Consumer-Aware Load Control to Provide Contingency Reserves using Frequency Measurements and Inter-load Communication

Jonathan Brooks and Prabir Barooah
University of Florida
Gainesville, Florida USA

***Abstract*— We consider the problem of smart and flexible loads providing contingency reserves to the electric grid based on using local frequency measurements. The impact on consumers must be minimized at the same time. A recent paper by Zhao *et al.* proposed a solution to this optimization problem that was based on solving the dual problem in a distributed manner: local measurements and information exchanged with nearby loads are used to make decisions. In this paper, we provide a distributed algorithm to solve the primal problem. In contrast to the "dual algorithm" of Zhao *et al.*, the proposed algorithm is applicable when consumer disutility is a convex, but not necessarily strictly convex, function of consumption changes; for example, a model of consumer behavior that is insensitive to small changes in consumption. Simulations show the proposed method aids the grid in arresting frequency deviations in response to contingency events. We provide a proof of convergence of the proposed algorithm, and we compare its performance to that of the dual algorithm, when applicable, through simulations.**

I. Introduction

For stable and reliable operation of the power grid, generation must match consumption at all time-scales [1]. Traditionally, generation is matched to consumption through controllable generators that provide not only energy but also ancillary services. Contingency reserves are one such service provided after a sudden change in generation. With the increasing penetration of volatile renewable energies into the power grid, more resources are required to provide contingency reserves. Conventional fossil-fuel generators are often operated at part-load in order to provide spinning reserves (fast-acting contingency reserves). However, generators may be less efficient when rapidly ramping and when operating at part-load, which results in increased emission rates [2]. Part-loading requires additional generators to supply the needs of the grid as well. Building additional fossil-fuel generators to mitigate renewable volatility will reduce the environmental benefits of the renewable energies.

It has been recognized in recent years that an attractive alternative exists: loads can be used to provide spinning reserves by changing their consumption without increasing emissions [3, 4]. Due to the size of the grid, centralized solutions to the load control problem are not practical. A distributed solution is more attractive and is possible by utilizing the cyber-physical nature of the electric grid whereby "information can be transmitted through actuation and sensing" [5]. In particular, loads can provide primary control by using local frequency measurements [6–9]. The value of information contained in frequency measurements has been recognized much earlier [10].

Any changes in consumption to help the grid, however, may incur some cost or disutility for the consumer—such as deviation of the indoor temperature from a comfortable range. Thus there is a need to balance the two—service to the grid and cost to the consumer. In this paper, we consider the problem of designing decision-making algorithms that provide spinning reserves through control of loads while striking this balance.

This paper is inspired by the recent work by Zhao *et al*. [11]. We adopt the problem formulation from [11]: minimize total consumer disutility while returning the consumption-generation mismatch in the grid to zero after a sudden change in generation. The consumption-generation mismatch is estimated by each load from noisy local frequency measurements using a state estimator. The algorithm in [11] is based on solving the dual optimization problem. The dual variable, which is constant across the grid, is iteratively estimated using consensus averaging through inter-node communication.

The algorithm proposed by Zhao *et al.* requires the consumers' disutilities to be strictly convex functions of changes in consumption. Quantifying consumers' disutility in response to consumption changes is challenging, and work in this area is limited. In [12], an exponential function is used to model disutility, while [13] proposes a dynamic disutility model. A study of an industrial aluminum smelting plant providing ancillary services suggests that there may be no disutility for several hours when changing consumption within some threshold of a nominal value, but there *is* significant disutility if consumption is varied too much or for too long [14]. Likewise, [15] showed that consumption in commercial air-conditioning loads can be varied to provide ancillary services without any disutility (adverse effect on indoor climate) as long as the changes in consumption are small in amplitude and band-limited. Based on these studies, we hypothesize that an appropriate model of disutility for many consumers is like the function f_1 shown in Figure 1. The disutility is zero for small changes in consumption, but there *is* non-zero disutility for larger changes in consumption. Such a consumer's disutility is modeled by a convex—not

This research is supported by NSF award ECCS-0955023.

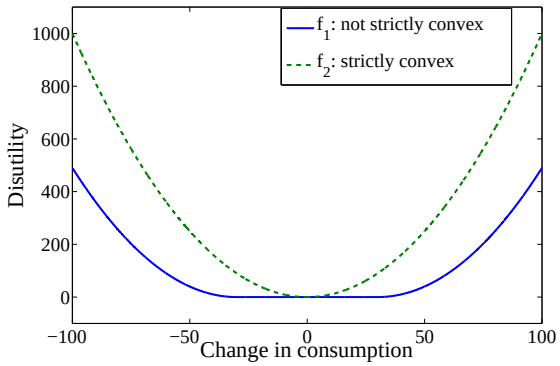

Fig. 1. Alternate models of consumer disutility vs. consumption change.

strictly convex—function of consumption change.

In this work, we propose a method to solve the *primal* optimization problem in a distributed manner, which we call the Distributed Gradient Projection (DGP) algorithm. The main contribution over that of [11] is that the method proposed here is applicable to disutility functions that are not strictly convex. The algorithm proposed in [11] requires the disutility functions to be strictly convex since the inverse image of the gradient is used in the computation. The inverse image does not exist if there is a linear region in the disutility function (such as in f_1 in Figure 1). We prove that the DGP algorithm converges to an optimal solution almost surely. In this preliminary work, our convergence proof is limited to the case where there are no upper and lower bounds on how much a load can change its consumption. However, we test our algorithm through simulations in the more realistic case where each load has an upper and a lower bound. Simulations indicate our proposed algorithm performs well in both scenarios: with and without bounds on changes in consumption. It is able to reduce frequency excursions following step changes in generation. Simulation comparisons, in those scenarios where comparison is possible, show that the proposed DGP algorithm performs similarly or slightly better than the dual algorithm of [11].

Our work is also closely related to [16], which proposed a distributed algorithm to solve the unit-commitment problem for generators: determine setpoints for generators to match consumption while respecting individual generator constraints and minimizing the total generation cost. However, the algorithm proposed in [16] requires generators to know part of the total load such that the total load is fully known among the entire generation network (even if no single generator knows the total load). In contrast, the DGP algorithm requires no loads to know the total mismatch; rather the mismatch is estimated by each load independently via local frequency measurements.

