Energy management of PV-storage systems: ADP approach with temporal difference learning

Conference Paper · June 2016

DOI: 10.1109/PSCC.2016.7540924

3 authors:

Chanaka Keerthisinghe

University of Washington Seattle

Gregor Verbic

The University of Sydney

Archie C. Chapman

The University of Sydney

Energy management of PV-storage systems: ADP approach with temporal difference learning

Chanaka Keerthisinghe, *Student MIEEE*, Gregor Verbič, *Senior MIEEE*, and Archie C. Chapman, *MIEEE*
School of Electrical and Information Engineering
The University of Sydney, Sydney, New South Wales, Australia
Email:{chanaka.keerthisinghe, gregor.verbic, archie.chapman}@sydney.edu.au

***Abstract*—In the future, residential energy users can seize the full potential of demand response schemes by using an automated home energy management system (HEMS) to schedule their distributed energy resources. In order to generate high quality schedules, a HEMS needs to consider the stochastic nature of the PV generation and energy consumption as well as its inter-daily variations over several days. However, extending the decision horizon of proposed optimisation techniques is computationally difficult and moreover, these approaches are only computationally feasible with a limited number of storage devices and a low-resolution decision horizon. Given these existing shortcomings, this paper presents an *approximate dynamic programming* (ADP) approach with *temporal difference learning* for implementing a computationally efficient HEMS. In ADP, we obtain policies from value function approximations by stepping forward in time, compared to the value functions obtained by backward induction in DP. We use empirical data collected during the Smart Grid Smart City project in NSW, Australia, to estimate the parameters of a Markov chain model of PV output and electrical demand, which are then used in all simulations. To evaluate the quality of the solutions generated by ADP, we compare the ADP method to stochastic mixed-integer linear programming (MILP) and dynamic programming (DP). Our results show that ADP computes a solution much quicker than both DP and stochastic MILP, while providing better quality solutions than stochastic MILP and only a slight reduction in quality compared to the DP solution. Moreover, unlike the computationally-intensive DP, the ADP approach is able to consider a decision horizon beyond one day while also considering multiple storage devices, which results in a HEMS that can capture additional financial benefits.**

***Index Terms*—demand response, home energy management, distributed energy resources, approximate dynamic programming, dynamic programming, stochastic mixed-integer linear programming, value function approximation, temporal difference learning.**

NOMENCLATURE

k	Time-step
N	Total number of time-steps
i	Index of storage devices
I	Total number of storage devices
j	Index of stochastic variables
J	Total number of stochastic variables
$s_k^{\{i,j\}}$	State of i or j at time-step k
x_k^i	Decision of storage device i at time-step k
ω_k^j	Variation of stochastic variable j at time-step k
C_k	Cost or reward at time-step k
π	Policy
r	Realisation of random information
R	Total number of random realisations
α	Stepsize
V_k^π	Expected future cost/reward for following π from k
$s^{i,\max}$	Maximum state of a storage device
$s^{i,\min}$	Minimum state of a storage device
μ^i	Efficiency of a storage device [%]
l^i	Losses of a storage device per time-step
a	Index of a segment in the VFA
A_k	Total number of segments at time-step k
z_{ka}	Capacity of a segment
$\bar{v}_{ka}$	Slope of a segment

I. INTRODUCTION

INCREASING penetration of intermittent renewable energy sources (RES) in future electrical power systems will drive a paradigm shift from "supply-following-demand" to "demand-following-supply". A cornerstone of this shift is demand response (DR), which can reduce residential electricity cost on one side and provide system services on the other. At the consumer side, DR can harness the flexibility of residential distributed energy resources (DER), including distributed generation, energy storage and flexible loads, by using an automated *home energy management system* (HEMS).

In Australia, the penetration of rooftop PV and battery storage systems has increased dramatically in response to rising electricity costs, government incentives, decreasing capital costs and growing concerns about climate change. In March 2015, Australia reached 4.2 GW solar PV from residential and commercial users [1], while global solar PV had increased from 4 GW in 2003 to 150 GW in 2014 [2]. Tesla's Powerwall batteries, which were announced in May 2015, provide an economical solution for storage in residential buildings [3]. A study carried out in USA demonstrates that residential users with PV-storage systems should reach grid parity within the next decade [4].

Given this context, this paper focuses on residential PV-storage systems. An illustration of electrical power flow in a residential building is depicted in Fig. 1. When the PV generation is higher than the electrical demand, extra electrical power will be either stored, consumed by shiftable and controllable loads or/and fed back to the electrical grid. However, in Australia, selling power back to the electrical grid is uneconomical since feed-in tariffs (FiTs) are significantly less than the retail tariffs paid by the households. Coupled

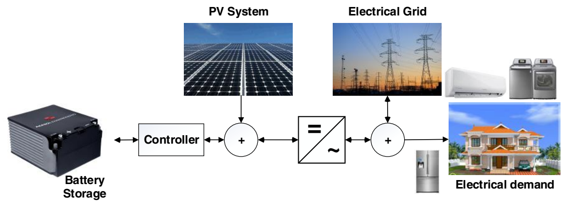

Fig. 1. Illustration of electrical energy flow in a residential PV-storage system

with ever dropping PV costs, there is a strong incentive for PV owners to self-consume as much locally generated power as possible. Our conjecture is that in the near future this may happen in other parts of the world too. Therefore, in order to maximise the benefits of PV-storage system, residential energy users can use an automated HEMS to schedule and coordinate their energy use.

