Preventing Cascading Failures in Microgrids with One-sided Support Vector Machines

Matt Wytock, Srinivasa Salapaka, and Murti Salapaka

Abstract— **Microgrids formed by a network of power sources and power consumers yield significant advantages over the conventional power grid including proximity of power consumption to power generation, distributed generation, resiliency against wide area blackouts and ease of incorporation of renewable energy sources. On the other hand, unlike the conventional grid, microgrids are compliant where a single load or a single generation unit can often form a significant fraction of the total generation capacity. Here large excursions from the nominal operating conditions are possible motivating the need for safety mechanisms which isolate power electronic equipment from damage. Breakers serve the purpose of protecting equipment from surge conditions by shutting off, for example, generation units. However in microgrids, a loss of a single generation unit can have catastrophic impact on the viability of the entire system. Here settings on breakers cannot be chosen too conservatively to protect the equipment at the expense of system viability or too liberally which will result in equipment damage. The ensuing problem of striking a suitable compromise tends to be combinatoric in nature due to numerous states of breakers which is further exacerbated by an uncertain load profile and nonlinear nature of system dynamics. In this article we provide a methodology to determine current thresholds and *guard times*, the time interval when current is allowed to exceed threshold value, for each inverter for fail-safe operation of microgrid. We employ a machine learning approach to address the problem where we first demonstrate that conventional support vector machine (SVM) methodology does not yield a satisfactory solution. We then develop a one-sided SVM method and generalize it to yield nonlinear support boundaries which captures the need for fail-safe operation against system blackouts while protecting equipment. A simulation engine is developed to model a real microgrid which is used to generate data for assessing and guiding our approach.**

I. Introduction

In the recent past a confluence of factors has occurred which will revolutionize the way in which energy is sourced, distributed and consumed. An enabling impetus that is changing the energy infrastructure is the rapid advance of the power electronics area which facilitates sophisticated means of controlling and managing flow of power. Another challenge to the conventional power grid is posed by the emphasis being laid on renewable sources of energy which lessen environmental concerns and can be generated in a distributed manner, often at the site of power consumption. Distributed generation provides the advantages of resiliency, modularity and proximity between the sources of power and its consumption. A key framework for future means of managing energy is provided by microgrids; small networks (relative to the conventional power grid) formed by power generation units and power consuming devices.

Microgrids differ from the conventional grid in significant ways. The conventional grid has a centralized architecture and has significant inertia which ensures that voltage and frequency remain tightly regulated and thus the deviation of operating parameters from the nominal values remains small. In contrast microgrids when not connected to the conventional grid are significantly more compliant where large excursion from the nominal and desired behavior is possible. In microgrids it is often the case that power generation capacity is close to power consumption needs, where a single load or generation unit can form a considerable fraction of the total capacity. Here, loads or generation units being switched on or off can lead to large transient response. Thus in microgrids it becomes crucial to introduce mechanisms, such as breakers, to safeguard equipment and switch them off when necessary to avoid damage due to overload conditions. However, breakers' settings can have significant influence on the performance of the system as a whole. Here, an examination of similar issues in the conventional grid can provide useful insights.

The conventional power grid is a complex system of many individual subsystems. It employs protective relays to isolate subsystems from the entire grid and to shut them off when needed. However, relays utilized to protect components can often be the cause of blackouts. The opening of a relay, which might be justifiable at the local scale can lead to a rapid cascade of failures that result in a large geographic area losing power. Considerable focus has been placed on mitigating the adverse effects of relays where if the settings are too conservatively placed to protect the equipment, the system viability is compromised. The conservative usage of conventional relays can be relaxed using smart relays which incorporate knowledge of the system learned from data. For example, in [1] adaptive relays are presented which are capable of differentiating between normal and fault conditions. These adaptive relays minimize the probability of large-scale blackouts.

Issues relating to protective relays in the conventional grid have parallels in the microgrid. A key component of

M. Wytock, is with Google Inc., Mountain View, CA 94043, email: mwytock@google.com

S. Salapaka is with Mechanical Science and Engineering at the University of Illinois, Urbana, IL 61801, email: salapaka@illinois.edu

M. Salapaka is with Electrical and Computer Engineering at the University of Minnesota,Twin Cities, MN 55455, email: salapakam@google.com

the microgrid framework are inverters which are power electronic devices that condition power. It is typical to include mechanisms (for example, breakers) that switch the inverter off if currents exceed imposed thresholds. Moreover breakers are configured to tolerate excursion from the imposed threshold for a limited time which is termed *guard time* in this article. Thus a breaker trips only if the threshold is violated throughout a guard-time interval. Evidently low threshold values and small guard-times are beneficial for the safety of the inverter. However, aggressive choice of these parameters can lead to unacceptable system performance with the probability of blackouts being higher.

A blackout condition caused by a cascade of breakers tripping is a complex phenomena which can be caused in a multitude of ways. The sequence of trip events depends on the threshold values and guard times placed on each inverter, the load profile over the time horizon of concern and the power generating capacities of the associated inverter units. Moreover tripping is often caused by transient conditions in the highly interconnected system with loads which are often time-varying and nonlinear. Here the transient behavior is intractably hard to model quantitatively. Thus the combinatoric nature of possible causes, the stochasticity of the load profile and power generation capacity (for example, due to changing availability of power from the sun in a PV source) and the nonlinear and time-varying nature of the system makes it hard to arrive at a rational choice of breaker threshold values and corresponding guard-times via conventional analytical means.

