1. Introduction

The worldwide interest in renewable energy such as wind and solar is driven by pressing environmental problems, energy supply security and nuclear power safety concerns. The energy production from these renewable sources is \textit{variable}: uncontrollable, intermittent, uncertain. Variability is a challenge to deep renewable integration.

A central problem is that of economically balancing demand and supply of electricity in the presence of large amounts of variable generation. The current \textit{supply side} approach is to absorb the variability in operating reserves. Here, renewables are treated as negative demand, so the variability appears as uncertainty in net load which is compensated by scheduling fast-acting reserve...
generation. This strategy of tailoring supply to meet demand works at today's modest penetration levels. But it will not scale. Recent studies in California, e.g., CAISO (2010), project that the load-following capacity requirements will need to increase from 2.3 GW to 4.4 GW. These large increases in reserves will significantly raise electricity cost, and diminish the net carbon benefit from renewables as argued in several papers in the literature (Ortega-Vazquez and Kirschen (2010), Negrete-Pincetic et al. (2013)).

There is an emerging consensus that demand side resources must play a key role in supplying zero-emissions regulation services that are necessary for deep renewable integration (e.g., Callaway (2009), Galus et al. (2010), Papavasiliou and Oren (2010), Mathieu et al. (2012), Subramanian et al. (2012)). These include thermostatically controlled loads (TCLs), electric vehicles (EVs), and smart appliances. Some of these loads are deferrable: they can be shifted over time. For example, charging of electrical vehicles (EVs) may be postponed to some degree. Other loads such as HVAC units can be modulated within limits. The core idea of demand side approaches to renewable integration is to exploit load flexibility to track variability in supply, i.e., to tailor demand to match supply. For this a cluster manager or aggregator offers a control and business interface between the loads and the system operator (SO).

The demand side approach has led to two streams of work: (a) indirect load control (ILC) where flexible loads respond, in real-time, to price proxy signals, and (b) direct load control (DLC) where flexible loads cede physical control of devices to operators who determine appropriate actions. The advantage of DLC is that with greater control authority the cluster manager can more reliably control the aggregate load. However, DLC requires a more extensive control and communication infrastructure and the manager must provide economic incentives to recruit a sufficient consumer base. The advantage of ILC is that the consumer retains authority over her electricity consumption.

Both ILC and DLC require appropriate economic incentives for the consumers. In ILC, the real-time price signals provides the required incentives. However, the quantification of those prices, the feasibility of consumer response and the impact on the system and market operations in terms of price volatility and instabilities is a matter of concern, as recent literature suggest (Roozbehani et al. (2010), Barbose et al. (2004), Wang et al. (2011)).

DLC also requires the creation of economic signals, but unlike real-time pricing schemes DLC can use forward markets. For DLC to be effective, it is necessary to offer consumers who present greater demand flexibility a larger discount. The discounted pricing can be arranged through flexibility-differentiated electricity markets Tan and Varaiya (1993). Here, electricity is regarded as a set of differentiated services as opposed to a homogeneous commodity. Consumers can purchase
an appropriate bundle of services that best meets their electricity needs. From the producer’s perspective, providing differentiated services may better accommodate supply variability. This paper is concerned with electric power services differentiated by the duration $h$ for which power is supplied. We explore balancing supply and demand for such services through forward markets. There is a growing body of work Varaiya et al. (2011), Negrete-Pincetic and Meyn (2012), Bitar and Xu (2013) on differentiated electricity services.

1.1. Main Contributions

This paper considers a stylized version of flexible loads. The service interval is divided into $T$ slots, indexed $t = 1, \cdots, T$. We assume the available power $p_t$ kW in any slot $t$ is given. A flexible load demands 1 kW of power for a duration of $h$ slots. The flexibility resides in the fact that any $h$ of the available $T$ slots will satisfy the load. There are $N$ loads, defined by their service durations $h = \{h_i, i = 1, \cdots, N\}$.

We take a demand side approach with direct load control. Our objective is to study the allocation of available supply to the various loads in a market context. Loads are induced to participate in this market by receiving lower electricity prices in exchange for greater flexibility. If the supplier contracts ex ante to deliver power for $h$ slots to a particular flexible load, it is obligated to do so. The supplier selects $h$ of the the available $T$ slots to supply power to the load. This scheme requires certain technology infrastructure (communication, power electronics), a treatment of which is outside the scope of this paper. The load is not informed much in advance which $h$ slots it will receive power. Thus the load must assume the burden of planning its consumption without knowing exactly when power will be available. The available power is principally drawn from zero-marginal cost renewable sources. Because of the variability in these sources, the supplier may be compelled to use supplemental generation such as on-site gas turbine to meet its obligations.

Our first set of results are contained in Section 2. Here, we study the problems of adequacy. We first give a necessary and sufficient condition under which the available power $p = (p_1, p_2, \ldots, p_T)$ is adequate to meet the loads $h$. Under this condition, we describe a Least Laxity First (LLF) algorithm that constructs an appropriate allocation to services the loads. In the event the available supply profile $p$ does not meet the adequacy condition, it becomes necessary for the supplier to purchase additional power. We characterize the minimum cost power purchase decisions under (a) oracle information, and (b) run-time information about the supply. We show that the optimal purchase decisions are identical in both cases. Finally, we treat the more general case where load $i$ must be supplied $E_i$ kW-slots of power at a maximum rate of $m_i$ kW per slot.
Our second set of results may be found in Section 3. Here, we consider a forward market for duration-differentiated services, which are bundles of \( h \) 1-kW slots sold at prices \( \pi(h) \), \( h = 1, 2, \ldots \). Consumer \( i \) selects the service \( h \) that maximizes her net utility \( U(h) - \pi(h) \), and the supplier bundles its supply (both its renewable generation and any forward purchases made from the grid) into \( n_h \) units of service for \( h \) slots, so as to maximize its revenue. We show that there is a competitive market equilibrium, and it maximizes social welfare. The competitive duration-differentiated market equilibrium is then compared with a sequential real-time market, in which the price of power \( \zeta(t) \) is the market clearing price for slot \( t \). The comparison reveals that the real-time markets may not be efficient.

All proofs are collated in the Appendix.

1.2. Prior Work

Supply side approaches

Here, variable supply from renewable sources is regarded as negative demand. The objective is to arrange for reserve power generation to compensate for fluctuations in net demand. The problem is formulated from the viewpoint of the system operator (SO) who must purchase reserve generation capacity and energy to meet the random demand while minimizing the risk of mismatch and the cost of reserves. Reserve generation can be purchased in forward markets with different time horizons (day-ahead, hour-ahead, 5 minutes-ahead). With shorter time horizons, the uncertainty in the forecast net demand is reduced but the cost of reserves increases. The SO's optimal decision can be formulated as a stochastic control problem known as risk-limiting dispatch presented in Varaiya et al. (2011). When SO's decisions include unit commitment and transmission constraints, the problem is a mixed-integer nonlinear stochastic programming problem that is computationally challenging as discussed in Bouffard and Galiana (2008). A number of papers address the computational aspects of stochastic unit commitment (e.g., Carpentier et al. (1996), Nowak and Romisch (2000), Papavasiliou et al. (2011)). Alternatively, Bertsimas et al. (2013) present a robust optimization formulation of the unit commitment problem. If unit commitment and transmission constraints are omitted, the resulting stochastic dispatch problem has an analytical solution as shown in Rajagopal et al. (2013, 2012).

Demand side approaches

Current research in direct load control focuses on developing and analyzing algorithms for coordinating resources (e.g., Chen et al. (2011), Lee et al. (2008), Hsu and Su (1991), Mets et al.
(2010)). For example, Gan et al. (2012) develops a distributed scheduling protocol for electric vehicle charging; Papavasiliou and Oren (2010) uses approximate dynamic programming to couple wind generation with deferrable loads; and Ma et al. (2011), Hug-Glanzmann (2010), Galus et al. (2010) suggest the use of receding horizon control approaches for resource scheduling.

Recent studies in indirect load control have developed real-time pricing algorithms Ilic et al. (2011) and quantified operational benefits Lijesen (2007). There has also been research focused on economic efficiency in Borenstein (2005), Spees and Lave (2007), feedback stability of price signals in Roozbehani et al. (2010), volatility of real-time markets in Wang et al. (2011) as well as the practical issues associated with implementing ILC programs presented in Barbose et al. (2004).

An early exposition of differentiated energy services is offered in Oren and Smith (1993). There are other approaches to such services that naturally serve to integrate variable generation sources. Reliability differentiated energy services where consumers accept contracts for $p$ MW of power with probability $\rho$ are developed in Tan and Varaiya (1993). More recently, the works of Bitar and Low (2012), Bitar and Xu (2013) consider deadline differentiated contracts where consumers receive price discounts for offering larger windows for the delivery of $E$ MWh of energy.

