Extensions of the Dynamic Programming Framework: Battery Scheduling, Demand Charges, and Renewable Integration

Morgan Jones, Matthew M. Peet

***Abstract*—We consider a general class of Dynamic Programming (DP) problems with non-separable objective functions. We show that for any problem in this class, there exists an augmented-state DP problem which satisfies the Principle of Optimality and the solutions to which yield solutions to the original problem. Furthermore, we identify a subclass of DP problems with Naturally Forward Separable (NFS) objective functions for which this state-augmentation scheme is tractable. We extend this framework to stochastic DP problems, proposing a suitable definition of the Principle of Optimality. We then apply the resulting algorithms to the problem of optimal battery scheduling with demand charges using a data-based stochastic model for electricity usage and solar generation by the consumer.**

I. Introduction

The optimal use of battery storage can be formulated as a discrete time process combined with decision variables and an objective function - a formulation commonly known as Dynamic Programming (DP) [1]. DP is a class of algorithms that break down complex optimization problems into simpler sequential subproblems, each of which is solved using Bellman's Equation. For DP to work, however, we require that the optimization problem satisfies the *Principle of Optimality* [2] (a.k.a time-consistency [3], [4]); from any point on an optimal trajectory, the remaining portion of the optimal trajectory is also optimal for the problem initiated at that point [5]. DP problems commonly have an additively separable objective function of the form $J(\mathbf{u},\mathbf{x}) = \sum_{t=0}^{T-1} c_t(x(t),u(t)) + c_T(x(T))$. Problems of this form can be shown to satisfy the Principle of Optimality. However in problems such as optimal battery scheduling, we find non-additively separable objective functions. For example, if the objective is of the form $J(\mathbf{u},\mathbf{x}) = \sum_{t=0}^{T-1} c_t(x(t),u(t)) + \max_{t_0 \le k \le T} d_k(x(k))$ then the problem does not satisfy the Principle of Optimality. In this paper we propose a method for solving DP problems with non-separable objective functions by constructing equivalent DP problems with additively separable objective functions. Such reformulated problems then satisfy the Principle of Optimality and can therefore be solved using Bellman's Equation. Moreover, we identify a class of problems, defined by Naturally Forward Separable (NFS) objectived, wherein state augmentation method does not substantially increase the complexity of the problem.

For stochastic DP we generalize our framework and propose a suitable definition of the *Principle of Optimality*. As discussed in [6] such a definition is non trivial. Inspired by [7], we construct probability measures on the sets the state variable can take at each time stage induced by the underlying random variables. We then say a stochastic problem satisfies the *Principle of Optimality* if from any point on a trajectory followed using an optimal policy, π, the policy π is also optimal for the problem initiated from that point with probability one.

Dynamic programming for problems which do not satisfy the Principle of Optimality has received relatively little attention. The only general approach to the problem seems to be that taken in [8] which considered the use of multi-objective optimization in the case where the objective function is "backward separable". Our approach differs from [8] as we consider the class of "forward separable" objective functions. In this paper we show that almost any objective function is forward separable in a certain sense and that for such problems there exists an additively separable augmented-state dynamic programming problem that satisfies the Principle of Optimality and from which solutions to the original forward separable problem can be recovered - See Section III. However, the resulting augmented-state DP problem has a higher dimensional state space than the original DP problem - an issue that can potentially render the augmented problem intractable due to the "curse of dimensionality". For this reason, we propose a complexity metric for the forward separable representation and show that in certain cases the dimensionality of the augmented system does not significantly exceed the dimensionality of the original problem - a case where Bellman's equation can be used effectively [9] and which we refer to as *Naturally Forward Separable* (NFS).

Note that although state augmentation has been used in the context of DP [10], [11], the only tractable use of state augmentation to recover the Principle of Optimality appears to be [12] and [13], who considered a DP problem with objective function of the form $J(\mathbf{u},\mathbf{x}) = U(\sum_{t=0}^{T-1} c_t(x(t),u(t)))$; and [14], who considered a DP problem with variance-type objective function. Both these results can be considered special cases of the NFS class of objective functions proposed in this paper.

In practice it is rare to be able to analytically solve Bellman's Equation. Therefore, once the augmented-state DP problem is formulated we propose a map to an approximated DP problem that can be analytically solved. Using an optimal solution from the approximated DP problem a feasible policy for the original problem is then constructed using a discretization scheme based on [15] and [16]. We show that as the number of discrete points increases, the resulting policies converge to the optimal policy.

To illustrate the proposed methods, we consider battery scheduling for mitigating the effect of variability in renewable energy resources. Specifically, renewable energy sources are

M. Jones is with the School for the Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, 85298 USA. e-mail: morgan.c.jones@asu.edu

M. Peet is with the School for the Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, 85298 USA. e-mail: mpeet@asu.edu

most accurately modeled as an uncontrollable Gauss-Markov (G-M) process and the battery (for both consumers and utilities) attempts to minimize energy costs based on time-of-use while also minimizing the maximum rate of energy consumption. Based on this model, we formulate the battery storage problem as a DP with a non-separable objective function consisting of both integrated time-of-use charges and a maximum term representing the demand charge. Furthermore, we propose a model of solar generation as a G-M process and minimize the expected value of the proposed objective. The fundamental mathematical challenge with dynamic programming problems of this form is that, as shown in Section II, problems which include maximum terms in the objective do not satisfy the *Principle of Optimality* and thus the recursive solution of the Bellman equation ([1]) does not yield an optimal policy. To overcome this difficulty, we show that the battery scheduling problem is a special case of a forward separable DP problem with an NFS objective function. We then apply our state-augmentation technique to numerically solve the deterministic battery scheduling problem for given forecast solar data. In section IX-E we apply our approach to the battery scheduling problem using a Gauss-Markov model of solar generation extracted from data provided by local utility SRP. Note that this result extends previous work which considered a more limited non-separable DP framework as applied to battery scheduling in [17].

Remarkably, almost no work has been done on optimal use of batteries for reduction of demand charges. The exceptions include the heuristic algorithms of [18] and the pioneering work of [19] , which considered *only* a demand charge. Recently this group used an ad-hoc algorithm to consider a combined demand/consumption charge in [20] using detailed models of cooling/load. Furthermore, in [21] a similar energy storage problem is solved using optimized curtailment and load shedding. An L_p approximation of the demand charge was used in combination with multi-objective optimization in [22] and, in addition, the optimal use of building mass for energy storage was considered in [23], wherein a bisection on the demand charges was used. We note that none of these approaches resolve the fundamental mathematical problem of DP with a non-separable cost function.

The paper is organized as follows. In Section II we introduce a formal definition of the DP problem along with associated notation and use this framework to define a *Principle of Optimality*. Next, we consider a class of objective functions we refer to as forward separable. In Section III, we show that for any DP problem with forward separable objective, there exists an augmented-state DP problem with separable objective for which the Principle of Optimality holds and from which solutions to the original DP with FS objective can be recovered. In Section IV we define a class of objective functions, termed Naturally Forward Separable (NFS). We show DP problems with naturally forward separable objective functions can be tractable solved using state augmentation. In Section V we show how to approximate and numerically solve augmented-state dynamic programming problems. Furthermore, we extend our framework to stochastic DP problems in Section VI and give a discretization scheme to solve stochastic DP's with additively separable objective functions in Section VII. We summarize how state augmentation can be used with discretization methods to solve DP problems with NFS objective functions in Section VIII. In Section IX we formulate and solve the battery scheduling problem as a DP with NFS objective function.

II. Background: Dynamic Programming

In this paper, we propose a framework for representing a general class of Dynamic Programming (DP) problems. Specifically, we define a general DP problem as a sequence of optimization problems $\mathscr{G}(t_0,x_0)$, indexed by $t_0 \in \mathbb{N}$, and defined by an indexed sequence of objective functions $J_{t_0} : \mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T-t_0+1)} \to \mathbb{R}$ where we say that the state and input sequence, $\mathbf{u}^* \in \mathbb{R}^{m\times(T-t_0)}$ and $\mathbf{x}^* \in \mathbb{R}^{n\times(T-t_0+1)}$, solve $\mathscr{G}(t_0,x_0)$ if,

$$(\mathbf{u}^*,\mathbf{x}^*) \in \arg\min_{\mathbf{u},\mathbf{x}} J_{t_0}(\mathbf{u},\mathbf{x}) \tag{1}$$

$$\begin{aligned} &\text{subject to:}\\ &x(t+1) = f(x(t),u(t),t) \text{ for } t=t_0,..,T,\\ &x(t_0)=x_0,\ x(t)\in X_t \subset \mathbb{R}^n \text{ for } t=t_0,..,T,\\ &u(t)\in U\subset \mathbb{R}^m \text{ for } t=t_0,..,T-1,\\ &\mathbf{u}=(u(t_0),...,u(T-1)) \text{ and } \mathbf{x}=(x(t_0),...,x(T)),\end{aligned}$$

where $f:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{N}\to\mathbb{R}^n$, $U\subset\mathbb{R}^m$ and $X_t\subset\mathbb{R}^n$ for all t. We denote $J^*_{t_0} = J_{t_0}(\mathbf{u}^*,\mathbf{x}^*)$.

We call $\{x(t)\}_{t_0\le t\le T}$ the set of state variables and n the state space dimension. Similarly we will call $\{u(t)\}_{t_0\le t\le T-1}$ the input (control) variables and $m=\dim(U)$ the input (control) space dimension. For cases where the dimension of the state variable, $x(t)$, varies with time, we slightly abuse notation and define the state space dimension as $\max_{t_0\le t\le T}\dim(X_t)$.

Definition 1. *The function* $J_{t_0} : \mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T-t_0+1)} \to \mathbb{R}$ *is said to be additively separable if there exist functions,* $c_T(x):\mathbb{R}^n\to\mathbb{R}$*, and* $c_t(x,u):\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ *for* $t=t_0,\cdots T-1$ *such that,*

$$J_{t_0}(\mathbf{u},\mathbf{x}) = \sum_{t=t_0}^{T-1} c_t(x(t),u(t)) + c_T(x(T)), \tag{2}$$

where $\mathbf{u}=(u(t_0),...,u(T-1))$ *and* $\mathbf{x}=(x(t_0),...,x(T))$.

To illustrate, we note that the average value of a function $a_t:\mathbb{R}^n\to\mathbb{R}$, defined as $J(\mathbf{u},\mathbf{x}) = \frac{1}{T}\sum_{t=0}^{T} a_t(x(t))$, is clearly an additively separable function. Variance type functions (11), however, are not additively separable.

Definition 2. *We say the sequence of inputs* $\mathbf{u} = (u(t_0),...,u(T-1)) \in \mathbb{R}^{m\times(T-t_0)}$ *is feasible if* $u(t)\in U$ *for* $t=t_0,..,T-1$ *and for* $x(t+1)=f(x(t),u(t),t)$ *and* $x(t_0)=x_0$*, then* $x(t)\in X$ *for all* t*. For a given* x*, we denote by* $\Gamma_{t,x}$*, the set* $u\in U$ *such that* $f(x,u,t)\in X_{t+1}$*. In this paper we only consider problems where* $\Gamma_{t,x}$ *is nonempty for all* x *and* t*.*

Note that for this class of DP problems, feasibility is inherited. That is, if $\mathbf{u}=(u(t),....,u(T-1))$ is feasible with $\mathbf{x}=(x(t),\cdots,x(T))$ for $\mathscr{G}(t,x(t))$ and $\mathbf{v}=(v(s),....,v(T-1))$ is feasible with $\mathbf{h}=(h(s),\cdots,h(T))$ for $\mathscr{G}(s,x(s))$ where

$s>t$, then $\mathbf{w}=(u(t),\cdots,u(s-1),v(s),....,v(T-1))$ with $\mathbf{z}=(x(t),\cdots,x(s-1),h(s),\cdots,h(T))$ is feasible for $\mathscr{G}(t,x(t))$.

Definition 3. *A (Markov) policy is any map from the present state and time to a feasible input* $(x,t)\mapsto u(t)\in\Gamma_{x,t}$, *as* $u(t)=\pi(x,t)$. *We denote the set of policies consistent with some DP problem as* Π. *We say that* π^* *is an optimal policy for Problem* (1) *if*

$$\mathbf{u}^*=(\pi^*(x_0,t_0),....,\pi^*(x(T-1),T-1))$$

where $x(t+1)^*=f(x(t)^*,\pi^*(x(t)^*,t),t)$ *for all* t.

We now define a "Principle of Optimality" consistent with our DP formulation and which provides a necessary condition for such DP problems to be solvable using Bellman's equation (4). If a DP problem is solvable using Bellman's equation, then this equation yields an optimal policy.

Definition 4. *We say a DP problem,* $\mathscr{G}(t_0,x_0)$*, of the Form (1) satisfies the Principle of Optimality if the following holds. For any* s *and* t *with* $t_0\le t<s<T$, *if* $\mathbf{u}^*=(u(t),...,u(T-1))$ *and* $\mathbf{x}^*=(x(t),...,x(T))$ *solve* $\mathscr{G}(t,x(t))$ *then* $\mathbf{v}=(u(s),...,u(T-1))$ *and* $\mathbf{h}=(x(s),...,x(T))$ *solve* $\mathscr{G}(s,x(s))$.

The standard form of DP, equivalent to that defined in [1], solves indexed DP problems of the Form (1) with an additively separable objective function. We denote this class of DP problems by $\mathscr{P}(t_0,x_0)$:

$$\min_{\mathbf{u},\mathbf{x}} J_{t_0}(\mathbf{u},\mathbf{x})=\sum_{t=t_0}^{T-1}c_t(x(t),u(t))+c_T(x(T)) \tag{3}$$

subject to:

$$\begin{aligned}&x(t+1)=f(x(t),u(t),t)\ \text{for}\ t=t_0,..,T,\\ &x(t_0)=x_0,\ x(t)\in X_t\subset\mathbb{R}^n\ \text{for}\ t=t_0,..,T,\\ &u(t)\in U\subset\mathbb{R}^m\ \text{for}\ t=t_0,..,T-1,\\ &\mathbf{u}=(u(t_0),...,u(T-1))\ \text{and}\ \mathbf{x}=(x(t_0),...,x(T)).\end{aligned}$$

Note that $J_T(x)=c_T(x)$. We will refer to $x_0\in\mathbb{R}^n$ as the initial state, J_{t_0} is the objective function, $c_t:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ for $t=t_0,..,T-1$, $c_T\ \mathbb{R}^n\to\mathbb{R}$ are given functions and $f:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{N}\to\mathbb{R}^n$ is a given vector field. The following lemma shows that this class of problems satisfies the proposed Principle of Optimality.

