Shapley Value Estimation for Compensation of Participants in Demand Response Programs

Gearóid O'Brien, Abbas El Gamal, and Ram Rajagopal

***Abstract*—Designing fair compensation mechanisms for demand response (DR) is challenging. This paper models the problem in a game theoretic setting and designs a payment distribution mechanism based on the Shapley value (SV). As exact computation of the SV is in general intractable, we propose estimating it using a reinforcement learning algorithm that approximates optimal stratified sampling. We apply this algorithm to a DR program that utilizes the SV for payments and quantify the accuracy of the resulting estimates.**

***Index Terms*—Economics, power system economics.**

I. INTRODUCTION

DEMAND response (DR) refers to the increased flexibility in usage of electricity due to the intentional adjustment of end users' consumption. Such flexibility helps to alleviate many challenges with modern power systems by providing an inexpensive source of capacity. Motivated by a number of long-term considerations such as the growing number of electric vehicles, the proliferation of renewable energy generated from nondispatchable sources, and the worrying trajectory of climate change and its consequences, DR can play a role in balancing electricity supply and demand by appropriately adjusting the aggregate load profile.

Designing, implementing, and operating large-scale DR programs is a nontrivial task and has been extensively investigated in the literature; see for example [1]. Several papers have investigated mechanisms for various types of response capabilities [2]–[7].

The administration of DR programs is often left to a third party operator known as a "load serving entity" (LSE) or "aggregator." The LSE coordinates the participants, where each participant may be a user or a user load if such granularity is available. By employing an appropriate incentive, the LSE can expect to adjust the aggregate consumption profile of the participants. The net effect of DR (the difference between the baseline and the new expected aggregate consumption profile as illustrated in Fig. 1) may be considerably equivalent to electricity generation, and so the LSE can participate in the wholesale electricity marketplace.

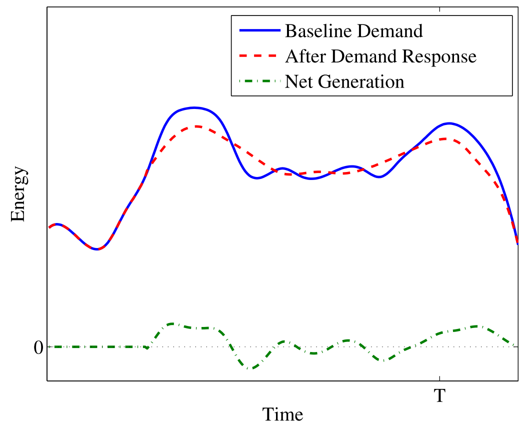


Fig. 1. Net effect of DR can be considered equivalent to electricity generation.

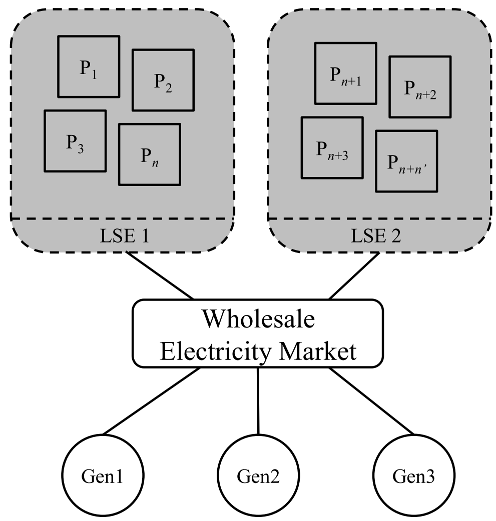


Fig. 2. LSEs participating in the wholesale electricity marketplace alongside traditional generators.

Fig. 2 illustrates this concept. The figure depicts two LSEs together with three generators, interacting with the wholesale electricity marketplace. The first LSE has n participants ($P_1, P_2, \ldots, P_n$) and the second has n' participants ($P_{n+1}, P_{n+2}, \ldots, P_{n+n'}$).

There is an inherent uncertainty present in this formulation, as net generation due to DR is contracted *a priori* but measured and verified based on real-time metering. This may

Manuscript received June 4, 2014; revised October 20, 2014; accepted November 16, 2014. Date of publication May 5, 2015; date of current version October 17, 2015. The work of G. O'Brien is supported by the Stanford Graduate Fellowship. The work of A. El Gamal is supported by the TomKat Center. The work of R. Rajagopal is supported in part by the Powell Foundation Fellowship and in part by the TomKat Center. Paper no. TSG-00535-2014.

G. O'Brien and A. El Gamal are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: gobrien@stanford.edu).

R. Rajagopal is with the Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2015.2402194

be manifested in two ways. First, the participants themselves may agree to a certain level of load adjustment in advance, but the real-time metering may well reveal a different level of adjustment. Such a setup is often how DR schemes are currently operated. Second, the LSE may attempt to forecast the aggregate load adjustment in order to determine the optimal quantity to offer to the market. Actual adjustments are again observed in real time. Such uncertainty increases the difficulty and cost of integrating DR into the wholesale market. Electric utility companies hold the LSE responsible for most of the program details, including performance, marketing, enrollment, and compensation to the participants.

This paper, which is a more complete and detailed version of [8], focuses on the last problem, which is, how to appropriately compensate the participants in a DR program. We model the interaction between participants in a DR program as a cooperative game and propose a payment distribution scheme based on the Shapley value (SV) [9], a well-known concept in cooperative games. Examining Fig. 2 again, participants in each shaded area take part in cooperative games. Ensuring that participants remain enrolled in a DR scheme—and that it is also appealing to new participants—relies in part on a fair and attractive compensation mechanism. The SV can often be shown to be such a mechanism.

Compensation mechanisms based on game theory have been investigated recently. Zhu *et al.* [10] and Saad *et al.* [11] investigated schemes based on noncooperative games in which consumers make choices independently under parameterized utility functions. In this paper, we focus on cooperative games in which groups of consumers work together to reach the goal of the DR program. Haring *et al.* [12] also detailed an analysis based on cooperative game theory where the authors proposed to optimize the LSEs offering of daily reserve capacity contracts. In contrast to their approach, we present a practical method for distributing compensation to participants in DR programs via the SV.