**Notation**

Bold letters denote vectors. We reserve subscripts $t$ to index time and $i$ to index loads. For a vector $a = (a_1, a_2, \ldots, a_T)$, $a^{\downarrow}$ denotes the non-increasing rearrangement of $a$, so, $a^{\downarrow}_t \geq a^{\downarrow}_{t+1}$ for $t = 1, 2, \ldots, T - 1$. For an assertion $A$, $\mathbb{1}_A$ denotes 1 if $A$ is true and 0 if $A$ is false.

2. **Adequacy Results**

We consider loads whose energy needs are flexible. Examples include electric vehicles that allow flexible charging over an 8 hour service interval, aluminum smelters that might operate for $h$ hours out of 24, appliances such as washing machines that require a fixed power for any 1 hour out of the next 8. We use a common abstraction for such loads. While this abstraction does not account for many important practical constraints, it serves to formulate and study the central mathematical problems in scheduling/control and economics for flexible loads.

Time is segmented into $T$ slots, indexed by $t$. The (constant) power available in time slot $t$ is $p_t$, and $p = (p_1, p_2, \ldots, p_T)$ is called the *supply profile*.

There are $N$ flexible loads, indexed by $i$. Load $i$ requires 1 kW for *any* $h_i$ of $T$ time slots. The vector $h = (h_1, h_2, \ldots, h_N)$ is called the *demand profile*. For a demand profile $h$, we define an associated *demand-duration vector* $d = (d_1, d_2, \ldots, d_T)$ where $d_t := \sum_i \mathbb{1}_{t \leq h_i}$. Equivalently, the
number of consumers that need $t$ slots is $d_t - d_{t+1}$ (where $d_{T+1} := 0$). There is a bijection between the demand profile and the demand-duration vector as $\mathbf{h}$ specifies $\mathbf{d}$ uniquely and vice versa.

Since the loads are flexible, many supply profiles can service a given load profile. We next consider several versions of the problem of allocating supply (the available power) to demand (the flexible loads).

![Figure 1](characterizing duration-differentiated loads (left), adequate supply profile (center), inadequate supply profile (right).)

### 2.1. Adequacy conditions

Any allocation of supply to loads can be specified by a binary allocation rule $A \in \{0, 1\}^{N \times T}$ where $A(i, t) = 1$ if and only if load $i$ receives power in slot $t$. We define two notions of supply adequacy.

**Definition 1 (Adequacy).** The supply profile $\mathbf{p} = (p_1, \cdots, p_T)$ is *adequate* for the demand profile $\mathbf{h} = (h_1, \cdots, h_N)$ if there exists an allocation rule $A(\cdot, \cdot)$ such that

$$
\sum_t A(i, t) = h_i, \\
\sum_i A(i, t) \leq p_t.
$$

If further, $\sum_i A(i, t) = p_t$, we will say that $\mathbf{p} = (p_1, \cdots, p_T)$ is *exactly adequate* for $\mathbf{h} = (h_1, \cdots, h_N)$.

**Example 1.** Consider $T = 6$ time slots and $N = 5$ flexible loads as illustrated in Figure 1. If the demand profile is $\mathbf{h} = (1, 2, 2, 3, 6)$, the associated demand-duration vector is $\mathbf{d} = (5, 4, 2, 1, 1, 1)$. The supply profile shown in the center panel is exactly adequate to service the loads. The supply profile shown in the right panel has the same total energy, but it is inadequate to service the loads.

The following lemma is a direct consequence of the above definition.
Lemma 1. If \( p \) is (exactly) adequate for a demand profile \( h \), then any temporal rearrangement of \( p \) is also (exactly) adequate for the same demand profile \( h \).

We will characterize adequacy more directly via the demand-duration vector. For this we employ some notions from majorization theory.

Definition 2 (Majorization). Let \( a = (a_1, \ldots, a_T) \) and \( b = (b_1, \ldots, b_T) \) be two non-negative vectors. Denote by \( a^↓ \), \( b^↓ \) the non-increasing rearrangements of \( a \) and \( b \) respectively. We say that \( a \) majorizes \( b \), written \( a \prec b \), if

(i) \[ \sum_{s=1}^{T} a^↓_s \leq \sum_{s=1}^{T} b^↓_s , \quad \text{for } t = 1, 2, \ldots, T , \text{ and} \]

(ii) \[ \sum_{s=1}^{T} a^↓_s = \sum_{s=1}^{T} b^↓_s . \]

If only the first condition holds, we say that \( a \) weakly majorizes \( b \), written \( a \prec^w b \).

Remark 1. The inequalities in our definition of majorization are reversed from standard use in majorization theory. This departure from convention allows us to write our adequacy conditions as \( d \prec p \) and \( d \prec^w p \) which suggests the intuitive adequacy condition of demand being “less than” supply .

Our next result characterizes adequacy.

Theorem 1 (Adequacy). (a) The supply profile \( p \) is exactly adequate for a demand profile \( h \) with the associated demand-duration vector \( d \) if and only if \( d \prec p \).

(b) The supply profile \( p \) is adequate for a demand profile \( h \) with the associated demand-duration vector \( d \) if and only if \( d \prec^w p \).

Proof. See Appendix B.

2.2. Least Laxity Allocation

We now describe an allocation rule that will play a key role in this paper. Given an allocation rule \( A \), we define the laxity of load \( i \) at time \( t \) as \( x_i(t) := T - t + 1 - \left( h_i - \sum_{s=1}^{t-1} A(i, s) \right) \).

Definition 3 (LLF Allocation). Fix the supply profile \( p = (p_1, p_2, \ldots, p_T) \). The Least Laxity Allocation rule \( A(i, t) \) is defined by

(i) At time 1, \( x_i(1) = T - h_i \). Arrange the loads in non-decreasing order of \( x_i(1) \) and let \( A_1 \) be the collection of the first \( p_1 \) loads from this order. Set \( A(i, 1) = 1 \) if and only if \( i \in A_1 \).

(ii) At time \( t \), \( x_i(t) = T - t + 1 - \left( h_i - \sum_{s=1}^{t-1} A(i, s) \right) \). Arrange loads in non-decreasing order of \( x_i(t) \) and let \( A_t \) be the collection of the first \( p_t \) loads from this order. Set \( A(i, t) = 1 \) if and only if \( i \in A_t \).
As its name suggests, at each time $t$ LLF gives priority to loads with smaller laxity. Our next result shows that the LLF allocation successfully services the loads when the supply profile is adequate. We have:

**Theorem 2.** If the supply profile $p = (p_1, p_2, \ldots, p_T)$ is adequate, then the Least Laxity First allocation rule satisfies all the demands, i.e.

$$\sum_i A(i, t) = h_i, \quad \sum_i A(i, t) \leq p_t$$

**Proof** See Appendix C.

### 2.3. Supplemental Power Purchases

It may happen that the supply profile is not adequate for a given demand profile. In this case, the supplier will have to purchase additional power to serve the loads. We determine the least costly increment in supply profile to make it adequate. We consider two scenarios: (a) *Oracle information*: the entire supply profile $p$ is revealed in advance, (b) *Run-time information*: the power available in slot $[t, t+1)$ is revealed at time $t$, i.e. immediately before the beginning of the slot.

**Case (a):** The supplier needs to serve a demand profile $h$ with the associated demand-duration vector $d = (d_1, d_2, \ldots, d_T)$. Before the time of delivery, the supplier learns the true realization of the entire supply profile $p = (p_1, \ldots, p_T)$. If $d \prec_w p$, the supply is adequate and there is no need for supplemental power. In the case of inadequate supply, the additional power to be purchased at minimum cost while ensuring that all demands are met is given by the solution of the following optimization problem:

$$\min_{a \geq 0} \sum_{t=1}^{T} c \cdot a_t \quad \text{subject to} \quad d \prec_w (p + a),$$

where $c \geq 0$ is the unit price of supplemental power. This is a linear programming problem since the majorization inequalities are linear.

**Case (b):** The power available in each slot is revealed just before the beginning of that slot. The supplier now faces a sequential decision-making problem where the information available to make the purchase decision $a_t$ at time $t$ is $d, p_1, p_2, \ldots, p_t$. The supplier’s objective is to minimize the total cost of additional power $\sum_{t=1}^{T} c \cdot a_t$ while ensuring that all load demands are met.

Clearly the supplier’s optimal cost in Case (b) is lower bounded by its optimal cost in Case (a). Surprisingly, it happens that the optimal costs and corresponding decision strategies are identical in both situations. More precisely, we have:
Theorem 3. Consider the following decision strategy for the supplier:

(i) The additional power purchased at $t = 1$ is $a_1 = (d_T - p_1)^+$. The total power $(p_1 + a_1)$ is allocated to consumers according to the LLF policy described in Definition 3.