Lemma 5. *Any problem of form* $\mathscr{P}(t_0,x_0)$ *in* (3) *satisfies the Principle of Optimality.*

Proof. Suppose $\mathbf{u}^*=(u(t),...,u(T-1))$ and $\mathbf{x}^*=(x(t),...,x(T))$ solve $\mathscr{P}(t,x(t))$ in (2). Now we suppose by contradiction that there exists some $s>t$ such that $\mathbf{v}=(u(s),...,u(T-1))$ and $\mathbf{h}=(x(s),...,x(T))$ do not solve $\mathscr{P}(s,x(s))$. We will show that this implies that $\mathbf{u}^*$ and $\mathbf{x}^*$ do not solve $\mathscr{P}(t,x)$ in (2), thus verifying the conditions of the Principle of Optimality. If $\mathbf{v}$ and $\mathbf{h}$ do not solve $\mathscr{P}(s,x(s))$, then there exist feasible $\mathbf{w}$, $\mathbf{z}$ such that $J_s(\mathbf{w},\mathbf{z})<J_s(\mathbf{v},\mathbf{h})$. i.e.

$$\begin{aligned}J_s(\mathbf{w},\mathbf{z})&=\sum_{t=s}^{T-1}c_t(z(t),w(t))+c_T(z(T))\\ &<\sum_{t=s}^{T-1}c_t(x(t),u(t))+c_T(x(T))=J_s(\mathbf{v},\mathbf{h})\end{aligned}$$

Now consider the proposed feasible sequences $\hat{\mathbf{u}}=(u(t),...,u(s-1),w(s),...,w(T-1))$ and $\hat{\mathbf{x}}=(x(t),...,x(s-1),z(s),...,z(T-1))$. It follows:

$$\begin{aligned}&J_t(\hat{\mathbf{u}},\hat{\mathbf{x}})\\ &=\sum_{k=t}^{s-1}c_k(x(k),u(k))+\sum_{k=s}^{T-1}c_k(z(k),w(k))+c_T(z(T))\\ &<\sum_{k=t}^{s-1}c_k(x(k),u(k))+\sum_{k=s}^{T-1}c_k(x(k),u(k))+c_T(x(T))\\ &=J_t(\mathbf{u}^*,\mathbf{x}^*)\end{aligned}$$

which contradicts optimality of $\mathbf{u}^*,\mathbf{x}^*$. Therefore, this class of problems satisfies the Principle of Optimality. □

Proposition 6 ([24])**.** *For DP problems of the form* $\mathscr{P}(t,x)$ *in* (3) *with optimal objective,* $J_t^*=J_t(\mathbf{u}^*,\mathbf{x}^*)$*, define the function* $F(x,t)=J_t^*$*. Then the following holds.*

$$\begin{aligned}F(x,t)&=\inf_{u\in\Gamma_{t,x}}\{c_t(x,u)+F(f(x,u,t),t+1)\} \qquad (4)\\ &\qquad\forall x\in X_t\ \text{and}\ \forall t\in\{t_0,..,T-1\},\\ F(x,T)&=c_T(x)\qquad\forall x\in X_T\end{aligned}$$

Equation (4) is often referred to as Bellman's equation and a function F which satisfies Bellman's equation is often referred to as the "optimal cost-to-go" function. Prop. 6 shows that problems of the Form $\mathscr{P}(t_0,x_0)$ define a solution to Bellman's equation which in turn indexes the optimal objective to the problem. Furthermore, for problems $\mathscr{P}(t_0,x_0)$, the solution to Bellman's equation can be obtained recursively backwards in time using a minimization on u. A solution to Bellman's equation provides a state-feedback law or optimal policy as follows.

Corollary 7 ([24])**.** *Consider* $\mathscr{P}(t_0,x_0)$ *in* (3)*. Suppose* $F(x,t)$ *satisfies Equation* (4) *for* $\mathscr{P}(t_0,x_0)$*. Then if there exists a policy such that,*

$$\theta(x,t)\in\arg\min_{u\in\Gamma_{t,x}}\{c_t(x,u)+F(f(x,u,t),t+1)\},$$

then θ *is an optimal policy for the problem* $\mathscr{P}(t_0,x_0)$.

Dynamic Programming with Maximum Terms In this paper we consider the special class of indexed DP problems, denoted by $\mathscr{S}(t_0,x_0)$ and given in (5). In contrast to problems of the form $\mathscr{P}(t_0,x_0)$ in (1), class $\mathscr{S}(t_0,x_0)$ has maximum terms in the objective. Specifically, these problems have the following form:

$$\min_{\mathbf{u},\mathbf{x}} J_{t_0}(\mathbf{u},\mathbf{x}):=\sum_{t=t_0}^{T-1}c_t(x(t),u(t))+c_T(x(T))+\max_{t_0\le k\le T}d_k(x(k))$$

subject to: (5)

$$\begin{aligned}&x(t+1)=f(x(t),u(t),t)\ \text{for}\ t=t_0,..,T,\\ &x(t_0)=x_0,\ x(t)\in X_t\subset\mathbb{R}^n\ \text{for}\ t=t_0,..,T,\\ &u(t)\in U\subset\mathbb{R}^m\ \text{for}\ t=t_0,..,T-1,\\ &\mathbf{u}=(u(t_0),...,u(T-1))\ \text{and}\ \mathbf{x}=(x(t_0),...,x(T)),\end{aligned}$$

where $c_T(x):\mathbb{R}^n\to\mathbb{R}$; $c_t(x,u):\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ for $t_0\le t\le T-1$; $d_t(x):\mathbb{R}^n\to\mathbb{R}$ for $t=t_0,\cdots T$; $f:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{N}\to\mathbb{R}^n$.

Table I
THIS TABLE SHOWS THE CORRESPONDING COST OF EACH FEASIBLE POLICY USED IN THE COUNTER EXAMPLE IN LEMMA 1

feasible $\mathbf{u}$	objective value	feasible $\mathbf{u}$	objective value
$(0,0,0)$	0	$(h,0,-h)$	h/2
$(0,0,h)$	h/2	$(h,0,0)$	0
$(0,h,0)$	2h	$(h,-h,0)$	-h
$(0,h,-h)$	(5/2)h	$(h,-h,h)$	-(3/2)h

Counterexample 8. *The class of DP problems of the form $\mathscr{S}(t_0,x_0)$ in (5) does not satisfy the Principle of Optimality.*

Proof. We give a counterexample. For $h>0$, we consider the following problem $\mathscr{S}(0,0)$:

$$\begin{aligned}&\min_{\mathbf{u}\in\mathbb{R}^3,\mathbf{x}\in\mathbb{R}^4} \quad \sum_{t=0}^{2} c_t(u(t)) + \max_{0\le k\le 3} x(k)\\ &\text{subject to: } x(t+1)=x(t)+u(t),\\ &x(0)=0,\ 0\le x_t\le h,\ u(t)\in\{-h,0,h\},\end{aligned}$$

where here we define $c_0(u(0)) = -u(0)$, $c_1(u(1)) = u(1)$, $c_2(u(2)) = -u(2)/2$.
Since $\mathbf{u}\in\{-h,0,h\}^3$, there are 27 input sequences, only 8 of which are feasible. In Table I, we calculate the objective value of each feasible input sequence and deduce the optimal input is $\mathbf{u}^*=(h,-h,h)$, yielding an optimal trajectory of $\mathbf{x}^*=\{0,h,0,h\}$. Following this input sequence until $t=2$ we examine the problem $\mathscr{S}(2,0)$.

$$\begin{aligned}&\min_{u(2)\in\mathbb{R},0\le x(3)\le h} c_2(u(2)) + \max_{2\le k\le 3} x(k)\\ &\text{subject to: } x(t+1)=x(t)+u(t),\\ &x(2)=0,\ 0\le x(t)\le h,\ u(t)\in\{-h,0,h\}.\end{aligned}$$

For this sub-problem, there are two feasible inputs: $u(2)\in\{0,h\}$. Of these, the first is optimal (objective value 0 vs $h/2$). Thus, although $\mathbf{u}^*=\{h,-h,h\}$ and $\mathbf{x}^*=\{0,h,0,h\}$ solve $\mathscr{S}(0,0)$, $\mathbf{v}=\{h\}$ and $\mathbf{h}=\{0,h\}$ do not solve $\mathscr{S}(2,0)$. ☐

III. CONVERTING FORWARD SEPARABLE DP TO ADDITIVELY SEPARABLE DP

In this section we define the class of forward separable objective functions. We will show that for dynamic programming problems with a forward separable objective function, augmenting the state variables allows us to use Bellman's equation to obtain an optimal policy.

Forward separable functions were first defined in [25]. Intuitively, this is the class of functions that can be separated into a nested composition of maps ordered forward in time. In the next definition we build upon the concept of forward separability by introducing the notion of augmented dimension.

Definition 9. *The function $J:\mathbb{R}^{m\times(T-t_0)}\times\mathbb{R}^{n\times(T+1-t_0)}\to\mathbb{R}$ is said to be forward separable if there exist representation maps $\phi_{t_0}:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^{d_{t_0}}$, $\phi_T:\mathbb{R}^n\times\mathbb{R}^{d_{T-1}}\to\mathbb{R}$, and $\phi_i:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^{d_{i-1}}\to\mathbb{R}^{d_i}$ for $i=t_0+1,\cdots T-1$ such that*

$$\begin{aligned}J(\mathbf{u},\mathbf{x}) = \phi_T(x(T),\phi_{T-1}(x(T-1),u(T-1),\phi_{T-2}(....,\\ \phi_{t_0+1}(x(t_0+1),u(t_0+1),\phi_{t_0}(x(t_0),u(t_0)))),....,))),\end{aligned} \tag{6}$$

where $\mathbf{u}=(u(t_0),...,u(T-1))\in\mathbb{R}^{m\times(T-t_0)}$ and $u(i)\in\mathbb{R}^m$ for $i\in\{t_0,...,T-1\}$; $\mathbf{x}=(x(t_0),...,x(T))\in\mathbb{R}^{n\times(T+1-t_0)}$ and $x(i)\in\mathbb{R}^n$ for $i\in\{t_0,...,T\}$; $d_i\in\mathbb{N}$ for $i\in\{t_0,...,T-1\}$.

Moreover we say $J(\mathbf{u},\mathbf{x})$ is forward separable and has a representation dimension of l if there exists $\{\phi_i\}$ that satisfies (6) and $l=\max_{i\in\{t_0,...,T-1\}}\{d_i\}$ where $d_i=\dim(Im\{\phi_i\})$.

Note: The representation dimension of a forward separable function is a property of the set $\{\phi_i\}$ chosen and not the function. The representation dimension of a forward separable function is not unique. Moreover, the forward separable property of an objective function is independent of the DP problem it is associated with; forward separability is solely a property of the function only.

Clearly, any additively separable objective function of the form $J(\mathbf{u},\mathbf{x})=\sum_{t=t_0}^{T-1}c_t(u(t),x(t))+c_T(x(T))$ is forward separable and has a representation dimension of 1 using,

$$\begin{aligned}&\phi_{t_0}(x,u)=c_{t_0}(x,u)\\ &\phi_i(x,u,w)=c_i(x,u)+w \quad \text{for } i=t_0+1,\cdots,T-1\\ &\phi_T(x,w)=c_T(x)+w.\end{aligned} \tag{7}$$

A. How State Augmentation Solves Forward Separable DP Problems

We now define the class of forward separable problems $\mathscr{H}(t_0,x_0)$. Such problems are a special case of $\mathscr{G}(t_0,x_0)$ (1), but not of $\mathscr{P}(t_0,x_0)$ (3). Specifically, $\mathscr{H}(t_0,x_0)$ has the form:

$$\begin{aligned}&\min_{\mathbf{u},\mathbf{x}} J_{t_0}(\mathbf{u},\mathbf{x})\\ &\text{subject to:}\\ &x(t+1)=f(x(t),u(t),t) \text{ for } t=t_0,..,T,\\ &x(t_0)=x_0,\ x(t)\in X_t\subset\mathbb{R}^n \text{ for } t=t_0,..,T,\\ &u(t)\in U\subset\mathbb{R}^m \text{ for } t=t_0,..,T-1,\\ &\mathbf{u}=(u(t_0),...,u(T-1)) \text{ and } \mathbf{x}=(x(t_0),...,x(T)),\end{aligned} \tag{8}$$

where J_{t_0} is forward separable with associated representation maps ϕ_i. For any forward separable DP problem $\mathscr{H}(t_0,x_0)$, we may associate a new augmented-state DP problem of form $\mathscr{A}(t_0,x_0)$, which is equivalent to $\mathscr{H}(t_0,x_0)$ and which satisfies the Principle of Optimality. $\mathscr{A}(t_0,x_0)$ is defined as

$$\begin{aligned}&\min_{\mathbf{u},\mathbf{z}} L_{t_0}(\mathbf{u},\mathbf{z}) := z_2(T+1)\\ &\text{subject to:}\\ &\begin{bmatrix}z_1(t+1)\\ z_2(t+1)\end{bmatrix}=\begin{bmatrix}f(z_1(t),u(t),t)\\ \phi_t(z_1(t),u(t),z_2(t))\end{bmatrix} t_0\le t<T\\ &\begin{bmatrix}z_1(T+1)\\ z_2(T+1)\end{bmatrix}=\begin{bmatrix}f(z_1(T),u(T),T)\\ \phi_T(z_1(T),z_2(T))\end{bmatrix}\\ &\begin{bmatrix}z_1(t_0)\\ z_2(t_0)\end{bmatrix}=\begin{bmatrix}x_0\\ 0\end{bmatrix},\ z_1(t)\in X_t,\ u(t)\in U \text{ for } t=t_0+1,..,T\\ &\mathbf{u}=(u(t_0),...,u(T-1)) \text{ and } \mathbf{z}=\left(\begin{bmatrix}z_1(t_0)\\ z_2(t_0)\end{bmatrix},...,\begin{bmatrix}z_1(T)\\ z_2(T)\end{bmatrix}\right)\end{aligned} \tag{9}$$

where $f:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{N}\to\mathbb{R}^n$, $z_1(t)\in\mathbb{R}^n$, $z_2(t)\in\mathbb{R}^{d_t}$, $d_t=\dim(Im\{\phi_{t-1}\})$ and $u(t)\in\mathbb{R}^m$ for all t.