Methods based on machine learning can play an important role in providing viable settings for breakers. Such approaches are available to address issues that arise in the conventional grid. In [2], machine learning based methods are developed to assist in prioritization of repair and maintenance tasks. In [3], in the conventional grid setting machine learning methods are developed to deal with temporal aspects that arise in the context of power reliability.

In this article the problem of ascertaining the threshold values and the associated guard-times of inverters that are sourcing power to a shared load is studied in a microgrid setting. It is shown that using a machine learning framework, the function which maps the threshold values and guard-times to the status of inverters being on and off can be learned. Such a function can then be used to provide a non-conservative choice of breaker parameters that decrease the probability of blackouts. Since the breaker parameters are critical to functioning of the microgrid, the learning should be such that the number of parameters wrongly identified as safe (when in actuality they may lead to microgrid failure) is very small. Equivalently, the false positive rate (FPR), that is, the number of outcomes predicted positive when observed data shows they are negative should be small. We first present the standard support vector machine (SVM) and its relaxation that admits feasible solutions for larger class of scenarios. The relaxed problem remains inadequate in reflecting the concerns of cascading failure where very small FPR values need to be emphasized. The article formulates a *one-sided SVM* which severely penalizes false positives. The one-sided SVM results in higher true positive rates (TPR, the ratio of the number of correctly predicted positive outcomes to the number of observed positive outcomes) than standard SVM for small FPR values. The dual formulation of the one-sided SVM with kernels is also considered, admitting nonlinear support boundaries which result in better accuracy (percentage of correctly predicted outcomes) and performance in terms of TPR. The article also provides a simulation framework that is made specific for the design of microgrids formed for residential purposes.

The article is organized as follows: Section II describes the microgrid for which the breaker parameters are being sought is detailed. A description of machine learning models and the formulation of the one-sided SVM is given in Section III. The results applied to data obtained from microgrid model described in Section II are presented in IV. Conclusions are presented in Section V.

II. System and problem description

In this section, a model of a real microgrid is presented (see Figure 2). The data used by the machine learning algorithms proposed in this paper are obtained using a simulation engine based on this model. This simulation engine employs high-fidelity models of inverters and loads, and droop-based control architectures [4]. The simulation models and control architecture described in this section are not the goals of this paper; they provide the context under which the data for machine learning is obtained and emphasize the details of the microgrid that the model captures. It is to be noted that even though the study is focused on a specific microgrid the framework being developed is applicable for broader class of microgrids.

A microgrid comprises multiple power sources, multiple loads that consume power, power electronic devices (such as inverters), and controllers (see Figure 2). The power generated from various sources is conditioned (and shared) by inverters that provide and distribute power to the loads consistent with requirements of the load. For instance an inverter whose input is provided by a DC power source can output an AC current through appropriate control at a fast time scale of related electronics. After averaging the fast time scale dynamics, an inverter can be modeled as a controllable voltage source, v_{inv}, with an inductor and a capacitor as shown in Figure 1 [5]. A control system manages power sharing and voltage regulation at the outputs of the inverters as different loads come on or go off the grid. More specifically, the control system can feedback measured signals such as inductor current, inverter output voltage and current for determining the switching control (or equivalently the v_{inv} in the averaged model).

In the microgrid considered in this paper, two power sources provide energy to a shared load where the power flow is conditioned by inverters. For each inverter the operation is realized using standard PWM based operation

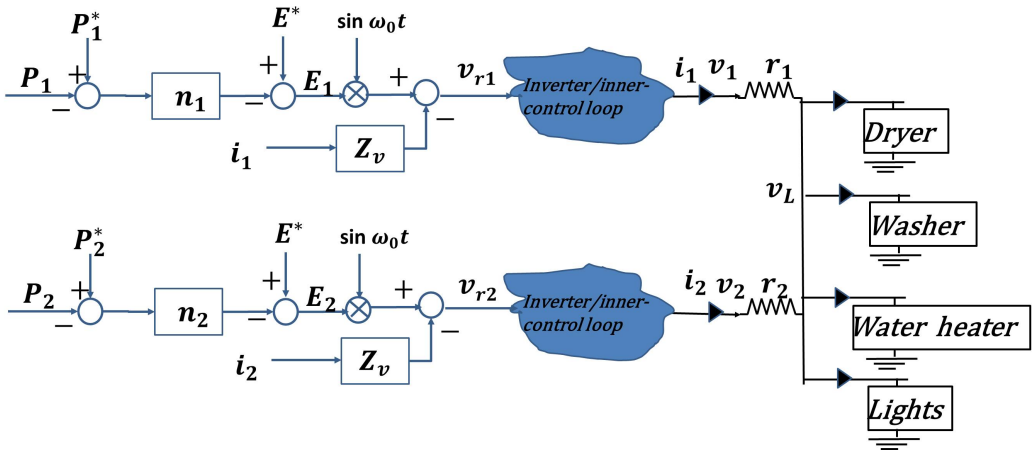

Fig. 2. (a) Schematic describing the system shows the outer-control loop. The loads being serviced are a dryer, washer, water heater and lights.

