^T P\frac{\partial\Phi(x)}{\partial x}W\Phi(x) \\ &+ \Phi(x)^T P\frac{\partial\Phi(x)}{\partial x}\Big(\sum_{j=1}^{m} W_j\Phi(x)u_j\Big) + \Phi(x)^T W^T\frac{\partial\Phi(x)}{\partial x}^T P\Phi(x) \\ &+ \Big(\sum_{j=1}^{m} W_j\Phi(x)u_j\Big)^T\frac{\partial\Phi(x)}{\partial x}^T P\Phi(x)\Bigg). \end{aligned} \quad (3.37)$$

By employing these simplifications, we can look for the minimum value of the Hamiltonian, that would represent the optimal policy. In order to do that, we can seek the critical points of the Hamiltonian concerning the control signal by setting the derivative with respect to the control signal equal to zero. This is possible because we know that there is only one critical point, and it is the minimum value of the Hamiltonian.

$$0 = \frac{\partial H}{\partial u_j} = -e^{-\gamma t}\Big(2r_{j,j}u_j + 2\Phi(x)^T P\frac{\partial\Phi(x)}{\partial x}W_j\Phi(x)\Big). \quad (3.38)$$

The existence of a unique solution for all components of the control signal relies on the condition that none of the diagonal components of the R matrix are equal to zero, i.e., $r_{j,j} \neq 0$.

$$u_j^* = -\frac{1}{r_{j,j}}\Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} W_j \Phi(x). \tag{3.39}$$

In the other hand side of the HJB equation, we have that the derivative with respect to time can be written as

$$\frac{\partial V(t,x)}{\partial t} = -\gamma e^{-\gamma t}\Phi(x)^T P \Phi(x) + e^{-\gamma t}\Phi(x)^T \dot{P}\Phi(x) \tag{3.40}$$

By substituting the solution of the optimal control into the Hamiltonian equation and equating both sides of the equation in the HJB (Hamilton-Jacobi-Bellman) equation, we can determine the dynamics of the Value Function P. To simplify certain parts of the equation, we can leverage the fact that u_j is a scalar and thus equivalent to its transpose.

$$\begin{aligned}
-\gamma e^{-\gamma t}\Phi(x)^T P\Phi(x) + e^{-\gamma t}\Phi(x)^T \dot{P}\Phi(x) = -e^{-\gamma t}\Bigg(& \Phi(x)^T \bar{Q}\Phi(x) \\
& + \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x}\Big(\sum_{j=1}^{m} r_{j,j}^{-1} W_j \Phi(x)\Phi(x)^T W_j^T\Big)\frac{\partial \Phi(x)}{\partial x}^T P\Phi(x) \\
& + \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} W\Phi(x) + \Phi(x)^T W^T \frac{\partial \Phi(x)}{\partial x}^T P\Phi(x) \\
& - \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x}\Big(\sum_{j=1}^{m} r_{j,j}^{-1} W_j \Phi(x)\Phi(x)^T W_j^T \frac{\partial \Phi(x)}{\partial x}^T P\Phi(x)\Big) \\
& - \Big(\sum_{j=1}^{m} \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} r_{j,j}^{-1} W_j \Phi(x)\Phi(x)^T W_j^T\Big)\frac{\partial \Phi(x)}{\partial x}^T P\Phi(x)\Bigg).
\end{aligned} \tag{3.41}$$

After taking the common factor $\frac{\partial \Phi(x)}{\partial x}^T P\Phi(x)$ or $\Phi(x)^T P\frac{\partial \Phi(x)}{\partial x}$ in all of the summations, we can simplify some of the addends and simplify the exponential factor $e^{-\gamma t}$ leading to the following expression:

$$\begin{aligned}
-\gamma\Phi(x)^T P\Phi(x) + \Phi(x)^T \dot{P}\Phi(x) = & -\Phi(x)^T \bar{Q}\Phi(x) \\
& + \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x}\Big(\sum_{j=1}^{m} r_{j,j}^{-1} W_j \Phi(x)\Phi(x)^T W_j^T\Big)\frac{\partial \Phi(x)}{\partial x}^T P\Phi(x) \\
& - \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} W\Phi(x) - \Phi(x)^T W^T \frac{\partial \Phi(x)}{\partial x}^T P\Phi(x).
\end{aligned} \tag{3.42}$$

As all the addend terms has $\phi(x)^T$ as left side factor and $\Phi(x)$ as right factor, we can take the common factor out and finally obtain the dynamics of the P matrix

$$
\begin{aligned}
-\gamma P + \dot{P} = &- \bar{Q} + P\frac{\partial \Phi(x)}{\partial x}\Big(\sum_{j=1}^{m} r_{j,j}^{-1} W_j \Phi(x)\Phi(x)^T W_j^T\Big)\frac{\partial \Phi(x)}{\partial x}^T P \\
&- P\frac{\partial \Phi(x)}{\partial x} W - W^T \frac{\partial \Phi(x)}{\partial x}^T P.
\end{aligned} \tag{3.43}
$$

After developing the learning model for the Value Function, we can derive the expected long-term cost and the optimal control policy by utilizing a linear combination of non-linear functions over the state as a parameterization of the dynamics. However, to effectively capture these dynamics, it is necessary to learn them from a set collected information.

3.3 System identification

Numerous approaches exist for determining the linear coefficients of dynamical data for a set of non linear basis functions $\Phi(x)$.

3.3.1 Least Squares

Among the commonly used methods is the Least Squares approach, which aims to minimize the squared difference between the dynamics predicted by the parameterized model and the actual dynamics observed in a given dataset.

$$
\min_{w} \frac{1}{2} \sum_{i=1}^{N_s} (w\Phi(x^i) - \dot{x}^i)^T (w\Phi(x^i) - \dot{x}^i), \tag{3.44}
$$

where w was explicitly written in lower case, as in this case they does not represent the controlled system. And the supper-index in each x^i represents the $i-th$ sample of the dataset.

3.3.2 Recursive Least Squares

Alternate approaches involve online learning of model parameters, where the parameters are updated as new data is collected from sample to sample. One example of such an approach is the Recursive Least Squares algorithm (RLS), which continuously refines the optimal parameters based on a set of samples up to the $k-th$ sample [FL23].

$$
\hat{w}_k = \Big(\sum_{i=1}^{k} \dot{x}^i \Phi(x^i)^T\Big)\Big(\sum_{i=1}^{k} \Phi(x^i)\Phi(x^i)^T\Big), \tag{3.45}
$$

we can state a recursive correction each time that a new samples is introduced. First we introduce R_k as

$$
R_k = \sum_{i=1}^{k} \Phi(x^i)\Phi(x^i)^T = \sum_{i=1}^{k-1} \Phi(x^i)\Phi(x^i)^T + \Phi(x^k)\Phi(x^k)^T = R_{k-1} + \Phi(x^k)\Phi(x^k)^T. \tag{3.46}
$$

Accordingly, we can write

$$\begin{aligned}\sum_{i=1}^{k} \dot{x}^i \Phi(x^i)^T &= \sum_{i=1}^{k-1} \dot{x}^i \Phi(x^i)^T + \dot{x}^k \Phi(x^k)^T = \hat{w}_{k-1} R_{k-1} + \dot{x}^k \Phi(x^k)^T \\ &= \hat{w}_{k-1}(R_k - \Phi(x^k)\Phi(x^k)^T) + \dot{x}^k \Phi(x^k)^T \\ &= \hat{w}_{k-1} R_k + (\dot{x} - \hat{w}_{k-1}\Phi(x^k))\Phi(x^k)^T.\end{aligned} \tag{3.47}$$

And the recursive algorithms can be developed as follows

$$\hat{w}_k = \hat{w}_{k-1} + (\dot{x} - \hat{w}_{k-1}\Phi(x^k))\Phi(x^k)^T R_k^{-1}. \tag{3.48}$$

3.3.3 Gradient Descent

In addition to the Recursive Least Squares algorithm, there are other recursive methods available, such as Gradient Descent (GD). This method updates the parameters incrementally, taking steps based on the derivative of the current squared error, with the aim of minimizing it.

If we define the squared error as $J(w) = \frac{1}{2} e^T e = (w\Phi(x) - \dot{x})^T (w\Phi(x) - \dot{x})$, the GD algorithm would update the coefficients with the new data

$$\begin{aligned}\hat{w}_k =& \hat{w}_{k-1} - \gamma \left(\frac{\partial J}{\partial w}\right)\Big|_{\hat{w}_{k-1}, \Phi(x_k), \dot{x}_k} \\ &= \hat{w}_{k-1} - \gamma e \frac{\partial e}{\partial w}^T \Big|_{\hat{w}_{k-1}, \Phi(x_k), \dot{x}_k} = (w\Phi(x) - \dot{x})^T \Phi(x)^T\end{aligned} \tag{3.49}$$

were γ is the learning rate of the algorithm.

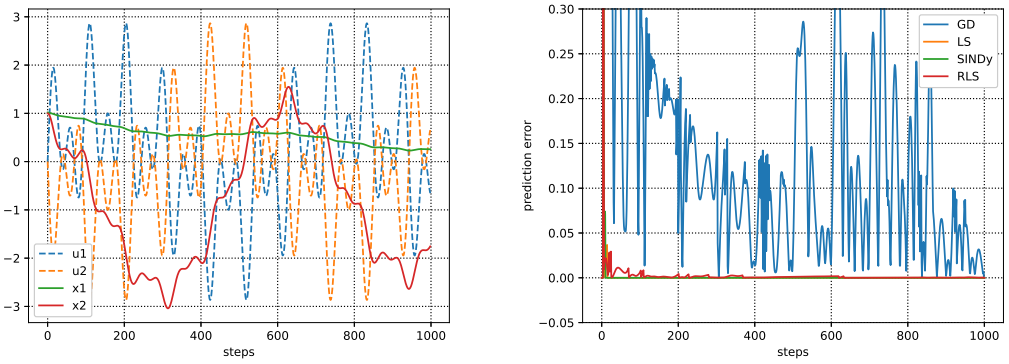
3.3.4 Sparse Regression

Furthermore, the Sparse Identification of Non-linear Dynamics (SINDy) is another approach worth considering. In SINDy, the objective function to be minimized incorporates the weight of the parameter size. Typically, this discourages the presence of large parameters and promotes model sparsity. This emphasis on sparsity enhances learning stability, as the model tends to rely on a smaller number of basis functions.

$$J(w) = \sum_{i=1}^{N_s} e_i^T e_i + \lambda ||w||_1 \tag{3.50}$$

3.3.5 System identification examples

To evaluate the performance of the different system identification algorithms, we can check it with some naive Systems. In the first one, we will have linear model with two states and two inputs



(a) Control signal generated and the subsequent dynamics of the system.

(b) Progression of the error made by the dynamical predictors.

Figure 3.1: Linear System Identification performance

$$\dot{x} = \begin{pmatrix} -0.1 & 0 \\ 0 & -0.1 \end{pmatrix} x + \begin{pmatrix} 0.9 & 1 \\ -0.9 & 0.9 \end{pmatrix} u, \tag{3.51}$$

where the state space representation will have some losses proportional to the state values as represented with the state matrix, and the input matrix will couple both input values to both of the system states.

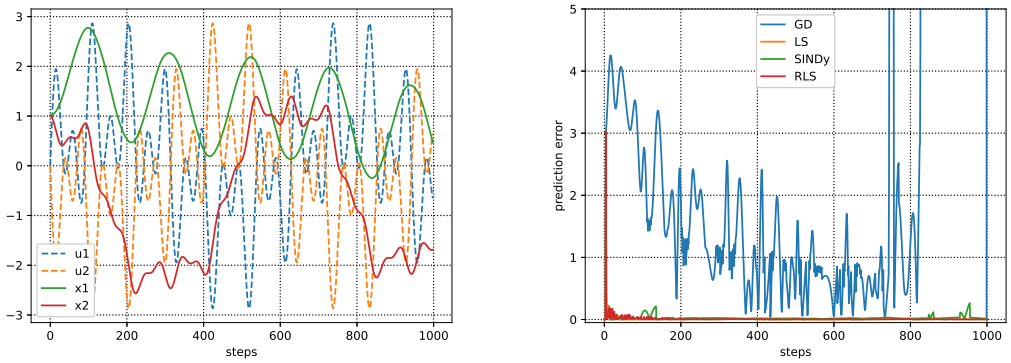
In this case, we have to adjust the hyperparameters of the SINdy and GD algorithms. In order to do it, we have to take into account that the greater the λ in the SINDy algorithm, the greater the weight of having more parameters and the lower the accuracy. As we have almost all the parameters, and we will evaluate mostly the predictions accuracy, we will choose a small value such as $\lambda = 0.05$. In the GD case the learning rate adjustments becomes a trade off between the learning speed of the algorithm, and the stability towards a fixed set of parameters. The greater the λ, the greater the shifts as each sample enters the learning algorithm. As we will have many input samples, w priorized the leaning stability, and selected a learning rate of $\lambda = 0.001$.

Figure [3.1a] shows an excitation of the model with a control signal generated with three sinusoidal with different frequencies added. As we can see in the Figure[3.1b], this excitation is enough for the predictors to learn the dynamics of the system with a set of linear basis functions. The Gradient Descend is the the algorithm that learns the dynamics slower and with a worse final error, and the Least Squares algorithm commits the smaller accumulated error.

We can also check the efficiency of the algorithm with a nonlinear system, that depends on the sinus function of some states

$$\dot{x} = \begin{pmatrix} -0.1x_1 + 3sin(x_3) + 0.9u_1 + u_2 \\ -0.1x_2 + sin(x_3) - 0.9u_1 + 0.9u_2 \\ 3 \end{pmatrix} \tag{3.52}$$

In this scenario, prior knowledge of sinusoidal components in the system non-linearities becomes



(a) Control signal generated and the subsequent dynamics of the nonlinear.