The main objective of such a HEMS is to minimise energy costs while providing suitable levels of comfort to the inhabitants. The underlying optimisation problem can be thought of a sequential decision making process under uncertainty. Specifically, the problem contains stochastic variables such as PV output and electrical demand. Considering these stochastic variables in a HEMS yields better quality schedules compared to the ones obtained from a deterministic optimisation using off-the-shelf solvers [5]. These schedules can be further improved by capturing inter-daily variations in consumption and solar-insolation patterns over several days [6]. To illustrate this, consider a few sunny days with low demand (e.g. a weekend) followed by a few cloudy days with high demand; in anticipation of this, the HEMS may adjust the end-of-day battery state of charge (SOC) to reap significant financial benefits [6].

Currently, stochastic MILP [5]–[8], particle swarm optimisation (PSO) [9] and dynamic programming (DP) approaches [10], [11] have all been proposed for addressing the stochastic nature of this energy management problem. In [7], we identified the drawbacks associated with these existing solution techniques by comparing a HEMS using stochastic MILP to one using a DP approach. The paper showed that the DP approach results in a better quality schedule, because the non-linear characteristics of the storage devices can be modelled in detail, while the stochastic MILP approach requires linear models. On the other hand; DP is computationally difficult because of the dimensionalities of state, action and outcome spaces, while stochastic MILP is computationally feasible over short horizons because it uses scenario reduction techniques to reduce the scenario set so that the problem can be solved within a desired time limit.

Given these insights, this paper presents a computationally efficient HEMS using an *approximate dynamic programming* (ADP) approach with *temporal difference learning*, an approach that has successfully been applied for the control of grid level storage in [12], and provides guidelines for its practical use. A range of ADP methods have already been adopted in several power engineering applications [13], [14], [14]–[19], such as a multidimensional wind energy storage problem in [17] and managing a building cooling system in [18]. In this paper, policies are extracted from value function approximations (VFAs)[1] while stepping forward in time. We use VFA as it has been suggested to work best in a time-dependent problem with daily load, energy and price patterns, relatively high noise and less accurate forecasts (errors grow with the horizon) [17]. Other methods can be found in [16].

In the context of a HEMS, ADP is expected to improve on the computational performance of DP while providing similar solution quality (i.e. DP results in close to optimal solutions when finely discretised). In order to validate this, the performance of ADP over a two-day horizon is benchmarked against DP and stochastic MILP over a daily horizon (i.e. end-of-day one SOC is same for all three techniques) in two different scenarios (i.e. different electrical demand patterns and PV output).

The proposed ADP method enable us to:

1) incorporate stochastic nature of the input variables with less computational difficulty.
2) extend the decision horizon with less computational burden to consider uncertainties over several days (i.e. up to a week), which results in financial benefits.
3) enable integration of multiple storage devices with less computational burden.
4) integrate the HEMS into an existing smart meter as it uses less computer memory compared to existing methods.

In addition, throughout the paper we evaluate the algorithms and show the benefits of residential PV-storage systems with a HEMS using real data collected during the *Smart Grid Smart City* project in NSW, Australia [20].

The paper is structured as follows: Section II states the stochastic energy management problem. This is followed by a description of the ADP formulation in Sections III. Implemented HEMSs are described in Section IV. Section V presents the simulation results and the discussion.

II. Home energy management problem

In this section, we formulate the stochastic optimisation problem and describe the stochastic variables of a residential PV-storage system. In particular, we focus on the nature of the stochasticity and possible ways to model it. Finally, a short description of the existing solution techniques including stochastic MILP and DP is provided.

A. Optimisation problem

The general objective of a HEMS is to minimise energy costs and user discomfort over a decision horizon. In this paper, the problem is formulated as a Markov decision process (MDP), which consists of a state space, $(s \in \mathcal{S})$, a decision space, $(x \in \mathcal{X})$, random information, transition functions and contribution functions. A state variable, $s_k^{\{i,j\}} \in \mathcal{S}$, contains the information that is necessary and sufficient to make the decisions and compute costs, rewards and transitions. Note that i and j relates to storage devices and stochastic variables,

[1] A value function describes the expected future discounted cost of following a policy from a given state.

respectively. The decision/control variable, $x_k^i \in X$, is a control action for each storage device, $i \in \mathscr{I}$, over the decision horizon for all time steps $k \in \{1 \dots N\}$, where k and N denote a particular time-step and the total number of time-steps, respectively. The random variable, $\omega_k^j \in \Omega$, depends on either weather or inhabitants' behavioral pattern. Let: $\mathbf{s}_k = [\ s_k^{\{i,j\}} \dots s_k^{\{1 \dots I, 1 \dots J\}}\ ]^\mathsf{T}$, $\mathbf{x}_k = [\ x_k^i \dots x_k^I\ ]^\mathsf{T}$, and $\boldsymbol{\omega}_k = [\ \omega_k^j \dots \omega_k^J\ ]^\mathsf{T}$. The problem is given by:

$$\begin{aligned}
\min_{\pi} \quad & \mathbb{E}\left\{ \sum_{k=0}^{N} C_k^{\pi}(\mathbf{s}_k, \mathbf{x}_k, \boldsymbol{\omega}_k) \right\}, \\
\text{s.t.} \quad & \text{satisfies storage device, comfort, power flow} \\
& \text{and energy balance constraints,} \\
& \forall k \in \{1 \dots N\}, \\
& \forall i \in \{1 \dots I\}, \qquad (1)
\end{aligned}$$

