Approximate Dynamic Programming Methods for Residential Water Heating

by

Matthew H. Motoki

A thesis submitted in partial fulfillment for the degree of Master’s of Science in the Department of Electrical Engineering

December 2015
“There’s a way to do it better - find it.”

Thomas A. Edison
In this thesis, we address the problem of minimizing the operating cost of a residential water heater while maintaining a desired level of comfort for the customer. We state the problem as a discrete-time finite-state average cost Markov decision problem (MDP). We view hot water usage as a random process and develop a model of the water heater system. We develop approximate dynamic programming algorithms to solve the MDP. Specifically, we use aggregation to obtain a simplified, but related problem, we use density estimation to calculate transition probabilities, and we consider the Q-Learning algorithm that can be used when a model of the water heater is not known and/or the transition probabilities are not available. We prove that our algorithms can be at least as good as existing methods, in terms of minimizing the objective cost. Using numerical simulations, we evaluate our algorithms’ performance. Our simulations suggest that our algorithms can decrease operating costs by about 15% while maintaining a specified level of comfort. Finally, we discuss modifications to the basic water heater optimization problem that apply to solar water heating and automated demand response.
*Acknowledgements*

I would like to express my sincere gratitude to my advisor Professor Anthony Kuh, not only for the continuous guidance throughout my Master’s study and related research, but also for granting me the freedom to explore different methodologies and techniques. Under his direction, I gained a much deeper understanding of machine learning and optimization algorithms. His counsel has given me an appreciation of how these algorithms can be applied throughout the smart grid.

In addition to Professor Kuh, I would like to thank Dr. Matthias Fripp, who sparked my interest in renewable energy and smart grid technologies. I am grateful that he allowed me to join his research group, where I was able to learn about much more than just water heater optimization. The numerous discussions we had lead to new ways of approaching and understanding the problem.

I would also like to thank Dr. Gürdal Arslan, who introduced me to the field of optimization. I thank him for his insightful comments on my thesis, as well for asking the hard questions that motivated me to delve deeper into the details of the problem; much of the progress that I made during this thesis is due to meticulous thinking about the details.

A special thanks goes to Dr. Maxim Ballmer, who mentored me as an undergrad and provided my first exposure to the ins and outs of academia. Under his supervision, I learned a general framework for conducting scientific research, which I continue to use today.

I would like to thank the University of Hawai‘i at Mānoa and in particular the entire Electrical Engineering Department, for providing me the opportunity to further my knowledge.

Lastly, I would like to express my appreciation to my family and friends for their continuous support; especially my mother, who provided me numerous valuable resources, and Karissa, who made it possible for me to focus on this thesis during my last semester.
Contents

Abstract ii

Acknowledgements iii

List of Figures vii

List of Tables viii

Abbreviations ix

1 Introduction 1

2 Context 5
  2.1 Conventional Power Production .............................................. 5
  2.1.1 Conventional Power Grid .................................................. 5
  2.1.2 Conventional Power Generation ........................................... 6
  2.2 Renewable Energy Sources ................................................... 7
  2.3 The Smart Grid ................................................................. 8
    2.3.1 Smart Metering ........................................................... 9
    2.3.2 Solar Water Heating ..................................................... 9
    2.3.3 Automated Demand Response ......................................... 11

3 Problem Formulation 13
  3.1 Model Variables ............................................................. 13
    3.1.1 State Variable .......................................................... 13
    3.1.2 Decision Variable ...................................................... 14
    3.1.3 Hot Water Usage (Disturbance Variable) .......................... 15
  3.2 State Dynamics .............................................................. 17
    3.2.1 Discrete-Time Dynamics ............................................... 17
    3.2.2 Usage History Dynamics ............................................... 17
    3.2.3 Temperature Dynamics ............................................... 17
  3.3 Objective Function .......................................................... 18
    3.3.1 Operating Cost ......................................................... 19
    3.3.2 Discomfort Cost ....................................................... 19
  3.4 Problem Statement ........................................................... 20

4 Methodology 21
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The conventional power system [1]</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>The amount of renewable generation in Hawai‘i during the past seven years [2]</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>Schematics of passive and active solar water heating systems [3]</td>
<td>10</td>
</tr>
<tr>
<td>4.1</td>
<td>A comparison of hard aggregation, coarse aggregation, and theoretical trajectories.</td>
<td>25</td>
</tr>
<tr>
<td>5.1</td>
<td>Visualization of the a typical model with three household members.</td>
<td>39</td>
</tr>
<tr>
<td>5.2</td>
<td>Finite-horizon and infinite horizon average cost DP daily cumulative costs and temperature trajectories for $\alpha = 0.01$.</td>
<td>41</td>
</tr>
<tr>
<td>5.3</td>
<td>Finite-horizon and infinite horizon average cost DP daily cumulative costs and temperature trajectories for $\alpha = 0.99$.</td>
<td>42</td>
</tr>
<tr>
<td>5.4</td>
<td>Objective costs for the finite-horizon solution ($M = N, 2N$).</td>
<td>43</td>
</tr>
<tr>
<td>5.5</td>
<td>Stochastic and deterministic DP weekly cumulative costs and temperature trajectories for $\alpha = 0.005$.</td>
<td>45</td>
</tr>
<tr>
<td>5.6</td>
<td>Stochastic and deterministic DP weekly cumulative costs and temperature trajectories for $\alpha = 0.995$.</td>
<td>46</td>
</tr>
<tr>
<td>5.7</td>
<td>Tradeoff curve for the no-feature DP and the simple set-point method under a time-varying price for power.</td>
<td>48</td>
</tr>
<tr>
<td>5.8</td>
<td>Tradeoff curve for the no-feature DP solution and the simple set-point solution under a constant price for power.</td>
<td>50</td>
</tr>
<tr>
<td>5.9</td>
<td>Average weekly cumulative costs and temperature trajectories for $\alpha = 0.005$.</td>
<td>52</td>
</tr>
<tr>
<td>5.10</td>
<td>Average weekly cumulative costs and temperature trajectories for $\alpha = 0.995$.</td>
<td>53</td>
</tr>
<tr>
<td>5.11</td>
<td>Tradeoff curve for aggregation schemes ($\Delta T = 1, 1/3, 1/10$).</td>
<td>54</td>
</tr>
<tr>
<td>5.12</td>
<td>Average weekly cumulative costs and temperature trajectories for $\alpha = 0.005$.</td>
<td>55</td>
</tr>
<tr>
<td>5.13</td>
<td>Average weekly cumulative costs and temperature trajectories for $\alpha = 0.995$.</td>
<td>56</td>
</tr>
<tr>
<td>5.14</td>
<td>Tradeoff curve for features measuring durations.</td>
<td>57</td>
</tr>
<tr>
<td>6.1</td>
<td>Block diagram of the customer-utility optimization problem. In the static version of the problem, the stage $k$ is fixed.</td>
<td>60</td>
</tr>
<tr>
<td>A.1</td>
<td>A block diagram of the water heater hardware.</td>
<td>66</td>
</tr>
<tr>
<td>C.1</td>
<td>Two approaches for obtaining a truncated Gaussian distribution.</td>
<td>73</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Average Capacity Factors for Renewable Resources [4] .......................... 8
4.1 Features Summarizing Usage History .................................................. 26
5.1 Numerical Simulation Parameters ......................................................... 37
B.1 The parameters of the water heater model for different family sizes. ............. 70
C.1 Shower Probability Distribution Parameters. ........................................ 72
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADP</td>
<td>Approximate Dynamic Programming</td>
</tr>
<tr>
<td>ADR</td>
<td>Automated Demand Response</td>
</tr>
<tr>
<td>DP</td>
<td>Dynamic Programming</td>
</tr>
<tr>
<td>DR</td>
<td>Demand Response</td>
</tr>
<tr>
<td>KDE</td>
<td>Kernel Density Estimation</td>
</tr>
<tr>
<td>LP</td>
<td>Linear Programming</td>
</tr>
<tr>
<td>MDP</td>
<td>Markov Decision Problem</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>VI</td>
<td>Value Iteration</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

In this thesis, we address the general problem of optimizing a residential water heater. By adding sensors to monitor a water heater, we are able to develop an approximate dynamic programming (ADP) algorithm that minimizes a weighted sum of a customer discomfort cost and the operating cost of the water heater. The details of the ADP algorithm are worked out in detail and we prove, mathematically, that the ADP control strategy is better than existing methods. We carry out numerical simulations to quantify the increased energy efficiency obtained by using the ADP algorithm. We discuss modifications of the basic ADP algorithm that address the related problems of solar water heating and automated demand response.

Water heaters contribute to a significant portion of residential electricity bills and according to [5], nearly 17% of the average US household’s electricity bill is due to water heating. It is also fair to say that the amount of energy used towards water heating is higher than 17% during the winter season and in cold climates. Because water heaters experience heat losses to the environment that are proportional to the temperature difference between the water heater and the environment, maintaining a high water temperature will waste energy. On the other hand, maintaining a low water temperature will likely result in discomfort for the user. Finding a balance between minimizing losses and maintaining comfort is a primary focus of this thesis.

In the absence of any optimization procedure, the set-point of the water heater can be chosen subjectively, possibly based on average customer hot water needs. Choosing a set-point(s), with the purpose of maximizing energy efficiency, has been the subject of previous work on water heater optimization. The simplest way to optimize a water heater is to use a timing switch that automatically shuts off the water heater at preprogrammed times. When the water heater is
not turned off by the timer, it operates as normal. This strategy can be easily implemented using a store-bought water heater timer, such as those in [6]. The effectiveness of this approach depends on the customer’s expert knowledge of household usage patterns. When choosing the timer settings, the customer must also be mindful of reheating times following the period when the water heater is shut off.

More sophisticated approaches with time-varying set-points have been studied [7], [8], [9], [10] and [11]. For example, [7] develop a linear programming (LP) algorithm that minimizes the operating cost of the water heater, subject to comfort constraints explicitly given by a customer-defined operating range. However, this method also requires expert knowledge of the customer when choosing the customer-defined operating range. It is unlikely that the customer will be able to choose an operating range that accurately reflects their desired tradeoff between savings and comfort. An easier way for the customer to choose the right tradeoff is to reduce the set of parameters (set-point operating range) defining the tradeoff into one or two parameters. One way to do this is to minimize the convex combination of the expected operating cost of the water heater and some function (with possibly some customer defined parameter) representing the discomfort of the customer. The tradeoff parameter is then the weight in the convex combination. In addition to reducing customer effort, this approach also offers more flexibility when modeling the objective. For example, we can choose the discomfort function to be such that there is discomfort only when hot water is being used. Thus, it is possible to define an operating range that the temperature may fall outside of without incurring a discomfort cost. We develop this approach in more detail in Section 3.3. To make the process of choosing the right tradeoff parameter more intuitive, it would be good to visualize the effects of varying the tradeoff parameter. In Section 5.2, we construct a visualizations that shows the tradeoff curve between savings and comfort, parameterized by the tradeoff parameter (e.g., see Figure 5.7).

The research in this thesis is similar to that of [12] and [13], in that the primary means of optimization is approximate dynamic programming, but with the notable difference that we optimize a water heater rather than an entire home. Also similar is the work of [14], where approximate dynamic programming is used to optimize a cooling system for a household. The research in this thesis is very similar to that of the researchers at the University of Leuven who also optimize a water heater using approximate dynamic programming (specifically variants of the model-free Q-Learning algorithm) [15], [16], and [17]. The main difference between the algorithms considered here and those considered in the aforementioned papers is that, here, we use a model of the system, while the algorithms in the previous work implicitly learn the
model from experience. Specifically, by focusing our efforts on a single major appliance, we are able derive a scalar first order linear difference equation modeling the system. In addition, we develop a random process to model the hot water usage of a household. With a model of the usage, we can perform numerous numerical simulations to evaluate our algorithms.

In the current power market, utilities supply power at a fixed rate. However, the cost of producing power can vary from one day to another depending on the demand or the amount of operational generating units [18]. Furthermore, on the demand side, fixed rates do not offer customers any incentive to shift or curtail power consumption. Thus, in the current power market, there is a mismatch between the cost of producing power and the cost of consuming power. Demand response can be used to align these costs. By charging customers more during peak load times and less during low load times, customers are encouraged to shift power consumption of the peak load time on to the low load times. Because water heaters are well insulated, they are able to store energy and function as a battery. Thus, in the presence of a time-varying price of power, an intelligent water heater strategy can possibly save money by heating when the price of power is low and storing this energy (while incurring some heat losses to the environment) until it is needed. This shift of energy consumption is, indeed, a simple form of demand response. We discuss using optimized water heaters to do automated demand response in more detail in Section 2.3.3 and we evaluate the water heaters using numerical simulations in Section 6.2.

In summary, we formulate a residential water heater optimization problem with the objective of minimizing a weighted sum of the operating cost of the water heater and a customer discomfort cost. We allow the customer to set the weighting parameter and, therefore, specify their desired tradeoff between the objectives. We derive a scalar first order linear difference equation to model the thermodynamics of a water heater. We model hot water usage as a cyclostationary random process and estimate the probability density function of the usage over time. Our goal is to find an on/off policy for the water heater that minimizes the objective cost. Once a water heater is optimized, we vary the price of electricity to assess the water heater’s potential for demand response; in particular, we look at the expected discomfort of a customer, the expected use and the expected operating cost of the water heater.

In Chapter 2, we discuss how the research work of this thesis fits into the bigger picture of sustainability and energy efficiency. In Chapter 3, we model the residential hot water heater dynamics by defining the key variables and equations and use them to state our problem as an average cost Markov Decision Process. Next, in Chapter 4, we develop approximate dynamic
programming (DP) algorithms to solve the problem. In Chapter 5, we compare our algorithms to existing strategies and show that the approximate DP methods are asymptotically no worse than (and in fact, superior to) many of the existing methods. In Chapter 6, we discuss possible directions for further research. Finally, in Chapter 7, we provide a detailed summary of the work included in this thesis.
Chapter 2

Context

In this chapter, we will focus on why the research in this thesis is important. First, we discuss the flaws in conventional power production and why it is necessary to change the way we currently produce power. Next, we discuss renewable energy sources that are an attractive alternative to conventional power production. Lastly, we discuss the smart grid, which will aid in increasing energy efficiency and is necessary to incorporate renewables into the existing power system.

2.1 Conventional Power Production

In this section, we discuss the conventional power system and the current challenges it faces.

2.1.1 Conventional Power Grid

In the existing power grid (see Figure 2.1), power is generated at a few power plants. Power can be distributed with less losses at high voltages. Thus, after generation, power is efficiently sent along high voltage transmission lines to distribution substations. The voltage is decreased at these substations and transmitted along distribution lines to distribution transformers. At distribution transformers, the voltage is further reduced and sent along secondary distribution lines to customers. In this system, demand is assumed to be more or less predictable so that generation can be increased or decreased to adequately meet demand. Furthermore, in this system, power generally flows in one direction (from generation down to where it is needed).
2.1.2 Conventional Power Generation

In conventional power systems, power is generated by burning fossil fuels such as coal, natural gas, and crude oil. These resources—although natural—are being depleted faster than they are being replenished, and are thus, deemed unsustainable. Furthermore, burning fossil fuels at a global scale releases copious amounts of CO$_2$ into the Earth’s atmosphere. It is strongly believed that this release of CO$_2$ contributes significantly to global warming [19]. Global warming is defined as “an increase in the earth’s average atmospheric temperature that causes corresponding changes in climate and that may result from the greenhouse effect” [20]. This warming can increase heat stress, disease, severity of tropical storms, ocean acidity, sea levels, and the melting of glaciers, snow pack, and sea ice [21]. The large scale burning of fossil fuels also creates air pollution, which can, in turn, cause and increase the severity of public health problems such as asthma and other respiratory illnesses [19]. As a result, the combustion of fossil fuels for power
Figure 2.2: The amount of renewable generation in Hawai`i during the past seven years [2].

generation is being phased-out by many of the world’s top industrialized nations; e.g., the U.S. is aiming to cut emissions to 80% below 1990 levels by 2050 [22]. As the most oil dependent state in the nation [23], Hawai`i has strengthened its commitment to promoting clean energy by pledging to accomplish a 100% renewable portfolio standard by the end of 2045 [24].

2.2 Renewable Energy Sources

Renewable energy sources (renewables), defined as “any naturally occurring, theoretically inexhaustible source of energy that is not derived from fossil or nuclear fuel” [25], are an attractive alternative to burning fossil fuels. In Hawai`i, about 21% of the state’s power comes from renewable energy sources [26]. Examples of renewables are solar power (photovoltaics and solar water heating), wind power, and geothermal power. Figure 2.2 shows the different amounts of renewable generation in Hawai`i during the past seven years. Capacity factors, defined as the percent of time a facility is expected to operate at full capacity (or its equivalent, in terms of energy production) over a one year period [2], are a common way to quantify the efficiency of renewable energies. The capacity factors for various types of renewable energies are shown in Table 2.1.