based on high bandwidth periodic switching which admits the average model described by Figure 1 [5]. It is assumed

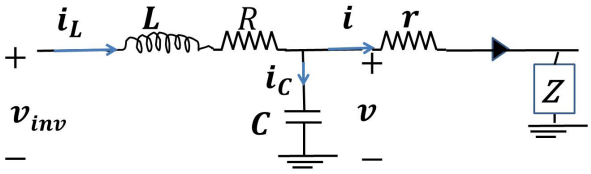

Fig. 1. Averaged model of a single inverter with a linear load Z.

for each inverter that the inverter output current i, inductor current i_L and inverter output voltage v are measured variables. The control system for each inverter consists of an outer-control loop that implements the voltage-active power droop law to generate a reference $v_{lref,k}$ where

$$v_{lref,k} = [E^* - n_k(P_k - P_k^*)] \sin \omega_0 t \tag{1}$$

where E^* is set to $120\sqrt{2}$, $\omega_0 = 2\pi f_0$ with $f_0 = 60Hz$, P_k^* is the setpoint active power to be sourced by the inverter, P_k is the actual power being delivered by the inverter and n_k is the droop coefficient which dictates the change in the voltage magnitude desired for a given error $P_k - P_k^*$ (see Figure 2). The inner control loop generates the reference voltage $v_{r,k}$ to be tracked by the k^{th} inverter given by

$$v_{r,k} = v_{lref,k} - Z_v i_k \tag{2}$$

where Z_v is the virtual resistance [4]. The reference voltage $v_{r,k}$ is tracked using a inner voltage and current controller (see Figure 3). Note that the reference voltage $v_{r,k}$ is tracked using a inner-control current loop. Here, the output voltage is compared with the reference $v_{r,k}$ which is used to generate a reference current $i_{ref,k}$ to be tracked by the current controller K_{cur}. The tracking of the voltage reference via a current controller allows for placing safety measures to limit unsafe magnitudes of i_{ref}. Furthermore, to safeguard the inverters from damage, if the rms of the k^{th} inverter inductor current, $i_{L,k}$ crosses a specified threshold

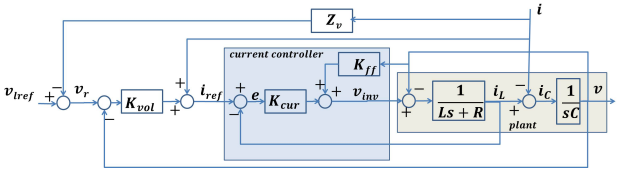

Fig. 3. The inner-control loop for voltage regulation where a virtual resistance is incorporated.

$i_{th,k}$ and remains above the threshold for a specified guard-time $t_{g,k}$ then the k^{th} inverter is shut down.

The loads consist of a washer, dryer, water heater (rated at 1.7 KW) and lights (rated at 1.3 KW). Washer and dryer pose load profiles which vary with time. The models used for washer and dryer are representative of the behavior of generic washer and dryers and are realistic. The total commanded generation is 4 KW with one inverter sourcing 1 KW while the other sources 3 KW. In the setup above the choice of $(i_{th,1}, t_{g,1})$ and $(i_{th,2}, t_{g,2})$ has a significant influence on the overall viability of the system for a demanding load profile. Aggressive choices of $(i_{th,1}, t_{g,1})$ and $(i_{th,2}, t_{g,2})$ motivated by objectives of protecting devices can lead to unacceptable probability of blackout, whereas, on the other hand large values can damage the inverters.

We consider the following functions to provide a guidance on the choice of the breaker settings. The function $f_0 : \mathbb{R}^4 \to \{-1, 1\}$ which takes the input $(i_{th,1}, t_{g,1}, i_{th,2}, t_{g,2})$ and provides an output s where $s = 1$ implies that at least one inverter remains operational and $s = -1$ implies that both inverters have shut down. Similarly $f_1 : \mathbb{R}^4 \to \{-1, 1\}$ represents the survivability of inverter 1 where the input remains the same as for f_0 and the output is 1 if inverter 1 remains operational else the output is -1; f_2 is defined similar to f_1 to describe the survivability of inverter 2.

Remark 1: Note that occurrence of a blackout (where all inverters shut down) in a microgrid depends on a complex interaction of many factors of which $i_{th,1}, t_{g,1}, i_{th,2}, t_{g,2}$ forms only a small subset. These factors include dynamics of

loads, the on-and-off time schedules of loads, the controller architecture, and power commanded from each inverter. The framework for learning functions f_k ($k = 0, 1$ and 2) has to consolidate the variability of factors not provided as inputs to f_k.

III. Machine learning model

In the following a machine learning approach is taken to learn the functions f_0, f_1 and f_2. This approach is appealing as it makes few assumptions on the underlying system—in particular, we do not attempt to model the complex time-varying and nonlinear switched dynamics that describe the microgrid comprising the inverters, the loads, and the control system. We also make limited assumptions on the load schedule which in practice is controlled by the user of the microgrid who expects faultless operation. Our approach learns the function f_k ($k = 0, 1, 2$) from test data and in the large sample limit will converge to the true underlying function, subject to weak smoothness constraints on f_k [6].