While noncooperative game theory is concerned mainly with the moves participants should rationally make in response to other participants, cooperative game theory is primarily focused on interactions among coalitions of participants, the value of each coalition, and especially how this value can be distributed among the participants of such a coalition. In general, cooperative games differ from noncooperative games in that binding agreements are possible before the start of the game. For DR programs, the LSE puts in place such a binding agreement (i.e., the value function), hence, the programs can be well modeled using cooperative game theory.

The SV has been previously used in power systems to study electrical energy generation and transmission. In [13], the SV is used to allocate transmission service costs among network users in energy markets. In [14], the aggregation of wind power producers is studied using coalition game theory and show that in this case, the nature of the resulting game does not lend itself to use the SV.

The contributions of this paper are as follows.

1) In Section II, we introduce the concept of using the SV as a method for LSEs to compensate DR participants.
2) As the SV can be difficult to compute, in Section III, we propose a new method of estimating it using a reinforcement learning algorithm. This is a general contribution, and can be readily applied outside the realm of DR schemes.
3) We apply our proposed SV estimation method to an example of DR program that provides reserve in Section IV. This also helps to demonstrate the effectiveness of our method.

TABLE I
LIST OF KEY SYMBOLS

$\mathcal{X}$	Set of all n DR participants, indexed i.
$\mathcal{X}_{-i}$	Set of $n-1$ DR participants, removing i.
$\mathcal{S}$	Subset (of size $j, 0 \leq j \leq n-1$) of participants in $\mathcal{X}_{-i}$.
$v(\mathcal{S})$	Function calculating the worth of the participants in $\mathcal{S}$.
X_i	Contribution of participant i to the DR scheme.
ϕ_i	SV for i.
$\rho_i(\mathcal{S})$	Marginal contribution of i to $\mathcal{S}$. (Note: if $\mathcal{S}$ is chosen randomly, $\rho_i(\mathcal{S})$ is (RV).)
N_j^i	Number of samples taken from stratum j.
$\rho_{k,j}^i$	k^{th} random sample from stratum j for i. (RV)
$\overline{\rho}_j^i$	Sample mean of the samples taken from stratum j. (RV)
$\mu_{j,i},\ \sigma_{j,i}^2$	Mean and variance of $\overline{\rho}_j^i$.
T_i	Average of the n values of $\overline{\rho}_j^i$. (RV)
σ_i^2	Variance of T_i.
$\hat{\phi}_i$	Maximum likelihood estimate of ϕ_i. (RV)

To help the reader, we provide definitions of the key variables used throughout this paper are shown in Table I. Random variables are indicated by RV.

II. Problem Setup

Consider a set $\mathcal{X} = \{1, 2, \ldots, n\}$ of n participants in a DR scheme. For the set $\mathcal{S} \subseteq \mathcal{X}$, define the value function (also known as the characteristic function) $v(\mathcal{S})$ as a function $v : 2^n \to \mathbb{R}$. The function $v(\mathcal{S})$ represents the total expected payoff the participants in $\mathcal{S}$ achieve together. As an example, assume the goal of the DR program is to provide reserve by reducing load levels and let $X_i \in \mathbb{R}$ be the amount by which participant i reduces its load. Then, the value function may be taken as

$$v(\mathcal{S}) = q \sum_{i \in \mathcal{S}} X_i \tag{1}$$

where $q \geq 0$ is a constant that converts energy into a dollar value. This is only an illustrative example and a more practical choice of the value function will be presented in Section IV.

The LSE wishes to distribute the total payoff $v(\mathcal{X})$ among the n participants in a fair manner that depends on their contribution to the goal of the DR scheme. We denote the payoff assigned to participant i by $\phi_i(v)$. Hence, the total payoff is

$$v(\mathcal{X}) = \sum_{i=1}^{n} \phi_i(v). \tag{2}$$

Shapley proposed a solution to the distribution of the total payoff that is both unique and fair [15]. Defining $\mathcal{X}_{-i}$ to be the set of all participants after removing participant i, the marginal contribution of participant i to a coalition $\mathcal{S}$, $\mathcal{S} \subseteq \mathcal{X}_{-i}$, is

$$\rho_i(\mathcal{S}) = v(\mathcal{S} \cup \{i\}) - v(\mathcal{S}). \tag{3}$$

Furthermore, we define $\mathcal{R}$ to be one of the $n!$ permutations of the participants in $\mathcal{X}$, and $\mathcal{P}_i^{\mathcal{R}}$ to be the ordered set preceding i in $\mathcal{R}$. The SV is defined as the mean marginal contribution of that participant to all possible coalitions of the other participants, that is

$$\phi_i = \frac{1}{n!}\sum_{\mathcal{R}} \rho_i\left(\mathcal{P}_i^{\mathcal{R}}\right). \tag{4}$$

Remark 1: For brevity, we suppress the v and denote the SV by ϕ_i. The value function should be clear from the context.

The SV is an intuitive concept as participants who contribute more to the coalitions that include themselves should receive a higher compensation. A cursory examination of the value function in (1) shows that $\phi_i = qX_i$, although for nontrivial value functions the SV is often intractable.

Understanding the advantage of using the SV when compared to other conventional distribution methods is a key contribution of this paper. The LSE, by participating in the wholesale electricity marketplace, realizes a revenue stream (or payoff) which must then be distributed fairly among the participants. The SV solution concept provides a fair and unique method for distributing this revenue. Fairness is defined as satisfying the following four axioms, which a payment distribution scheme could reasonably be expected to satisfy.

1) *Efficiency:* The entire payoff is divided among the participants (no excess remains).
2) *Symmetry:* Two participants that contribute equally are rewarded equally.
3) *Null Player:* Participants that do not contribute receive no payoff.
4) *Linearity:* The total payoff rewarded for contributing to two DR programs is the sum of the payoffs that would be awarded for contributing to each of the two programs individually.

Surprisingly, the SV can be shown to be the only payment distribution method that satisfies these four axioms, with the added benefit that the solution is unique.