Renewables can be cheap and relatively easy to produce and are, by definition, much more sustainable than fossil fuels [2]. Furthermore, renewable energy generation releases much less
Table 2.1: Average Capacity Factors for Renewable Resources [4]

<table>
<thead>
<tr>
<th>Renewable Energy Source</th>
<th>Capacity Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomass</td>
<td>80%</td>
</tr>
<tr>
<td>Wind</td>
<td>35% – 45%</td>
</tr>
<tr>
<td>Geothermal</td>
<td>95.5%</td>
</tr>
<tr>
<td>Hydro</td>
<td>44.2%</td>
</tr>
<tr>
<td>Residential Solar</td>
<td>22.5%</td>
</tr>
</tbody>
</table>

CO$_2$ compared to fossil fuels [27]. Studies have concluded that much of the general public is willing to use renewables even if its cost is slightly more than power from conventional sources [28].

Although renewables have clear benefits over fossil fuels, there are complications when using renewables. One of the main drawbacks of renewable energy sources, such as solar power and wind power, is that they are intermittent; hence, forecasting the amount of generation from these sources is a complex but necessary process. Most renewables are also distributed throughout the power system and, thus, do not follow the structure of the existing power grid (see Figure 2.1). Since the infrastructure of the current power system was designed to serve relatively steady power load, it is not readily prepared to adequately accommodate the large amounts of renewables sought by many policy makers and environmentalists. Too much renewable generation during a period of low load can lead to voltage fluctuations, which can damage the power grid and negatively affect service to customers. Hence, integrating renewable energies into the existing power system will require innovative technologies [29].

2.3 The Smart Grid

The smart grid is defined as the collection of technologies, hardware, software, or practices that collectively make the delivery infrastructure of the grid more reliable, more versatile, more secure, more accommodating, more resilient, and ultimately, more useful to consumers [30]. In this thesis, the smart grid application of interest will be how an optimized water heater can be used to incorporate renewables into the power system, promote energy efficiency, adapt to dynamic pricing, and even act as a smart meter and a load forecasting device. Two specific smart grid applications of optimized water heaters are solar water heating and automated demand response.

\(^1\)An intermittent energy source is any source of energy that is not continuously available due to some factor outside direct control.
2.3.1 Smart Metering

Sensors play a significant role in the smart grid. In particular, sensors are an integral part of smart meters. A smart meter is a metering device that participates in two-way communication with a centralized system; in the context of the smart grid, the centralized system is a power plant. Smart meters can be used to gather and report data about the device to the power plant; thus, smart meters make assessing the state of the power system more accessible. Furthermore, plant operators can then send information, such as the future price for power, to the meter, where that information can then be used by the customer.

An optimized water heater can not only save a customer money, but it can also be used within the smart grid as a smart meter. To do this, we take advantage of the temperature sensor(s) built-in to the water heater to determine the state of the water heater. By adding a sensor to the water heater to actively monitor the hot water usage, we can infer not only the state of the water heater, but also the state of the customer (e.g., whether the customer will be likely to use hot water and draw power in the near future). Therefore, an optimized water heater can be used as a forecasting device that predicts loads for the water heater. In Section 2.3.3, we consider automated demand response, which also takes advantage of the smart meter properties of the water heater.

2.3.2 Solar Water Heating

One way of incorporating solar power into the smart grid is solar water heating. Although using photovoltaic panels to directly produce energy for water heating is a form of solar water heating, it is typically much less efficient than the alternative solar thermal water heating. It is for this reason that we focus exclusively on solar thermal water heating. Solar thermal water heaters work by pumping water up to a collection area where it is heated by the sun; heated water is then transferred down into a storage container where it can be used for consumption. Solar water heaters come in many types; a main distinction among them is whether the solar water heater is active or passive. An active solar water heater pumps a fluid between the collection and storage container (see Figure 2.3(a)). On the other hand, in a passive solar water heater, once heated water enters the storage container, it does not leave until it is needed for consumption (see Figure 2.3(b)). Active solar water heaters are typically more expensive than passive solar

\[^2\text{See Appendix A for a discussion on the physical implementation of these sensors.}\]
water heaters, but passive solar water heaters usually require less maintenance. During periods of low solar insolation, both types of systems use a conventional water heater as a backup system. Solar thermal water heating can be an efficient way of reducing energy consumption. The efficiency of the system depends on many factors, such as the ambient temperature, amount of solar insolation, the type of the system, including the size of the collection area and the size of the storage tank, etc. The maximum efficiency of solar thermal water heaters is roughly 75% [31]. Since power is consumed directly where it is produced, solar thermal water heaters promote distributed generation.

One of the main disadvantages of solar water heating systems are the relatively high upfront costs. Furthermore, overheating and freezing can cause damage to the solar water heating systems. Thus, solar water heating systems require routine maintenance and excellent overheating and freeze protection. Moreover, since the solar water heater requires a conventional water heater as a backup, a natural response is to question when the back-up water heater should be operated. We note now that it may be possible to improve operations of solar water heating systems using the dynamic programming framework that we will consider in this thesis. Specifically, with slight modifications to the main DP algorithm considered in this thesis, we can design an algorithm with the purpose of controlling the back-up water heater while minimizing the operating cost of the water heater, the discomfort of the customer, and over/under heating of the solar water heating components.

\(\text{Figure 2.3: Schematics of passive and active solar water heating systems [3].}\)
2.3.3 Automated Demand Response

While integrating distributed renewable energy resources will play a big role in changes to the electrical grid, demand response (DR) has largely been overlooked. Demand response is defined as the process of customers shifting and/or decreasing their electricity usage patterns to adapt to time-varying electricity prices. There are many forms of demand response as defined by [32]. The most common form of demand response is manual demand response, where customers are directly informed about electricity prices and change their usage patterns accordingly; e.g., a customer can receive notification that the price of power will be high in the next hour and the customer can then take actions to conserve energy and save money. Because manual demand response relies on customer initiative, the tradeoff between maximizing customer savings and minimizing customer discomfort is hard to quantify. Semi-automatic demand response aims at reducing the amount of customer input needed to change usage patterns; e.g., a building owner can program an air conditioning thermostat to use less energy during peak electricity usage periods. Fully automatic demand response (or simply automatic demand response) aims at minimizing the need for customer input. An example of automatic demand response involving water heaters occurs when customers allow the utility the capability of turning off their water heater during peak load times in return for a monetary compensation. Although this example requires minimal customer input, the customer’s comfort preferences are not taken into account; as a result, the customer can experience significant discomfort.

The impact that manual demand response has on a power system has been well studied. Qualitatively, demand response can help maintain grid stability, avert blackouts and brownouts, and reduce the impact of extreme weather events, while customers participating in DR can enjoy lower prices for power. A study by [33] shows that by decreasing/increasing prices when demand is low/high, price-aware customers decrease energy consumption by at least 10%. Similarly, simulations using real-world data performed by [34] suggest that demand response can decrease hourly peak expenditures of utilities by 12.1%. On the other hand, the effects of semi-automatic and automatic demand response have not been as carefully studied. Nevertheless, the effectiveness of automating demand response techniques are surveyed in [18]; they find that automation can increase average peak-demand reductions by 10% to 20%.

Despite its benefits, demand response has been slow to adopt. This may be due to a disconnect between understanding DR’s capacity to deliver reductions and its value as a grid resource. More specifically, finding the price that maximizes the benefit of both parties is complicated by
not knowing how the customer will react to changes in the prices. Furthermore, it is difficult to quantify the inconvenience that the customers experience from having to shift their load. To address these issues, new technologies and methods have been developed to exploit the potential of demand response. We propose a method that uses optimized water heaters to automate the demand response process. This framework allows the customer to quantify the level of discomfort (inconvenience) that they are willing to tolerate. Moreover, we can effectively determine how water heaters (customers) will react to changes in prices (they react optimally in the sense that they minimize operating cost subject to maintaining specified level of comfort); that is, we are able to calculate the expected load and operating cost of the water heater, and the expected discomfort of the customer.
Chapter 3

Problem Formulation

In this chapter, we discuss the key components of the water heater optimization problem. In particular, we discuss the details of the state and decision variables, the random process modeling hot water usage, the state equation, and the objective function. These quantities collectively define a periodic average cost Markov Decision Problem (MDP).

In what follows, we use the following rules of thumb when developing the problem:

1. Continuous-time is better than discrete-time.
2. Piecewise linear is better than piecewise constant.

3.1 Model Variables

In this section, we develop the so-called dynamic programming triple \((x, u, w)\), where \(x\) is the state of the system, \(u\) is the decision variable, and \(w\) is the power consumption for hot water usage.

3.1.1 State Variable

The state of the system \(x\), contains all the information needed to make an optimal decision. The key components of the state are time \(t\), the water heater’s average temperature \(T\), and the usage history \(h\).
In this thesis, we consider a periodic system and use the convention that the start of a cycle occurs at time \( t = 0 \). We uniformly discretize a cycle into \( N \) time stages of length \( \Delta t \). We view the set \( \Omega_t := \{0, \Delta t, \ldots, (N - 1)\Delta t\} \) as a set of sample times in which we are allowed to view the state and make a decision. To capture the periodic nature of sample times, we use \( t_k := \text{mod}(k, N) \Delta t \), where \( k \) is a non-negative integer, referred to as the simulation time stage. We otherwise use \( k \) as a subscript to denote an instance of a variable at time \( t_k \); e.g., \( T_k \) is the temperature at time \( t_k \).

We assume \( T_k \) must lie within some operating range \([T_{amb}, T_{max}]\), where \( T_{amb} \) is the ambient temperature of the water heater’s environment and \( T_{max} \) is the maximum allowable temperature that the water heater can safely operate at. We treat both \( T_{amb} \) and \( T_{max} \) as given constants.

The usage history \( h \) is defined as the set of all previous hot water usage during the current cycle; that is, \( h_k := \{(t_i, w_i) \mid 0 \leq i < \text{mod}(k, N)\} \), with the convention that \( h_k \) is the empty set when \( \text{mod}(k, N) = 0 \). An alternative interpretation of the usage history is that it is the length-\( k \) usage trajectory of a cycle; that is, it is a sequence of time and hot water usage pairs. In sequence notation (with \( k < N \)), we write \( h_k := (w_i)_{i=0}^{k-1} \).

### 3.1.2 Decision Variable

The decision variable is

\[
    u_k := \begin{cases} 
        1, & \text{if the water heater is on} \\
        0, & \text{if the water heater is off.} 
    \end{cases}
\]

We assume that the decision \( u_k \) is constant during the interval \([t_k, t_{k+1})\). Recall the constraint \( T_k \in [T_{amb}, T_{max}] \). We can explicitly impose this constraint by only optimizing over decisions that are feasible. In this case, let the set of feasible decisions at the current state be \( \Omega_u(x_k) \).

In order to say more about \( \Omega_u(x_k) \), we first need to develop the relationship between \( u_k \) and \( T_{k+1} \) (see Section 3.2.3 for a discussion on the water heater thermodynamics). An alternative to this approach is to assign infinite costs to decisions that are infeasible (see Section 3.3). In this case, all decisions are feasible; that is, \( \Omega_u = \{0, 1\} \).

We define a policy to be a function \( \mu \), that maps a state \( x_k \) into a feasible decision \( u_k \in \Omega_u(x_k) \).

---

1. We use the symbol := to mean “defined as.”
3.1.3 Hot Water Usage (Disturbance Variable)

Here, we develop a continuous-time cyclostationary stochastic process\(^2\) that models the instantaneous power of hot water demand \(W(t)\).\(^3\) In this thesis, we use the terms “hot water demand” and “hot water usage” interchangeably; we shorten this to just “demand” or “usage” when no ambiguity can occur. In the context of control theory, \(W(t)\) is sometimes called the disturbance variable or uncertainty in the system.

We assume that \(W(t)\) is, independent of the decisions that we make. Although this assumption does not hold in reality (e.g., the policy of always turning off the water heater will probably lead to less hot water usage), it is likely to hold true for all “good” policies, which are policies that provide an adequate amount of hot water at an acceptable monetary cost to the customer. Since we are interested only in finding “good” policies, the assumption is reasonable.

Let the set of possible usage events be \(\Omega_\tau := \{\text{shower, bath, \ldots, dishwasher}\}\). Let \(N_{\text{people}}\) be the number of people in a household and let \(i\) denote the \(i\)th person in a household. For a given day, and a given type \(\tau \in \Omega_\tau\), we define the number of events of each type with the random variable \(N_{\tau,i}\). We characterize the amount of hot water used by the \(j\)th event of type \(\tau\) for person \(i\), denoted \(E_{\tau,i}^{(j)}\), with the following random variables

\[
S_{\tau,i}^{(j)} := \text{the start time of } E_{\tau,i}^{(j)} \\
D_{\tau,i}^{(j)} := \text{the duration of } E_{\tau,i}^{(j)} \\
F_{\tau,i}^{(j)} := \text{the flow rate of } E_{\tau,i}^{(j)} \\
T_{\tau,i}^{(j)} := \text{the desired temperature of } E_{\tau,i}^{(j)},
\]

We require all distributions defining usage events to have finite support and that, for all \(i\) and \(\tau\), the densities of \(S_{\tau,i}^{(j)}\) and \(D_{\tau,i}^{(j)}\) are periodic with a period of \(N\Delta t\). Furthermore, to ensure \(W(t)\) are independent during different cycles, we require that all events finish before the end of a cycle (that is, \(S_{\tau,i}^{(j)} + D_{\tau,i}^{(j)} \leq N\Delta t\) for all \(i, j, \text{ and } \tau\)).

---

\(^2\)Loosely speaking, a random processes whose statistical characteristics vary periodically with time and are called cyclostationary processes. For a more rigorous definition, we say that a random process \(W\) is strictly cyclostationary if the distribution function of \(W\) is such that

\[
F_{W(t+\tau_1+N\Delta t),\ldots,W(t+\tau_{k-1}+N\Delta t),W(t+N\Delta t)}(\xi_1,\ldots,\xi_{k-1},\xi_k) = F_{W(t+\tau_1),\ldots,W(t+\tau_{k-1}),W(t)}(\xi_1,\ldots,\xi_{k-1},\xi_k),
\]

for all \(t \in \mathbb{R}\), all \((\tau_1,\ldots,\tau_{k-1}) \in \mathbb{R}^{k-1}\), and all \((\xi_1,\ldots,\xi_{k-1},\xi_k) \in \mathbb{R}^k\).\(^3\) There is also definition for wide-sense cyclostationary processes, which includes more general processes; however, in this thesis we use only the definition for strictly cyclostationary processes.

\(^3\)We use calligraphic letters (e.g., \(W\)) to denote random quantities.
Using the definitions of the random variables defined above, the amount of power drawn by the $j^{th}$ use of type $\tau$ from person $i$ is

$$ q = f_{\tau,i}^{(j)} \cdot \text{specific heat} \cdot (T_{\tau,i}^{(j)} - T_{\text{amb}}), $$

where specific heat is the specific heat of water (4.186 J/(g°C)), and $T_{\tau,i}^{(j)} - T_{\text{amb}}$ is the temperature increase (from ambient temperature) needed to supply the use (cf. Eq. (B.1)). Summing over all possible uses we obtain

$$ W(t) := \text{specific heat} \sum_{\tau \in \Omega} \sum_{i=1}^{N_{\text{people}}} \sum_{j=1}^{N_{\tau,i}} f_{\tau,i}^{(j)} \cdot (T_{\tau,i}^{(j)} - T_{\text{amb}}) \cdot I\{S_{\tau,i}^{(j)} \leq t < S_{\tau,i}^{(j)} + D_{\tau,i}^{(j)}\}, $$

To simulate $W(t)$, we use further require the usage events to be realistic⁴ (see Algorithm 1 Appendix D).

In practice, we can observe $W(t)$ only at the prespecified sample times $t \in \Omega_t$. Accordingly, we approximate $W(t)$ using the piecewise linear interpolation

$$ \tilde{W}(t) := W(t_k) + \frac{t - t_k}{\Delta t} [W(t_k + \Delta t) - W(t_k)], $$

(3.1)

for all $t \in [t_k, t_k + \Delta t)$. We take the discrete-time analog of $W(t)$ to be the average of $\tilde{W}(t)$ over $t \in [t_k, t_k + \Delta t)$,

$$ W_k := \frac{1}{\Delta t} \int_{t_k}^{t_k + \Delta t} \tilde{W}(t) \, dt = \frac{1}{2} [W(t_k) + W(t_k + \Delta t)]. $$

For further motivation for defining $W_k$ this way, see the subsection on water heater thermodynamics (Section 3.2.3). We denote particular realizations of $\tilde{W}(t)$ and $W_k$ using $\tilde{w}(t)$ and $w_k$, respectively. We write the conditional probability density/mass function of $W_k$ given $h_k$ as $p_{W_k}(w_k \mid h_k)$.

⁴ An example of an unrealistic event is if we are simulating usage for a household with one shower and two (or more) shower events occur at the same time.
3.2 State Dynamics

To design model-based algorithms, we state the equations governing the state dynamics. In particular, we define the state equation

\[ x_{k+1} = f(x_k, u_k, w_k) := (f_T(T_k, u_k, w_k), f_t(t_k), f_h(t_k, h_k, w_k)) \]  \hspace{1cm} (3.2)

where \( f_T, f_t, \) and \( f_h \) describe how the temperature, time, and usage history variables evolve over the time stage \( k \).

