Distributed Monitoring of Voltage Collapse Sensitivity Indices

John W. Simpson-Porco, Member, IEEE, and Francesco Bullo, Fellow, IEEE

Abstract—The assessment of voltage stability margins is a promising direction for wide-area monitoring systems. Accurate monitoring architectures for long-term voltage instability are typically centralized and lack scalability, while completely decentralized approaches relying on local measurements tend towards inaccuracy. Here we present distributed linear algorithms for the online computation of voltage collapse sensitivity indices. The computations are collectively performed by processors embedded at each bus in the smart grid, using synchronized phasor measurements and communication of voltage phasors between neighboring buses. Our algorithms provably converge to the proper index values, as would be calculated using centralized information, but do not require any central decision maker for coordination. Modifications of the algorithms to account for generator reactive power limits are discussed. We illustrate the effectiveness of our designs with a case study of the New England 39 bus system.

Index Terms—voltage stability, wide-area monitoring and control, voltage collapse

I. INTRODUCTION

Power grids are transitioning from a paradigm of centralized monitoring and control to one based on decentralized decisions and consumer interaction. When coupled with waning infrastructure investment, rapidly growing urban load centers, and the wide-spread adoption of intermittent distributed generation, this structural shift will lead to a broader and more uncertain range of operating conditions for the grid and an erosion of system stability margins if not properly coordinated.

Simultaneously however, the rise of cheap and ubiquitous sensing, communication, and computational capability suggests a future where the physical grid is strongly coupled to many accompanying layers of data and control. Unlike classic grids where sparsely available information is telemetered to a control center and processed, copious amounts of information are distributed throughout the smart grid along with the computational capabilities to process measured data and make coordinated decisions in real-time for wide-area monitoring, protection, and control (WAMPAC) [1], [2]. The distributed nature of these smart grid resources suggests we explore and evaluate the effectiveness of different information architectures for WAMPAC, ranging the spectrum from centralized to decentralized.

In this article we consider the problem of online monitoring for long-term voltage instability (LTVI) within a smart grid, which has recently been identified as an area of urgent interest for industry [3]. While in general voltage stability is a complex, multi time-scale phenomena, long-term voltage instability is a quasi-static bifurcation instability [4], [5] associated with an inability of the combined generation/transmission system to transmit sufficient power to loads [6]. After an increase in load, or a disturbance such as generation failure or load shedding, the grid’s long-term operating point can vanish leading to the tripping of protection equipment and potentially to a large-scale voltage collapse blackout [6]. Robustness margins against LTVI are quantified via voltage collapse proximity indicators (VCPIs), which produce measures of distance to instability [7]. Accurate indicator estimates are key for distinguishing vulnerable system conditions from stable conditions exhibiting low voltages [8], and we now review various architectures for the calculation of such indicators.

A. Monitoring of Long-Term Voltage Instability

Many monitoring solutions for LTVI have been proposed, and can be broadly classified by architecture (centralized, decentralized, distributed), measurement rate/complexity (timeskewed SCADA vs. time-stamped phasor measurement unit data), and theoretical rigor (heuristic vs. exact) [8]. The most important distinction for our purposes is the first, which we now expand on.

Centralized Monitoring: In a centralized monitoring approach, relevant data is telemetered to a central computer in a control center and potentially combined with state estimation to calculate relevant indices; see [8]–[11] and the many references therein. From the perspective of this paper, the main drawback of a centralized approach is that it results in a single point of failure for the monitoring system, and potentially requires data to be sent over large distances. Data privacy issues may also come into consideration. Moreover, in emerging applications such as microgrids, centralized supervision may be untenable or prohibitive. In this case, a more modular, scalable monitoring approach is desirable.

Decentralized Monitoring:* In contrast to centralized monitoring, decentralized techniques rely only on locally measured information to estimate voltage stability margins. Monitoring techniques based on PMU data and/or Thévenin equivalent circuits were proposed in [12]–[16], among others. Decentralized VCPIs offer low implementation complexity and

*We use the term decentralized here for what is sometimes called “completely decentralized” or “completely distributed” — the VCPI calculated at bus \( i \) will depend only on information measured locally at bus \( i \), such as phasor voltage \( V_i + \theta_i \) and complex power injection \( P_i + jQ_i \).
easy scalability, with the additional advantage that (typically) no communication or state estimation is required. The price paid for these advantages is accuracy: decentralized VCPIs are invariably heuristic, often inspired by single-line power flow results, and are always too optimistic since they do not explicitly account for coupling between buses [17]. A notable exception for decentralized approaches is [17], where sensitivities of voltages with respect to on-load tap changer ratios are used to monitor the system for instability.

A monitoring approach for LTVI was presented in [18], where the grid is partitioned into overlapping monitoring areas (akin to control areas in automatic generation control). The indicators used however are somewhat unconventional, and the approach relies on selecting monitoring areas which are only weakly coupled. Neglecting this inter-area coupling then results in a centralized assessment problem within each area. While being termed “distributed”, in our terminology the approach in [18] is a hybrid of centralized and decentralized ideas. In Remark 3 we comment on the extension of our results to similar multi-area monitoring architecture.