where: π is a policy, a choice of action for each state, $\pi : \mathcal{S} \to \mathcal{X}$; and $C_k(\mathbf{s}_k, \mathbf{x}_k, \boldsymbol{\omega}_k)$ is the contribution (i.e cost/reward of energy, but may include a discomfort penalty) incurred at a given time-step k, which accumulates over time. The problem is formulated as an optimisation of the expected contribution because the contribution is generally a random variable due to the effect of $\boldsymbol{\omega}_k$. The transition function $\mathbf{s}_{k+1} = \mathbf{s}^M(\mathbf{s}_k, \mathbf{x}_k, \boldsymbol{\omega}_k)$ describes the evolution of states from k to $k+1$, where $\mathbf{s}^M(.)$ is the system model that consists of storage device i's operational constraints such as power flow limits, efficiencies and losses. Note that the transition functions are only required for the storage devices.

B. Stochastic variables

In order to optimise performance, it is important for a HEMS to incorporate stochastic input variables, and to do so over a horizon of several days. In the following, a list of three sources of random variation are enumerated and their effects on the scheduling problem are discussed.

1) PV output: PV output depends on solar insolation, which can be obtained before the horizon starts with a reasonable accuracy from weather forecasting services. PV output is important to the HEMS problem as it is a key source of energy and is expected to be closely coupled with the battery storage profile. Failing to accommodate for variation in PV generation would be expected to increase costs to the household as more power is imported from the grid.

2) Electrical demand of the household: Electrical demand of the household depends on the number of occupants and their behavioral patterns, which is difficult to predict in the real world. In the context of HEMS, electrical demand should be supplied from the DG units, storage units and the electrical grid. Failure to accommodate variations in electrical demand may result in additional costs to the household.

3) Electricity price: In this paper, we assume that the exact electricity prices, $s_k^{\text{p}} \in \{s_k^{\text{p,buy}}, s_k^{\text{p,sell}}\}$, are available before the start of the decision horizon from an residential DR aggregator/retailer.

PV output and the electrical demand varies depending on type of day and therefore, the required battery SOC at end of the day should depend on the PV output and electrical demand in future days. As we identified in [6], controlling the end-of-day battery SOC in order to optimise over a longer decision horizon can reduce the average weekly financial costs.

In this paper, empirical data is used to learn the transition probabilities associated with the uncertain variables, which is more realistic than assuming parametric approaches such as a Gaussian or skew-Laplace distribution. More details are provided in Section IV-A.

C. Solution techniques

The first method used to solve the optimisation problem at hand is a scenario-based MILP approach, which we referred to as *stochastic* MILP in [7]. A MILP approach assumes a linear objective function and linear constraints in (1), and scenarios are used to incorporate stochasticity. The optimisation problem is solved for the whole horizon at once, so the solution time grows exponentially with the length of the horizon. As such, in the existing literature, a one day optimisation horizon is typically assumed. Moreover, the solutions are lower quality because of the linear approximations made and the inability to incorporate all the probability distributions. In response to these limitations, DP was proposed in [11] to improve the solution quality.

DP solves the optimisation problem of the form in (1) by computing a value function $V^{\pi}(\mathbf{s}_k)$, which is the expected future discounted cost of following a policy, π, starting in state, $\mathbf{s}_k$, and is given by:

$$V^{\pi}(\mathbf{s}_k) = \sum_{\mathbf{s}' \in \mathcal{S}} \mathbb{P}(\mathbf{s}'|\mathbf{s}_k, \mathbf{x}_k, \boldsymbol{\omega}_k) \left[C(\mathbf{s}_k, \mathbf{x}_k, \mathbf{s}') + V^{\pi}(\mathbf{s}') \right], \qquad (2)$$

where $\mathbb{P}(\mathbf{s}'|\mathbf{s}_k, \mathbf{x}_k, \boldsymbol{\omega}_k)$ is the transition probability of landing on state $\mathbf{s}'$ from $\mathbf{s}_k$ if we take action $\mathbf{x}_k$.

An optimal policy, π^*, is one that optimises (1). It can be found by recursively computing the optimal value function, $V_k^{\pi^*}(\mathbf{s}_k)$, by using Bellman's optimality condition:

$$V_k^{\pi^*}(\mathbf{s}_k) = \min_{\mathbf{x}_k \in \mathcal{X}_k} \left(C_k(\mathbf{s}_k, \mathbf{x}_k(\mathbf{s}_k)) + \mathbb{E}\left\{ V_{k+1}^{\pi^*}(\mathbf{s}')|\mathbf{s}_k \right\} \right). \qquad (3)$$

The expression in (3) is typically computed using backward induction, a procedure called *value iteration*, and then an optimal policy is extracted from the value function by selecting a minimum value action for each state. However, the required computation grows exponentially with the size of the state, action and outcome spaces. One possible solution to overcome this problem is to approximate the value function.

III. Approximate dynamic programming (ADP)

ADP, also known as *forward DP*, is based on an algorithmic strategy that steps forward through time, compared to backward induction used in *value iteration*. As such, the optimisation problem in (1) is solved by approximating the optimal value function (3) while stepping forward in time, and policies in ADP are extracted from these VFAs [16].