3.2.1 Discrete-Time Dynamics

Note that although the time stage \( k \) proceeds forward infinitely, because we assume our problem is periodic, \( t_k \) can only take values within \( \Omega_t \). Thus, discrete-time proceeds according to

\[ t_{k+1} = f_t(t_k) := \text{mod}(t_k + \Delta t, N\Delta t), \]

for all \( k \in \mathbb{Z}^+ \), starting with \( t_0 = 0 \).

3.2.2 Usage History Dynamics

Similar to discrete time, the usage history is periodic. The usage history dynamical equation is thus

\[ h_{k+1} = f_h(t_k, h_k, w_k) := \begin{cases} \emptyset, & t_k = (N - 1)\Delta t \\ \{(t_k, w_k)\} \cup h_k, & \text{otherwise.} \end{cases} \]

3.2.3 Temperature Dynamics

In Appendix B, we derive the following thermodynamic equation governing the water heater:

\[ T_{k+1} = f_T(T_k, u_k, w_k) := \max \left\{ T_k - r_{\text{cool}} \Delta t (T_k - T_{\text{amb}}) + r_{\text{heat}} \Delta t u_k - r_{\text{loss}} \Delta t w_k, T_{\text{amb}} \right\} \]  \hspace{1cm} (3.3)

where \( r_{\text{cool}} \) is the rate at which the water heater loses temperature to the environment, \( r_{\text{heat}} \) is the rate at which the water heater heats water, and \( r_{\text{loss}} \Delta t \) is a conversion factor from power to
a temperature. Derivations for \( r_{\text{cool}} \), \( r_{\text{heat}} \), \( r_{\text{loss}} \) can also be found in Appendix B; the respective values that we use in this thesis can be found in Table B.0(a). Eq. (3.3) is comparable to the temperature equations used in [7], [8], and [17].

### 3.3 Objective Function

The objective in (3.8) is to minimize the following function over all policies \( \mu \):

\[
J_\mu(x_0) = \lim_{K \to \infty} \mathbb{E} \left[ \frac{1}{K} \sum_{k=0}^{K-1} g(\mathcal{X}_k, \mu(\mathcal{X}_k), \mathcal{W}_k; \theta) \mid x_0 \right],
\]

where \( \mathcal{X}_0 = x_0 \) is given and \( \mathcal{X}_k = f(\mathcal{X}_{k-1}, \mu(\mathcal{X}_{k-1}), \mathcal{W}_{k-1}) \) for all \( k \in \mathbb{Z}^+ \). Notice that the randomness in the state at stage \( k = 1 \) is due solely to \( \mathcal{W}_0 \) since \( \mathcal{X}_1 = f(x_0, \mu(x_0), \mathcal{W}_0) \) and \( x_0 \) is given. Similarly, the randomness in the state at stage \( k = 2 \) is due to both \( \mathcal{W}_0 \) and \( \mathcal{W}_1 \) since \( \mathcal{X}_2 = f(\mathcal{X}_1, \mu(\mathcal{X}_1), \mathcal{W}_1) \). Thus, the expected value in the objective function can be written out explicitly as

\[
J_\mu(x_0) = \lim_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} \mathbb{E}_{\mathcal{W}_0, \mathcal{W}_1, \ldots, \mathcal{W}_k} \left[ g(\mathcal{X}_k, \mu(\mathcal{X}_k), \mathcal{W}_k; \theta) \mid x_0 \right].
\]

In the above equation, we define the stage cost \( g \) as the convex combination of two conflicting terms: the operating cost of the water heater \( g_{\text{operating}} \) and a discomfort cost \( g_{\text{discomfort}} \). We allow the customer to encode their tradeoff preferences by specifying a parameter set \( \theta := \{\alpha, T_{\text{min}}\} \), where \( \alpha \in [0, 1] \) is the relative weighting of the objectives, and \( T_{\text{min}} \) is the minimum desirable temperature during a hot water use. We write the stage cost as

\[
g(x_k, u_k, w_k; \theta) := \alpha g_{\text{discomfort}}(x_k, u_k, w_k; T_{\text{min}}) + (1 - \alpha) g_{\text{operating}}(x_k, u_k).
\]

For computational convenience, we can require \( g \) to assign infinite cost to decisions that are infeasible. We can do this, for example, by dividing the right-hand side of Eq. (3.5) by \( \mathbb{I}\{f_T(T_k, u_k, w_k) \leq T_{\text{max}}\} \). Throughout the remainder of this thesis, we use \( g \) with this modification and we take the set of feasible decisions to be \( \Omega_u = \{0, 1\} \).

The objective cost that we consider has an easy to understand parameterization; there are only two parameters: \( T_{\text{min}} \), which is relatively intuitive, and \( \alpha \), which can be set with the help of an objective tradeoff curve such as that in Figure 5.7. In contrast, the related work of [7] and [8] model discomfort as an explicit temperature constraint; e.g., the temperature is
required to be within some desirable range \([T_{low,k}, T_{high,k}]\) for all \(k \in \{0\} \cup \mathbb{Z}^+\). The disadvantage of this approach is that it requires the customer to specify the desirable range \([T_{low,k}, T_{high,k}]\), which is essentially an optimization problem in itself. Other work, such as [17], calculate discomfort as the temperature deviation below a desirable value; that is, the discomfort cost is \(\max\{T_{min} - T, 0\}\). The objective cost that we consider betters this approach by penalizing below desirable temperatures only if hot water is being used.

3.3.1 Operating Cost

The operating cost we use is

\[
g_{\text{operating}}(t_k, u_k) := \frac{1}{\Delta t} \int_{t_k}^{t_k + \Delta t} C(t) \text{ rating } u_k \, dt,
\]

where \(C(t)\) is the price of power at time \(t\) (assumed to be periodic) and \text{ rating} is the power rating of the water heater. In words, \(g_{\text{operating}}\) is simply the average monetary cost of operating during \([t_k, t_k + \Delta t]\). Although we consider deterministic price for power \(C(t)\), the algorithms to be discussed can be easily generalized to address the case when \(C(t)\) is stochastic.

3.3.2 Discomfort Cost

Let \(\tilde{T}(t)\) be the following linear interpolation

\[
\tilde{T}(t) := T_k + \frac{t - t_k}{\Delta t} [f_T(T_k, u_k, w_k) - T_k].
\]

Consider the following discomfort cost

\[
g_{\text{discomfort}}(x_k, u_k, w_k; T_{min}) := \frac{1}{\Delta t} \int_{t_k}^{t_k + \Delta t} \max\{T_{min} - \tilde{T}(t), 0\} \cdot I\{w_k > 0\} \, dt.
\]

This discomfort cost can be interpreted as the average temperature below \(T_{min}\) during \([t_k, t_k + \Delta t]\) when there is a hot water use; thus, temperature is allowed to decay below \(T_{min}\) without penalty as long as no hot water is being used.\(^5\)

\(^5\) We assumed that \(T_{min}\) is a given constant; however, when \(T_{min}(t)\) is a function of time, the discomfort cost is easily modified. In particular,

\[
g_{\text{discomfort}}(x_k, u_k, w_k; T_{min}(t)) := \frac{1}{\Delta t} \int_{t_k}^{t_k + \Delta t} \max\{T_{min}(t) - \tilde{T}(t), 0\} \cdot I\{w_k > 0\} \, dt.
\]
3.4 Problem Statement

With the key components of the problem now defined, we can formally state the problem that we seek to solve:

\[
\begin{align*}
\text{minimize} \quad & \lim_{K \to \infty} \mathbb{E}_{W} \left[ \frac{1}{K} \sum_{k=0}^{K-1} g(X_k, \mu(X_k), W_k; \theta) \bigg| x_0 \right] \\
\text{subject to} \quad & X_{k+1} = f(X_k, \mu(X_k), W_k), \quad \mu(x_k) \in \{0, 1\}, \\
& T_k \in [T_{amb}, T_{max}], \text{ for all } k = 0, 1, \ldots
\end{align*}
\] (3.8)

In words, the objective of (3.8) is to minimize over all on/off policies \(\mu\), the long-term average expected (w.r.t. \(W\)) cost \(g\), subject to the temperature (given by \(f\)) being within a desired operating range \([T_{amb}, T_{max}]\).
Chapter 4

Methodology

In this chapter, we discuss our methods for theoretically solving the water heater optimization problem given in (3.8). These methods are computationally prohibitive in practice, and so we also discuss ways of obtaining practical approximate solutions. To get an idea of how close to optimal the approximate solutions are, we develop a lower bound of the optimal expected objective cost.

4.1 Dynamic Programming

A common method for discrete-time MDP’s is dynamic programming (DP) [36]. The optimal solution is given in terms of an optimal cost-to-go function $J^*(x_k)$, which is the minimum expected objective cost of finishing the simulation starting from state $x_k$. The optimal policy, denoted $\mu^*(x_k)$, maps the current state $x_k$ into an optimal decision; that is, using the decision $\mu^*(x_k)$ at the current state $x_k$ will result in a cost-to-go of $J^*(x_k)$.

4.1.1 Finite-Horizon Dynamic Programming

Consider a finite-horizon MDP whose objective is to minimize over all policies $\mu$, the total cost starting from a given initial state $x_0$:

$$J_\mu(x_0) = \mathbb{E}_{\mathcal{W}} \left[ g_{\text{terminal}}(X_M) + \sum_{k=0}^{M-1} g(X_k, \mu_k(X_k), \mathcal{W}_k; \theta) \middle| x_0 \right].$$
Here, $M$ is known as the horizon and is defined as the number of times during the problem that a control is applied, and $g_{\text{terminal}}$ is a terminal cost function that assigns a cost to a state at the end of the simulation horizon. We can use the solution to the finite-horizon problem as suboptimal approximation of the solution to the original problem. However, the quality of the suboptimal solution depends crucially on the choice of $M$ and $g_{\text{terminal}}$. In fact, in later sections, we will show how the original problem can be reformulated as a finite-horizon problem with $M = N$ and $g_{\text{terminal}}$ chosen appropriately.

The finite-horizon MDP’s solution can be obtained by appealing to the principle of optimality, which states that the optimal cost-to-go of a state $x_k$ can be obtained by minimizing the current expected stage costs $g(x_k, u_k, w_k; \theta)$ plus the expected cost-to-go at the next stage $J^*(f(x_k, u_k, W_k))$. Stated mathematically, we have the following set of equations, known as Bellman’s equations:

$$
J^*(x_M) = g_{\text{terminal}}(x_M),
J^*(x_k) = \min_{u_k \in \{0, 1\}} \mathbb{E}_{W_k} \left[ g(x_k, u_k, w_k; \theta) + J^*(f(x_k, u_k, W_k)) \mid x_k \right],
$$

for all $k \in \{1, \ldots, M - 1\}$. The optimal policy is the minimizer of Bellman’s equation. We can solve Bellman’s equations using backwards recursion [36].

### 4.1.2 Average Cost Dynamic Programming

Relative value iteration (VI) can be used to solve periodic average cost MDP. The main idea is to use feedback to find relative terminal costs of a finite-horizon MDP. The following procedure loosely explains the relative VI algorithm [37]:

1. Initialize $J$ and $\mu$ arbitrarily and fix a reference state $x_{\text{ref}}$.
2. Calculate the new cost-to-go function $J'$ by solving an $N$-horizon MDP using $J(x_0)$ for all $x_0 \in \{(T_0, 0, \emptyset) \mid T_0 \in [T_{\text{amb}}, T_{\text{max}}]\}$ as the current terminal costs.
3. Update the current cost-to-go function using $J(x_k) \leftarrow J'(x_k) - J'(x_{\text{ref}})$ for all possible $x_k$.
4. Repeat steps 2 and 3 until convergence is achieved; e.g., stop if the 2-norm of the difference between subsequent cost-to-go vectors is less than some tolerance.
The relative VI algorithm terminates with $J$ being a differential cost function, interpreted as the minimum expected $N$-stage costs relative to the reference state $x_{ref}$. Furthermore, $J(x_{ref})$ is interpreted as the average cost of completing a cycle.

4.2 Approximate Dynamic Programming

Although, in theory, dynamic programming provides a way of obtaining the optimal cost and policy of the MDP in (3.8), its implementation in practice is unrealistic. In this section, we develop approximate DP algorithms that are suitable for practical implementation. In particular, we first discuss aggregation methods for decreasing the effective size of the state-space. We then consider a related problem with a lower-dimensional state-space. We develop a method for calculating the transition probabilities for the lower-dimensional states and we use the relative value iteration algorithm to solve for the corresponding policy. Finally, we discuss the model-free and the model-based versions of the relative $Q$-learning algorithm, both of which are able to handle a larger state-space.

4.2.1 Aggregation

State aggregation is a common tool that effectively decreases the size of the state-space, thereby making the original problem more tractable [38].

4.2.1.1 Aggregating Temperature

So far we have viewed $T_k$ as a continuous variable. To implement our algorithms on a computer, we convert our original problem into a discrete-valued temperature problem using a specialization of the coarse grid aggregation scheme found in [39, p. 483]. First, we uniformly discretize the temperature space $[T_{amb}, T_{max}]$ into $n$ values to obtain $\bar{\Omega}_T := \{T_{amb}, T_{amb} + \Delta T, \ldots, (n-1)\Delta T\}$.\(^1\) To distinguish between continuous and discrete temperatures, we use $T_k$ and $\bar{T}_k$, respectively. We develop an aggregate problem by viewing $\bar{T}_k \in \bar{\Omega}_T$ as aggregate states. Let $A(T_k)$ be the

\(^1\)Note that $T_{max} = (n - 1)\Delta T$. 
following random function of $T_k$:

$$A(T_k) := \begin{cases} 
\text{sgn}(T_k - \bar{T}) \Delta T, & \text{w.p. } |T_k - \bar{T}_k|/\Delta T \\
0, & \text{w.p. } 1 - |T_k - \bar{T}_k|/\Delta T
\end{cases} \tag{4.1}$$

where $\bar{T}_k = \text{round}(T_k/\Delta T) \Delta T$. The aggregate problem has the following modified thermodynamics:

$$\bar{T}_{k+1} = \bar{f}_T(T_k, u_k, w_k) := \text{round}(T_{k+1}/\Delta T) \Delta T + A(T_{k+1}), \tag{4.2}$$

where $T_{k+1} = f_T(T_k, u_k, w_k)$. Thus, for any given $(T_k, u_k, w_k)$, there are two possible transitions: a primary transition (which happens with a probability of at least 0.5) that rounds $f_T(T_k, u_k, w_k)$ to the nearest aggregate state, and a compensation transition that compensates for rounding.

An alternative interpretation of Eq (4.2) is that $A(T_{k+1})$ is a random correction term, and thus, $\bar{f}_T$ is the “corrected” state dynamics. It can be shown that the transitions of the aggregate equation are unbiased and have minimal variance (see Proposition 5.2).

Let $\bar{J}$ be the aggregate state equation obtained by using the aggregate thermodynamics in Eq. (3.2); that is,

$$\bar{x}_k = (\bar{T}, u, w)$$

Similarly, let $\bar{x} = (\bar{T}, u, w)$ be the state with an aggregate temperature component $\bar{T}$. Define $\hat{J}$ to be the optimal cost-to-go of the aggregate problem:

$$\begin{align*}
\min_{\mu} & \frac{1}{K} \sum_{k=0}^{K-1} E_{W_k} \left[ g(\bar{x}_k, \mu(\bar{x}_k), W_k; \theta) \right] \\
\text{subject to} & \bar{x}_{k+1} = \bar{f}(\bar{x}_k, \mu(\bar{x}_k), w_k), \quad \mu(\bar{x}_k) \in \{0, 1\} \\
& \bar{T}_k \in \{T_{\text{amb}}, T_{\text{amb}} + \Delta T, \ldots, T_{\text{max}}\}, \text{ for all } k = 0, 1, \ldots
\end{align*}$$

We then approximate the optimal cost-to-go of the original problem by

$$\hat{J}(T_k, h_k) := E[A(\bar{J}(\bar{T}_k + A(T_k), h_k))]$$

$$= \frac{|T_k - \bar{T}_k|}{\Delta T} \hat{J}(\bar{T}_k + \text{sgn}(T_k - \bar{T}_k) \Delta T, h_k) + \left(1 - \frac{|T_k - \bar{T}_k|}{\Delta T}\right) \bar{J}(\bar{T}_k, h_k),$$

where $\bar{T}_k = \text{round}(T_k/\Delta T) \Delta T$. Thus, $\hat{J}$ is a piecewise linear interpolation of $\bar{J}$.

A simpler way to discretize temperature is to use hard aggregation, in which case $A(T_k) \equiv 0$ with probability 1. Using hard aggregation amounts to a piecewise constant approximation of
aggregate cost-to-go $\bar{J}$ [36]. Although hard aggregation is simpler, we instead choose to use the coarse grid aggregation scheme for two reasons:

1. The optimal continuous cost-to-go function $J^*$, is a continuous strictly decreasing function of $T_k$ (see Proposition 5.1). Coarse grid aggregation preserves the monotonicity and continuity of the cost-to-go function, hard aggregation does not.