Lemma 10. *Suppose J_{t_0} is the objective function for the DP problem $\mathscr{H}(t_0,x_0)$ (8) and is forward separable with associated representation maps ϕ_i. Consider the augmented DP problem $\mathscr{A}(t_0,x_0)$ (9) and denote its objective* *function by L_{t_0}. Then $J^*_{t_0} = L^*_{t_0}$. Furthermore, suppose* **u** *and* $\mathbf{x} = (x(t_0),...,x(T))$ *solve* $\mathscr{H}(t_0,x_0)$ *and* **w** *and* $\mathbf{z} = \left(\begin{bmatrix} z_1(t_0) \\ z_2(t_0)\end{bmatrix},...,\begin{bmatrix} z_1(T) \\ z_2(T)\end{bmatrix}\right)$ *solve* $\mathscr{A}(t_0,x_0)$. *Then* $\mathbf{u} = \mathbf{w}$ *and* $x(t) = z_1(t)$ *for all t.*

Proof. Suppose **w** and **z** solve $\mathscr{A}(t_0,x_0)$. First we show that **w** and $\mathbf{z}_1 := (z_1(t_0),...,z_1(T))$ are feasible for $\mathscr{H}(t_0,x_0)$. Clearly $w(t) \in U$ for all t and if we let $\mathbf{u} = \mathbf{w}$ then $x(0) = x_0$ and $x(t+1) = f(x(t),u(t),t)$ for all t. Since likewise $z_1(t_0) = x_0$ and $z_1(t+1) = f(z_1(t),u(t),t)$, we have $x(t) = z_1(t) \in X_t$ for all t. Hence **u** and $\mathbf{x} = \mathbf{z}_1$ are feasible for $\mathscr{H}(t_0,x_0)$. Likewise, if **u** and **x** solve $\mathscr{H}(t_0,x_0)$, then if we let $\mathbf{w} = \mathbf{u}$ and $\mathbf{z}_1 = \mathbf{x}$ and define $z_2(t+1) = \phi_t(z_1(t),u(t),z_2(t))$, $z_2(t_0+1) = \phi_0(z_1(t_0),u(t_0))$, $z_2(t_0) = 0$, then **w** and **z** are feasible. Furthermore, since $\mathscr{H}(t_0,x_0)$ has a forward separable objective function we have,

$$J_{t_0}(\mathbf{u},\mathbf{x}) = \phi_T(z_1(T),\phi_{T-1}(z_1(T-1),w(T-1),\phi_{T-2}(....,\\ \phi_{t_0+1}(z_1(t_0+1),w(t_0+1),\phi_{t_0}(z_1(t_0),w(t_0)))),....,))).$$

However, we now observe

$$\begin{aligned} &z_2(T+1) = \phi_T(z_1(T),z_2(T)) \\ &z_2(T) = \phi_{T-1}(z_1(T-1),u(T-1),z_2(T-1)) \\ &\vdots \\ &z_2(t_0+1) = \phi_{t_0}(z_1(t_0),u(t_0)). \\ &z_2(t_0) = 0. \end{aligned}$$

Hence we have,

$$\begin{aligned} L_{t_0}(\mathbf{w},\mathbf{z}) &= z_2(T+1) \\ &= \phi_T(z_1(T),\phi_{T-1}(z_1(T-1),w(T-1),\phi_{T-2}(...., \\ &\quad \phi_{t_0+1}(z_1(t_0+1),w(t_0+1),\phi_{t_0}(z_1(t_0),w(t_0)))),....,))). \\ &= J_{t_0}(\mathbf{u},\mathbf{x}). \end{aligned}$$

Thus if **w** and **z** solve $\mathscr{A}(t_0,x_0)$ with objective $L^*_{t_0} = z_2(T+1)$, then **w** and $\mathbf{z}_1$ solve $\mathscr{H}(t_0,x_0)$ with objective value $J^*_{t_0}$. □

Proposition 11. *The augmented DP problem* $\mathscr{A}(t_0,x_0)$ *in* (9) *satisfies the Principle of Optimality.*

Proof. $\mathscr{A}(t_0,x_0)$ is a special case of $\mathscr{P}(t_0,x_0)$ (3) where $c_i = 0$ for $i \neq T$ and $c_T([z_1 z_2]^T) = z_2$. Lemma 5 shows DP problems of the form $\mathscr{P}(t_0,x_0)$ satisfy the Principle of Optimality. □

Lemma 10 tells us that for any DP problem with forward separable objective, $\mathscr{H}(t_0,x_0)$ (8), there exists an equivalent DP problem of the form $\mathscr{A}(t_0,x_0)$ (9). Furthermore Proposition 11 shows that $\mathscr{A}(t_0,x_0)$ satisfies the Principle of Optimality. Therefore a solution for $\mathscr{H}(t_0,x_0)$ can be found by recursively solving Bellman's equation (4) for $\mathscr{A}(t_0,x_0)$.

To understand the augmented approach intuitively, we note that DP breaks a multi-period planning problem into simpler sub-problems at each stage. However, for non-separable problems, to make the correct decision at each stage we need past information about the system. In this context, the augmented state contains the information from the trajectory history necessary to make the correct decision at the present time. However by adding augmented states we increase the state space dimension and the complexity of the DP problem.

Corollary 12. *Suppose the forward separable function,* $J : \mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T+1-t_0)} \to \mathbb{R}$, *is the objective function for DP problem* $\mathscr{H}(t_0,x_0)$ (8) *and has a representation dimension of l. Then the associated augmented DP problem with this representation,* $\mathscr{A}(t_0,x_0)$ (9), *has a state space of dimension* $l+n$ *and input space of dimension m.*

Proof. From the definition of $\mathscr{A}(t_0,x_0)$ (9), the state space dimension is $n + \max_{t_0\le t\le T}$, where $d_t = \dim(Im\{\phi_{t-1}\})$. From the definition of representation dimension, we have $\max_{t_0\le t\le T} d_t = l$ and hence it follows that the state space dimension is $n + \max_{t_0\le t\le T} d_t = n+l$. □

IV. A Class Of DP For Which The Use Of State Augmentation Is Tractable

It is well known that discretization of the state space combined with a solution of Bellman's equation become computationally intractable when the discretized dimension increases; this is often called "the curse of dimensionality". In the previous Section, we proved that any non-separable DP of state space dimension n can be converted to a separable augmented DP with state-space dimension $n+l$, where l is the representation dimension of the objective function. However, for some representations, l may increase as the time interval increases - thus triggering the curse of dimensionality. To address this problem, in this section, we define a class of forward separable objective functions, called Naturally Forward Separable (NFS) functions, with representation dimension, l, which is independent of the number of time steps and the dimension of the state and input space.

Before we define NFS functions we motivate this new class of functions by showing that it is possible to represent any function as a forward separable function. To do this we introduce some additional notation. Specifically, for a vector $v = (v_1,...,v_n)^T \in \mathbb{R}^n$ we define $[v]_i^j = (v_i,...,v_j)$ for $1 \le i < j \le n$.

Lemma 13. *Any function* $J : \mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T+1-t_0)} \to \mathbb{R}$ *is forward separable with a representation of dimension* $l(n,m,T-t_0) = (T-t_0)(n+m)$.

Proof. Consider a function $J : \mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T+1-t_0)} \to \mathbb{R}$. To show J is forward separable we define a forward separable representation $\{\phi_i\}_{i=t_0}^T$ which satisfy (6) as follows.

First, define $\phi_{t_0} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n+m}$ as

$$\phi_{t_0}(x,u) = [x^T,u^T] = \left[x_1,...,x_n,u_1,...,u_m\right].$$

For $i \in \{t_0+1,...T-1\}$ the define $\phi_i : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{(i-t_0)(n+m)} \to \mathbb{R}^{(i+1-t_0)(n+m)}$ as

$$\phi_i(x,u,w) = \left[[w]_1^{n(i-t_0)}, x^T, [w]_{n(i-t_0)+1}^{(i-t_0)(n+m)}, u^T\right].$$

Lastly, define $\phi_T : \mathbb{R}^n \times \mathbb{R}^{(T-t_0)(n+m)} \to \mathbb{R}$ as

$$\phi_T(x,w) = J([[w]_1^{n(T-t_0)},x],[w]_{n(T-t_0)+1}^{(n+m)(T-t_0)}).$$

Clearly, this definition of ϕ_i satisfies (6). Furthermore, it can be seen that the maximum dimension of the images of the maps $\{\phi_i\}_{i=t_0}^T$ is $(T-t_0)(n+m)$ showing the dimension of this representation of J is $l(n,m,T-t_0) = (T-t_0)(n+m)$. □

In the above approach to show that $J(\mathbf{u},\mathbf{x})$ is forward separable we naively took the strategy of using the functions $(\phi_i)_{t_0 \leq i \leq T}$ to act like memory functions; that is to store the entire historic state trajectory and input sequence used. If $J(\mathbf{u},\mathbf{x})$ is the objective function for some DP problem of form $\mathscr{H}(t_0,x_0)$ (8) then this approach would result in the associated augmented DP problem, $\mathscr{A}(t_0,x_0)$ (9), having a very large state space dimension. Specifically, Corollary 12 shows that $\mathscr{A}(t_0,x_0)$ has state space dimension $(T-t_0)(n+m)+n$. Clearly, for a large number of time-steps, $T-t_0$, $\mathscr{A}(t_0,x_0)$ is intractable. For this reason we next define a special class of forward separable functions that have a representation with dimension independent of the number of time-steps.

Definition 14. *We say a function $J:\mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T+1-t_0)} \to \mathbb{R}$ is a Naturally Forward Separable (NFS) function if there exists maps, $\{\phi_i\}_{i=t_0}^{T}$, that satisfy* (6) *with representation dimension independent of n, m and T.*

A. An Algebra Of Naturally Forward Separable Functions

Given a function, $J:\mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T+1-t_0)} \to \mathbb{R}$, there is no obvious way to determine whether J is NFS. Instead, in this section, we show that the set of NFS functions form an algebra closed under pointwise multiplication and which is preserved under nonlinear transformation - implying that simple NFS functions ('building blocks') can be combined to construct new, more complex, NFS functions. In this way, one might approach the problem of finding representation maps for a function, J, by combining known NFS "building blocks". Several examples of such "building blocks" can be found in Subsection IV-B. We first prove closure under addition and pointwise multiplication.

Lemma 15. *Consider the function $U:\mathbb{R}\to\mathbb{R}$ and the NFS function, $J_1:\mathbb{R}^{m_1\times(T_1-t_1)} \times \mathbb{R}^{n_1\times(T_1+1-t_1)} \to \mathbb{R}$ and $J_2:\mathbb{R}^{m_2\times(T_2-t_2)} \times \mathbb{R}^{n_2\times(T_2+1-t_2)} \to \mathbb{R}$, with representation dimensions l_1 and l_2 respectively. The functions $G_1(\mathbf{u},\mathbf{x}) = J_1(\mathbf{u},\mathbf{x})+J_2(\mathbf{u},\mathbf{x})$, $G_2(\mathbf{u},\mathbf{x}) = J_1(\mathbf{u},\mathbf{x})\cdot J_2(\mathbf{u},\mathbf{x})$ and $G_3(\mathbf{u},\mathbf{x}) = U(J_1(\mathbf{u},\mathbf{x}))$ are NFS functions with representation dimension less than or equal to l_1+l_2, l_1+l_2, and l_1, receptively.*

Proof. For simplicity let us consider the case where $t_1=t_2$ and $T_1=T_2$; other cases follow by the same argument. As J_1 and J_2 are forward separable functions there exist associated representations $\{g_i\}$ and $\{h_i\}$ such that J_1 and J_2 can be written in the form (6) and with associated representation dimensions l_1 and l_2, respectively. We now show that G_1 is forward separable by defining the associate representation $\{\phi_i\}$ such that G_1 can be written in the form (6). Specifically, let

$$\phi_{t_1}(x,u) = \begin{bmatrix} g_{t_1}(x,u) \\ h_{t_1}(x,u) \end{bmatrix}, \tag{10}$$

$$\phi_i(x,u,w) = \begin{bmatrix} g_i(x,u,[w]_1^{d_{i-1}}) \\ h_i(x,u,[w]_{d_{i-1}+1}^{d_{i-1}+s_{i-1}}) \end{bmatrix} \text{ for } i\in\{t_1+1,....,T_1-1\}$$

$$\phi_{T_1}(x,u,w) = g_T(x,u,[w]_1^{d_{T_1-1}}) + h_T(x,u,[w]_{d_{T_1-1}+1}^{d_{T_1-1}+s_{T_1-1}}),$$

where $d_i = \dim(Im\{g_i\})$ and $s_i = \dim(Im\{h_i\})$ for $i \in \{t_1,...,T_1-1\}$.

We conclude that G_1 has a representation dimension, denoted l_{G_1}, such that

$$\begin{aligned} l_{G_1} &= \max_{i\in\{t_1,...,T_1-1\}} \{d_i+s_i\} \\ &\leq \max_{i\in\{t_0,...,T-1\}} \{d_i\} + \max_{i\in\{t_0,...,T-1\}} \{s_i\} \\ &= l_1+l_2. \end{aligned}$$

Furthermore, by a similar argument it can be shown that G_2 and G_3 are NFS with representation dimension less than or equal to l_1+l_2. We are able to show this using the same representation maps $\{\phi_i\}_{t_1\leq i\leq T_1-1}$ from (10) with the terminal representation map for G_2 given by

$$\phi_{T_1}(x,u,w) = g_T\left(x,u,[w]_1^{d_{T_1-1}}\right)\cdot h_T\left(x,u,[w]_{d_{T_1-1}+1}^{d_{T_1-1}+s_{T_1-1}}\right),$$

and the terminal representation map for G_3 given by

$$\phi_{T_1}(x,u,w) = U\left(g_T\left(x,u,[w]_1^{d_{T_1-1}}\right)\right).$$

□

B. Simple Examples Of NFS Functions

The first example of a NFS function is found in problems involving risk measures and certainty equivalents [12]. In this case, we have the function $U(x)=\frac{1}{\gamma}e^{\gamma x}$ and apply the following:

Example 16. *For any functions $U:\mathbb{R}\to\mathbb{R}$ and $c_t:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$,*

$$J(\mathbf{u},\mathbf{x}) = U\left(\sum_{t=t_0}^{T-1} c_t(x(t),u(t))\right)$$

is NFS with representation dimension 1.

Proof. The additively separable function $\sum_{t=t_0}^{T-1} c_t(x(t),u(t))$ is NFS using the representation maps given in (7). It therefore follows J is NFS by Lemma 15. □

Example 17. *The p-norm function given by*

$$J(\mathbf{u},\mathbf{x}) = \left(\sum_{t=t_0}^{T-1} ||x(t)||_2^p\right)^{\frac{1}{p}},$$

where $||\cdot||_2$ is the euclidean norm and $p>0$, is NFS with representation dimension 1.

Proof. Follows by Example 16 using $U(x)=x^{\frac{1}{p}}$ and $c_t(x,u)=||x||_2^p$. □

We next give a NFS function that can be considered a discrete time version of the Green measure; when used as an objective function for a DP problem it measures the amount of time the state and input spend in some set.

Example 18. *Consider the function $J:\mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T+1-t_0)} \to \mathbb{R}$ defined as*

$$J(\mathbf{u},\mathbf{x}) = |\{i\in\{t_0,...,T\} : (x(i),u(i))\in S\}|$$

where $\mathbf{u}=(u(t_0),...,u(T-1))$, $u(t)\in\mathbb{R}^m$, $\mathbf{x}=(x(t_0),...,x(T))$, $x(t)\in\mathbb{R}^n$, $S\subset\mathbb{R}^n\times\mathbb{R}^m$ and for $B\subset\mathbb{N}$ we denote $|B|$ to be the *cardinality of the set B. Then J is NFS and has a representation of dimension 1.*

Proof. We present functions such that $J(\mathbf{u},\mathbf{x})$ that can be written in the form of (6).