Distributed Monitoring: In the intermediate between centralized and decentralized we arrive at distributed monitoring strategies, which rely on local measurements along with communicated data from directly adjacent buses or areas of the grid. Distributed strategies promise to combine the performance of centralized monitoring with the scalability of decentralized monitoring, requiring only sparse, localized communication without any global information. Reflecting this, the recent literature has witnessed a steady growth in the applications of distributed algorithms to power system monitoring and control, now including resource allocation [19], load shedding [20], economic dispatch [21], optimal power flow [22], voltage control [23], transfer capability assessment [24], and inverter coordination [25] in microgrids. To the authors knowledge however, distributed algorithms have not yet been designed for the monitoring of VCPIs. In particular, we focus on a subclass of VCPIs termed sensitivity indices, which quantify the sensitivity of grid states to changes in grid parameters. These sensitivities increase as voltage collapse is approached, and monitoring these sensitivities therefore provides information on the proximity to collapse.

B. Contributions

In this work we present the first distributed algorithms for the online computation of voltage collapse sensitivity indices. We demonstrate that the exact calculation of several standard centralized indices can be distributed among agents embedded within the smart grid, achieving centralized performance through only local measurements and short-range nearest-neighbor communication. The exact nature of these agents is left unconstrained; the software could be embedded in next generation power inverters, power electronic devices for voltage control, or implemented at generators or smart meters. Our algorithms do not rely on a state estimator processing sparse measurements to estimate state variables, but rather combine direct local measurements with measurements made and communicated by neighboring buses. Data is transmitted only over short distances, minimizing communication problems such as packet delays and measurement problems such as time-stamp drift. No centralized decision maker is required. Our approach does not rely on any pre-defined interfaces, on any representative sets of offline data used for learning, or on any Thévenin equivalent representations. After algorithm convergence, each bus recovers its exact sensitivity index along with the indices of neighboring buses. We demonstrate the efficacy of our algorithms via simulation in Section IV on the IEEE 39 bus system.

We assume that each bus in the system is equipped with a phasor measurement unit. While current power systems are not equipped with this level of observability, the smart grid eventually may, and demonstrations of the operational benefits of observability (such as those presented herein) will serve as incentive to invest in such measurement capabilities in the future. It seems plausible that our assumption of full observability can be relaxed, and that the approach can be extended to more detailed models of long-term voltage instability, but we defer further discussion of this to our concluding remarks in Section V. At the transmission level, centralized, state estimation-based voltage stability margins will continue to play a major role, but it is nonetheless important to assess the advantages and limitations of alternative methods. Our main message is that complicated sensitivity indices can in fact be computed using only localized information, without the need for centralized coordination or computation. An area where our algorithms may prove particularly useful is microgrids, where centralized monitoring, control and optimization architectures are often absent and must be implemented collectively by coordinating devices within the microgrid in a scalable way.

C. Preliminaries and Notation

Sets, vectors and functions: We let $\mathbb{R}$ (resp. $\mathbb{R}_{\geq 0}$) denote the set of real (resp. strictly positive real) numbers. Given $x \in \mathbb{R}^n$, $\|x\|_\infty = \max_{i \in \{1, \ldots, n\}} |x_i|$, and $[x] \in \mathbb{R}^{n \times n}$ is the associated diagonal matrix with $x$ on the diagonal. Throughout, $1_n$ and $0_n$ are the $n$-dimensional vectors of unit and zero entries, and $0$ is a matrix of all zeros of appropriate dimensions. The $n \times n$ identity matrix is $I_n$.

II. System Models and Sensitivity-Based Voltage Collapse Proximity Indicators

We begin by defining the grid models to be used in the paper before reviewing the relevant voltage collapse indices.

A. Power System Model

We model a balanced, quasi-synchronous power grid as a connected, undirected and weighted graph $(V, E)$, where $V$ is the set of nodes (buses) and $E \subseteq V \times V$ is the set of edges (branches). We partition the set of buses $V$ as $V = L \cup G$, with $n \geq 1$ load (PQ) buses $L = \{1, \ldots, n\}$ and $m \geq 1$ generator (PV) buses $G = \{n + 1, \ldots, n + m\}$. Each branch $\{i, j\} \in E$†