Although distributed optimization has been studied in the literature (e.g., [17–19]), the solutions proposed in these references do not exploit the special structure of our problem due to the cyber-physical nature of the power grid: the ability to estimate at each load the amount of violation of the constraint (consumption-generation mismatch) purely from local measurements (frequency).

This paper is organized as follows. Section II formally defines the problem that we solve. In Section III, we propose our solution method. We provide a proof of convergence in Section IV, and we describe the simulation parameters in Section V-A. In the remainder of Section V, we compare the simulation results to those in [11]. Finally, Section VI concludes this work and discusses avenues for future work.

II. Problem Formulation

As in [11], we consider an electric grid with a single frequency throughout the grid, whose nominal value is denoted by ω_*. This is the case when electrical distances are negligible—such as in a microgrid. There are n controllable loads, and the deviation of load i's consumption from its nominal value is denoted by x^i. Load i incurs a disutility $f^i(x^i)$ as a result of a consumption change x^i. The deviation must lie in $\Omega^i \triangleq [\underline{x}^i, \bar{x}^i]$, where $\underline{x}^i \leq 0 \leq \bar{x}^i$ are specified a-priori.

Let Δg be the generation deviation from the nominal value. The problem is for the loads to decide how much to change their own consumption so that the consumption-generation mismatch is diminished while the resulting disutility of the loads is minimized:

$$\min_{x^i,\ i=1,\ldots,n} \sum_{i=1}^{n} f^i(x^i), \quad \text{s. t. } \sum_{i=1}^{n} x^i = \Delta g, \quad x^i \in \Omega^i, \quad (1)$$

Load i can obtain a noisy measurement of the grid frequency and can use it to make a decision on x^i. In addition, the computation of the decision variables x^i must be distributed in the following sense. There is a connected communication graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the node set $\mathcal{V} = \{1, 2, \ldots, n\}$ is simply the loads and the edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, specified a-priori, determines which pairs of loads can exchange information. The set of neighbors $\mathcal{N}^i$ of load i, with which it can exchange information, is defined by $\mathcal{N}^i = \{\ j \mid (i,j) \in \mathcal{E}\}$. The frequency measurements are essential since every load can use them to estimate the equality constraint violation $u \triangleq \Delta g - \sum_{i=1}^{n} x^i$. How this is done is described in Section III-A.

Even though Problem (1) is not a dynamic optimization problem, time plays a role since the noise on frequency measurement is naturally modeled as a stochastic process, and consequently the estimates of u obtained by every node vary with time.

Time is measured by a discrete iteration counter: $k = 0, 1, \ldots$. The generation at time k is denoted by g_k so that the generation change from nominal is $\Delta g_k \triangleq g_k - g_*$, where g_* is the nominal generation. We assume that at $k = 0$ total load and total generation are equal, and we limit ourselves to step changes. That is, $\Delta g_0 = 0$ and $\Delta g_k = \bar{g}$ for $k \geq K$ for some K, where $\bar{g}$ is the step change.

III. Distributed Gradient Projection (DGP) Algorithm

To describe the algorithm, we define the consumption-generation mismatch at iteration k:

$$u_k \triangleq \Delta g_k - \sum_{i=1}^{n} x_k^i = \Delta g_k - \mathbf{1}^T \boldsymbol{x}_k, \tag{2}$$

where $\boldsymbol{x}_k \triangleq [x_k^1, \ldots, x_k^n]^T$ and $\mathbf{1} \in \mathbb{R}^n$ is a vector of all ones. Neither $\sum_{i=1}^n x_k^i$ nor Δg_k is known to any of the loads. However, load i can obtain a noisy measurement of the frequency deviation $\Delta\omega_k \triangleq \omega_k - \omega_*$, which is denoted by $\Delta\tilde{\omega}_k^i$. It uses this measurement to estimate the mismatch, which is denoted by $\hat{u}_k^i$,

The update law of the DGP algorithm comprises of three main operations: (i) a generation-matching step, (ii) a gradient descent step, and (iii) a projection step. The first step uses the estimated mismatch, $\hat{u}_k^i$, to compute a change in the consumption that will reduce the mismatch. Pure gradient descent to reduce the cost, though possible due to the separable cost function, will violate the equality constraint (consumption-generation matching). Therefore the gradient descent step is designed to be orthogonal to the generation-matching step, i.e., it does not change the total consumption. The updates computed by the first two steps are added and then projected onto Ω^i to respect the upper and lower bounds on consumption change.

The update law of the DGP algorithm at load i at time k is summarized below:

DGP Algorithm:

1) Obtain $\hat{u}_k^i$ from the measurement $\Delta\tilde{\omega}_k^i$ using a state estimator, which is described in Section III-A. The *generation-matching step* is then $b\gamma_k\hat{u}_k^i$, where γ_k is a step size and b is a positive constant.
2) Compute gradient $\frac{d}{dx^i}f^i(x_k^i)$, transmit gradient value to neighbors, and receive neighbors' gradient values. Compute the *gradient descent step* Δx_k^i as the i-th entry of $\Delta\boldsymbol{x}_k$, where
$$\Delta\boldsymbol{x}_k \triangleq -L\nabla f(\boldsymbol{x}_k)^T, \tag{3}$$
where L is the Laplacian matrix of the communication graph $\mathcal{G}$ [20].
3) Compute $x_{k+1}^i = P_{\Omega^i}\left[x_k^i + a\alpha_k\Delta x_k^i + b\gamma_k\hat{u}_k^i\right]$, where $P_{\Omega^i}[\cdot]$ denotes the standard projection operator, α_k is a step size, and a is a positive constant.

Since L is a Laplacian matrix, the only entries of the i-th row of L that are non-zero are those that correspond to the neighbors of i in $\mathcal{G}$ [20]. It follows from (3) that load i requires only $\frac{d}{dx^j}f^j(x_k^j),\ j \in \mathcal{N}^i$. That is, the iterates can be computed by every load in a distributed manner.

A. Estimation of consumption-generation mismatch using frequency measurements

We borrow the estimation method proposed in [11] for use in this paper, though it is possible to use any estimator in the DGP algorithm. The power grid is modeled as a discrete-time LTI system with consumption-generation mismatch u_k as the input and frequency deviation from nominal $\Delta\omega_k$ as the output. At each time k, load i obtains the noisy measurement $\Delta\tilde{\omega}_k^i$ to estimate the state of the plant by using the estimator in [21], which was developed for estimating the state of a system with an unknown input. Once the state estimate is obtained, each load estimates the unknown input by essentially assuming that the most recent output is error-free and solving for the previous input from the state-update equation. We omit the details here; the interested reader is referred to [11].