Define $\phi_{t_0}:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ as

$$\phi_{t_0}(x,u)=\begin{cases}1 & \text{if } (x,u)\in S\\ 0 & \text{otherwise}\end{cases}.$$

Define $\phi_t:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}\to\mathbb{R}$ for $1\le t\le T-1$ as

$$\phi_t(x,u,w)=\begin{cases}w+1 & \text{if } (x,u)\in S\\ w & \text{otherwise.}\end{cases}$$

Define the function $\phi_T:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ as

$$\phi_T(x,w)=\begin{cases}w+1 & \text{if } (x,u)\in S\\ w & \text{otherwise.}\end{cases}$$

This definition of ϕ_i satisfies (6). Moreover it can be seen that the maximum dimension of the images of the maps $\{\phi_i\}_{i=t_0}^T$ is 1 implying that the dimension of this representation of J is 1. □

Example 19. *Consider the variance type function,* $J:\mathbb{R}^{m\times T}\times\mathbb{R}^{n\times(T+1)}\to\mathbb{R}$ *defined as*

$$J(\mathbf{u},\mathbf{x})=\sum_{t=0}^{T}\left[a_t(x(t))-\frac{1}{T}\sum_{s=0}^{T}a_s(x(s))\right]^2 \tag{11}$$

where $\mathbf{u}=(u(0),...,u(T-1))$, $u(t)\in\mathbb{R}^m$, $\mathbf{x}=(x(0),...,x(T))$, $x(t)\in\mathbb{R}^n$, *and* $a_t:\mathbb{R}^n\to\mathbb{R}$. *Then J is NFS and has a representation dimension of 2.*

Proof. Expanding the right hand side of (11) as in [14] we get,

$$\begin{aligned}&J(\mathbf{u},\mathbf{x})\\&=\sum_{t=0}^{T}\left[a_t^2(x(t))-\frac{2}{T}a_t(x(t))\sum_{s=0}^{T}a_s(x(s))+\frac{1}{T^2}\left(\sum_{s=0}^{T}a_s(x(s))\right)^2\right]\\&=\sum_{t=0}^{T}a_t^2(x(t))-\frac{1}{T}\left[\sum_{s=0}^{T}a_s(x(s))\right]^2.\end{aligned}$$

We now present functions $J(\mathbf{u},\mathbf{x})$ that can be written in the form of (6). We define $\phi_{t_0}:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^2$ as

$$\phi_0(x,u)=\begin{bmatrix}a_1^2(x)\\a_1(x)\end{bmatrix}.$$

We define $\phi_i:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^2\to\mathbb{R}^2$ as

$$\phi_i(x,u,[w_1,w_2]^T)=\begin{bmatrix}w_1+a_i^2(x)\\w_2+a_i(x)\end{bmatrix}\text{ for } 1\le i\le T-1.$$

Finally, $\phi_T:\mathbb{R}^n\times\mathbb{R}^2\to\mathbb{R}$ is given by,

$$\phi_T(x,[w_1,w_2]^T)=(w_1+a_T^2(x))-\frac{1}{T}\left(w_2+a_T(x)\right)^2.$$

This definition of ϕ_i satisfies (6). Moreover it can be seen that the maximum dimension of the images of the maps $\{\phi_i\}_{i=t_0}^T$ is 2 showing the dimension of this representation of J is 2. □

We now show that the maximum function, that appears in the objective function of the battery scheduling problem in Section IX, is NFS.

Example 20. *Consider the function* $J:\mathbb{R}^{m\times T}\times\mathbb{R}^{n\times(T+1)}\to\mathbb{R}$ *such that,*

$$J(\mathbf{u},\mathbf{x})=\max\{\max_{0\le k\le T-1}\{c_k(u(k),x(k))\},c_T(x(T))\}$$

where $\mathbf{u}=(u(0),...,u(T-1))$, $u(t)\in\mathbb{R}^m$, $\mathbf{x}=(x(0),...,x(T))$, $x(t)\in\mathbb{R}^n$, $c_k:\mathbb{R}^m\times\mathbb{R}^n\to\mathbb{R}$ *for* $0\le k\le T-1$ *and* $c_T:\mathbb{R}^n\to\mathbb{R}$. *Then J is NFS and has a representation dimension of 1.*

Proof.

$$\begin{aligned}J(\mathbf{u},\mathbf{x})&=\max\{\max_{0\le k\le T-1}\{c_k(u(k),x(k))\},c_T(x(T))\}\\&=\max\{c_T(x(T)),\max\{c_{T-1}(u(T-1),x(T-1)),\cdots\\&\quad\max\{..,\max\{c_1(u(1),x(1)),\max\{c_0(u(0),x(0))\}\},..\}\}.\end{aligned}$$

It is now clear we can write J in the form of (6) as follows. The function $\phi_{t_0}:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ is defined by,

$$\phi_{t_0}(x,u)=c_{t_0}(x,u).$$

The function $\phi_i:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}\to\mathbb{R}$ is defined by,

$$\phi_i(x,u,w)=\max(c_i(x,u),w)\text{ for } t_0+1\le i\le T-1.$$

The function $\phi_T:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$ is defined by,

$$\phi_T(x,w)=\max(c_T(x),w).$$

This definition of ϕ_i satisfies (6). Moreover it can be seen that the maximum dimension of the images of the maps $\{\phi_i\}_{i=t_0}^T$ is 1 showing the dimension of this representation of J is 1. □

V. Solving Deterministic Additively Separable DP Problems

In Section III we showed that all forward separable problems of the form $\mathscr{H}(t_0,x_0)$ have an equivalent DP problem of the form $\mathscr{A}(t_0,x_0)$. Problems of the form $\mathscr{A}(t_0,x_0)$ are special cases of problems of the form $\mathscr{P}(t_0,x_0)$. In this section we address the problem of implementation by numerically solving problems of the form $\mathscr{P}(t_0,x_0)$.

For implementation, we use an approximation scheme that maps our class of DP problems to a much simpler class of DP problems with finite state and control spaces. It is known for DP problems with finite state and control spaces that the infimum in Bellman's equation (4) is attained and the optimal cost to go function, $F(x,t)$, can be computed by enumeration. Similar numerical schemes with convergence proofs can be found in [16] and [15].

A. Construction Of Approximated Tractable DP Problems

Consider the DP problem $\mathscr{P}(t_0,x_0)$ (3) with compact state and control spaces of the form $X=[\underline{x},\bar{x}]^n$ and $U=[\underline{u},\bar{u}]^m$. For DP problems of this form it is not generally possible to solve Bellman's Equation (4). We thus need to consider a sequence of "close" DP problems with countable state and control spaces. We define a sequence of approximated DP problems indexed by k and denoted by $\mathscr{P}_k(t_0,x_0)$,

$$\min_{\mathbf{u},\mathbf{x}} J_{t_0}(\mathbf{u},\mathbf{x}) = \sum_{t=t_0}^{T-1} c_t(x(t),u(t)) + c_T(x(T)) \quad (12)$$

subject to:

$x(t+1) = argmin_{y\in X_k}\{||y - f(x(t),u(t),t)||_2\}$,

$x(t_0) = x_0,\ x(t) \in X_k \subset \mathbb{R}^n,\ u(t) \in U_k \subset \mathbb{R}^m$ for $t = t_0,..,T$,

$\mathbf{u} = (u(t_0),...,u(T-1))$ and $\mathbf{x} = (x(t_0),...,x(T))$,

where $X_k = \{x_1,...,x_k\}^n$ such that $\underline{x} = x_1 < x_2 < ... < x_k = \bar{x}$ and $||x_{i+1} - x_i||_2 = \frac{\bar{x}-\underline{x}}{k}$ for $1 \le i \le k-1$, and $U_k = \{u_1,...,u_k\}^m$ such that $\underline{u} = u_1 < u_2 < ... < u_k = \bar{u}$ and $||u_{i+1} - u_i||_2 = \frac{\bar{u}-\underline{u}}{k}$ for $1 \le i \le k-1$.

Note that our approximation scheme is based on the discretization of the state and input space and as such is subject to the "curse of dimensionality". However, discretization-based schemes for solving Bellman's equation have efficient parallel implementations - see [26]. Alternatively, other Approximate Dynamic Programming (ADP) schemes such as [27] use grid sampling and can be shown to converge with respect to expected cost.

B. Constructing A Feasible Policy From The Solution Of The Approximated DP Problem

By iteratively solving Bellman's equation (4) we can find an optimal solution to $\mathscr{P}_k(t_0,x_0)$ which we denote as $(\mathbf{x}_k^*,\mathbf{u}_k^*)$. Because the vector fields that define the underlying dynamics of $\mathscr{P}(t_0,x_0)$ and $\mathscr{P}_k(t_0,x_0)$ are different, the solution $(\mathbf{x}_k^*,\mathbf{u}_k^*)$ is not necessarily feasible for $\mathscr{P}(t_0,x_0)$. However using an optimal policy for $\mathscr{P}_k(t_0,x_0)$, π_k^*, we can construct a feasible policy for $\mathscr{P}(t_0,x_0)$ in the following way,

$$\theta_k(x,t) = \arg\min_{u\in\Gamma_{t,x}} ||\pi_k^*(\arg\min_{y\in X_k}\{||y-x||_2\},t) - u||_2 \in \Pi \quad (13)$$

where we recall $\Gamma_{t,x}$ is the set of feasible inputs such that if $u \in \Gamma_{t,x}$ then $u \in U$ and $f(x,u,t) \in X$ for the DP problem $\mathscr{P}(t_0,x_0)$ (3).

C. Convergence Of Our Constructed Policy

Suppose $\theta_k(x,t)$, from (13), is a feasible policy for $\mathscr{P}(t_0,x_0)$ constructed from an optimal policy of $\mathscr{P}_k(t_0,x_0)$ using (13). Let $\mathbf{u}_k = (\theta_k(x_0,t),...,\theta_k(x_k(T-1),T-1))$ and $\mathbf{x}_k = (x_k(t_0),...,x_k(T))$ where $x_k(t_0) = x_0$, $x_k(t+1) = f(x_k(t),\theta_k(x_k(t),t),t)$ and f is the vector field from $\mathscr{P}_k(t_0,x_0)$. From Theorem 2 in [16] if $\mathscr{P}(t_0,x_0)$ satisfies certain continuity assumptions then it is known that

$$\lim_{k\to\infty} ||J_{t_0}(\mathbf{u}_k,\mathbf{x}_k) - J_{t_0}^*|| = 0, \quad (14)$$

where $J_{t_0}(\mathbf{u}_k,\mathbf{x}_k)$ is the resulting value objective function of $\mathscr{P}(t_0,x_0)$ when the policy θ_k is used and $J_{t_0}^*$ is the optimal value of the objective function.

VI. Stochastic DP And Optimal Policies

As shown in Counterexample 8 there exist non-separable deterministic DP problems that do not satisfy the Principle of Optimality as formulated in Definition 4. This Principle Of Optimality stated that the optimality of an input sequence for any instantiation of the sequence of DP problems is inherited by every subsequent instantiation - implying that recursive use of Bellman's equation will eventually return an input which solves the DP problem for every instantiation.

In the stochastic case, however, additively separable DP problems may not satisfy this version of Principle of Optimality. Specifically, we show that even in the case of an additively separable objective function, stochastic perturbations can drive the system to a state wherein the original optimal input sequence is no longer optimal. This implies that for problems which include stochastic perturbations, a new criterion must be proposed for the use of Bellman's equation. This reformulation of the Principle Of Optimality is non-trivial and requires us to re-work our definitions of DP. Specifically, we formulate the stochastic DP problem as a time-indexed sequence of DP problems, wherein the variable is not the input sequence, but is rather replaced by the policy. This policy is then used to generate a *set* of possible trajectories and input sequences, indexed by the set of possible instantiations of the random variables.

Using this new formulation of the stochastic DP problem, our definition of the Principle Of Optimality requires that optimality of the *policy* is inherited by every possible instantiation of the trajectory with probability one - i.e. the policy may be sub-optimal, but only on a set of instantiations of measure zero. We then show that the state-augmentation strategy we propose for DP problems with stochastic perturbations and NFS objective functions results in a stochastic DP problem for which this revised the Principle Of Optimality holds. Finally, we show how Bellman's equation can be used in this case to recover an optimal policy.

We begin by defining the map from a chosen policy (π), initial condition (x_0) and instantiation of the random variables ($[\mathbf{v}]_{t_0}^{T-1}$) to the resulting trajectory, x; this will allow us to state precisely which random variables the expectation in the objective function is respect to. For simplicity we only consider random variables with Gaussian distributions.

Definition 21. *For a vector field $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{N} \times \mathbb{R}^q \to \mathbb{R}^n$, a set of optimal polices Π associated with some DP problem, a starting time $t_0 \in \mathbb{N}$, and terminal time $T \in \mathbb{N}$, let us denote* ***the state map*** *by $\psi_{f,t_0} : \Pi \times \mathbb{R}^n \times \mathbb{N} \times \mathbb{R}^{q\times(T-t_0)} \to \mathbb{R}^n$. We say that $x = \psi_{f,t_0}(\pi,x_0,T,[\mathbf{v}]_{t_0}^{T-1})$ if $x = x(T)$ where $x(T)$ is a solution to the following recursion equations $x(t_0) = x_0$, $x(t+1) = f(x(t),\pi(x(t),t),t,v(t))$ for $t \in \{t_0,...,T-1\}$ and $[\mathbf{v}]_{t_0}^{T-1} = [v(t_0),..,v(T-1)] \in \mathbb{R}^{q\times(T-t_0)}$. We denote the image of the state vector under a set of instantiations $Y \subset \mathbb{R}^{q\times(T-t_0)}$ by $\psi_{f,t_0}(\pi,x_0,T,Y) = \{\psi_{f,t_0}(\pi,x_0,T,[\mathbf{v}]_{t_0}^{T-1}) \in \mathbb{R}^n : [\mathbf{v}]_{t_0}^{T-1} \in Y\}$.*

We also denote ***the trajectory map*** *by $\Phi_{f,t_0} : \Pi \times \mathbb{R}^n \times \mathbb{N} \times \mathbb{R}^{q\times(T-t_0)} \to \mathbb{R}^{m\times(T-t_0)} \times \mathbb{R}^{n\times(T-t_0+1)}$. We say that $(\mathbf{u},\mathbf{x}) = \Phi_{f,t_0}(\pi,x_0,T,[\mathbf{v}]_{t_0}^{T-1})$ if $\mathbf{u} = (\pi(x(t_0),t_0),...,\pi(x(T-$*

$1), T-1))$, *and* $\mathbf{x} = (x(t_0), ..., x(T))$ *is such that* $x(t) = \psi_{f,t_0}(\pi, x_0, t, [\mathbf{v}]_{t_0}^{T-1})$ *for* $t \in \{t_0, ..., T-1\}$.