(b) Progression of the error made by the dynamical predictors.

Figure 3.2: Nonlinear System Identification performance

crucial. This knowledge allows to achieve near-perfect prediction of the system dynamics using predictors based on Least Squares, as illustrated in Figure [3.2b].

With these two examples, we have learned that those system identification algorithms are perfectly capable of learning the parameters for non-linear when the basis of functions are correctly set up, and the relevance that the hyperparameters have for the learning process using SINdy and GD.

To determine the algorithms deserving further investigation, we must consider the RLS and SINDy algorithm ability to efficiently incorporate new knowledge gained from experience. Time efficiency is prioritized to avoid the computational burden associated with relearning all data points from scratch each time new input data is introduced for the LS method. This considerations guides our selection process towards the first two ones.

It should be noted that in the toy examples used to demonstrate the learning capabilities of the System Identification problem, certain hyperparameters needed to be tuned. For instance, in the Gradient Descent algorithms, the learning rate γ had to be adjusted to ensure stable learning. Additionally, the weight of the parameters norm had to be considered in the SINDy algorithm to balance precision in the learning process. These hyperparameters play a crucial role in achieving optimal performance and should be carefully fine-tuned accordingly.

3.4 Structured Online Learning-Based Control

A novel approach in System Identification involves combining the previously learned algorithms with the model-based controller discussed in subsection 3.2. This integration results in an adaptive control algorithm capable of simultaneously learning the model for non-linear dynamics using collected data and generating control signals based on the acquired knowledge. In the book [FL23], a method known as Structured Online Learning-Based Control (SOL Control) is proposed. SOL Control involves a series of iterative steps throughout the control process to effectively govern the system.

The SOL control algorithm initiates by collecting a small set of samples comprising state, control signal, and dynamical output triples. These samples are used to initialize the System Identi-

fication algorithm. Simultaneously, an initialization of the Value Function parameter $P(0)$ is also performed to set the initial state of the algorithm. These initialization are crucial steps to kick-start the learning and control processes effectively.

Once the controller parameters are initialized with certain values, the learning process can commence through iterative steps. The iterations involve the following steps:

1. **Generate a control signal** $u(t)$ given the current learned Value Function parameters $P(t)$ and Model coefficients W with the Equation (3.39).

2. **Apply the generated control signal** and observe the behavior of the system.

3. If necessary, **save the state, control, and dynamics trio** for later use in **updating the model parameters** based on the chosen System Identification (SI) algorithm.

4. **Integrate** $P(t)$ using Equation (3.43), which updates the Value Function using the collected data.

These steps are repeated iteratively to continually improve the learned model and control strategy based on the observed system behavior.

It is worth to remark that the System Identification problem can be extended to include the control signal by introducing it to the basis functions. In this case, the dynamical model can be rewritten as follows

$$\dot{x} = W\Phi(x) + \sum_{j=1}^{m} W_j \Phi(x) u_j = \begin{pmatrix} W & W_1 & \dots & W_m \end{pmatrix} \begin{pmatrix} \Phi(x) \\ \Phi(x)u_1 \\ \vdots \\ \Phi(x)u_m \end{pmatrix} \tag{3.53}$$

In order to adapt the controlled system to the system identification notation (Section 3.3), the following substitutions can be made:

- The notation w can be replaced with $w \triangleq \begin{pmatrix} W & W_1 & \dots & W_m \end{pmatrix}$. This indicates that the parameter vector w is composed of the model coefficients W along with additional coefficients W_1 to W_m.

- The notation $\Phi(x)$ can be substituted with $\Phi(x) \triangleq \begin{pmatrix} \Phi(x) \\ \Phi(x)u_1 \\ \vdots \\ \Phi(x)u_m \end{pmatrix}$. This means that the feature vector $\Phi(x)$ is augmented with additional terms $\Phi(x)u_1$ to $\Phi(x)u_m$, where u_1 to u_m represent the control signals.

These substitutions allow for an expanded representation of the parameters and feature vectors, taking into account the additional terms associated with the control signals.

Furthermore, it is crucial to highlight that for improved time performance of the controller, we

have the flexibility to determine when to update the learned system model. Usually, if the current model demonstrates accurate predictions or if the system is in a steady state condition, there may be no need to incorporate additional data and update the model. This selective approach to model updates allows for more efficient utilization of computational resources and can contribute to overall controller performance.

4 Proposed approach

In this section, we are going to put together the elements described in Section 2 in order to establish the system that we will try to manage.

4.1 The microgrid components

In this case, we will model one household with a combined storage system composed of a battery and a super capacitor (see Figure 4.1). The higher battery capacity can give us the flexibility to store the energy that we produce in the long term and cover the power needs of the household. The supercapacitor might have a lower capacity due to the high costs and space needs for higher power capacity. This last element will give us the capacity to absorb in the short term all the power needed or generated due to its high power density, but will not be able to maintain it in the long term.

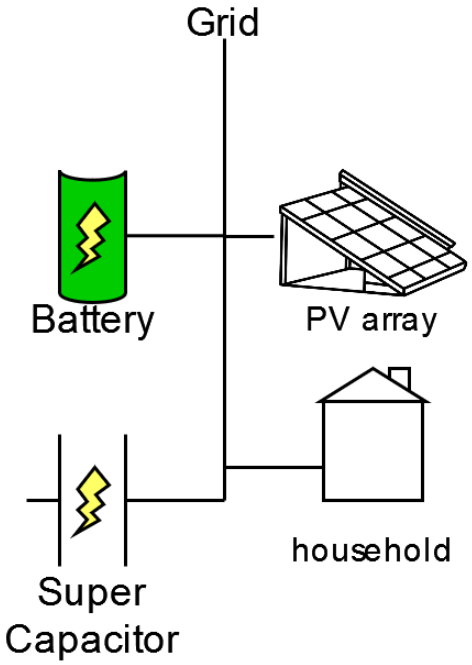


Figure 4.1: Household Grid

The house elements will be considered as power drawers only, and the demand profiles will depend on the complexity of the problem that we can solve. The simplest ones might be modeled as a constant consumption or a known sinusoidal behavior that represents the two peaks of consumption in the morning and the late evening. But we can also address the challenge of having real data once the model is well defined and solved for simpler cases.

For this problem, we are additionally considering a source of electrical power. As discussed earlier, it will consist in an array of Photovoltaic pannels, that will be cheaper to install and service due to the lack of moving parts, and the easy access to it in any circumstances opposed to the wind turbines, whose optimal deployment would require high altitude installations.

Finally, all of the energy elements will be connected to the same power line, so any lack or

excess of power can be sent or subtracted from the wider grid. This set up will ensure the well functioning of the system even when the batteries are full or drained out, or when the system cannot provide the necessary power capacities.

Note that this will add an extra cost to the installation, as it will need to be able to handle enough power peaks, but more importantly, the cost of buying electricity from the grid will have a greater impact on the cost of operation, as the prices cannot be determined by the household members.

4.2 Working hypothesis

In our simplified grid, we are going to consider a linear behavior of the Energy Storage Systems, what means that the charging and discharging of the batteries will be proportional to the amount of power introduced or drained from the ESS.

$$\frac{dSOC}{dt} = \eta Power, \tag{4.1}$$

where the SOC is the state of charge of the Storage System, and η is the efficiency factor, that we are going to consider constant for both the battery and the SC, and symmetrical for charging and discharging. In a real system, this factor could depend in many variables such as the temperature, the amount of power and even the life cycle of the system in a nonlinear fashion. Although we have simplified matters, we are going to consider different efficiencies for the two ESS that we have available due to the discussed properties, the battery will have a lower efficiency coefficient due to the sensibility to changes, and in the other hand, the SC will be pretty much efficient both charging and discharging operations.

The SOC of ESS will also have some leaks only from having it turned on, so some loses proportional to the SOC itself can be considered. In this case, the battery tends to be more stable, and therefore, the leaked energy is lower than that from the Super Capacitor. Given this two ESS, the model of the system will have the following structure:

$$\begin{aligned} dSOC_{batt}/dt &= -\alpha_{batt} SOC_{batt} + \eta_{batt} u_{batt} \\ dSOC_{SC}/dt &= -\alpha_{SC} SOC_{SC} + \eta_{SC} u_{SC} \end{aligned} \tag{4.2}$$

where the control inputs u_{batt} and u_{SC} are the delivered power to the Storage Systems, that can be positive if charging the ESSs or negative if draining them. And the suffixes of the SOC, coefficients and inputs represents the effects on that term.

In addition, we have to considerate the equilibrium of the grid condition, that relates the power that can be delivered to the Storage Systems with the generation, the demand, and the needed power that we have to draw from the electrical grid.

$$u_{batt} + u_{SC} = Generation - Demand + u_{grid}, \tag{4.3}$$

where a positive *Generation* indicates an increase on the power of the grid that has to be absorbed by the ESSs or the electrical grid if the previous fails in that. The *Demand* represents the

Parameters	Values
η_{batt}	0.9
η_{SC}	0.99
α_{batt}	0.01
α_{SC}	0.1

Table 4.1: Grid parameters

consumption of the household and will drain out the stored energy. Finally, the u_{grid} represents the power that is taken out of the electrical grid, so a positive value may charge the ESSs or compensate the high demand peak, whereas a negative value indicates a excess of the power that the system can handle, and therefore the need to deliver it to he grid.

Given the complexity to forecast the price of the electricity from the grid, we are going to consider that the price of borrowing power from it will be constant, and the price of delivering power will also be positive, as the surplus of electricity must be absorbed by a third party with some costs.

To simplify the notation, we are going to introduce the *disparity* as the difference between the *Generation* and the *Demand*, so it can be treated as a single variable that the system have to reject in order to stabilize the SOC of the ESSs.

$$u_{batt} + u_{SC} = Disparity + u_{grid}. \tag{4.4}$$

In the Table 4.1, we can see a summary of the parameter values that we are going to use in order to simulate the problem and the costs of operation parameters.

$$\begin{bmatrix} dSOC_{batt}/dt \\ dSOC_{SC}/dt \end{bmatrix} = \begin{bmatrix} -\alpha_{batt} & 0 \\ 0 & -\alpha_{SC} \end{bmatrix} \begin{bmatrix} SOC_{batt} \\ SOC_{SC} \end{bmatrix} + \begin{bmatrix} \eta_{batt} & 0 \\ 0 & \eta_{SC} \end{bmatrix} \begin{bmatrix} u_{batt} \\ u_{SC} \end{bmatrix} \tag{4.5}$$

4.2.1 Economic MPC

A natural way to solve the problem is by presenting a multi modal Economic Model Predictive Control where the objective cost function balances different control objectives cost for the optimization problem. The knowledge of the system evolution is leveraged in order make accurate predictions of the system behavior and therefore optimize accurately the objective function. The Economic part stands for the type of function that has to be optimized where, opposed to a tracking problem, does not have reference that the system must follow, but and accumulated cost of operation that has to be minimized. This objective usually has various costs that competes to be minimized, as different objectives might not be aligned, and an optimal equilibrium must be reached according to the design of the cost function.

In this case, several types of costs can be assigned to the operation of the system. The principal one that has to be considered is the economical cost that extracting and delivering electrical power to the grid, as buying energy or storing it in third party Storage Systems does come with a price.

Other factors might include the amount of power that the battery and SC systems has to handle,

and the higher it is, the more it affects to the lifespan of the components. This cost will be higher for the battery component and be progressive. They can be translated to an economical cost as the degradation of the efficiency can be taken into account as well as the need to change the battery sooner implies an economical cost to the owners.

A third cost can be associated to the State of Charge of the elements, as having some room in the higher level of the SOC will be useful to absorb more energy and some room in the lower part is advised to have a save operation. It can be also taken into account the fact that some battery technologies like the Li^+ ions based ones, suffer some degree of degradation if they operates in the extremes of its capacities.

That last type of costs can be addressed with some soft constraints over the SOC of the components. That add a new parameter to the system, but has an optimal value of zero if the desired restriction is met by the SOC

$$\begin{aligned} L(SOC)_{safety} = \quad & c_{safety} \cdot z_{safety} \\ s.t. \quad & z_{safety} \geq 0 \\ & SOC \leq SOC_{maxdesired} + z_{safety}, \end{aligned} \tag{4.6}$$

where $SOC_{maxdesired}$ is the limit that we wished to be fulfilled. Eventhough the described cost is linear, in a more general case, it can be considered to describe any other positive function for positive values of z_{safety}, but coming with its computational burden.

Another great advantage of using EMPC methods comes from the possibility to establish hard limits to the system states, controls or any relation between them, what bring us the flexibility to work and operate in safe environments if there is a solution that meets the restrictions.

Despite all those advantages, there are also some drawbacks in employing MPC methods. The first one comes with the hyperparameters that we have to tune in order to obtain the desired equilibrium between the different cost functions. But the hardest choice to make is the equilibrium between the control horizon, that will affect to the computational burden as the horizon is increased, and the time constant, that will allow a more detailed control if short, but at the cost of being able to predict shorter in time.

This decision can be found quite hard, as the power demand and generation forecasts are done in a daily basis, so the relevant fluctuations can last for hours, whereas the controller must be able to handle changes in a sub-minute basis, as switching appliances like the light, the oven or the microwave happen in the scale of minutes.

In order to address the dichotomy between the time performance (with lower prediction horizon steps), reliability (having a small time intervals at each step) and optimality (leveraging the knowledge of the forecast by expanding the predicted time horizon), there are several proposals. Among the most popular ones, there is the differentiation between a EMPC based planner, with higher time horizon that leverages the forecasting knowledge and generates the trajectory that the SOC of the ESSs must follow, and the tracker, with shorter time horizon, but higher reliability in the control, that will be in charge of following the desired trajectory despite the sudden changes on production or demand.

4.2.2 Reinforcement Learning

Another approach that has been growing in popularity is the data driven approach, that can learn the policy to control the system. These methods does also rely on a cost function, that will be optimized in the long term as seen in the Section 3.