Algorithm 1 : ADP using Temporal Difference Learning TD(1)

```
1: Initialize V̄_k^0,   k ∈ N,
2: Set r = 1 and k = 1,
3: Set s_1^i.
4: while r ≤ R do
5:     Choose a sample path ω^r.
6:     for k = 0, ..., N do
7:         Solve the deterministic problem (10).
8:         Find right and left marginal contributions (11).
9:         if k < N then
10:            Find the post decision state (5) and the next pre decision state (6).
11:        end if
12:    end for
13:    for k = N, ..., 0 do
14:        Calculate marginal values (12).
15:        Update the marginal values (13).
16:        Update the VFA using CAVE algorithm.
17:    end for
18:    r = r + 1.
19: end while
20: Return the value function approximations V̄_k^R   ∀k.
```

A. Policy-based value function approximation

VFAs are obtained iteratively and here the focus is on approximating the value function around a post decision state vector, $\mathbf{s}_k^x$, the state of the system at discrete time, k, soon after making the decisions but before the realisation of random variables [16]. This is because approximating the expectation within the max or min operator in (3) is difficult to do in large practical applications as we need to obtain transition probabilities from all the possible states. Given this the original transition function:

$$\mathbf{s}_{k+1} = \mathbf{s}^M\left(\mathbf{s}_k, \mathbf{x}_k, \boldsymbol{\omega}_{k+1}\right), \tag{4}$$

is divided into the post-decision state and the next pre-decision state:

$$\mathbf{s}_k^x = \mathbf{s}^{M,x}\left(\mathbf{s}_k, \mathbf{x}_k\right), \tag{5}$$

$$\mathbf{s}_{k+1} = \mathbf{s}^{M,\omega}\left(\mathbf{s}_k^x, \boldsymbol{\omega}_{k+1}\right), \tag{6}$$

which are used in line 10 of Algorithm 1. Note that the mean and variation of the stochastic variables are use to obtain the post-decision and next pre-decision states, respectively. The new form of the value function is written as:

$$\bar{V}_k^\pi(\mathbf{s}_k) = \min_{\mathbf{x}_k}\left(C_k(\mathbf{s}_k, \mathbf{x}_k) + \bar{V}_k^{\pi,x}(\mathbf{s}_k^x)\right), \tag{7}$$

where $\bar{V}_k^{\pi,x}(\mathbf{s}_k^x)$ is the value function approximation around the post-decision state $\mathbf{s}_k^x$, given by:

$$\bar{V}_k^{\pi,x}(\mathbf{s}_k^x) = \mathbb{E}\left\{V_{k+1}^\pi(\mathbf{s}_{k+1})|\mathbf{s}_k^x\right\}. \tag{8}$$

This method is computationally feasible because $\mathbb{E}\left\{V_{k+1}^\pi(\mathbf{s}_{k+1})|\mathbf{s}_k^x\right\}$ is a function of the post-decision state $\mathbf{s}_k^x$, which is a deterministic function of $\mathbf{x}_k$. However, in order to solve (7), we still need to calculate the value functions for every possible state $\mathbf{s}_k^x$ for all k. This can be computationally difficult since $\mathbf{s}_k^x$ is continuous and in general, multidimensional[2]. In order to overcome this, we construct a lookup table for VFAs that are concave and piecewise linear in the resource dimension of all the state variables. Accordingly, the VFA is given by:

$$\bar{V}_k(\mathbf{s}_k^x) = \sum_{a=1}^{A_k} \bar{v}_{ka} z_{ka}, \tag{9}$$

where $\sum_a z_{ka} = \mathbf{s}_k^x$ and $0 \le z_{ka} \le \bar{z}_{ka}$ for all a. z_{ka} is the resource coordinate variable for segment $a \in (1 \dots A_k)$, $A_k \in \mathcal{A}$, $\bar{z}_{ka}$ is the capacity of the segment and $\bar{v}_{ka}$ is the slope. Other strategies that could be used for this step are parametric and non-parametric approximations of the value functions [16].

Pseudo-code of the method used to approximate the value function is given in Algorithm 1, which is a double pass algorithm referred to as *temporal difference learning* with a discount factor $\lambda = 1$ or TD(1). It proceeds as follows:

1) Set the initial VFAs to zero (i.e. all the slopes to zero) or to an initial guess to speed up the convergence (lines 1-3). Estimates for the initial VFAs can be obtained by solving the deterministic problem using MILP. The value of the initial starting state s_1^i is assumed.
2) For each sample path, step forward in time by solving the deterministic problem at each time-step, using the VFA of the future state from the last iteration (line 7). The optimal action is found using:

$$\begin{aligned}\mathbf{x}_k^r =& \arg\min_{\mathbf{x}_k \in X_k}\left(C(\mathbf{s}_k^r, \mathbf{x}_k) + \bar{V}_k^{r-1}(\mathbf{s}_k^{x,r})\right), \\ =& \arg\min_{\mathbf{x}_k \in X_k}\left(C(\mathbf{s}_k^r, \mathbf{x}_k) + \sum_{a=1}^{A_k^{r-1}} \bar{v}_{ka}^{r-1} z_{ka}\right).\end{aligned} \tag{10}$$