We denote the estimation error at time k by $\boldsymbol{\epsilon}_k \triangleq \hat{\boldsymbol{u}}_k - u_k\mathbf{1}$, where $\hat{\boldsymbol{u}}_k$ is the column vector of $\hat{u}_k^i$'s. Define the σ-algebra $\mathcal{F}_{K-1} := \sigma(\epsilon_{k-1}^i |\ i \in \mathcal{V},\ 1 \le k \le K)$. It was shown in [21] that

$$\mathbb{E}[\epsilon_k^i | \mathcal{F}_{k-1}] = 0. \tag{4}$$

In [11], it was shown that the estimation error converges in m.s. for the power system model considered. This combined with (4) implies that the estimation error sequence $\boldsymbol{\epsilon}_k$ is a martingale-difference sequence.

IV. Convergence analysis

A. Main Results

We make the following assumptions for our analysis.

Assumption 1. (Technical assumptions)**.**

1) $\alpha_k = c\gamma_k$ for some positive constant c.
2) The function $\gamma_k \to 0$ satisfies $\sum_{k=0}^{\infty}\gamma_k = \infty$ and $\sum_{k=0}^{\infty}\gamma_k^2 < \infty$.
3) The estimation error sequence, $\boldsymbol{\epsilon}_k$, is a martingale-difference sequence.

Assumption 2. (Assumptions on disutility)**.**

1) $f^i(x^i)$ is convex for each i with a (not necessarily unique) minimum at $x^i = 0$.
2) $f^i(x^i)$ is coercive for each i; i.e., $\{x^i | f^i(x^i) \le F\}$ is compact for every $F \ge 0$ for each i.
3) $f^i(x^i)$ is continuously differentiable for each i.
4) $\nabla f^i(x^i)$ is Lipschitz for each i.

Assumption 3. (Assumptions on loads and generators)**.**

1) $\Omega^i = \mathbb{R}$ for each i.
2) $\mathcal{G}$ is connected.
3) $\Delta g_k \equiv \bar{g}$ for all $k \ge 0$.

Assumption 4 (Additional assumption on estimation error)**.** $\sup_{i,k}|\epsilon_k^i| < \bar{\epsilon} < \infty$.

Assumptions 1(1) and 1(2) are satisfied by choice of α_k, and γ_k, and Assumptions 1(2) and 1(3) are standard technical assumptions in the field of stochastic approximation. For the estimator used in this work, Assumption 1(3) is satisfied as discussed in Section III-A. Assumption 2 is readily met because $f(\boldsymbol{x})$ is a modeling choice. Assumption 3(1) is the main limiting one: it states that there are no upper and lower limits on possible changes in consumption. Assumption 3(3) means that we only consider a step-change in generation.

The main convergence result is the following.

Theorem 1. *If Assumptions 1, 2, and 3 hold, $\boldsymbol{x}_k$ converges to a solution of Problem (1) in the mean. If in addition, Assumption 4 holds, $\boldsymbol{x}_k$ converges to a solution to Problem (1) almost surely.*

The technique used to prove this result is known as the o.d.e. method of stochastic approximation, which establishes a rigorous connection between noisy discrete iterations and a continuous-time o.d.e. [22].

Proposition 1 (Theorem 2 (Chapter 2) in [22])**.** *Consider the sequence $\{\boldsymbol{y}_k\}$ generated by the iteration*

$$\boldsymbol{y}_{k+1} = \boldsymbol{y}_k + \gamma_k[h(\boldsymbol{y}_k) + \boldsymbol{\epsilon}_k],$$

where $h(\boldsymbol{y}) : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz and $\{\boldsymbol{\epsilon}_k\}$ is a martingale-difference sequence. If γ_k satisfies Assumption 1(2) and $\sup_k \|\boldsymbol{y}_k\| < \infty$ almost surely (a.s.), then $\boldsymbol{y}_k$ converges a.s. to a (possibly sample-path dependent) compact, connected, internally chain-transitive invariant set of the o.d.e.

$$\dot{\boldsymbol{y}}(t) = h(\boldsymbol{y}(t)).$$

As in many applications of the o.d.e. method, the main hurdle in analyzing convergence of the DGP algorithm is to establish boundedness of the iterates $\boldsymbol{x}_k$. Presence of the projection step guarantees boundedness trivially, but the corresponding o.d.e. can create spurious, undesired equilibria. In this preliminary work, we have therefore limited ourselves to the case where there is no projection, i.e., no bounds on the changes in consumption, but boundedness is no longer guaranteed. However, by taking expectation of both sides of the update law, noise can be eliminated and the analysis can be carried out in the deterministic setting. In this case, boundedness can be established by Assumption 2. For the stochastic case, we make Assumption 4 to prove boundedness of $\boldsymbol{x}_k$.

However, if the disutilities are quadratic, boundedness of the iterates is achieved via a technique in [22]. In that case we can remove the assumption on the estimation error being bounded:

Theorem 2. *Let Assumptions 1, 2, and 3 hold, and let $f^i(x^i) = q^i(x^i)^2/2$, where $q^i > 0$. Then $\boldsymbol{x}_k \to \boldsymbol{x}_*$ a.s., where $\boldsymbol{x}_*$ is the unique optimal solution to the optimization Problem 1.*

Due to lack of space, we omit the proof of this result, which is provided in [23].

B. Proof of Theorem 1

We must now introduce some notation. For a given ℓ, define the $(n-1)$-dimensional hyperplane $H(\ell) \triangleq \{\boldsymbol{x} | \mathbf{1}^T\boldsymbol{x} = \ell\}$ and $X(\ell) = \{\boldsymbol{x} \in H(\ell) | f(\boldsymbol{x}) \leq f(\boldsymbol{y}),\ \boldsymbol{y} \in H(\ell)\}$. It follows that $H(\bar{g})$ is the set of all feasible solutions, and $X(\bar{g}) \subset H(\bar{g})$ is the set of all solutions to Problem (1). Since f is convex and the equality constraint is linear, necessary conditions for $\boldsymbol{x}_*$ to be optimal are also sufficient; they are

$$\nabla f(\boldsymbol{x}_*) + \lambda_* \mathbf{1} = 0, \quad \bar{g} - \mathbf{1}^T\boldsymbol{x}_* = 0, \tag{5}$$

for some scalar λ_* [24]. The interpretation of (5) is that $\nabla f(\boldsymbol{x}_*) \parallel \mathbf{1}$ and $u = 0$.