As in the previous case, one of the principal problems is to determine the cost function that we want to optimize in order to describe a well functioning system. We decided to make an analogy to a quadratic cost in order to leverage the SOL method described in the Section 3.4 from the book [FL23].

The first limitation that we can find applying those learning techniques is the lack of hard constraints that can limit or couple different variables, like in the the equation (4.4). As stated before, this links the elements in the same power-line inlcuding the inputs from the grid, the energy production and consumption disparity and the Storage Systems ones. This input signal can be freed if we substitute the power input of one of the Energy Storage Systems, for instance, the one related to the Super Capacitor

$$u_{SC} = Disparity + u_{grid} - u_{batt}. \tag{4.7}$$

The model becomes a little bit different by substituting the previous condition. Now, the production and demand profiles implications on the system can be explicitly observed, as well as the power drawn from the grid.

$$\begin{aligned} dSOC_{batt}/dt &= -\alpha_{batt} SOC_{batt} + \eta_{batt} u_{batt} \\ dSOC_{SC}/dt &= -\alpha_{SC} SOC_{SC} + \eta_{SC}(Disparity + u_{grid} - u_{batt}). \end{aligned} \tag{4.8}$$

In this context, we want to penalize having great power inputs to the Storage Systems due to the degradation that it may generate. But the cost will be much lower for the Super Capacitor compared to the Battery one. In the other hand, the power obtained and delivered to the grid will have a higher cost due to the economical implications that it must convey. This description leads us to a quadratic cost with different 3 parameters for the control inputs

$$L(u) = r_{batt} u_{batt}^2 + r_{SC} u_{SC}^2 + r_{grid} u_{grid}^2, \tag{4.9}$$

where the specific proportion between each one of them will generate a different behavior.

As we have done before with the system equation, the grid stability condition can be embedded in the control cost as well

$$\begin{aligned} L(u) &= r_{batt} u_{batt}^2 + r_{SC}(Disparity + u_{grid} - u_{batt})^2 + r_{grid} u_{grid}^2 \\ &= r_{batt} u_{batt}^2 + r_{SC} Disparity^2 + r_{SC} u_{grid}^2 + r_{SC} u_{batt}^2 + 2 r_{SC} u_{batt} Disparity \\ &\quad + 2 r_{SC} u_{grid} Disparity + 2 r_{SC} u_{batt} u_{grid} + r_{grid} u_{grid}^2 \\ &\approx (r_{batt} + r_{SC}) u_{batt}^2 + 2 r_{SC} u_{batt} Disparity \\ &\quad + 2 r_{SC} u_{grid} Disparity + 2 r_{SC} u_{batt} u_{grid} + (r_{grid} + r_{SC}) u_{grid}^2, \end{aligned} \tag{4.10}$$

where the last equivalence has been done joining the terms and eliminating the one that only depends on the power *Disparity* between generation and consumption. If we assume that the remaining terms related to the *Disparity* won't have much effect on the total cost thanks to the small value that r_{SC} represents and that the PV array is well dimensioned such as the mean of the *Disparity* is null, the cost of control can be estimated as

$$L(u) \simeq (r_{batt} + r_{SC})u_{batt}^2 + 2r_{SC}u_{batt}u_{grid} + (r_{grid} + r_{SC})u_{grid}^2. \tag{4.11}$$

That can also be rewritten in a quadratic non diagonal matrix

$$R = \begin{pmatrix} r_{batt} + r_{SC} & r_{SC} \\ r_{SC} & r_{grid} + r_{SC} \end{pmatrix}. \tag{4.12}$$

such as $L(u) = u^T Ru$, and the control vector can be expressed as a two dimensional vector

$$u = \begin{pmatrix} u_{batt} \\ u_{grid} \end{pmatrix}. \tag{4.13}$$

In order to elongate the life time of the battery, we stated that it was better to operate in a narrower path that avoids reaching the top and the bottom of the electrical capacity. And better if we could discourage operating near the limits. This condition can be translated to a quadratic condition

$$L(SOC_{batt}) = q_{batt}(SOC_{batt} - SOC_{batt,middle})^2, \tag{4.14}$$

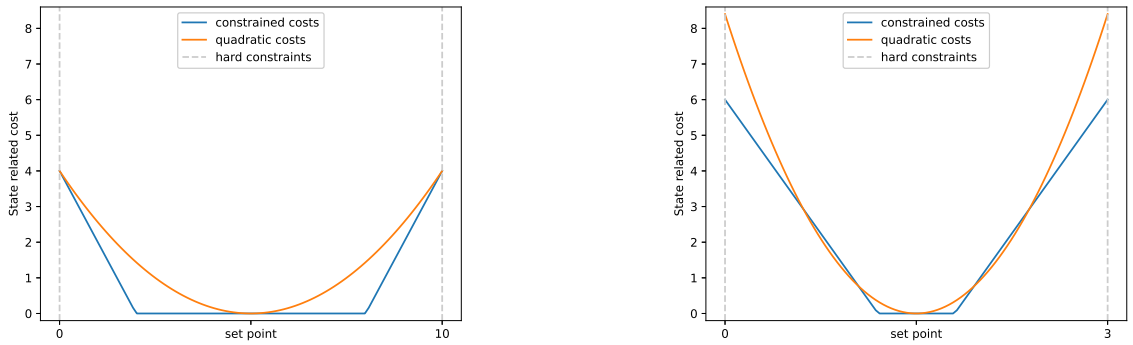
where the coefficient q_{batt} can determinate how loose the condition is if we get far from the desired point, and next to the operational limits of the battery. The $SOC_{batt,middle}$ term is the operational point of the battery, around which we have decided that can be easier to operate around safely. The easier operational point can be the 50% of the battery capacity, as it will have the same room up to the top and lower limits.

In the case of the *SC*, we will also want to operate around the middle point of the capacity, as it will allow us to be more flexible in the operation and be able to absorb the power peaks more efficiently.

$$L(SOC_{SC}) = q_{SC}(SOC_{SC} - SOC_{SC,middle})^2, \tag{4.15}$$

In this case, as the *SC* has a lower capacity and due to the relevance that having some availability in the *SC* capacity for the grid reliability, this q_{SC} coefficient might take a higher value compared to the battery's coefficient.

If we join the two terms, we can now define a quadratic cost for the states $L(x) = x^T Qx$, where



(a) Battery SOC cost comparison.

(b) *SC* SOC cost comparison.

Figure 4.2: Comparison between the soft constraints and quadratic behavior

$$Q = \begin{pmatrix} q_{batt} & 0 \\ 0 & q_{SC} \end{pmatrix}, \tag{4.16}$$

and the state is a normalized one that goes around zero

$$x = \begin{pmatrix} SOC_{batt} - SOC_{batt}^{ref} \\ SOC_{SC} - SOC_{SC}^{ref} \end{pmatrix}. \tag{4.17}$$

Note that this method does not contemplate any hard constraints, neither in the control power signals nor in the states. The only way to avoid it is by adjusting correctly the parameters in order to avoid the limits and in case that it cannot be handled, having some recovering policies such as cropping the power inputs for the battery an *SC* to the specifications in case the controller's signal gets out of bounds. And having some policy of extracting and delivering the excess power to the grid despite the greater cost in case the limits are reached.

In the Figure [4.2] we can see how the quadratic cost can be compared to the soft constraints if the parameters are chosen in order to match them. The Battery will have a wider operational range, so the cost curve can be softer, but as the *SC* operational point consideration is due to reliability issues, the range will be smaller, and the quadratic cost curve must also be more pronounced.

In order to describe all the parameters that we will use for the learning algorithm parameters, they will be described in the Table [4.2]. The capacity of the systems are designed to represent a real installation with a significantly higher volume for the battery compared to the SC, and adjusted to the production and consumption profiles in the Figure 2.3. The references will be set at the middle point of the capacity, so the penalization Will be higher next to the working limits and symmetrical in the upper and lower bounds. The values of the Q matrix are adjusted to represent the higher costs of the soft constraints of the SC, that were set to ensure some safety measurements. Lastly, the parameters of the R costs matrix are scaled to represent the higher cost of taking electricity from the grid with respect to the battery and the SC. And also represent

Parameters	value
Battery Capacity	$10\,kW \cdot h$
SOC_{batt}^{ref}	$5\,kW \cdot h$
SC Capacity	$3\,kW \cdot h$
SOC_{SC}^{ref}	$1.5\,kW \cdot h$
r_{SC}	$0.01 kW^{-2}$
r_{batt}	$0.1 kW^{-2}$
r_{grid}	$1 kW^{-2}$
q_{batt}	$1(kW \cdot h)^{-2}$
q_{SC}	$23(kW \cdot h)^{-2}$

Table 4.2: Controller hyperparameters

the higher power limits with the degradation costs of the battery components with respect to the SC's.

4.3 Generalization of the learning algorithm

Given that we have a non diagonal matrix for the control cost, we cannot employ the method developed in 3.4 straightforwardly, as they assumed that it was diagonal. However, the assumption was only used to state the dynamics of the $P(t)$ matrix, and we can take the same approximation with a more general matrix.

We can begin with the same definition of the Hamiltonian functional

$$H = -e^{-\gamma t}\big(\Phi(x)^T \bar{Q} \Phi(x) + u^T R u\big) - e^{-\gamma t} \frac{\partial\big(\Phi(x)^T P \Phi(x)\big)}{\partial x} \Big(W\Phi(x) + \sum_{j=1}^{m} W_j \Phi(x) u_j\Big), \quad (4.18)$$

so the maximum value of it needs to be equal to the time derivative of the Value Function (3.34) according to the Hamilton-Jacobi Bellman equation (3.30).

In this case, we are going to derive the equation by components as before

$$\frac{\partial H}{\partial u_j} = -e^{-\gamma t}\Big(R_{[j,:]} u + u^T R_{[:,j]} - 2\Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} W_j \Phi(x)\Big), \quad (4.19)$$

where u_j stands for the $j-th$ element of the control signal, $R_{[j,:]}$ and $R_{[:,j]}$ represents the $j-th$ row and column of the R matrix respectively. As the R matrix is symmetrical, and the partial derivative must be zero, we can obtain some conditions over the optimal control inputs

$$0 = R_{[:,j]} u_j - \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} W_j \Phi(x) \quad (4.20)$$

And stacking all the dimensions, we can get

$$0 = \begin{pmatrix} u_1 & \dots & u_m \end{pmatrix} R - \begin{pmatrix} \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} W_1 \Phi(x) & \dots & \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} W_m \Phi(x) \end{pmatrix} \tag{4.21}$$

leading to the following definition of the optimal control signal

$$u^T = \Phi(x)^T P \frac{\partial \Phi(x)}{\partial x} \begin{pmatrix} W_1 \Phi(x) & \dots & W_m \Phi(x) \end{pmatrix} R^{-1}, \tag{4.22}$$

where we have leveraged the common pre-factor of the independent term vector. After that, we can substitute it in the HJB equation (3.30) and return to the expected result

$$\begin{aligned} -\gamma e^{-\gamma t} \Phi(x)^T P \Phi(x) + e^{-\gamma t} \Phi(x)^T \dot{P} \Phi(x) = -e^{-\gamma t} \Bigg(& \Phi(x)^T \bar{Q} \Phi(x) \\ & - \Phi(x)^T P \frac{d\Phi(x)}{\partial x} \mathcal{W}\Phi(x) R^{-1} \big(\mathcal{W}\Phi(x)\big)^T \frac{d\Phi(x)}{\partial x}^T P \Phi(x) \\ & + \Phi(x)^T P \frac{d\Phi(x)}{\partial x} W \Phi(x) + \Phi(x)^T W^T \frac{d\Phi(x)}{\partial x}^T P \Phi(x). \end{aligned} \tag{4.23}$$

where $\mathcal{W}\Phi(x) = \begin{pmatrix} W_1 \Phi(x) & \dots & W_m \Phi(x) \end{pmatrix}$. And finally, the dynamics of the $P(t)$ matrix can be stated as

$$\begin{aligned} \dot{P} = -\bar{Q} + \gamma P + P \frac{d\Phi(x)}{\partial x} \mathcal{W}\Phi(x) R^{-1} \big(\mathcal{W}\Phi(x)\big)^T \frac{d\Phi(x)}{\partial x}^T P \\ - P \frac{d\Phi(x)}{\partial x} W - W^T \frac{d\Phi(x)}{\partial x}^T P \end{aligned} \tag{4.24}$$

With this generalization, we can also recover the previous dynamics if the matrix R is diagonal. But similarly, the matrix must be not singular.

5 Results

5.1 Simulation results

In this section, we are going to test the performance of the SOL algorithms compared with other classical solutions like the Linear Quadratic Regulators (LQR) in different scenarios. The main goals of the problem and the structure of the cost was already given in the previous chapter [4], and we are going to variate the profiles of the disparity.

Firstly, we will assume a null disparity, what will reduce the problem into a classical tracking problem with a single tracking reference. Secondly, we can consider it as a random white noise, what will make the problem similar to a noise rejection problem in a tracking scenario. And later on, the problem will be stated with a synthetic disparity profile, were the new signal must be managed adequately and cannot be longer considered a noise rejection problem.

5.1.1 Stabilizing Control

In our first approach, we will compare the different methods in a simple stabilizing method with online learning. To have some ground truth, to compare it with, we will have the optimal Ricatti solution so the cost of the tacking problem should be lower bounded by it. In this case, we will consider a noise free environment, where the disparity between the production and the consumption will be null. So, the system must be driven to the reference point the most efficiently possible and maintain the State of Charge despite the power leaks of the Storage Systems.

Regarding the SOL algorithm, we have to make several choices in order to define correctly the solution. The first one is related to the choice of basis functions. This decision can be defined by the knowledge that we have about the system and the complexity that we expect. If there are high nonlinearities an periodic parameters, we could introduce polynomial and sinusoidal basis with respect to the system states, but those would come with a higher computational costs and even instabilities in the System Identification process.