2. Hard aggregation does not model small losses of temperature ot the environment well. On the other hand, coarse grid aggregation accounts for these losses using $A(T_k)$.

4.2.1.2 Aggregating Usage History Using Features

Here, our goal is to exploit our prior intuition about the usage history to construct a low-dimensional feature vector $\phi_k$, such that $p_{W_k}(w_k \mid h_k) \approx p_{W_k}(w_k \mid \phi_k)$. Thus, $\phi_k$ summarizes the usage history $h_k$ and can be interpreted as a so-called approximate sufficient statistic for $W_k$. Moreover, we are interested in $\phi_k$ with simple update rules $\phi_{k+1} = f_\phi(\phi_k, w_k)$. We describe possible features with these characteristics in Table 4.1. The aggregate problem uses $\hat{x}_k = (T_k, t_k, \phi_k)$ in place of $x_k$. In Table 4.1, we note that $\phi_k^{(2a)}$ and $\phi_k^{(3a)}$ are both more general than $\phi_k^{(1a)}$ in the sense that we can obtain $\phi_k^{(1a)}$ from $\phi_k^{(2a)}$ or $\phi_k^{(3a)}$, but not the other way around. Thus, $\phi_k^{(2a)}$ and $\phi_k^{(3a)}$ capture more information about the usage history and we would expect the solution using one of these two features to be better than the solution using
Table 4.1: Features Summarizing Usage History

<table>
<thead>
<tr>
<th>Feature Definition</th>
<th>Update Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \phi_{k+1}^{(1a)} = \mathbb{1}{w_{k-1} &gt; 0} ) represents whether hot water was used during the previous stage and approximates the immediate history.</td>
<td>( \phi_{k+1}^{(1a)} = \mathbb{1}{w_k &gt; 0} )</td>
</tr>
<tr>
<td>( \phi_{k+1}^{(2a)} = \sum_{i=i_{\text{StartUse}}}^{k-1} \mathbb{1}{w_i &gt; 0} ) is the duration of the last use (( i_{\text{StartUse}} ) is the first stage of a use) and approximates the short-term history.</td>
<td>( \phi_{k+1}^{(2a)} = \mathbb{1}{w_k \neq 0} \cdot (\phi_k + \mathbb{1}{w_k &gt; 0}) )</td>
</tr>
<tr>
<td>( \phi_{k+1}^{(3a)} = \sum_{i=i_{\text{StartCycle}}}^{k-1} \mathbb{1}{w_i &gt; 0} ) is the total duration of all uses during the current period (( i_{\text{StartCycle}} ) is the first stage of the cycle) and approximates the long-term history.</td>
<td>( \phi_{k+1}^{(3a)} = \mathbb{1}{\text{mod}(k, N) = 0} \cdot (\phi_k + \mathbb{1}{w_k &gt; 0}) )</td>
</tr>
</tbody>
</table>

Note: All features discussed in this table, can be modified by changing every instance of \( \mathbb{1}\{w > 0\} \) to \( w \), in which case the features keep track of the intensities of the hot water usage rather than the durations. We denote these derivative features \( \phi_{k}^{(1b)}, \phi_{k}^{(2b)}, \phi_{k}^{(3b)} \).

\( \phi_{k}^{(1a)} \). On the other hand, the simplicity of \( \phi_{k}^{(1a)} \) is an advantage if a quick algorithm is desired; it takes less effort to store and calculate this feature.

Another noteworthy feature that is not discussed in Table 4.1 is the unit-feature given by \( f_\phi \equiv 1 \) and denoted simply as 1. This is the simplest possible feature (along with any constant feature) and does not tell us anything about the usage history, and so, we do not expect that using this feature will give exactly the same results as the no-feature case. However, the unit-feature is useful semantically. For example, we use the unit-feature as a placeholder in algorithms to refer to the no-feature case; thus, the algorithm does not need to be restated with trivial modifications just to include this case.

4.2.2 Relative Value Iteration Using Approximate Transition Probabilities

Another difficulty with DP is that, even for relatively simple \( \phi_k \), it may be difficult to obtain \( p_{W_k}(w_k \mid \phi_k) \) in closed-form. Here, rather than obtain an explicit formula for \( p_{W_k}(w_k \mid \phi_k) \), we aim to estimate it using kernel density estimation. However, it is difficult to estimate high-dimensional densities [40]. We therefore consider a simpler but related model by using the approximate state \( \tilde{x}_k = (T_k, u_k) \); equivalently, we can assume that \( W_k \) is independent of \( h_k \). Under this setting, we have \( p_{W_k}(w_k \mid h_k) = p_{W_k}(w_k) \) and so, we only need to perform a
1-dimensional density estimate for each time stage in a cycle. Once we have an estimate of the density, we can derive the state transition probabilities. Ultimately, we solve Bellman’s equations using the relative VI given in Algorithm 2 in Appendix D.

Before discussing the details of Algorithm 2, we make some definitions. We denote matrices using bold font, we denote transposition using $^\top$, and we assume that all vectors are column vectors unless otherwise stated. We define the length-$n$ vector $T := [T_{amb}, T_{amb} + \Delta T, \ldots, T_{max}]^\top$. We denote the components of a matrix (with a starting index at 0) using superscript notation; e.g., the $i$th component of $T$ is $T^{(i)} = T_{amb} + i\Delta T$, for all $i \in \{0, \ldots, n-1\}$. Similarly, we define the length-$m$ vector $w := [0, \Delta w, \ldots, w_{max}]^\top$ whose $l$th given by $w^{(l)} = l\Delta w$, for all $l \in \{0, \ldots, m-1\}$. Here, $w_{max}$ is the maximum observed value of hot water usage and $\Delta w$ is a step-size parameter. Finally, we define the $n \times N \times 2 \times m$ matrix $g_{\theta}$ whose $(i, k, u, l)^{th}$ component is $g(T^{(i)}, t_k, u, w^{(l)}; \theta)$, for all $i \in \{0, \ldots, n-1\}$, $t_k \in \Omega_t$, $u \in \{0, 1\}$, and $l \in \{0, \ldots, m-1\}$.

4.2.2.1 Kernel Density Estimation

We use kernel density estimation to estimate the probability density of hot water usage $W_k$. Assuming $W_k$ is independent of its history, we have, by the law of total probability,

$$p_{W_k}(w_k) = p_{W_k}(w_k \mid W_k = 0) \cdot \Pr(W_k = 0) + p_{W_k}(w_k \mid W_k > 0) \cdot \Pr(W_k > 0)$$

$$= \mathbb{1}\{w_k = 0\} \cdot \Pr(W_k = 0) + p_{W_k}(w_k \mid W_k > 0) \cdot [1 - \Pr(W_k = 0)].$$

Because hot water is used relatively sparsely throughout the day, it is typically the case that $\Pr(W_k = 0)$ is much larger than $1 - \Pr(W_k = 0)$. Furthermore, since demand is either nonexistent or some positive value (sufficiently larger than zero), it is likely that $p_{W_k}(w_k)$ is discontinuous at $w_k = 0$. Therefore, to eliminate unnecessary smoothing by the kernel density estimation, we do the estimation in two steps: first, we estimate $\Pr(W_k = 0)$, then we estimate $p_{W_k}(w_k \mid W_k > 0)$. In our simulations, we use the standard Gaussian kernel with the bandwidth parameter chosen optimally (see [40]).

4.2.2.2 Aggregate Transition Probabilities

Let $P_A$ be the $n \times n \times 2 \times m$ matrix of aggregation probabilities; that is, the $(i, j, u, l)^{th}$ entry of $P_A$ is $P^{(i,j,u,l)} = \Pr\left[\bar{f}_T(T^{(i)}, u, w^{(l)}) = T^{(j)} \mid i, j, u, l, \right]$. 

27
4.2.2.3 Temperature Transition Probabilities

Define the $n \times n \times N \times 2$ transition matrix $P$ whose $(i,j,k,u)^{th}$ entry is given by

$$P(i,j,k,u) := \Pr[\tilde{f}_T(T^{(i)}, u, W_k) = T^{(j)} \mid i, j, k, u]$$

$$= \frac{1}{m} \sum_{l=0}^{m-1} p_{W_k}(w^{(l)}) \mathbb{E}_{A}[\mathbb{I}\{\tilde{f}_T(T^{(i)}, u, w^{(l)}) = T^{(j)}\} \mid i, j, u, l]$$

$$= \frac{1}{m} \sum_{l=0}^{m-1} p_{W_k}(w^{(l)}) \Pr[\tilde{f}_T(T^{(i)}, u, w^{(l)}) = T^{(j)} \mid i, j, u, l]$$

$$= \sum_{l=0}^{m-1} p_{W_k}(w^{(l)}) P(i,j,u,l)^{(i,j,k,u)}; \quad (4.3)$$

that is, $P(i,j,k,u)$ is the probability of transitioning to $T^{(j)}$ given that the current stage, temperature, and decision are $k$, $T^{(i)}$, and $u$, respectively. From (4.3), we note that in order to calculate $P(i,j,k,u)$, we only need to consider $w^{(l)}$ that have non-zero probability at time $t_k$ and are such that $\tilde{f}_T(T^{(i)}, u, w^{(l)}) = T^{(j)}$. With Eq. (4.3) in mind, we calculate $P$ using Algorithm 4 in Appendix D.

Loosely speaking, the temperature of the water heater cannot change much within a small interval of time $\Delta t$. Hence, there are only a few transitions from a given temperature $T_k$, under a decision $u$, that have non-zero probability; that is, the vector $P^{(i::k,u)}$ is sparse. In particular, $P^{(i::k,u)}$ is a banded matrix (see Section 5.1). The special structure of $P^{(i::k,u)}$ allows for efficient storage and quicker computations.

For a given policy $\mu$, we obtain the policy-dependent temperature transition matrix $P_\mu$ using

$$P_\mu^{(i::k)} \leftarrow P^{(i::k,u)}$$

where $u = \mu(T^{(i)}, t_k)$, for all $i \in \{0, 1, \ldots, n - 1\}$ and $k \in \{0, 1, \ldots, N - 1\}$.

4.2.3 Relative $Q$-Learning

Here, we discuss the $Q$-Learning algorithm for average cost problems; specifically, this is the relative $Q$-Learning algorithm [39]. A particular advantage of the (relative) $Q$-Learning algorithm
is that it can be used when a model of the system and/or the state transition probabilities are not available.

### 4.2.3.1 Model-Free Relative $Q$-Learning

When a model of the system is not available, the relative $Q$-Learning algorithm is referred to as being model-free. For the model-free relative $Q$-Learning algorithm to converge to the optimal policy, all possible state decision pairs are visited infinitely often and the stepsize $\gamma$ must be such that

$$\sum_{j=1}^{\infty} \gamma_j^{(i_k,k,\phi_k)} = \infty \quad \text{and} \quad \sum_{j=1}^{\infty} (\gamma_j^{(i_k,k,\phi_k)})^2 < \infty,$$

where $\gamma_j^{(i_k,k,\phi_k)}$ is the stepsize during the $j^{th}$ visit to the state $\tilde{x}_k = (T_k, t_k, \phi_k)$ [41]. The stepsize used in Algorithm 6 in Appendix D is the reciprocal of the number visits to a state and thus, satisfies (4.4).

Notice that in the model-free relative $Q$-Learning algorithm, only the $Q$-factors corresponding to the trajectory $(T_k)_{k=0}^{N-1}$ and the decision made by the current policy are updated. Thus, the model-free relative $Q$-learning algorithm suffers from the problem of “exploitation versus exploitation”: that is, we need to tradeoff between exploring the space of state-decision pairs and using the best policy known so far. A simple method for addressing this problem is known as $\epsilon$-optimal decision making. In this method, we choose a random decision with some exploration probability $\epsilon$, and with probability $1 - \epsilon$, we choose the best decision so far $u_k = \mu(\tilde{x}_k)$.

### 4.2.3.2 Model-Based Relative $Q$-Learning

When a model of the system is available, it is better to use the model-based version of the $Q$-Learning algorithm [36]. The model-based version of the relative $Q$-Learning algorithm does not have the problem of “exploitation versus exploitation”. This is because we can take advantage of the system model to simulate all possible state-decision pair trajectories corresponding to a given a hot water usage trajectory $(w_k)_{k=0}^{N-1}$. To see this, we define the following operator:

$$\mathcal{F}(u, w, J) := E_A \left[ J(\text{round}(\hat{f}(T,u,w)-T_{amb}/\Delta T)) \right] \approx P_{\mu(u|J)}^{\epsilon} J,$$

where $l = \text{round}(w/\Delta w)\Delta w$. Here, $\mathcal{F}(u, w, J)$ maps the current decision $u$, usage realization $w$, and cost-to-go vector $J$ into the expected (w.r.t. $A$) cost-to-go vector at the next stage. We
can update an entire vector of $Q$-factors using

$$Q^{(i,k,\phi)} \leftarrow \left[1 - \gamma^{(k,\phi)}\right] Q^{(i,k,\phi)} + \gamma^{(k,\phi)} \left[g(T, t, u, w; \theta) + \mathcal{F}(u, w, J)\right],$$

where $g(T, t, u, w; \theta)$ is the length-$n$ vector whose $i$th entry is $g(T^{(i)}, t, u, w; \theta)$. This step can be approximated using precalculated values of $g^{(i,k,u,l)}$ and $P^{(i,,u,l)}_A J$ instead of the exact values for $g(T, t, u, w; \theta)$ and $\mathcal{F}(u, w, J)$. The approximate update is given by

$$Q^{(i,k,\phi)} \leftarrow \left[1 - \gamma^{(k,\phi)}\right] Q^{(i,k,\phi)} + \gamma^{(k,\phi)} \left[g^{(i,k,u,l)} + P^{(i,,u,l)}_A J\right].$$

In practice, as long as the spacing $\Delta w$ is sufficiently small, there will be no appreciable difference between the exact and approximate updates. For this reason, we state the model-based relative $Q$-learning using the approximate updates (see Algorithm 7 in Appendix D).

### 4.3 Prescient Lower Bound

Recall that our original problem is not in the standard MDP framework; particularly, the random process defining the hot water usage is not conditionally independent of the previous disturbances. We resolve this difficulty by augmenting the state with the previously observed hot water uses (that is, the usage history); this results in a state-space dimension that increases as time progresses. We obtain a suboptimal solution by assuming conditional independence between uses and solving the resulting problem. We can better this elementary solution by replacing the usage history with approximate sufficient statistic and solving the resulting problem. To get an idea how well these approximate methods solve the original problem, we consider a lower bound on the cost-to-go function known as the prescient lower bound (PLB) [42]. The bound is obtained using the following procedure:

1. Generate a series of usage trajectories.
2. Solve the finite-horizon problem corresponding to these trajectories exactly.
3. The average of the optimal costs is a lower bound for the objective function Eq. (3.4).

The intuition in the above procedure is that reasoning with certainty will outperform reasoning under uncertainty. Specifically, the bound represents the minimum possible costs given that the
hot water usage is known in advance. Thus, the objective function cannot be smaller than the optimal cost when the hot water usage is known with certainty.
Chapter 5

Results and Discussion

Here, we prove/motivate a priori results used in Chapter 4. We provide numerical simulations designed to verify the theoretical results.

5.1 Theoretical Results

In this section, we derive the result for unbiased aggregate thermodynamics with minimal variance. We also conjecture that the transition probability matrices are aperiodic; hence, the relative VI and Q-Learning\(^1\) algorithms will converge to the optimal solution. Finally, we prove that the optimal on/off policy is at least as good as any set-point policy, and even as good as a policy of randomized decisions.

5.1.1 Aggregation

The following proposition motivates the use of coarse grid aggregation rather than hard aggregation (see Section 4.2.1.1).

**Proposition 5.1.** The optimal cost-to-go function \(J^*(x_k)\) is continuous and non-increasing in \(T_k\).

\(^1\)As always, the Q-Learning algorithm must satisfy (4.4) to converge to the optimal solution.
Proof. The single stage costs are continuous non-increasing functions of $T_k$ (see Eq. (3.5), Eq. (3.6), and Eq. (3.7)). Recall Bellman’s equation:

$$J^*(x_k) = \min_{u_k \in \{0, 1\}} E \left[ g(x_k, u_k, W_k; \theta) + J^*(f(x_k, u_k, W_k)) \mid x_k \right].$$

The sum of continuous non-increasing functions is continuous and non-increasing. Moreover, the minimization in Bellman’s equation preserves continuity and monotonicity. We are free to initialize our algorithms with an arbitrary (finite) value for $J$. Thus, if we start with a continuous and monotonic function of $J$, our algorithms will converge to a continuous and monotonic function $J^*$. Since the solution to Bellman’s equation is unique, the result holds when $J$ is initialized arbitrarily.

In Section 4.2.1.1, we stated that the random function given in Eq. (4.1) can be used to construct unbiased aggregate state transitions with minimal variance. We prove this result now.

**Proposition 5.2.** The random function $A(T_k)$ in Eq. (4.2) is chosen such that the aggregate state transitions are unbiased and have minimal variance.