†For our purposes, PV generator buses $G$ may represent synchronous generators, frequency-dependent loads, or grid-forming inverters implementing droop controllers [21], [26]. Similarly, load buses $L$ may represent standard static loads as well as inverters performing maximum power point tracking.
is weighted by a transfer admittance $y_{ij} = g_{ij} + j b_{ij}$, where $g_{ij} \geq 0$ and $b_{ij} \leq 0$. We encode the weights and topology in the bus admittance matrix $Y$, with elements $Y_{ij} = -y_{ij}$ and $Y_{ii} = -\sum_{j=1}^{n+m} y_{ij} + y_{shunt,i}$, where $y_{shunt,i}$ is the shunt element at bus $i$. The conductance matrix $G$ and susceptance matrix $B$ are defined by $G = \text{Re}(Y)$ and $B = \text{Im}(Y)$. To each bus we associate a phasor voltage $V_i \angle \theta_i$ and a complex power injection $P_i + j Q_i$, which are related by the power flow equations

$$P_i = \sum_{j \in V} V_j G_{ij} \cos(\theta_i - \theta_j) + B_{ij} \sin(\theta_i - \theta_j), \quad (1a)$$

$$Q_i = \sum_{j \in V} V_j G_{ij} \sin(\theta_i - \theta_j) - B_{ij} \cos(\theta_i - \theta_j), \quad (1b)$$

The unknowns in (1a)–(1b) are the phase angles $\theta = (\theta_1, \ldots, \theta_{n+m})$ and the load voltages $V_L = (V_1, \ldots, V_n)$. With shunt admittances absorbed into the admittance matrix, we assume that the remainder of the load at each PQ bus can be described by a constant power load model. Extending our approach to more general voltage-dependent static load models $P_i(V_i)$ and $Q_i(V_i)$ is straightforward, and requires only a few additional terms in the formulae and algorithms which follow; we omit these extensions for notational simplicity. Further comments on grid modeling are deferred to Section V.

B. Cyber Layer Model

We assume that devices (agents) are embedded at each bus which can measure local information, communicate information with devices at nearby buses, and perform basic computations on the measured data.

Regarding measurement, we assume that each bus $i \in V$ is equipped with a PMU, yielding accurate synchronized measurements of voltage phasors $V_i \angle \theta_i$. In addition we assume that the processor at each bus has knowledge of (or access to measurements of) power system infrastructure incident to the bus, such as local power consumption/generation $(P_i, Q_i)$, the admittances of incident electrical lines $(y_{ij})$, and local shunt elements $(y_{shunt,i})$. Admittances may be know a priori, or estimated in an initialization phase using ranging technologies over power line communication (PLC) channels. Similarly, shunt susceptances designed to support voltage magnitudes are often either fixed or switched by local controllers which could be integrated into the processing equipment under consideration. In contrast to SCADA systems which sample data every few seconds, PMUs under the IEEE Standard C37.118 [27], [28] are synchronized and able to take 10’s of samples per second. On the time-scales of interest for long-term voltage stability, this fast sampling is well approximated as continuous-time measurement. Throughout we assume high-quality measurements and do not distinguish between measured and true values: if PMU measurement quality is determined to be an issue, a distributed state-estimation and filtering layer [29], [30] can be implemented between the raw measurements and our algorithms to improve signal-to-noise ratios and reject bad data.

Regarding communication, we assume that each agent can communicate bidirectionally with the agents at adjacent buses to which it is electrically connected. Said differently, the topology of the communication layer mimics the physical grid topology. This communication could be achieved through power line communication (PLC), limited-range wireless, or Ethernet. We emphasize that here our focus is not on detailed communication protocols, but on highlighting the sufficiency of local information exchange for the exact calculation of standard sensitivity indicators. To streamline our mathematical developments throughout, we will therefore assume generous communication capabilities which in effect permit continuous-time communication. Due to the large separation of time-scales between PMU sampling rates and LTVI, and the fact that our algorithms require only short-distance communications, throughout we assume that delays are negligible. Nonetheless, in Remark 2 we comment on theoretical extensions to less restrictive communication assumptions.

C. Sensitivity-Based Voltage Collapse Proximity Indicators

As long-term voltage instability and voltage collapse is associated with saddle-node bifurcation of the power flow equations, singularity of the power flow Jacobian or related matrices has long been used as an indicator of voltage collapse [31]. Related approaches include modal analysis, singular value and condition number indicators, sensitivity indices, continuation methods, optimization, and energy VCPIs. Surveys, classifications, and comparative studies of various VCPIs are available in [6]–[8], [32]–[39].

Here we focus on one of the oldest classes of VCPIs, the sensitivity indices, which are based on the sensitivity of the system operating point to variations in parameters. The idea is that small variations in, for example, load demands, will produce large variations in bus voltages near bifurcation [32], [40], [41]. While many sensitivity-based VCPIs have been superseded in practical power system operations by more accurate, more computationally intensive techniques, they nonetheless provide intuitive actionable information, and are relatively straightforward to define and interpret. We recall three basic indices [42, Sec. 8.2.3] and then comment on the information needed to compute them.

a) The “$dV/dQ$” Index: This index measures the sensitivity of load bus voltage magnitudes with respect to changes in reactive power demands. For a multi-bus network, we may formulate the appropriate indicator $I_i$ as

$$I_i \triangleq \sum_{j \in L} \frac{Q_j}{V_i} \frac{\delta V_i}{\delta Q_j}, \quad i \in L.$$  