We first formulate the problem in the classification framework; the *feature set* x and *predicted outcome* y respectively refer to the input and the output of the function f_k to be learned. A labeled example is a tuple consisting of a feature set x_i and the corresponding *observed* outcome y_i. In the standard setting, given a set of n labeled examples (x_i, y_i) for $i = 1, \ldots, n$ with $x_i \in \mathbb{R}^p$ and $y_i \in \{-1, 1\}$, the classification problem seeks $f(x_i)$ which minimizes a composite objective including the empirical loss between $f(x_i)$ and y_i and structural penalties on f. For each $k \in \{0, 1, 2\}$, we aim at solving a classification problem to obtain f_k.

We first consider the standard support vector machine (SVM) [7] for solving the classification problem; SVMs are appealing because they have a natural geometric interpretation which we adapt to our problem and because they yield efficient algorithms for learning nonlinear functions through dual formulation and kernels. The result of a basic SVM is a hyperplane that separates *positive* examples (where features x_i lead to outcome $y_i = 1$) and *negative* examples (where features x_i lead to outcome $y_i = -1$). The basic SVM seeks the optimal hyperplane separating positive and negative examples by solving the convex optimization problem

$$\begin{aligned} &\underset{w,b}{\text{minimize}} \;\; \frac{1}{2}\|w\|^2 \\ &\text{subject to} \;\; y_i(w^T x_i - b) \geq 1, \quad 1 \leq i \leq n, \end{aligned} \tag{3}$$

where the learned parameters $w \in \mathbb{R}^p$ and $b \in \mathbb{R}$ determine a hyperplane in p dimensions. In the above setting the constraint $y_i(w^T x_i - b) \geq 1$ characterizes two conditions

$$\begin{aligned} &(w^T x_i - b) \geq 1 \text{ for all } (x_i, y_i) \text{ with } y_i = 1 \\ &(w^T x_i - b) \leq -1 \text{ for all } (x_i, y_i) \text{ with } y_i = -1; \end{aligned}$$

thus the condition $y_i(w^T x_i - b) \geq 1$ forces the features corresponding to positive and negative examples respectively to lie on one side of the hyperplane H_a given by $w^T x - b = 1$ and the opposite side of the hyperplane H_b given by $w^T x - b = -1$. The distance between these parallel hyperplanes is given by $2/\|w\|$; this distance is maximized (or equivalently $\|w\|^2$ is minimized) to find the maximum margin classifier which is the unique solution to the SVM optimization problem. Assuming that the examples are *linearly separable*—a set of labeled examples is linearly separable if there exists (w, b) such that $y_i(w^T x_i - b) \geq 1$ for *all* examples in the set—the resulting function is given by

$$f(x) = sign((w^\star)^T x - b^\star) \tag{4}$$

where $(w^\star, b^\star)$ is the solution to (3).

In general, labeled examples are not linearly separable which renders problem (3) infeasible; the standard approach to overcome this issue is to introduce slack variables ξ_i which allow for the possibility of misclassification of i^{th} example. The resulting optimization problem is given by

$$\begin{aligned} &\underset{w,b,\xi_i}{\text{minimize}} \;\; \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \\ &\text{subject to} \;\; y_i(w^T x_i - b) \geq 1 - \xi_i, \;\; 1 \leq i \leq n \\ &\qquad\qquad \xi_i \geq 0, \end{aligned} \tag{5}$$

where $\xi_i > 0$ indicates that example i was misclassified which is penalized in the objective function with weight C. For our case, problem formulation (5) is still not adequate, which we illustrate as follows. Note that in view of Remark 1, it is possible to have labeled examples (x_i, y_i) and (x_j, y_j) such that $x_i = x_j$ and $y_i \neq y_j$. That is, for the same feature set it is possible to have a blackout outcome for one load schedule and no-blackout outcome for another load schedule. The ratio ρ_s of blackout to no-blackout outcomes for a given feature set x_s is fixed by the true distribution of the load schedule. Now suppose a function f_0 is learnt such that $f_0(x_s) = -1$ (that is predicted outcome of x_s is a blackout), then the proportion of correctly predicted outcomes by f_0 to all the outcomes for examples that have x_s as feature set is ρ_s; no extra data will improve the performance better than ρ_s. Furthermore (5) places equal emphasis on positive and negative examples. These issues can be addressed either by improving the classifier (e.g. by extending the feature set to include more comprehensive information, such as load schedules) or by modifying the classification approach to produce more desirable results for our specific application of preventing blackouts. In the sequel we consider the latter as (in general) producing perfect classifications for complex systems is intractable.

Specifically, since choice of current threshold and guard time parameters are critical to ensure failsafe operation of the microgrid, it is necessary that classification scheme selects parameters that have no failures (no negatives) over all observed data. Ideally the classification scheme should not characterize any parameter set as safe if it leads to a negative outcome even for one specific schedule of loads. In this aspect, the standard classification criterion described in (5) is lacking since it does not emphasize exclusion of negative examples. Accordingly, we adapt the SVM by