Proof. Recall the aggregate temperature equation given in Eq. (4.2):

$$\bar{T}_{k+1} = \bar{f}_T(T_k, u_k, w_k) = \text{round}(T_{k+1}/\Delta T)\Delta T + A(T_{k+1}).$$

Throughout the remainder of this proof, we omit the subscript $k$ for clarity. Define $\bar{T} = \text{round}(T/\Delta T)\Delta T$. The equation for $A(T)$ given in Eq. (4.1) can be rewritten as

$$A(T) = \begin{cases} 
\Delta T, & \text{w.p. } a^+(T) = \max\{T - \bar{T}, 0\}/\Delta T \\
-\Delta T, & \text{w.p. } a^-(T) = \max\{\bar{T} - T, 0\}/\Delta T \\
0, & \text{w.p. } 1 - a^+(T) - a^-(T),
\end{cases}$$

where $a^+(T)$ and $a^-(T)$ are known as the aggregation probabilities. Since the first term in $\bar{f}_T$ does not contribute to the variance, it suffices to show that $a^+(T)$ and $a^-(T)$ are chosen so that $\bar{f}_T$ is unbiased and $A(T)$ has minimal variance.

For the expected value of $\bar{T} + A(T)$ to equal $T$, we must have

$$T = E[\bar{T} + A(T)] = \bar{T} + a^+(T)\Delta T - a^-(T)\Delta T;$$
or equivalently,
\[ a^+(T) = a^-(T) - \frac{(\bar{T} - T)}{\Delta T}. \] (5.1)

The fact that \( a^+(T) \) and \( a^-(T) \) are valid probabilities implies
\[ \max \left\{ \frac{(\bar{T} - T)}{\Delta T}, 0 \right\} \leq a^-(T) \leq \min \left\{ 1 + \frac{(\bar{T} - T)}{\Delta T}, 1 \right\}. \] (5.2)

Now, we note that
\[
\text{var} \left[ \bar{T} + A(T) \right] = a^+(T) (\Delta T)^2 + a^-(T) (-\Delta T)^2 - T^2 \\
= (T - \bar{T}) \Delta T + 2a^-(T) (\Delta T)^2 - T^2.
\]
Thus, the variance is an increasing affine function of \( a^-(T) \). If \( \bar{T} - T > 0 \), then Eq. (5.1) and Eq. (5.2) imply that the minimal variance occurs when \( a^-(T) = \frac{(\bar{T} - T)}{\Delta T} \) and \( a^+(T) = 0 \), respectively. Similarly, if \( \bar{T} - T < 0 \), then \( a^-(T) = 0 \) and \( a^+(T) = \frac{(T - \bar{T})}{\Delta T} \). Combining the above gives the desired result. \( \square \)

5.1.2 Relative VI and Relative Q-Learning

The applicability of the relative VI and relative Q-Learning algorithm depend on the aperiodicity of the \( N \)-step transition probability matrix for all polices.

**Proposition 5.3.** The transition matrices \( P^{(:,;k,0)} \) and \( P^{(:,;k,1)} \) are banded\(^2\).

**Proof.** Recall, \( u = 0 \) is always a feasible decision. Since the only way to increase a temperature is if \( u = 1 \), \( P^{(:,;k,0)} \) is a lower triangular matrix and its upper bandwidth \( l_0 \) is zero. An upper bound of the lower bandwidth is
\[
l_0 \leq \max_{T \in \Omega_T, \ w \in \Omega_w} \left[ T - f(T, 0, w) \right] / \Delta T \\
= \left[ (T_{\max} - T_{amb}) \ r_{cool} \ + \ r_{loss} \ w_{\max} / \Delta T \right].
\]
Recall we defined \( T' \) as the largest value of \( T \) where \( u = 1 \) is a feasible decision; in particular,
\[
T' = \frac{T_{\max} - T_{amb} - \frac{r_{heat}}{\Delta t} T_{amb}}{1 - \frac{r_{cool}}{\Delta t}} + T_{amb}.
\]

\(^2\)A banded matrix is a sparse matrix whose non-zero entries are confined to a diagonal band, comprising the main diagonal and zero or more diagonals on either side.
An upper bound of the upper bandwidth of $P^{(\vdash, k, 1)}$ is

$$l_1 \leq \left\lceil \max_{T \in \Omega_T, w \in \Omega_w} \text{I}(T \leq T') \frac{f(T, 1, w) - T}{\Delta T} \right\rceil = \lfloor r_{heat} \Delta t / \Delta T \rfloor.$$ 

An upper bound of the lower bandwidth is

$$l_1 \leq \left\lceil \max_{T \in \Omega_T, w \in \Omega_w} \text{I}(T \leq T') \frac{T - f(T, 1, w)}{\Delta T} \right\rceil \leq \left\lceil \left[ (T' - T_{amb}) r_{cool} + r_{loss} w_{max} - r_{heat} \right] \Delta t / \Delta T \right\rceil.$$ 

Conjecture 5.4. For any policy $\mu$, and any $k = 0, 1, \ldots$, the $N$-step transition probability matrices

$$P^{(k, k + N)}_\mu := \prod_{k' = 0}^N P^{(\vdash, k')}_\mu = P^{(\vdash, k + N)}_\mu \cdots P^{(\vdash, k + 1)}_\mu P^{(\vdash, k)}_\mu,$$ 

are aperiodic.

To motivate Conjecture 5.4, we note that $P^{(\vdash, k, 0)}$ is a lower triangular banded matrix with lower bandwidth $l_0$, and $P^{(\vdash, k, 1)}$ is a banded matrix with lower and upper bandwidth $l_1$ and $l_1$, respectively (Proposition 5.3). Given any initial matrix $P_0$, multiplication by $P^{(\vdash, k, 0)}$ moves the entries of $P_0$ to the left by $l_0$ places. Similarly, multiplication by $P^{(\vdash, k, 0)}$ moves the entries of $P_0$ to the left by $l_1$ places and to the right $l_1$. Therefore, as long as the there is one $u_k = 1$, all entries of

$$P_\infty := \lim_{M \to \infty} \left[ P^{(k, k + N)}_\mu \right]^M P_0$$

will be non-zero, which means $P_\infty$ is aperiodic. If all $u_k = 0$, then the first column of $P_\infty$ is contains only ones and the rest of the entries are zero. This means $P_\infty$ has a period of 1 which by definition means it is aperiodic.

If any $N$-step probability matrix associated with any admissible policy is aperiodic, then the conditions in [37] are met. Therefore, we can be assured that using relative VI will asymptotically lead to the optimal policy for the water heater optimization problem (3.8).
5.1.3 Generalized Set-Point Method

A common strategy used to optimize residential water heaters is to optimize the set-points of a water heater [7], [8], [9], [10] and [11]. We write these set-points mathematically as the function \( T_{\text{set}}(t_k) \). A generalization of this method allows for a temperature deadband centered at the set-point. The half-width of the deadband denoted \( \delta(t_k) \) is then viewed as an additional parameter to optimize.

Let the set of all possible set-point and half-width parameters combinations be \( \Theta \). For a given \( \vartheta \in \Theta \), the policy of a generalized set-point method can be written as

\[
\mu_{\text{set-point}}(T_k, t_k, u_{k-1}; \vartheta) := \begin{cases} 
0, & \text{if } T_k > T_{\text{set}}(t_k) + \delta(t_k) \\
1, & \text{if } T_k < T_{\text{set}}(t_k) - \delta(t_k) \\
u_{k-1}, & \text{otherwise,}
\end{cases}
\]

(5.3)

where \( \vartheta \) is the set of set-point and half-width parameters.

**Proposition 5.5.** Let \( \Pi_\Theta \) be the set of all policies of the form given in Eq. (5.3). Define \( \mu^*_{\text{set-point}} \) as the policy in \( \Pi_\Theta \) corresponding to the smallest expected objective cost. Consider the Q-Learning algorithm with state \( \tilde{x}_k = (T_k, t_k, u_{k-1}) \) (i.e., with feature \( \phi_k = u_{k-1} \)), and let \( \mu^* \) be the policy obtained asymptotically from the relative Q-Learning algorithm. Then, \( \mu^* \) is at least as good as \( \mu^*_{\text{set-point}} \) in terms of minimizing the expected objective cost.

**Proof.** Let \( \Pi \) be the set of all policies mapping \( (T_k, t_k, u_{k-1}) \) into \( u_k \). Clearly, \( \Pi_\Theta \) is a subset of \( \Pi \). Since \( \mu^* \) is an optimal policy in \( \Pi \), it follows that \( \mu^* \) is at least as good as \( \mu^*_{\text{set-point}} \). \( \Box \)

A simple case occurs when \( \delta(t_k) \equiv 0 \). We refer to the corresponding policy as the simple set-point policy and write its policy succinctly as

\[
\mu_{\text{simple}}(T_k, t_k; T_{\text{set}}) := \mathbb{1}\{T_k < T_{\text{set}}(t_k)\},
\]

for all \( k \in \{0\} \cup \mathbb{Z}^+ \).

An analogous result to Proposition 5.5 holds for simple set-point policies: Let \( \mu^* \) be the optimal mapping \( (T_k, t_k) \) into \( u_k \), then \( \mu^* \) is at least as good as \( \mu^*_{\text{simple}} \).

---

3A notable exception is the work of the researchers at the University of Leuven, who also optimize a water heater’s on/off policy using model-free Q-Learning [15], [16], and [17].
5.1.4 Randomized Decision Making

Consider the randomized decision \( v_k \), defined as the probability of turning on at time \( t_k \). Let \( \nu \) be a policy of randomized decisions; that is, \( \nu : x_k \mapsto v_k \).

**Proposition 5.6.** An optimal randomized decision policy \( \nu^* \) is no better than the optimal on/off policy \( \mu^* \).

**Proof.** Consider Bellman’s equation for a random decision

\[
J^*(x_k) = \min_{0 \leq v_k \leq 1} \left( E_{W_k} \left[ g(x_k, v_k, W_k; \theta) + J^*(f(x_k, v_k, W_k)) \right] \right)
\]

\[
= \min_{0 \leq v_k \leq 1} \left( v_k E_{W_k} \left[ g(T_k, t_k, 1, W_k; \theta) + J^*(f(T_k, 1, W_k)) \right] \right.
\]

\[
\left. + (1 - v_k) E_{W_k} \left[ g(T_k, t_k, 0, W_k; \theta) + J^*(f(T_k, 0, W_k)) \right] \right)
\]

\[
= \min_{0 \leq v_k \leq 1} [v_k Q^*(x_k, 1) + (1 - v_k)Q^*(x_k, 0)].
\]

Thus, the minimization in Bellman’s equation is of an affine function of \( v_k \). If \( Q^*(x_k, 0) \neq Q^*(x_k, 1) \), then the minimum must occur at either \( v_k = 0 \) or \( v_k = 1 \). If \( Q^*(x_k, 0) = Q^*(x_k, 1) \), then any \( v_k \in [0, 1] \) is optimal; therefore, we have that the optimal on/off policy is at least as good as the optimal randomized decision policy.

5.2 Numerical Results

In this section, we compare the algorithms previously discussed using numerical simulations involving a “typical household”.

**Table 5.1: Numerical Simulation Parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta t )</td>
<td>time step</td>
<td>1/12 h</td>
</tr>
<tr>
<td>( N \Delta T )</td>
<td>length of a cycle</td>
<td>1 day</td>
</tr>
<tr>
<td>( \Delta T )</td>
<td>temperature step</td>
<td>1/3 ( ^\circ )C</td>
</tr>
<tr>
<td>( T_{amb} )</td>
<td>ambient temperature</td>
<td>25 ( ^\circ )C</td>
</tr>
<tr>
<td>( T_{max} )</td>
<td>maximum operating temperature</td>
<td>50 ( ^\circ )C</td>
</tr>
<tr>
<td>( T_{min} )</td>
<td>minimum desirable temperature</td>
<td>40 ( ^\circ )C</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>objective tradeoff</td>
<td>varied</td>
</tr>
</tbody>
</table>
The typical household is defined as a three person household with $\Omega_{\tau} = \{\text{shower}, \text{handwash}\}$. The price of power is taken from [43] and is given by

$$C(t) := \begin{cases} 
18.2 & \text{if } 9 \text{ p.m.} \leq t < 7 \text{ a.m.} \\
23.7 & \text{if } 7 \text{ a.m.} \leq t < 5 \text{ p.m.} \\
26.7 & \text{if } 5 \text{ p.m.} \leq t < 9 \text{ p.m.},
\end{cases}$$

(5.4)

see Figure 5.1(c).

A plot of the average household usage and the corresponding probability densities are shown in Figure 5.1(a) and 5.1(b).

Given a price $C(t)$ and a corresponding policy $\mu_C$, we determine a water heater’s objective cost $J$ using a method known as policy evaluation. We can calculate the expected operating costs by setting $\alpha = 0$ and then performing policy evaluation on $\mu_C$. Similarly, we can calculate the expected discomfort cost by setting $\alpha = 1$ and then performing policy evaluation on $\mu_C$.

One way to perform policy evaluation is to use the $\mu$-dependent transition matrix $P_\mu$ to do a single cycle of value iteration. We refer to this method as Policy Evaluation using Value Iteration (PEVI), see Algorithm 8. The problem with this method is that $P$ (and hence, $P_\mu$) obtained using kernel density estimation is an estimate. Furthermore, since we are evaluating the policy on the aggregate problem, $J$ will an estimate of the true objective cost.

An alternative to using PEVI is to use the actual data to evaluate a policy. We refer to this method as Monte Carlo Policy Evaluation (MCPE). In PEVI, we generate/observe usage trajectories; in our numerical simulations, we generate 10 usage trajectories of length 500N. Then, given a policy, we simulate the corresponding temperature trajectory. We allow temperature to be analog, and so, to obtain decisions, we round temperatures to the nearest value in $\Omega_T$. We use these usage, temperature, and decision trajectories to calculate the objective costs. We then average appropriately to get the results that follow. Since we are evaluating the policy on the original problem (with analog states), compared to PEVI, MCPE will produce a more reliable estimate of $J$. The disadvantage of using MCPE is that it takes considerably more time to evaluate the policy.
5.2.1 Finite Horizon Dynamic Programming

In Section 4.1.1, we claimed that the choice of the simulation horizon \( M \) and terminal cost function \( g_{\text{terminal}} \) crucially affect the quality of the finite-horizon dynamic programming solution. Recall that the terminal cost function \( g_{\text{terminal}}(x_M) \) should be chosen to reflect the cost of completing a cycle starting from a state \( x_M \); that is, \( g_{\text{terminal}}(x_M) = J^*(x_M) \). If we only consider \( M \) that are integer multiples of \( N \) (the number of time stages in a cycle), then \( x_M = (T_M, 0, \emptyset) \) and \( g_{\text{terminal}} \) becomes a function of a single variable, namely \( T_M \).
The effects of incorrectly choosing $g_{\text{terminal}}$ increase over time and are most noticeable near the end of the horizon. Thus, a good idea is to choose $M$ large so the effect of incorrectly choosing $g_{\text{terminal}}$ is spread out over time. The suboptimal policy is then the policy of the first cycle. The following example highlights what happens when $M = N$ (the worst case value) and $g_{\text{terminal}}$ is poorly chosen.

Example 5.1. A naive approach for choosing the terminal costs is to set $g_{\text{terminal}} \equiv 0$; we refer to this assignment as case 1. A more informed approach takes into account the shape of the cost-to-go function. From Proposition 5.1, we know that the cost-to-function $J^*$ is a continuous non-increasing function of $T$. Thus, a heuristic for choosing the terminal cost function is $g_{\text{terminal}}(T) = (T_{\text{max}} - T)/(T_{\text{max}} - T_{\text{amb}})$; we refer to this as case 2.

The results for this example are shown in Figure 5.2 and 5.3. As expected, the infinite horizon average cost is better at solving the problem than both the finite-horizon cases. It is also interesting to note that for both values of $\alpha$, the more sophisticated terminal cost yielded a better solution.

As evidenced by Figure 5.2 and 5.3, when $\alpha$ is small, the optimizer focuses on minimizing then operating cost. In case 1 of the finite-horizon DP solution, the temperature deteriorates near the end of a cycle. The reason for this is that the terminal cost makes it seem beneficial to end a cycle without any residual temperature since $g_{\text{terminal}}$ assigns a cost of zero no matter what the state. Implementing this suboptimal policy will lead to significant discomfort costs near the end of the cycle and also large operating costs corresponding to reheating at the beginning of a cycle. When $\alpha$ is large, minimizing discomfort cost is the primary objective. Here, both cases of the finite-horizon DP method do reasonably well.
Figure 5.2: Finite-horizon and infinite horizon average cost DP daily cumulative costs and temperature trajectories for $\alpha = 0.01$.
Figure 5.3: Finite-horizon and infinite horizon average cost DP daily cumulative costs and temperature trajectories for $\alpha = 0.99$
The following example illustrates how $M$ affects the quality of the solution when the terminal cost is $g_{\text{terminal}} \equiv 0$.

**Example 5.2.** Recall in the previous example, using the terminal cost $g_{\text{terminal}} \equiv 0$ gave a suboptimal result when $M = N$; here, we also consider a longer horizon $M = 2N$.