3) Determine the positive and the negative marginal contributions $\hat{c}_k^{r+}(s_k^r)$ and $\hat{c}_k^{r-}(s_k^r)$, respectively (line 8), using:

$$\begin{aligned}\hat{c}_k^{r+}(s_k^r) &= \frac{c_k^{r+}(s_k^{r+}, x_k^{r+}) - c_k^r(s_k^r, x_k)}{\delta s}, \\ \hat{c}_k^{r-}(s_k^r) &= \frac{c_k^r(s_k^r, x_k) - c_k^{r-}(s_k^{r-}, x_k^{r-})}{\delta s},\end{aligned} \tag{11}$$

where $s_k^{r+} = (s_k^r + \delta s)$ and $x_k^{r+} = X_k^\pi\left(s_k^{r+}\right)$, similarly for s_k^{r-} and x_k^{r-}.
4) Find the post-decision and the next pre-decision states using (5) and (6), respectively. Transition functions of the storage devices can be non-linear (line 10).
5) Starting from N, step backward in time to compute the slopes, which are then used to update the VFA (line 14). Compute $\hat{v}_k^{r+}$:

$$\hat{v}_k^{r+}(s_k^r) = \begin{cases}\hat{c}_N^{r+}(s_N^r), & \text{if } k = N \\ \hat{c}_k^{r+}(s_k^r) + \Delta_k^{r+}\hat{v}_{k+1}^{r+}(s_{k+1}^r) & \text{otherwise}\end{cases}, \tag{12}$$

where $\Delta_k^{r+} = \frac{1}{\delta s} S^M(x_k^r - x_k^{r+})$, which is a variable that tells whether or not there is a change in energy in the storage as a result of the random perturbation. We do this similarly for $\hat{v}_k^{r-}(s_k^r)$.

[2] Note that we only consider one storage device in this paper.

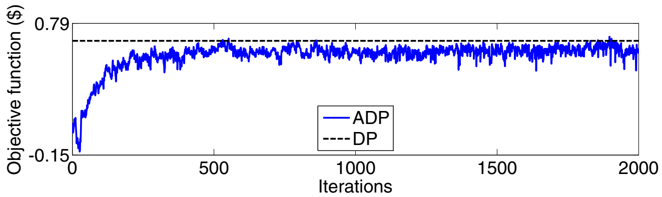

Fig. 2. Value of the objective function (i.e. reward) vs. iterations for ADP approach and the expected value from DP.

6) Obtain an approximation of the value function around a post-decision state [13]:

$$\bar{V}^{r+}_{k-1}(s^{x,r}_{k-1}) = \left(1 - \alpha^{r-1}\right) \bar{V}^{r-1+}_{k-1}(s^{x,r}_{k-1}) + \alpha^{r-1} \hat{v^{r+}_k}, \tag{13}$$

where α is a "stepsize"; $\alpha \in (0,1]$, and similarly for $\bar{V}^{r-}_{k-1}(s^{x,r}_{k-1})$ (line 15). In this research, a harmonic step-size formula is used, $\alpha = b/(b{+}r)$, where b is a constant. This step-size formula satisfies conditions ensuring that the values will converge as $r \to \infty$ [21].

7) The concave adaptive value estimation (CAVE) algorithm is use to update the VFAs [22] (line 16).

8) Repeat this procedure over R iterations. We choose $R = 1000$ random realisations for the proposed ADP approach as it is enough for the objective function to come within an acceptable accuracy even for the worst possible scenario. We investigated a range of scenarios and an example is depicted in Fig. 2.

To illustrate the performance of the ADP, we now use the three solution techniques to solve the energy management problem.

IV. Implementation

The DER in the residential building considered in this paper comprise a PV-battery system. A comparison is made of the battery SOC and the total electricity cost over a day for two different scenarios (i.e. different electrical demand patterns and PV output) using the three solution techniques. One-day and two-day decision horizons are divided into $N = 48$ and $N = 96$ time-steps, respectively, which gives a 30 minutes resolution. The decision horizon can be extended beyond one-day with ADP as it is computationally efficient (discussed in the Section V-B) so the ADP based HEMS is used over a two-day horizon (i.e. the total electricity cost for day one is calculated using uncertainties over the next day). The end-of-day one SOC from ADP is used as a constraint in DP and stochastic MILP approaches.

The implemented HEMSs operate according to the following:

- The battery is charged with the highest possible charge rate if the electrical power generated from the PV system exceeds the electrical demand of the household.
- Only send electrical power back to the grid if the electrical power from the PV system exceeds the highest possible charge rate of the battery, and the electrical demand of the household.

TABLE I
Daily optimisation results for the two scenarios.

	Scenario 1	Scenario 2
Total demand	11.92 kWh	23.45 kWh
Total PV generation	7.11 kWh	9.48 kWh
Cost - neither PV or storage	\$1.83	\$4.04
Cost - with PV but no storage	\$1.52	\$3.28
Cost - HEMS (DP)	\$0.85	\$2.27
Cost - HEMS (ADP)	\$0.91	\$2.35
Cost - HEMS (Stochastic MILP)	\$1.09	\$2.38
Cost - HEMS (DP 60%)	\$0.97	\$2.56

Now we present the system model and the contribution function used in the case study, and discuss computational aspects of the approaches.