(ii) At time $t$, knowing the supply $p_1, p_2, \ldots, p_t$ and the purchases $a_1, \ldots, a_{t-1}$, the power purchased $a_t$ is the solution of the following optimization problem:

\[
\begin{align*}
\min_{a_t} & \quad a_t \\
\text{s.t.} & \quad (p_1 + a_1, p_2 + a_2, \ldots, p_{t-1} + a_{t-1}, p_t + a_t) > w (d_{T-t+1}, d_{T-t+2}, \ldots, d_T) \quad (1)
\end{align*}
\]

The total power $(p_t + a_t)$ is allocated to consumers according to the LLF policy. Then,

(a) This strategy is optimal under both the oracle information and run-time information cases.

(b) The optimal cost is $c \left( \max_t \left( \sum_{s \geq t} (d_s - p_s^+) \right) \right)$.

Proof: See Appendix D.

2.4. Rate-constrained energy services

We have thus far assumed that consumers demand 1 kW for a specified number of time slots. We now consider a more general energy service, which provides a specified quantity of energy to be delivered over $T$ time slots. However, the power level in each time slot can assume any integer value up to a maximum rate $m$. We will argue that these products can be viewed as a combination of the duration-differentiated services at a fixed power level of 1 kW presented before.

Consumer $i$ specifies two quantities: $(E_i, m_i)$. $E_i$ is the total energy to be consumed over $T$ time slots at a maximum rate of $m_i$ kW per time slot. Both $E_i$ and $m_i$ are integer-valued. For example, a consumer specifying $(100, 10)$ requires 100 kW-slots of energy with the constraint that the consumption rate in each time slot is 0, 1, 2, … or 10 kW. A consumer specifying $(E_i, m_i)$ is satisfied with any supply allocation $A_t \in \{0, 1, \ldots, m_i\}$ such that $\sum A_t = E_i$. The consumer model of section 2 corresponds to the case where $m_i = 1$.

Consider a consumer whose energy requirement and rate constraint are specified by the pair $(E, m)$. Let $k, r$ be such that $E = km + r$ with $r < m$. The following result shows that for the purpose of allocating supply, a consumer specifying $(E, m)$ is equivalent to a collection of $m$ consumers with $(E_n = k + 1, m_n = 1)$ for $n = 1, 2, \ldots, r$ and $(E_n = k, m_n = 1)$ for $n = r + 1, \ldots, m$. Hence, the adequacy and allocation results of Section 2 (which were derived for consumers with $m = 1$) can be used for the variable rate consumers as well.
Theorem 4. Consider a consumer whose energy requirement and rate constraint are specified by the pair \((E, m)\). Let \(k, r\) be such that \(E = km + r\) with \(r < m\). Then, a supply allocation satisfies

\[ A_t \in \{0, 1, \ldots, m\} \sum_t A_t = E \tag{2} \]

if and only if

\[ A_t = \sum_{n=1}^{m} A^n_t, \tag{3} \]

where \(A^n_t \in \{0, 1\}\) and \(\sum_t A^n_t = k + 1\) for \(n \leq r\) and \(\sum_t A^n_t = k\) for \(n > r\).

Proof. See Appendix G.

3. Forward market for duration-differentiated contracts

In this section we consider a forward market for duration-differentiated services and investigate its properties. Duration-differentiated services couple supply and consumption across different time slots. A natural way to capture this is to consider a forward market for the whole operational period where services of different durations are bought and sold. In this way, both consumers and suppliers can effectively quantify the value/cost of consuming/producing these services. All the market transactions are completed before the delivery time. Thus, the market decisions are made prior to the operational decisions required for the delivery of the products which is a characteristic of direct load control. In order to illustrate the advantages of forward markets, we perform a comparison with a stylized real-time spot market implementation. The results shed light about potential inefficiencies resulting from the implementation of spot markets in which the inter-temporal dimension of duration-differentiated contracts is hard to capture.

The forward market has three elements:

(a) Services: The services are differentiated by the number \(h\) of time slots within \(T\) slots during which 1 kW of electric power is delivered. Service \(h\) is sold at price \(\pi(h)\).

(b) Consumers: The benefit to a consumer who receives \(h\) slots is \(U(h)\); all consumers have the same utility function \(U\).

(c) Supplier: The supplier receives for free the power with profile \(p = (p_1, \ldots, p_T)\). The supplier knows the profile \(p\). The supplier may purchase additional energy at price \(C\) per kW-slot.

The information flow of the market is depicted in Fig 2. Facing a menu of services with associated prices \(\mathcal{M} = \{k, \pi(k)\}\), consumer \(n\) selects a service \(h_n\) that maximizes her net benefit, while the supplier selects the number \(n_t\) of services of duration \(t\) to produce that maximize her net profit.
We first characterize the decisions that maximize total social welfare, defined as aggregate consumer utility minus the cost of purchased energy. We then show that the optimum decisions can be sustained as a competitive equilibrium. Lastly, we compare the competitive equilibrium in a real-time spot market with the equilibrium for our duration-differentiated forward market.

3.1. Social Welfare Problem

We consider a set of homogenous consumers, consumer \( i = 1, \ldots, N \) enjoys utility \( U(h) \) upon consuming 1 kW of power for \( h \) slots. The supplier has available for free a quantity of power with profile \( p = (p_1, \ldots, p_T) \), and can also purchase additional energy at \( C \) per kW-slot. The social welfare optimization problem is

\[
\max_{h \geq 0, y \geq 0} \sum_{i=1}^{N} U(h_i) - \sum_{j=1}^{T} C y_j \\
\text{subject to } d_t = \sum_{i=1}^{N} \mathbb{1}_{\{t \leq h_i\}} \\
y + p >^w d
\]

**Theorem 5 (Social Welfare).** Assume the profile \( p \) is arranged in decreasing order: \( p = p^\downarrow \).

(a) (Convex utility) Suppose \( U(h) - U(h-1) \) is a non-negative, non-decreasing function of \( h \) (with \( U(0) = 0 \)) and the number of consumers \( N \) is larger than \( p_1 \). Define

\[
k^* = \begin{cases} 
\min k : k \in \{0, 1, \ldots, (T-1)\}, & \frac{U(T)-U(k)}{T-k} \geq C \\
T & \text{otherwise}
\end{cases}
\]
If $k^* \geq 1$, the optimum demand duration is
\[
d^*_t = \begin{cases} p_t & \text{if } t < k^* \\ p_{k^*} & \text{if } t \geq k^* \end{cases}
\]
(5)

If $k^* = 0$, the optimum demand duration vector is $d^*_t = N$ for all $t$.

(b) (Concave utility) Suppose $U(h) - U(h-1)$ is a non-negative, non-increasing function of $h$ (with $U(0) = 0$) and the number of consumers $N$ is larger than $\sum_t p_t$. Define
\[
k^* = \begin{cases} \min k : k \in \{1, \ldots, T\}, & U(k) - U(k-1) \geq C \text{ if this exists} \\ 0 & \text{otherwise} \end{cases}
\]
(6)

If $k^* \geq 1$, the optimum demand duration is
\[
d^*_t = \begin{cases} N & \text{if } t \leq k^* \\ 0 & \text{if } t > k^* \end{cases}
\]
(7)

If $k^* = 0$, the social welfare maximizing demand duration vector is
\[
d^*_t = \begin{cases} \sum_{i=1}^T p_i & \text{if } t = 1 \\ 0 & \text{if } t > 1 \end{cases}
\]
(8)

Proof See Appendix E.

In the convex case, the utility increments are non-decreasing in $h$ and the optimal allocation favors longer duration contracts. In the concave case, the utility increments are non-increasing in $h$ and the optimal allocation favors the shortest durations.

Remark 2. In standard commodity markets, the usual setting is to consider concave utility functions which reflects the decreasing marginal utility of many goods. In the case of duration-differentiated loads, the concave case could represent situations in which additional hours of consumption does not increase the marginal utility, for example the filtering of a pool beyond the minimum numbers of hours. However, in this case convex utility functions are also of interest. That could represent loads for which interruptions of the consumption is material. Examples include industrial mining processes, power supply for computational applications, air flow in hospitals.

Example 2. For example, take $p = (5, 4, 2, 1, 1, 0)$, the number of consumers $N = 14$ and the time period $T = 6$. In addition, the utility function for the convex case is such that
\[
U(6) - U(5) \geq C.
\]
(9)
and for the concave case
\[
U(1) - U(0) \geq C.
\]
(10)
In the convex case the optimal allocation is $h = (1, 2, 2, 3, 6)$ as in Fig 3 (left). Note that an additional unit of supply is utilized and used to create a contract of duration 6 hours. In the concave case only contracts of duration 1 slots are required as in Fig 3 (middle). Note that in this case, also an additional unit of supply is utilized to create an additional contract of duration 1.
3.2. Competitive equilibrium

We now analyze a stylized forward market for the production and consumption of duration-differentiated services. We focus on a perfect-competition setting in which prices are assumed to be *exogenous* to the players’ decisions. Consequently, all players are *price takers*, i.e., they cannot influence the prices. The perfect-competition setting is certainly an idealization but it provides valuable insights in terms of market design. In particular, these outcomes are usually used as a benchmark for analysis in which perfect-competition is not considered, e.g., monopolistic or oligopolistic settings.