Figure 5.4 shows that the longer horizon is effective in improving the solution. In fact, a closer inspection of the solutions indicates that the finite-horizon solution is exactly the same as the infinite-horizon average cost solution. This is not surprising since the relative VI algorithm (which essentially solves the finite-horizon problem with longer and longer horizons) has exponential convergence [37] and usually converges within a single iteration.

![Figure 5.4](image.png)

(a) Cumulative objective cost for $\alpha = 0.01$.

(b) Cumulative objective cost for $\alpha = 0.99$.

**Figure 5.4:** Objective costs for the finite-horizon solution ($M = N, 2N$).

### 5.2.2 Prescient Lower Bound

As discussed in Section 4.3, the PLB is a lower bound on the objective function that is obtained by assuming knowledge of the usage trajectories and averaging the corresponding costs. A problem with using this bound is that it applies to cost-to-go of finite-horizon problems. To
obtain an approximate PLB, we solve the deterministic finite-horizon problem corresponding to the 10 usage trajectories of length $500N$.\footnote{We solve the deterministic finite-horizon problem using a finite grid of temperatures. Thus, the solution that we obtain is only an approximation of the optimal solution.} In an attempt to remove transient behavior, we discard the first two cycles from the beginning of the trajectories. Appropriate averages of the truncated trajectories yield the approximate PLB. \textbf{Example 5.2}, suggests that the transient state behavior dies out quickly; hence, we expect the average of the truncated cost trajectories to be very similar to the PLB for the average cost MDP.

We now compare the no-feature relative VI algorithm to the PLB. Notice that in \textbf{Figure 5.5(d)} and 5.6(d), the deterministic DP solution varies more from day to day compared to the stochastic DP solution. This makes sense since the deterministic DP solution can adapt to the individual uses, while the stochastic DP solution addresses the average outcome. It is interesting to note that although the temperature trajectories of the two methods are quite different, the cost trajectories are similar; e.g., for $\alpha = 0.005$, the deterministic DP solution can supply reliable hot water for $8 per week (see \textbf{Figure 5.6(b)}), while the stochastic DP solution can achieve the same level of comfort for only about $1 more per week or about $0.15 per day (see \textbf{Figure 5.5(b)}). For both values of $\alpha$, the deterministic DP solution offers only marginal improvement over the stochastic DP solution.
Figure 5.5: Stochastic and deterministic DP weekly cumulative costs and temperature trajectories for \(\alpha = 0.005\)
Figure 5.6: Stochastic and deterministic DP weekly cumulative costs and temperature trajectories for $\alpha = 0.995$
5.2.3 Relative VI Versus Set-Point Methods

Here, we compare the no-feature relative VI algorithm to a simple set-point method. Figure 5.7 displays the tradeoff between objectives for the water heater optimized with the relative VI algorithm with no features (referred to as the no-feature DP method) and a water heater optimized via the simple set-point method (see Eq. 5.1.3). In Figure 5.7, the dynamic programming curve is parameterized by $0 \leq \alpha \leq 1$. The simple set-point method curve is parameterized by a single set-point varied between $25^\circ C \leq T_{set} \leq 55^\circ C$.

Example 5.3. In this example, we consider the objective tradeoff when the price of power is time-varying and given by Eq. (5.4).

As expected (Proposition 5.5), the no-feature DP method achieves better costs than the single set-point method for all choices of parameters.

Notice that the DP solution can reliably supply hot water at about $1.40 per day. To achieve approximately the same level of comfort, the simple set-point would cost nearly $1.65. Thus, the DP solution can save the customer roughly $90 a year. The PLB reveals that, even if hot water usage is known exactly, an optimal strategy can supply reliable hot water at a cost of $1.30 per day. Thus, the relative VI method is not too far from optimal.

Another important observation is that the simple set-point method cannot reduce operating costs below about $1.50 without incurring large discomfort costs. To highlight this, suppose a customer on a budget wants to spend, on average, no more than $1.25 per day on water heating. The relative VI solution can operate with this constraint while incurring a discomfort cost of about $0.5^\circ C$ per use. On the other hand, the simple set-point method operating with this constraint will incur a discomfort cost of more than $10^\circ C$ per use. Notice that the relative VI solution is, again, not too far from the PLB, which says that at this level of operating cost, the minimum possible discomfort is at least $0.15^\circ C$ per use.
Example 5.4. In this example, we consider the objective tradeoff when the price of power is a constant $0.22/kWh. Here, the average daily price for power is the same as that in Eq. (5.4).

Again, as expected, the no-feature DP method achieves better costs than the single set-point method for all choices of parameters. However, for this price of power, the no-feature DP method betters the simple set-point method only slightly. To see this, consider the following instances.
The no-feature DP solution can reliably supply hot water at about $1.47 per day, while the simple set-point method operating at this level of comfort costs only about $0.05 more per day. The amount of savings that the DP solution can offer the customer is then, roughly $20 a year, which is not nearly as significant of a savings as when the price is time-varying. This makes sense since the only way to save money is to minimize losses to the environment. Because water heaters are well insulated, the loss rate is relatively small (see Table B.0(a)), and so the losses to the environment are also small. This is in contrast to when the price of power is time-varying; in this case, savings can occur by smart preheating during periods of low prices. There, is again, not too large of a difference between the PLB and the relative VI method.

Similar to the observation made in the previous example, the simple set-point method is very ineffective at reducing operating costs below about $1.40 without also incurring large discomfort costs. As a concrete example, suppose the customer wanted to pay no more than $1.25 per day on water heating. The no-feature DP algorithm can do this while incurring a discomfort cost of about 2 °C per use. For the set-point method to operate at this level, about 9.5 °C per use discomfort cost would need to be incurred. The PLB in this case says that the minimum possible discomfort cost at this level of operating cost is at least 1.4 °C per use. Thus, the no-feature DP algorithm is relatively close to optimal.
The preceding examples motivate a particularly useful application of a tradeoff curve, such as that shown in Figure 5.7. The tradeoff curve can be used as a visual tool for a customer to choose an acceptable tradeoff weight $\alpha$. This can be done by having the customer decide upon on a minimum desirable temperature $T_{\text{min}}$ based on their personal preference. Next, the customer can compare the costs on the tradeoff curve and pick the most preferable point. This point can then be mapped back to the corresponding value of $\alpha$.

**Figure 5.8:** Tradeoff curve for the no-feature DP solution and the simple set-point solution under a constant price for power.
5.2.4 Effect of the Number of Samples

All the ADP algorithms previously discussed depend on the number of sample cycles. In the relative VI algorithm, the more cycles that are observed, the better the estimate of the probability distribution. Similarly, in the relative QL algorithm, the sample averages converge to their respective expected values as the number of sample cycles are increased. In this example, the step size was taken to be the reciprocal of the number of visits to a state. Figure 5.9 and 5.10 show that average weekly cumulative costs and temperature trajectories of the typical model. The colors correspond to the algorithm used and the number of days used to obtain the solution (e.g., VI100 used 100 training days).
Figure 5.9: Average weekly cumulative costs and temperature trajectories for $\alpha = 0.005$
Figure 5.10: Average weekly cumulative costs and temperature trajectories for $\alpha = 0.995$.
Both Figure 5.9 and 5.10 suggest that the VI algorithm converges quicker than the QL algorithm with respect to the number of samples. Furthermore, there is not much improvement in the VI algorithm from 500 to 1000 samples.

5.2.5 Aggregation Schemes

Figure 5.11, 5.12, and 5.13 show that for sufficiently small temperature step sizes, there is no large difference between the aggregation schemes. This suggests that the solutions are converging to the continuous temperature solution. When $\Delta T = 1$, there is a noticeable difference between the coarse grid aggregation method and hard aggregation; coarse grid aggregation provides a better result. Thus, the coarse aggregation method can be used when $\Delta T$ is relatively large to obtain a good approximate solution quickly. A small but noteworthy observation is that the hard aggregation scheme tends to minimize discomfort at the expense of incurring higher operating costs; conversely, the coarse grid aggregation scheme tends to minimize the operating cost at the expense of incurring larger discomfort costs.

![Figure 5.11: Tradeoff curve for aggregation schemes ($\Delta T = 1, 1/3, 1/10$).](image)

Figure 5.11: Tradeoff curve for aggregation schemes ($\Delta T = 1, 1/3, 1/10$).
(a) Average cumulative objective cost.

(b) Average cumulative operating cost.

(c) Average cumulative discomfort cost.

(d) Average temperature trajectory

Figure 5.12: Average weekly cumulative costs and temperature trajectories for $\alpha = 0.005$
Figure 5.13: Average weekly cumulative costs and temperature trajectories for $\alpha = 0.995$
5.2.6 Feature Comparison

In this section, we experimentally compare some of the features discussed in Section 4.2.1.2; in particular, we compare the features measuring durations. Figure 5.14 shows the tradeoff curve for these features. As expected, all features improve upon the unit-feature case (equivalently, the no-feature case). Also expected is that using $\phi^{(2a)}$ or $\phi^{(3a)}$ is better than using $\phi^{(1a)}$. We were already able to derive these results in Section 4.2.1.2, but what is not apparent in that section is whether $\phi^{(2a)}$ or $\phi^{(3a)}$ is better. For this particular household, Figure 5.14 suggests that $\phi^{(3a)}$ is uniformly better than $\phi^{(2a)}$.

![Figure 5.14: Tradeoff curve for features measuring durations.]

5.2.7 Algorithm Complexity

The ADP algorithms can be sped up by precalculating and storing the cost matrix $g_\theta$. The relative VI algorithm can be further sped up by also precalculating and storing the transition probability matrix $P$. The relative VI algorithm converges within a few cycles (about 2-3), whereas the relative $Q$-Learning algorithm takes on the order of a few hundred cycles to obtain a good solution (cf. Section 5.2.4). Indeed, in the numerical simulations that were conducted, after precalculation, the relative VI algorithm completed in less than 0.5s whereas the relative $Q$-Learning algorithm, which trained on 5000 cycles, took a few minutes to complete. In Section 6.2, we take advantage of the speed of the relative VI algorithm to construct an automated demand response algorithm.
Chapter 6

Problem Extensions

In this chapter, we discuss applications of the basic water heater optimization problem. In particular, we consider the problem of supplying hot water when a solar water heater is used; we address the issue of using water heaters for automated demand response; we discuss possible modifications to our algorithms when hot water demand is not assumed to be cyclostationary.

6.1 Solar Water Heating

In order to incorporate solar water heating into our model, we append to the state a random variable $V_k$, representing the solar irradiance at time $t_k$. In practice, we will not have access to the realization $v_k$, but we can estimate it using forecasting methods. Let $\text{eff}(v_k)$ convert irradiance into usable energy. The water heater thermodynamics can now be recast as

$$f_T(T_k, u_k, w_k, v_k) = \min \{T_k - r_{\text{cool}} \Delta t (T_k - T_{\text{amb}}) + r_{\text{heat}} \Delta t u_k - r_{\text{loss}} \Delta t w_k + r_{\text{solar}} \Delta t \text{eff}(v_k), T_{\text{max}} \},$$

where $r_{\text{solar}}$ is a conversion factor from power to temperature. To modify (3.8), we also take the expected value with respect $V_k$.

$$\min_{\mu} \lim_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K-1} \mathbb{E}_{V_k} \left[ g(x_k, \mu(x_k), W_k; \theta) \mid x_k \right]$$

subject to $x_{k+1} = f(x_k, v_k, \mu(x_k), w_k)$, $\mu(x_k) \in \{0, 1\}$

$T_k \in \left[ T_{\text{amb}}, T_{\text{max}} \right]$, for all $k = 0, 1, \ldots$
6.2 Automated Demand Response

Approximate dynamic programming is well-suited to developing automatic demand response algorithms. For example, [12] and [13] develop a general reinforcement learning algorithm (a type of ADP algorithm) to obtain an asymptotically optimal schedule for household devices. Related to this, [14] and [44] design specialized reinforcement learning algorithms to optimize the heating and cooling of a building in the presence of time-varying prices for electricity. Our work is similar to that of [17], who develop model-free Q-learning algorithms to solve water heater optimization problems framed as finite-horizon MDP’s.

We extend the previous work by implementing the following modifications. Since the horizon $M$ and terminal cost function $g_{\text{terminal}}$ are difficult to objectively select in finite-horizon problems, we consider an average cost MDP instead. In contrast to the previous work, we explicitly account for customers’ comfort preferences. Finally, we develop and use a model of the water heater dynamics. It is well known that if a model of the system is known, then it can be used to develop more powerful algorithms [36].

6.2.1 Utility Optimization Problem

We formulate the goal of the utility as a continuous-time optimization problem. The state of the utility $L(t)$, is the expected load to be served during a cycle of length $t_{N-1}$. We assume that the generation exactly matches load. The utility’s marginal cost is typically modeled as an upwards sloping function of power generated [45]. Here, we assume the simplest such function, which is the linear function $\beta_1 + \beta_2 L(t)$ with constant coefficients $\beta_1$ and $\beta_2$. We define the utility’s objective function to be the average total cost (the total cost is the integral of the marginal cost with respect to generation) per cycle; that is, $\frac{1}{t_{N-1}} \int_0^{t_{N-1}} (\beta_1 L(t) + \beta_2 L(t)^2) \, dt$. The decision variable of the utility is the price of power $C(t)$ during a cycle. The utility dynamics are given by $L(t) = f_{\text{Utility}}(C(t))$. We do not assume to know the specific form of $f_{\text{Utility}}$; however, by solving a water heater optimization problem, we can evaluate $f_{\text{Utility}}(C(t))$ for any given $C(t)$. To keep the utility from overcharging customers, we require the price of power to be no more than some maximum value $C_{\text{max}}$. This constraint alone is not enough to ensure a fair price for power; e.g., there is nothing stopping the utility from setting $C(t) \equiv C_{\text{max}}$. Accordingly, we add the additional constraint that the average cost of power for a given period must equal some predefined value $C_{\text{avg}}$. However, the utility can still produce devious prices; e.g., to meet the
average cost constraint, the utility can set the price to be extremely low at some obscure time when the demand for power is minimal. Hence, we also require the price to be at least some minimum value $C_{\text{min}}$. Altogether, the utility optimization problem is

$$\begin{align*}
\text{minimize} \quad & C(t) \\
\text{subject to} \quad & L(t) = f_{\text{utility}}(C(t)), \\
& C_{\text{avg}} = \frac{1}{t_{N-1}} \int_0^{t_{N-1}} C(t) \, dt, \\
& C(t) \in [C_{\text{min}}, C_{\text{max}}], \text{ for all } t \in [0, t_{N-1}].
\end{align*}$$

The interplay between water heater and utility is summed up in Figure 6.1.

We model multiple water heaters independently; that is, water heaters are allowed to have different usage patterns and customer preferences. The only assumption we make is that all water heaters have the same cycle length, in particular, $t_{N-1}$. 

\textbf{Figure 6.1:} Block diagram of the customer-utility optimization problem. In the static version of the problem, the stage $k$ is fixed.
6.2.2 Water Heater Expected Load

Although the explicit form of the utility dynamics are unknown, we can approximately evaluate \( f_{\text{Utility}}(C(t)) \) using the following procedure (see Algorithm 9). We assume the utility optimization problem starts at \( k = 0 \). Let \( x_0 = (T_0, t_0) \) be the initial state of a water heater. Let \( L \) be the length-\( N \) vector with components \( L(k) = L_k \), for \( k = 0, \ldots, N - 1 \). The state probability vector at time \( t_k \) is \( \pi_k \); that is, the components of \( \pi_k \) are \( \pi_{i,k} := \Pr(T_k = T_i \mid \mu, x_0) \), for \( i = 0, \ldots, n - 1 \). The algorithm applied to one water heater can be described as follows (see Figure 6.1(a)): First, we solve the water heater optimization problem 3.8 for a given \( C(t) \) to obtain a policy \( \mu_C \) for the water heater. Next, we calculate the \( \mu_C \)-dependent probability matrix \( P_{\mu C} \) (see Algorithm 5). We then initialize the initial probability vector \( \pi_0 \) and calculate the state probability vectors using the recursion \( \pi_{k+1} = \pi_{k} P_{\mu C}^{(k)} \). The expected load at stage \( t_k \) is given by \( L_k = \pi_{k} \mu_C^{(k)} \). We approximate \( f_{\text{Utility}}(C(t)) \) with \( \tilde{L}(t) \), which is the linear interpolation of \( \{L_k \mid k = 0, 1, \ldots, N - 1\} \). The value of \( \tilde{L}(t) \) is added to the approximation of the total load \( \hat{L}(t) \). For a multiple water heater problem, we simply repeat the single water heater process for every water heater (see Figure 6.1(b)).