It should be noted that although the SV is a fair method for distributing the total payoff, it does not by itself address issues of stability regarding the grand coalition of all n participants. The set of all payoff vectors that results in a stable grand coalition forms the core (note that the core can be empty). To be precise, consider a payoff vector x where x_i is the payoff participant i receives. If there exists a coalition $\mathcal{S}$ whose combined payoff is less than what the coalition can achieve by acting alone [i.e., $\sum_{i\in\mathcal{S}} x_i < v(\mathcal{S})$], then the members of coalition $\mathcal{S}$ will have a tendency to break away from the grand coalition and form the set $\mathcal{S}$. Therefore, the payoff vector x is unstable and is not in the core.

When the cooperative game is convex, it can be guaranteed that the core is nonempty, that is, there exists a payoff vector that results in a stable grand coalition. Intuitively, a convex game means that the incentive for a participant joining a coalition increases as the number of participants in the coalition grows. The SV can be shown to be in the core for convex games, hence, defines a stable payoff vector for such games.

A. Shapley Value and Demand Response Schemes

In order to utilize the SV as the distribution mechanism of a DR scheme, the scheme itself must be representable as a value function. This function, defined over subsets of the participants, returns the payoff that will be received by the LSE. Value functions are particularly suited to schemes where loads are controllable to some degree. The precise formulation of the value function is left to the designer of the DR scheme.

When the SV is used to distribute a payoff to the participants, the value function must be supermodular. The corollary is if the SV is used to distribute a penalty rather than a payoff, it must be submodular. Such a situation may arise if the participants are given an ex-ante payment for joining the DR program, from which a penalty for noncompliance is deducted.

The assumption is the SV is calculated after each DR event has occurred. Since it is expected that such events will become more frequent and will involve an increasing number of participants, it is necessary to be able to efficiently compute the SV.

B. Computational Complexity

The most challenging aspect in utilizing the SV is its computational intractability. For a DR program with n participants, the value function may need to be evaluated $n\,2^n$ times. A modest DR program with $n = 500$ requires about 1.5×10^{153} function evaluations. Approximation approaches that rely on simple schemes to selectively perform function evaluations have been proposed to mitigate this problem. Shapley proposed a Monte Carlo random sampling technique [16], which was subsequently extended in [17] and [18] to achieve desired accuracy levels in polynomial time. Such mechanisms neither exploit relevant properties of the value function nor enforce important constraints such as the efficiency axiom. Section III will detail a computationally efficient method for estimating the SV via sampling.

III. Estimating the Shapley Value

Before describing our algorithm for estimating the SV, we need the following alternative formulation. Grouping the terms in (4) in which the participants to the left of i are the same gives the alternative form for the SV

$$\phi_i = \sum_{\mathcal{S}\subseteq\mathcal{X}_{-i}} \frac{|\mathcal{S}|!(|\mathcal{X}|-|\mathcal{S}|-1)!}{n!}\rho_i(\mathcal{S}). \tag{5}$$

Further grouping by the number of terms in $\mathcal{S}$, defining $j = |\mathcal{S}|$, and recalling that $n = |\mathcal{X}|$ we obtain

$$\begin{aligned}\phi_i &= \sum_{j=0}^{n-1} \sum_{\substack{\mathcal{S}\subseteq\mathcal{X}_{-i}\\|\mathcal{S}|=j}} \left(\frac{j!(n-j-1)!}{n!}\right)\rho_i(\mathcal{S})\\ &= \frac{1}{n}\sum_{j=0}^{n-1} \sum_{\substack{\mathcal{S}\subseteq\mathcal{X}_{-i}\\|\mathcal{S}|=j}} \left(\frac{(n-1)!}{j!(n-1-j)!}\right)^{-1}\rho_i(\mathcal{S})\\ &= \frac{1}{n}\sum_{j=0}^{n-1} \sum_{\substack{\mathcal{S}\subseteq\mathcal{X}_{-i}\\|\mathcal{S}|=j}} \binom{n-1}{j}^{-1}\rho_i(\mathcal{S}).\end{aligned}$$

The inner sum can be considered as an expectation over a uniform probability mass function, hence, we can write

$$\phi_i = \frac{1}{n}\sum_{j=0}^{n-1}\mathbb{E}[\rho_i(\mathcal{S})]. \tag{6}$$

This form of the SV suggests an estimation approach based on stratified sampling [19]. For participant $i \in \{1, 2, \ldots, n\}$, let stratum j be the set of marginal contributions of that participant to every subset $\mathcal{S} \subseteq \mathcal{X}_{-i}$ of size $|\mathcal{S}| = j$. We randomly and independently draw N_j^i samples $\rho_{1,j}^i, \ldots, \rho_{N_j^i,j}^i$ from each stratum j. Define the sample mean for participant i as the random variable

$$T_i = \frac{1}{n}\sum_{j=0}^{n-1}\frac{1}{N_j^i}\sum_{k=0}^{N_j^i}\rho_{k,j}^i = \frac{1}{n}\sum_{j=0}^{n-1}\overline{\rho}_j^i \tag{7}$$

where the random variable $\overline{\rho}_j^i$ is the sample mean of the data drawn from stratum j. This sample mean is a linear, unbiased estimate of ϕ_i and would be a reasonable estimate of ϕ_i except for the fact that the sum of the estimates may not be equal to the total budget $v(\mathcal{X})$, which would violate the efficiency axiom of the SV. We, therefore, use the sample averages as basis for computing the maximum likelihood (ML) estimates of the SVs with the budget constraint as follows.

Assume that the number of samples from each stratum is sufficiently large so that we can use the central limit theorem to approximate the distribution of $\overline{\rho}_j^i$ by a Gaussian with mean $\mu_{j,i} = \mathbb{E}[\rho_i(\mathcal{S})]$ and variance $\sigma_{j,i}^2$. By independence of the sample averages, it follows that the variance of T_i:

$$\sigma_i^2 = \frac{1}{n^2}\sum_{j=0}^{n-1}\sigma_{j,i}^2$$

and $T_i \sim \mathcal{N}(\phi_i, \sigma_i^2)$.

