Learning Aided Optimization for Energy Harvesting Devices with Outdated State Information

Hao Yu and Michael J. Neely
University of Southern California

***Abstract*—This paper considers utility optimal power control for energy harvesting wireless devices with a finite capacity battery. The distribution information of the underlying wireless environment and harvestable energy is unknown and only outdated system state information is known at the device controller. This scenario shares similarity with Lyapunov opportunistic optimization and online learning but is different from both. By a novel combination of Zinkevich's online gradient learning technique and the drift-plus-penalty technique from Lyapunov opportunistic optimization, this paper proposes a learning-aided algorithm that achieves utility within $O(\epsilon)$ of the optimal, for any desired $\epsilon > 0$, by using a battery with an $O(1/\epsilon)$ capacity. The proposed algorithm has low complexity and makes power investment decisions based on system history, without requiring knowledge of the system state or its probability distribution.**

I. Introduction

Energy harvesting can enable self-sustainable and perpetual wireless devices. By harvesting energy from the environment and storing it in a battery for future use, we can significantly improve energy efficiency and device lifetime. Harvested energy can come from solar, wind, vibrational, thermal, or even radio sources [1], [2], [3]. Energy harvesting has been identified as a key technology for wireless sensor networks [4], internet of things (IoT) [5], and 5G communication networks [6]. However, the development of harvesting algorithms is complex because the harvested energy is highly dynamic and the device environment and energy needs are also dynamic. Efficient algorithms should learn when to take energy from the battery to power device tasks that bring high utility, and when to save energy for future use.

There have been large amounts of work developing efficient power control policies to maximize the utility of energy harvesting devices. In the highly ideal case where the future system state (both the wireless channel sate and energy harvesting state) can be perfectly predicted, optimal power control strategies that maximize the throughput of wireless systems are considered in [7], [8]. In a more realistic case with only the statistics and causal knowledge of the system state, power control policies based on Markov Decision Processes (MDP) are considered in [9], [10]. In the case when the statistical knowledge is unavailable but the current system state is observable, work [11] develops suboptimal power control policies based on approximation algorithms.

Hao Yu and Michael J. Neely are with the Department of Electrical Engineering, University of Southern California, Los Angeles, USA. This work is supported in part by grant NSF CCF-1718477.

However, there is little work on the challenging scenario where neither the distribution information nor the system state information are known. In practice, the amount of harvested energy on each slot is known to us only after it arrives and is stored into the battery. Further, the wireless environment is often unknown before the power action is chosen. For example, the wireless channel state in a communication link is measured at the receiver side and then reported back to the transmitter with a time delay. If the fading channel varies very fast, the channel state feedback received at the transmitter can be outdated. Another example is power control for sensor nodes that detect unknown targets where the state of targets is known only after the sensing action is performed.

In this paper, we consider utility-optimal power control in an energy harvesting wireless device with outdated state information and unknown state distribution information. This problem setup is closely related to but different from the Lyapunov opportunistic power control considered in works [12], [13], [14] with instantaneous wireless channel state information. The policies developed in [12], [13], [14] are allowed to adapt their power actions to the instantaneous system states on each slot, which are unavailable in our problem setup. The problem setup in this paper is also closely related to online convex optimization where control actions are performed without knowing instantaneous system states [15], [16], [17]. However, existing methods for online convex learning require the control actions to be chosen from a fixed set. This does not hold in our problem since the power to be used can only be drained from the battery whose backlog is time-varying and dependent on previous actions.

By combining the drift-plus-penalty (DPP) technique for Lyapunov opportunistic optimization [18] and the online gradient learning technique for online convex optimization [15], we develop a novel learning aided dynamic power control algorithm that can achieve an $O(\epsilon)$ optimal utility by using a battery with an $O(1/\epsilon)$ capacity for energy harvesting wireless devices with outdated state information.

II. Problem Formulation

Consider an energy harvesting wireless device that operates in normalized time slots $t \in \{1, 2, \ldots\}$. Let $\omega[t] = (e[t], \mathbf{s}[t]) \in \Omega$ represent the system state on each slot t, where

- $e[t]$ is the amount of harvested energy for slot t (for example, through solar, wind, radio signal, and so on).
- $\mathbf{s}[t]$ is the wireless device state on slot t (such as the vector of channel conditions over multiple subbands).

- Ω is the state space for all $\omega[t] = (e[t], \mathbf{s}[t])$ states.

Assume $\{\omega[t]\}_{t=1}^{\infty}$ evolves in an independent and identically distributed (i.i.d.) manner according to an unknown distribution. Further, the state $\omega[t]$ is *unknown* to the device *until the end of slot* t. The device is powered by a finite-size battery. At the beginning of each slot $t \in \{1, 2, \ldots\}$, the device draws energy from the battery and allocates it as an n-dimensional power decision vector $\mathbf{p}[t] = [p_1[t], \ldots, p_n[t]]^{\mathsf{T}} \in \mathcal{P}$ where $\mathcal{P}$ is a compact convex set given by

$$\mathcal{P} = \{\mathbf{p} \in \mathbb{R}^n : \sum_{i=1}^{n} p_i \leq p^{\max}, p_i \geq 0, \forall i \in \{1, 2, \ldots, n\}\}.$$

Note that $p^{\max}$ is a given positive constant (restricted by hardware) and represents the maximum total power that can be used on each slot. The device receives a corresponding utility $U(\mathbf{p}[t]; \omega[t])$. Since $\mathbf{p}[t]$ is chosen without knowledge of $\omega[t]$, the achieved utility is unknown until the end of slot t. For each $\omega \in \Omega$, the utility function $U(\mathbf{p}; \omega)$ is assumed to be continuous and concave over $\mathbf{p} \in \mathcal{P}$. An example is:

$$U(\mathbf{p}; \omega) = \sum_{i=1}^{n} \log(1 + p_i[t] s_i[t]) \tag{1}$$

where $\mathbf{s}[t] = (s_1[t], \ldots, s_n[t])$ is the vector of (unknown) channel conditions over n orthogonal subbands available to the wireless device. In this example, $p_i[t]$ represents the amount of power invested over subband i in a rateless coding transmission scenario, and $U(p[t]; \omega[t])$ is the total throughput achieved on slot t. We focus on fast time-varying wireless channels, e.g., communication scenarios with high mobility transceivers, where $\mathbf{s}[t]$ known at the transmitter is outdated since $\mathbf{s}[t]$ must be measured at the receiver side and then reported back to the transmitter with a time delay.

A. Further examples

The above formulation admits a variety of other useful application scenarios. For example, it can be used to treat power control in cognitive radio systems. Suppose an energy limited secondary user harvests energy and operates over licensed spectrum occupied by primary users. In this case, $\mathbf{s}[t] = (s_1[t], \ldots, s_n[t])$ represents the channel activity of primary users over each subband. Since primary users are not controlled by the secondary user, $\mathbf{s}[t]$ is only known to the secondary user at the end of slot t.

Another application is a wireless sensor system. Consider an energy harvesting sensor node that collects information by detecting an unpredictable target. In this case, $\mathbf{s}[t]$ can be the state or action of the target on slot t. By using $\mathbf{p}[t]$ power for signaling and sensing, we receive utility $U(\mathbf{p}[t]; \omega[t])$, which depends on state $\omega[t]$. For example, in a monitoring system, if the monitored target performs an action $\mathbf{s}[t]$ that we are not interested in, then the reward $U(\mathbf{p}[t]; \omega[t])$ by using $\mathbf{p}[t]$ is small. Note that $\mathbf{s}[t]$ is typically unknown to us at the beginning of slot t and is only disclosed to us at the end of slot t.

B. Basic assumption

Assumption 1.