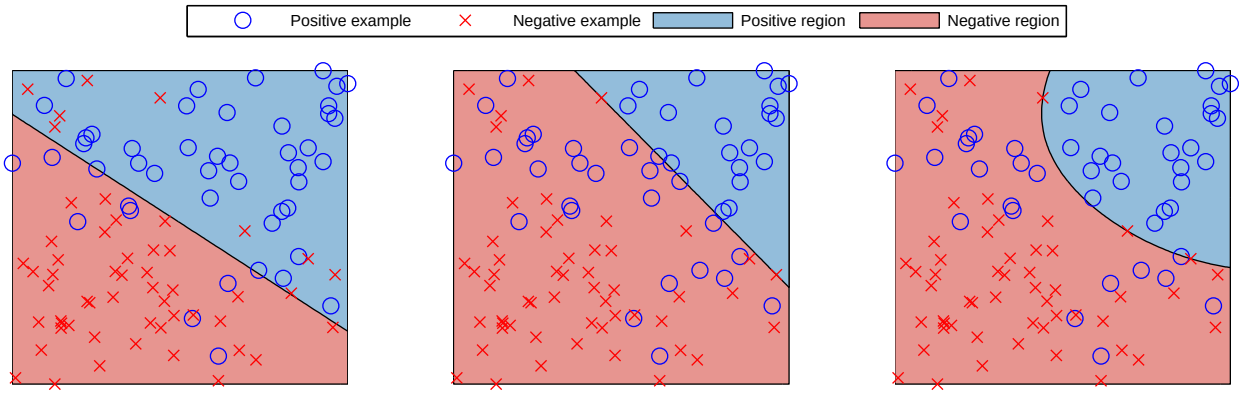

Fig. 4. Example of method on synthetic data with linear SVM (left), one-sided SVM (center) and one-sided SVM with RBF kernel (right).

adding a constraint that ensures such exclusion by posing the following optimization problem,

$$\begin{aligned} \underset{w,b,\xi_i}{\text{minimize}} \;& \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ \text{subject to} \;& y_i(w^T x_i - b) \geq 1 - \xi_i,\; 1 \leq i \leq n \\ & \xi_i \geq 0 \\ & \xi_i \leq 1 \text{ for } i \in \mathcal{N}, \end{aligned} \tag{6}$$

where $\mathcal{N}$ denotes the set of negative examples; this new formulation, the *one-sided SVM*, ensures through the added constraint $\xi_i \leq 1$ that the resulting classifier in (4) will label all negative training examples correctly. Analogous to the standard SVM, additional benefits are gained by considering the dual formulation

Theorem 1: The dual problem to (6) is given by

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \;& \sum_{i=1}^{n}\min(\alpha_i, C) - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i^T x_j \\ \text{subject to} \;& \alpha_i \geq 0 \\ & \alpha_i \leq C \text{ for } i \in \mathcal{P} \\ & \sum_{i=1}^{n} y_i\alpha_i = 0 \end{aligned} \tag{7}$$

where $\mathcal{P}$ denotes the set of positive examples.

Proof: see Appendix for proof

The dual formulation above is appealing for two reasons: first, it allows us to apply computationally efficient algorithms (e.g. sequential minimal optimization [8]); second, it allows us to consider nonlinear classification functions. Indeed the classifier that results from problem (6) is restricted to be linear with the number of parameters determined by the dimension p of the feature set. Specifically, the separating hyperplanes defined by w and b constitute $p+1$ (equal to five in our case) unknown parameters. Better fitting of data can be accomplished by nonlinear classifiers. Consider a positive definite kernel $k(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to R$ where $\mathcal{X}$ is the original feature space. Given such a kernel, it can be shown that there exists a mapping $\phi : \mathcal{X} \to Z$ where Z is an inner-product space with an inner-product $< \cdot, \cdot >$ such that for all elements $x, \tilde{x}$ in $\mathcal{X}$, $< \phi(x), \phi(\tilde{x}) >= k(x, \tilde{x})$. In our case the original features space is $\mathcal{X} = R^4$. The dimension of Z can be much larger than that of $\mathcal{X}$ and possibly infinite. The classification problem is now cast in the new Hilbert space where hyperplanes are sought to separate the positive examples from the negative examples. The resulting optimization problem is described by (6) with x_i replaced by $\phi(x_i)$ with the appropriate inner-product

$$\begin{aligned} \underset{w,b,\xi_i}{\text{minimize}} \;& \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ \text{subject to} \;& y_i(< w, \phi(x_i) > -b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \\ & \xi_i \leq 1 \text{ for } i \in \mathcal{N}. \end{aligned} \tag{8}$$

Even though problem (8) involves the mapping ϕ, we can efficiently solve it by using the dual formulation

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \;& \sum_{i=1}^{n}\min(\alpha_i, C) - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j k(x_i, x_j) \\ \text{subject to} \;& \alpha_i \geq 0 \\ & \alpha_i \leq C \text{ for } i \in \mathcal{P} \\ & \sum_{i=1}^{n} y_i\alpha_i = 0 \end{aligned} \tag{9}$$

which is equivalent to the previous dual (7) except that it depends on the features only through the kernel function $k(x_i, x_j)$. In addition, determining how the optimal hyperplane classifies x (more precisely $\phi(x)$), does not require knowledge of the map ϕ because the signed distance from the optimal hyperplane determined by $< w, \phi(x) >$ and b can be written in terms of $< \phi(x), \phi(x_i) >$ which in turn is given by the known kernel $k(x, x_i)$. Indeed the following can shown

Theorem 2: For the optimal solution $(w^\star, b^\star)$ of (8), the

following hold true

$$< w^\star, \phi(x) > = \sum_{i=1}^{m} \alpha_i y_i < \phi(x_i), \phi(x) >$$
$$b^\star = \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} [< w^\star, \phi(x_i) > - y_i] \tag{10}$$

where $\mathcal{A} = \{i : 0 < \alpha_i < C\}$ denotes the set of support vectors which lie on the interior of the $[0, C]$ constraint. Thus, given the optimal α_i's, we can compute the distance from the hyperplane for any new example x with an expression that depends only $k(x_i, x)$. In this article we use the Gaussian radial basis function (RBF) kernel given by

$$k(x, \tilde{x}) = \exp(-\|x - \tilde{x}\|^2 / 2\sigma^2). \tag{11}$$

which is a standard choice for kernel support vector machines.