By independence of the sample averages $f(T_i|\phi_i)$, $i \in \mathcal{X}$, the likelihood function can be written as

$$f(T_1, \ldots, T_n|\phi_1 \ldots \phi_n) = \prod_{i=1}^{n} f(T_i|\phi_i). \tag{8}$$

Since the sample averages T_i are Gaussian, we consider the log likelihood function

$$\sum_{i=1}^{n}\log(f(T_i|\phi_i)) = \zeta - \sum_{i=1}^{n}\frac{(T_i - \phi_i)^2}{2\sigma_i^2} \tag{9}$$

where ζ is not a function of ϕ_i. To obtain the ML estimates of the SVs, we then need to solve the optimization problem

$$\begin{aligned} \underset{\phi_i}{\text{maximize}} \quad & \zeta - \sum_{i=1}^{n}\frac{(T_i - \phi_i)^2}{2\sigma_i^2} \\ \text{subject to} \quad & \sum_{i=1}^{n}\phi_i = v(\mathcal{X}). \end{aligned} \tag{10}$$

This is a convex optimization problem and has the following analytical solution.

Theorem 1: The ML estimates of ϕ_i are given by

$$\hat{\phi}_i = T_i - \frac{\sigma_i^2}{\sum_{m=0}^{n}\sigma_m^2}\big(\hat{v}(\mathcal{X}) - v(\mathcal{X})\big) \tag{11}$$

where $\hat{v}(\mathcal{X}) = \sum_{i=1}^{n} T_i$.

Note that all properties of the SV (efficiency, symmetry, null player, and linearity [20]) hold in expectation in (11), with the added benefit that the budget is always balanced, as the constraint in the optimization problem (10) is satisfied.

A. Sample Allocation

We now turn our attention to the question of how many samples we should select from each stratum. Suppose, we have a total budget of N samples per participant, i.e., $\sum_{j=0}^{n-1} N_j^i = N$ for every $i \in \mathcal{X}$. How do we divide them among the strata? One reasonable approach would be to allocate the samples for each participant i to minimize the variance of the sample mean T_i subject to $\sum_{j=0}^{n-1} N_j^i = N$. The following shows that the optimal sample allocation is the Neyman allocation [19] for equal weighting.

Lemma 1: The minimum variance of T_i subject to $\sum_{j=0}^{n-1} N_j^i = N$ is

$$\sigma_{i,\mathrm{SD}}^2 = \frac{1}{N}\mathrm{mean}\big(\sigma_{j,i}\big)^2 \tag{12}$$

where $\sigma_{j,i}$ is the standard deviation of the population in stratum j for participant i. The value of $\mathrm{mean}(\sigma_{j,i})$ is calculated by averaging over the n values of $\sigma_{j,i}$ for participant i.

The values of N_j^i that achieve this minimum are

$$N_j^i = \frac{\sigma_{j,i}}{\sum_{m=0}^{n-1}\sigma_{m,i}}, \; j \in \{0, 1, \ldots, n-1\}.$$

The proof of this lemma is straightforward and hence is not included.

It is interesting to compare the achievable variance of the sample means using the above optimal stratified sampling to the more commonly used uniform sampling. With uniform sampling, we draw N samples independently at random from the set of marginal contributions of participant i without taking strata into consideration. The variance of the sample average for this approach can be easily shown to be

$$\sigma_{i,\mathrm{RS}}^2 = \frac{1}{N}\Big[\mathrm{mean}\Big(\sigma_{j,i}^2\Big) + \mathrm{var}\big(\mu_{j,i}\big)\Big] \tag{13}$$

where $\mu_{j,i}$ is the mean value of the population in stratum j for participant i. The value $\mathrm{var}(\mu_{j,i})$ is calculated as the variance of the n values of $\mu_{j,i}$ for participant i.

Sampling according to the Neyman allocation (Lemma 1) requires prior knowledge of the standard deviation of each stratum for each participant, which is not realistic. A more practical approach would be to sample equally from each of the n stratum, i.e., $N_j^i = N/n$. With this allocation, it can be easily shown that the variance of the sample average is

$$\sigma_{i,\mathrm{ES}}^2 = \frac{1}{N}\mathrm{mean}\Big(\sigma_{j,i}^2\Big). \tag{14}$$

Comparing the variances for the above three allocation strategies, we can clearly see that

$$\sigma_{i,\mathrm{SD}}^2 \le \sigma_{i,\mathrm{ES}}^2 \le \sigma_{i,\mathrm{RS}}^2. \tag{15}$$

Hence, it is always better to sample in proportion to standard deviations. In the following section we describe a reinforcement learning algorithm for estimating these standard deviations during sampling.

B. Approximating Optimum Stratified Sampling

Implementing an approximation to SD sampling is a typical reinforcement learning problem in which the algorithm seeks to exploit the information it has about the standard deviations of the strata to sample correctly, but must at the same time explore in order to accurately calculate these very standard deviations. In our setting, the goal is to sample a specific (but unknown) number of times from each stratum. This differs from the usual reinforcement learning problems where the goal is to converge on a single optimum action that maximizes the total reward. This contrast means that some techniques (such as ϵ-greedy, pursuit, and reinforcement comparison) are not suitable, and other approaches must be altered to make them suitable for the problem at hand, see [21] for information on reinforcement learning. By comparison, stochastic methods [21] which assign a probability to each action in accordance with the expected reward (or standard deviation in this case) are quite suitable to our setting.