- *There exist a constant* $e^{\max} > 0$ *such that* $0 \leq e[t] \leq e^{\max}, \forall t \in \{1, 2, \ldots\}$.
- *Let* $\nabla_{\mathbf{p}} U(\mathbf{p}; \omega)$ *denote a subgradient (or gradient if* $U(\mathbf{p}; \omega)$ *is differentiable) vector of* $U(\mathbf{p}; \omega)$ *with respect to* $\mathbf{p}$ *and let* $\frac{\partial}{\partial p_i} U(\mathbf{p}; \omega), \forall i \in \{1, 2, \ldots, n\}$ *denote each component of vector* $\nabla_{\mathbf{p}} U(\mathbf{p}; \omega)$. *There exist positive constants* $D_1, \ldots, D_n$ *such that* $|\frac{\partial}{\partial p_i} U(\mathbf{p}; \omega)| \leq D_i, \forall i \in \{1, 2, \ldots, n\}$ *for all* $\omega \in \Omega$ *and all* $\mathbf{p} \in \mathcal{P}$. *This further implies there exists* $D > 0$, *e.g.,* $D = \sqrt{\sum_{i=1}^{n} D_i^2}$, *such that* $\|\nabla_{\mathbf{p}} U(\mathbf{p}; \omega)\| \leq D$ *for all* $\omega \in \Omega$ *and all* $\mathbf{p} \in \mathcal{P}$, *where* $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2}$ *is the standard* l_2 *norm.*

Such constants $D_1, \ldots, D_n$ exist in most cases of interest, such as for utility functions (1) with bounded $s_i[t]$ values. [1]

C. Power control and energy queue model

The finite size battery can be considered as backlog in an *energy queue*. Let $E[0]$ be the initial energy backlog in the battery and $E[t]$ be the energy stored in the battery at the **end** of slot t. The power vector $\mathbf{p}[t]$ must satisfy the following *energy availability constraint*:

$$\textstyle\sum_{i=1}^{n} p_i[t] \leq E[t-1], \forall t \in \{1, 2, \ldots\}. \tag{2}$$

which requires the consumed power to be no more than what is available in the battery.

Let $E^{\max}$ be the maximum capacity of the battery. If the energy availability constraint (2) is satisfied on each slot, the energy queue backlog $E[t]$ evolves as follows:

$$E[t] = \min\{E[t-1] - \textstyle\sum_{i=1}^{n} p_i[t] + e[t], E^{\max}\}, \forall t. \tag{3}$$

D. An upper bound problem

Let $\omega[t] = (e[t], s[t])$ be the random state vector on slot t. Let $\mathbb{E}[e] = \mathbb{E}[e[t]]$ denote the expected amount of new energy that arrives in one slot. Define a function $h : \mathcal{P} \to \mathbb{R}$ by

$$h(\mathbf{p}) = \mathbb{E}[U(\mathbf{p}; \omega[t])].$$

Since $U(\mathbf{p}; \omega)$ is concave in $\mathbf{p}$ and has bounded gradients/subgradients for each $\omega \in \Omega$ by Assumption 1, it can be shown that $h(\mathbf{p})$ is concave and continuous.

The function h is typically unknown because the distribution of ω is unknown. However, to establish a fundamental bound, suppose both h and $\mathbb{E}[e]$ are known and consider choosing a fixed vector $\mathbf{p}$ to solve the following deterministic problem:

$$\max_{\mathbf{p}} \ h(\mathbf{p}) \tag{4}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} p_i - \mathbb{E}[e] \leq 0 \tag{5}$$

$$\mathbf{p} \in \mathcal{P} \tag{6}$$

where constraint (5) requires that the consumed energy is no more than $\mathbb{E}[e]$.

[1] This is always true when $s_i[t]$ are wireless signal strength attenuations.

Let $\mathbf{p}^*$ be an optimal solution of problem (4)-(6) and U^* be its corresponding utility value of (4). Define a *causal policy* as one that, on each slot t, selects $\mathbf{p}[t] \in \mathcal{P}$ based only on information up to the start of slot t (in particular, without knowledge of $\omega[t]$). Since $\omega[t]$ is i.i.d. over slots, any causal policy must have $\mathbf{p}[t]$ and $\omega[t]$ independent for all t. The next lemma shows that no causal policy $\mathbf{p}[t], t \in \{1, 2, \ldots\}$ satisfying (2)-(3) can attain a better utility than U^*.

Lemma 1. *Let $\mathbf{p}[t] \in \mathcal{P}, t \in \{1, 2, \ldots\}$ be yielded by any causal policy that consumes less energy than it harvests in the long term, so $\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\sum_{i=1}^{n} p_i[t]\right] \leq \mathbb{E}[e]$. Then,*

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[U(\mathbf{p}[t]; \omega[t])] \leq U^*.$$

Proof. Fix a slot $t \in \{1, 2, \ldots\}$. Then

$$\mathbb{E}\left[U(\mathbf{p}[t]; \omega[t])\right] \overset{(a)}{=} \mathbb{E}\left[\mathbb{E}\left[U(\mathbf{p}[t]; \omega[t])|\mathbf{p}[t]\right]\right] \overset{(b)}{=} \mathbb{E}\left[h(\mathbf{p}[t])\right] \tag{7}$$

where (a) holds by iterated expectations; (b) holds because $\mathbf{p}[t]$ and $\omega[t]$ are independent (by causality).

For each $T > 0$ define $\bar{\mathbf{p}}[T] = [\bar{p}_1[T], \ldots, \bar{p}_n[T]]^{\mathsf{T}}$ with

$$\bar{p}_i[T] = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[p_i[t]\right], \forall i \in \{1, 2, \ldots, n\}.$$

We know by assumption that:

$$\limsup_{T\to\infty} \sum_{i=1}^{n} \bar{p}_i[T] \leq \mathbb{E}[e] \tag{8}$$

Further, since $\mathbf{p}[t] \in \mathcal{P}$ for all slots t, it holds that $\bar{\mathbf{p}}[T] \in \mathcal{P}$ for all $T > 0$. Also,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[U(\mathbf{p}[t]; \omega[t])] &\overset{(a)}{=} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[h(\mathbf{p}[t])\right] \\ &\overset{(b)}{\leq} h\left(\mathbb{E}\left[\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\mathbf{p}[t]\right]\right) \\ &= h(\bar{\mathbf{p}}[T])\end{aligned}$$

where (a) holds by (7); (b) holds by Jensen's inequality for the concave function h. It follows that:

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[U(\mathbf{p}[t]; \omega[t])] \leq \limsup_{T\to\infty} h(\bar{\mathbf{p}}[T]).$$

Define $\theta = \limsup_{T\to\infty} h(\bar{\mathbf{p}}[T])$. It suffices to show that $\theta \leq U^*$. Since $\bar{\mathbf{p}}[T]$ is in the compact set $\mathcal{P}$ for all $T > 0$, the Bolzano-Wierstrass theorem ensures there is a subsequence of times T_k such that $\bar{\mathbf{p}}[T_k]$ converges to a fixed vector $\mathbf{p}_0 \in \mathcal{P}$ and $h(\bar{\mathbf{p}}[T_k])$ converges to θ as $k \to \infty$:

$$\lim_{k\to\infty} \bar{\mathbf{p}}[T_k] = \mathbf{p}_0 \in \mathcal{P}$$
$$\lim_{k\to\infty} h(\bar{\mathbf{p}}[T_k]) = \theta$$

Continuity of h implies that $h(\mathbf{p}_0) = \theta$. By (8) the vector $\mathbf{p}_0 = [p_{0,1}, \ldots, p_{0,n}]^{\mathsf{T}}$ must satisfy $\sum_{i=1}^{n} p_{0,i} \leq \mathbb{E}[e]$. Hence, $\mathbf{p}_0$ is a vector that satisfies constraints (5)-(6) and achieves utility $h(\mathbf{p}_0) = \theta$. Since U^* is defined as the optimal utility value to problem (4)-(6), it holds that $\theta \leq U^*$. □

Note that the U^* utility upper bound of Lemma 1 holds for any policy that consumes no more energy than it harvests in the long term. Policies that satisfy the physical battery constraints (2)-(3) certainly consume no more energy than harvested in the long term. However, Lemma 1 even holds for policies that violate these physical battery constraints. For example, U^* is still a valid bound for a policy that is allowed to "borrow" energy from an external power source when its battery is empty and "return" energy when its battery is full.