In our case, a simple linear model is enough to represent the variability of the dynamics, and assuming the representation of the states as in the Equation (4.17), we could choose the constant term and the linear parts as basis functions:

$$\Phi(x) = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}. \tag{5.1}$$

And the second choice that we can make is related to the computational efficiency and the stability of the solution. A constant update of the learned model parameters would lead to an unstable update of the dynamics of the Value Function $\dot{P}(t)$. In order to enforce the stability of the system, we could limit the updates of the database of points, and therefore the updates of the learned model parameters when the prediction diverge from the actual dynamics.

We could estimate the actual dynamics with and Eulerian equation from two consecutive samples

$$\dot{x}(t) = \frac{x(t+\Delta) - x(t)}{\Delta}, \tag{5.2}$$

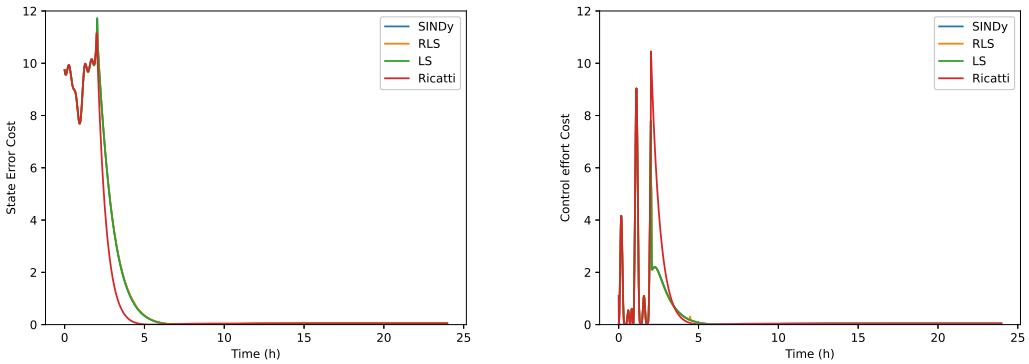


Figure 5.1: State error costs evolution on the left side and Control effort ones on the right side of the system with a initial system identification phase of two hours.

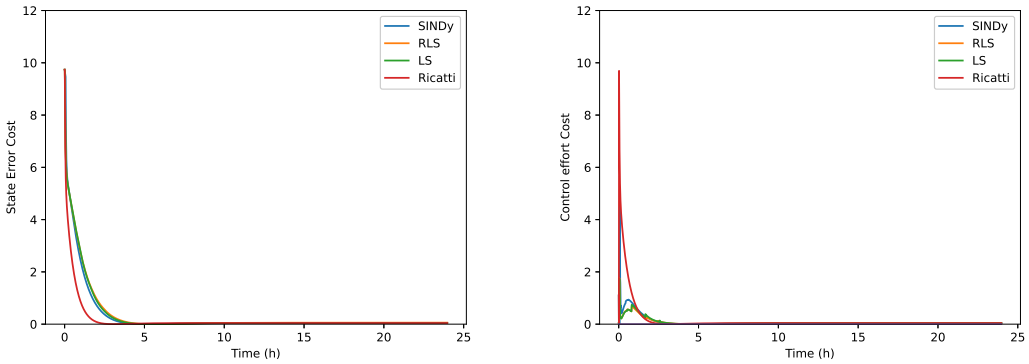


Figure 5.2: State error costs evolution on the left side and control effort ones on the right side of the system without an initial system identification phase.

and the modeled prediction will depend on the current set of parameters and the basis evaluation

$$\hat{\dot{x}}(t) = W(t)\Phi(x(t)) + \sum_j W_j(t)\Phi(x(t))u_j(t) \tag{5.3}$$

And the updates will only take place if the difference between them reaches a certain value $||\dot{x}(t) - \hat{\dot{x}}(t)|| > \epsilon(t)$ that may also be variable as discussed at the end of the Section 3.4.

We will keep the choices that we have made relating the basis functions and the System Identification updates in this section along all the experimentation part to make a more simple tracking of the results. Except if we say so, the basis functions will be limited to the constant and linear parts thanks to the simplicity of the problems that we have proposed, and the updates condition will be kept in order to be more efficient, and avoid unnecessary updates of the system parameters.

As we can see in the Figures [5.1] and [5.2], the learning algorithms are able to produce the correct control actions in order to reduce the tracking error except for the Gradient Descent method, that behaves more unstably, and requires more iterations to learn the dynamics.

We can observe that the SINDy and LS System Identification methods gives us an almost perfect tracking compared to the costs generated by the Ricatti solution in both cases, with and without an initial phase with some system excitation. However, the RLS system Identification algorithm overlaps the control signal with the SINDy's if we provide an initial period of system excitation. But if we do not have that initial phase, the system is not behaving in a stable way, due to the corrections that a recursive method might introduce when new data is introduced.

Regarding the precision of the model learning, we have to take into account that the system is learned around the tracking objective

$$x_s = x - x_{ref}, \tag{5.4}$$

Therefore, there will be some constants in the linear model. In this case, the states will be the SOC of the storing elements. After performing the learning, the SINDy and LS's parameters coincide up to the third decimal as

$$\begin{bmatrix} \dot{x}_{s,1} \\ \dot{x}_{s,2} \end{bmatrix} = \begin{bmatrix} -0.050 & -0.010 & 0 \\ -0.150 & 0 & -0.100 \end{bmatrix} \Phi(x) + \begin{bmatrix} 0.900 & 0 & 0 \\ 0.990 & 0 & 0 \end{bmatrix} \Phi(x)u(1) + \begin{bmatrix} 0 & 0 & 0 \\ 0.990 & 0 & 0 \end{bmatrix} \Phi(x)u(2). \tag{5.5}$$

This can be transcribed to a more standard form of the system dynamics

$$\begin{bmatrix} \dot{x}_{s,1} \\ \dot{x}_{s,2} \end{bmatrix} = \begin{bmatrix} -0.010 & 0 \\ 0 & -0.100 \end{bmatrix} \begin{bmatrix} x_{s,1} \\ x_{s,2} \end{bmatrix} + \begin{bmatrix} 0.900 & 0 \\ 0.990 & 0.990 \end{bmatrix} u + \begin{bmatrix} -0.050 \\ -0.150, \end{bmatrix} \tag{5.6}$$

that matches the expected values of the modeled microgrid $0.050 = \alpha_{batt} \cdot SOC_{batt}^{ref}$, $0.150 = \alpha_{batt} \cdot SOC_{batt}^{ref}$, the efficiency parameters are also equal to $\eta_{batt} = 0.90$ and the super capacitor one $\eta_{SC} = 0.990$. This means that the system identification module does converge with a high precision to the expected values along the iterative online learning and control process.

As the model is pretty simple, only a few points are necessary to estimate the parameters. As we stated before, the updating strategy will depend on the estimated error of the predictions, and once the error is lower than some threshold, we will stop collecting samples and updating the model parameters. For instance, the SINDy method has only used 9 samples of the dynamics in the database, the RLS method has required 11, and the LS one needed only 8 samples.

In this context, the time constraints of learning de dynamics does not affect the computation of the control signals, and the similarities between the RLS and LS methods are quite clear.

In the Figure [5.3], we can observe the evolution of the states along the reference tracking. We can observe that the Ricatti solution is the smothest one, but all the methods seem reliable.

In these cases, the optimal solution will fall a bit below the tracking signal due to the energy leaks in the storage systems, where maintaining the level of charge also has some costs on the power grid.

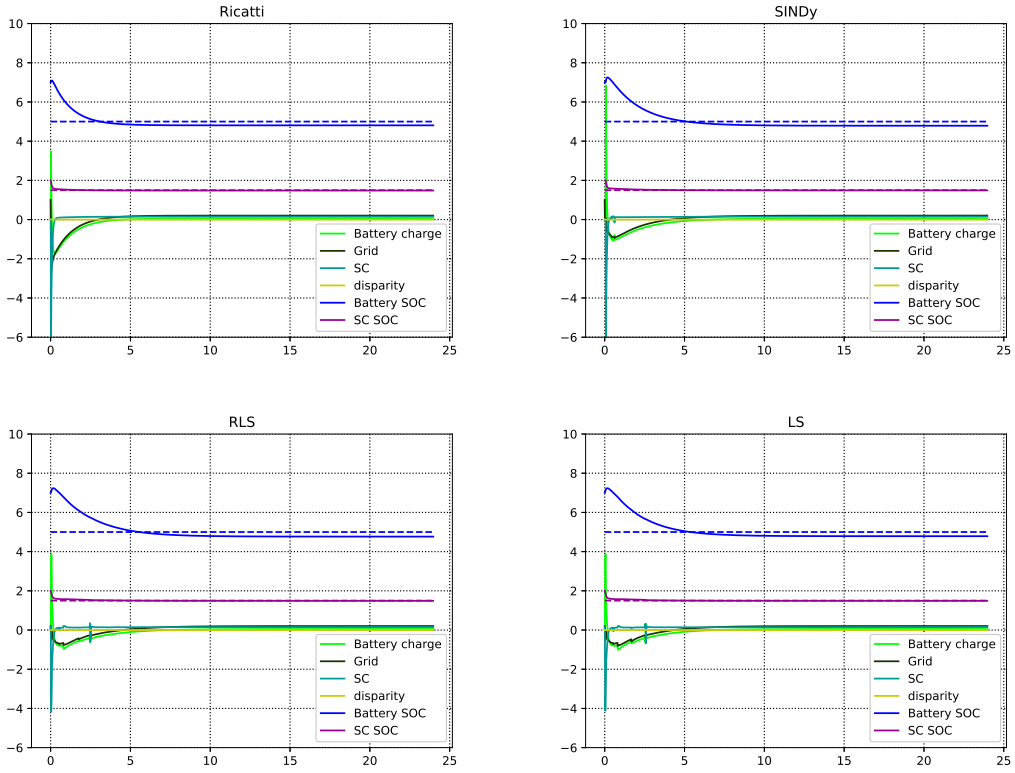


Figure 5.3: Evolution of the SOC and control actions for the steady state reference and no demand.

	Ricatti	SINDy	RLS	LS
Steady state cost	0.0857	0.0891	0.0993	0.0897
Total cost	778.52	926.18	958.06	940.15
max u_{batt}	6.849	3.856	3.863	3.488
max u_{SC}	0.169	0.352	0.323	0.149
max u_{grid}	1.0	0.203	0.203	1.0

Table 5.1: System control without identification phase KPIs.

Lastly, we will compare the steady state cost of each one of the methods, the total cost in all the process and the maximum control actions that will affect to estimate the cost of operation, the longevity of the batteries and the installation needed.

As we can see in the Table [5.1], the best steady state cost is found by the Ricatti equation as expected by the theoretical results. But only by a small margin with respect to the SINDy and LS method.

Regarding the Total Cost of the operation, the best one is again the Ricatti method by a great margin, as it starts approaching the tracking to the reference since the beginning. The other methods must learn the system model parameters and converge the dynamics of the value function at the same time, so it is expected some initial gap.

The Key Performance Indicator that can be improved by the SOL models is the conservativeness at imposing great charging loads to the battery, mostly at the beginning. This may assure a better battery longevity, but ideally, those costs were already taken into account in the modeling.

In our experiments, the SINDy is the one that provides a better and more stable performance if the parameter λ that accounts for the cost of having more parameters is well tuned so that the prediction performance is no hurt. In the other algorithms, despite having a small value, there are some residual components on the rest of the parameters.

5.1.2 Noise rejection

After checking the stabilization probabilities, and the optimality of the learned solutions, we can study how the system will behave if there is some unmodeled disturbance.

In this section, we will evaluate the robustness of the control methods based on data points against some random noise signal of the mean power of about $E[disparity^2] = 0.2kW$ embedded in the disparity between the energy consumption and a the energy production.

In the Figure [5.4], we can see the evolution of the systems under the noisy conditions. The yellow line represents the presence of the power disparity disturbance that must be compensated by the control signals, whereas the control inputs are in different tones of green.

We can observe that in this case, the states matches better the reference, and the control signal will absorb the disturbances. As they are of a very high frequency, and a zero mean, the noise will be integrated in a very short period of time, and therefore not affect to the evolution of the system.

As the dynamics of the battery SOC is not affected by the power disparity explicitly, the system

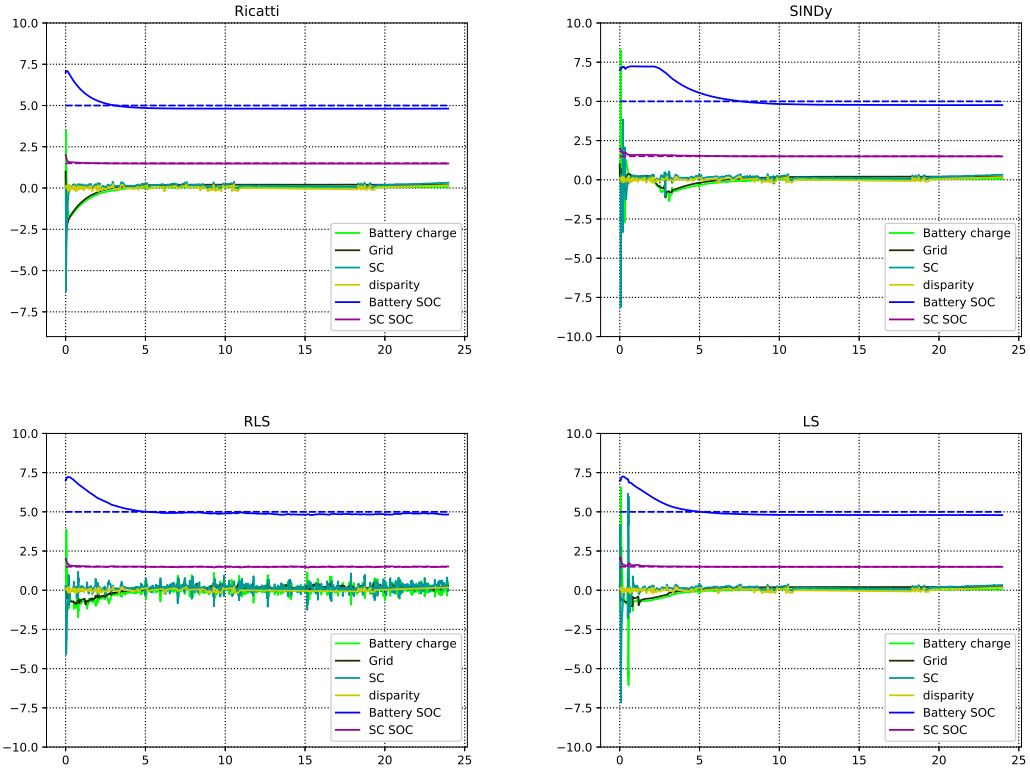


Figure 5.4: Evolution of the SOC and control actions with a zero mean noisy demand disparity.

identification algorithm will not be affected by the presence of noise. However, in the dynamics of the SOC of the supercapacitor, the estimated dynamics depends a lot on the samples chosen due to the relevance of noise. In this scenario, broader samplings are encouraged as well as a wider view of the data. This will favour the LS algorithm in front of the RLS and the SINDy method, that iteratively takes each sample as ground truth.