A. System model

The stochastic energy management problem is formulated using:

- State variables to represent the battery SOC, s^{b}_k, mean PV output, s^{pv}_k, mean electrical demand, $s^{\mathrm{d,e}}_k$, and electricity tariff, s^{p}_k, for each time-step, k, in the decision horizon.
- A control variable to represent the charge and discharge rates of the battery, x^{b}_k, at a given time-step k.
- Transition functions governing how the state variables evolve over time. For example the battery SOC described by the state variable $s^{\mathrm{b}}_k \in \left\{s^{\mathrm{b,min}}, s^{\mathrm{b,max}}\right\}$ will progress by:

$$s^{\mathrm{b}}_{k+1} = \left(1 - l^{\mathrm{b}}(s^{\mathrm{b}}_k)\right)\left(s^{\mathrm{b}}_k - x^{\mathrm{b}^-}_k + \mu^{\mathrm{b}^+}(x^{\mathrm{b}^+}_k)\right), \tag{14}$$

where $\mu^{\mathrm{b}^+}_k(x^{\mathrm{b}^+}_k)$ is the efficiency of the charging process and $l^{\mathrm{b}}(s^{\mathrm{b}}_k)$ models the self-discharging process of the battery.
- Random variables capturing variations in PV output, ω^{pv}_k, and variations in electrical demand, $\omega^{\mathrm{d,e}}_k$. Transition probabilities associated with PV output and electrical demand are obtained as follows: (i) first we cluster historical data collected during the *Smart Grid Smart City* trial using k-means algorithm (i.e. three clusters each for the PV output and electrical demand), and (ii) for every time-step in each cluster we estimate transition probabilities using Epanechnikov kernel estimating technique.

Characteristics of the storage devices found in [7] indicate that DP and ADP can properly incorporate non-linear characteristics, while linear approximations have to be made with stochastic MILP. The maximum and minimum battery SOC are 2 kWh and 10 kWh, respectively. The charging efficiency of the battery is $\mu^{\mathrm{b+}}(x^{\mathrm{b+}}_k) = 1$.

B. Objective function

The objective of the implemented HEMSs is to minimise energy costs/maximise rewards over the decision horizon. The cost incurred at a given time-step, $C_k(\mathbf{s}_k, \mathbf{x}_k, \boldsymbol{\omega}_k)$, is given by:

$$C_k(\mathbf{s}_k, \mathbf{x}_k, \boldsymbol{\omega}_k) = s^{\mathrm{p}}_k \left(\omega^{\mathrm{d,e}}_k - \mu^{\mathrm{i}} x^{\mathrm{i}}_k\right), \tag{15}$$

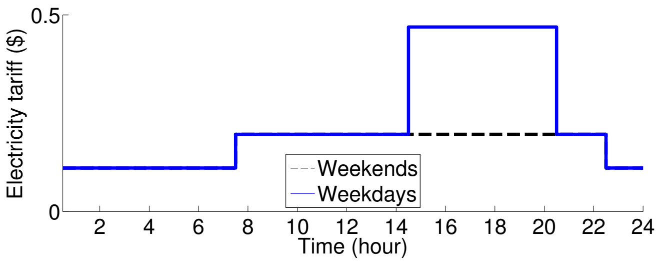

Fig. 3. Electricity tariff over a day for all the scenarios.

where $x_k^{\mathrm{i}} = \omega_k^{\mathrm{pv}} - \mu^{\mathrm{b}} x_k^{\mathrm{b}}$ is the output of the inverter.

In all the implemented HEMSs, the values of the control variable, x_k^{b}, at every time-step are determined depending on the state variables $\mathbf{s}_k = [s_k^{\mathrm{b}}, s_k^{\mathrm{pv}}, s_k^{\mathrm{d,e}}, s_k^{\mathrm{p}}]$ and realisations of random variables, $\boldsymbol{\omega}_k = [\omega_k^{\mathrm{pv}}, \omega_k^{\mathrm{d,e}}]$. The stochastic MILP approach uses 1200 scenarios [7].

V. Simulation results and discussion

Now we report and discuss the quality of the solutions and computational aspects of the three solution techniques. The daily electricity tariff used in all the simulations is given in Fig. 3 and we choose scenarios with different demand patterns and PV outputs from real data obtained during *Smart Grid Smart City* project [20], [23]. First we show the performance of the three HEMSs over a day (Section V-A) using two scenarios shown in Fig. 4:

1) Scenario 1 is on 1st of August, 2010, for a Central Coast, NSW, Australia, based residential building with a daily PV output of 7.11 kWh and a total demand of 11.92 kWh.
2) Scenario 2 is on 31st of December, 2010, for a residential building in Central Coast, NSW, Australia, with a total daily PV output of 9.48 kWh and a demand of 23.45 kWh.

Second we discuss the computational performance of the three solution techniques (Section V-B).

Before showing the quality of the solutions from the three solution techniques, we highlight the significant benefits of residential PV-storage systems with a DP based HEMS ($s_1^{\mathrm{b}} = s_N^{\mathrm{b}} = 6$ kWh) in Table 1. For the sake of identifying benefits of residential PV-battery systems, we decided to use half of the available battery capacity as the start-of-day and end-of-day battery SOC ($s_1^{\mathrm{b}} = s_N^{\mathrm{b}} = 6$ kWh) constraint in DP. On days with a low morning demand, high PV output and medium-high evening demand, $s_1^{\mathrm{b}} = s_N^{\mathrm{b}} = 2$ kWh gives the best results because the battery can be used to supply the evening demand and there is no need to charge it back. However, next days electricity cost can significantly increase if we are anticipating a high morning demand and low PV output. In such situations, it is beneficial to control the end-of-day battery SOC to account for uncertainties over several days, which can be done using ADP.