6.2.3 Heuristic Utility Optimization Procedure

We solve the utility optimization problem (6.1) heuristically. In particular, we discretize time into the set \( \Omega_t \). We denote the estimated expected load vector as \( \hat{L} \), where the components of \( \hat{L} \) are \( \hat{L}_k := \frac{1}{\Delta t} \int_{t_k}^{t_{k+1}} \hat{L}(t) \, dt \). Similarly, we denote the price of power at time \( t_k \) as \( C_k \) and we define \( C \) as the length-\( n \) vector with components \( C^{(k)} = C_k \). The state of the utility is \( \hat{L} \) and its control variable is \( C \). Let \( \hat{C}(t) \) be the linear interpolation of \( \{C_k \mid k = 0, 1, \ldots, N - 1\} \). The utility’s objective function is \( J_{\text{Utility}}(L) = \beta_1 \|L\|_1 + \beta_2 \|L\|_2^2 \). The utility dynamics are given by \( \hat{L} = f_{\text{Utility}}(\hat{C}(t)) \).

If \( f_{\text{Utility}} \) is differentiable with respect to \( C \), then we can attempt to find a local optima of (the unconstrained) \( J_{\text{Utility}} \) using gradient descent. The update rule is given by

\[
C \leftarrow C + \eta \frac{dJ(L)}{dC}(2\beta_2 L + \beta_1 1),
\]

where \( \eta \) is the step size parameter and \( \frac{dJ(L)}{dC} \) is the Jacobian matrix of \( J_{\text{Utility}} \). Notice that the gradient descent update is an affine function of \( L \). Guided by this insight, we consider the
A heuristic update,
\[ C \leftarrow (1 - \eta)C + \eta(\gamma_2^* L + \gamma_1^*), \]

where \( \gamma_1^* \) and \( \gamma_2^* \) are such that the constraints in (6.1) are satisfied and \( \gamma_2^* \) is maximal. Notice that the heuristic update is also an affine function of \( L \). Here, maximizing \( \gamma_2^* \) is based on the intuition that increasing the price of power \( C_k \) will likely decrease the load \( L_k \).

To derive expressions for \( \gamma_1^* \) and \( \gamma_2^* \), we first combine the second constraint in (6.1) and the price equation \( C^{(k)} = \gamma_2 L^{(k)} + \gamma_1 \) to yield
\[ \gamma_1 = -\gamma_2 L_{\text{avg}} + C_{\text{avg}}, \tag{6.3} \]
where \( L_{\text{avg}} \) is the average element in \( L \). We can now eliminate \( \gamma_1 \) from the price equation, which is now
\[ C^{(k)} = \gamma_2 (L^{(k)} - L_{\text{avg}}) + C_{\text{avg}}. \]

The goal now is to maximize \( \gamma_2 \) subject to \( C_{\text{min}} \leq C^{(k)} \leq C_{\text{max}} \) or
\[ C_{\text{min}} - C_{\text{avg}} \leq \gamma_2 (L^{(k)} - L_{\text{avg}}) \leq C_{\text{max}} - C_{\text{avg}}. \]

There are two cases\(^1\) that need to be considered:

\[
\begin{align*}
L^{(k)} - L_{\text{avg}} > 0 & \implies \frac{C_{\text{min}} - C_{\text{avg}}}{L^{(k)} - L_{\text{avg}}} \leq \gamma_2 \leq \frac{C_{\text{max}} - C_{\text{avg}}}{L^{(k)} - L_{\text{avg}}} \\
L^{(k)} - L_{\text{avg}} < 0 & \implies \frac{C_{\text{max}} - C_{\text{avg}}}{L^{(k)} - L_{\text{avg}}} \leq \gamma_2 \leq \frac{C_{\text{min}} - C_{\text{avg}}}{L^{(k)} - L_{\text{avg}}},
\end{align*}
\]

for all (respectively) \( k = 1, \ldots, N \). Thus, we have that \( \gamma_2 \) can be at most

\[
\begin{align*}
L^{(k)} - L_{\text{avg}} > 0 & \implies \gamma_2^* \leq \frac{C_{\text{max}} - C_{\text{avg}}}{L_{\text{max}} - L_{\text{avg}}} \\
L^{(k)} - L_{\text{avg}} < 0 & \implies \gamma_2^* \leq \frac{C_{\text{min}} - C_{\text{avg}}}{L_{\text{min}} - L_{\text{avg}}},
\end{align*}
\]

where we defined \( L_{\text{max}} \) and \( L_{\text{min}} \) to be the maximum and minimum element in \( L \), respectively.

Overall, we have
\[ \gamma_2^* = \min \left\{ \frac{C_{\text{max}} - C_{\text{avg}}}{L_{\text{max}} - L_{\text{avg}}}, \frac{C_{\text{min}} - C_{\text{avg}}}{L_{\text{min}} - L_{\text{avg}}} \right\}. \]

\(^1\)The case when \( L^{(k)} - L_{\text{avg}} = 0 \) cannot be used to impose restrictions on \( \gamma_2 \).
Substituting the above equation into Eq (6.3) gives

\[ \gamma_1^* = C_{avg} - \gamma_2^* L. \]

The benefit of this heuristic method is that a simple closed-form expression for the price update exists. However, the disadvantage is that it is essentially oblivious to the objective function of the utility. Randomized algorithms, such as genetic algorithms and simulated annealing, offer an alternative to our method for determining the best price. Another possibility would be to restrict the form of the price in which we optimize over; e.g., we can try to find the best piecewise constant function for price.

6.3 Problems With Weekly, Seasonal, and Nonstationary Usage Patterns

In this thesis, we considered a periodic system. In the numerical simulations, we took the period to be one day. However, it is likely that a period longer than a day would model hot water usage patterns better. For example, many households have weekday hot water usage patterns that differ from their weekend usage patterns. More generally, hot water usage patterns are likely to vary by day of the week and, in many cases, usage patterns vary with the month of the year (especially in regions far from the equator, where there are large seasonal changes in climate). Furthermore, usage patterns may be nonstationary. In these cases, we should extend the period of the problem to reflect these usage patterns. However, if the length of the period is too long, then it is unlikely that we will be able to observe enough data to calculate a reliable estimate of the usage pdf. Instead, we can introduce a rolling window of data that can be used for the estimation procedure. The correlation in usage patterns most likely decays over time. For example, we would expect last week’s usage patterns to be similar to the current week’s patterns; however, we do not expect the usage patterns from several months ago to be very much like the current week’s patterns. To address this case, we can weight the observations so that the most current data receives more attention.
Chapter 7

Conclusion

We formulate the problem of determining an on/off policy for a residential water heater that minimizes the weighted sum of an operating cost and a discomfort cost, as an average cost Markov decision problem. Dynamic programming offers a framework for an optimal solution, but its implementation is difficult for all but the most simplified problems. We hence use approximate versions of dynamic programing. In particular, we use state-aggregation to develop an aggregate problem with a smaller state space. We develop a model of the water heater dynamics which we use along with kernel density estimation to calculate approximate transition probabilities for the aggregate problem. Relative value iteration provides a relatively fast way of solving the aggregate problem. We also consider the Q-Learning algorithm which does not require a model of the water heater or the transition probabilities. When compared to a water heater with a single set-point, the ADP algorithms reduce heating costs by about 15%. A theoretical lower bound for the objective costs of the original problem is given in terms of the prescient lower bound. The approximate dynamic programming methods come relatively close to achieving this lower bound. The methodologies considered in this paper create a solid foundation in which other exciting problems can be solved; e.g., optimizing a solar water heater and automated demand response via residential water heater control can be accomplished by extending the algorithms and models considered here. Due to the challenges presented with the integration of renewables, it is technological innovations such as these that will help towards reaching ambitious renewable energy goals.
Appendix A

Physical Implementation

In order to calibrate the coefficients in Eq. (3.3) and determine the effectiveness of the control strategies developed, an electric water heater was outfitted with temperature sensors, flow sensors, and a relay for on/off control. Rather than employing a standard 40-gallon capacity water heater with two heating elements, a 20-gallon water heater with just one heating element was chosen to simplify the thermodynamic modeling of the heat in the tank.

The on-off function of the water heater was controlled by an AC solid-state relay. A solid-state relay was chosen over a mechanical relay due to increased speed, a longer life, and reduced electromagnetic interference. However, since solid-state relays produce a considerable amount of heat, a heat sink was also installed. The relay was controlled by a Raspberry Pi, an ARM-based single-board computer, to transmit data to the base server for data aggregation via Wi-Fi. Also connected to the Raspberry Pi were three digital temperature sensors that lined the sidewall of the water heater and a Hall effect flow sensor connected to the input of the water heater. The DS18B20 temperature sensors utilize the Dallas Semiconductors 1-Wire protocol, allowing them to be connected in parallel, using one port pin for communication. The flow sensor featured an internally rotating wheel that pulses a Hall effect sensor; the rate of the pulses were then plotted against and related to the volume of water in the input pipe. The output valve for the water heater was controlled by a servo controlled by the Raspberry Pi. When the output valve is opened, water flows into the water heater to maintain a constant water level in the tank. In this way, the volume of water drawn from the output of the water heater can be determined by the flow sensor on the input pipe. A block diagram of the physical hardware implementation can be seen in Figure A.1.
Figure A.1: A block diagram of the water heater hardware.
Appendix B

Water Heater Thermodynamics

Here we derive the thermodynamic equation governing the evolution of the temperature of the water heater. To derive such an equation, we start by assuming radial symmetry of the water heater. Furthermore, we assume that the temperature at a given height is uniform. Thus, we can treat the water heater one dimensional object since the temperature now only depends on the depth of the water in the water heater $y \in [0,1]$. These assumptions agree well with intuition, since the water heater is radially symmetric (except for the heating coils and inlet and outlet valves). The conservation of energy equation is

$$\frac{\partial T(t,y)}{\partial t} + \frac{\partial F(t,y)}{\partial y} = 0,$$

where $F(t,y)$ is the temperature flux through a circular cross-section of the water heater at a depth of $y$ given by

$$F(t,y) = w(t)T_y(t,y) - a\frac{\partial F(t,y)}{\partial y}.$$

Here the first term is the advective flux that results as temperature is moved via the physical flow of water. The second term comes from Fourier’s Law which states that the diffusive flux is proportional to the temperature gradient (cf. [46, Eq. (1.8)]) where $a$ is the thermal diffusivity of water. Because of losses to the environment, temperature is not conserved; instead, it decays according to Newton’s Law of Cooling (cf. [46, Eq. (1.3)]). Thus, combining the above and using shorthand notation for derivatives, we have

$$\dot{T}(t,y) + w(t)T_y(t,y) - aT_{yy}(t,y) = -r_{cool}[T(t) - T_{amb}],$$

67
which is known as the Advection-Diffusion-Decay equation (also known as the Advection-Diffusion-Reaction equation when the sign in front of $r_{\text{cool}}$ is positive).

We can simplify the above vector equation into a scalar equation by assuming that the water heater is well-mixed and so there is no thermal stratification throughout the tank (that is, the temperature throughout the water heater is uniform). This assumption is equivalent to assuming that $a$ is sufficiently large so that $T_{yy} \approx 0$ and therefore $T_y$ is essentially a constant with respect to $y$, let this constant (w.r.t. $y$) be $r_{\text{loss}}(t)$. Since there is no physical reason why $r_{\text{loss}}(t)$ should be a function of time, we further assume that $r_{\text{loss}}$ is also a constant with respect to $t$. We now have

$$\dot{T}(t, y) = -r_{\text{cool}}[T(t) - T_{\text{amb}}] - r_{\text{loss}}(t)w(t).$$

By grouping the terms involving $y$, we see that $T(t, y)$ is also a constant with respect to $y$. Adding the control action as a temperature source term we get

$$\dot{T}(t) = -r_{\text{cool}}[T(t) - T_{\text{amb}}] + r_{\text{heat}}u(t) - r_{\text{loss}}w(t),$$

where $r_{\text{heat}}$ is the rate at which the water heater heats water.

To solve the above equation we proceed as follows. By multiplying by $e^{r_{\text{cool}}t}$ and rearranging, we get

$$\frac{d}{dt}[e^{r_{\text{cool}}t}(T(t) - T_{\text{amb}})] = e^{r_{\text{cool}}t}[r_{\text{heat}}u(t) - r_{\text{loss}}w(t)].$$

Since we do not know $w(t)$, we substitute its piecewise linear approximation given in Eq. (3.1), and then integrate from $t_k$ to $t_k + \Delta t$ to get

$$e^{r_{\text{cool}}(t_k + \Delta t)}(T_{k+1} - T_{\text{amb}}) - e^{r_{\text{cool}}t_k}(T_k - T_{\text{amb}}) = \left(e^{r_{\text{cool}}(t_k + \Delta t)} - e^{r_{\text{cool}}t_k}\right)\frac{r_{\text{heat}}}{r_{\text{cool}}}u_k - \int_{t_k}^{t_k + \Delta t} e^{r_{\text{cool}}t} r_{\text{loss}} \tilde{w}(t) \, dt,$$

where the fact that $u(t)$ is constant on $[t_k, t_k + \Delta t]$ was used. Using the definition in Eq. (3.1), we solve for $T_{k+1}$ to get

$$T_{k+1} = T_k - \left(1 - e^{-r_{\text{cool}}\Delta t}\right) \left[(T_k - T_{\text{amb}}) + \frac{r_{\text{heat}}}{r_{\text{cool}}}u_k\right] - \frac{r_{\text{loss}}}{r_{\text{cool}}} \left[\tilde{w}(t_k + \Delta t) - e^{-r_{\text{cool}}\Delta t} \tilde{w}(t_k) - \frac{\tilde{w}(t_k + \Delta t) - \tilde{w}(t_k)}{r_{\text{cool}}\Delta t} \left(1 - e^{-r_{\text{cool}}\Delta t}\right)\right].$$

68
Using the Taylor Series expansion of the exponential and then dropping all \( O((\Delta t)^2) \) terms gives

\[
T_{k+1} = T_k - r_{cool} \Delta t (T_k - T_{amb}) + r_{heat} \Delta t u_k - r_{loss} \Delta t w_k,
\]

where we recall \( w_k = \frac{1}{2}[w(t_k) + w(t_k + \Delta t)] \). Finally, we assume that water enters the water heater at ambient temperature; thus,

\[
T_{k+1} = \max \{T_k - r_{cool} \Delta t (T_k - T_{amb}) + r_{heat} \Delta t u_k - r_{loss} \Delta t w_k, T_{amb}\}.
\]

Recall the constraint \( T_k \in [T_{amb}, T_{max}] \) and the corresponding set of feasible states \( \Omega_u(x_k) \).

From Eq. (3.3), \( u_k = 0 \) is always a feasible decision; on the other hand, \( u_k = 1 \) is only feasible for \( T \) such that \( f_T(T_k, 1, 0) \leq T_{max} \). We can write the set of feasible decisions available at the current state as

\[
\Omega_u(T_k) = \{0\} \cup \{I\{T_k \leq T'\}\},
\]

where \( T' \) is the largest value of \( T_k \) such that \( f_T(T_k, 1, 0) \leq T_{max} \), specifically

\[
T' = \arg \max_{T_k} \{T_k \cdot I\{f_T(T_k, 1, 0) \leq T_{max}\}\} = \frac{T_{max} - T_{amb} - r_{heat} \Delta t}{1 - r_{cool} \Delta t} + T_{amb}.
\]

In the remainder of this section, we derive formulas for the coefficients of Eq. (3.3). In our numerical simulations, we use the coefficient values given in Table B.0(a). The physical parameters of the water heater used to calculate coefficients are given in Table B.0(b) are taken from [47].