Despite the presence of noise, and the bad approximations that it triggers in the System Identification algorithms, the generated control signal can stabilize the states to the desired reference and with similar results in the tracking as the noiseless situation.

Figure [5.5] compares the performance of the different methods with the LQR criteria. As we expected, the Ricatti solution provides the more stable and smooth solutions, but the learned methods don not fall short in comparison despite not knowing the real model a priori.

The RLS and LS curves are, again, very similar one to another, and the SINDy methods does take a while longer to reach the desired reference as improving the model predictions is not its unique focus.

Table [5.2] compares the performance of the models. One surprising finding is the performance of the steady state cost with the RLS method. We can observe that the tracking of the battery SOC reference in this case was better, than in the noise free version. This was achieved with a more responsive control signal to the noise, where higher spikes are observed in the charging signals.

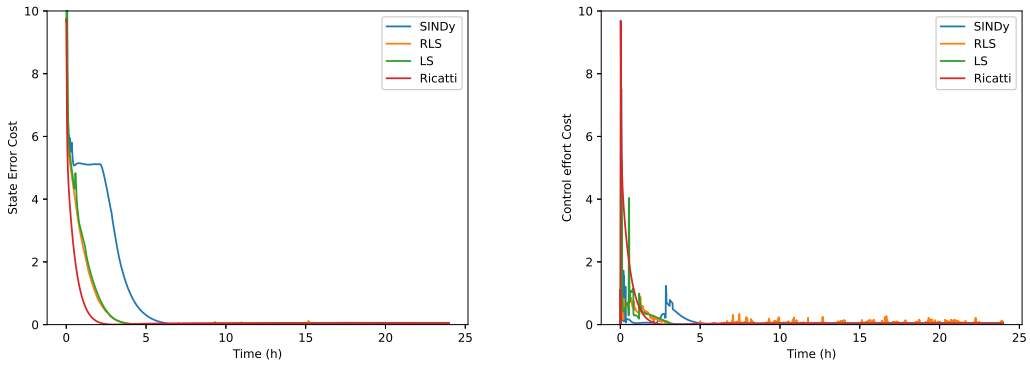


Figure 5.5: State error costs evolution on the left side and control effort ones on the right side of the system with a zero mean noisy disparity.

	Ricatti	SINDy	RLS	LS
Steady state cost	0.0883	0.0911	0.0427	0.0975
Total cost	778.69	2047.46	896.17	963.90
max u_{batt}	3.496	8.276	3.852	6.559
max u_{SC}	6.242	8.145	4.089	7.165
max u_{grid}	2.842	1.654	0.904	2.755

Table 5.2: System control with noisy disturbance KPIs.

However, the total cost of the Ricatti solution is still lower than the other methods due to it's previous knowledge. That compared with the time that the convergence of the system identification parameters added to the convergence of the value function puts them in some disadvantageous position.

5.1.3 Simplified consumption-production profiles

The best simplification that we can find will be given by the composition of a set of sinusoidal functions, that can be easily tracked and predicted. In our case, we will simulate the power consumption and generation disparity with a basal negative power consumption, to negative peaks, one in the morning and another in the evening, and another positive peak at noon representing the power generation of the solar panels.

The formula that we have found roughly can be given by

$$Disparity(t) = 1.7 \cdot sin(\omega_1 t - \pi/4) - sin(\omega_2 t - 7\pi/12) - 0.5 \tag{5.7}$$

where the time t is expressed in hours, and the $Disparity$ is in kW. The frequencies of the Disparity will be $\omega_1 = 3/2\pi/24$ and $\omega_2 = 4\pi/24$. The total balance of the power will be slightly positive of about $0.226kW \cdot h$. This will not be able to compensate the loses due to the battery inefficiencies, but can contribute a little in that sense.

As we can see in the Figure [5.6], it will be clearly prominent the peak in the middle of the day, as the production exceeds by far the consumption. We have also introduced the consumption

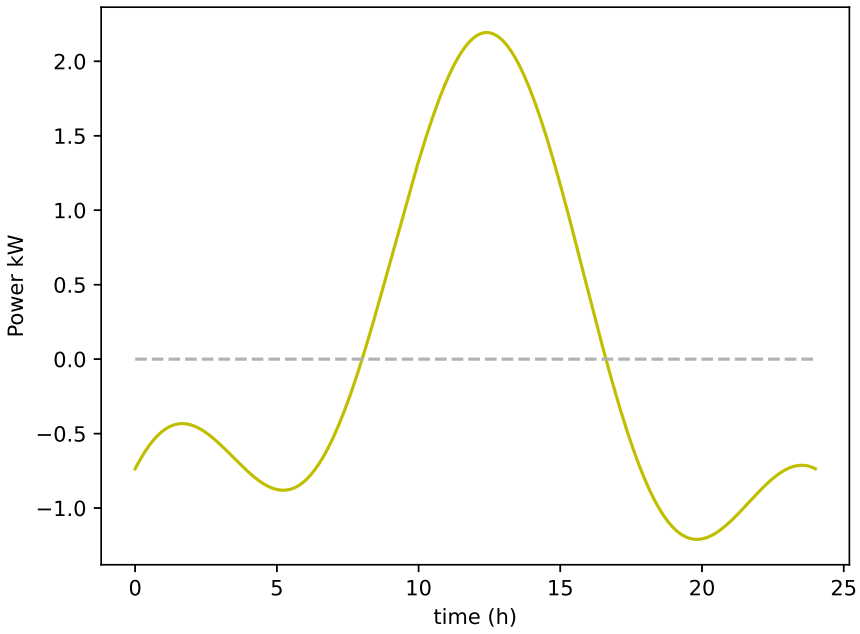


Figure 5.6: Energy disparity profile.

peaks early in the morning at awaking time, and late at night since the diner.

We can note that we have adjusted the function with only two sinusoid signals and a steady state consumption, so the representations will contain some obvious simplifications. However, we have captured the main characteristics of the functions that we wanted to portray.

As the dynamics will no correspond to the learned model, the parameters learned will not correspond any real ones, so we are only going to evaluate the dynamics the the continuous learning process will generate in the control policy and therefore the states transitions.

As we can see in the Figure [5.7], the accumulative information of the LS method drives the control away from the optimal point, and a drastic shift on the control signal is observed at noon as the production starts to decrease. This shift is caused by the accumulated error on the predicted dynamics, that will trigger the new introduction of learning data-points in order to adapt the model to the newly observed data.

The SINDy method will generate a similar evolution as the reference Ricatti equation, where the battery absorbs must of the excess and deficiencies of energy. This will help to reduce the peaks of power that we extract from the grid and therefore, the total cost of getting and producing energy.

The RLS method is the one that adapts the fastest to changing conditions, and in this case there will be no exception. It will have a more nervous systems, where the control actions pikes every now and then, but the recursive nature of the model will make the learning faster and therefore, the tracking more accurate.

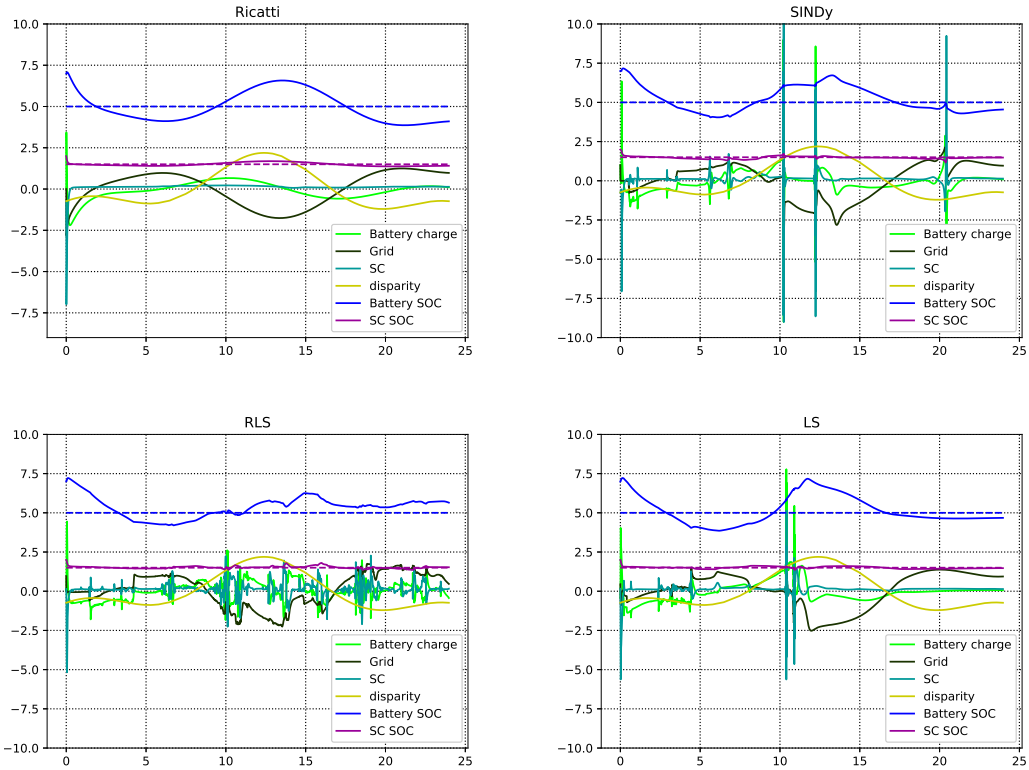


Figure 5.7: Evolution of the SOC and control actions with a simplified demand disparity.

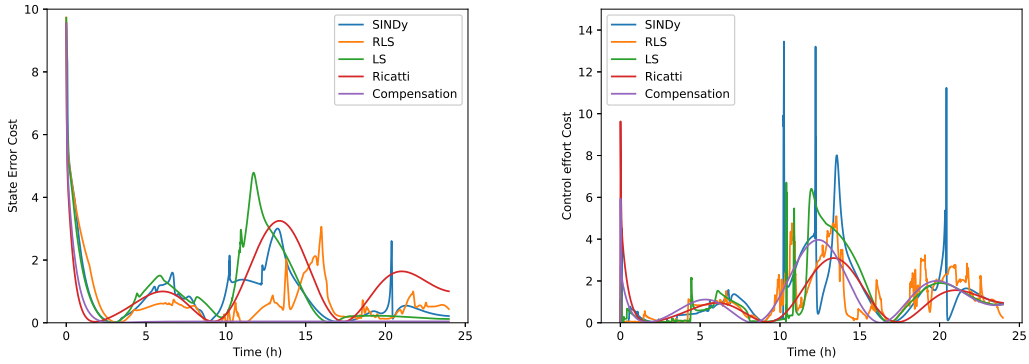


Figure 5.8: State Error costs evolution on the left side and Control effort ones on the right side of a simplified demand disparity.

	Ricatti	SINDy	RLS	LS
Total cost	5156	5161	4433	5444
max u_{batt}	3.433	9.0	4.443	7.779
max u_{SC}	6.987	10.793	5.170	5.629
max u_{grid}	2.822	3.630	2.257	2.523

Table 5.3: System control with simplified demand KPIs.

In the cost curves displayed in the Figure [5.8], we can match the insights that have discussed above. The LS model returns many undesired spikes along the day in both the tracking performance and the control input. The SINDy method also has one single spike that would be needed to be handled by some bound in order to preserve the safety of the components. And the RLS method will lead to a lower tracking cost than the Ricatti solution once the system is stabilized. However, this will come with a higher control cost in most of the cases.

Now we can discuss about some Key Performance Indicators in the Table [5.3]. In this case, there will be no steady state cost, as the system will be more dynamic. We can observe, as we expected, that the total cost of operation will be similar in the SINDy and the Ricatti reference one due to the similarities that we have observed before. But surprisingly, the total accumulated cost of all the day will be lower for the RLS method. This means that the greater flexibility and learning capabilities of this system allowed this model to outperform the Ricatti solution by rejecting a predictable signal of the disparity, despite not having the formulation of the signal nor the basis functions that are adequate for representing it.

The LS is clearly the one that provided worst performance from all the previous KPIs as we would expect from the previous graphics.

Regarding the maximum control signal registered, we have that the u_{grid} is especially stable, as all the methods will remain almost in the same level, except for the LS. The SINDy algorithm suffers from a peak in the other signals in the Battery and the SC controls. But the RLS method is the one that maintains the maximum levels of power under some threshold.