The following lemma states that the iterates $\boldsymbol{x}_k$ are asymptotically feasible a.s. Note that Assumption 4 (boundedness of estimation error) is not required for this result.

Lemma 1. *Let Assumptions 1, 2, and 3 hold, then $\boldsymbol{x}_k \to H(\bar{g})$ a.s. Furthermore, all trajectories of the o.d.e.*

$$\dot{\boldsymbol{x}}(t) = -L\nabla f(\boldsymbol{x}(t))^T + (-\mathbf{1}^T\boldsymbol{x}(t) + \bar{g})\mathbf{1} \tag{6}$$

converge to $H(\bar{g})$. Consequently, $u(t) \to 0$, where

$$u(t) \triangleq \bar{g} - \mathbf{1}^T\boldsymbol{x}(t). \tag{7}$$

Proof. See Appendix. □

The following lemma states conditions for the DGP step direction to be a descent direction.

Lemma 2. *Let Assumptions 1, 2, 3, and 4 hold. If $\|\boldsymbol{x}_k\|$ and k are sufficiently large, then $d(\boldsymbol{x}_k) \triangleq -L\nabla f(\boldsymbol{x}_k) + u_k\mathbf{1} + \boldsymbol{\epsilon}_k$ is almost surely a descent direction; that is, $d(\boldsymbol{x}_k)^T\nabla f(\boldsymbol{x}_k)^T < 0$ a.s.*

Proof. See Appendix. □

Lemma 3 below is a consequence of Lemma 2. The proof can be found in [23]. Due to lack of space, we omit the proof here.

Lemma 3. *Let Assumptions 1, 2, 3, and 4 hold. If $\|\boldsymbol{x}_k\|$ and k are sufficiently large, then $f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k)$ a.s.*

Proof. See [23]. □

An immediate consequence of Lemma 3 is the following corollary which establishes boundedness of the iterates—a condition needed for applying Proposition 1.

Corollary 1. *Let Assumptions 1, 2, 3, and 4 hold. Then $\sup_k \|\boldsymbol{x}_k\| < \infty$ a.s.*

Proof. By Lemma 3, for large enough $\|\boldsymbol{x}_k\|$ and large enough k, $f(\boldsymbol{x}_{k+1}) < f(\boldsymbol{x}_k)$ a.s. Therefore $\sup_k f(\boldsymbol{x}_k) < \infty$ a.s. It follows that $\sup_k \|\boldsymbol{x}_k\| < \infty$ a.s. because $f(\boldsymbol{x})$ is coercive. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. Proving convergence in the mean is very similar to the proof of a.s. convergence but simpler, so we only provide the proof of a.s. convergence.

By Corollary 1, $\sup_k \|\boldsymbol{x}_k\| < \infty$ a.s. Therefore, by Proposition 1, the iterates of the DGP algorithm converge almost surely to a compact, connected, internally chain-transitive invariant set of the o.d.e. (6). We call this set I.

Our proof consists of two main parts: (i) we show $I \subseteq E$, where E is the set of equilibrium points of (6); (ii) we show $E = X(\bar{g})$; that is, the set of equilibrium points of (6) is precisely the set of solutions to Problem (1).

If E is globally attractive (i.e., if all trajectories $\boldsymbol{x}(t) \to E$ for any $\boldsymbol{x}(t_0)$ for some t_0), then all internally chain-transitive invariant sets of (6) must be contained within E. Therefore,

it suffices to show $\boldsymbol{x}(t) \to E$. From Lemma 1, $\boldsymbol{x}(t) \to H(\bar{g})$—the set of all feasible points. Denote the RHS of (6) by $h(\boldsymbol{x}(t))$. Note that $\nabla f(\boldsymbol{y})h(\boldsymbol{y}) \leq 0$ for all $\boldsymbol{y} \in H(\bar{g})$ because L is positive semidefinite [20]. Because $\nabla f(\boldsymbol{x})h(\boldsymbol{x})$ is a continuous function of $\boldsymbol{x}$, $\boldsymbol{x}(t) \to H(\bar{g})$ implies that $\dot{f}(\boldsymbol{x}(t)) = \nabla f(\boldsymbol{x}(t))h(\boldsymbol{x}(t)) \to \mathbb{R}_{\leq 0}$.

Next we show by contradiction that $\dot{f}(\boldsymbol{x}(t)) \to 0$. Suppose $\dot{f}(\boldsymbol{x}(t)) \nrightarrow 0$; that is, suppose $\nabla f(\boldsymbol{x}(t))h(\boldsymbol{x}(t)) \to \mathbb{R}_{<0}$. Then there exists $\delta > 0$ such that for all $T > 0$ there exists $t > T$ with $|\nabla f(\boldsymbol{x}(t))h(\boldsymbol{x}(t))| \geq \delta$. Because $\nabla f(\boldsymbol{x}(t))h(\boldsymbol{x}(t)) \to \mathbb{R}_{\leq 0}$, this implies

$$\dot{f}(\boldsymbol{x}(t)) \equiv \nabla f(\boldsymbol{x}(t))h(\boldsymbol{x}(t)) \leq -\delta$$

infinitely often. Therefore, $f(\boldsymbol{x}(t))$ is decreasing without bound, which is a contradiction because $f(\boldsymbol{x})$ has a minimum by Assumption 2. Therefore, $\boldsymbol{x}(t) \to E$ (i.e., E is a global attractor). Hence, $I \subseteq E$.

E is the set of points where the RHS of (6) is zero. Because $-L\nabla f(\boldsymbol{x})^T \perp \mathbf{1}$, the RHS of (6) is zero if and only if $-L\nabla f(\boldsymbol{x}) = \mathbf{0}$ and $u = 0$. Because $\mathcal{G}$ is connected, $-L\nabla f(\boldsymbol{x}) = \mathbf{0}$ if and only if $\nabla f(\boldsymbol{x}) \parallel \mathbf{1}$, and $u = 0$ if and only if $\boldsymbol{x} \in H(\bar{g})$. These are precisely the necessary and sufficient conditions (5). Therefore, $E = X(\bar{g})$. Combining this result with the previous result, we have $\boldsymbol{x}_k \to I \subseteq E = X(\bar{g})$ by Proposition 1, which proves the theorem.