### B.1 Heating Rate

The First Law of Thermodynamics for a closed system with no work states

\[
q = \frac{dU}{dt} = mass \cdot specific \ heat \cdot \frac{dT}{dt}, \tag{B.1}
\]

where \( q \) is the rate of heat convection from the water heater, \( U \) is the energy contained in the object, \( mass \) is the mass of the object, \( specific \ heat \) is the specific heat of the object (assuming the object is incompressible and homogeneous) [46, Eq. (1.3)]. The amount of energy contained in the water heater’s tank at time \( t_k \) is equal to the amount of convective heat that would be
(a) Coefficients of the Water Heater Thermodynamic Equation

<table>
<thead>
<tr>
<th>Family Size</th>
<th>$r_{\text{heat}}$ (°C/h)</th>
<th>$r_{\text{cool}}$ (%/h)</th>
<th>$\hat{r}_{\text{loss}}$ (°C/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.53</td>
<td>2.01</td>
<td>67.50</td>
</tr>
<tr>
<td>1-2</td>
<td>24.62</td>
<td>1.90</td>
<td>45.00</td>
</tr>
<tr>
<td>2-3</td>
<td>23.49</td>
<td>1.69</td>
<td>33.75</td>
</tr>
<tr>
<td>3-4</td>
<td>18.58</td>
<td>1.64</td>
<td>27.00</td>
</tr>
<tr>
<td>5+</td>
<td>11.74</td>
<td>1.35</td>
<td>16.88</td>
</tr>
</tbody>
</table>

(b) Physical Parameters of the Water Heater

<table>
<thead>
<tr>
<th>Family Size</th>
<th>V (gal.)</th>
<th>r (in.)</th>
<th>z (in.)</th>
<th>efficiency</th>
<th>rating (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>9.5</td>
<td>26</td>
<td>0.92</td>
<td>3500</td>
</tr>
<tr>
<td>1-2</td>
<td>30</td>
<td>9</td>
<td>45.25</td>
<td>0.93</td>
<td>3500</td>
</tr>
<tr>
<td>2-3</td>
<td>40</td>
<td>10</td>
<td>47.25</td>
<td>0.92</td>
<td>4500</td>
</tr>
<tr>
<td>3-4</td>
<td>50</td>
<td>10</td>
<td>58.5</td>
<td>0.91</td>
<td>4500</td>
</tr>
<tr>
<td>5+</td>
<td>80</td>
<td>12</td>
<td>59.25</td>
<td>0.92</td>
<td>4500</td>
</tr>
</tbody>
</table>

Table B.1: The parameters of the water heater model for different family sizes.

lost to the environment over the interval $[t_k, \infty)$. Integrating over $[t_k, \infty)$ yields

$$U_k = \text{mass} \cdot \text{specific heat} \cdot (T_k - T_{\text{amb}}),$$

where we defined the reference energy at $\lim_{t \to \infty} T(t) = T_{\text{amb}}$ to be $\lim_{t \to \infty} U(t) = 0$. Turning on the water heater for $\Delta t$ time units increases the energy in the tank by

$$U_{\text{on}} = \text{efficiency} \cdot \text{rating} \cdot \Delta t,$$

where efficiency is the percent efficiency of the water heater (that is, the amount of energy that gets converted into heat) and rating is the power rating of the water heater. Using the equations above we have

$$U_{k+1} = U_k + U_{\text{on}}$$

or

$$T_{k+1} = T_k + \frac{\text{efficiency} \cdot \text{rating}}{\text{mass} \cdot \text{specific heat}} \Delta t.$$

Therefore, the heating rate is given by

$$r_{\text{heat}} = \frac{\text{efficiency} \cdot \text{rating}}{\text{mass} \cdot \text{specific heat}}.$$
B.2 Cooling Rate

Consider Newton’s Law of Cooling

\[
\frac{d(T - T_{amb})}{dt} = -\frac{\kappa A}{\rho c V} (T - T_{amb}),
\]

where \(\kappa\) is known to be the convective heat transfer coefficient and is an intrinsic property of an object; \(A\) and \(V\) is the surface area and volume of the water heater, respectively [46, Eq. (1.20)].

Let \(r\) and \(z\) be the radius and height of the water heater, respectively. Substituting the equations for the surface area and volume of a cylinder into (B.2) we obtain (after simplification)

\[
r_{cool} = \frac{2\kappa(h + r)}{\rho c r}.
\]

We experimentally determined that for \(r = 9.5\) in and \(z = 26\) in the cooling rate is \(r_{cool} \approx 2.01\,\%/h\). Hence \(2\kappa/(\rho c) \approx 5.77\) in.\%/h. The cooling rates used in simulations assume a 2in. shell of insulation and so the dimensions of the water heater were modified appropriately before using (B.3) (see Table B.0(b)).

B.3 Loss Rate

Recall that we defined \(w_k\) to be the average power used during \([t_k, t_{k+1}]\). To convert an energy to a temperature we divide by mass \(\times\) specific heat. Thus, by inspection of Eq. (3.3), we have \(r_{loss} = 1/(\text{mass }\times\text{ specific heat})\). Hence, \(r_{loss}\) is not really a rate but rather the conversion factor from power to a temperature loss rate. In order to easily compare \(r_{heat}\) to the other rates, we define the reference loss rate

\[
r_{loss} = r_{loss} [\text{specific heat} \times (1.5\,\text{gal}/\text{min}) \times (40 - 25)\,\text{°C}],
\]

which is the rate at which temperature is lost (°C/h) during a hot water draw with flow rate 1.5 gal/min, desired temperature 40°C, and ambient temperature 25°C.
Appendix C

Hot Water Simulation

The random variables defining an event are assumed to have a truncated normal distribution; that is, a truncated normal random variable $X$ with mean $\bar{x}$, standard deviation $\sigma_x$ and support $[a, b]$ has the following pdf

$$p_X(x) = \begin{cases} 
  c_X \exp \left[ -\frac{(x-\bar{x})^2}{2\sigma_x^2} \right] & x \in [a, b], \\
  0 & \text{otherwise}; 
\end{cases}$$

where $c_X$ is a constant chosen so that the pdf integrates to 1. Furthermore, we take $\bar{x}$ to be a random sample from another truncated normal random variable $\bar{X}$. We refer to the parameters defining the distribution of $\bar{X}$ as the population parameters; and for a given value of $\bar{x}$, we refer to the parameters defining a particular random variable as the personal parameters. The parameters used in simulations are shown in Table C.1.

Table C.1: Shower Probability Distribution Parameters.

<table>
<thead>
<tr>
<th></th>
<th>Personal</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Support</td>
<td>SD\textsuperscript{a}</td>
</tr>
<tr>
<td>$S$ (morning)</td>
<td>[0, 24]</td>
<td>0.5/2</td>
</tr>
<tr>
<td>$S$ (night)</td>
<td>[0, 24]</td>
<td>0.5/2</td>
</tr>
<tr>
<td>$D$</td>
<td>[2.5, 30]</td>
<td>(2.5/60)/2</td>
</tr>
<tr>
<td>$F$</td>
<td>[0.75, 2.5]</td>
<td>0.25/2</td>
</tr>
<tr>
<td>$T$</td>
<td>[36, 48]</td>
<td>1/2</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The standard deviation is reported as $\sigma = y/2$ so that the random variable $X$ is such that $P[|x - \bar{x}| \leq y] \approx 95.4$.

\textsuperscript{b}The averages and ranges for the support are taken directly from [48].
The problem with using a TGauss distribution is that it is discontinuous at its endpoints. This does not make much practical sense. To methods of creating distributions that are continuous everywhere are discussed next.

One way to obtain a distribution that is zero at its endpoints is to treat the probability distribution as a temperature within a rod of length \([a, b]\). If the TGauss distribution is used as an initial condition, then solving a boundary value problem with the endpoints held at zero, produces a smoothed TGauss distribution. Normalizing turns the temperature distribution into a probability distribution.
One way to mimic the behavior of the solution to the above PDE it to treat the boundaries as reflecting. In which case we can obtain an approximate solution the BVP by simply extending the support of the TGauss distribution to \([2a - b, 2b - a]\), then reflecting the distribution on \([2a - b, a]\) onto the distribution on \([a, b]\). Similarly, we reflect the distribution on \([b, 2b - a]\) on to the distribution on \([b, 2b - a]\). Again, normalizing the temperature ensures a probability distribution.
Appendix D

Water Heater Algorithms

Here, we provide MATLAB-like psuedo code for the algorithms used in this thesis.

D.1 Hot Water Simulation

In this section we state algorithm used to simulate hot water usage from the random quantities defined in section 3.1.3.

Algorithm 1: Hot Water Usage Simulation

```plaintext
Initialize $W(t) \equiv 0$

for $\tau \in \Omega$ do

for $i \in \{1, \ldots, N_{people}\}$ do

for $j \in \{1, \ldots, N_{\tau,i}\}$ do

Obtain samples of $S_{\tau,i}^{(j)}, D_{\tau,i}^{(j)}, F_{\tau,i}^{(j)}, T_{\tau,i}^{(j)}$;

while $E_{\tau,i}^{(j)}$ is unrealistic do

Resample $S_{\tau,i}^{(j)}, D_{\tau,i}^{(j)}, F_{\tau,i}^{(j)}, T_{\tau,i}^{(j)}$;

end

$W(t) \leftarrow W(t) + \text{specific heat} \cdot F_{\tau,i}^{(j)} \cdot (T_{\tau,i}^{(j)} - T_{amb}) \cdot I\{S_{\tau,i}^{(j)} \leq t < S_{\tau,i}^{(j)} + D_{\tau,i}^{(j)}\}$;

end

end

end
```

Notes: Here $W(t) \equiv 0$ means $W(t)$ is identically zero, that is, $W(t) = 0$ for all $t$; furthermore, $I\{\cdot\}$ denotes the indicator function; i.e., $I\{\cdot\} = 1$ if the event $\{\cdot\}$ is true, otherwise $I\{\cdot\} = 0$. Here the constraints defining the realistic events are implicit.
D.2 Dynamic Programming

In this section, we state the algorithms relating to dynamic programming.

Algorithm 2: Relative Value Iteration

1. Initialize $\mu \leftarrow 0^{n \times N}$; $J \leftarrow T_{\text{max}} - T$; $\mu_{\text{ref}} \leftarrow \mu + 1$;
2. while $\mu \neq \mu_{\text{ref}}$ do
3. \hspace{1em} $\mu_{\text{ref}} \leftarrow \mu$;
4. \hspace{1em} for $k = N - 1, \ldots, 1, 0$ do
5. \hspace{2em} $[J, \mu^{(i,k)}] \leftarrow \text{eimin}_{u \in \{0,1\}} [g^{(i,k,u)}_b p^{(i,k)}_{W} + P^{(i,k,u)} J]$;
6. \hspace{1em} end
7. \hspace{1em} $J_{\text{ref}} \leftarrow J^{(n)}$; $J \leftarrow J - J_{\text{ref}}$
8. end

Notes: Here, eimin is the element-wise minimum that returns both the vector of minimum values $J$ and the vector of minimizers $\mu^{(i,k)}$. We can perform every iteration of $k$ by keeping in memory only current the values of $J$; this reduces the memory requirement by a factor of $N$.

Algorithm 3: Aggregate Transition Matrix Calculation

1. Initialize $P_A \leftarrow 0^{n \times n \times N \times 2}$;
2. for $u \in \{0,1\}$ do
3. \hspace{1em} for $l \in \{0, \ldots, m - 1\}$ do
4. \hspace{2em} for $i \in \{0, \ldots, n - 1\}$ do
5. \hspace{3em} $T \leftarrow f(T^{(i)}, u, w^{(l)})$; $\bar{T} \leftarrow \text{round}(T/\Delta T)\Delta T$; $a = |T - \bar{T}|/\Delta T$
6. \hspace{3em} if $T > T_{\text{max}}$ then $P_{A}^{(i,u,l)} \leftarrow 1$, continue;
7. \hspace{3em} $j \leftarrow (\bar{T} - T_{\text{amb}})/\Delta T$; $P_{A}^{(i,j,u,l)} \leftarrow 1 - a$; // Primary transition
8. \hspace{3em} $j \leftarrow j + \text{sgn}(T - \bar{T})$; $P_{A}^{(i,j,u,l)} \leftarrow P_{A}^{(i,j,u,l)} + a$; // Compensation transition
9. \hspace{2em} end
10. end
11. end
12. Store $P_{A}^{(i,j,u,l)}$ as a sparse matrix;

Notes: The continue command passes control to the next iteration of a loop.

D.3 Load Foresting

In this section we state the algorithm used in section 6.2 to obtain the expected load of a water heater for a given policy $\mu_C$. 

76
Algorithm 4: Temperature Transition Matrix Calculation

1. Initialize $P \leftarrow 0^{n \times n \times N \times 2}$;
2. for $u \in \{0, 1\}$ do
   3. for $k \in \{0, \ldots, N - 1\}$ do
      4. for $l \in \{0, \ldots, m - 1\}$ do
         5. if $P_{WL}^{(l,k)} = 0$ then continue;
         6. $P^{(:,k,u)} \leftarrow P^{(:,k,u)} + P^{(:,u,l)} A^{(l,k)} W$;
      7. end
   8. Store $P^{(:,k,u)}$ as a banded matrix;
3. end
4. end

Notes: We use the MATLAB-like notation $P^{(:,k,u)}$ to represent the $n \times n$ transition probability matrix given that the current stage and decision is $k$ and $u$, respectively.

Algorithm 5: Policy-Dependent Transition Matrix Calculation

1. Initialize $P_{\mu} \leftarrow P^{(:,0)}$;
2. for $k \in \{0, \ldots, N - 1\}$ do
   3. for $i \in \{0, \ldots, n - 1\}$ do
      4. if $\mu^{(i,k)} = 1$ then $P^{(i,k)} \leftarrow P^{(i,k,1)}$;
   5. end
3. end

Algorithm 6: Model-Free Relative Q-Learning

1. Initialize $\mu \leftarrow 0^{n \times N \times M}$, $Q \leftarrow 0^{n \times N \times M \times 2}$, $\gamma \leftarrow 1^{n \times N \times M}$, $J \leftarrow T_{max} - T$;
2. for $k_{cycle} = 1, \ldots, N_{cycle}$ do
   3. Apply the current policy $\mu$;
   4. Observe $(w_k)_{k=0}^{N-1}$ and $(T_k)_{k=0}^{N-1}$;
   5. Calculate $(\phi_k)_{k=0}^{N-1}$ and $i_N \leftarrow \text{round}((T_{N-1} - T_{amb})/\Delta T)\Delta T$;
   6. for $k = N - 1, \ldots, 1, 0$ do
      7. $l_k \leftarrow \text{round}(w_k/\Delta w)$; $i_k \leftarrow \text{round}((T_k - T_{amb})/\Delta T)\Delta T$;
      8. for $u_k \in \{0, 1\}$ do
         9. $Q^{(i_k,k,\phi_k,u_k)} \leftarrow [1 - \gamma^{(i_k,k,\phi_k)}] Q^{(i_k,k,\phi_k,u_k)} + \gamma^{(i_k,k,\phi_k)} [g(T_k, k, u, l_k; \theta) + J^{(i_{k+1},k+1,\phi_k+1)}]$;
      10. end
      11. $\gamma^{(i_k,k,\phi_k)} \leftarrow (1/\gamma^{(i_k,k,\phi_k)} + 1)^{-1}$; $[J, \mu^{(i_k,k,\phi_k)}] \leftarrow \text{elem}_u^{\mu^{(i_k,k,\phi_k)}}$;
      12. end
   13. $J_{\text{ref}} \leftarrow J^{(n)}$; $J \leftarrow J - J_{\text{ref}}$
3. end

Notes: We can approximate the update in line using the precalculated $g_{\theta}^{(:,k,u,l)}$ instead of $g(T_k, k, u, l_k; \theta)$.
Algorithm 7: Model-Based Relative Q-Learning

1. Initialize $Q \leftarrow 0^{n \times N \times M \times 2}$, $\gamma \leftarrow 1^{N \times M}$, $J \leftarrow T_{\text{max}} - T$;

2. for $k_{\text{cycle}} = 1, \ldots, N_{\text{cycle}}$ do

3. Observe $(w_k)_{k=0}^{N-1}$ and calculate $(\phi_k)_{k=0}^{N-1}$;

4. for $k = N - 1, \ldots, 1, 0$ do

5. $l_k \leftarrow \text{round}(w_k / \Delta w)$;

6. for $u_k \in \{0, 1\}$ do

7. $Q(:\,, k, \phi_k, u_k) \leftarrow [1 - \gamma(k, \phi_k)] Q(:\,, k, \phi_k, u_k) + \gamma(k, \phi_k) [g^0_{\phi_k}(:\,, k, u_k, l_k) + P_{\text{A}}(\cdot, u_k, l_k) J]$;

8. end

9. $\gamma(k, \phi_k) \leftarrow \frac{1}{\gamma(k, \phi_k) + 1} - 1$;

10. $J \leftarrow \min_{u_k \in \{0, 1\}} Q(\cdot, k, \phi_k, u_k)$;

11. end

12. $J_{\text{ref}} \leftarrow J^{(n)}$;

13. $J \leftarrow J - J_{\text{ref}}$;

14. end

15. $\mu \leftarrow \mathbb{I}\{Q(\cdot, 1) < Q(\cdot, 0)\}$

**Notes:** The update to the stepsize vector $\gamma$ in line 9 can be different as long as the convergence criteria in (4.4) are met.

Algorithm 8: Policy Evaluation using Value Iteration

1. Initialize $J \leftarrow J_\mu$;

// Solve finite horizon problem

2. for $k = N - 1, \ldots, 1, 0$ do

3. $J \leftarrow \mu^{(k)} \odot [g^0_{\phi_k}(\cdot, k, 1) P_{\text{W}}^{(k)}] + [1 - \mu^{(k)}] \odot [g^0_{\phi_k}(\cdot, k, 0) P_{\text{W}}^{(k)}] + P^{(\cdot, k)} J$;

4. end

**Notes:** Here $\odot$ denotes element-wise multiplication (Hadamard product).

Algorithm 9: Expected Load Calculation

1. Initialize $L(t) \equiv 0$;

2. for all water heaters do

3. Initialize $L$ arbitrarily;

4. Using Algorithm 2, obtain a policy $\mu_C$;

5. Using Algorithm 5, calculate $P_{\mu_C}$;

6. $\pi_0 \leftarrow 0^{n \times 1}$, $\pi_0^{(\text{round}([T_0 - T_{\text{amb}}]/\Delta T))} \leftarrow 1$;

7. for $k = 0, \ldots, N - 1$ do

8. $L^{(k)} \leftarrow \pi^{(k)}_{\mu_C}$;

9. end

10. $\hat{L}(t) \leftarrow \text{LinInterp}(L)$; $\hat{L}(t) \leftarrow \hat{L}(t) + L(t)$;

11. end
Bibliography