III. New Algorithm

This subsection proposes a new learning aided dynamic power control algorithm that chooses power control actions based on system history, without requiring the current system state or its probability distribution.

A. New Algorithm

Algorithm 1 New Algorithm

Let $V > 0$ be a constant algorithm parameter. Initialize virtual battery queue variable $Q[0] = 0$. Choose $\mathbf{p}[1] = [0, 0, \ldots, 0]^{\mathsf{T}}$ as the power action at slot 1. At the *end* of each slot $t \in \{1, 2, \ldots\}$, observe $\omega[t] = (e[t], \mathbf{s}[t])$ and do the following:

- **Update virtual battery queue** $Q[t]$: Update $Q[t]$ via:

$$Q[t] = \min\{Q[t-1] + e[t] - \sum_{i=1}^{n} p_i[t], 0\}. \tag{9}$$

- **Power control**: Choose

$$\mathbf{p}[t+1] = \text{Proj}_{\mathcal{P}}\Big\{\mathbf{p}[t] + \frac{1}{V}\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t]) + \frac{1}{V^2}Q[t]\mathbf{1}\Big\} \tag{10}$$

as the power action for the next slot $t+1$ where $\text{Proj}_{\mathcal{P}}\{\cdot\}$ represents the projection onto set $\mathcal{P}$, $\mathbf{1}$ denotes a column vector of all ones and $\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])$ represents a subgradient (or gradient if $U(\mathbf{p}; \omega[t])$ is differentiable) vector of function $U(\mathbf{p}; \omega[t])$ at point $\mathbf{p} = \mathbf{p}[t]$. Note that $\mathbf{p}[t]$, $Q[t]$ and $\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])$ are given constants in (10).

The new dynamic power control algorithm is described in Algorithm 1. At the end of slot t, Algorithm 1 chooses $\mathbf{p}[t+1]$ based on $\omega[t]$ without requiring $\omega[t+1]$. To enable these decisions, the algorithm introduces a (nonpositive) *virtual* battery queue process $Q[t] \leq 0$, which shall later be shown to be related to a shifted version of the physical battery queue $E[t]$.

Note that Algorithm 1 does not explicitly enforce the energy availability constraint (2). Let $\mathbf{p}[t+1]$ be given by (10), one may expect to use

$$\hat{\mathbf{p}}[t+1] = \frac{\min\{\sum_{i=1}^{n} p_i[t+1], E[t]\}}{\sum_{i=1}^{n} p_i[t+1]}\mathbf{p}[t+1] \tag{11}$$

that scales down $\mathbf{p}[t+1]$ to enforce the energy availability constraint (2). However, our analysis in Section IV shows that

if the battery capacity is at least as large as an $O(V)$ constant, then directly using $\mathbf{p}[t+1]$ from (10) is ensured to always satisfy the energy availability constraint (2). Thus, there is no need to take the additional step (11).

B. Algorithm Inuitions

Lemma 2. *The power control action* $\mathbf{p}[t+1]$ *chosen in* (10) *is to solve the following quadratic convex program*

$$\max_{\mathbf{p}} \; V(\nabla_{\mathbf{p}} U(\mathbf{p}[t];\omega[t]))^{\mathsf{T}}(\mathbf{p}-\mathbf{p}[t]) + Q[t]\mathbf{1}^{\mathsf{T}}\mathbf{p} - \frac{V^2}{2}\|\mathbf{p}-\mathbf{p}[t]\|^2 \quad (12)$$

$$\textit{s.t.} \quad \mathbf{p} \in \mathcal{P} \quad (13)$$

Proof. By the definition of projection, equation (10) is to solve $\min_{\mathbf{p}\in\mathcal{P}} \|\mathbf{p} - \big(\mathbf{p}[t] + \frac{1}{V}\nabla_{\mathbf{p}} U(\mathbf{p}[t];\omega[t]) + \frac{1}{V^2}Q[t]\mathbf{1}\big)\|^2$. By expanding the square, eliminating constant terms and converting the minimization to a maximization of its negative object, it is easy to show this problem is equivalent to problem (12)-(13). □

The convex projection (10), or equivalently, the quadratic convex program (12)-(13) can be easily solved. See e.g., Lemma 3 in [19] for an algorithm that solves an n-dimensional quadratic program over set $\mathcal{P}$ with complexity $O(n\log n)$. Thus, the overall complexity of Algorithm 1 is low.

1) *Connections with the drift-plus-penalty (DPP) technique for Lyapunov opportunistic optimization*: The Lyapunov opportunistic optimization solves stochastic optimization without distribution information by developing dynamic policies that adapt control actions to the current system state [20], [21], [22], [23], [24], [18]. The dynamic policy from Lyapunov opportunistic optimization can be interpreted as choosing control actions to maximize a DPP expression on each slot. Unfortunately, the problem considered in this paper is different from the conventional Lyapunov opportunistic optimization problem since the power decision cannot be adapted to the unknown current system state. Nevertheless, if we treat $V(\nabla_{\mathbf{p}} U(\mathbf{p}[t];\omega[t]))^{\mathsf{T}}(\mathbf{p}-\mathbf{p}[t]) - \frac{V^2}{2}\|\mathbf{p}-\mathbf{p}[t]\|^2$ as a penalty term and $Q[t]\mathbf{1}^{\mathsf{T}}\mathbf{p}$ as a drift term, then Lemma 2 suggests that the power control in Algorithm 1 can still be interpreted as maximizing a (different) DPP expression. However, this DPP expression is significantly different from those conventional ones used in Lyapunov opportunistic optimization [18]. Also, the penalty term $VU(\mathbf{p}[t+1];\omega[t+1])$ used in conventional Lyapunov opportunistic optimization of [18] is unavailable in our problem since it depends on the unknown $\omega[t+1]$.
2) *Connections with online convex learning*: Online convex learning is a multi-round process where a decision maker selects its action from a *fixed* set at each round before observing the corresponding utility function [15], [16], [17]. If we assume the wireless device is equipped with an external free power source with infinite energy, i.e., the energy availability constraint (2) is dropped, then the problem setup in this paper is similar to an online learning problem where the decision maker selects $\mathbf{p}[t+1] \in \mathcal{P}$ on each slot $t+1$ to maximize an unknown reward function $U(\mathbf{p}[t+1];\omega[t+1])$ based on the information of previous reward functions $U(\mathbf{p}[\tau];\omega[\tau]), \tau \in \{1,\ldots,t\}$. In this case, Zinkevich's online gradient method [15], given by

$$p[t+1] = \text{Proj}_{\mathcal{P}}\{\mathbf{p}[t] + \gamma\nabla_{\mathbf{p}} U(\mathbf{p}[t];\omega[t])\} \quad (14)$$

where γ is a learning rate parameter, can solve this idealized problem. In fact, if we ignore $\frac{1}{V^2}Q[t]\mathbf{1}$ involved in (10), then (10) is identical to Zinkevich's learning algorithm with $\gamma = 1/V$. However, Zinkevich's algorithm and its variations [15], [25], [17] require actions to be chosen from a *fixed* set. Our problem requires $\mathbf{p}[t]$ chosen on each slot t to satisfy the energy availability constraint (2), which is time-varying since $E[t]$ evolves over time based on random energy arrivals and previous power allocation decisions.