□

V. Simulation results

In the sequel, we refer to the algorithm proposed in [11] as the "dual algorithm."

A. Simulation Setup

Figure 2 shows the system architecture used for design and simulation. The generator dynamics block shown in Figure 2 also includes local controls that are usually present in generators. The loss of generation is modeled as an exogenous disturbance $\bar{g}$ in the figure. The estimator in the figure is the one described in Section III-A to estimate the consumption-generation mismatch from local, noisy frequency measurements. The process disturbance, ζ, and measurement noise, ξ^i, at each load are modeled as wide-sense stationary white noise. For ease of comparison between the proposed DGP algorithm and the dual algorithm, we use the same generator dynamics, noise statistics, and communication graph as in [11], and the reader is referred to that work for more detailed information about the simulation model or implementation of the state estimator.

Even without the use of intelligent loads, the local generator control will change the generator setpoint in response to frequency deviation to match consumption, which will restore the frequency to its nominal value on its own. Intelligent loads are supposed to help the generator in reacting to frequency deviations faster so that large excursions of system frequency are avoided.

For each load i, we consider both constrained and unconstrained changes in consumption. For the constrained

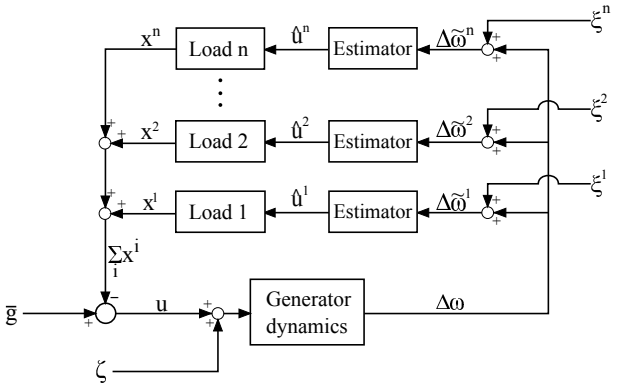

Fig. 2. System architecture for simulations. Inter-load communication is not shown.

case, $\Omega^i = [-\bar{x}^i, \bar{x}^i]$, where $\bar{x}^i$ is chosen from a uniform distribution and then normalized so that $\sum_{i=1}^n \bar{x}^i = 60$ MW (as in [11]).

We test the performance of the DGP algorithm with two distinct disutility functions. The first is a convex but not strictly convex function:

$$f^i(x^i) = \begin{cases} 0, & |x^i| \leq a^i \\ q^i(x^i - a^i)^2, & |x^i| \geq a^i, \end{cases} \tag{8}$$

where $a^i = 0.1\bar{x}^i$; for the unconstrained case, we use a^i from the constrained case. The consumer does not experience any disutility as long as the load variation is within $\pm a^i$. The second disutility function *is* strictly convex:

$$f^i(x^i) = \frac{q^i}{2}(x^i)^2. \tag{9}$$

For both disutility functions, we pick q^i to be an arbitrary positive number such that $1/q^i$ is chosen from a uniform distribution on the interval $[0.1,\ 0.3]$. This is chosen for comparison with [11], which makes a similar choice for disutility functions.

The initial conditions are $g_0 = 200$ MW and $u_0 = 0$. Generation undergoes two contingencies modeled as step changes:

$$g_k = \begin{cases} 200 \text{ MW}, & 0 \text{ s} \leq kT < 20 \text{ s} \\ 190 \text{ MW}, & 20 \text{ s} \leq kT < 50 \text{ s} \\ 170 \text{ MW}, & 50 \text{ s} \leq kT, \end{cases}$$

where $T = 0.1$ seconds is the discretization interval.

The simulations are conducted with a 1D-grid communication network, where each load i communicates with loads from $\max\{1,\ i - n_0\}$ to $\min\{n,\ i + n_0\}$, where $n_0 \leq n$. For the simulation results reported here, we use $n = 1000$ and $n_0 = 1$.

Additionally, we use $a = 5$, $b = 1.5$, $c = 1$, and $\gamma_k = \gamma_0/(k^{0.8})$ for $k > 0$, with $\gamma_0 = 4\underline{q}/n$, where $\underline{q} \triangleq \min_i q^i$.

With all of these parameter choices, Assumptions 1, 2, 3(2), and 3(3) are satisfied. Note that Assumption 1(3) is satisfied from the discussion in Section III-A. Theorem 1 is applicable to both disutility

functions (8) and (9). Theorem 2 applies to disutility function (9).

B. Results with non-strictly convex disutility function

Here we report simulation results with the consumer disutility function (8). The dual algorithm is not applicable because the inverse of $\nabla f(\boldsymbol{x})$ must exist in Ω to implement the dual algorithm, which is not the case when $|x^i| \leq a^i$.

Although the analysis presented in this work is for the scenario with no projection, Figure 3 shows simulation results for both the projected and non-projected case (i.e., without and with Assumption 3(1), respectively); the system frequency without smart loads (i.e., with generator-only control) is shown in red as well.

System frequency is similar both with and without projection. There is a lower disutility for the scenario with projection; this may be caused by the algorithm reaching a "wall" and then having slower convergence thereafter compared to the scenario without projection. However, using the DGP algorithm, the loads are able to assist the generator in avoiding large frequency deviations from the nominal when each contingency occurs—even with projection.