(2)

The summands are the point elasticities of the voltage at load bus $i$ with respect to the reactive demand at load bus $j$. The sum then evaluates the total elasticity of the voltage at bus $i \in L$. The index ranges from 0 at open-circuit conditions to $+\infty$ when the system reaches the point of collapse.

b) The “$dV_i/dE$” Index: Also called the “$dV_i/dE$” index, this index measures the sensitivity of load bus voltages to changes in generator network voltage set points. The appropriate multi-bus index $J_i$ is

$$J_i \triangleq \sum_{k \in G} \frac{\delta V_i}{\delta V_k}, \quad i \in L.$$  

(3)
Near open-circuit conditions \( J \), should be near unity, indicating that changes in load voltages track changes in generator voltages with unity gain (a “controllability” property [32]). The index tends to \(+\infty\) at the point of collapse.

c) The \( \frac{dQ_G}{dQ_L} \) Index: This index measures the incremental reactive power generation required to supply an incremental amount of additional load, and therefore quantifies the (inverse) efficiency of reactive power transport through the network. The appropriate multi-bus definition is

\[
K_i \triangleq \sum_{k \in G} \frac{\delta Q_k}{\delta Q_i}, \quad i \in \mathcal{L}.
\] (4)

With our sign conventions, the \( K_i \) ranges from \(-1\) at open-circuit to \(-\infty\) at bifurcation, indicating that the network transports reactive power inefficiently near voltage collapse.

The important observations regarding the indices (2)–(4) are (i) that the matrices of derivatives defining them are generally dense matrices, and (ii) that the matrix elements take into account the global state of the network. For example, for a processor at bus \( i \in \mathcal{L} \) to directly compute \( K_i \), it would need to not only be directly aware of all generators connected to the network, but also know numerically how the set point \( V_k \) of each influences the local voltage \( V_i \). This sensitivity is in turn influenced by the presence (or absence) of loading/compensation at all other buses. Ostensibly then, (2)–(4) incorporate non-local information, and it would appear then that only an operator with centralized or near-centralized state information can evaluate them.

III. DISTRIBUTED COMPUTATION OF SENSITIVITY-BASED VCPIs

We now detail our approach for distributing the computation of the VCPIs presented in Section II-C. We present our approach pedagogically for the \( dV/dQ \) index (2) before formally defining our distributed protocols for all indices (2)–(4) and addressing protocol convergence. To begin, note that in vector notation the dV/dQ index (2) becomes

\[
I = [V_L]^{-1} \frac{\delta V_L}{\delta Q_L} Q_L,
\]

where \( I = (I_1, \ldots, I_n) \), \( Q_L = (Q_1, \ldots, Q_n) \), \( [V_L] \) is the diagonal matrix of load bus voltages, and \( \delta V_L/\delta Q_L \) is the matrix with elements \( \delta V_i/\delta Q_j \), \( i, j \in \mathcal{L} \). Away from the point of collapse the matrix \( \delta V_L/\delta Q_L \) is invertible, and we may equivalently write

\[
\frac{\delta Q_L}{\delta V_L} [V_L] I = Q_L,
\] (5)

which is a (dense) system of equations. Returning briefly to the power flow (1a)–(1b), around an operating point \((\theta, V_L) \in \mathbb{R}^{n+m} \times \mathbb{R}^n\) incremental changes \( (\delta \theta, \delta V_L, \delta V_G) \) in phase angles, load voltages, and generator voltages are related to incremental changes \( (\delta P, \delta Q_L, \delta Q_G) \) in active power injections and reactive power injections (load and generator) by

\[
\begin{pmatrix}
\frac{\delta P}{\delta \theta} \\
\frac{\delta Q_L}{\delta \theta} \\
\frac{\delta Q_G}{\delta \theta}
\end{pmatrix}
= 

\begin{pmatrix}
\frac{\partial P}{\partial \theta} & \frac{\partial P}{\partial V_L} & \frac{\partial P}{\partial V_G} \\
\frac{\partial Q_L}{\partial \theta} & \frac{\partial Q_L}{\partial V_L} & \frac{\partial Q_L}{\partial V_G} \\
\frac{\partial Q_G}{\partial \theta} & \frac{\partial Q_G}{\partial V_L} & \frac{\partial Q_G}{\partial V_G}
\end{pmatrix}
\begin{pmatrix}
\delta \theta \\
\delta V_L \\
\delta V_G
\end{pmatrix},
\] (6)

where all partial derivatives are evaluated at the operating point. If variations in active power injections \( \delta P \) and generator voltages \( \delta V_G \) are held at zero, the first two block-rows of equations in (6) may be solved to yield

\[
\frac{\delta Q_L}{\delta V_L} = \frac{\partial Q_L}{\partial V_L} - \frac{\partial Q_L}{\partial \theta} \left( \frac{\partial P}{\partial \theta} \right)^\dagger \frac{\partial P}{\partial V_L},
\] (7)

where \( \dagger \) denotes the Moore-Penrose pseudoinverse (see Remark 1). The system of equations (5) for \( I \) then becomes

\[
\frac{\partial Q_L}{\partial V_L} [V_L] I = \frac{\partial Q_L}{\partial \theta} \left( \frac{\partial P}{\partial \theta} \right)^\dagger \frac{\partial P}{\partial V_L} [V_L] = Q_L.
\]