Now, it is clear why Algorithm 1 is called a learning aided dynamic power control algorithm: Algorithm 1 can be viewed as an enhancement of the DPP technique originally developed for Lyapunov opportunistic optimization by replacing its penalty term with an expression used in Zinkevich's online gradient learning.

C. Main Results

While the above subsection provides intuitive connections to prior work, note that existing techniques cannot be applied to our problem. The next section develops a novel performance analysis (summarized in Theorems 1 and 3) to show that if $E[0] = E^{\max} = O(V)$, then the power control actions from Algorithm 1 are ensured to satisfy the energy availability constraint (2) and achieve

$$\frac{1}{t}\sum_{\tau=1}^{t}\mathbb{E}[U(\mathbf{p}[\tau];\omega[\tau])] \geq U^* - O(\frac{V}{t}) - O(\frac{1}{V}).$$

That is, for any desired $\epsilon > 0$, by choosing $V = 1/\epsilon$ in Algorithm 1, we can attain an $O(\epsilon)$ optimal utility for all $t \geq \Omega(\frac{1}{\epsilon^2})$ by using a battery with capacity $O(1/\epsilon)$.

IV. Performance Analysis of Algorithm 1

This section shows Algorithm 1 can attain an $O(\epsilon)$ close-to-optimal utility by using a battery with capacity $O(1/\epsilon)$.

A. Drift Analysis

Define $L[t] = \frac{1}{2}(Q[t])^2$ and call it a *Lyapunov function*. Define the *Lyapunov drift* as

$$\Delta[t] = L[t+1] - L[t]$$

Lemma 3. *Under Algorithm 1, for all* $t \geq 0$*, the Lyapunov drift satisfies*

$$\Delta[t] \leq Q[t](e[t+1] - \sum_{i=1}^{n} p_i[t+1]) + \frac{1}{2}B \quad (15)$$

with constant $B = (\max\{e^{\max}, p^{\max}\})^2$*, where* $e^{\max}$ *is the constant defined in Assumption 1.*

Proof. Fix $t \geq 0$. Recall that for any $x \in \mathbb{R}$ if $y = \min\{x, 0\}$ then $y^2 \leq x^2$. It follows from (9) that

$$(Q[t+1])^2 \leq (Q[t] + e[t+1] - \sum_{i=1}^{n} p_i[t+1])^2.$$

Expanding the square on the right side, dividing both sides by 2 and rearranging terms yields $\Delta[t] \leq Q[t](e[t+1] - \sum_{i=1}^{n} p_i[t+1]) + \frac{1}{2}(e[t+1] - \sum_{i=1}^{n} p_i[t+1])^2$.

This lemma follows by noting that $|e[t+1] - \sum_{i=1}^{n} p_i[t+1]| \leq \max\{e^{\max}, p^{\max}\}$ since $0 \leq \sum_{i=1}^{n} p_i[t+1] \leq p^{\max}$ and $0 \leq e[t+1] \leq e^{\max}$. □

Recall that a function $f : \mathcal{Z} \mapsto \mathbb{R}$ is said to be *strongly concave* with modulus α if there exists a constant $\alpha > 0$ such that $f(\mathbf{z}) + \frac{1}{2}\alpha\|\mathbf{z}\|^2$ is concave on $\mathcal{Z}$. It is easy to show that if $f(\mathbf{z})$ is concave and $\alpha > 0$, then $f(\mathbf{z}) - \frac{\alpha}{2}\|\mathbf{z} - \mathbf{z}_0\|^2$ is strongly concave with modulus α for any constant $\mathbf{z}_0$. The maximizer of a strongly concave function satisfies the following lemma:

Lemma 4 (Corollary 1 in [26])**.** *Let $\mathcal{Z} \subseteq \mathbb{R}^n$ be a convex set. Let function f be strongly concave on $\mathcal{Z}$ with modulus α and $\mathbf{z}^{opt}$ be a global maximum of h on $\mathcal{Z}$. Then, $f(\mathbf{z}^{opt}) \geq f(\mathbf{z}) + \frac{\alpha}{2}\|\mathbf{z}^{opt} - \mathbf{z}\|^2$ for all $\mathbf{z} \in \mathcal{Z}$.*

Lemma 5. *Let U^* be the utility upper bound defined in Lemma 1 and $\mathbf{p}^*$ be an optimal solution to problem* (4)-(6) *that attains U^*. At each iteration $t \in \{1, 2, \ldots\}$, Algorithm 1 guarantees*

$$V\mathbb{E}[U(\mathbf{p}[t]; \omega[t])] - \Delta[t] \geq VU^* + \frac{V^2}{2}\mathbb{E}[\Phi[t]] - \frac{D^2 + B}{2}$$

where $\Phi[t] = \|\mathbf{p}^* - \mathbf{p}[t+1]\|^2 - \|\mathbf{p}^* - \mathbf{p}[t]\|^2$, D is the constant defined in Assumption 1 and B is the constant defined in Lemma 3.

Proof. Note that $\sum_{i=1}^{n} p_i^* \leq \mathbb{E}[e]$. Fix $t \in \{1, 2, \ldots\}$. Note that $V(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t]))^{\mathsf{T}}(\mathbf{p} - \mathbf{p}[t]) + Q[t]\sum_{i=1}^{n} p_i$ is a linear function with respect to $\mathbf{p}$. It follows that

$$V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p} - \mathbf{p}[t]) + Q[t]\sum_{i=1}^{n} p_i - \frac{V^2}{2}\|\mathbf{p} - \mathbf{p}[t]\|^2 \tag{16}$$

is strongly concave with respect to $\mathbf{p} \in \mathcal{P}$ with modulus V^2. Since $\mathbf{p}[t+1]$ is chosen to maximize (16) over all $\mathbf{p} \in \mathcal{P}$, and since $\mathbf{p}^* \in \mathcal{P}$, by Lemma 4 we have

$$\begin{aligned}
&V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p}[t+1] - \mathbf{p}[t]) + Q[t]\sum_{i=1}^{n} p_i[t+1] \\
&- \frac{V^2}{2}\|\mathbf{p}[t+1] - \mathbf{p}[t]\|^2 \\
\geq& V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p}^* - \mathbf{p}[t]) + Q[t]\sum_{i=1}^{n} p_i^* \\
&- \frac{V^2}{2}\|\mathbf{p}^* - \mathbf{p}[t]\|^2 + \frac{V^2}{2}\|\mathbf{p}^* - \mathbf{p}[t+1]\|^2 \\
=& V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p}^* - \mathbf{p}[t]) + Q[t]\sum_{i=1}^{n} p_i^* + \frac{V^2}{2}\Phi[t].
\end{aligned}$$

Subtracting $Q[t]e[t+1]$ from both sides and rearranging terms yields

$$\begin{aligned}
&V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p}[t+1] - \mathbf{p}[t]) \\
&+ Q[t](\sum_{i=1}^{n} p_i[t+1] - e[t+1]) \\
\geq& V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p}^* - \mathbf{p}[t]) + Q[t](\sum_{i=1}^{n} p_i^* - e[t+1]) \\
&+ \frac{V^2}{2}\Phi[t] + \frac{V^2}{2}\|\mathbf{p}[t+1] - \mathbf{p}[t]\|^2.
\end{aligned}$$