We define the class of stochastic DP problems with forward separable objective as $\mathscr{H}_s(t_0, x_0)$,

$$\pi^{\mathscr{H}_s *} = \arg\min_{\pi \in \Pi} \; \mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}} \left(J_{t_0}^{\mathscr{H}_s}(\Phi_{f,t_0}(\pi, x_0, T, [\mathbf{v}]_{t_0}^{T-1})) \right) \tag{15}$$
$$\text{subject to: } \psi_{f,t_0}(\pi, x_0, t, [\mathbf{v}]_{t_0}^{t-1}) \in X_t \text{ for } t = t_0, .., T$$
$$\pi(x,t) \in U_t \text{ and } v(t) \in \mathbb{R}^q \sim \mathscr{N}(\mathbf{0}, I_{q \times q}) \; \forall x \in X_t, \forall t = t_0, .., T-1,$$

where $J_{t_0}^{\mathscr{H}_s} : \mathbb{R}^{m \times (T-t_0)} \times \mathbb{R}^{n \times (T-t_0+1)} \to \mathbb{R}$ is a forward separable function with associated representation $\{\phi_i\}_{i=t_0}^T$; $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{N} \times \mathbb{R}^q \to \mathbb{R}^n$; ψ_{f,t_0} is measurable and Φ_{f,t_0} are the state and trajectory map respectively defined in Definition 21; U_i is assumed to be some compact subset of $\mathbb{R}^{m \times (i-t_0)}$; $X_i \subset \mathbb{R}^{n \times (i-t_0+1)}$; $[\mathbf{v}]_{t_0}^{T-1} = [v(t_0), .., v(t-1)] \in \mathbb{R}^{q \times (T-t_0)}$; $\mathbb{E}_\mathbf{v}$ is the expectation with respect to the random variable $\mathbf{v}$. Define $J_{t_0}^{\mathscr{H}_s *} = \mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}} \left(J_{t_0}^{\mathscr{H}_s}(\Phi_{f,t_0}(\pi^{H_s *}, x_0, T, [\mathbf{v}]_{t_0}^{T-1})) \right)$ as the expected cost of using an optimal policy when applied to $\mathscr{H}_s(t_0, x_0)$.

Change in Variables for Stochastic Problems: Unlike in the deterministic case the solution to stochastic DP problems, such as (15), is now a policy $\pi \in \Pi$ and not a definite input and state sequence $\mathbf{u}^* \in \mathbb{R}^{m \times (T-t_0)}$ and $\mathbf{x}^* \in \mathbb{R}^{n \times (T-t_0+1)}$, such as in (1). This is because the optimal sequence of inputs, $\mathbf{u}^*$, that results in an optimal trajectory, $\mathbf{x}^*$, will depend on the instantiation of the random variables. This change of notation demonstrates that the solution to DP problems involving stochastic dynamics no longer belongs to some finite dimensional space, $(\mathbf{u}^*, \mathbf{x}^*) \in \mathbb{R}^{m \times (T-t_0)} \times \mathbb{R}^{n \times (T-t_0+1)}$, but rather an infinite dimensional functional space $\pi^* \in \Pi$.

A. Stochastic Additively Separable DP Problems

In the special case when the objective function of (15) is an additively separable function, as per Definition 1, given as

$$J_{t_0}^{\mathscr{Q}}(\mathbf{u}, \mathbf{x}) = \sum_{t=t_0}^{T-1} c_t(x(t), u(t)) + c_T(x(T)), \tag{16}$$

we denote the DP problem (15) by $\mathscr{Q}(t_0, x_0)$.

stochastic DP problems of form $\mathscr{Q}(t_0, x_0)$ are often referred to as Markov Decision Processes (MDP), and are sometimes denoted by the tuple $\{\{X_t\}_{t_0}^T, \{U_t\}_{t_0}^{T-1}, \psi, \{Q_t\}_{t_0}^{T-1}, \{c\}_{t_0}^T\}$; where $\psi(x, v, t) = \{u \in U_t : f(x, u, t, v) \in X_{t+1}\}$, $Q_t(B|x, u) = \int_B \mathbb{1}_B(f(x, u, t, v)) \phi(v) dv$, and $\phi(v)$ is the probability density function of the random variable v.

B. The Principle Of Optimality For Stochastic Problems

As discussed in [6] the extension of the Principle of Optimality to the stochastic case is non-trivial. We first give an example from [28] of a stochastic DP problem which shows that an optimal policy may not be optimal for every instantiation of the random variables at future time steps.

Let us consider the following stochastic DP problem $\mathscr{W}(0, x_0)$,

$$\pi^* = \arg\min_{\pi \in \Pi} \mathbb{E}_{v(0)} \left(J_0(\Phi_{f,0}(\pi, x_0, 1, [v(0)])) \right) \tag{17}$$
$$\text{subject to: } v(0) \sim U[0,1], \quad x(0) = x_0.$$

Here $J_t(\mathbf{u}, \mathbf{x}) = -\sum_{n=t}^1 u(n)$, $f(x, u, t, v) = v$, and $\pi \in \Pi \iff \pi(x,t) \in \{0,1\} \forall x \in \mathbb{R}, t = 0, 1$.

Counterexample 22. *The policy* $\pi(x,t) = \begin{cases} 1 \text{ if } x \in [0,1) \\ 0 \text{ if } x = 1 \end{cases}$ *is optimal for the problem* $\mathscr{W}(0,0)$ *(17) but not optimal for the problem* $\mathscr{W}(1,1)$.

Proof. Clearly $J_0(\mathbf{u}, \mathbf{x}) \geq -2$ for all $(\mathbf{u}, \mathbf{x}) \in \{0,1\}^2 \times \mathbb{R}^3$ and $J_0(\mathbf{u}, \mathbf{x}) = -2$ is attainable using the input $(u(0), u(1)) = (1,1)$; therefore any solution of $\mathscr{W}(0,0)$ will minimize the objective function to a value of -2. Now using the law of total expectation we get,

$$\begin{aligned}
&\mathbb{E}_{v(0)} \left(J_0(\Psi_{f,0}(\pi, 0, 1, [v(0), v(1)])) \right) \\
&= -\mathbb{E}_{v(0)} \left(\pi(0,0) + \pi(v(0), 1) \right) \\
&= -\pi(0,0) - \mathbb{E}_{v(0)} \left(\pi(v(0),1) | v(0) \in [0,1) \right) \mathbb{P}_{v(0)}(v(0) \in [0,1)) \\
&\quad - \mathbb{E}_{v(0)} \left(\pi(v(0),1) | v(0) = 0 \right) \mathbb{P}_{v(0)}(v(0) = 0) \\
&= -2,
\end{aligned}$$

since the probability of a continuous random variable (such as a uniformly distributed random variable) taking a particular value is 0. Thus it follows the policy π is optimal for $\mathscr{W}(0,0)$. Trivially π is not optimal for $\mathscr{W}(1,1)$ as the value of the objective functions becomes 0 under π whereas the input $u(1) = 1$ produces a smaller objective function value of -1. □

Clearly, for the stochastic DP problems of form $\mathscr{H}_s(t_0, x_0)$ (15), such as $\mathscr{W}(0,0)$ (17), any optimal policy π^* does not always result in the same trajectory $\mathbf{x} = (x(t_0), ..., x(T))$; as this is dependent on the instantiations of the underlying random variables, $[\mathbf{v}]_{t_0}^{T-1}$. As Counterexample 22 has shown there exist stochastic DP problems, with additively separable objective functions, that have optimal policies that are no longer optimal for future timesteps if certain instantiations of the underlying random variables are realized. Therefore, it is too restrictive to extend Definition 4, the Principle of Optimality for the deterministic case, to the stochastic case by requiring stochastic DP problems satisfying the Principle of Optimality to be such that their optimal policies are also optimal for each instantiation at any future time step. With this in mind and motivated by the work of [7] we now give a probabilistic definition of the Principle of Optimality for stochastic DP problems.

Definition 23. *For a stochastic DP problem* $\mathscr{H}_s(t_0, x_0)$ *(15) with an optimal policy* $\pi^* \in \Pi$ *and associated state map* ψ_{f,t_0}*, defined in definition 21, let us denote the set indexed by* $k \geq t_0$*,*

$$\begin{aligned}
Y_k = \{ & [\mathbf{v}]_{t_0}^{k-1} \in \mathbb{R}^{q \times (k-t_0)} : \\
& \pi^* \text{ does not solve } \mathscr{H}_s(k, \psi_{f,t_0}(\pi^*, x_0, k, [\mathbf{v}]_{t_0}^{k-1})) \}
\end{aligned}$$

where $[\mathbf{v}]_{t_0}^{k-1} = [v(t_0), ..., v(k-1)] \in \mathbb{R}^{q \times (k-t_0)}$*. We say stochastic DP problems of the form* $\mathscr{H}_s(t_0, x_0)$ *(15)* ***satisfy the Principle of Optimality*** *if for any* $k \geq t_0$ *we have*

$$\mathbb{P}_{[\mathbf{v}]_{t_0}^{k-1}}([\mathbf{v}]_{t_0}^{k-1} \in Y_k) = 0.$$

Here $\mathbb{P}_{[\mathbf{v}]_{t_0}^{k-1}}$ *is the probability measure associated with the random variable* $[\mathbf{v}]_{t_0}^{k-1} \in \mathbb{R}^{q \times (k-t_0)}$*,* $v(t) \sim \mathscr{N}(\mathbf{0}, I_{q \times q})$ *for* $t \in \{t_0, ..., k-1\}$.

We next show that stochastic DP problems with additively separable objective functions (MDP's) satisfy the Principle of Optimality as formulated in Definition 23.

Lemma 24. *A stochastic DP problem of Form $\mathscr{Q}(t_0,x_0)$ (16) satisfies the Principle of Optimality as formulated in Definition 23.*

Proof. Suppose π^* solves $\mathscr{Q}(t_0,x_0)$. For $k>t_0$ and the state map ψ_{f,t_0} associated with $\mathscr{Q}(t_0,x_0)$ let us recall the set defined in Definition 23,

$$Y_k := \{[\mathbf{v}]_{t_0}^{k-1} \in \mathbb{R}^{q\times(k-t_0)} : \pi^* \text{ does not solve } \mathscr{Q}(k,x_{\pi^*}([\mathbf{v}]_{t_0}^{k-1}))\}.$$

Here $[\mathbf{v}]_{t_0}^{k-1} := [v(t_0),...,v(k-1)] \in \mathbb{R}^{q\times(k-t_0)}$, and we use the short-hand $x_{\pi^*}([\mathbf{v}]_{t_0}^{k-1}) := \psi_{f,t_0}(\pi^*,x_0,k,[\mathbf{v}]_{t_0}^{k-1})$.

Now for contradiction suppose there exists $k\in\{t_0,...,T\}$ such that $\mathbb{P}_{[\mathbf{v}]_{t_0}^{k-1}}([\mathbf{v}]_{t_0}^{k-1}\in Y_k)>0$; where $v(t)\sim\mathscr{N}(\mathbf{0},I_{q\times q})$ for $t\in\{t_0,...,k-1\}$. For $[\mathbf{v}]_{t_0}^{k-1}\in Y_k$ we know the policy π^* is not optimal for $\mathscr{Q}(k,x_{\pi^*}([\mathbf{v}]_{t_0}^{k-1}))$ and thus there exists a feasible policy $\theta\in\Pi$ such that,

$$\begin{aligned}&\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_k^{\mathscr{Q}}(\Phi_{f,k}(\theta,x_{\pi^*}([\mathbf{v}]_{t_0}^{k-1}),T,[\mathbf{v}]_k^{T-1}))\middle|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right)\\ &< \\ &\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_k^{\mathscr{Q}}(\Phi_{f,k}(\pi^*,x_{\pi^*}([\mathbf{v}]_{t_0}^{k-1}),T,[\mathbf{v}]_k^{T-1}))\middle|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right).\end{aligned} \tag{18}$$

Now let us consider the map,

$$\hat{\pi}_t([x(t_0),...,x(t)]) = \begin{cases}\theta(x(t),t) \text{ if } t\geq k, x(k)\in\psi_{f,t_0}(\pi^*,x_0,k,Y_k)\\ \pi^*(x(t),t) \text{ otherwise }.\end{cases} \tag{19}$$

Using Lemma 28, there exists a policy $\alpha\in\Pi$ such that (36) holds for $\{\hat{\pi}_t\}$ defined in (19). We will now show the policy α contradicts that π^* be an optimal policy for $\mathscr{Q}(t_0,x_0)$. We first note using (36) and the law of total probabilities,

$$\begin{aligned}&\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\alpha,x_0,T,[\mathbf{v}]_{t_0}^{T-1}))\right)\\ &=\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\hat{\pi},x_0,T,[\mathbf{v}]_{t_0}^{T-1}))\right)\\ &=\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\hat{\pi},x_0,T,[\mathbf{v}]_{t_0}^{T-1})|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right)\\ &\qquad\mathbb{P}_{[\mathbf{v}]_{t_0}^{k-1}}([\mathbf{v}]_{t_0}^{k-1}\in Y_k)\\ &\quad+\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\hat{\pi},x_0,T,[\mathbf{v}]_{t_0}^{T-1}))|[\mathbf{v}]_{t_0}^{k-1}\notin Y_k\right)\\ &\qquad\mathbb{P}_{[\mathbf{v}]_{t_0}^{k-1}}([\mathbf{v}]_{t_0}^{k-1}\notin Y_k).\end{aligned} \tag{20}$$

We recall the additive structure of $J_{t_0}^{\mathscr{Q}}$

$$J_{t_0}^{\mathscr{Q}}(\mathbf{u},\mathbf{x}) = \sum_{t=t_0}^{T-1} c_t(x(t),u(t)) + c_T(x(T)),$$

where $\mathbf{u}=(u(t_0),...,u(T-1))$ and $\mathbf{x}=(x(t_0),...,x(T))$ and $c_T(x):\mathbb{R}^n\to\mathbb{R}$, $c_t(x,u):\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ for $t=t_0,\cdots T-1$.

Now using the fact $\hat{\pi}_t([x(t_0),...,x(t)])=\pi^*(x(t),t)$ for all $t<k$, $\hat{\pi}_t([x(t_0),...,x(t)])=\theta(x(t),t)$ if $t\geq k$ and $x(k)\in\psi_{f,t_0}(\pi^*,x_0,k,Y_k)$, linearity of the expectation and the inequality (18) we have,

$$\begin{aligned}&\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,k}(\pi^*,x_0,T,[\mathbf{v}]_{t_0}^{T-1}))\middle|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right)\\ &=\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(\sum_{t=t_0}^{k-1}c_t(x_{\pi^*}([\mathbf{v}]_{t_0}^{t-1}),\pi^*(x_{\pi^*}([\mathbf{v}]_{t_0}^{t-1}),t))\middle|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right)\\ &\quad+\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_k^{\mathscr{Q}}(\Phi_{f,k}(\theta,x_{\pi^*}([\mathbf{v}]_{t_0}^{k-1}),T,[\mathbf{v}]_k^{T-1}))\middle|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right)\\ &<\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(\sum_{t=t_0}^{k-1}c_t(x_{\pi^*}([\mathbf{v}]_{t_0}^{t-1}),\pi^*(x_{\pi^*}([\mathbf{v}]_{t_0}^{t-1}),t))\middle|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right)\\ &\quad+\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_k^{\mathscr{Q}}(\Phi_{f,k}(\pi^*,x_{\pi^*}([\mathbf{v}]_{t_0}^{k-1}),T,[\mathbf{v}]_k^{T-1}))\middle|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right)\\ &=\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\pi^*,x_0,T,[\mathbf{v}]_{t_0}^{T}))\middle|[\mathbf{v}]_{t_0}^{k-1}\in Y_k\right).\end{aligned} \tag{21}$$

Therefore using (20); the fact $\hat{\pi}_t([x(t_0),...,x(t)])=\pi^*(x(t),t)$ if $x(k)\notin\psi_{f,t_0}(\pi^*,x_0,k,Y_{k,\pi^*})$; the total law of probability; the above inequality (21); and the assumption $\mathbb{P}_{[\mathbf{v}]_{t_0}^{k-1}}([\mathbf{v}]_{t_0}^{k-1}\in Y_k)>0$ (so the inequality remains strict) we derive,

$$\begin{aligned}&\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\alpha,x_0,T,[\mathbf{v}]_{t_0}^{T-1}))\right)\\ &\qquad<\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\pi^*,x_0,T,[\mathbf{v}]_{t_0}^{T-1}))\right).\end{aligned} \tag{22}$$

This contradicts the fact π^* is an optimal policy for $\mathscr{Q}(t,x)$. Therefore we conclude $P_{[\mathbf{v}]_{t_0}^{k-1}}([\mathbf{v}]_{t_0}^{k-1}\in Y_k)=0$ showing DP problems of the form $\mathscr{Q}(t_0,x_0)$ satisfy Definition 23 and hence satisfy the Principle of Optimality. ☐

We will now state Bellman's equation for stochastic DP problems of the form $\mathscr{Q}(t,x)$.