A. Quality of the solutions from ADP, DP and stochastic MILP

ADP and DP results in better quality schedules than stochastic MILP, which is evident from the total electricity cost of the two scenarios given in Table 1. This is because DP and ADP enable us to incorporate stochastic nature of the input variables using appropriate probabilistic models, and non-linear constraints. However, ADP results in lower quality schedules than DP over a day because the value functions used are approximations. Note that DP produces close to optimal solutions when finely discretised; however, we have only considered a PV-storage system, and adding more storage devices will make the DP based HEMS extremely difficult to compute even over a daily horizon.

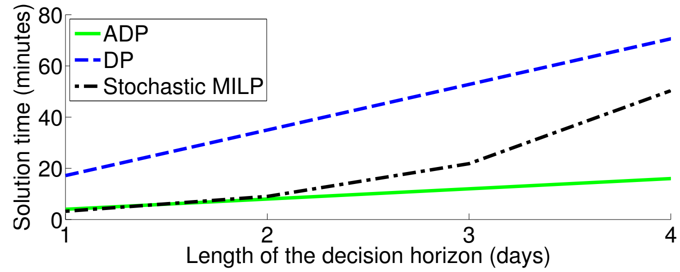

Fig. 4. Computational time of ADP, DP and stochastic MILP against the length of the decision horizon.

B. Computational aspects

ADP is computationally efficient compared to DP and stochastic MILP, as shown in Fig. 5. A daily and a weekly optimisation for a residential PV-battery system takes approximately 4 and 28 minutes, respectively, using ADP while DP approach takes approximately 17.1 minutes and 124 minutes, respectively. The computational time of both HEMSs using ADP and DP increases linearly as we increase the decision horizon, however, ADP has a lesser slope. This linear increase with DP is because the state transitions in this problem are only between two adjacent time-steps, so time does not contribute to an exponential increase in state space size. However, the computational time of DP will increase exponentially when more storage devices are added, where's ADP will have only a linear increase because the state space is factorised. The solution time of stochastic MILP grows exponentially with the length of the horizon because the optimisation problem is solved for the whole horizon at once and the number of possible scenarios increases as the number of time-steps increases [7]. The computational burden can be improved at the expense of solution quality by using scenario reduction techniques.

In summary, we can see that ADP is computationally efficient with good quality solutions. Moreover, we can extend the decision horizon beyond a day and add more storage devices with less computational burden than other approaches. Our preliminary simulation results indicate that the average yearly electricity cost can be reduced by up to 12% with a two-day decision horizon.

VI. Conclusion

This paper presented an ADP approach for implementing a computationally efficient HEMS with good quality schedules.

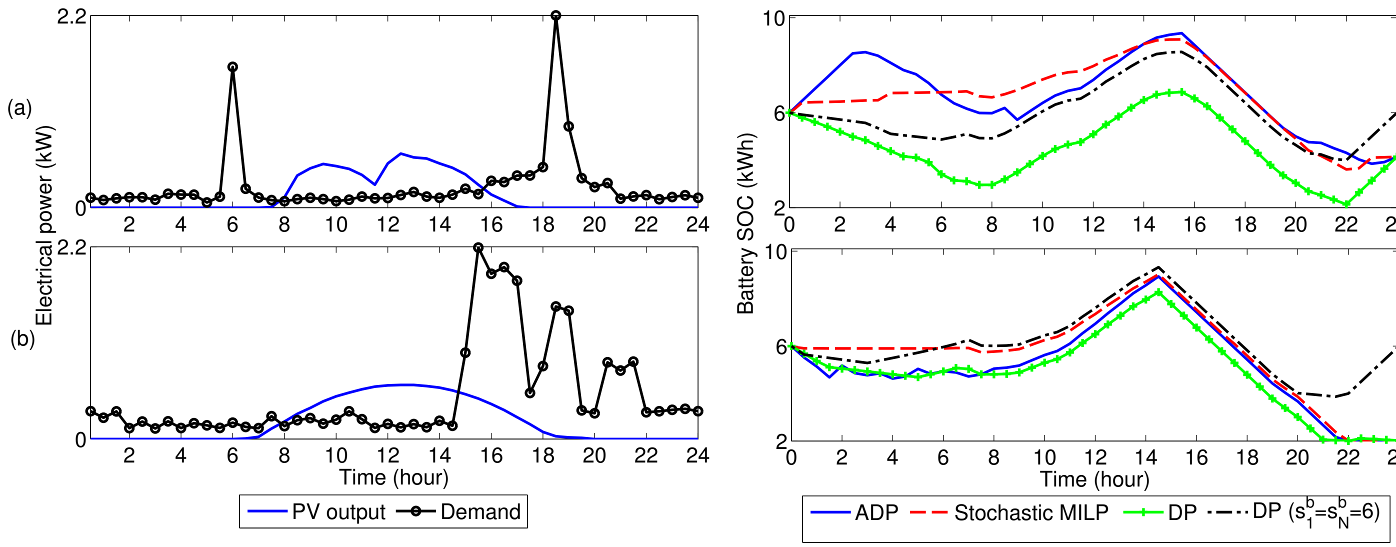

Fig. 5. PV output, electrical demand and battery SOC using HEMSs based on ADP, DP and stochastic MILP for (a) Scenario 1 and (b) Scenario 2. ADP is used over a two-day horizon and its end-of-day one SOC is used in DP and stochastic MILP approaches over a day. This gives a valid comparison among the three solution techniques as the start-of-day and end-of-day SOC is the same. Additionally, DP is used over a day with 60% start-of-day and end-of-day SOC in each scenarios to identify the benefits of PV-storage systems with a HEMS.