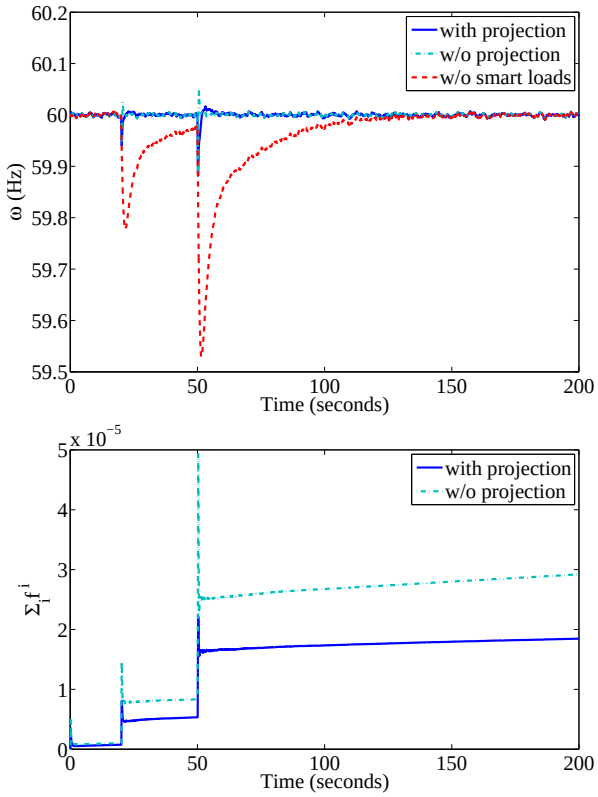

Fig. 3. Performance of the DGP algorithm with consumer disutility that is not strictly convex. Step changes in generation occur at 20 and 50 seconds.

C. Comparison with dual algorithm

Figure 4 shows the results of the DGP and dual algorithms with quadratic disutility functions (9) *with* projection. As the figure shows, the DGP algorithm results in a significantly smaller frequency drop compared to both generator-only control and the dual algorithm. Although the dual algorithm returns the frequency to the nominal value faster than generator-only control, it does not reduce the initial frequency drop as much as the DGP algorithm.

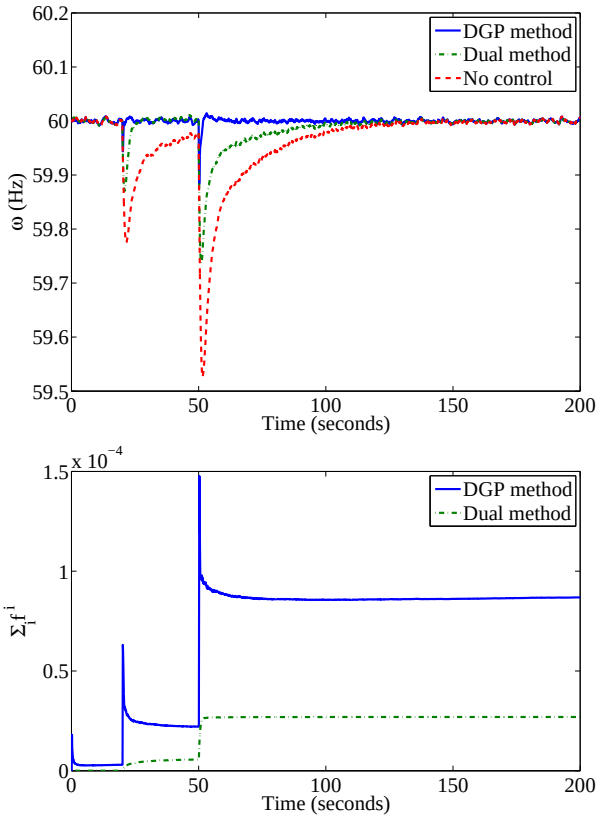

Fig. 4. Performance of the DGP and dual algorithms with quadratic consumer disutility with projection. Step changes in generation occur at 20 and 50 seconds.

However, the consumer disutility is significantly lower for the dual algorithm than for the DGP algorithm. This is because the dual algorithm is responding more slowly than the DGP algorithm, so the equality constraint is not being satisfied—resulting in a lower cost. The slower response of the dual algorithm is due to the inversion of the derivative of each load's disutility function. Because the derivative of each disutility function is rather steep, the inverse is quite flat, so large changes in its argument still result in small changes in its value—leading to small changes in consumption. Conversely, the DGP algorithm aggressively meets the equality constraint because it uses the gradient direction, which is rather steep. This results in a lower frequency deviation but more disutility as the loads are changing consumption more aggressively.

The dual algorithm appears to have a significantly lower steady-state disutility because the generator control restores much of the frequency. The loads interpret the restored

frequency as a smaller consumption-generation mismatch, which results in less change in consumption and therefore in lower disutility.

Although we do not report them here, simulations with varying number of loads ($n = 10,\ 100$) and varying amount of communication ($n_0 = 10,\ 100,\ 1000$) showed similar trends as in the $n = 1000$, $n_0 = 1$ case. It was observed in [11] that the dual algorithm showed similar behavior. We conclude that performance of both algorithms is largely unaffected as the number of loads and degree of communication increases.

VI. Conclusion

The proposed DGP algorithm solves a constrained optimization problem in a distributed manner to aid a power grid in maintaining system frequency near its nominal value while minimizing consumers' disutilitiy. The DGP algorithm solves the primal problem, whereas prior work solved the dual problem [11]. The advantage of the proposed primal method is that it is not restricted to strictly convex disutility functions; rather it is applicable to generally convex disutility functions that capture a consumer behavior that may be quite common. Simulations show that the algorithm is effective in reducing frequency excursions after contingency events while keeping the consumer disutility low. Simulations also show that the DGP algorithm performed either better than or similar to the dual algorithm from [11] in maintaining frequency.

In this preliminary work, we proved that the DGP algorithm converges to the optimal solution under two idealized assumptions. The first one is that there is no upper or lower bound on possible consumption change. The reason for this assumption is ease of analysis. The projection step of the algorithm that enforces bounds leads to potentially spurious equilibria, making the analysis more challenging. Future work will focus on removing this assumption. Simulation results with and without upper and lower bounds imposed through projections are promising: there is hardly any difference in the behavior of the algorithm between the two cases. The second is the assumption that the estimation errors are bounded. This is due to difficulty in proving the iterates are bounded without projection, so removing the first assumption automatically removes this assumption. In this paper, we have been able to remove this assumption for a specific disutility function even in the projection-free case (Theorem 2).

Other interesting paths for future work include extension of the DGP algorithm to time-varying communication networks and time-varying changes in generation.

Acknowledgment

The authors thank C. Zhao and S. Low for their assistance in implementing the power system model used in this work and in reproducing the results of [11].

References

[1] B. Kirby, "Ancillary services: Technical and commercial insights," 2007, prepared for Wärtsilä North America Inc.

[2] D. Lew, G. Brinkman, E. Ibanez, B. Hodge, and J. King, "The western wind and solar integration study phase 2," *National Renewable Energy Laboratory, NREL/TP-5500*, vol. 55588, 2013.