Another factor that we could measure is the average slew rate of the control signals, as a higher slew rate will also affect to the battery performance. But, as we did not take it into account in the definition of the problem, we are going to keep that factor aside.

5.1.4 Internal Model

If the system disturbance have a known structure - either they are generated by another system, or the signal follows a known linear differential equation-, the controller can reject it by incorporating the model of the disturbance within the original model. This approach is known as Internal Model Principle (IMP) first proposed by Francis and Wonham [FW76].

In a State Space representation, the Internal Model can be represented as an augmentation of the plant. The extra dimensions will follow the expected dynamics and can affect linearly the dynamics of the original variables. This characteristics can affect and improve the design of the new controllers thanks to the more accurate knowledge of the system dynamics.

As the synthetic disparity profile is generated by a composition of two periodic signals and a constant, it can be represented with a combination of variables of an Internal Model whose dynamics are known.

The constant term does not need any extra variable, as the original basis of functions already accounts for them. But each one of the periodic terms will need the introduction of two at least two variables. Let consider we have a single frequency sinusoidal function to learn as a disturbance. Then, the augmented plant would follow the subsequent dynamics:

$$\begin{bmatrix} \dot{x}_{1d} \\ \dot{x}_{2d} \end{bmatrix} = \begin{bmatrix} 0 & \omega \\ -\omega & 0 \end{bmatrix} \begin{bmatrix} x_{1d} \\ x_{2d} \end{bmatrix} \tag{5.8}$$

such as the solution of the dynamics can be obtained like a system of Ordinary Differential Equations (ODE) in the Laplace transformed space

$$\begin{cases} sX_{1d} - x_{1d}(0) = \omega X_{2d} \\ sX_{2d} - x_{2d}(0) = -\omega X_{1d} \Rightarrow X_{2d} = \dfrac{-\omega X_{1d} + x_{2d}(0)}{s} \end{cases} \tag{5.9}$$

What can be substituted in the first equation

$$s^2 X_{1d} - s x_{1d}(0) = -\omega^2 X_{1d} + \omega x_{2d}(0) \Rightarrow X_{1d} = \frac{s x_{1d}(0) + \omega x_{2d}(0)}{s^2 + \omega^2} \tag{5.10}$$

whose solution in the temporal domain will be

$$x_{1d}(t) = x_{1d}(0) cos(\omega t) + x_{2d}(0) sin(\omega t). \tag{5.11}$$

The solution of the second variable will be also given by the derivative of the first one

$$\frac{\dot{x}_{1d}(t)}{\omega} = x_{2d}(t) \Rightarrow x_{2d}(t) = -x_{1d}(0) sin(\omega t) + x_{2d}(0) cos(\omega t). \tag{5.12}$$

With those functions, we can generate an arbitrary sinusoidal disturbance as a linear combination of those two new variables of the augmented plant

$$\begin{aligned} d(t) &= c_1 \cdot x_{1d}(t) + c_2 \cdot x_{2d}(t) \\ &= \Big(c_1 x_{1d}(0) + c_2 x_{2d}(0)\Big) cos(\omega t) + \Big(- c_2 x_{1d}(0) + c_1 x_{2d}(0)\Big) sin(\omega t) \end{aligned} \tag{5.13}$$

using this decomposition, we can reproduce the two frequencies that we chose in the first place into those terms by selecting the adequate parameters. Firstly, we can start with the

$$1.7 \cdot sin(3/2\pi/24t - \pi/4) = 1.7 sin(3/2\pi/24t) cos(\pi/4) - 1.7 cos(3/2\pi/24t) sin(\pi/4) \tag{5.14}$$

and a simple identification of the terms can be translated into

$$\begin{aligned} 3/2\pi/24 &= \omega_1 \\ \Big(c_1 x_{1d}(0) + c_2 x_{2d}(0)\Big) &= -1.7 sin(\pi/4) \\ \Big(- c_2 x_{1d}(0) + c_1 x_{2d}(0)\Big) &= 1.7 cos(\pi/4) \end{aligned} \tag{5.15}$$

In order to simplify the terms we can state that the initial conditions can be $[x_{1d}(0), x_{2d}(0)] = [1, 0]$, so the constants can be easily assigned $c_1 = -1.7sin(\pi/4)$ and $c_2 = -1.7cos(\pi/4)$.

To represent the other periodic signal, we can add two more dimensions to the problem in the disturbance hidden dynamics

$$-sin(4\pi/24t - 7\pi/12) = -sin(4\pi/24t)cos(7\pi/12) + cos(4\pi/24t)sin(7\pi/12) \tag{5.16}$$

$$\begin{aligned} 4\pi/24 &= \omega_2 \\ \Big(c_3 x_{3d}(0) + c_4 x_{4d}(0)\Big) &= sin(7\pi/12) \\ \Big(-c_4 x_{3d}(0) + c_3 x_{4d}(0)\Big) &= -cos(7\pi/12) \end{aligned} \tag{5.17}$$

And, as before, simplifying the initial conditions to $[x_{3d}(0), x_{4d}(0)] = [1, 0]$, we can get $c_3 = sin(7\pi/12)$ and $c_4 = cos(7\pi/12)$.

With this signal, we can obtain exactly the same disparity that we had before in the Figure [5.6]. And the plant that we will want to obtain will be

$$\begin{bmatrix} \dot{x}_{states} \\ \dot{x}_d \end{bmatrix} = \begin{bmatrix} A_{states} & 0 \\ 0 & A_d \end{bmatrix} + \begin{bmatrix} B_{states} \\ 0 \end{bmatrix} u \tag{5.18}$$

where the *states* subscript denotes the original linear system, and the d subscript represents the dynamics of the auxiliary terms. In this case, x_d will have four dimensions and an initial state of $[1, 0, 1, 0]$, whereas

$$A_d = \begin{bmatrix} 0 & \omega_1 & 0 & 0 \\ -\omega_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_2 \\ 0 & 0 & -\omega_2 & 0 \end{bmatrix}. \tag{5.19}$$

to make a fair comparison, we are going to introduce a new way of controlling the system. In this case, we will assume that there is a system that is fast enough in order to recognize the current demand in Real Time. And, in order to correct the variable disparity, this system will ask from the grid the power needed to compensate that. This decoupling of the tracking problem with respect to the disparity rejection will make the system behave as with a constant reference tracking problem.

As we can see in the Figure [5.9], the compensated model will have a much better tracking performance, but at the cost of increasing the maximum power extracted from the grid. For the other models, we can see how the knowledge obtained from the hidden model is leveraged to predict the future profile of the disparity, and therefore take more knowledgeable decisions and reduce the total cost in the longer term.

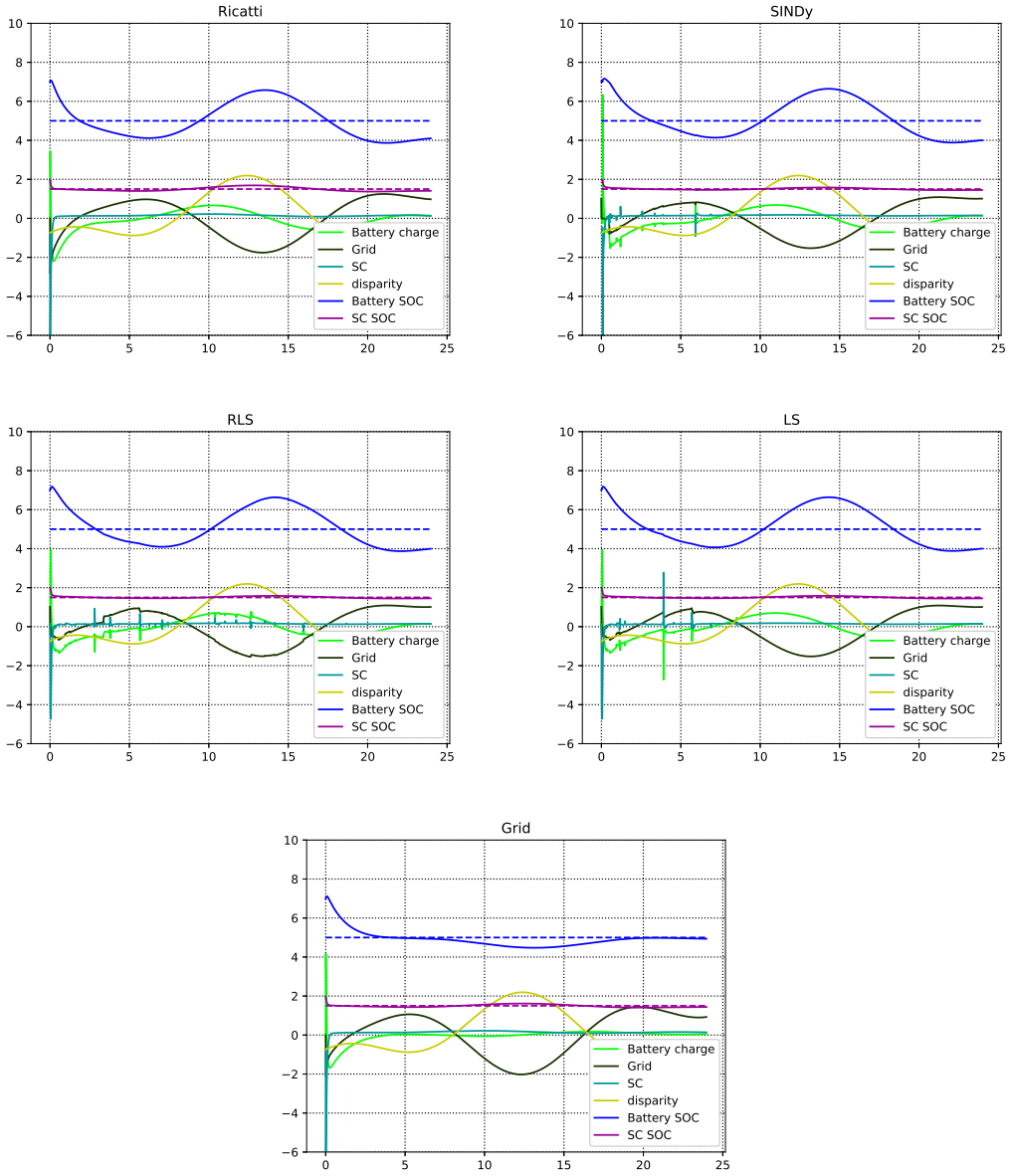


Figure 5.9: Evolution of the SOC and control actions with an augmented plant for the disparity.

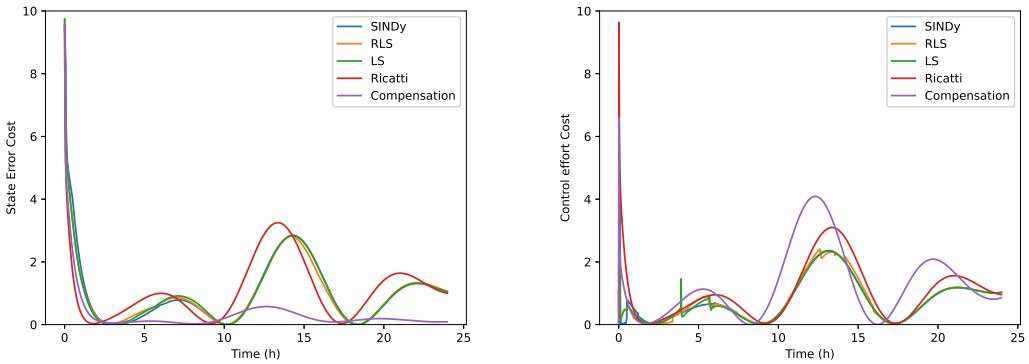


Figure 5.10: State error costs evolution on the left side and control effort ones on the right side of a system with an augmented plant for the disparity.

	Ricatti	SINDy	RLS	LS	Ricatti+FF
Total cost	5166	4241	4227	4201	3968
max u_{batt}	3.433	6.330	3.954	3.947	4.165
max u_{SC}	6.987	7.041	4.733	4.724	6.987
max u_{grid}	2.822	1.531	1.545	1.526	2.090

Table 5.4: System control with an augmented demand KPIs.

All three learning models does obtain a reasonable parametrization of the real system, and are able to generate a good control signal that drives correctly the battery charges. In this case, the learned models with the three methods will also coincide at least up to the third decimal, and will take the form

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \end{bmatrix} = \begin{bmatrix} -0.050 & -0.010 & 0 & 0 & 0 & 0 & 0 \\ -0.646 & 0 & -0.100 & -1.192 & -1.189 & 0.956 & -0.259 \\ 0 & 0 & 0 & 0 & 0.196 & 0 & 0 \\ 0 & 0 & 0 & -0.196 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.524 \\ 0 & 0 & 0 & 0 & 0 & -0.524 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} + \begin{bmatrix} 0.900 & 0 \\ 0.990 & 0.990 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{5.20}$$

From the constants, we can identify the terms that we calculated theoretically, and see how the terms match the ones we predicted like $\omega_1 = 3/2 * \pi/24 \simeq 0.196$, $\omega_2 \simeq 0.524$, $c_1 = c_2 = 1.7sin(\pi/4) \simeq 1.192 \simeq 1.189$, $c_3 = sin(7\pi/12) \simeq 0.956$ and $c_4 = cos(7\pi/12) \simeq 0.259$.