Adding $VU(\mathbf{p}[t]; \omega[t])$ to both sides and noting that $U(\mathbf{p}[t]; \omega[t]) + (\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t]))^{\mathsf{T}}(\mathbf{p}^* - \mathbf{p}[t]) \geq U(\mathbf{p}^*; \omega[t])$ by the concavity of $U(\mathbf{p}; \omega[t])$ yields

$$\begin{aligned}
&VU(\mathbf{p}[t]; \omega[t]) + V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p}[t+1] - \mathbf{p}[t]) \\
&+ Q[t](\sum_{i=1}^{n} p_i[t+1] - e[t+1]) \\
\geq& VU(\mathbf{p}^*; \omega[t]) + Q[t](\sum_{i=1}^{n} p_i^* - e[t+1]) + \frac{V^2}{2}\Phi[t] \\
&+ \frac{V^2}{2}\|\mathbf{p}[t+1] - \mathbf{p}[t]\|^2.
\end{aligned}$$

Rearranging terms yields

$$\begin{aligned}
&VU(\mathbf{p}[t]; \omega[t]) + Q[t](\sum_{i=1}^{n} p_i[t+1] - e[t+1]) \\
\geq& VU(\mathbf{p}^*; \omega[t]) + Q[t](\sum_{i=1}^{n} p_i^* - e[t+1]) + \frac{V^2}{2}\Phi[t] \\
&+ \frac{V^2}{2}\|\mathbf{p}[t+1] - \mathbf{p}[t]\|^2 \\
&- V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p}[t+1] - \mathbf{p}[t])
\end{aligned} \tag{17}$$

Note that

$$\begin{aligned}
&V\big(\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\big)^{\mathsf{T}}(\mathbf{p}[t+1] - \mathbf{p}[t]) \\
\overset{(a)}{\leq}& \frac{1}{2}\|\nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])\|^2 + \frac{V^2}{2}\|\mathbf{p}[t+1] - \mathbf{p}[t]\|^2 \\
\overset{(b)}{\leq}& \frac{1}{2}D^2 + \frac{V^2}{2}\|\mathbf{p}[t+1] - \mathbf{p}[t]\|^2
\end{aligned} \tag{18}$$

where (a) follows by using basic inequality $\mathbf{x}^{\mathsf{T}}\mathbf{y} \leq \frac{1}{2}\|\mathbf{x}\|^2 + \frac{1}{2}\|\mathbf{y}\|^2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with $\mathbf{x} = \nabla_{\mathbf{p}} U(\mathbf{p}[t]; \omega[t])$ and $\mathbf{y} = V(\mathbf{p}[t+1] - \mathbf{p}[t])$; and (b) follows from Assumption 1. Substituting (18) into (17) yields

$$\begin{aligned}
&VU(\mathbf{p}[t]; \omega[t]) + Q[t](\sum_{i=1}^{n} p_i[t+1] - e[t+1]) \\
\geq& VU(\mathbf{p}^*; \omega[t]) + Q[t](\sum_{i=1}^{n} p_i^* - e[t+1]) + \frac{V^2}{2}\Phi[t] - \frac{1}{2}D^2
\end{aligned} \tag{19}$$

By Lemma 3, we have

$$-\Delta[t] \geq Q[t](\sum_{i=1}^{n} p_i[t+1] - e[t+1]) - \frac{B}{2} \tag{20}$$

Summing (19) and (20); and cancelling common terms on both sides yields

$$\begin{aligned}&VU(\mathbf{p}[t];\omega[t])-\Delta[t]\\ \geq&VU(\mathbf{p}^*;\omega[t])+Q[t](\sum_{i=1}^n p_i^*-e[t+1])+\frac{V^2}{2}\Phi[t]\\ &-\frac{D^2+B}{2}\end{aligned} \tag{21}$$

Note that each $Q[t]$ (depending only on $e[\tau], p[\tau]$ with $\tau \in \{1,2,\ldots,t\}$) is independent of $e[t+1]$. Thus,

$$\begin{aligned}&\mathbb{E}[Q[t](\sum_{i=1}^n p_i^*-e[t+1])]\\ =&\mathbb{E}[Q[t]]\mathbb{E}[\sum_{i=1}^n p_i^*-e[t+1]]\\ \overset{(a)}{\geq}&0\end{aligned} \tag{22}$$

where (a) follows because $Q[t] \leq 0$ and $\sum_{i=1}^n p_i^* \leq \mathbb{E}[e]$ (recall that $e[t+1]$ is an i.i.d. sample of e).

Taking expectations on both sides of (21) and using (22) and $\mathbb{E}[U(\mathbf{p}^*;\omega[t])] = U^*$ yields the desired result. □

B. Utility Optimality Analysis

The next theorem summarizes that the average expected utility attained by Algorithm 1 is within an $O(1/V)$ distance to U^* defined in Lemma 1.

Theorem 1. *Let U^* be the utility bound defined in Lemma 1. For all $t \in \{1,2,\ldots\}$, Algorithm 1 guarantees*

$$\frac{1}{t}\sum_{\tau=1}^t \mathbb{E}[U(\mathbf{p}[\tau];\omega[\tau])] \geq U^* - \frac{V(p^{\max})^2}{2t} - \frac{B}{2Vt} - \frac{D^2+B}{2V} \tag{23}$$

where D is the constant defined in Assumption 1 and B is the constant defined in Lemma 3. This implies,

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^t \mathbb{E}[U(\mathbf{p}[\tau];\omega[\tau])] \geq U^* - \frac{D^2+B}{2V}. \tag{24}$$

In particular, if we take $V = 1/\epsilon$ in Algorithm 1, then

$$\frac{1}{t}\sum_{\tau=1}^t \mathbb{E}[U(\mathbf{p}[\tau];\omega[\tau])] \geq U^* - O(\epsilon), \forall t \geq \Omega(\frac{1}{\epsilon^2}). \tag{25}$$

Proof. Fix $t \in \{1,2,\ldots\}$. For each $\tau \in \{1,2,\ldots,t\}$, by Lemma 5, we have

$$\mathbb{E}[VU(\mathbf{p}[\tau];\omega[\tau])]-\mathbb{E}[\Delta[\tau]] \geq VU^*+\frac{V^2}{2}\mathbb{E}[\Phi[\tau]]-\frac{D^2+B}{2}.$$

Summing over $\tau \in \{1,2,\ldots,t\}$, dividing both sides by Vt and rearranging terms yields

$$\begin{aligned}&\frac{1}{t}\sum_{\tau=1}^t \mathbb{E}[U(\mathbf{p}[\tau];\omega[\tau])]\\ \geq&U^*+\frac{V}{2t}\sum_{\tau=1}^t \mathbb{E}[\Phi[\tau]]+\frac{1}{Vt}\sum_{\tau=1}^t \mathbb{E}[\Delta[\tau]]-\frac{D^2+B}{2V}\\ \overset{(a)}{=}&U^*+\frac{V}{2t}\mathbb{E}[\|\mathbf{p}^*-\mathbf{p}[t+1]\|^2-\|\mathbf{p}^*-\mathbf{p}[1]\|^2]\\ &+\frac{1}{2Vt}\mathbb{E}[(Q[t+1])^2-(Q[1])^2]-\frac{D^2+B}{2V}\\ \geq&U^*-\frac{V}{2t}\mathbb{E}[\|\mathbf{p}^*-\mathbf{p}[1]\|^2]-\frac{1}{2Vt}\mathbb{E}[(Q[1])^2]-\frac{D^2+B}{2V}\\ \overset{(b)}{\geq}&U^*-\frac{V(p^{\max})^2}{2t}-\frac{B}{2Vt}-\frac{D^2+B}{2V}\end{aligned}$$

where (a) follows by recalling that $\Phi[\tau] = \|\mathbf{p}^*-\mathbf{p}[\tau+1]\|^2 - \|\mathbf{p}^*-\mathbf{p}[\tau]\|^2$ and $\Delta[\tau] = \frac{1}{2}(Q[\tau+1])^2 - \frac{1}{2}(Q[\tau])^2$; and (b) follows because $\|\mathbf{p}^*-\mathbf{p}[1]\| = \|\mathbf{p}^*\| = \sqrt{\sum_{i=1}^n (p_i^*)^2} \leq \sum_{i=1}^n p_i^* \leq p^{\max}$ and $|Q[1]| = |Q[0]+e[1]-\sum_{i=1}^n p_i[1]| = |e[1]-\sum_{i=1}^n p_i[1]| \leq \max\{e^{\max}, p^{\max}\} = \sqrt{B}$ where B is defined in Lemma 3. So far we have proven (23).