The market determines a price for each service, every consumer selects the service she wants to maximize her net benefit, and the supplier decides how much of each service to produce. If the demand and supply for each service match, a competitive equilibrium is said to exist. Mathematically, a competitive equilibrium is defined as follows.

**Definition 4** (Competitive Equilibrium). The supplier’s service production vector \( \mathbf{n} = (n_1, n_2, \ldots, n_T) \), the consumers’ demand profile \( \mathbf{h} = (h_1, h_2, \ldots, h_N) \), and a set \( \{\pi(h)\} \) of prices constitute a competitive equilibrium if three conditions hold:

(i) **Consumer Surplus Maximization**: \( h_i \in \arg\max_h U(h) - \pi(h) \), all \( i \).

(ii) **Profit Maximization**: The services produced maximize profit, i.e. \( \mathbf{n} \) solves this optimization problem:

\[
\max_{\mathbf{n} \geq 0, \mathbf{y} \geq 0} \sum_{t=1}^{T} (n_t \pi(t) - Cy_t)
\]

subject to \( d_t = \sum_{i=t}^{T} n_i \)
\[ \mathbf{p} + \mathbf{y} \succ \mathbf{d} \]

(iii) Market Clearing: The supply and demand for each service are equal: \( n_t = \sum_{i=1}^{N} 1\{h_i=t\} \).

**Definition 5 (Efficiency).** The competitive equilibrium is *efficient* if the resulting allocation maximizes social welfare.

We next characterize efficient competitive equilibria for convex and concave utility functions.

**Theorem 6 (Efficient Competitive Equilibrium).** Assume the profile \( \mathbf{p} \) is arranged in decreasing order: \( \mathbf{p} = \mathbf{p}^\downarrow \).

(a) (Convex utility) Suppose \( U(h) - U(h-1) \) is non-decreasing (with \( U(0) = 0 \)) and the number of consumers \( N \) is larger than \( p_1 \). Then an efficient competitive equilibrium exists and is described as follows:

**Prices:** \( \pi(h) = U(h) \).

**Supplier’s production:** Define \( k^* \) as in (4). Then, if \( k^* \geq 1 \), \( n_t = p_t - p_{t+1} \) for \( t < k^* \), \( n_t = 0 \), for \( k^* \leq t < T \), and \( n_T = p_{k^*} \). If \( k^* = 0 \), the supplier’s production is \( n_T = N \).

**Consumption:** \( n_t \) consumers purchase duration \( t \) service, so market is cleared.

(b) (Concave utility) Suppose \( U(h) - U(h-1) \) is non-increasing (with \( U(0) = 0 \)) and the number of consumers \( N \) is larger than \( \sum_t p_t \). Then an efficient competitive equilibrium exists and is described as follows:

**Prices:** \( \pi(h) = \min(C, U(1))h \).

**Supplier’s production:** Define \( k^* \) as in (6). If \( k^* \geq 1 \), \( n_k = N \). If \( k^* = 0 \), \( n_1 = \sum_t p_t \) and \( n_k = 0 \) for \( k > 1 \).

**Consumption:** \( n_t \) consumers purchase duration \( t \) service, so market is cleared.

**Proof** See Appendix F.

**Remark 3.** In the above analysis, the supply side represents an aggregation of many suppliers. Note that individual price taking suppliers may benefit by coordinating to offer longer service contracts in order to take advantage of higher prices for longer durations. We assume that the suppliers are able to identify and exploit all such opportunities. We further assume that even with such coordination the number of effective suppliers is large enough to justify the assumption of a competitive market.
Remark 4. In the convex case, the marginal price of the \( h \)th slot, \( \pi(h) - \pi(h - 1) \) is increasing. This is due to the fact that from a given supply profile it may be possible to produce a \( m \)-slot and a \( k \)-slot service but not a \((m + k)\)-slot service, as can be seen from the definition of adequacy. This contrasts with the assumption in Chao et al. (1986) that, for conventional generator technology, the marginal price of producing electricity for slot \( h \) decreases with \( h \). In the concave case above this issue does not arise since service of a single duration is produced.

Remark 5. Suppose the actual chronological profile of the power is \( q = (q_1, \cdots, q_T) \). Suppose \( q \) has \((m + 1)\) local minima. Then, in the market allocations in the convex case, each service will have at most \( m \) interruptions. For the same example as before Fig 3 (right) shows that there is at most one interruption since \( q = (5, 2, 4, 1, 1, 1) \) has two local minima.

3.3. Real-time vs duration-differentiated markets

We compare the duration-differentiated market with a real-time spot market that operates as follows. At the beginning of slot \( t \) the market determines a price \( \pi_t \) for 1kW of power delivered during slot \( t \). The price \( \pi_t \) equates the supply function \( s_t(\pi) \) and demand function \( q_t(\pi) \), i.e., \( s_t(\pi_t) = q_t(\pi_t) \). We model these functions. The supply function is straightforward. At the beginning of slot \( t \), the supplier receives for free power \( p_t \) and offers it for sale inelastically. She can also supply as much power as she wishes at price \( C \) per kW-slot. Thus the supply function is

\[
 s(\pi) = \begin{cases} 
 \infty & \text{if } \pi > C \\
 [p_t, \infty) & \text{if } \pi = C \\
 p_t & \text{if } \pi < C 
\end{cases} 
\]  

(11)

The demand function is a bit more complicated. At the beginning of slot \( t \), consumer \( n \) determines her willingness to pay \( v_n(t) \) for 1 kW in slot \( t \). So the (aggregate) demand function is

\[
 q_t(\pi) = \sum_{i=1}^{N} I_{\{\pi \leq v_n(t)\}}. 
\]  

(12)

The willingness to pay \( v_n(t) \) will depend on the power that consumer \( n \) acquired in slots \( 1, \cdots, t - 1 \). All consumers have the same utility function \( U(h) \). If consumer \( n \) had acquired \( x \) kW-slots (out of a total of \( t - 1 \)), her willingness to pay would equal the additional utility she will gain,

\[
 v_n(t) = U(x + 1) - U(x). 
\]  

(13)

Thus the consumer is myopic: her willingness to pay is unaffected by her opportunities to make future purchases in slots \( t + 1, \cdots, T \). The demand function is obtained from (12) and (13). Note
that although consumers are myopic, their purchasing decisions depend on previous decisions. Let \( x_i, i = 0, \ldots, t - 1 \), be the number of consumers that have purchased \( i \) kW-slots during the first \((t - 1)\) spot markets.

Consider the convex case: \( U(h) - U(h - 1) \) is non-decreasing. Here, the demand function is

\[
q_t(\pi) = \begin{cases} 
0 & \text{if } \pi > U(t) - U(t - 1) \\
x_{t-1} & \text{if } U(t) - U(t - 1) \geq \pi > U(t - 1) - U(t - 2) \\
x_{t-1} + x_{t-2} & \text{if } U(t - 1) - U(t - 2) \geq \pi > U(t - 2) - U(t - 3) \\
\vdots & \\
x_{t-1} + \cdots + x_1 & \text{if } U(2) - U(1) \geq \pi > U(1) - U(0) \\
x_{t-1} + \cdots + x_0 = N & \text{if } U(1) \geq \pi 
\end{cases}
\]  

(14)

Now consider the concave case: \( U(h) - U(h - 1) \) is non-increasing. Then the demand function is

\[
q_t(\pi) = \begin{cases} 
0 & \text{if } \pi > U(1) - U(0) \\
x_0 & \text{if } U(1) - U(0) \geq \pi > U(2) - U(1) \\
x_0 + x_1 & \text{if } U(2) - U(1) \geq \pi > U(3) - U(2) \\
\vdots & \\
x_0 + \cdots + x_{t-2} & \text{if } U(t - 1) - U(t - 2) \geq \pi > U(t - 2) - U(t - 3) \\
x_0 + \cdots + x_{t-1} = N & \text{if } U(t) - U(t - 1) \geq \pi 
\end{cases}
\]  

(15)

Figure 4 shows the supply and demand functions for the convex case (left) and the concave case (right).