Proposition 25 ([24])**.** *For stochastic DP problems of the form $\mathscr{Q}(t,x)$ in* (16) *with optimal objective values $J_t^{\mathscr{Q}*}$, define the function $F(x,t)=J_t^{\mathscr{Q}*}$. Then the following hold for all $x\in X_t$,*

$$\begin{aligned}&F(x,t)=\inf_u\{c_t(x,u)+\mathbb{E}_v[F(f(x,u,t,v),t+1)]\}.\\ &F(x,T)=c_T(x).\end{aligned} \tag{23}$$

We see in the next corollary that if we are able to solve the stochastic Bellman equation (23) then we are able to construct an optimal policy that solves (16).

Corollary 26 ([24])**.** *Consider a stochastic DP problem of the form $\mathscr{Q}(t_0,x_0)$ in* (16)*. Suppose $F(x,t)$ satisfies Equation* (23) *and suppose there exists a policy such that,*

$$\theta(x,t)\in\arg\min_{u\in\Gamma_{t,x}}\{c_t(x,u)+\mathbb{E}_v[F(f(x,u,t,v),t+1)]\}.$$

Then the policy θ solves $\mathscr{Q}(t_0,x_0)$.

C. State Augmentation For Stochastic DP problems

Analogous to the deterministic case shown in Lemma 10, for a stochastic DP problem of the form $\mathscr{H}_s(t_0,x_0)$ (15) we can use the separable representation maps $\{\phi_i\}_{i=t_0}^T$ of the objective function $J_{t_0}^{\mathscr{H}_s}$ to construct an equivalent DP problem of form $\mathscr{Q}(t_0,x_0)$ (16) by using state augmentation.

VII. Numerically Solving Additively Separable Stochastic DP Problems

In Section VI, we proposed a state-augmentation scheme for converting a stochastic forward separable DP problem, of form $\mathscr{H}_s(t_0,x_0)$ (15), to an equivalent additively separable DP problem, of form $\mathscr{Q}(t_0,x_0)$ (16). In this section we propose a scheme to numerically solve stochastic additively separable DP problems of the form $\mathscr{Q}(t_0,x_0)$ (16).

To numerically solve problems of the form $\mathscr{Q}(t_0,x_0)$, we propose a discretization scheme similar to the one detailed in Section V. However, unlike in the deterministic case, the presence of random variables, $v \sim \mathscr{N}(0,1)$, implies a non-compact state space. This requires the use of state projection onto appropriately constructed approximating compact sets.

A. Constructing An Approximated DP Problem With Compact State Space

Consider the stochastic DP problem $\mathscr{Q}(t_0,x_0)$ with compact control space $U=[\underline{u},\bar{u}]^m$ and underlying random variables $v \sim \mathscr{N}(\mathbf{0},I_{q\times q})$. As in [15], we assume $\forall \varepsilon>0$ that there exists a compact set $H_{\varepsilon,t}=[\underline{x}_{\varepsilon,t},\bar{x}_{\varepsilon,t}]^n \subset X$ (that depends on ε and t) such that $x_0 \in H_{\varepsilon,0}$ and,

$$\sup_{x\in H_{\varepsilon,t},u\in U} \mathbb{P}_v(f(x,u,t,v)\notin H_{\varepsilon,t+1})<\varepsilon. \tag{24}$$

We then construct the associated compact stochastic DP approximation to $\mathscr{Q}(t_0,x_0)$ denoted by $\mathscr{Q}_{\varepsilon,k}(t_0,x_0)$,

$$\begin{aligned}&\arg\min_{\pi\in\Pi}\ \mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{\tilde{f},t_0}(\pi,x_0,T,[\mathbf{v}]_{t_0}^{T-1}))\right) \qquad (25)\\ &\text{subject to: } \ \psi_{\tilde{f},t_0}(\pi,x_0,t,[\mathbf{v}]_{t_0}^{t-1})\in\tilde{X}_{\varepsilon,t,k} \text{ for } t=t_0,..,T,\\ &\pi(x,t)\in\tilde{U}_k \text{ and } v(t)\in\mathbb{R}^q\sim\mathscr{N}(\mathbf{0},I_{q\times q})\forall x\in X_t,\forall t=t_0,..,T-1,\end{aligned}$$

where $\tilde{f}(x,u,t,v) = \arg\min_{y\in X_{\varepsilon,t+1,k}}\{||y-f(x,u,t,v)||_2\}$, $\tilde{X}_{\varepsilon,t,k}=\{x_{1,t},...,x_{k,t}\}^n$ such that $\underline{x}_{\varepsilon,t}=x_{1,t}<x_{2,t}<...<x_{k,t}=\bar{x}_{\varepsilon,t}$ and $||x_{i+1,t}-x_{i,t}||_2=\frac{\bar{x}_{\varepsilon,t}-\underline{x}_{\varepsilon,t}}{k}$ for $1\le i\le k-1$, $\tilde{U}_k=\{u_1,...,u_k\}^m$ such that $\underline{u}=u_1<u_2<...<u_k=\bar{u}$ and $||u_{i+1}-u_i||_2=\frac{\bar{u}-\underline{u}}{k}$ for $1\le i\le k-1$, and $[\mathbf{v}]_{t_0}^{T-1}=[v(t_0),..,v(T-1)]\in\mathbb{R}^{q\times(T-t_0)}$.

Analogous to the deterministic case, an optimal policy $\pi^*_{\varepsilon,k}$ for $\mathscr{Q}_{\varepsilon,k}(t_0,x_0)$ can be found exactly by iteratively solving Bellman's equation (23). One can then construct a feasible policy for $\mathscr{Q}(t_0,x_0)$ using,

$$\theta_{\varepsilon,k}(x,t)=\arg\min_{u\in\Gamma_{t,x}}||\pi^*_{\varepsilon,k}(\arg\min_{y\in X_{\varepsilon,t,k}}\{||y-x||_2\},t)-u||_2\in\Pi \tag{26}$$

where $\Gamma_{t,x}$ is the set of feasible controls at time $t\in\{0,...,T-1\}$ and state position $x\in\mathbb{R}^n$ for $\mathscr{Q}(t_0,x_0)$ (16) and $X_{\varepsilon,t,k}$ is the state grid constraint in the problem $\mathscr{Q}_{\varepsilon,k}(t_0,x_0)$ (25).

If $\mathscr{Q}(t_0,x_0)$ satisfies assumption (A1) to (A4) from Theorem 3.5 [15] then

$$\lim_{\varepsilon\to0,k\to\infty}\left|\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{\tilde{f},t_0}(\theta_{\varepsilon,k},x_0,T,[\mathbf{v}]_{t_0}^{T-1}))\right)-J_{t_0}^{\mathscr{Q}*}\right|=0, \tag{27}$$

where $J_{t_0}^{\mathscr{Q}*}=\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}}\left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\pi^{\mathscr{Q}*},x_0,T,[\mathbf{v}]_{t_0}^{T-1}))\right)$ is the expected cost of using an optimal policy when applied to $\mathscr{Q}(t_0,x_0)$.

VIII. Summary: Solving NFS DP Problems Using Augmentation And Discretization

Given a DP problem with a NFS objective function, with known representation maps, we have shown in Section III and Section VI how to construct equivalent DP problems with additively separable objective functions. We have furthermore proposed discretization schemes in Section V, for the deterministic case, and Section VII-A, for the stochastic case, to solve DP problems with additively separable objective functions. We now summarize these results by proposing the following steps for solving a non-separable DP problem. Given a DP problem of the from $\mathscr{H}(t_0,x_0)$ (8), or $\mathscr{H}_s(t_0,x_0)$ (15) if stochastic, we do the following:

1) Find a NFS representation of the objective function (Eqn. (6)) with associated representation maps. One approach to this is to use Section IV-A which details how to combine known NFS functions, with known representation maps, in order to find potential representation maps for other NFS functions.
2) Construct the associated augmented DP problem of form $\mathscr{P}(t_0,x_0)$ (3), if deterministic, or $\mathscr{Q}(t_0,x_0)$ (16), if stochastic.
3) Use discretization to approximate the augmented DP problem using Form $\mathscr{P}_k(t_0,x_0)$ (12), if deterministic, or $\mathscr{Q}_{\varepsilon,k}(t_0,x_0)$ (25), if stochastic.
4) Numerically solve $\mathscr{P}_k(t_0,x_0)$ or $\mathscr{Q}_{\varepsilon,k}(t_0,x_0)$ for a sufficiently large $k\in\mathbb{N}$.
5) Construct a feasible policy for the original DP problem from an optimal policy of $\mathscr{P}_k(t_0,x_0)$ or $\mathscr{Q}_{\varepsilon,k}(t_0,x_0)$ using (13) or (26).

To illustrate how we use state augmentation and discretization methods we consider the following DP problem from [8].

$$\begin{aligned}\min J = x(3)^2[u(0)^2+u(1)^2+u(1)u(2)^2]^{\frac{1}{2}}&\\ +[u(0)^2+u(1)^2+u(1)u(2)^2]^2& \qquad (28)\\ \text{subject to,}\quad x(t+1)=\frac{x(t)}{u(t)} \quad \text{for } t\in\{1,2,3\}&\\ x(0)=10,\quad u(0),u(1),u(2)\ge0.&\end{aligned}$$

In [8] an analytic solution for (28) was found to be:

$$\mathbf{x}^*=\begin{bmatrix}10\\6.3943938\\5.782475\\3.8882658\end{bmatrix},\ \mathbf{u}^*=\begin{bmatrix}1.5638699\\1.105823\\1.4871604\end{bmatrix},\ J^*=74.767439.$$

The objective function J in (28) is NFS and has a representation dimension of 2. This can be shown by writing J in the form of (6) using the functions,

$$\phi_0(x,u)=u^2,\quad \phi_1(x,u,w)=\begin{bmatrix}w+u^2\\u\end{bmatrix}$$
$$\phi_2\left(x,u,\begin{bmatrix}w_1\\w_2\end{bmatrix}\right)=w_1+w_2^2u^2,$$
$$\phi_3(x,w)=x^2\sqrt{w}+w^2.$$

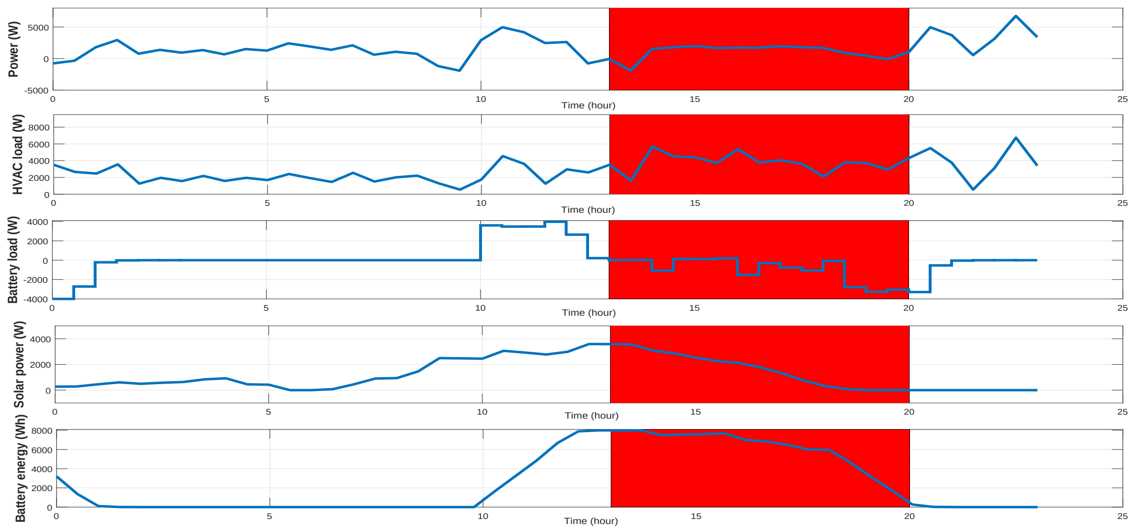

Figure 1. The trajectory the algorithm produces for randomly generated stochastic solar data. The supremun of the power is 1.05788(kw) and the cost is \$47.7211.

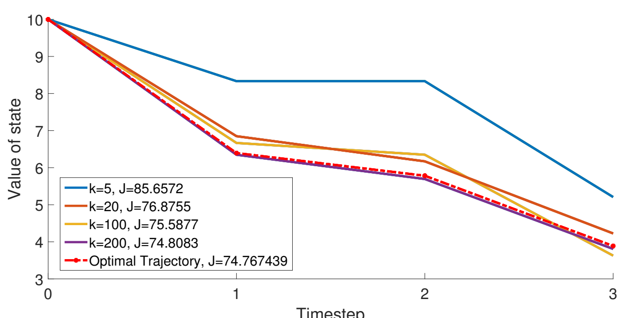

Figure 2. The resulting state trajectories from using the policy constructed from $P_k(t_0,x_0)$ in the DP Problem (28).

The DP Problem (28) can now be written in the form of $\mathscr{A}(t_0,x_0)$ using state augmentation, as

$$\min z_3(4) \tag{29}$$
subject to,
$$z_1(t+1)=\frac{z_1(t)}{u(t)},\; z_2(t+1)=\begin{cases}u(t) \text{ if t=1}\\ 0 \text{ otherwise}\end{cases} \quad \forall t\in\{1,2,3\},$$
$$z_3(1)=u(1)^2, \quad z_3(2)=z_3(1)+u(1)^2,$$
$$z_3(3)=z_3(2)+z_2(2)^2u(2), \quad z_3(4)=z_1(3)^2\sqrt{z_3(3)}+z_3(3)^2,$$
$$z_1(0)=10,\; z_2(0)=0,\; z_3(0)=0 \quad u(0),u(1),u(2)\geq 0.$$

The DP Problem (29) is now a special case of $\mathscr{P}(t_0,x_0)$ and equivalent to the original DP Problem (28). The associated approximated DP problem of the form $\mathscr{P}_k(t_0,x_0)$ (12) can now be found by selecting appropriate compact state and control spaces; $X\subset\mathbb{R}^3$ and $U\subset\mathbb{R}$. A feasible policy for (28) is then constructed from an optimal policy of the associated $\mathscr{P}_k(t_0,x_0)$ using (13). Figure 2 shows the state trajectories by following different constructed policies for various values of k. It is seen that for $k=200$ the algorithm produces a solution within three significant figures of the analytic optimal objective function for (28).