Regarding the cost curves in the Figure [5.10], we can note that there is clear shifting between the standard Ricatti control law and the learned ones. The latter methods also presents lower peaks in each of the curves, that could denote some slight better performance in general. In the other hand, the Ricatti law added to a Feed Forward (FF) law that compensates for the disparity will have a much better performance while tracking the reference with the Ricatti law as we can deduce form the purple curve on the chart. However, this comes with a worse cost in the control

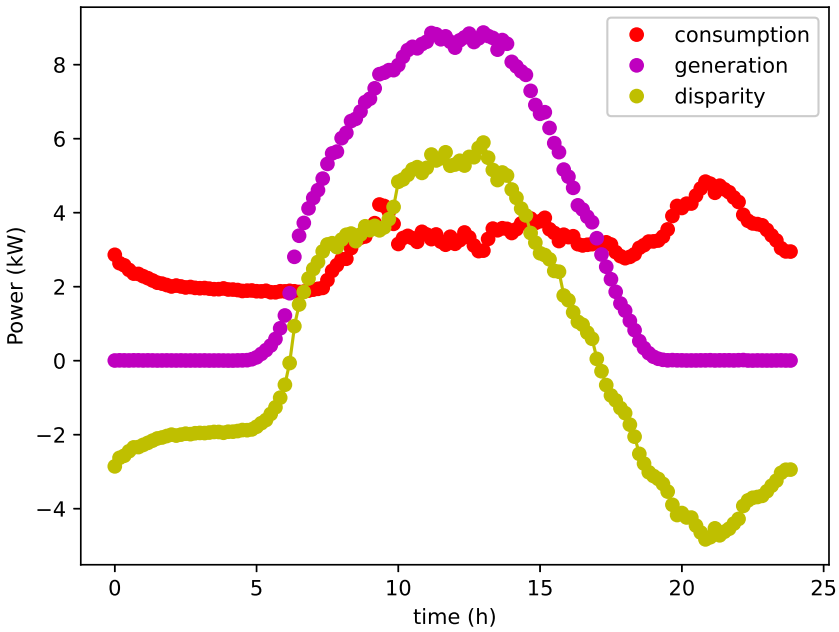


Figure 5.11: Real dataset for generation and consumption profiles

signal.

In the Table [5.4] we have gathered the same KPIs as before. We can see that the compensation of the Disparity with the grid control will result in a better control law as the Total cost is the lower. But this advantage might be a result from the fast control at the initial phase, as no model must be learned since the beginning.

It is very worth noting that the learning systems performed similarly in the three cases, and notably better than the simple Ricatti law. This demonstrate that the extra knowledge of the system can be well leveraged by this controller in order to reduce the long term cost.

Regarding the maximum control signals, the learning model will be more conservative than the Ricatti models when it comes to taking power from the grid. And the SINDy will have the biggest peak in the battery power.

5.2 Experimetal Results

Finally, we tried to implement those systems with a set of real data. The consumption dataset was collected in a regular household of four members in Spain, courtesy of the Automatic Control group at the IRI [NSS+21b]. This dataset is composed by a registration every ten minutes of the power consumed at the moment along several years. This generates a very noisy dataset, as the consumption may drop or skyrocket very easily at any point of time even though it does not average all the ten minutes straight. For example, the microwave use can represent a very high power consumption, but the average time of use is only for a few minutes.

To avoid the highly noisy data, we averaged the consumption over two years. In this way, we

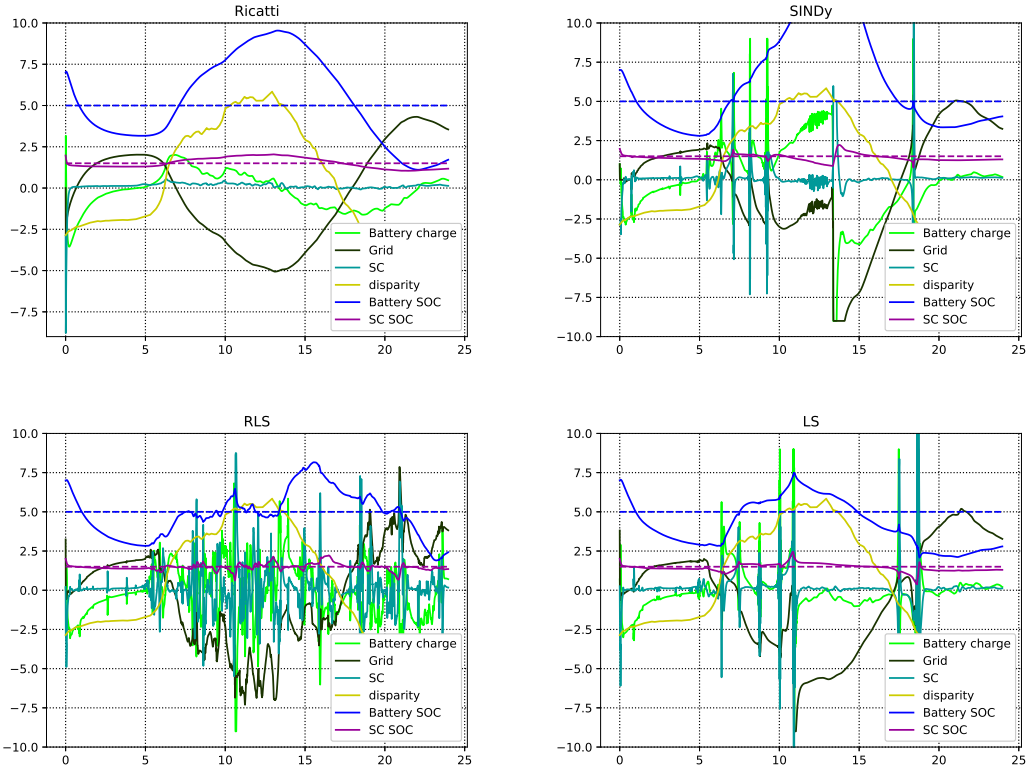


Figure 5.12: Evolution of the SOC and control actions with the real data.

can filter most of the irregular spikes in the consumption, but in the other hand, we might also loose the details about the seasonal profile of the consumption.

Regarding the Generation Power datset, we chose a grid conformed by a few Photovoltaic Panels destined to cover the consumption of a farm in Germany [NSS+21a]. As the power generated along the day was higher than the consumption of the household, we scaled the data in order to match the consumption that we needed.

The total power that the household consumes along the average day is of about $72kW \cdot h$, and we will scale the daily energy generation to about $78kW \cdot h$, so there is room to loose against the inefficiencies of the batteries.

As the disparity data is very sparse for the control purposes that we wanted (the control signal must react in under a minute basis), we have interpolated the data points so the integration can be performed naturally.

In this case, we do not know the main frequencies of the system disparity, so we could not implement the augmented plant that we discussed in the previous context. For that reason, the simple model, with some disparity profile will be implemented for the control law. As we discussed before, this will flavor models that can adapt to new dynamics faster, such as the RLS System Identification method. But it will discourage other methods that wants to balance between the prediction accuracy and the other parameters, such as the SINDy method.

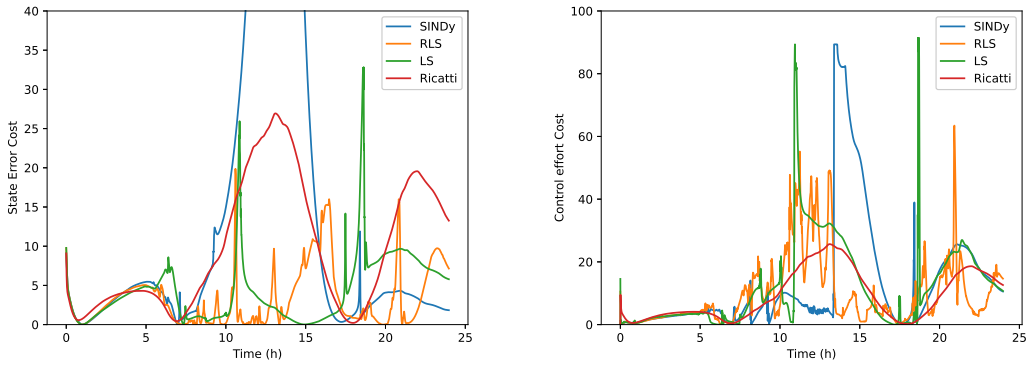


Figure 5.13: Tracking and Control cost the system with real data for the disparity.

	Ricatti	SINDy	RLS	LS
Total cost	45305	78484	31497	39271
max u_{batt}	3.553	9	9	9
max u_{SC}	8.783	12.12	8.755	15.244
max u_{grid}	5.063	9	7.864	9

Table 5.5: System control with real data KPIs.

As we can see in the Figure [5.12], the SOC of the Battery even gets out of bounds when the production is getting much higher than the consumption. For that reason, this method can be discarded in order to preserve the safety operation of the system.

As the Ricatti control law can not be adapted to the current situation, it will only reject the disparity as a disturbance in the states. And due to the unbounded nature of it, the system states does fluctuate dangerously near the limits. The good part is that the control signal will behave smoothly as the system evolves, and it will be able to absorb the excess of consumption and production with the batteries capacity.

In the other hand, the RLS and LS methods present a tighter tracking of the reference at the cost of having more important control signals. Specially in the case of the LS method. The RLS will have a very nervous control, but it can be seen that it does adapt fast to the changing conditions of the system.

From Figure [5.13], we can corroborate the previous hypothesis that the RLS and LS methods will generate a much smaller tracking error than the Ricatti control law at the cost of increasing the operation cost. And, as before, we are going to discard the SINDy method due to the state limits surpassing.

From Table [5.5], it is confirmed that the significant reduction on the tracking cost translates also in a significant reduction in the Total Cost of operation along all the day. But we can see that the hard limitations on the battery power and the grid power are reached in almost all the cases except for the Ricatti control law.

6 Conclusions

In this project, our focus was on investigating methods for controlling a system without possessing specific knowledge about the system's plant and dynamics. We delved into the depths of various methods, particularly highlighting their relevance to the control of electrical systems using reinforcement learning (RL) approaches. Through these approaches, we aimed to generate control policies that minimize a predefined cost function.

During our study, we encountered certain limitations regarding the permissible form of the cost function, as analytical solutions to the equations were required. Specifically, the control signal component had to be quadratic to facilitate differentiation and obtain an equation for solving the optimal control law.

As the model is learned online, and the control signal is generated over the learned model, this techniques can resemble the more established adaptive control techniques, where the controller is parameterized and adjusted online. However, this particular method is based on the Hamilton-Jacobi Bellman equations in order to define the optimal control for each scenario, and the learned model is employed to update the value function that describes the best action. This technique allowed us to achieve remarkably good performance in the tracking problem, even in the presence of noise, and the model could be also employed to learn and control more complex systems with embedded non linearities.

In our study case, the problem that we solved was not an exact match of the original problem. Several simplifications of the problem modeling allowed us to have a more schematic problem without the internal dynamics of the components or the circuits. More over, the parameters that we employed in the simulation model were not tuned with a real system, but employed to illustrate the learning and controlling capabilities of the SOL instead. And the Economical problem was transformed into a reference tracking problem with a set of approximations of the costs, that might not portrait the reality of the problem. For instance, the symmetry in the control costs of each component might not be realistic, as the charging and depleting of storage systems might not be equivalent in terms of efficiency nor battery degradation. And regarding the power extracted from the grid, the price of storing it in a third party station will probably not be at the same cost as buying it from the grid. However, those approximations made the problem tractable for the purposes of this project.

Remarkably, we observed that it is indeed possible to control a system with nearly flawless performance, even when compared to the Ricatti control law in tracking problems. This success was achieved despite the initial learning phase required to acquire the model and converge the Value Function. Furthermore, we discovered that the prediction capabilities incorporated in the model learning process provided an advantage to the SOL algorithm. Leveraging the future dynamics of the system allowed us to minimize the control signal or preemptively compensate for potential tracking disturbances.

Overall, this project served as a significant step towards our broader objective, showcasing promising advancements in our understanding of system control methods employing Robust Reinforcement Learning methods.

In the future, we could work on the approximations that we have taken to solve this problem, make the operation cost more resembling to a solution for the original Economic MPC problem, or even forecast the costs and benefits derived from extracting and injecting power to the grid.

What would shift the current problem of reducing the peaks of power energies into a time shifting of the energy storing, consumption and release to achieve the optimal economic performance.

We expect to develop a further investigation on the RL algorithms oriented to control situations. Regarding this approaches, we have already worked with some model free methods that can learn the control signal without a system identification module. Those works were based on the assumption over the basis functions that can represent the real Value Function, and the learning of the parameters with the sampled values of the cost function and the states. Then, the control can be obtained by optimizing the associated cost in the Value Function, and a function estimator that generates those control signals.

We are also expecting to explore other branches that are also related with the study of RL algorithms that can reproduce and improve the EMPC methods like the line of studies traced by Sebastien Gros and Mario Zanon [GZ20],[ZG21]. Where they demonstrated the equivalence between the EMPC optimality and the RL solutions, and leveraged the best characteristics of both worlds: the easy implementation of hard limits on the control and states signals of the MPC approaches, and the learning adaptability of the RL methods.

7 Time Schedule

As in any research project, a generic time schedule is needed in order to organize and set the intermediate goals towards the finalization of a good thesis on time. In this regard, we proposed an initial time line that scheduled the different tasks involved in the process.

We need to highlight that this project was developed as an extended Master's thesis, what means that the work load must of 30 European Credits Transfer system (ECTS). As for each credit it is expected a dedication of at least 30 hours, the total workload is about 900 hours. Those were split from the $6th$ of February until the $3rd$ of March with a total count of 21 weeks.

The distribution of the expected week dedication can be followed in the Gantt chart of Figure [7.1].

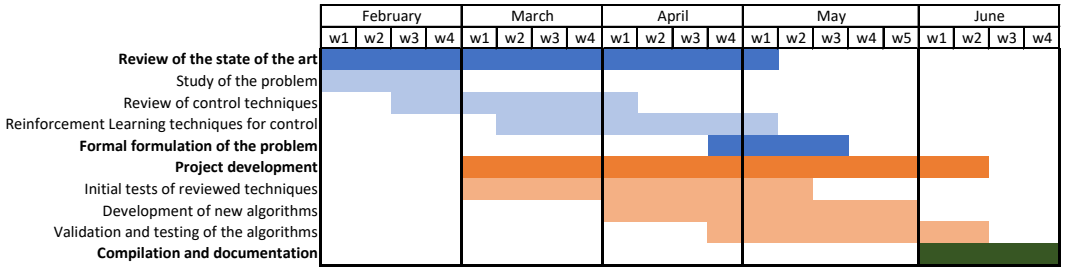


Figure 7.1: Gantt chart of the weeks load distribution.