**Theorem 7 (Real-time market).** (a) Suppose \( U(h) - U(h - 1) \) is a non-negative, non-decreasing function of \( h \) (with \( U(0) = 0 \)). Then, the spot price at \( t \) is

\[
\pi_t = \begin{cases} 
\min(C, U(t) - U(t - 1)) & \text{if } p_t < x_{t-1} \\
\min(C, U(t - 1) - U(t - 2)) & \text{if } x_{t-1} \leq p_t < x_{t-1} + x_{t-2} \\
\min(C, U(t - 2) - U(t - 3)) & \text{if } x_{t-1} + x_{t-2} \leq p_t < x_{t-1} + x_{t-2} + x_{t-3} \\
\vdots & \\
\min(C, U(1)) & \text{if } x_{t-1} + x_{t-2} + \cdots + x_1 \leq p_t < N \\
0 & \text{if } N \leq p_t 
\end{cases}
\]  

(16)

---

**Figure 4** Demand and supply functions for spot market: convex utility (left), concave utility (right).
(b) Suppose \( U(h) - U(h - 1) \) is a non-negative, non-increasing function of \( h \) (with \( U(0) = 0 \)).

Then the spot price at \( t \) is

\[
\pi_t = \begin{cases} 
\min(C, U(1) - U(0)) & \text{if } p_t < x_0 \\
\min(C, U(2) - U(1)) & \text{if } x_0 \leq p_t < x_0 + x_1 \\
\cdots & \\
\min(C, U(t) - U(t - 1)) & \text{if } x_0 + \cdots + x_{t-2} \leq p_t < x_0 + \cdots + x_{t-1} = N \\
0 & \text{if } N \leq p_t 
\end{cases}
\]  \hspace{1cm} (17)

**Proof**  From the intersection of the demand and supply functions in Figure 4 one can read off the formulas (16) and (17) for the spot price.

**Example 3.** An example shows that the allocation achieved by the real time spot market may not be efficient. Consider one consumer and two periods, \( t = 1, 2 \), the supplier’s free power is \( p_1 = 0, p_2 = 1 \). For the convex utility case take \( U(0) = U(1) = 0, U(2) = 10, C = 8 \). In slot 1, since \( C > U(1) \), \( \pi_1 = U(1) \), the consumer will demand 0 power. In slot 2, the price will be \( \pi_2 = U(1) = 0 \), the consumer’s net benefit will be 0, the producer will receive \( \pi_2 = 0 \). In the duration-differentiated market (Theorem 6), at the competitive allocation the consumer gets 1 kW for 2 slots and pays \( U(2) = 10 \) (so her net benefit is 0), the producer will purchase 1 kW in the first slot and pay \( C = 8 \), and will use the free power \( p_1 \) in the second slot and get a net revenue of \( U(2) - 8 = 2 \).

For the concave utility case take \( U(0) = 0, U(1) = U(2) = 5, C = 2 \). Since \( U(1) > C \), the producer will supply 1 kW in slot 1 at price \( \pi_1 = C = 2 \). In slot 2, the consumer is indifferent between 1 kW and 0 kW. She may or not consume \( p_2 = 1 \) at price \( \pi_2 = 0 \). So the consumer’s net benefit will be \( U(1) - C = 3 \), the producer’s net profit is 0. In the duration-differentiated competitive allocation, the consumer will only purchase 1 kW in slot 2 at price 0, resulting in her net utility of 5, the producer’s net profit will again be zero.

**Remark 6.** Because of the inter-temporal nature of a consumer’s utility, her demand in slot \( t \) is contingent on her consumption in other slots. This contingent demand cannot be met by the system of spot markets, which is therefore not complete. This system of markets can be completed through forward markets, like the duration-differentiated market.

### 4. Conclusions

Flexible loads play a critical role in enabling deep renewable integration. They enable demand shaping to balance supply variability, and thus offer an effective alternative to conventional generation reserves. In this paper, we study a stylized model of flexible loads. These loads are modeled as requiring constant power level for a specified duration within a delivery window. The flexibility resides in the fact that the power delivery may occur at any subset of the total period.
We offered a complete characterization of supply adequacy, and an algorithm for allocating supply to meet these loads. We studied the implementation of a forward market for these duration differentiated loads. We presented the centralized solution that maximizes social welfare and characterized efficient competitive equilibria in this forward market. We then contrasted the outcomes in the forward market with those from a stylized real-time market for these loads. Our study suggests possible inefficiencies of the real-time market implementation.

The results in this paper can be expanded in several ways. In particular, the study of flexible loads with different power levels, the consideration of heterogeneous suppliers and consumers and the market operation in the presence of supply uncertainty are important directions for future work.

**Appendix A: Preliminary Majorization Based Results**

Let \( \mathbf{a} = (a_1, a_2, \ldots, a_T) \) and \( \mathbf{b} = (b_1, b_2, \ldots, b_T) \) be two non-negative vectors arranged in non-increasing order (that is, \( a_t \geq a_{t+1} \) and \( b_t \geq b_{t+1} \)). The majorization conditions for \( \mathbf{a} \prec \mathbf{b} \) in Definition 2 can equivalently be written as

(i) \( \sum_{s=1}^{t} b_s \leq \sum_{s=1}^{t} a_s \), for \( t = 1, 2, \ldots, T-1 \), and

(ii) \( \sum_{s=1}^{T} b_s = \sum_{s=1}^{T} a_s \).

The first condition is equivalent to the following condition: Let \( \mathbf{c} \) be any rearrangement of \( \mathbf{b} \); then, for any \( S \subset \{1, 2, \ldots, T\} \), there exists \( S' \subset \{1, 2, \ldots, T\} \) of the same cardinality as \( S \) such that

\[
\sum_{s \in S} c_s \leq \sum_{s \in S'} a_s.
\]

**Definition 6.** We define a 1 unit Robin Hood (RH) transfer on \( \mathbf{a} \) as an operation that:

(i) Selects indices \( t, s \) such that \( a_t > a_s \),

(ii) Replaces \( a_t \) by \( a_t - 1 \) and \( a_s \) by \( a_s + 1 \).

(iii) Rearranges the resulting vector in a non-increasing order.

**Lemma 2.** Let \( \tilde{\mathbf{a}} \) be a vector obtained from \( \mathbf{a} \) after a 1 unit RH transfer. Then, \( \mathbf{a} \prec \tilde{\mathbf{a}} \).

**Proof.** Note \( \tilde{\mathbf{a}} \) is a rearrangement of the vector \( \hat{\mathbf{a}} = (a_1, a_2, \ldots, a_t-1, \ldots, a_s+1, \ldots, a_T) \). For any subset \( S \) of \( \{1, 2, \ldots, T\} \),

(i) if \( t \in S, s \in S \) (or if \( t \notin S, s \notin S \)), then \( \sum_{r \in S} \hat{a}_r = \sum_{r \in S} a_r \).

(ii) if \( t \in S, s \notin S \), then \( \sum_{r \in S} \hat{a}_r = \sum_{r \in S} a_r - 1 \).

(iii) if \( t \notin S, s \in S \), then \( \sum_{r \in S} \hat{a}_r = a_s + 1 + \sum_{r \in S \setminus \{s\}} \hat{a}_r \leq a_t + \sum_{r \in S \setminus \{s\}} \hat{a}_r = a_t + \sum_{r \in S \setminus \{s\}} a_r \).

Therefore, \( \mathbf{a} \prec \tilde{\mathbf{a}} \) (see the equivalent condition in the definition of majorization). \( \square \)

**Lemma 3.** Suppose \( \mathbf{a} \prec \mathbf{b} \). If for any \( t, 1 \leq t \leq T \), the following conditions hold:
(i) \( a_j = b_j \) for all \( j < t \),
(ii) \( a_t - a_T \leq 1 \).

Then, (a) \( a_t = b_t \), and (b) \( a = b \).

**Proof of Lemma 3.** \( a \prec b \) and \( a_j = b_j \) for all \( j < t \) imply that
\[
\sum_{i=t}^s b_i \leq \sum_{i=t}^s a_i, \quad s = t, t+1, \ldots, T - 1, \\
\sum_{i=1}^T a_i = \sum_{i=1}^T b_i
\]
In particular, \( b_t \leq a_t \). If \( b_t < a_t \), then \( \sum_{i=t}^T b_i \leq (T - t + 1)(a_t - 1) \). On the other hand, since \( a_T \geq a_t - 1 \), \( \sum_{i=t}^T a_i > (T - t + 1)(a_t - 1) \). Therefore, \( \sum_{i=t}^T a_i \neq \sum_{i=t}^T b_i \), which contradicts the conditions of the lemma. Thus, \( b_t \) must be equal to \( a_t \). Reapplying the first part of the lemma for \( i = t + 1 \), then gives \( a_{t+1} = b_{t+1} \). Proceeding sequentially till \( T \) proves the second part of the lemma. \( \square \)

**Lemma 4.** Let \( a \prec b \) and \( a \neq b \). Then, there exists a 1 unit RH operation on \( a \) that gives a vector \( \tilde{a} \neq a \) satisfying \( a \prec \tilde{a} \prec b \).