Our proposed Algorithm 1 explicitly "explores" the problem space initially before gradually moving to an "exploit" phase in which it uses the results of the exploration to improve the sampling allocations. For participant i, the probability of sampling from stratum j at sample $t \leq N$ is

$$\pi_{j,i}(t) = \epsilon(t)\frac{1}{n} + (1-\epsilon(t))\frac{\hat{\sigma}_{j,i}}{\sum_{m=0}^{n-1}\hat{\sigma}_{m,i}} \tag{16}$$

where $\hat{\sigma}_{j,i}$ is the current estimate of the standard deviation of stratum j. The choice of $\epsilon(t)$ is left to the user, but should be a decreasing function of t with $\epsilon(0) = 1$. We implemented a number of such functions (including the stepped function described in [22]) and found the most accurate to be the double sigmoid function

$$\epsilon(t) = \kappa - \frac{1}{1 + e^{-\frac{t-\gamma N}{\beta N}}} \tag{17}$$

where κ is chosen to ensure $\epsilon(0) = 1$. Increasing γ in the above equation reduces the percentage of samples used for exploration, and increasing β increases the transition time from exploration to exploitation.

At each step t, Algorithm 1 chooses stratum j with probability $\pi_{j,i}(t)$ for participant i. The probabilities are then updated for the next iteration. The vector of standard deviations is updated in each step using a numerically stable algorithm from [23]. The algorithm returns the sample mean T_i for participant i as well as the variance of that statistic, σ_i^2. Once this has been calculated for all n participants, the MLE can be computed using (11) to ensure that the budget is balanced.

If $\sigma_{i,\text{SD}}^2 << \sigma_{i,\text{ES}}^2$, then implementing Algorithm 1 will significantly reduce the variance of the sample mean. If, however, $\sigma_{i,\text{SD}}^2 \approx \sigma_{i,\text{ES}}^2$, then the benefit of the algorithm may well be outweighed by the complexity involved in the implementation and time involved in its execution. Comparing $\sigma_{i,\text{SD}}^2$ to $\sigma_{i,\text{ES}}^2$, we have

$$\frac{\sigma_{i,\text{SD}}^2}{\sigma_{i,\text{ES}}^2} = 1 + \frac{\text{var}\left(\sigma_{j,i}^2\right)}{\text{mean}\left(\sigma_{j,i}\right)^2}. \tag{18}$$

Hence, if $\text{var}(\sigma_{j,i}^2)/\text{mean}(\sigma_{j,i})^2 \approx 0$, sampling equally from each strata would be preferable.

TABLE II
OVERHEAD (IN SECONDS PER MILLION SAMPLES) OF EMPLOYING EACH OF THE THREE METHODS DISCUSSED IN THIS PAPER

Sampling Method	Time (secs) per million
Random sampling	00.20
Equal sampling from strata	00.13
Sigmoid $\epsilon(t)$	62.36

C. Computational Overhead

Employing the proposed learning algorithm adds an overhead to the computation time needed to estimate the SV. Table II quantifies this overhead for a dual core 2.1 GHz processor. These times assume that the value function can be calculated instantaneously, i.e., the times are independent of the value function used.

Although the learning algorithm adds significant overhead to the SV estimation time, it does not take into consideration the estimation accuracy (i.e., variance) of the SV value. In Section IV, we show that when accuracy is equalized our learning algorithm requires much less samples and is in fact much faster than the other estimation methods. Also, the implementation of our learning algorithm can be optimized to achieve much faster speed than reported in the table.

IV. DR PROGRAM PROVIDING RESERVE

In this example DR program, the LSE expects participant i to reduce their load by $X_i \in \mathbb{R}_+$ (loads can be reduced for example by dimming lights or controlling heating, ventilation, and air conditioning in a building) during a DR event. The LSE then offers a quantity M of "spinning reserve" to a utility where $\sum_{i=1}^n X_i \geq M, M \in \mathbb{R}_+$. This ensures that the program has a leeway of $\Delta M = \sum_{i=1}^n X_i - M$. When a DR event is requested, each participant responds appropriately. There may be a discrepancy between a participant's actual reduction in consumption, $\tilde{X}_i$, and the expected reduction amount X_i. In this example, we assume that if the aggregate DR meets or exceeds the level M, the LSE receives a payoff according to an affine function of the DR supplied, otherwise the LSE receives no payoff. Hence, the value function for this DR program is

$$v(\mathcal{S}) = \begin{cases} q\sum_{i\in\mathcal{S}}\tilde{X}_i - M, & \sum_{i\in\mathcal{S}}\tilde{X}_i \geq M \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

The form of this value function is motivated by the "weighted voting" cooperative game, a classic problem in the literature of cooperative game theory. A proof of the convexity of this game is given in the Appendix. As an electricity marketplace has a minimum bid quantity, such a value function could well

Algorithm 1 Approximating SD Sampling

procedure STANDARDDEVIATIONSAMPLING(N, i)
  $t \leftarrow 1$
  $\hat{\mu}_{j,i} \leftarrow \underline{0}$ ▷ Estimate of $\mu \in \mathbb{R}^n$
  $\hat{\sigma}_{j,i} \leftarrow \underline{0}$ ▷ Estimate of $\sigma \in \mathbb{R}^n$
  $c \leftarrow \underline{0}$ ▷ Vector where c_j is the number of samples taken from stratum j, $c \in \mathbb{R}^n$
  $m2 \leftarrow \underline{0}$ ▷ Vector of the sum of squared differences from the current mean of stratum j, $m2 \in \mathbb{R}^n$
  while $t \leq N$ **do**
    $\pi_{j,i}(t) \leftarrow \epsilon(t)\frac{1}{n} + (1-\epsilon(t))\frac{\hat{\sigma}_j}{\sum_{m=0}^{n-1}\sigma_{m,i}}$
    Choose stratum j at random, weighted by $\pi_{j,i}(t)$.
    Choose a random coalition, $\mathcal{S} \subseteq \mathcal{X}_{-i}$ where $|\mathcal{S}| = j$.
    $x \leftarrow \rho_i(\mathcal{S})$ ▷ x is a sample from stratum j
    $c_j \leftarrow c_j + 1$ ▷ Update the count for stratum j
    $\Delta \leftarrow x - \hat{\mu}_{j,i}$
    $\hat{\mu}_{j,i} \leftarrow \hat{\mu}_{j,i} + \frac{\Delta}{c_j}$ ▷ Online update for estimate of $\mu_{j,i}$
    $m2_j \leftarrow m2_j + \Delta(x - \hat{\mu}_{j,i})$
    $\hat{\sigma}_{j,i} \leftarrow \sqrt{\frac{m2_j}{c_j - 1}}$ ▷ Online update for estimate of $\sigma_{j,i}$
  end while
  $T_i \leftarrow \text{mean}(\hat{\mu}_{j,i})$
  $\sigma_i^2 = \frac{1}{n^2}\sum_{j=0}^{n-1}\hat{\sigma}_{j,i}^2$
  return T_i, σ_i^2
end procedure