Equation (24) follows directly by taking $\limsup$ on both sides of (23). Equation (25) follows by substituting $V = \frac{1}{\epsilon}$ and $t = \frac{1}{\epsilon^2}$ into (23). □

C. Lower Bound for Virtual Battery Queue $Q[t]$

Note that $Q[t] \leq 0$ by (9). This subsection further shows that $Q[t]$ is bounded from below. The projection $\text{Proj}_{\mathcal{P}}\{\cdot\}$ satisfies the following lemma:

Lemma 6. *For any $\mathbf{p}[t] \in \mathcal{P}$ and vector $\mathbf{b} \leq \mathbf{0}$, where $\leq$ between two vectors means component-wisely less than or equal to, $\tilde{\mathbf{p}} = Proj_{\mathcal{P}}\{\mathbf{p}[t]+\mathbf{b}\}$ is given by*

$$\tilde{p}_i = \max\{p_i[t]+b_i, 0\}, \forall i \in \{1,2,\ldots,n\}. \tag{26}$$

Proof. Recall that projection $\text{Proj}_{\mathcal{P}}\{\mathbf{p}[t]+\mathbf{b}\}$ by definition is to solve

$$\min_{\mathbf{p}} \quad \sum_{i=1}^n (p_i-(p_i[t]+b_i))^2 \tag{27}$$

$$\text{s.t.} \quad \sum_{i=1}^n p_i \leq p^{\max} \tag{28}$$

$$p_i \geq 0, \forall i \in \{1,2,\ldots,n\} \tag{29}$$

Let $\mathcal{I} \subseteq \{1,2,\ldots,n\}$ be the coordinate index set given by $\mathcal{I} = \{i \in \{1,2,\ldots,n\} : p_i[t]+b_j < 0\}$. For any $\mathbf{p}$ such that $\sum_{i=1}^n p_i \leq p^{\max}$ and $p_i \geq 0, \forall i \in \{1,2,\ldots,n\}$, we have

$$\begin{aligned}&\sum_{i=1}^n (p_i-(p_i[t]+b_i))^2\\ =&\sum_{i\in\mathcal{I}} (p_i-(p_i[t]+b_i))^2+\sum_{i\in\{1,2,\ldots,n\}\setminus\mathcal{I}} (p_i-(p_i[t]+b_i))^2\end{aligned}$$

$$\geq \sum_{i\in\mathcal{I}}(p_i - (p_i[t]+b_i))^2$$
$$\overset{(a)}{\geq} \sum_{i\in\mathcal{I}}(p_i[t]+b_i)^2$$

where (a) follows because $p_i[t]+b_i < 0$ for $i \in \mathcal{I}$ and $p_i \geq 0, \forall i \in \{1,2,\ldots,n\}$. Thus, $\sum_{i\in\mathcal{I}}(p_i[t]+b_i)^2$ is an object value lower bound of problem (27)-(29).

Note that $\tilde{\mathbf{p}}$ given by (26) is feasible to problem (27)-(29) since $\tilde{p}_i \geq 0, \forall i \in \{1,2,\ldots,n\}$ and $\sum_{i=1}^n \tilde{p}_i \leq \sum_{i=1}^n p_i[t] \leq p^{\max}$ because $\tilde{p}_i \leq p_i[t]$ for all i and $\mathbf{p}[t] \in \mathcal{P}$. We further note that

$$\sum_{i=1}^n (\tilde{p}_i - (p_i[t]+b_i))^2 = \sum_{i\in\mathcal{I}}(p_i[t]+b_i)^2.$$

That is, $\tilde{\mathbf{p}}$ given by (26) attains the object value lower bound of problem (27)-(29) and hence is the optimal solution to problem (27)-(29). Thus, $\tilde{\mathbf{p}} = \text{Proj}_{\mathcal{P}}\{\mathbf{p}[t]+\mathbf{b}\}$. □

Corollary 1. *If $Q[t] \leq -V(D^{\max}+p^{\max})$ with $D^{\max} = \max\{D_1,\ldots,D_n\}$, then Algorithm 1 guarantees*

$$p_i[t+1] \leq \max\{p_i[t] - \frac{1}{V}p^{\max}, 0\}, \forall i \in \{1,2,\ldots,n\}.$$

where $D_1,\ldots,D_n$ are constants defined in Assumption 1.

Proof. Let $\mathbf{b} = \frac{1}{V}\nabla_{\mathbf{p}}U(\mathbf{p}[t];\omega[t]) + \frac{1}{V^2}Q[t]\mathbf{1}$. Since $\frac{\partial}{\partial p_i}U(\mathbf{p}[t];\omega[t]) \leq D_i, \forall i \in \{1,2,\ldots,n\}$ by Assumption 1 and $Q[t] \leq -V(D^{\max}+p^{\max})$, we know $b_i \leq -\frac{1}{V}p^{\max}, \forall i \in \{1,2,\ldots,n\}$. By Lemma 6, we have

$$\begin{aligned} p_i[t+1] &= \max\{p_i[t]+b_i, 0\} \\ &\leq \max\{p_i[t] - \frac{1}{V}p^{\max}, 0\}, \forall i \in \{1,2,\ldots,n\}. \end{aligned}$$

□

By Corollary 1, if $Q[t] \leq -V(D^{\max}+p^{\max})$, then each component of $\mathbf{p}[t+1]$ decreases by $\frac{1}{V}p^{\max}$ until it hits 0. That is, if $Q[t] \leq -V(D^{\max}+p^{\max})$ for sufficiently many slots, Algorithm 1 eventually chooses $\mathbf{0}$ as the power decision. By virtual queue update equation (9), $Q[t]$ decreases only when $\sum_{i=1}^n p_i[t] > 0$. These two observations suggest that $Q[t]$ yielded by Algorithm 1 should be eventually bounded from below. This is formally summarized in the next theorem.

Theorem 2. *Let V in Algorithm 1 be a positive integer. Define positive constant Q^l, where superscript l denotes "lower" bound, as*

$$Q^l = V(D^{\max} + 2p^{\max} + e^{\max}) \tag{30}$$

where $e^{\max}$ is the constant defined in Assumption 1 and $D^{\max}$ is the constant defined in Corollary 1. Algorithm 1 guarantees

$$Q[t] \geq -Q^l, \forall t \in \{0,1,2,\ldots\}.$$

Proof. By virtual queue update equation (9), we know $Q[t]$ can increase by at most $e^{\max}$ and can decrease by at most $p^{\max}$ on each slot. Since $Q[0]=0$, we know $Q[t] \geq -Q^l$ for all $t \leq V$. We need to show $Q[t] \geq -Q^l$ for all $t > V$. This can be proven by contradiction as follows:

Assume $Q[t] < -Q^l$ for some $t > V$. Let $\tau > V$ be the *first* (smallest) slot index when this happens. By the definition of τ, we have $Q[\tau] < -Q^l$ and

$$Q[\tau] < Q[\tau-1]. \tag{31}$$

Now consider the value of $Q[\tau-V]$ in two cases (note that $\tau - V > 0$).