[3] P. Siano, "Demand response and smart gridsa survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.

[4] M. Milligan and B. Kirby, "Utilizing load response for wind and solar integration and power system reliability," in *Wind Power Conference, Dallas, Texas*, 2010.

[5] S. Bolognani and S. Zampieri, "A distributed control strategy for reactive power compensation in smart microgrids," *IEEE Trans. on Automatic Control*, vol. 58, no. 11, November 2013.

[6] Prepared by NERC RS Committee, "Balancing and Frequency Control: A Technical Document Prepared by the NERC Resources Subcommittee," NERC Technical Report, pp. 1–53, January 26 2011.

[7] Y. V. Makarov, L. S., J. Ma, and T. B. Nguyen, "Assessing the value of regulation resources based on their time response characteristics," Pacific Northwest National Laboratory, Richland, WA, Tech. Rep. PNNL-17632, June 2008.

[8] J. A. Short, D. G. Infield, and L. L. Freris, "Stabilization of grid frequency through dynamic demand control," *Power Systems, IEEE Transactions on*, vol. 22, no. 3, pp. 1284–1293, 2007.

[9] A. Molina-García, F. Bouffard, and D. S. Kirschen, "Decentralized demand-side contribution to primary frequency control," *Power Systems, IEEE Transactions on*, vol. 26, no. 1, pp. 411–419, 2011.

[10] F. Schweppe, R. Tabors, J. Kirtley, H. Outhred, F. Pickel, and A. Cox, "Homeostatic utility control," vol. PAS-99, no. 3, pp. 1151 –1163, May 1980.

[11] C. Zhao, U. Topcu, and S. H. Low, "Optimal load control via frequency measurement and neighborhood area communication," *Power Systems, IEEE Transactions on*, pp. 3576–3587, 2013.

[12] P. Khadgi, L. Bai, and G. Evans, "Modeling demand response using utility theory and model predictive control," in *IIE Annual Conference. Proceedings*. Institute of Industrial Engineers-Publisher, 2014, p. 1262.

[13] B. Zhang, M. C. Caramanis, and J. Baillieul, "Optimal price-controlled demand response with explicit modeling of consumer preference dynamics," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 2481–2486.

[14] D. Todd, M. Caufield, B. Helms, A. P. Generating, I. M. Starke, B. Kirby, and J. Kueck, "Providing reliability services through demand response: A preliminary evaluation of the demand response capabilities of alcoa inc," *ORNL/TM*, vol. 233, 2008.

[15] Y. Lin, P. Barooah, S. Meyn, and T. Middelkoop, "Experimental evaluation of frequency regulation from commercial building HVAC systems," *IEEE Transactions on Smart Grid*, vol. 6, pp. 776 – 783, 2015.

[16] A. Cherukuri and J. Cortes, "Distributed coordination for economic dispatch with varying load and generator commitment," in *Allerton Conference on Communications, Control, and Computing*, 2014.

[17] A. Nedic and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1757–1780, 2009.

[18] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *Automatic Control, IEEE Transactions on*, vol. 57, no. 1, pp. 151–164, 2012.

[19] M. Zhu and S. Martínez, "An approximate dual subgradient algorithm for multi-agent non-convex optimization," *Automatic Control, IEEE Transactions on*, vol. 58, no. 6, pp. 1534–1539, 2013.
[20] C. Godsil and G. Royle, *Algebraic Graph Theory*, ser. Graduate Texts in Mathematics. Springer, 2001.
[21] P. K. Kitanidis, "Unbiased minimum-variance linear state estimation," *Automatica*, vol. 23, no. 6, pp. 775–778, 1987.
[22] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.
[23] J. Brooks and P. Barooah, "Consumer-aware load control to provide contingency reserves using frequency measurements and inter-load communication: Technical report," University of Florida, Department of Mechanical and Aerospace Engineering, Tech. Rep., 2015. [Online]. Available: plaza.ufl.edu/brooks666
[24] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.
[25] V. S. Borkar and S. P. Meyn, "The ode method for convergence of stochastic approximation and reinforcement learning," *SIAM Journal on Control and Optimization*, vol. 38, no. 2, pp. 447–469, 2000.

APPENDIX

The following proposition is an algebraic relationship that is easily verified and is used in the proof of Lemma 2.

Proposition 2. *Let $a^{(i)}$, $i = 1, \dots, n$ be scalars such that $a^{(1)} \geq a^{(2)} > 0$ and $|a^{(1)}| \geq |a^{(i)}|$ for all i. Let $b^{(i)}$, $i = 1, \dots, n$ be scalars such that $|b^{(i)}| < B$ for all i. If $a^{(2)} > 2a^{(3)} + n\frac{a^{(1)}}{a^{(2)}}B$, then $\left(a^{(2)} - a^{(3)}\right)^2 > \sum_{i=1}^{n} b^{(i)}a^{(i)}$.*

Proof of Lemma 1. From (7), we have

$$\begin{aligned} u_{k+1} &= \bar{g} - \mathbf{1}^T\boldsymbol{x}_k + \gamma_k\mathbf{1}^T L\nabla f(\boldsymbol{x}_k)^T \\ &\quad - \gamma_k u_k\mathbf{1}^T\mathbf{1} - \gamma_k\mathbf{1}^T\boldsymbol{\epsilon}_k \\ &= u_k + \gamma_k[v(u_k) + \mathbf{1}^T\boldsymbol{\epsilon}_k], \end{aligned}$$

where $v(u_k) = -nu_k$; it was shown in [25] that if the scaled o.d.e. $\dot{u}(t) = \lim_{r\to\infty} v(ru(t))/r$ is asymptotically stable, then $\sup_k |u_k| < \infty$ a.s. Therefore, we may apply Proposition 1, and u_k converges a.s. to an internally chain-transitive invariant set of the o.d.e. $\dot{u}(t) = -nu(t)$.