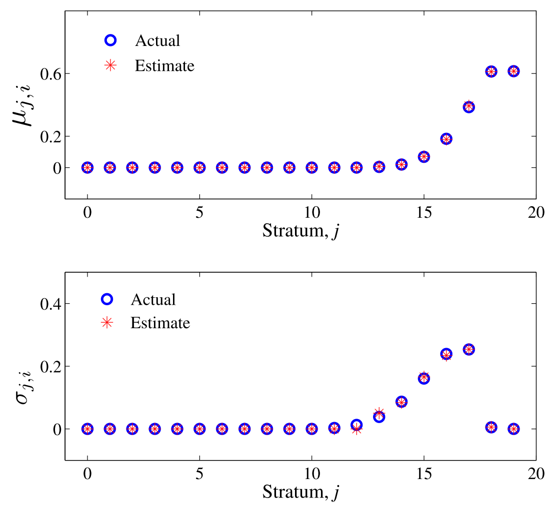


Fig. 3. Mean (top) and standard deviation (bottom) for each stratum. A red "x" indicates the final estimate from employing Algorithm 1 with $N = 5000$ samples.

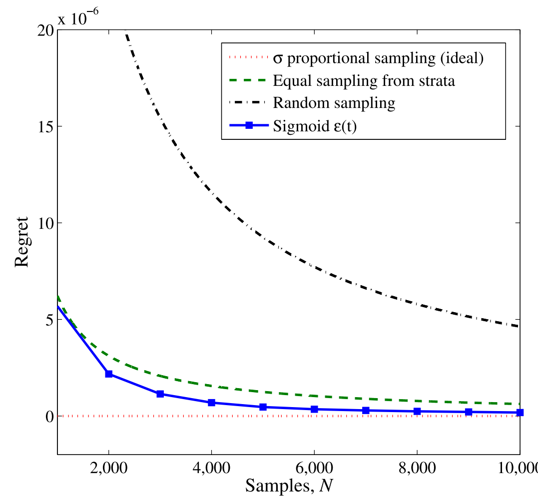


Fig. 4. Decrease in regret as sample size increases for various $\epsilon(t)$ functions. The upper dashed line is the regret from uniform sampling and the lower dashed line at zero regret corresponds to ideal sampling.

model the LSEs DR program and is therefore justifiable in its use. Using the SV as a means of determining compensation is of great relevance to such LSEs.

To compare the performance of the sampling techniques we discussed in Section III, we consider a small set of $n = 20$ participants so that we can compute the exact SV (ground truth).

Fig. 3 plots the sample mean and standard deviation for each stratum when using the value function (19) for a representative participant i. As can be seen, our stratified sampling algorithm which approximates sampling in proportion to the standard deviations will show a significant improvement over both uniform and random sampling because strata 0–10 have zero mean (and standard deviation) and as such do not contribute to the SV and the samples taken from these strata in the uniform and random sampling methods are wasted.

Remark 2: As mentioned in Section III-B, we implemented the reinforcement learning algorithm using various $\epsilon(t)$ functions. Fig. 4 plots the sample size against "regret," defined as the difference between the variance of the SV estimate for a given $\epsilon(t)$ and that of the estimate calculated using exact SD sampling. As can be seen, the sigmoid function (with $\gamma = 0.2$ and $\beta = 0.075$) which we use in all numerical results closely approximates ideal sampling.

In Fig. 5, we compare the distribution of the differences between the estimated and actual SVs for all participants, i.e., the error distributions. The estimates of the SVs were calculated using $N = 5000$ samples. It can be seen that employing stratified sampling reduces the error significantly when compared to both random sampling and equal sampling. In fact, the distribution of errors for stratified sampling is approaching that of ideal sampling.

We determine the accuracy of the sampling methods using the mean squared prediction error (MSPE)

$$\text{MSPE} = \mathbb{E}\left[\left(\phi_i - \hat{\phi}_i\right)^2\right]. \tag{20}$$

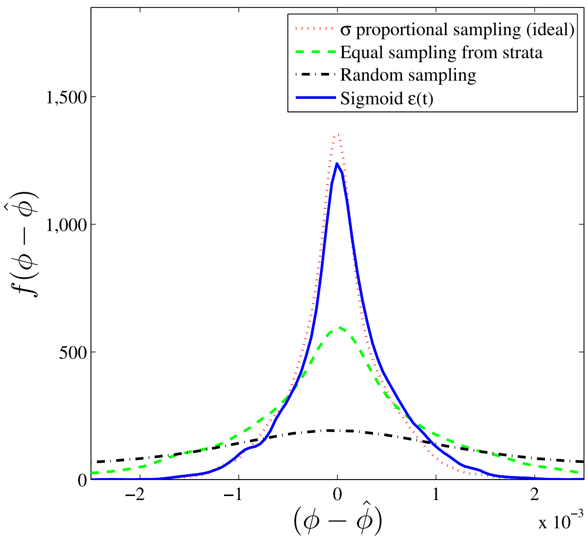


Fig. 5. Estimated densities of the error in the estimated SV, i.e., the difference between the actual SVs for all participants and the values estimated using the various sampling techniques. The estimates are calculated using $N = 5000$.