**Proof.** Let \( t \) be the smallest index such that \( a_t \neq b_t \). Then, \( b_t < a_t \) since \( \sum_{i=t}^t b_i \leq \sum_{i=t}^t a_i \) and the two vectors have the same first \( t - 1 \) elements. Let \( s > t \) be the smallest index such that \( a_i - a_s > 1 \). Such \( s \) must exist, otherwise Lemma 3 would imply that \( a = b \). Consider a 1 unit RH transfer from \( t \) to \( s \). Let \( \tilde{a} \) be the resulting vector. Then, by Lemma 2, \( a \prec \tilde{a} \).

Also, if \( k \) is the number of elements of \( a \) with value equal to \( a_t \), then the number of elements of \( \tilde{a} \) with value equal to \( a_t \) is \( k - 1 \). Therefore, \( \tilde{a} \neq a \).

Further, it is clear that \( a \) and \( \tilde{a} \) have the same first \( t - 1 \) elements (since the RH operation depleted 1 unit from \( a_t \) and added it to \( a_s < a_t - 1 \), the non-increasing rearrangement would not change the \( t - 1 \) highest elements.) Similarly, \( a \) and \( \tilde{a} \) have the same elements from index \( s \) to \( T \). Further, from the definition of \( s \), \( a_j \geq a_s - 1 \) for \( t \leq j < s \). Since, \( \tilde{a}_t, \tilde{a}_{t+1}, \ldots, \tilde{a}_{s-1} \), must be a rearrangement of \( a_t - 1, a_{t+1}, \ldots, a_{s-1} \), it follows that \( \tilde{a}_j \geq a_t - 1 \) for \( t \leq j < s \).

We now prove that \( \tilde{a} \prec b \).
(i) If \( j < t \) or \( j \geq s \), then \( \sum_{i=1}^j \tilde{a}_i = \sum_{i=1}^j a_i \geq \sum_{i=1}^j b_i \).
(ii) If \( t \leq j < s \), then
\[
\sum_{i=1}^j \tilde{a}_i = \sum_{i=1}^{t-1} a_i + \sum_{i=t}^j \tilde{a}_i \\
= \sum_{i=1}^{t-1} b_i + \sum_{i=t}^j \tilde{a}_i \geq \sum_{i=1}^{t-1} b_i + (j - t + 1)(a_t - 1),
\]
(19)
where the first and second equalities are true because the first \(t - 1\) elements of \(a, \tilde{a}\) and \(b\) are the same, and last inequality follows from the fact that for \(t \leq i < s\), \(\tilde{a}_i \geq (a_t - 1)\). Moreover,

\[
\sum_{i=1}^{j} b_i = \sum_{i=1}^{t-1} b_i + \sum_{i=t}^{j} b_i \leq \sum_{i=1}^{t-1} b_i + (j - t + 1)b_t \leq \sum_{i=1}^{t-1} b_i + (j - t + 1)(a_t - 1),
\]

equation (20)

where the last inequality follows from \(b_t < a_t\). Equations (19) and (20) imply that \(\sum_{i=1}^{j} b_i \leq \sum_{i=1}^{t} \tilde{a}_i\), or \(\tilde{a} \prec b\).

\[\square\]

**Claim 1.** Let \(a \prec b\) with \(a \neq b\). Then, there exists a finite sequence of 1 unit RH transfers that can be applied on \(a\) to get \(b\).

**Proof.** The claim is established using Lemmas 3 and 4. Let \(a^0 = a\).

(i) For \(n = 1, 2, \ldots\), if \(a^{n-1} \neq b\), use Lemma 4 to construct \(a^n \neq a^{n-1}\) such that \(a^{n-1} \prec a^n \prec b\).

Then, \(a^n \neq a^m\) for any \(m < n - 1\) (otherwise, we would have \(a^n = a^{n-1} \prec a^n \implies a^n = a^{n-1}\)).

(ii) If \(a^{n-1} = b\), stop.

Since there are only finitely many non-negative integer valued vectors that majorize \(b\), this procedure must eventually stop and it can do so only if \(a^{n-1} = b\), proving the claim. \[\square\]

**Appendix B: Proof of Theorem 1**

We require with the following intermediate result.

**Lemma 5.** For a given demand profile, if \(a\) is an exactly adequate supply profile, then any supply profile \(b\) satisfying \(a \prec b\) is also exactly adequate.

**Proof.** Without loss of generality, we will assume that \(a\) and \(b\) are arranged in non-increasing order. Since \(b\) can be obtained from \(a\) by a sequence of 1 unit RH transfers (see Claim 1, Appendix A), we simply need to prove that a 1 unit RH transfer preserves exact adequacy. Consider an exactly adequate profile \(a\) and let \(A\) be the corresponding allocation function. Consider a 1 unit RH transfer from time \(t\) to \(s\) (without rearrangement) that gives a new profile \(\tilde{a}\). Let \(i\) be a load for which \(A(i, t) = 1\) but \(A(i, s) = 0\). Such an \(i\) must exist because \(\sum_j A(j, t) = a_t > a_s = \sum_j A(j, s)\).

Under the new profile, the allocation rule

\[
\tilde{A}(j, r) = \begin{cases} 
A(j, r) & r \neq t, s \text{ or } j \neq i \\
0 & r = t, j = i \\
1 & r = s, j = i 
\end{cases}
\]

establishes exact adequacy. \[\square\]
Proof of Theorem 1 (a): Observe that d is exactly adequate: the allocation function \( A(i, t) = \mathbb{1}_{\{t \leq h_i\}} \) meets the exact adequacy requirements under d. Therefore, any p satisfying p \( \succ d \) is also exactly adequate.

To prove necessity, suppose p is exactly adequate and A is the corresponding allocation function. Consider any set \( S \subset \{1, 2, \ldots, T\} \) of cardinality s. Then, because \( A(i, t) \in \{0, 1\} \) and \( \sum_{t=1}^{T} A(i, t) = h_i \), it follows that
\[
\sum_{t \in S} A(i, t) \leq \min(s, h_i) = \sum_{t=1}^{s} \mathbb{1}_{\{t \leq h_i\}}. \tag{22}
\]

Summing over i,
\[
\sum_{i=1}^{N} \sum_{t \in S} A(i, t) \leq \sum_{i=1}^{N} \sum_{t=1}^{s} \mathbb{1}_{\{t \leq h_i\}}
\]
\[
\implies \sum_{t \in S} \sum_{i=1}^{N} A(i, t) \leq \sum_{i=1}^{s} \sum_{t=1}^{N} \mathbb{1}_{\{t \leq h_i\}}
\]
\[
\implies \sum_{t \in S} p_t \leq \sum_{t=1}^{s} d_t \quad \text{for all } S \text{ with } |S| = s
\]
\[
\implies \sum_{t=1}^{s} p_t^j \leq \sum_{t=1}^{s} d_t \tag{23}
\]

For \( s = T \), the inequality in (22) becomes an equality resulting in an equality in (23). □

Proof of Theorem 1 (b): To prove necessity, suppose p is adequate and A is the corresponding allocation function. Then, clearly, for any set \( S \subset \{1, \ldots, T\} \) of cardinality s
\[
\sum_{t \in S} p_t \geq \sum_{t \in S} \sum_{i=1}^{N} A(i, t) \tag{24}
\]

Further, because \( A(i, t) \in \{0, 1\} \) and \( \sum_{t=1}^{T} A(i, t) = h_i \), it follows that
\[
\sum_{t \in S} A(i, t) \geq \max(h_i - s + 1, 0) = \sum_{t=s}^{T} \mathbb{1}_{\{t \leq h_i\}}
\]

Summing over i,
\[
\sum_{i=1}^{N} \sum_{t \in S} A(i, t) \geq \sum_{i=1}^{N} \sum_{t=s}^{T} \mathbb{1}_{\{t \leq h_i\}}
\]
\[
\implies \sum_{t \in S} \sum_{i=1}^{N} A(i, t) \geq \sum_{t=s}^{T} \sum_{i=1}^{N} \mathbb{1}_{\{t \leq h_i\}}
\]
\[
\implies \sum_{t \in S} \sum_{i=1}^{N} A(i, t) \geq \sum_{t=s}^{T} d_t. \tag{25}
\]
Combining (24) and (25) proves the necessity conditions.