As we can see, most of the weeks will have overlapped tasks, as the research process is intrinsically a back a forward process between the literature review and the development of the project.

the review of the state of the art will begin with the study of the problem itself, that includes the definition of electrical grids, the study of the working principles of the components and the contextualization of microgrids inside the hierarchical structure of the electric grid management.

Once the problem complexities are overviewed, we can start studying the more classical methods of controlling it, and after that, we can almost simultaneously start with the more modern and unconventional proposals of control.

Alongside with that review of the already existing proposals, we can start developing our part of the project that may include the development of new control algorithms, the adaptation of previous proposals to our specific case and the final validation of them.

The last part of the thesis will be dedicated to the documentation of all the work done along those months, that may include theoretical and experimental results of the problem.

8 Budget

This section is devoted to the description of the economical costs that the development of this thesis has generated. This can be divided into several sources like the working force, the software licensing, the material costs and the energetic costs.

8.1 Working fore

Regarding the working force that has been gathered with the purpose of developing this thesis, we can compute the time of one student and two tutors.

The student has dedicated a full time effort during the four 5 months from the beginning of February until the 3rd of July. This can be computed as the hours worked by week (40 h/week) times the 52 weeks of a year and multipliying it by the fraction of the year worked 12/4. This returns a total of roughly 866 hours worked in the project.

In this case, the project was done in the frame of a PhD program, so the student was perceiving the salary of a regular PhD candidate, in this case of 19.788, 26€ as brute salary per year. This give us a total of 8245€ along the 4 months dedicated to it.

With respect to the directors of the thesis, it can be assumed that a 5% of the total hours dedicated by the student must also be dedicated by each one of the directors including meetings and the consequent followup of the work. This can be translated as dedication of roughly 43.3 hours for each one of the directors. As the average salary of a university professor in Spain is of 50, 600 € a year [sal23], we can compute the percentage of the salary that was dedicated to this particular project.

Asset	Decication (h)	Yearly salary (€)	Year hours	cost (€)
Student	866	19.788,26	2080	8245
First director	43.3	50,600	2080	1053
Second Director	43.3	50,600	2080	1053

Table 8.1: Thesis working force costs

So we can conclude that the total cost associated to the dedicated working force is of roughly 10351 € along all the process.

8.2 Software licensing

In this case, the majority of the code was developed in the python 3.10 programming language, that is under an Open Source licensing. This means that no cost is needed to assume in order to enjoy the use of it.

8.3 Material costs

Regarding the materials used along the development of the thesis, the main asset was the use of a computer in all the phases of the development. Since the initial research and learning of the topic until the code writing and results gathering.

The estimation of the upper bound of a laptop lifetime is about 5 years. Following this, we can compute linearly the depreciation of this particular asset along the used time of 5 months.

As the total cost of the laptop was 1800 €, the depreciation cost can be computed as

$$1800€ \times \frac{5 months}{12 months/year \cdot 5 years} = 150€. \tag{8.1}$$

8.4 Energetic costs

Regarding the energetic expenses, the main costs can be associated to the laptop, as the illumination and conditioning of the working environment can be split in all the workers of the office.

As the average consumption of a high end laptop is of $180W$ and the average electricity price along the office hours is of $0.185€/kWh$, we can compute the total consumption of electricity and the associated cost to it.

$$180W \times 693h \frac{1kW}{1000W} \times 0.185€/kWh = 23.08€ \tag{8.2}$$

8.5 Total associated cost

Finally, the total associated cost devoted to the development of the project can be estimated as the summation of all the previous costs.

This gives us a total cost of

$$10351€ + 0€ + 120€ + 23€ = 10524€. \tag{8.3}$$

Where the majority of the costs are associated to the staff involved in the development of the project

9 Environmental impact

Regarding the environmental impact of the thesis, most of consumption is related to the building and transportation of the needed materials for the development of the thesis, and the other part is related to the production of the energy resources employed.

Despite its inaccuracies, one generally accepted way of calculating the environmental impact of a project can be associated to the carbon footprint that is measured in Kg of $CO2$ emitted to the atmosphere.

9.1 Material footprint

In this regard, we will only study the emissions that the production and transportation of a laptop may generate. The average emissions during the building of the product is estimated as $331KgCO2$[Hau22] and usually constitutes between the 75% and 85% of the total footprint of the laptop. As to the transportation of the product and the assembling pieces, until reaching the final consumer, it is estimated of about $33KgCO2$ [Hau22].

If we add them and take the proportional part of the lifetime of the asset, we can have an estimation of the carbon footprint associated to the employed materials.

$$(331KgCO2 + 33KgCO2) \times \frac{4months}{60months} = 24.27KgCO2. \tag{9.1}$$

9.2 Energetic footprint

Regarding the carbon footprint that the energy production has in Spain, we can observe a descendant tendency over the last twenty years thanks to the introduction and consolidation of renewable sources inside the energy mix[Tis23]. This website also provides an estimation of $0.166KgCO2$ per kWh for the Spanish electrical mix, so we can compute the total electrical footprint associated to the laptop consumption of electricity

$$180W \times 693h \frac{1kW}{1000W} \times 0.166KgCO2/kWh = 20.7KgCO2. \tag{9.2}$$

9.3 Environmental impact of DER

Regardless of the previous carbon footprint, we can also consider the negative carbon footprint that the implementation of and efficient system that include Distributed Energy Resources, with smaller losses thanks to the proximity between production and consumption, and the wider installation of renewable electrical sources.

The development and evolution of the knowledge in this field can bring major changes in the electrical industry, not only in the local scope, but even in the global one. But it is hard to estimate the impact that one single thesis may have in any of those developments.

10 Social impact

The social impact of this thesis is tightly related to the development of new control techniques that can make a more reliable and more economical efficient management of the Distributed Energy Resources.

As we cannot asses the particular impact of the current thesis inside this topic, any work towards this objective may produce a positive influence on the development of the field.

We have stated that a robust management of the microgrids that contains production, consumption and storage elements can make viable the correct working of the system under harsh circumstances. This can affect positively to communities that are isolated from a wider electrical grid, as the understanding of the needs and the management conditions can help to design the installation of the electrical system and bring that resource to them.

This robustness in the supply may also be of vital relevance under natural disasters, where some communities may suffer from cutoffs from the grid, and it may be important to maintain the supply to the most critical infrastructure like hospitals or communication systems.

In the other hand, the economical benefits can come from a correct operation, that can optimize the costs of every household by selling the electricity when the consumption is high and buying when the price is low. This personal benefit from each one of the "prosumers" can affect entire communities and therefore depict a generalized saving in the electrical bill of all the participants.

The economical benefit may also attract the investment in this types of systems of all types of particulars. This could change the landscape of the electrical production, where the majority of the population may take part and conforming a social change in the electrical consumption and production habits.

Acknowledgments

I would like to express my heartfelt gratitude to everyone who contributed to the completion of this thesis, marking the culmination of my Master's Degree and the beginning of a new chapter in my life.

To my family, for their unwavering support throughout my journey. Their constant encouragement and sacrifices have made it possible for me to pursue higher education and reach this milestone.

To Ramon Costa and Vicenç Puig, the thesis directors who guided me throughout the entire process of research and personal growth required to achieve this significant accomplishment. Their invaluable assistance and mentorship have been instrumental in my success.

To all the teachers I have had throughout my academic journey, from my primary education to high school, bachelor's degree, and now Master's degree. Your dedication and commitment to education shape the future of our society, and I am grateful for the knowledge and skills you have imparted upon me.

To my friends, both those from my hometown who have been with me since childhood and continue to support me, and those I have made in Barcelona, my current city—I am eternally grateful for your companionship. Your presence brings joy, serenity, and countless memorable experiences filled with board games, muixeranga, museum visits, trips, and get-togethers. To los cuates, I have never imagined encountering such great people during our Master's degree. To the friends I have yet to meet in the future, I am excited to embark on new connections and create wonderful memories together.

To La Rioja, my birthplace, which has witnessed my growth and transition into adulthood. It is within your borders that my passions took shape, and where I discovered my current pursuits. I am indebted to your people, your rivers, your mountains, and your fields for shaping who I am today.

To each and every one of you, thank you.

Ce Xu

References

[AA22] Mohammed H. Alabdullah and Mohammad A. Abido. Microgrid energy management using deep q-network reinforcement learning. *Alexandria Engineering Journal*, 61(11):9069–9078, 2022.

[Bel52] Richard Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences*, 38(8):716–719, 1952.

[CCR+20] Andreu Cecilia, Javier Carroquino, Vicente Roda, Ramon Costa-Castelló, and Félix Barreras. Optimal energy management in a standalone microgrid, with photovoltaic generation, short-term storage, and hydrogen production. *Energies*, 13(6), 2020.

[CI12] Aranya Chakrabortty and Marija D. Ilić. *Control and Optimization Methods for Electric Smart Grids*, volume 3. Springer Nature, 2012.

[DMBB23] Arindam Dutta, Shirsendu Mitra, Mitali Basak, and Tamal Banerjee. A comprehensive review on batteries and supercapacitors: Development and challenges since their inception. *Energy Storage*, 5(1):e339, 2023.

[FL23] Milad Farsi and Jun Liu. *Model-Based Reinforcement Learning*. IEEE Press Series on Control Systems Theory and Applications, 2023.

[FOF13] P. Finn, M. O'Connell, and C. Fitzpatrick. Demand side management of a domestic dishwasher: Wind energy gains, financial savings and peak-time load reduction. *Applied Energy*, 101:678–685, 2013. Sustainable Development of Energy, Water and Environment Systems.

[FW76] B.A. Francis and W.M. Wonham. The internal model principle of control theory. *Automatica*, 12(5):457–465, 1976.

[GZ20] Sebastien Gros and Mario Zanon. Data-driven economic nmpc using reinforcement learning. *IEEE Transactions on Automatic Control*, 65:636–648, 2 2020.

[Hau22] Nathan Haughton. What is the carbon footprint of a laptop?, Jul 2022.

[HPG18] Adam Hirsch, Yael Parag, and Josep Guerrero. Microgrids: A review of technologies, key drivers, and outstanding issues. *Renewable and Sustainable Energy Reviews*, 90:402–411, 2018.

[JWX+19] Ying Ji, Jianhui Wang, Jiacan Xu, Xiaoke Fang, and Huaguang Zhang. Real-time energy management of a microgrid using deep reinforcement learning. *Energies*, 12(12), 2019.

[LTS+19] Ll. Lledó, V. Torralba, A. Soret, J. Ramon, and F.J. Doblas-Reyes. Seasonal forecasts of wind power generation. *Renewable Energy*, 143:91–100, 2019.

[MKS+13] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *DeepMind Technologies*, 2013. cite arxiv:1312.5602Comment: NIPS Deep Learning Workshop 2013.

[MVG20] E.C. Malz, V. Verendel, and S. Gros. Computing the power profiles for an airborne wind energy system based on large-scale wind data. *Renewable Energy*, 162:766–778, 2020.

[Nas20] Mohamadou Nassourou. *Robust Economic Model Predictive Control of Smart Grids*. PhD thesis, Universitat Politècnica de Catalunya, 2020.

[NBP20] Mohamadou Nassourou, Joaquim Blesa, and Vicenç Puig. Optimal energy dispatch in a smart micro-grid system using economic model predictive control. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 234(1):96–106, 2020.

[NSS+21a] Unnikrishnan Raveendran Nair, Monika Sandelic, Ariya Sangwongwanich, Tomislav Dragičević, Ramon Costa-Castelló, and Frede Blaabjerg. An analysis of multi objective energy scheduling in pv-bess system under prediction uncertainty. *IEEE Transactions on Energy Conversion*, 36(3):2276–2286, 2021.

[NSS+21b] Unnikrishnan Raveendran Nair, Monika Sandelic, Ariya Sangwongwanich, Tomislav Dragičević, Ramon Costa-Castelló, and Frede Blaabjerg. Grid congestion mitigation and battery degradation minimisation using model predictive control in pv-based microgrid. *IEEE Transactions on Energy Conversion*, 36(2):1500–1509, 2021.

[Phi22] Phi4tech. Batteries and supercapacitors: What is the difference?, Apr 2022.

[Rue21] Gero Rueter. Where do large solar power plants pay off? – dw – 07/26/2021, Jul 2021.

[sal23] salaryexplorer. How much money does a person working as professor - education make in spain?, 2023.

[Sha20] Umair Shahzad. Significance of smart grids in electric power systems: A brief overview. *Journal of Electrical Engineering, Electronics, Control and Computer Science - JEEECCS*, 6(19):7–12, 2020.

[SLP+22] Jigar S. Sarda, Kwang Lee, Hirva Patel, Nishita Patel, and Dhairya Patel. Energy management system of microgrid using optimization approach. *IFAC-PapersOnLine*, 55(9):280–284, 2022. 11th IFAC Symposium on Control of Power and Energy Systems CPES 2022.

[SSM+22] Vikash Kumar Saini, Ravindra Singh, Dinesh Kumar Mahto, Rajesh Kumar, and Akhilesh Mathur. Learning approach for energy consumption forecasting in residential microgrid. In *2022 IEEE Kansas Power and Energy Conference (KPEC)*, pages 1–6, 2022.

[Tis23] Ian Tiseo. Spain: Power sector carbon intensity 2000-2021, Feb 2023.

[ZG21] Mario Zanon and Sebastien Gros. Safe reinforcement learning using robust mpc. *IEEE Transactions on Automatic Control*, 66:3638–3652, 8 2021.