Introducing an auxiliary variable \( I_{aux} \in \mathbb{R}^{n+m} \), this dense system of equations is equivalent to the expanded system

\[
\begin{pmatrix}
\frac{\partial Q_L}{\partial V_L} [V_L] \\
\frac{\partial P}{\partial \theta}
\end{pmatrix}
\begin{pmatrix}
I_{aux} \\
Q_L
\end{pmatrix}
= 

\begin{pmatrix}
\frac{\partial Q_L}{\partial \theta} \left( \frac{\partial P}{\partial \theta} \right)^\dagger \frac{\partial P}{\partial V_L} [V_L] \\
\frac{\partial P}{\partial \theta}
\end{pmatrix}
\begin{pmatrix}
Q_L \\
0_{n+m}
\end{pmatrix},
\] (8)

The coefficient matrix in (8) is sparse, its sparsity pattern closely related the physical grid topology. Indeed, the sparsity of such matrices has long been used as an aid for fast computation of stability margins [43]. While sparsity of the unreduced Jacobian (6) could also be exploited for computing the desired indices, (6) will typically contain unnecessary information, which for our algorithms would lead to unnecessary communication and computation. For example, the third block-row in (6) contains unnecessary information for the \( dV/dQ \) index. We therefore find the reduced Jacobian-like matrix in (8) more useful to work with. To propose the simplest, most intuitive distributed algorithm for calculating the stability index \( I \), we make the following assumptions.

Assumption 1 (System Matrix Stability): All eigenvalues of the matrices with positive real parts, with the exception of a simple zero eigenvalue for each with respective right eigenvectors \( 1_{n+m} \), \( (1_{n+m}, 0_n) \), and \( (0_{n+m}, 1_n) \).

Remark 1 (Comments on Assumption 1): The simple zero eigenvalues of the matrices in Assumption 1 correspond to a uniform shift \( \delta \theta \rightarrow \delta \theta + \alpha 1_{n+m} \) of all phase angle deviations \( \delta \theta \). Since phase is defined only up to a reference, this trivial degree of freedom may be removed by restricting \( I_{aux} \) to lie in the subspace orthogonal to \( 1_{n+m} \), in which case all three matrices are effectively invertible. In practice Assumption 1 holds away from the point of collapse [34], and the non-zero eigenvalues of these matrices are often found to be real or have small imaginary parts [44, Appendix B.3]. Moreover, Assumption 1 is quite natural since (1) the matrices under consideration describe stable small-signal behavior for certain classes of power system dynamics, and (2) these dynamics are known to not exhibit Hopf bifurcations, and hence the respective matrices can only become singular when an eigenvalue reaches the origin during saddle-node bifurcation [5]. In this sense then, Assumption 1 is “necessary and sufficient” for the linear system (8) to be well-posed.
Our key observation is that the matrix elements in (8) are determined by localized information: the $ij$th element depends only on the voltage phasors at buses $i$ and $j$ and on the admittances of the adjoining branches. It follows that by using phasor measurements and communication among adjacent buses, the solution of (8) for $(\mathbb{I}, \ldots)$ can be distributed among processors embedded at each bus. With this goal in mind, to each load bus $i \in \mathcal{L}$ we associates a pair of scalar states $(x_i, y_i) \in \mathbb{R}^2$, while to each generator bus $i \in \mathcal{G}$ we associate a scalar state $y_i \in \mathbb{R}$. We assume that these states can be communicated bidirectionally between directly adjacent buses. Our first formal result gives a simple distributed algorithm in continuous-time such that $\lim_{t \to \infty} x_i(t) \to \bar{x}_i$ for each $i \in \mathcal{L}$. For notational convenience we define the data coefficients

$$
d_{ij} \triangleq V_i V_j (G_{ij} \sin(\theta_i - \theta_j) - B_{ij} \cos(\theta_i - \theta_j)), \quad (9a)
$$

$$
D_{ij} \triangleq V_i V_j (G_{ij} \cos(\theta_i - \theta_j) + B_{ij} \sin(\theta_i - \theta_j)), \quad (9b)
$$

for each $i, j \in \mathcal{V}$, which depend only on known constants, locally measured PMU data, and PMU data communicated between adjacent buses.