Because this o.d.e. is exponentially stable, the only internally chain-transitive invariant set of the o.d.e. is the origin. Therefore, $u_k \to 0$ a.s. Hence, $\boldsymbol{x}_k \to H(\bar{g})$ a.s. Additionally, $u(t) \to 0$, so $\boldsymbol{x}(t) \to H(\bar{g})$ from (7). □

Proof of Lemma 2. Note that, because $u_k \to 0$ a.s. by Lemma 1, $\boldsymbol{x}_k$ is far from $X(\ell)$ a.s. for all ℓ. Consider:

$$\begin{aligned} \nabla f(\boldsymbol{x}_k)d(\boldsymbol{x}_k) &= \nabla f(\boldsymbol{x}_k)\big(-L\nabla f(\boldsymbol{x}_k)^T + u_k\mathbf{1} + \boldsymbol{\epsilon}_k\big) \\ &= -\nabla f(\boldsymbol{x}_k)L\nabla f(\boldsymbol{x}_k)^T \\ &\quad + u_k\nabla f(\boldsymbol{x}_k)\mathbf{1} + \nabla f(\boldsymbol{x}_k)\boldsymbol{\epsilon}_k. \end{aligned}$$

For $d(\boldsymbol{x}_k)$ to be a descent direction, we require $\nabla f(\boldsymbol{x}_k)d(\boldsymbol{x}_k) < 0$, which is equivalent to showing

$$\begin{aligned} &\nabla f(\boldsymbol{x}_k)L\nabla f(\boldsymbol{x}_k)^T > u_k\mathbf{1}^T\nabla f(\boldsymbol{x}_k)^T \\ \equiv &\sum_{(i,j)\in\mathcal{E}} \left(\nabla f^i(x_k^i) - \nabla f^j(x_k^j)\right)^2 > u_k\sum_{i=1}^{n}\nabla f^i(x_k^i) \\ &\qquad + \sum_{i=1}^{n}\epsilon_k^i\nabla f^i(x_k^i), \end{aligned} \tag{10}$$

The quadratic representation of the RHS comes from the fact that L is the Laplacian of the communication graph [20].

A sufficient condition to satisfy (10) is

$$\left(\nabla f^i(x_k^i) - \nabla f^j(x_k^j)\right)^2 > \sum_{i=1}^{n}(u_k + \epsilon_k^i)\nabla f^i(x_k^i), \tag{11}$$

for some $(i,j) \in \mathcal{E}$ because the sum of the LHS over $(i,j) \in \mathcal{E}$ is the LHS of (10). Note that this is identical to the result in Proposition 2.

Let $u_k \geq 0$; this is not restrictive because arguments are symmetric for $u_k \leq 0$. Because the LHS of (10) is nonnegative and $u_k + \epsilon_k^i$ is bounded a.s. for all i, it follows that the inequality in (11) holds a.s. for large enough k if $\sup_{i,k} |\nabla f^i(x_k^i)| < \infty$. Therefore, suppose $\sup_{i,k} \nabla f^i(x_k^i) = \infty$. (Symmetric arguments are made if $\inf_{i,k} \nabla f^i(x_k^i) = -\infty$.) This implies $\sup_{i,k} x_k^i = \infty$ because each $f^i(x^i)$ is coercive. Because $u_k \to 0$ a.s. and $\sup_{i,k} x_k^i = \infty$, we have $\inf_{i,k} x_k^i = -\infty$ a.s., which implies $\inf_{i,k} \nabla f^i(x_k^i) < 0$ a.s. because $f^i(x^i)$ is convex with a minimum at 0. Therefore, for all $C > 0$, there exists some κ a.s. such that $\max_i \nabla f^i(x_\kappa^i) > C$ and $\min_i \nabla f^i(x_\kappa^i) < 0$.

We now show via contradiction that there exist $(i,j) \in \mathcal{E}$ such that (11) is satisfied by Proposition 2. Consider $q_\kappa \triangleq \arg\max_i |\nabla f^i(x_\kappa^i)|$. Suppose $\nabla f^{q_\kappa}(x_\kappa^{q_\kappa}) > 0$; this is not restrictive because arguments are symmetric for $\nabla f^{q_\kappa}(x_\kappa^{q_\kappa}) < 0$. By Proposition 2, if there exists some $i \in \mathcal{N}^{q_\kappa}$ such that

$$\nabla f^{q_\kappa}(x_\kappa^{q_\kappa}) > 2\nabla f^i(x_\kappa^i) + n(|u_\kappa| + \bar{\epsilon}),$$

then $d(\boldsymbol{x}_\kappa)$ is a descent direction. Suppose no such $i \in \mathcal{N}^{q_\kappa}$ exists. Then $\nabla f^i(x_\kappa^i) > \frac{1}{4}\nabla f^{q_\kappa}(x_\kappa^{q_\kappa})$ for all $i \in \mathcal{N}^{q_\kappa}$ because $\frac{1}{4}\nabla f^{q_\kappa}(x_\kappa^{q_\kappa}) > n(|u_\kappa| + \bar{\epsilon})$.

Choose any $p_\kappa \in \mathcal{N}^{q_\kappa}$. Once again, by Proposition 2, if there exists some $i \in \mathcal{N}^{p_\kappa}$ such that

$$\nabla f^{p_\kappa}(x_\kappa^{p_\kappa}) > 2\nabla f^i(x_\kappa^i) + n\frac{\nabla f^{q_\kappa}(x_\kappa^{q_\kappa})}{\nabla f^{p_\kappa}(x_\kappa^{p_\kappa})}(|u_\kappa| + \bar{\epsilon}),$$

then $d(\boldsymbol{x}_\kappa)$ is a descent direction. If no such $i \in \mathcal{N}^{p_\kappa}$ exists, then $\nabla f^i(x_k^i) > \frac{1}{4}\nabla f^{p_\kappa}(x_\kappa^{p_\kappa}) > (\frac{1}{4})^2\nabla f^{q_\kappa}(x_\kappa^{q_\kappa})$ for all $i \in \mathcal{N}^{p_\kappa}$.

Continue examining neighbors of each node until an edge is found that satisfies the sufficient condition in Proposition 2. Suppose no such edge exists. Then $\nabla f^i(x_\kappa^i) > (\frac{1}{4})^n\nabla f^{q_\kappa}(x_\kappa^{q_\kappa}) > 0$ for all $i \in \mathcal{V}$ because the maximum diameter of the communication graph is n. This is a contradiction because there exists some $i \in \mathcal{V}$ such that $\nabla f^i(x_\kappa^i) < 0$. Therefore, there exists an edge satisfying the sufficient condition, and $d(\boldsymbol{x}_\kappa)$ is a descent direction a.s. □