IX. Application To The Energy Storage Problem

We apply our augmented DP methodology to the scheduling of batteries in the presence of demand charges and show that our proposed algorithm outperforms existing heuristics, such as [22] (approximately \$0.98 savings). To do this, we propose a simple model for the dynamics of battery storage. We then formulate the objective function using electricity pricing plans which include demand charges. We see that the system described becomes a DP problem of the form $\mathscr{S}(t_0,x_0)$ (5); which can be tractably solved as it has a NFS objective function. We will first solve the battery scheduling problem in the deterministic case based on real historical solar data. Later, we develop a stochastic Markov model that generates similar solar data to that seen in Tempe, Az.

A. Battery Dynamics

We model the energy stored in the battery using the difference equation:

$$e(k+1)=\alpha(e(k)+\eta u(k)\Delta t), \tag{30}$$

where $e(k)$ denotes the energy stored in the battery at time step k, α is the bleed rate of the battery, η is the efficiency of the battery, $u(k)$ denotes the charging/discharging $(+/-)$ at time step k and Δt is the amount of time passed between each time step. Moreover we denote the maximum charge and discharge rate by $\bar{u}$ and $\underline{u}$ respectively. Thus we have the constraint that $u(k)\in[\underline{u},\bar{u}]:=U$ for all k. Similarly we also add the constraint $e(k)\in[\underline{e},\bar{e}]:=X$ for all k where $\underline{e}$ and $\bar{e}$ are the capacity constraints of the battery (typically $\underline{e}=0$).

B. The Objective Function

Let us denote $q(k)$ as the power supplied by the grid at time step k.

$$q(k)=q_a(k)-q_s(k)+u(k), \tag{31}$$

where $q_a(k)$ and $q_s(k)$ are the power consumed by HVAC/appliances and the power supplied by solar photovoltaics at time step k respectively. For now, it is assumed that both are known a priori.

To define the cost of electricity we divide the day $t \in [0,T]$ into on-peak and off-peak periods. We define an off peak period starting from 12am till t_{on} and t_{off} till 12am. We define an on-peak period between t_{on} till t_{off}. The Time-of-Use (TOU, \$ per kWh) electricity cost during on-peak and off-peak is denoted by p_{on} and p_{off} respectively. We further simplify this as $p_k = p_{on}$ if $k \in T_{on}$ and $p_k = p_{off}$ if $k \in T_{off}$ where T_{on} and T_{off} are the on-peak and off-peak hours, respectively. These TOU charges define the first part of the objective function as:

$$J_{\text{E}}(\mathbf{u},\mathbf{e}) = p_{\text{off}} \sum_{k=0}^{t_{\text{on}}-1} q(k)\Delta t + p_{\text{on}} \sum_{k=t_{\text{on}}}^{t_{\text{off}}-1} q(k)\Delta t + p_{\text{off}} \sum_{k=t_{\text{off}}}^{T} q(k)\Delta t$$
$$= \sum_{k\in[0,T]} p_k(q_a(k)-q_s(k))\Delta t + \sum_{k\in[0,T]} p_k u(k)\Delta t$$

where the daily terminal timestep is $T = 24/\Delta t$.

We also include a demand charge, which is a cost proportional to the maximum rate of power taken from the grid during on-peak times. This cost is determined by p_d which is the price in \$ per kW. Thus it follows the demand charge will be:

$$J_D(\mathbf{u},\mathbf{e}) = p_d \max_{k\in\{t_{\text{on}},\dots,t_{\text{off}}-1\}} \{q_a(k)-q_s(k)+u(k)\}.$$

C. 24 hr Optimal Residential Battery Storage Problem

We may now define the problem of optimal battery scheduling in the presence of demand and Time-of-Use charges, denoted $\mathscr{D}(0,e_0)$.

$$\begin{aligned} &\min_{\mathbf{u},\mathbf{e}} \{J_E(\mathbf{u},\mathbf{e}) + J_D(\mathbf{u},\mathbf{e})\} \text{ subject to} \qquad (32)\\ &e(k+1) = \alpha(e(k)+\eta u(k)\Delta t) \text{ for } k=0,\dots,T\\ &e_0 = e0 \;, e(k)\in X,\; u(k)\in U \text{ for } k=0,\dots,T,\\ &\mathbf{u} = (u(0),\dots,u(T-1)) \text{ and } \mathbf{e} = (e(0),\dots,e(T)) \end{aligned}$$

where recall $U := [\underline{u},\bar{u}]$ and $X := [\underline{e},\bar{e}]$.

Proposition 27. *Problem $\mathscr{D}(0,e_0)$ is a special case of $\mathscr{S}(t_0,x_0)$ (5).*

Proof. Let $c_i = p_i(q_a(i)-q_s(i)+u(i))\Delta t$

$$d_i = \begin{cases} p_d(q_a(k)-q_s(k)+u_k) & k\in T_{on}\\ 0 & \text{otherwise.}\end{cases}$$ □

We conclude that our approach to solving NFS DP problems can be applied to battery scheduling. That is, battery scheduling can be represented as an augmented DP problem of Form $\mathscr{A}(t_0,x_0)$.

D. Numerically Solving The Deterministic Battery Scheduling Problem

Our proposed approximation scheme can be applied to solve the battery scheduling problem, $\mathscr{D}(0,e_0)$. This is done by creating an augmented state variable based on the maximum function in the objective function, as in Section III, and thus constructing an equivalent DP problem of the form $\mathscr{A}(0,x_0)$ (9); which is a special case of $\mathscr{P}(t_0,x_0)$. Figure 3 shows how the monthly cost decreases when we use policies constructed from the associated discretized DP problems, $\mathscr{P}_k(t_0,x_0)$, as k is increased. Although we do not get a monotonically decreasing sequence of costs, the error does decrease as $k\to\infty$. Figure 4 also shows that augmenting and then following our proposed discretization scheme for the battery scheduling problem results in a policy that reduces the consumption demand peak as k is increased. Figure 5 shows how the computational time required to solve the discretized battery scheduling problem appears to be of exponential nature with respect to the number of grid points.

We used solar and usage data obtained by local utility Salt River Project (SRP) in Tempe, AZ, for power variables q_s and q_a. We also use pricing data from SRP for the parameters p_{on}, p_{off} and p_d. Battery data obtained for the Tesla Powerwall was used to determine the parameters α, η, $\bar{u}$, $\underline{u}$ and $\bar{e}$. The results of the simulation are shown in Figure 6. The policy used for this simulation was created using our augmentation and approximation scheme with $k=20$. Interpolation was used

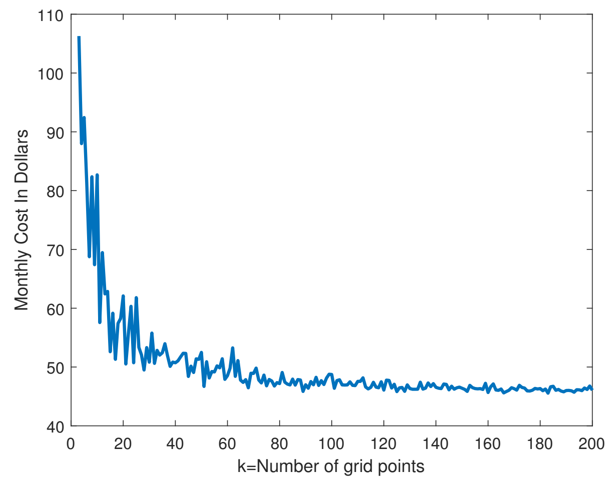

Figure 3. The resulting monthly cost from using the policy found by solving the discretized problem, of form $P_k(t_0,x_0)$, for optimal battery scheduling.

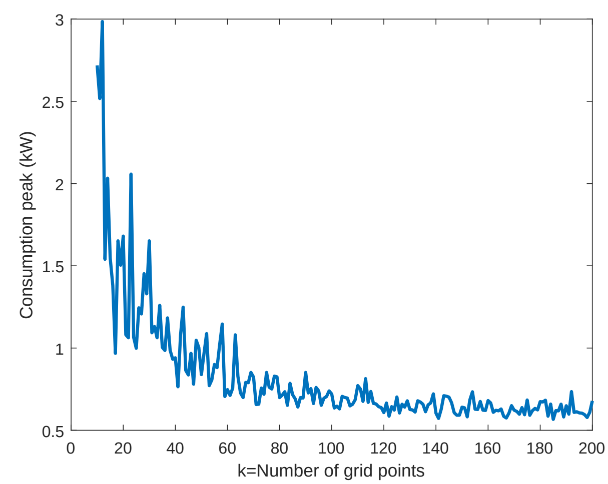

Figure 4. The resulting maximum demand from using the policy found by solving the discretized problem, of form $P_k(t_0,x_0)$, for optimal battery scheduling.

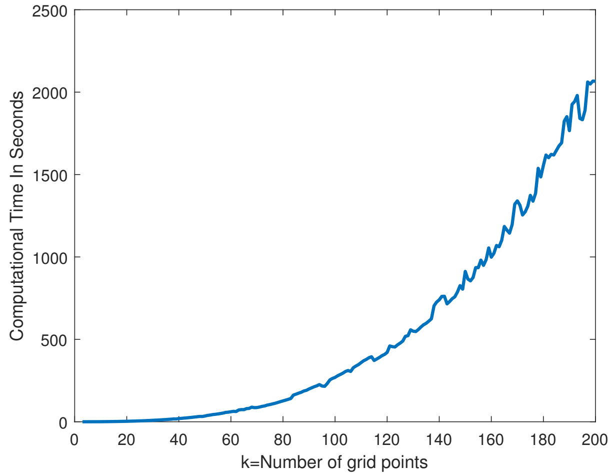

Figure 5. The computational time in seconds required to solve the discretized battery scheduling problem, of form $P_k(t_0, x_0)$.

Table II
LIST OF CONSTANT VALUES (PRICES CORRESPOND TO SALT RIVER PROJECT E21 PRICE PLAN)

Constant	Value	Constant	Value
α	0.999791667 (W/h)	t_{off}	41
η	0.92 (%)	p_{on}	0.0633×10^{-3} (\$/KWh)
$\bar{u}$	4000 (Wh)	p_{off}	0.0423×10^{-3} (\$/KWh)
$\underline{u}$	-4000 (Wh)	p_{d}	0.2973 (\$/KWh)
$\bar{e}$	8000 (Wh)	Δt	0.5 (h)
t_{on}	27		

to aid in solving Bellman's equation (4) and decrease the approximation error. These results show an improvement in accuracy over results obtained for a similar problem in [22] (approximately \$0.98 savings). As expected, we see the battery charges during off-peak and then discharges during on peak times to reduce ToU charges, while maintaining a reserve which it uses to keep consumption flat during on peak times, thereby minimizing the demand charge. As a result the power stabilizes during on peak times - becoming constant.

E. Solving The Stochastic Battery Scheduling Problem

To evaluate the effect of stochastic uncertainty on battery scheduling, we identify a Gauss-Markov model of solar generation based on SRP data. We construct the battery scheduling problem of the form $\mathscr{H}_s(t_0, x_0)$ (15) and then use our proposed state augmentation approach to construct an equivalent DP problem of form $\mathscr{Q}(t_0, x_0)$ (16). The problem of form $\mathscr{Q}(t_0, x_0)$ is then solved approximately using the methodology of Section VII-A.

F. Solar Generation Model

Our approach to modeling the dynamics of solar generation is based on [29]. Our Markov type model can be used to generate high resolution data over large time horizons. The Markov property of the model results in deviation from the mean being correlated time to time, helping represent the physical phenomena of clouds gradually passing overhead rather than instantaneously appearing.

Our model is a type of autoregressive-moving-average model (ARMAX) [30]. In [31] it is seen ARMAX models preform better than auto-regressive integrated moving average (ARIMA) and in [32] it is shown ARMAX models can produce data similar to real data for local sites in Califonia and Colorado.

Exogenous variables, temperature and humidity, are included as state variables in addition to the primary variable - solar radiance. Cross correlations between state variables are computed from data. Specifically, we take time-series data of these quantities, denoted $\mathbf{W}(t)$ and normalize this data as,

$$w_i(t) = \frac{W_i(t) - \mu_i(t)}{\sigma_i(t)},$$

where $\mu_i(t)$ is the average historic and clear-sky mean of the variable W_i at time step t and $\sigma_i(t)$ is the standard deviation of variable W_i at time step t.
The generating process is then given by:

$$\begin{aligned} &\mathbf{w}(t) = A\mathbf{w}(t-1) + B\mathbf{v}(t-1) \text{ for } t = 1,..,T \qquad (33)\\ &\text{where } \mathbf{w}(t) \in \mathbb{R}^3, \mathbf{w}(0) = \mathbf{0}\\ &\mathbf{v}(t) \sim \mathscr{N}(\mathbf{0}, I_{3\times 3}),\end{aligned}$$

where the matrices A and B are chosen to preserve the lag 0 and lag 1 cross-correlations seen in the collected data. Specifically, we can compute these matrices as ([29])

$$A = M_1 M_0^{-1} \qquad BB^T = M_0 - M_1 M_0^{-1} M_1^T, \qquad (34)$$

where M_i is the i-lag cross correlation matrix. So $(M_i)_{m,n} = \rho_i(m,n)$ where $\rho_i(m,n)$ is the cross-correlation coefficient between variables m and n with variable n lagged by i time steps. Then, adding back in the mean and deviation, we obtain the power supplied by solar at time step k as

$$q_s(k) = w_1(k)\sigma_1(k) + \mu_1(k).$$

Figure 7 shows simulated irradiance data from our solar model when compared to actual recorded irradiance data. For this numerical implementation the mean and standard deviation, $(\mu_i(t))_{0\le t\le T}$ and $(\sigma_i(t))_{0\le t\le T}$, were calculated using data from Wunderground for a weather station in Tempe, AZ on October the 15th 2014 for each state variable. Cross correlations between the variables were also calculated from the same data set and (34) was solved giving the matrices A and B in (33). As seen in Figure 7 this solar generation model gives an output similar to what is observed in real data. Next we incorporate this model into our battery scheduling DP problems.