TABLE III
NORMALIZED MSPE FOR THE VARIOUS SAMPLING METHODS

Method	Normalized MSPE
Random Sampling	23.2239
Equal Sampling from strata	2.4900
σ proportional sampling (ideal)	1
Sigmoid $\epsilon(t)$	1.2368

TABLE IV
TOTAL TIME TAKEN TO ESTIMATE THE SV TO THE SAME LEVEL OF ACCURACY FOR EACH OF THE THREE METHODS

Sampling Method	Number of samples	Total time taken (secs)
Random Sampling	1000000	58.2
Equal Sampling from strata	157610	9.16
Sigmoid $\epsilon(t)$	51129	6.15

For ease of comparison, we normalize the MSPE for each method by the MSPE for σ proportional sampling (the ideal method). Table III contains the comparison results. It is clear that employing stratified sampling gives much better results than simple random sampling. The learning algorithm significantly outperforms uniform stratified sampling and approaches the accuracy of ideal stratified sampling.

A. Time Taken to Calculate Estimates

Table II lists the number of samples required and the time taken to ensure the accuracy (i.e., variance) of the SV estimates are equivalent. The random sampling method calculated using one million samples is used as a benchmark. As can be seen from Table IV, the learning algorithm can estimate the SV with significantly less samples and in considerably less time.

V. CONCLUSION

This paper proposes the use of the SV to distribute revenue among the participants in a DR program. As the SV is computationally intractable in general, we proposed a stratified sampling technique that reduces the number of samples needed to achieve a desired estimation accuracy while satisfying the budget balance constraint. We found that optimal stratified sampling requires prior knowledge of the standard deviations of the strata, which may not be available. As such, we proposed a reinforcement learning heuristic which estimates the standard deviations and uses them to adjust the sample allocation among the strata. We demonstrated the use of the SV in DR programs numerically, describing a scenario (DR providing reserve) where the reinforcement learning algorithm can significantly reduce the variance of the estimate.

It should be also noted that our reinforcement learning method is agnostic to the specifics of the characteristic function, and can therefore be used to estimate the SV for any cooperative games, not only for DR programs. Also, this method ensures that the "budget balancing" constraint is met. To our knowledge, this constraint has not previously been considered in other research on estimating the SV using random sampling techniques. Its importance is clear in DR programs in which a given payoff needs to be distributed in its entirety among participants.

The work in this paper can be extended in several important directions, particularly regarding the choice of the value function. The DR program presented in this paper may be rather simplistic for practical use, as it does not capture some aspects of real-world DR programs, such as multiple thresholds. Furthermore, the difference between the load changes promised by—or expected of—the participants and the load changes delivered by them is not directly modeled in the value function in (19). Including actual user behavior in the analysis is a topic that will make this paper more relevant and applicable. However, it should be noted that these points reflect limitations in the choice of the value function, not in the suitability of the SV to be used as a distribution scheme. As such, there remains a significant opportunity to extend this research by using more realistic value functions that either captures existing DR programs or model new possibilities for DR.

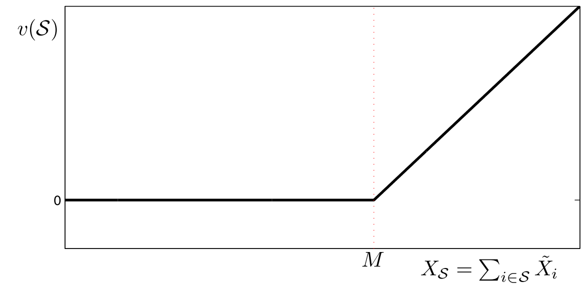


Fig. 6. Function $v(\mathcal{S})$, as described in (21).

APPENDIX

A. Convexity of the Game for the Example Value Function

The value function employed in this paper is reprinted here for convenience, (21) and plotted in Fig. 6

$$v(\mathcal{S}) = \begin{cases} q \sum_{i\in\mathcal{S}} \tilde{X}_i - M, & \sum_{i\in\mathcal{S}} \tilde{X}_i \geq M \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

We prove supermodularity by first noting that the value function can be rewritten as

$$v(\mathcal{S}) = -[\min(X_{\mathcal{S}}, M) - X_{\mathcal{S}}]$$

where $X_{\mathcal{S}} = \sum_{i \in \mathcal{S}} \tilde{X}_i$. If $v(\mathcal{S})$ is supermodular, then $-v(\mathcal{S})$ must be submodular, so it is sufficient to show that $\min(X_{\mathcal{S}}, M) - X_{\mathcal{S}}$ is submodular. This is equivalent to showing that $\min(X_{\mathcal{S}}, M)$ is submodular as the addition (or subtraction) of a linear function from a submodular function does not affect submodularity. Finally, the simple function $\min(X_{\mathcal{S}}, M)$ is nothing more than a budget-additive function, well-known to be submodular for $M \geq 0$ and each $\tilde{X}_i \geq 0$ [24]. Therefore, the value function $v(\mathcal{S})$ is a supermodular function, and the game is convex.

References

[1] S. Borenstein, M. Jaske, and A. Rosenfeld, "Dynamic pricing, advanced metering, and demand response in electricity markets," Center Study Energy Markets, Univ. California Energy Inst., Berkeley, CA, USA, Tech. Rep. CSEM WP 105, Oct. 2002.

[2] N. Li, L. Chen, and S. Low, "Optimal demand response based on utility maximization in power networks," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA, USA, 2011, pp. 1–8.

[3] N. Gatsis and G. Giannakis, "Cooperative multi-residence demand response scheduling," in *Proc. 45th Annu. Conf. Inf. Sci. Syst. (CISS)*, Baltimore, MD, USA, 2011, pp. 1–6.

[4] A. Conejo, J. Morales, and L. Baringo, "Real-time demand response model," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 236–242, Dec. 2010.

[5] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Gaithersburg, MD, USA, 2010, pp. 409–414.

[6] Q. Dong, L. Yu, W.-Z. Song, L. Tong, and S. Tang, "Distributed demand and response algorithm for optimizing social-welfare in smart grid," in *Proc. IEEE 26th Int. Parallel Distrib. Process. Symp. (IPDPS)*, Shanghai, China, 2012, pp. 1228–1239.

[7] T. Kim and H. Poor, "Scheduling power consumption with price uncertainty," *IEEE Trans. Smart Grid*, vol. 2, no. 3, pp. 519–527, Sep. 2011.