- Case $Q[\tau-V] \geq -V(D^{\max}+p^{\max}+e^{\max})$: Since $Q[t]$ can decrease by at most $p^{\max}$ on each slot, we know $Q[\tau] \geq -V(D^{\max}+2p^{\max}+e^{\max}) = -Q^l$. This contradicts the definition of τ.
- Case $Q[\tau-V] < -V(D^{\max}+p^{\max}+e^{\max})$: Since $Q[t]$ can increase by at most $e^{\max}$ on each slot, we know $Q[t] < -V(D^{\max}+p^{\max})$ for all $\tau - V \leq t \leq \tau-1$. By Corollary 1, for all $\tau - V \leq t \leq \tau-1$, we have

  $$p_i[t+1] \leq \max\{p_i[t] - \frac{1}{V}p^{\max}, 0\}, \forall i \in \{1,2,\ldots,n\}.$$

  Since the above inequality holds for all $t \in \{\tau-V, \tau-V+1, \ldots, \tau-1\}$, and since at the start of this interval we trivially have $p_i[\tau-V] \leq p^{\max}, \forall i \in \{1,2,\ldots,n\}$, at each step of this interval each component of the power vector either hits zero or decreases by $\frac{1}{V}p^{\max}$, and so after the V steps of this interval we have $p_i[\tau] = 0, \forall i \in \{1,2,\ldots,n\}$. By (9), we have

  $$\begin{aligned} Q[\tau] &= \min\{Q[\tau-1] + e[\tau] - \sum_{i=1}^n p_i[\tau], 0\} \\ &= \min\{Q[\tau-1] + e[\tau], 0\} \\ &\geq \min\{Q[\tau-1], 0\} \\ &= Q[\tau-1] \end{aligned}$$

  where the final equality holds because the queue is never positive (see (9)). This contradicts (31).

Both cases lead to contradictions. Thus, $Q[t] \geq -Q^l$ for all $t > V$. □

D. Energy Availability Guarantee

To implement the power decisions of Algorithm 1 for the physical battery system $E[t]$ from equations (2)-(3), we must ensure the energy availability constraint (2) holds on each slot. The next theorem shows that Algorithm 1 ensures the constraint (2) always holds as long as the battery capacity satisfies $E^{\max} \geq Q^l + p^{\max}$ and the initial energy satisfies $E[0] = E^{\max}$. It also explains that $Q[t]$ used in Algorithm 1 is a shifted version of the physical battery backlog $E[t]$.

Theorem 3. *If $E[0] = E^{\max} \geq Q^l + p^{\max}$, where Q^l is the constant defined in Theorem 2, then Algorithm 1 ensures the energy availability constraint* (2) *on each slot $t \in \{1,2,\ldots\}$. Moreover*

$$E[t] = Q[t] + E^{\max}, \forall t \in \{0,1,2,\ldots\}. \tag{32}$$

Proof. Note that to show the energy availability constraint $\sum_{i=1}^{n} p_i[t] \leq E[t-1], \forall t \in \{1,2,\ldots\}$ is equivalent to show

$$\sum_{i=1}^{n} p_i[t+1] \leq E[t], \forall t \in \{0,1,2,\ldots\}. \tag{33}$$

This lemma can be proven by inductions.

Note that $E[0] = E^{\max}$ and $Q[0] = 0$. It is immediate that (32) holds for $t = 0$. Since $E[0] = E^{\max} \geq p^{\max}$ and $\sum_{i=1}^{n} p_i[1] \leq p^{\max}$, equation (33) also holds for $t = 0$. Assume (33) and (32) hold for $t = t_0$ and consider $t = t_0+1$. By virtual queue dynamic (9), we have

$$Q[t_0+1] = \min\{Q[t_0] + e[t_0] - \sum_{i=1}^{n} p_i[t_0], 0\}$$

Adding $E^{\max}$ on both sides yields

$$\begin{aligned}
&Q[t_0+1] + E^{\max} \\
&= \min\{Q[t_0] + e[t_0+1] - \sum_{i=1}^{n} p_i[t_0+1] + E^{\max}, E^{\max}\} \\
&\overset{(a)}{=} \min\{E[t_0] + e[t_0+1] - \sum_{i=1}^{n} p_i[t_0+1], E^{\max}\} \\
&\overset{(b)}{=} E[t_0+1]
\end{aligned}$$

where (a) follows from the induction hypothesis $E[t_0] = Q[t_0] + E^{\max}$ and (b) follows from the energy queue dynamic (3). Thus, (32) holds for $t = t_0 + 1$.

Now observe

$$\begin{aligned}
E[t_0+1] &= Q[t_0+1] + E^{\max} \\
&\overset{(a)}{\geq} E^{\max} - Q^l \\
&\geq p^{\max} \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} p_i[t_0+2]
\end{aligned}$$

where (a) follows from the fact that $Q[t] \geq -Q^l, \forall t \in \{0,1,2,\ldots\}$ by Theorem 2; (b) holds since sum power is never more than $p^{\max}$. Thus, (33) holds for $t = t_0 + 1$.

Thus, this theorem follows by induction. □

E. Utility Optimality and Battery Capacity Tradeoff

By Theorem 1, Algorithm 1 is guaranteed to attain a utility within an $O(1/V)$ distance to the optimal utility U^*. To obtain an $O(\epsilon)$-optimal utility, we can choose $V = \lceil 1/\epsilon \rceil$, where $\lceil x \rceil$ represents the smallest integer no less than x. In this case, Q^l defined in (3) is order $O(V)$. By Theorem 3,we need the battery capacity $E^{\max} \geq Q^l + p^{\max} = O(V) = O(1/\epsilon)$ to satisfy the energy availability constraint. Thus, there is a $[O(\epsilon), O(1/\epsilon)]$ tradeoff between the utility optimality and the required battery capacity.

F. Extensions

Thus far, we have assumed that $\omega[t]$ is known with one slot delay, i.e., at the end of slot t, or equivalently, at the beginning of slot $t+1$. In fact, if $\omega(t)$ is observed with t_0 slot delay (at the end of slot $t+t_0-1$), we can modify Algorithm 1 by initializing $\mathbf{p}[\tau] = \mathbf{0}, \tau \in \{1,2,\ldots,t_0\}$ and updating $Q[t-t_0+1] = \min\{Q[t-t_0] + e[t-t_0+1] - \sum_{i=1}^{n} p_i[t-t_0+1], 0\}$, $\mathbf{p}[t+1] = \text{Proj}_{\mathcal{P}}\{\mathbf{p}[t-t_0+1] + \frac{1}{V}\nabla_{\mathbf{p}} U(\mathbf{p}[t-t_0+1]; \omega[t-t_0+1]) + \frac{1}{V^2} Q[t-t_0+1]\mathbf{1}\}$ at the end of each slot $t \in \{t_0, t_0+1, \ldots\}$. By extending the analysis in this section (from a $t_0 = 1$ version to a general t_0 version), a similar $[O(\epsilon), O(1/\epsilon)]$ tradeoff can be established.

V. Numerical Experiment

In this section, we consider an energy harvesting wireless device transmitting over 2 subbands whose channel strength is represented by $s_1[t]$ and $s_2[t]$, respectively. Our goal is to decide the power action $\mathbf{p}[t]$ to maximize the utility/throughput given by (1). Let $\mathcal{P} = \{\mathbf{p} : p_1 + p_2 \leq 5, p_1 \geq 0, p_2 \geq 0\}$. Let harvested energy $e[t]$ satisfy the uniform distribution over interval $[0,3]$. Assume both subbands are Rayleigh fading channels where $s_1[t]$ follows the Rayleigh distribution with parameter $\sigma = 0.5$ truncated in the range $[0,4]$ and $s_2[t]$ follows the Rayleigh distribution with parameter $\sigma = 1$ truncated in the range $[0,4]$.