Stochastic Battery Scheduling We now modify Problem $\mathscr{D}(0, e_0)$ (32) to give a stochastic version of the battery scheduling problem $\mathscr{D}_s(0, [e_0, 0])$,

$$\begin{aligned} \arg\min_{\pi\in\Pi}\ & \mathbb{E}_{[\mathbf{v}]_0^{T-1}} \Big[J_E(\Phi_{f,0}(\pi, [e_0, 0], T, [\mathbf{v}]_0^{T-1})) \qquad (35)\\ & + J_D(\Phi_{f,0}(\pi, [e_0, 0], T, [\mathbf{v}]_0^{T-1})) \Big]\\ \text{subject to: } & \psi_{f,0}(\pi, [e_0, 0], t, [\mathbf{v}]_0^{t-1}) \in E_t \times \mathbb{R}^3 \text{ for } t = 0,..,T\\ & \pi(x,t) \in U_t \text{ and } v(t) \in \mathbb{R}^3 \sim \mathscr{N}(\mathbf{0}, I_{3\times 3}) \forall x \in X_t, \forall t = 0,..,T-1,\end{aligned}$$

where J_E is the ToU cost function and J_D is the demand charge found in Section IX-A; $f([e,w],u,t,v) = \begin{bmatrix} \alpha(e + \eta u \Delta t) \\ Aw + Bv \end{bmatrix}$; $E_t =$

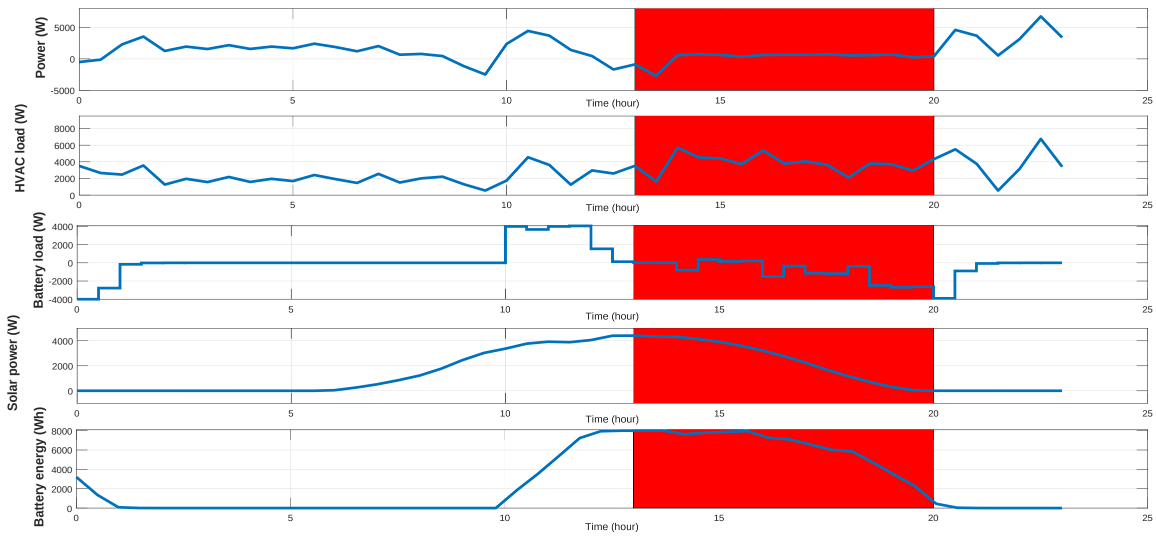

Figure 6. The trajectory the algorithm produces for deterministic solar data. The supremun of the power is 0.7033(kw) and the cost is \$46.389.

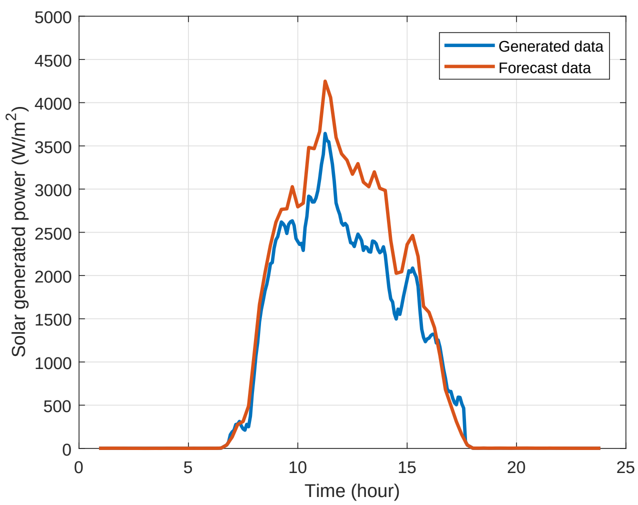

Figure 7. Solar data generated over 24 hours using data from Wunderground

$[\underline{e},\bar{e}]$ and $U_t = [\underline{u},\bar{u}]$ for all $t \in \{0,...,T\}$; ψ_{f,t_0} and Φ_{f,t_0} are the state and trajectory map respectively from Definition 21; $[\mathbf{v}]_0^{T-1} = [v(0),..,v(T-1)] \in \mathbb{R}^{3\times(T)}$; matrices A and B are calculated from weather data using equations (34); and all constants are found in Table II.

G. Numerically Solving The Stochastic Battery Scheduling Problem

Using the state augmentation procedure in Section III on the stochastic battery scheduling problem $\mathscr{D}_s(0,[e_0,0])$ (35), we may find a stochastic DP problem of the form $\mathscr{Q}(t_0,x_0)$ (16) such that an optimal policy for $\mathscr{D}_s(0,[e_0,0])$ can be constructed from an optimal policy of $\mathscr{Q}(t_0,x_0)$. We may then construct the approximated stochastic DP problem $\mathscr{Q}_{\varepsilon,k}(t_0,k)$ (25) and solve it using Bellman's equation (23). From an optimal policy of $\mathscr{Q}_{\varepsilon,k}(t_0,k)$ we then construct a feasible policy for $\mathscr{Q}(t_0,x_0)$ using (26). Figure 1 demonstrates a simulation of using the feasible policy obtained via augmenting and approximating the stochastic battery scheduling problem with a reasonably selected family of compact state spaces, $\{H_{\varepsilon,t}\}_{0\le t\le T}$, and discretization level $k=10$. To simplify computation we use a one state version of our solar model (33) and use interpolation while solving Bellman's equation. As expected the battery charges during the on peak times and conservatively discharges during the off-peak times. The solar data generated from this run are then used as input to the deterministic algorithm in order to compare performance. As anticipated, the deterministic case performs better than the stochastic case.

X. Conclusion

In this paper we propose a general formulation of the DP problem. We show that if the objective function is forward separable, DP problems may reformulated using state augmentation as an equivalent DP problem with additively separable objective function. Furthermore, we define a class of functions, called naturally forward separable (NFS) functions, such that DP problems with an objective function of this class can be tractably solved using state augmentation. Moreover, we show that the problem of optimal scheduling of battery storage in the presence of combined demand and time-of-use charges is a special case of this class of NFS DP problems. We further extend these results to stochastic DP problems with a NFS objective. The proposed algorithms are applied to a battery scheduling problem using first a deterministic and then Gauss-Markov model for solar generation and load.

Extensions of this work include the use of non-separable input constraints (such as those considered in [33]) and algorithms for finding the minimal dimension NFS representation of a given objective function.

XI. Appendix

Lemma 28. *Consider a DP problem of the form* $\mathscr{Q}(t_0,x_0)$ (16) *with additively separable objective function* $J_{t_0}^{\mathscr{Q}}$. *For any family of functions of the form* $\hat{\pi} : \mathbb{R}^{n\times(t-t_0+1)} \to \mathbb{R}^m$ *are such* $\hat{\pi}_t([(x(t_0),....,x(t))]) \in U_t$ *and* $f(x(t),\hat{\pi}_t([(x(t_0),....,x(t))]),t,v(t)) \in X_{t+1}$ *for all* $x(i) \in X_i$,

$i \in \{t_0,...,t\}$, $v(t) \in \mathbb{R}^q$ and $t \in \{t_0,...,T-1\}$ there exists $\alpha \in \Pi$ such that

$$\mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}} \left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\alpha, x_0, T, [\mathbf{v}]_{t_0}^{T-1})) \right) \qquad (36)$$
$$= \mathbb{E}_{[\mathbf{v}]_{t_0}^{T-1}} \left(J_{t_0}^{\mathscr{Q}}(\Phi_{f,t_0}(\hat{\pi}, x_0, T, [\mathbf{v}]_{t_0}^{T-1})) \right)$$

where we make a small abuse of notation to extend the trajectory map Φ_{f,t_0} to policies that use the entire state space history.

Proof. Proposition 8.1 [34] or Theorem 6.2 [7]. □

Acknowledgments

This research was supported by the NSF Grant CNS-1739990. We gratefully acknowledge data and expertise provided by Robert Hess of Salt River Project.

References

[1] R. E. Bellman, R. E. Kalaba, and T. Teichmann, "Dynamic programing," *Physics Today*, vol. 19, p. 98, 1966.

[2] R. H. Strotz, "Myopia and inconsistency in dynamic utility maximization," *The Review of Economic Studies*, vol. 23, no. 3, pp. 165–180, 1955.

[3] A. Shapiro, "On a time consistency concept in risk averse multistage stochastic programming," *Operations Research Letters*, vol. 37, no. 3, pp. 143–147, 2009.

[4] P. Carpentier, J.-P. Chancelier, G. Cohen, M. De Lara, and P. Girardeau, "Dynamic consistency for stochastic optimal control problems," *Annals of Operations Research*, vol. 200, no. 1, pp. 247–263, 2012.

[5] R. Weber, *Optimization and Control*. 2016.

[6] M. Sniedovich, *Eureka! Bellman's principle of optimality is valid!* Kluwer.

[7] K. Hinderer, *Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter*. Springer-Verlag.

[8] D. Li and Y. Y. Haimes, "Extension of dynamic programming to non-separable dynamic optimization problems," *Computers & Mathematics with Applications*, vol. 21, no. 11-12, pp. 51–56, 1991.

[9] W. B. Powell, *Approximate Dynamic Programming: Solving the curses of dimensionality*, vol. 703. John Wiley & Sons, 2007.

[10] E. Lourens, E. Reynders, G. De Roeck, G. Degrande, and G. Lombaert, "An augmented kalman filter for force identification in structural dynamics," *Mechanical Systems and Signal Processing*, vol. 27, pp. 446–460, 2012.

[11] J. Skaf, S. Boyd, and A. Zeevi, "Shrinking-horizon dynamic programming," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 17, pp. 1993–2002, 2010.

[12] N. Bäuerle and U. Rieder, "More risk-sensitive markov decision processes," *Mathematics of Operations Research*, vol. 39, no. 1, pp. 105–120, 2013.

[13] W. B. Haskell and R. Jain, "A convex analytic approach to risk-aware markov decision processes," *SIAM Journal on Control and Optimization*, vol. 53, no. 3, pp. 1569–1598, 2015.

[14] A. Domingo and M. Sniedovich, "Experiments with dynamic programming algorithms for nonseparable problems," *European Journal of Operational Research*, vol. 67, pp. 172–187, 1993.

[15] F. Dufour and T. Prieto-Rumeau, "Approximation of Markov decision processes with general state space," *J. Math. Anal. Appl.*, vol. 388, pp. 1254–1267, 2012.

[16] M. Jones and M. M.Peet, "A dynamic programming approach to evaluating multivariate gaussian probabilities," 2018.

[17] M. Jones and M. M. Peet, "Solving dynamic programming with supremum terms in the objective and application to optimal battery scheduling for electricity consumers subject to demand charges," in *Decision and Control (CDC), 2017 IEEE 56th Annual Conference on*, pp. 1323–1329, IEEE, 2017.

[18] D. Maly and K. Kwan, "Optimal battery energy storage system (bess) charge scheduling with dynamic programming," *IEE Proceedings-Science, Measurement and Technology*, vol. 142, no. 6, pp. 453–458, 1995.

[19] J. E. Braun and K.-H. Lee, "Assessment of demand limiting using building thermal mass in small commercial buildings," *ASHRAE Transactions*, vol. 112, no. 1, pp. 547–558, 2006.

[20] J. Cai, J. Braun, D. Kim, and J. Hu, "General approaches for determining the savings potential of optimal control for cooling in commercial buildings having both energy and demand charges," *Science and Technology for the Built Environment*, vol. 22, no. 6, pp. 733–750, 2016.

[21] A. Zeinalzadeh and V. Gupta, "Minimizing risk of load shedding and renewable energy curtailment in a microgrid with energy storage," *arXiv preprint arXiv:1611.08000*, 2016.

[22] R. Kamyar and M. M. Peet, "Multi-objective dynamic programming for constrained optimization of non-separable objective functions with application in energy storage," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*, pp. 5348–5353, IEEE, 2016.

[23] R. Kamyar and M. M. Peet, "Optimal thermostat programming and optimal electricity rates for customers with demand charges," in *American Control Conference (ACC), 2015*, pp. 4529–4535, IEEE, 2015.

[24] O. Hernandez-Lerma and J. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, vol. 42. 1999.

[25] D. Li and Y. Y. Haimes, "The envelope approach for multiobjeetive optimization problems," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 17, no. 6, pp. 1026–1038, 1987.

[26] J. Maidens, A. Packard, and M. Arcak, "Parallel dynamic programming for optimal experiment design in nonlinear systems," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 2894–2899, IEEE, 2016.

[27] J. Rust, "Using randomization to break the curse of dimensionality," *Econometrica: Journal of the Econometric Society*, pp. 487–516, 1997.

[28] E. Porteus, "An informal look at the principle of optimality," *Managment Science*, vol. 21, 1975.

[29] C. W. Richardson, "Stochastic simulation of daily precipitation, temperature, and solar radiation," *Water resources research*, vol. 17, no. 1, pp. 182–190, 1981.

[30] C. Wan and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *Journal of power and energy systems*, vol. 1, 2015.

[31] Y. Li and L. Shu, "An ARMAX model for forecasting the power output of a grid connected photovoltaic system," *Renewable Energy*, vol. 60, pp. 78–89, 2014.

[32] Y. Chen and L. Xie, "Multitime-scale data-driven spatiotemporal forecast of photovoltaic generation," *IEEE Transactions on Sustainable Energy*, vol. 6, pp. 104–112, 2015.

[33] S. Mannor, O. Mebel, and H. Xu, "Robust mdps with k-rectangular uncertainty," *Mathematics of Operations Research*, vol. 41, no. 4, pp. 1484–1509, 2016.

[34] D. Bertsekas and S. Shreve, *Stochastic Optiamal Control: The Discrete-Time case*. Athena Scientific, 1978.

Morgan Jones received the B.S. and Mmath in mathematics from The University of Oxford, England in 2016. He is a research associate with Cybernetic Systems and Controls Laboratory (CSCL) in the School for Engineering of Matter, Transport and Energy (SEMTE) at Arizona State University. His research primarily focuses on the estimation of attractors, reachable sets and regions of attraction for nonlinear ODE's. Furthermore, He has been studying the effects of optimal energy storage in smart grid environments.

Matthew M. Peet received the B.S. degree in physics and in aerospace engineering from the University of Texas, Austin, TX, USA, in 1999 and the M.S. and Ph.D. degrees in aeronautics and astronautics from Stanford University, Stanford, CA, in 2001 and 2006, respectively. He was a Postdoctoral Fellow at INRIA, Paris, France from 2006 to 2008. He was an Assistant Professor of Aerospace Engineering at the Illinois Institute of Technology, Chicago, IL, USA, from 2008 to 2012. Currently, he is an Associate Professor of Aerospace Engineering at Arizona State University, Tempe, AZ, USA. Dr. Peet received a National Science Foundation CAREER award in 2011.