In Figure 4 we construct a simple example using the one-sided SVM with kernels to learn a function on $\mathbb{R}^2$. First, we classify the training examples using the standard linear SVM—note that the separating hyperplane strikes a good balance between misclassification of positive and negative examples. Next, the one-sided SVM shifts this hyperplane so that we have no misclassification of negative examples while still attempting to classify the positive examples correctly. Finally, by using the RBF kernel, we find a nonlinear separator which captures a greater number of positive examples while still keeping the negative examples on one side.

IV. Results and Discussion

In this section we examine the ability of our approach to find safe parameter settings in the simulated microgrid system described previously. We find that given enough data, the model is able to find a set of parameters that avoid cascading failures; we also compare the classification performance of the one-sided SVM to that of the standard SVM and show that on this classification task, our method is better at minimizing the number of false positives while still capturing a large percentage of the true positives.

We generate data by simulating our model assuming a time horizon of $T = 3$ time units with varying load profiles for a total of $n = 5000$ training examples. The electrical signals reside primarily at 60Hz (time period of 16.7 ms) and for long times (which can exceed 30 minutes) of normal operation of washers and dryers not much new information is obtained. To restrict undue and unwarranted computational and simulation burden we run the simulation for scaled down time units. In addition, when evaluating our methods we use 5-fold cross validation whereby we train the functions on 4/5 of the training examples and report classification results for the held out 1/5. We repeat this procedure 5 times and report error results that are the average over these 5 experiments.

In each simulation, we assume that the lights and the water heater remain on throughout the time horizon and vary the load by randomly choosing a start time for the washer and dryer. In choosing the washer and dryer start times there are three three possible scenarios, depicted in Figure 5—since we want to emphasize the worst case scenarios for stability, in 2/5 of the simulations sample the washer and dryer and start times uniformly from $[0.5, 0.55]$, in 2/5 from $[0.5, 0.7]$ and in 1/5 we choose the start times so they do not overlap. In all scenarios the dryer and washer remain on for 1 time unit. Note that in each of these scenarios not only is the steady state current demand varying depending on which devices are on at a particular time, but also the instantaneous current drawn is driven by transients from the loads and the droop characteristic implemented by the inverters—all characteristics typical of a microgrid environment. Finally, in each simulated scenario we pick the parameter settings for each inverter uniformly at random with current thresholds in the range of $[10, 50]$ for inverter 1, $[18, 50]$ for inverter 2 and time limits between $[0.001, 0.04]$. We record which scenarios result in the failure of one or both of the inverters; our task is to classify these parameters in order to identify parameter settings that are stable under all load profiles.

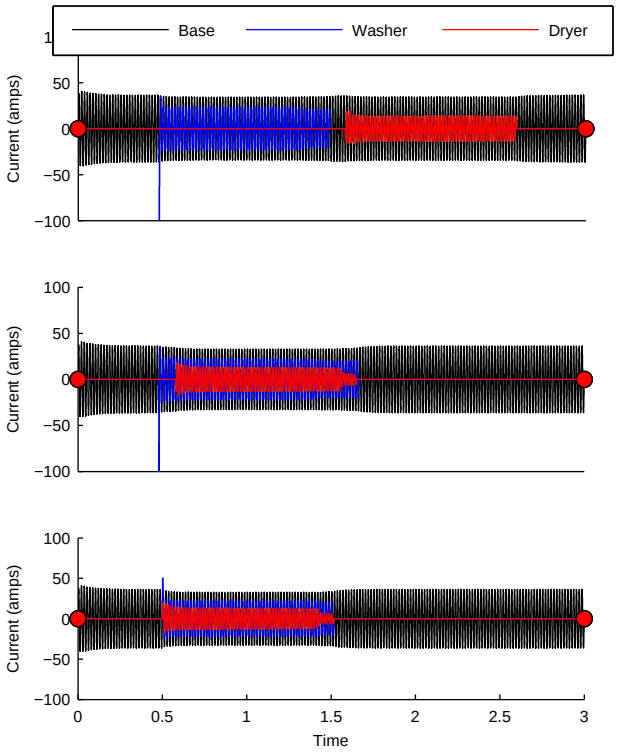

Fig. 5. Example scenarios showing varying simulation conditions: washer and dryer do not overlap (top), washer and dryer overlap in steady state (middle) and washer and dryer overlap during the transient start up (bottom).