**Theorem 3.1 (Distributed dV/dQ Index):** Consider the $dV/dQ$ indices $I_{ij}$ defined in (2) and let $d_{ij}$ and $D_{ij}$ be as in (9). Let each load bus $i \in \mathcal{L}$ execute

$$
\tau \dot{x}_i = Q_i (1 - x_i) - P_i y_i - \sum_{j \in \mathcal{L}} d_{ij} x_j + \sum_{j \in \mathcal{V}} D_{ij} y_j, \quad (10a)
$$

$$
\tau \dot{y}_i = Q_i y_i - P_i x_i - \sum_{j \in \mathcal{L}} D_{ij} x_j - \sum_{j \in \mathcal{V}} d_{ij} y_j, \quad (10b)
$$

for some chosen $\tau > 0$, while each generator bus $i \in \mathcal{G}$ executes

$$
\tau \dot{y}_i = Q_i y_i - \sum_{j \in \mathcal{L}} D_{ij} x_j - \sum_{j \in \mathcal{V}} d_{ij} y_j. \quad (11)
$$

Then for any initial condition $(x(0), y(0)) \in \mathbb{R}^n \times \mathbb{R}^{n+m}$ it holds for each $i \in \mathcal{L}$ that $\lim_{t \to \infty} x_i(t) = \bar{x}_i$.

**Proof:** Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_{n+m})$ be the state vectors associated with (10)–(11). Comparing the right-hand sides of (10)–(11) with the power flow Jacobian matrix elements in Lemma A.1, one finds that in vector notation (10)–(11) reads as

$$
\tau \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = - \begin{pmatrix} \frac{\partial Q_{ij}}{\partial \theta_i} & \frac{\partial Q_{ij}}{\partial \theta_j} \\ \frac{\partial P_{ij}}{\partial \theta_i} & \frac{\partial P_{ij}}{\partial \theta_j} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} Q_L \\ 0_{n+m} \end{pmatrix}. \quad (12)
$$

Conversely, (10)–(11) are obtained by writing out (12) in components, using Lemma A.1 and the definitions of $d_{ij}$ and $D_{ij}$ in (9). Comparing the dynamics (12) to the algebraic equation (8), it follows that the steady-states of (10)–(11) are one-to-one correspondence with the solutions $(\mathbb{I}, \ldots)$ of (8). The system matrix in (12) a permutation of the third matrix in Assumption 1, and is therefore Hurwitz except for a simple eigenvalue at zero with associated right eigenvector

$$
u_1 = (0_n, 1_{n+m}).$$

The component of the state which evolves parallel to $\nu_1$ only influences the (average value of the) auxiliary variable $y(t)$, and does not influence the index estimates $x(t)$. Since all other eigenvectors are associated with negative eigenvalues, it follows that $\lim_{t \to \infty} x(t) = I$, which completes the proof.

The monitoring architecture is depicted in Figure 1 for a simple power system. While the sums in (10)–(11) run over all loads or all buses, the coefficients $d_{ij}$ and $D_{ij}$ are zero when $\{i, j\}$ is not a physical branch of the network, and hence the only information needed at processor $i \in \mathcal{V}$ is that from adjacent buses. Said differently, the proposed monitoring architecture requires only peer-to-peer communication, without centralized coordination. Uniformity of the time-constant $\tau$ across all buses is formally required to infer stability, but as our case study in Section IV will demonstrate, nonuniform time-constants $\tau_i$ pose no difficulties when implemented.

We make five observations regarding the above algorithm. First, note that the storage and computational requirements for implementation are extremely low. Each agent stores only the local states $(x_i(t), y_i(t))$ or $(y_i(t))$, and integrates an ordinary differential equation; storing the time-history of states is not required, nor is it required that each agent maintains an estimate of the entire algorithm state. Second, the method relies on bus measurements, and no measurements of branch currents are required. Third, communication is required only between neighboring buses in the network, minimizing the effects of any communication delays. Fourth, the time variable $t$ in the algorithm should be interpreted as a computational time-scale; the time constant $\tau$ can be adjusted to achieve any desired convergence speed, limited ultimately by communication time scales, measurement sampling time, and system dynamics, but not by the algorithm itself. Fifth and finally, our method does not rely on a linearized power flow model; the linearity of (10)–(11) comes from examining sensitivities of the nonlinear power flow (1a)–(1b), with real-time measurements replacing a nonlinear power flow solver.

**Remark 2 (Relaxed Algorithms & Communications):** As one can see from (12), the distributed algorithm (10)–(11) is of the form $\tau \dot{v} = -Av + b$, and hence can be explic-
ity discretized for distributed synchronous implementation as 
\[ v(k+1) = (I_{2n+m} - hA/\tau)v(k) + hh/\tau \] for a time-step \( h > 0 \). Under Assumption 1, this discrete-time system is stable if and only if \( h < 2\tau \max_i \frac{\beta\sigma_i}{|\lambda_i|} \). Another option for distributed implementation is Jacobi iteration, where \( A \) is decomposed into its diagonal part \( T = \text{diag}(\alpha_{ii}) \) and off diagonal part \( R = A - T \), with the iteration taking the form 
\[ v(k+1) = T^{-1}(b - Ru(k)) \], This iteration converges if and only if \( \rho(T^{-1}R) < 1 \); the authors have observed numerically that this assumption often holds for the relevant matrices in Assumption 1, but in general this assumption and Assumption 1 are not equivalent. The easily verified diagonal dominance conditions for stability of the Jacobi iteration do not hold.