[8] G. O'Brien, A. E. Gamal, and R. Rajagopal, "Efficient computation of Shapley values for demand response programs," in *Proc. 4th Int. Conf. Future Energy Syst. (e-Energy)*, Berkeley, CA, USA, 2013, pp. 283–284.

[9] L. Shapley, "Cores of convex games," *Int. J. Game Theory*, vol. 1, no. 1, pp. 11–26, 1971.

[10] Q. Zhu, J. Zhang, P. Sauer, A. Dominguez-Garcia, and T. Basar, "A game-theoretic framework for control of distributed renewable-based energy resources in smart grids," in *Proc. Amer. Control Conf. (ACC)*, Montreal, QC, Canada, 2012, pp. 3623–3628.

[11] W. Saad, Z. Han, H. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 86–105, Sep. 2012.

[12] T. Haring, J. Mathieu, and G. Andersson, "Decentralized contract design for demand response," in *Proc. 10th Int. Conf. Eur. Energy Market (EEM)*, Stockholm, Sweden, May 2013, pp. 1–8.

[13] M. Junqueira *et al.*, "An Aumann–Shapley approach to allocate transmission service cost among network users in electricity markets," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1532–1546, Nov. 2007.

[14] E. Baeyens, E. Bitar, P. Khargonekar, and K. Poolla, "Wind energy aggregation: A coalitional game approach," in *Proc. 50th IEEE Conf. Decis. Control Eur. Control Conf. (CDC-ECC)*, Orlando, FL, USA, 2011, pp. 3000–3007.

[15] L. S. Shapley, "A value for n-person games," in *Contributions to the Theory of Games*, vol. 2. Princeton, NJ, USA: Princeton Univ. Press, 1953, pp. 307–317.

[16] I. Mann and L. S. Shapley, "Values for large games, IV: Evaluating the electoral college by Monte Carlo techniques," U.S. Air Force, Rand Corp., Santa Monica, CA, USA, Tech. Rep. RM 2651, 1960.

[17] Y. Bachrach *et al.*, "Approximating power indices: Theoretical and empirical analysis," *Auton. Agents Multi-Agent Syst.*, vol. 20, no. 2, pp. 105–122, Mar. 2010.

[18] J. Castro, D. Gómez, and J. Tejada, "Polynomial calculation of the Shapley value based on sampling," *Comput. Oper. Res.*, vol. 36, no. 5, pp. 1726–1730, 2009.

[19] P. S. Rao, *Sampling Methodologies*. London, U.K.: Chapman and Hall, 2000.

[20] G. Owen, *Game Theory*, 2nd ed. New York, NY, USA: Academic Press, 1982.

[21] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, vol. 1. Cambridge, MA, USA: MIT Press, 1998.

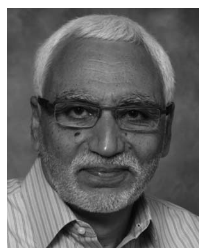
[22] P. Etoré and B. Jourdain, "Adaptive optimal allocation in stratified sampling methods," *Methodol. Comput. Appl. Probab.*, vol. 12, no. 3, pp. 335–360, 2010.

[23] D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*, vol. 2, 3rd ed. Boston, MA, USA: Addison-Wesley Longman, 1997.

[24] M. Kapralov, I. Post, and J. Vondrák, "Online submodular welfare maximization: Greedy is optimal," in *Proc. 24th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*, New Orleans, LA, USA, Jan. 2013, pp. 1216–1225. [Online]. Available: http://dx.doi.org/10.1137/1.9781611973105.88

Gearóid O'Brien received the B.E.E.E. and M.Eng.Sci. degrees in electrical engineering from University College Cork, Cork, Ireland, both in electrical engineering; the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, during which he was supported by a Stanford Graduate Fellowship and the China Electric Power Research Institute, in 2005, 2007, and 2014, respectively.

He is the Founding Data Scientist with Datasnap.io, San Francisco, CA, USA. He is currently with the Department of Electrical Engineering, Stanford University.



Abbas El Gamal received the B.Sc. (Hons.) degree in electrical engineering from Cairo University, Giza, Egypt, in 1972, and the M.S. degree in statistics and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1977 and 1978, respectively.

From 1978 to 1980, he was an Assistant Professor of Electrical Engineering with University of Southern California, Los Angeles, CA. From 2004 to 2009, he was the Director of the Information Systems Laboratory, Stanford University. He is the Hitachi America Professor with the School of Engineering and the Chair of the Department of Electrical Engineering, Stanford University. His current research interests include network information theory, field programmable gate arrays, and digital imaging devices and systems. He has authored/co-authored over 200 papers and holds 35 patents in the above areas. He has co-authored *Network Information Theory* (Cambridge Press, 2011).

Prof. El Gamal was a recipient of several honors and awards for his research contributions, including the 2012 Claude E. Shannon Award and the 2004 INFOCOM Paper Award. He serves on the Board of Directors of the Information Theory Society, where he is currently the President. He is a member of the U.S. National Academy of Engineering.

Ram Rajagopal received the Bachelor's degree in electrical engineering from the Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, in 1997; the Master's degree in electrical and computer engineering from the University of Texas, Austin, TX, USA, in 2002; and the M.A. degree in statistics, and the Ph.D. degree in electrical engineering and computer sciences from the University of California–Berkeley, Berkeley, CA, USA, in 2008 and 2009, respectively.

He was a Digital Signal Processing Research Engineer with National Instruments, Austin, and a Visiting Research Scientist with IBM Research, Austin. He is currently an Assistant Professor of Civil and Environmental Engineering with Stanford University, Stanford, CA, where he is the Director of the Stanford Sustainable Systems Laboratory (S3L), focused on large-scale monitoring, data analytics, and stochastic control for infrastructure networks, in particular energy and transportation. His current research interests include integration of renewables, smart distribution systems, and demand-side data analytics. He holds over 30 patents from his work, and has advised or founded various companies in the fields of sensor networks, power systems, and data analytics.

Prof. Rajagopal was a recipient of the Powell Foundation Fellowship, the Berkeley Regents Fellowship, and the Makhoul Conjecture Challenge Award.