TABLE I
Comparison of classification algorithms

SVM	Accuracy	TPR@95	TPR@99	TPR@99.9
Linear	90.78	86.58	72.45	56.48
Linear, OS	69.68	82.67	70.84	62.12
RBF Kernel	**93.08**	**90.42**	76.89	55.17
RBF Kernel, OS	82.28	87.82	**80.35**	**68.47**

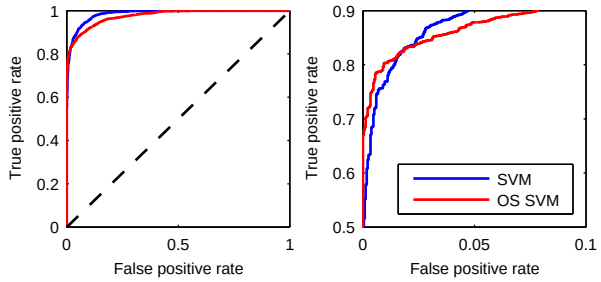

Fig. 6. Comparison of classifiers on the entire range over the entire ROC curve (left) and focused on a low false positive rate (right).

The data generated was used to obtain solutions from four optimization problems—the standard SVM (5) and one-sided SVM (6) are used to learn a linear and nonlinear classifier using the RBF kernel, with the kernel variants requiring solving the dual form (e.g. (7)). In Table I we compare SVM-based algorithms that classify the feature space in terms of no-blackout and blackout outcomes (corresponding to learning f_0). Note that the one-sided SVM with RBF kernel significantly outperforms the standard SVM approaches in maximizing the true positive rate (TPR) at a given false positive level. Here TPR refers to the fraction of no-blackout outcomes that were correctly predicted, and FPR refers to the fraction of blackout outcomes that were incorrectly predicted (as not blackout). As was described in the previous section, the one-sided SVM places larger emphasis on the negative examples which leads to lower a number of false positives at the expense of overall classification accuracy. For our application, ensuring viability of the electrical system is strongly desired as we would like to prevent cascading failures and microgrid collapse as much as possible. We see this tradeoff in Figure 6 which shows that the one-sided SVM achieves better performance on false positive rates less than 3% at the expense of performance on the rest of the curve.

We present results on safe parameter regions for each inverter (corresponding to learning f_1 and f_2) under different scenarios in Figure 7. In our setting, the classification function maps $\mathbb{R}^4 \to \{-1, 1\}$ and thus in order to visualize this function we fix two of the parameters and consider the classifier boundary as we vary the other two. The top row corresponds to fixing the parameters for inverter 2, first with aggressive settings $(0.01, 10)$ (top left) that will typically lead to that inverter failing. In this scenario, inverter 1 must source all of the current and thus the safe parameter settings are much higher that those required when inverter 2 has conservative settings $(0.05, 50)$ (top right). On the other hand, the situation depicted on the bottom row for inverter 2 is qualitatively different: due to the power sharing arrangement between the two inverters, it must be responsible for a significant portion of the load even when inverter 1 is fully operational. Thus even with low thresholds and guard times for inverter 1, the separating boundary is similar to the case where the inverter has high thresholds

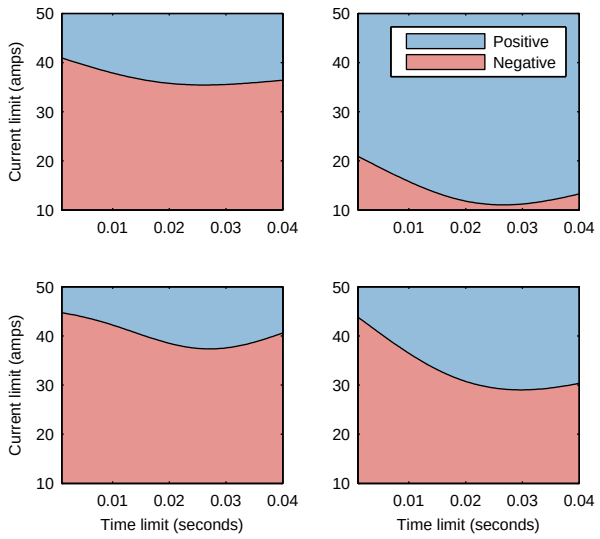

Fig. 7. Learned safe parameter regions for each inverter under different scenarios. The top row shows safe parameters for inverter 1 when inverter 2 has thresholds $(0.01, 10)$ (top left) and $(0.05, 50)$ (top right). Bottom row, vice versa.

and guard times.

The framework presented demarcates regions of safe parameters and unsafe parameters. These regions can be utilized by a micro-grid designer to arrive at breaker settings, where, the designer can incorporate data on the rating of the breaker while making decisions. Furthermore, the TPR and FPR curves can guide the level of robustness desired. Although the article has presented data from a two inverter system, the algorithms presented are scalable to more complex microgrids. Also, there are no restrictions on the topology of the network of inverters assumed. The two inverter system has the ease of presentation and can easily communicate the central ideas which is one of the reasons for the choice of this system.

The framework developed can be extended to expand the feature set. For example, microgrids are envisioned to have a communication layer, where controllers incorporating measurement devices can collect measured signals. These measurements can be added to the feature set which leads to a rich class of problems which can possibly enable of adaptation in real-time for microgrids.