Packets from neighbors may arrive asynchronously or with delays, and communication may be event-triggered based on sufficient changes in local measurements. While our focus is not on detailed communication protocols, we note a particular approach which is more complex but less restrictive. In [45] a discrete-time algorithm was developed for the distributed solution of linear equations such as (8). In the proposed approach, each bus is assumed to know its respective row of the coefficient matrix, and updates a local estimate of the entire system state \((x, y)\) by exchanging estimates with neighbors. Thus, the requirements on information and communication are similar to the ones required by Theorem 3.1, with slightly more local storage requirements. The approach extends to handle both asynchronism and delays [46], [47]. When Assumption 1 fails during extreme system conditions, the algorithm (10)–(11) will diverge, and one of these alternatives would be required to calculate the relevant indices (which may even change sign under such conditions).

**Remark 3 (Extension to Multi-Area Monitoring):** While we have presented the algorithm (10)–(11) with a direct peer-to-peer implementation, it is easily extended to the case of multi-area monitoring. To see this, partition the the buses \( V \) of the network into \( p \leq n+m \) non-overlapping monitoring areas \( V = A_1 \cup \cdots \cup A_p \). Depending on the specific problem setup, these areas could correspond to ISO regions, substations, phasor data concentrators (PDCs), or microgrids. Inside area \( A_k \), assume that a central processor \( P_k \) has (i) access to PMU measurements from each bus in area \( A_k \), (ii) knowledge of the grid topology and parameters in area \( A_k \), (iii) knowledge of the power lines which connect \( A_k \) to neighboring areas, and (iv) the ability to perform basic computations and communicate data with the processors in neighboring areas. In this case, the algorithm (12) would simply be block-partitioned according to the different areas, with central processors implementing the required blocks. The case of one area \( p = 1 \) would correspond to complete centralized monitoring, where a central processor aggregates all information and performs all computations, while \( p = n+m \) is the case described in main paper, where each bus (e.g., substation) constitutes an area, only local measurements are required, and information exchange is peer-to-peer. Depending on regional data disclosure policies and privacy concerns, one architecture may be preferable over another; these issues are outside the scope of this work.

Similar filters to (10)–(11) can be designed to calculate the \( dV_L/dV_G \) index (3) and the \( dQ_G/dQ_L \) index (4); the proofs may be found in Appendix A.

**Theorem 3.2 (Distributed \( \text{d}V_L/\text{d}V_G \) Index):** Consider the \( dV_L/dV_G \) indices \( J_i \) defined in (3) and let \( d_{ij} \) and \( D_{ij} \) be as in (9). Let each load bus \( i \in \mathcal{L} \) execute
\[ \tau \dot{x}_i = -\frac{Q_i}{V_i} x_i - P_i y_i - \sum_{j \in \mathcal{L}} d_{ij} x_j + \sum_{j \in \mathcal{V}} D_{ij} y_j - \sum_{j \in \mathcal{G}} D_{ij} y_j, \]
\[ \tau \dot{y}_i = Q_i y_i - P_i x_i - \sum_{j \in \mathcal{L}} D_{ij} x_j - \sum_{j \in \mathcal{V}} d_{ij} y_j - \sum_{j \in \mathcal{G}} D_{ij} y_j, \]
for some chosen \( \tau > 0 \), while each generator bus \( i \in \mathcal{G} \) executes
\[ \tau \dot{y}_i = -\frac{P_i}{V_i} + Q_i y_i - \sum_{j \in \mathcal{L}} D_{ij} x_j - \sum_{j \in \mathcal{V}} d_{ij} y_j - \sum_{j \in \mathcal{G}} D_{ij} y_j. \]
Then for any initial condition \((x(0), y(0)) \in \mathbb{R}^n \times \mathbb{R}^{n+m}\) it holds for each \( i \in \mathcal{L} \) that \( \lim_{t \to \infty} x_i(t) = J_i \).