In the future, we will evaluate ADP over a year using an extended decision horizon with a high resolution while considering multiple storage devices. Our preliminary results indicate that these improvements should provide financial benefits to households employing them in a HEMS. Moreover, we will incorporate learning methods to our model in order to make better VFAs, which will result in further improvements to the quality of the solutions.

References

[1] Australian PV Institute. Australian PV market since April 2001. [Online]. Available: http://pv-map.apvi.org.au/analyses

[2] Electric Power Research Institute (EPRI), "The integrated grid realizing the full value of central and distributed energy resources," Tech. Rep., 2014.

[3] C. Cooper. (2015, May) What the Tesla Powerwall battery means for households. RenewEconomy. Neutral Bay, NSW, 2089, Australia. [Online]. Available: http://reneweconomy.com.au/2015/what-the-tesla-powerwall-battery-means-for-households-61055

[4] C. T. E. Rocky Mountain Institute, Homer Energy. The economics of grid defection when and where distributed solar generation plus storage competes with traditional utility service.

[5] Z. Chen, L. Wu, and Y. Fu, "Real-time price-based demand response management for residential appliances via stochastic optimization and robust optimization," *Smart Grid, IEEE Transactions on*, vol. 3, no. 4, pp. 1822–1831, 2012.

[6] C. Keerthisinghe, G. Verbic, and A. Chapman, "Evaluation of a multi-stage stochastic optimisation framework for energy management of residential PV-storage systems," in *Power Engineering Conference (AUPEC), 2014 Australasian Universities*, Sept 2014, pp. 1–6.

[7] ——, "Addressing the stochastic nature of energy management in smart homes," in *Power Systems Computation Conference (PSCC), 2014*, Aug 2014, pp. 1–7.

[8] M. Bozchalui, S. Hashmi, H. Hassen, C. Canizares, and K. Bhattacharya, "Optimal operation of residential energy hubs in smart grids," *Smart Grid, IEEE Transactions on*, vol. 3, no. 4, pp. 1755–1766, 2012.

[9] M. Pedrasa, E. Spooner, and I. MacGill, "Robust scheduling of residential distributed energy resources using a novel energy service decision-support tool," in *Innovative Smart Grid Technologies (ISGT), 2011 IEEE PES*, 2011, pp. 1–8.

[10] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Athena Scientific, Bemont, Massachuttes, 2005, vol. 1.

[11] H. Tischer and G. Verbič, "Towards a smart home energy management system - a dynamic programming approach," in *Innovative Smart Grid Technologies Asia (ISGT), 2011 IEEE PES*, 2011, pp. 1–7.

[12] D. F. Salas and W. B. Powell, "Benchmarking a scalable approximate dynamic programming algorithm for stochastic control of multidimensional energy storage problems," Technical report, Working Paper, Department of Operations Research and Financial Engineering, Princeton, NJ, Tech. Rep., 2013.

[13] R. Anderson, A. Boulanger, W. Powell, and W. Scott, "Adaptive stochastic control for the smart grid," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 1098–1115, 2011.

[14] W. B. Powell, "Approximate dynamic programming for large-scale resource allocation problems," Princeton University, Princeton, New Jersey 08544, USA,, 2005.

[15] W. B. Powell, A. George, H. Simłło, W. Scott, A. Lamont, and J. Stewart, "SMART: A stochastic multiscale model for the analysis of energy resources, technology, and policy," *INFORMS Journal on Computing*, vol. 24, no. 4, pp. 665–682, 2012.

[16] W. B. Powell, *Approximate Dynamic Programming - Solving the Curses of Dimensionality*. John Wiley & Sons, Inc., 2007.

[17] D. Jiang, T. Pham, W. Powell, D. Salas, and W. Scott, "A comparison of approximate dynamic programming techniques on benchmark energy storage problems: Does anything work?" in *Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), 2014 IEEE Symposium on*, Dec 2014, pp. 1–8.

[18] F. Borghesan, R. Vignali, L. Piroddi, M. Prandini, and M. Strelec, "Approximate dynamic programming-based control of a building cooling system with thermal storage," in *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*, Oct 2013, pp. 1–5.

[19] M. Strelec and J. Berka, "Microgrid energy management based on approximate dynamic programming," in *Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES*, Oct 2013, pp. 1–5.

[20] Smart-Grid Smart-City Customer Trial Data. [Online]. Available: https://data.gov.au/dataset/smart-grid-smart-city-customer-trial-data

[21] T. Jaakkola, M. I. Jordan, and S. P. Singh, "Convergence of stochastic iterative dynamic programming algorithms," *Neural Computation*, vol. 6, pp. 1185–1201, 1994.

[22] G. A. Godfrey and W. B. Powell, "An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution," *Management Science*, vol. 47, no. 8, pp. 1101–1112, 2001.

[23] E. L. Ratnam, S. R. Weller, C. M. Kellett, and A. T. Murray, "Residential load and rooftop PV generation: an Australian distribution network dataset," *International Journal of Sustainable Energy*, vol. 0, no. 0, pp. 1–20, 0.