V. Conclusions

In this paper, we developed a machine-learning based framework for designing system parameters in power networks. In particular, we used this framework to design breaker parameters - current threshold and guard time, in a microgrid setting. It was seen that this approach is particularly relevant given the complexity of underlying dynamics in such a network, stochasticity of load schedules and combinatorial aspect in the sequence at which different loads are turned on or off, which make it impossible to deduce from dynamical analysis of the network. The

proposed learning approach is scalable with the number of inverters and loads and can easily incorporate experimental data as well as simulation data; making the accuracy better with more acquired data and therefore will become better with time. We show that with enough data, the breaker parameters can be learnt which will yield fail-safe operation (no cascade failures) with very low false positive rates.

APPENDIX

Proof of Theorem 1

Proof: We simplify notation by defining the sequence $n_i = 1, \ldots, |\mathcal{N}|$ to be the indices of the negative examples and define N to be the $|\mathcal{N}| \times n$ matrix with $N_{n_i,i} = 1$ and 0 everywhere else. Also, we define

$$X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}, \quad \xi = [\xi_1, \ldots, \xi_n], \quad Y = \begin{bmatrix} y_1 & 0 & \cdots \\ 0 & y_2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

and write (6) in Lagrange form

$$L(w, b, \xi, \alpha, \beta, \gamma) = \frac{1}{2}\|w\|^2 + C1^T\xi + \alpha^T(1 - \xi - Y(Xw - b)) - \beta^T\xi + \gamma^T(N\xi - 1)$$

where α, β, and γ are nonnegative Lagrange multpliers. By strong duality we have that

$$\min_{w,b,\xi} \max_{\alpha,\beta\gamma \geq 0} L(w, b, \xi, \alpha, \beta, \gamma) = \max_{\alpha,\beta,\gamma \geq 0} \min_{w,b,\xi} L(w, b, \xi, \alpha, \beta, \gamma)$$

and so we consider

$$\min_{w,b,\xi} L(w, b, \xi, \alpha, \beta, \gamma).$$

Observe that

$$\min_b \alpha^T Y b = \begin{cases} 0 & \text{if } \alpha^T Y = 0 \\ -\infty & \text{otherwise} \end{cases},$$

similarly for $\min_\xi \xi^T(C + N^T\gamma - \alpha - \beta)$; in addition, we optimize over w by taking

$$\nabla_w L = w - X^T Y \alpha = 0$$

giving $w^\star = X^T Y \alpha$. Combining these, we have the dual problem

$$\begin{aligned} \text{maximize} \quad & \alpha^T 1 - \gamma^T 1 - \frac{1}{2}\|X^T Y \alpha\|^2 \\ \text{subject to} \quad & \alpha^T Y = 0 \\ & C + N^T\gamma - \alpha - \beta = 0 \end{aligned}$$

which, written in elementwise notation is

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^n \alpha_i - \sum_{i \in \mathcal{N}} \gamma_{n_i} - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{subject to} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & C + \gamma_{n_i} - \alpha_i - \beta_i = 0 \text{ for } i \in \mathcal{N} \\ & C - \alpha_i - \beta_i = 0 \text{ for } i \in \mathcal{P} \\ & \alpha_i, \beta_i, \gamma_{n_i} \geq 0. \end{aligned}$$

Next, we drop β from the optimization problem since for $\beta_i \geq 0$, $C - \alpha_i - \beta_i = 0$ iff $\alpha_i \leq C$ and similarly for $C + \gamma_i - \alpha_i - \beta_i$. To drop γ, observe that for $i \in \mathcal{N}$, $\alpha_i > C$ implies that we will take $\gamma_{n_i} = \alpha_i - C$ in order to satisfy the constraints while maximizing the objective; thus we replace $\alpha_i - \gamma_{n_i}$ with $\min(\alpha_i, C)$. For $i \in \mathcal{P}$ replacing α_i with $\min(\alpha_i, C)$ has no effect since $\alpha_i \leq C$. This results in the dual (7). ■

REFERENCES

[1] Yi Zhang, Marija D Ilic, and Ozan K Tonguz. Mitigating blackouts via smart relays: a machine learning approach. *Proceedings of the IEEE*, 99(1):94–118, 2011.

[2] Cynthia Rudin, David Waltz, Roger N Anderson, Albert Boulanger, Ansaf Salleb-Aouissi, Maggie Chow, Haimonti Dutta, Philip N Gross, Bert Huang, Steve Ierome, et al. Machine learning for the new york city power grid. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(2):328–345, 2012.

[3] Pierre Geurts and Louis Wehenkel. Early prediction of electric power system blackouts by temporal machine learning. In *Proceedings of the ICML98/AAAI98 Workshop on "Predicting the future: AI Approaches to time series analysis*, pages 24–26, 1998.

[4] J. M. Guerrero, J. C. Vasquez, J. Matas, L. G. de Vicuña, and M. Castilla. Hierarchical control of droop-controlled ac and dc microgrids–A general approach toward standardization. *IEEE Trans. Ind. Electron.*, 58(1):158–172, 2011.

[5] A. Yazdani and R. Iravani. *Voltage-Sourced Converters in Power Systems*. Wiley, 2010.

[6] Régis Vert and Jean-Philippe Vert. Consistency and convergence rates of one-class svms and related algorithms. *The Journal of Machine Learning Research*, 7:817–854, 2006.

[7] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[8] John Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998.