By assuming the perfect knowledge of distributions, we solve the deterministic problem (4)-(6) and obtain $U^* = 1.0391$. To verify the performance proven in Theorems 1 and 3, we run Algorithm 1 with $V \in \{5, 10, 20, 40\}$ and $E[0] = E^{\max} = Q^l + p^{\max}$ over 1000 independent simulation runs. In all the simulation runs, the power actions yielded by Algorithm 1 always satisfy the energy availability constraints. We also plot the averaged utility performance in Figure 1, where the y-axis is the running average of expected utility. Figure 1 shows that the utility performance can approach U^* by using larger V parameter.

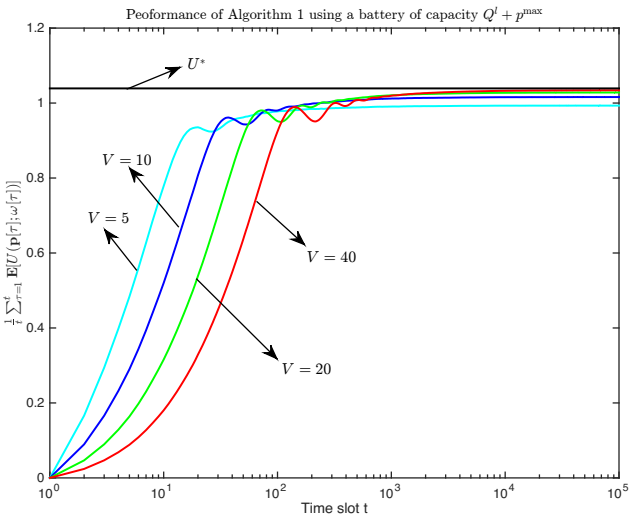

Fig. 1. Utility performance (averaged over 1000 independent simulation runs) of Algorithm 1 with $E[0] = E^{\max} = Q^l + p^{\max}$ for different V.

In practice, it is possible that for a given V, the battery capacity $E^{\max} = Q^l + p^{\max}$ required in Theorem 3 is too large. If we run Algorithm 1 with small capacity batteries such that $\sum_{i=1}^{n} p_i[t+1] \geq E[t]$ for certain slot t, a reasonable choice is to scale down $\mathbf{p}[t+1]$ by (11) and use $\hat{\mathbf{p}}[t+1]$ as the power action. Now, we run simulations by fixing $V = 40$ in Algorithm 1 and test its performance with small capacity batteries. By Theorem 3, the required battery capacity to ensure energy availability is $E^{\max} = 685$. In our simulations, we choose small $E^{\max} \in \{10, 20, 50\}$ and $E[0] = 0$, i.e., the battery is initially empty. If $\mathbf{p}[t+1]$ from Algorithm 1 violates energy availability constraint (2), we use $\hat{\mathbf{p}}[t+1]$ from (11) as the true power action that is enforced to satisfy (2) and update the energy backlog by $E[t+1] = \min\{E[t] - \sum_{i=1}^{n} \hat{p}_i[t+1] + e[t+1], E^{\max}\}$. Figure 2 plots the utility performance of Algorithm 1 in this practical scenario and shows that even with small capacity batteries, Algorithm 1 still achieves a utility close to U^*. This further demonstrates the superior performance of our algorithm.

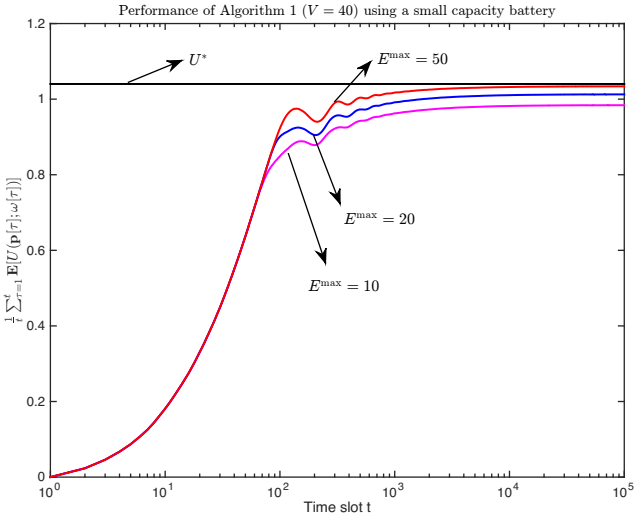

Fig. 2. Utility performance (averaged over 1000 independent simulation runs) of Algorithm 1 with $V = 40$ for different $E^{\max}$.

VI. Conclusion

This paper develops a new learning aided power control algorithm for energy harvesting devices, without requiring the current system state or the distribution information. This new algorithm can achieve an $O(\epsilon)$ optimal utility by using a battery with capacity $O(1/\epsilon)$.

References

[1] J. A. Paradiso and T. Starner, "Energy scavenging for mobile and wireless electronics," *IEEE Pervasive Computing*, vol. 4, no. 1, pp. 18–27, 2005.

[2] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, 2011.

[3] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communications: A review of recent advances," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 360–381, 2015.

[4] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Transactions on Embedded Computing Systems*, vol. 6, no. 4, 2007.

[5] P. Kamalinejad, C. Mahapatra, Z. Sheng, S. Mirabbasi, V. C. Leung, and Y. L. Guan, "Wireless energy harvesting for the internet of things," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 102–108, 2015.

[6] E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," *IEEE Instrumentation & Measurement Magazine*, vol. 18, no. 3, pp. 11–21, 2015.

[7] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Transactions on Communications*, vol. 60, no. 1, pp. 220–230, 2012.

[8] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 1180–1189, 2012.

[9] P. Blasco, D. Gunduz, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1872–1882, 2013.

[10] N. Michelusi, K. Stamatiou, and M. Zorzi, "Transmission policies for energy harvesting sensors with time-correlated energy supply," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2988–3001, 2013.

[11] W. Wu, J. Wang, X. Wang, F. Shan, and J. Luo, "Online throughput maximization for energy harvesting communication systems with battery overflow," *IEEE Transactions on Mobile Computing*, vol. 16, no. 1, pp. 185–197, 2017.

[12] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Control of wireless networks with rechargeable batteries," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 581–593, 2010.

[13] L. Huang and M. J. Neely, "Utility optimal scheduling in energy-harvesting networks," *IEEE/ACM Transactions on Networking*, vol. 21, no. 4, pp. 1117–1130, 2013.

[14] R. Urgaonkar, B. Urgaonkar, M. J. Neely, and A. Sivasubramaniam, "Optimal power cost management using stored energy in data centers," *Proceedings of ACM SIGMETRICS*, 2011.

[15] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proceedings of International Conference on Machine Learning (ICML)*, 2003.

[16] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[17] S. Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2011.

[18] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan & Claypool Publishers, 2010.

[19] H. Yu and M. J. Neely, "A new backpressure algorithm for joint rate control and routing with vanishing utility optimality gaps and finite queue lengths," in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, 2017.

[20] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, 1992.

[21] M. J. Neely, E. Modiano, and C. E. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89–103, 2005.

[22] A. Eryilmaz and R. Srikant, "Joint congestion control, routing, and mac for stability and fairness in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1514–1524, 2006.

[23] A. L. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Systems*, vol. 50, no. 4, pp. 401–457, 2005.

[24] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Transactions on Networking*, vol. 16, no. 2, pp. 396–409, 2008.

[25] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Machine Learning*, vol. 69, pp. 169–192, 2007.

[26] H. Yu and M. J. Neely, "A simple parallel algorithm with an $O(1/t)$ convergence rate for general convex programs," *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 759–783, 2017.