To prove sufficiency, let $\Delta = \sum_{t=1}^{T} p_t - \sum_{t=1}^{T} d_t$. Then, $\Delta \geq 0$. Consider a new demand profile $d^\Delta$ defined as $d^\Delta := d_t + \Delta \mathbb{1}_{t \leq 1}$. This new demand profile corresponds to the original demand profile augmented with $\Delta$ “fictitious loads” each requiring 1 unit of energy for 1 time slot. It is easy to see that $p \succ^w q$ implies $p \succ d^\Delta$. Therefore, by Theorem 1, $p$ is exactly adequate for the augmented demand profile $d^\Delta$ which implies that it must be adequate for the original demand profile $q$. □

Appendix C: Proof of Theorem 2

Since $p$ is adequate, there must exist at least one allocation function $B(i,t)$ such that

$$\sum_t B(i,t) = h_i, \quad \sum_i B(i,t) \leq p_t.$$ 

Let $B_1$ be the set of loads served at time 1 under the allocation rule $B(i,t)$. If $A_1 \neq B_1$, pick a load $i \in A_1 \setminus B_1$ and $j \in B_1 \setminus A_1$. Then, $h_i \geq h_j$. Therefore, there must exist a time $s > 1$ such that $B(i,s) = 1$ but $B(j,s) = 0$. Consider a new allocation rule $B^1(\cdot,\cdot)$ obtained by swapping the load $i$ at time $s$ and the load $j$ at time 1, that is, the new allocation rule $B^1$ is identical to $B$ except that for $t = 1, s$

$$B^1(i,t) = B(j,t) \quad \text{and} \quad B^1(j,t) = B(i,t).$$

It is straightforward to establish that the new allocation rule satisfies the adequacy requirements. One can proceed with this swapping argument one load at a time until the set of loads being served at time 1 is equal to $A_1$.

For time $t > 1$, the same swapping argument works after $h_i$ are replaced by the number of time slots each load needs to be served from time $t$ to the final time, which is precisely $x_i(t)$. □

Appendix D: Proof of Theorem 3

We first observe that the strategy described in the theorem is a valid strategy under Case B since it only uses the information available until time $t$. Clearly, the prescribed strategy is valid under Case A as well. We will now show that it yields an adequate supply in the end.

From the definition of $a_1$, we have $(p_1 + a_1) \succ^w d_T$. Now assume that $(p_t + a_1, \ldots, p_{t-1} + a_{t-1}) \succ^w (d_{T-t+2}, \ldots, d_T)$. Then, there exists an $a_t$ such that $(p_1 + a_1, p_2 + a_2, \ldots, p_{t-1} + a_{t-1}, p_t + a_t) \succ^w (d_{T-t+1}, d_{T-t+2}, \ldots, d_T)$. Thus, the supplier’s optimization problem at time $t$ has a solution. This is given by min $a_t$ subject to

$$p_t + a_t \geq d_T$$
$$p_s + a_s + p_t + a_t \geq d_T + d_{T-1} \quad \text{for all} \ s < T$$
$$\sum_{s=s_1, s_2, \ldots, s_k} (p_s + a_s) + (p_t + a_t) \geq d_{T-k} + \cdots + d_T \quad \text{for all} \ s_1 < \cdots < s_k < t \ \text{and} \ k \leq t - 1$$
It is clear that the prescribed decision strategy results in an adequate net supply since at time $T$, the supplier’s optimization ensures that $\mathbf{p} + \mathbf{a} \succ w \mathbf{d}$.

We now argue that the prescribed strategy is the optimal one. Let $\mathbf{b} = (b_1, \ldots, b_T) \geq 0$ be the optimal sequence of purchase decisions made in Case A. Then, adequacy constrain implies that $\mathbf{p} + \mathbf{b} \succ w \mathbf{d}$. Because, the total supply is now adequate, it can be allocated to consumers according to the LLDF rule.

Starting at time 1, we must have $b_1 \geq a_1 = (d_T - p_1)^+$. Otherwise, $(p_1 + b_1) < d_T$ which means that $\mathbf{p} + \mathbf{b}$ could not be adequate for the given demand. Suppose $b_1 > a_1$. Let $\mathcal{B}_1$ be the set of consumers served under a LLDF allocation rule at time 1 when the supply is $\mathbf{p} + \mathbf{b}$ and $\mathcal{A}_1$ be the set of consumers served under a LLDF allocation rule at time 1 when the supply is $\mathbf{p} + \mathbf{a}$. Since, $p_1 + b_1 \geq p_1 + a_1$, $\mathcal{B}_1 \supset \mathcal{A}_1$. Let $i \in \mathcal{B}_1 \setminus \mathcal{A}_1$. Because, we know that $\mathbf{p} + \mathbf{a}$ is adequate and that consumer $i$ is not being served at time 1, there must be a future time $s$ such that consumer $i$ is served at $s$ under the supply $\mathbf{p} + \mathbf{a}$ but not under $\mathbf{p} + \mathbf{b}$. We now consider a modification of the purchase decisions $\mathbf{b}$ constructed as follows:

1. Change $b_1$ to $b_1 - 1$ and not serve consumer $i$ at time 1.
2. At time $s$, if there was excess power available when the supply was $\mathbf{p} + \mathbf{b}$, then use it to serve consumer $i$ at time $s$. Otherwise, change $b_s$ to $b_s + 1$ with the new unit of power given to consumer $i$.

Denote this modified vector of decisions by $\tilde{\mathbf{b}}$. It is clear that this modified decision is also optimal since it ensures adequacy and its total cost does not exceed the cost of $\mathbf{b}$. We can repeat this argument until $\tilde{b}_1 = a_1$. Therefore, there is an optimal vector of purchase decisions of the form $(a_1, \tilde{b}_2, \ldots, \tilde{b}_T)$.

We now proceed by induction. Suppose that the optimal decision vector is $(a_1, \ldots, a_{t-1}, b_t, \ldots, b_T)$. Then, $b_t \geq a_t$ otherwise $\mathbf{p} + \mathbf{b} \not\succ w \mathbf{d}$. If $b_t > a_t$, we can use the same rearrangement argument used at time 1 to construct modifications of the optimal decision vector that is of the form $(a_1, \ldots, a_t, \tilde{b}_{t+1}, \ldots, \tilde{b}_T)$. Continuing sequentially till the final time $T$, we conclude that $(a_1, \ldots, a_T)$ must be optimal.

For any vector $\mathbf{x} = (x_1, \ldots, x_T)$ define $F(\mathbf{x}) := \max S \subseteq \{1, 2, \ldots, T\} \left( \sum_{s=T-|S|+1}^{T} d_s - \sum_{r \in S} x_r \right)^+$. If $a_1 > 0$, then $d_T > p_1$ and therefore the maximum achieving set $S$ in definition of $F(\mathbf{p})$ must contain time index 1. Let $\mathbf{p}(1) = (p_1 + a_1, p_2, \ldots, p_T)$. Then, $S$ also achieves the maximum in the definition of $F(\mathbf{p}(1))$. Therefore,

$$F(\mathbf{p}(1)) = F(\mathbf{p}) - a_1. \quad (26)$$
Let $\mathbf{p}(t-1)$ be the supply profile after the first $t-1$ purchase decisions. If $a_t > 0$, then there exists a set $\mathcal{T} \subset \{1, \ldots, t\}$ containing $t$ such that

$$\sum_{s=T-|\mathcal{T}|+1}^{T} d_s - \sum_{r \in \mathcal{T}} p_r(t-1) > 0$$  \hspace{1cm} (27)

Let $\mathcal{T}^*$ be the set for which the above difference is the largest. Then, the maximum achieving set $\mathcal{S}$ in the definition of $F(\mathbf{p}(t-1))$ must contain $\mathcal{T}^*$. Define $\mathbf{p}(t) = (p_1 + a_1, p_2 + a_2, \ldots, p_t + a_t, p_{t+1}, \ldots, p_T)$. Then, $\mathcal{S}$ also achieves the maximum in the definition of $F(\mathbf{p}(t))$. Therefore,

$$F(\mathbf{p}(t)) = F(\mathbf{p}(t-1)) - a_t.$$  \hspace{1cm} (28)

Combining (28) for all $t$ gives

$$F(\mathbf{p}(T)) = F(\mathbf{p}) - \sum_{t=1}^{T} a_t \quad \implies \quad \sum_{t=1}^{T} a_t = F(\mathbf{p}) - F(\mathbf{p}(T)).$$

Note that

$$F(\mathbf{p}) = \max_{\mathcal{S} \subset \{1, 2, \ldots, T\}} \left( \sum_{s=T-|\mathcal{S}|+1}^{T} d_s - \sum_{r \in \mathcal{S}} p_r \right)^+ = \max_{1 \leq t \leq T} \left( \sum_{s=T-t+1}^{T} d_s - \sum_{r=T-t+1}^{T} p_r^+ \right)^+ \hspace{1cm} (29)$$

Also note that since $\mathbf{p}(T) = \mathbf{p} + \mathbf{a}$, and $\mathbf{p} + \mathbf{a} \succeq^w \mathbf{d}$, it implies that

$$F(\mathbf{p}(T)) = \max_{1 \leq t \leq T} \left( \sum_{s=T-t+1}^{T} d_s - \sum_{r=T-t+1}^{T} p_r^+(T) \right)^+ = 0.$$  

Therefore, (29) amounts to

$$\sum_{t=1}^{T} a_t = \max_{1 \leq t \leq T} \left( \sum_{s=T-t+1}^{T} d_s - \sum_{r=T-t+1}^{T} p_r^+ \right)^+ = \max_{1 \leq t \leq T} \left( \sum_{s=t}^{T} d_s - \sum_{r=t}^{T} p_r^+ \right)^+$$

which proves the theorem. \hfill \Box

**Appendix E: Proof of Theorem 5**

**Proof of Part (a)** Consider any choice of $\mathbf{y}$ so that the total supply is $\mathbf{x} = \mathbf{p} + \mathbf{y}$. Without loss of generality, assume that $\mathbf{x}$ is arranged in non-increasing order. For $t = 1, \ldots, T$, define $\delta_t = \mathbf{p} ...
where the inequality follows from the fact that $\delta_0 = 0$. Then, $\delta_0 \leq \delta_1 \leq \delta_2 \leq \ldots \leq \delta_T$. For any demand duration vector that can be served with $x$, the total utility of consumers is

$$
\sum_{t=1}^{T-1} (d_t - d_{t+1})U(t) + d_T U(T) = \sum_{t=1}^{T} d_t(U(t) - U(t-1)) = \sum_{t=1}^{T} d_t \delta_t
$$

$$
= \delta_1(d_1 + d_2 + \ldots + d_T) + (\delta_2 - \delta_1)(d_2 + d_3 + \ldots + d_T)
$$

$$
+ (\delta_3 - \delta_2)(d_3 + \ldots + d_T) + \ldots + (\delta_T - \delta_{T-1})d_T
$$

$$
= \sum_{j=1}^{T} (\delta_j - \delta_{j-1}) \left[ \sum_{t=j}^{T} d_t \right]
$$

(30)

where the inequality follows from the fact that $\delta_j - \delta_{j-1} \geq 0$ and $x \succ^w d$. The profile $d_t = x_t$ for all $t$ is a valid demand profile that achieves the upper bound on consumer utility. Therefore, the optimal contracts offered when the supply is $x$ are: the first $x_T$ consumers get $T$ duration, the next $x_{T-1} - x_T$ get $T-1$ and in general $x_i - x_{i+1}$ get duration $i$. In total, $x_1$ consumers are served. If $x = p$, the first $p_T$ consumers get $T$ duration, the next $p_{T-1} - p_T$ get $T-1$ and in general $p_i - p_{i+1}$ get duration $i$. Since $x = p + y$ and $y \geq 0$, each consumer’s contract under $x$ is no less than its contract under $p$.

Therefore, the social welfare problem can be restated as follows: Let $(h_1, h_2, \ldots, h_N)$ be the utility maximizing contracts for the $N$ consumers under the supply $p$. Choose a vector $z = (z_1, z_2, \ldots, z_N)$ of contract increments to maximize

$$
\sum_{i=1}^{N} (U(h_i + z_i) - Cz_i)
$$

(31)

The summation can be maximized by maximizing each term separately which gives that $z_i = T - h_i$ if $h_i \geq k^*$ and 0 otherwise.

**Proof of part (b)** Consider any choice of $y$ so that the total supply is $x = p + y$. Let $\sum_{t=1}^{T} x_t = N t(x) + R(x)$ where $t(x)$ is a non-negative integer and $R(x) < N$. Then, the non-increasing increment property of the utility implies that total consumer utility is maximized if $R(x)$ consumers get contracts of duration $t(x) + 1$ and $N - R(x)$ consumers get contracts of duration $t(x)$. When $x = p$, $\sum_{t} p_t$ consumers get contract of duration 1. Thus, each consumer’s contract under $x$ is no less than its contract under $p$. Therefore, the social welfare problem can be restated as follows: Let $(h_1, h_2, \ldots, h_N)$ be the utility maximizing contracts for the $N$ consumers under the supply $p$. Choose a vector $z = (z_1, z_2, \ldots, z_N)$ of contract increments to maximize

$$
\sum_{i=1}^{N} (U(h_i + z_i) - Cz_i)
$$

(32)

The summation can be maximized term-by-term which yields $z_i = k^* - h_i$. 

□
Appendix F: Proof of Theorem 6

Proof of Part (a) Subject to prices $\pi(t) = U(t)$, the profit maximization problem is given by

$$\max_{n \geq 0, y \geq 0} \sum_{t=1}^{T} (n_i U(t) - Cy_t)$$

subject to $d_t = \sum_{t=1}^{T} n_i$

$p + y \succ w d$

but $n_i = \sum_{j=1}^{N} I_{(h_j = i)}$, then $\sum_{t=1}^{T} n_i U(t)$ is equal to $\sum_{i=1}^{N} U(h_i)$. Thus, the solution of the profit maximization problem is equivalent to the solution of the social welfare optimization problem. Subject to these prices, consumers obtain zero welfare under any allocation. Hence, an efficient competitive equilibrium exists. \qed

Proof of part (b) Prices are given by $\pi(h) = \min(C, U(1))h = \mu h$. The case in which $\pi(h) = U(1)h$ is straightforward. If $\pi(h) = Ch$, it is clear that consumers maximize their surplus by choosing contracts of duration $k^*$. This is a result of the non-increasing increments on $U(h)$ and the condition $U(k^*) - U(k^* - 1) \geq C$. Hence, the bundle of contracts $n_k^* = N$ maximizes consumers surplus. Subject to prices $\pi(h) = Ch$, supplier revenue is given by $C \sum_{t=1}^{T} n_i t - C \sum_{t=1}^{T} y_t = C \sum_{t=1}^{T} d_t - C \sum_{t=1}^{T} y_t = C(\sum_{t=1}^{T} d_t - \frac{1}{\mu} \sum_{t=1}^{T} p_t)$. The bundle, $n_k^* = N$, achieves the upper bound. Hence, an efficient competitive equilibrium exists. \qed

Appendix G: Proof of Theorem 4

Clearly, any supply allocation $A(t)$ that respects (3) also satisfies (2).

To prove the converse, first assume $r > 0$. Consider an $A(t)$ satisfying (2). Then, $A(t) > 0$ for at least $k + 1$ time slots and $A(t) = m$ for at most $k$ time slots. Pick $k + 1$ time slots with largest value of $A(t)$. Define $A^1(t) = 1$ at the selected slots and 0 otherwise. Let $B(t) = A(t) - A^1(t)$. Then, $B(t) \in \{0, 1, \ldots, m - 1\}$ and $\sum_{t} B(t) = k(m - 1) + (r - 1)$. If $r = 0$, then $A(t) > 0$ for at least $k$ time slots and $A(t) = m$ for at most $k$ time slots. Pick $k$ time slots with largest value of $A(t)$. Define $A^1(t) = 1$ at the selected slots and 0 otherwise. Let $B(t) = A(t) - A^1(t)$. Then, $B(t) \in \{0, 1, \ldots, m - 1\}$ and $\sum_{t} B(t) = k(m - 1)$. Thus, we can always write $A(t)$ as

$$A(t) = A^1(t) + B(t),$$

where $A^1(t) \in \{0, 1\}$, $\sum_{t} A^1(t) = k + \mathbb{1}_{(r > 0)}$, $B(t) \in \{0, 1, \ldots, m - 1\}$ and $\sum_{t} B(t) = k(m - 1) + (r - 1)^+$. Now, if $(r - 1) > 0$, then $B(t) > 0$ for at least $k + 1$ slots and $B(t) = m - 1$ for at most $k$ slots. Pick $k + 1$ time slots with largest value of $B(t)$. Define $A^2(t) = 1$ at the selected slots and 0 otherwise. Let
\(C(t) = B(t) - A^2(t)\). Then, \(C(t) \in \{0, 1, \ldots, m - 2\}\) and \(\sum_t B(t) = k(m - 2) + (r - 2)\). If \((r - 1)^+ = 0\), then \(B(t) > 0\) for at least \(k\) time slots and \(B(t) = m\) for at most \(k\) time slots. Pick \(k\) time slots with largest value of \(B(t)\). Define \(A^2(t) = 1\) at the selected slots and 0 otherwise. Let \(C(t) = B(t) - A^2(t)\). Then, \(C(t) \in \{0, 1, \ldots, m - 2\}\) and \(\sum_t C(t) = k(m - 2)\). Thus, we can write \(A(t)\) as

\[
A(t) = A^1_t + A^2_t + C(t), \tag{34}
\]

where \(A^2_t \in \{0, 1\}\), \(\sum_t A^2_t = k + 1_{(r > 1)}\), \(C(t) \in \{0, 1, \ldots, m - 2\}\) and \(\sum_t C(t) = k(m - 2) + (r - 2)^+\). Continuing sequentially, we can decompose \(A(t)\) into \(A^1(t), \ldots, A^m(t)\). \(\square\)
References