**Theorem 3.3 (Distributed \( dQ_G/dQ_L \) Index):** Consider the \( dQ_G/dQ_L \) indices \( K_i \) defined in (4) and let \( d_{ij} \) and \( D_{ij} \) be as in (9). Let each load bus \( i \in \mathcal{L} \) execute
\[ \tau \dot{x}_i = -\frac{Q_i}{V_i} x_i - P_i y_i - \sum_{j \in \mathcal{L}} d_{ij} x_j - \sum_{j \in \mathcal{V}} D_{ij} y_j + \sum_{j \in \mathcal{G}} D_{ij} y_j, \]
\[ \tau \dot{y}_i = Q_i y_i - P_i x_i + \sum_{j \in \mathcal{L}} D_{ij} x_j - \sum_{j \in \mathcal{V}} d_{ij} y_j, \]
\[ \tau \dot{z}_i = -P_i + Q_i z_i - \sum_{j \in \mathcal{V}} d_{ij} z_j + \sum_{j \in \mathcal{G}} D_{ij} z_j, \]
for some chosen \( \tau > 0 \), while each generator bus \( i \in \mathcal{G} \) executes
\[ \tau \dot{y}_i = Q_i y_i + \sum_{j \in \mathcal{L}} D_{ij} x_j - \sum_{j \in \mathcal{V}} d_{ij} y_j, \]
\[ \tau \dot{z}_i = -P_i + Q_i z_i - \sum_{j \in \mathcal{V}} d_{ij} z_j + \sum_{j \in \mathcal{G}} D_{ij} z_j. \]
Then for any initial condition \((x(0), y(0), z(0)) \in \mathbb{R}^n \times \mathbb{R}^{n+m} \times \mathbb{R}^{n+m}\) it holds for each \( i \in \mathcal{L} \) that \( \lim_{t \to \infty} x_i(t) = K_i \).

**A. Incorporating Generator VAR Limits**

The distributed algorithms presented in Theorems 3.1–3.3 ignore an important factor in LTVI studies, namely the reactive power limitations of generators [48]–[50]. When a synchronous generator exceeds these reactive power limits (derived from field and armature current limits) over medium time-scales, the AVR system becomes unable to regulate the network-side generator voltage and over-excitation limiters fix the reactive power output at its limit. On the long-time scales of interest for us, we can therefore approximate this behavior by replacing the PV bus model with a PQ bus model when the generator is at or above its reactive limit [44].

The approach for incorporating these limits into the distributed algorithm (10)–(11) of Theorem 3.1 is as follows (similar approaches hold for the remaining two algorithms). If the reactive power supplied by generator \( i \in \mathcal{G} \) satisfies...
If

\( Q_i < Q_i^{\text{max}} \), then the associated processor executes (11), just as before. When \( Q_i \geq Q_i^{\text{max}} \), the processor initializes an additional internal state \( x_i(t) \) and instead executes (10a)–(10b). If necessary, this can be accompanied with a binary alert message to its neighbors signaling that a switch has taken place. To avoid chattering due to oscillating reactive power injections during transients, a temporal hysteresis can be used which ensures that \( Q_i \) remains above or below \( Q_i^{\text{max}} \) for a sufficient amount of time before a switch in algorithm is made.

**B. Monitoring Thresholds and Worst-Case Indices**

The algorithms in Theorems 3.1–3.3 give the processor at load bus \( i \in \mathcal{L} \) a converging estimate of its stability index \( I_i, J_i \) or \( \mathbb{R}_i \), as well converging estimates of the same indices for any adjacent buses which are also load buses. Based on this information, we highlight two additional steps for monitoring that may be desirable. We discuss the \( dV/dQ \) algorithm (10)–(11); similar statements apply to the other algorithms.

**Monitoring Thresholds:** Suppose that to each load bus \( i \in \mathcal{L} \) we associate a threshold \( \gamma_i > 0 \) for the index \( I_i \). These thresholds may be determined by experience, offline trials, or determined online by yet another distributed algorithm. If during monitoring \( x_i(t) \) increases above \( \gamma_i \) and remains there, an alert is triggered and communicated to neighboring processors. This in turn could trigger localized control responses, or the alert could be propagated system-wide.

**Global Knowledge of Worst-Case Index:** Voltage stability of the network is ultimately limited by the weakest or most sensitive bus, quantified in our setup by the largest nodal value \( \|I\|_{\infty} = \max_{i \in \mathcal{L}} \|I_i \|_2 \) of the stability index. It may therefore be desirable for all buses to maintain an estimate \( w_i(t) \) of the worst-case index \( \|I\|_{\infty} \) and continuously update it. A simple distributed protocol for achieving this is called \( \text{max-consensus} \) [51], [52] where each processor executes (in discrete-time)

\[
    w_i(k+1) = \max \left\{ w_i(k), \max_{j \in \mathcal{E}} w_j(k) \right\},
\]

with the initialization \( w_i(t_0) = x_i(t_0) \) for \( i \in \mathcal{L} \), where \( t_0 \) is the time at which execution begins. As generator buses \( i \in \mathcal{G} \) do not carry a local state \( x_i \), each \( w_i(t_0) \) is initialized to a common value \( w^* \) for each \( i \in \mathcal{G} \), equaling the open-circuit value of the voltage stability index under consideration. For example, for the \( dV/dQ \) index \( w^* = 0 \), while for the \( dV_L/dV_G \) index \( w^* = 1 \). Each processor observes its own index and the indices of its neighbors and updates its estimate with the largest value it sees. By re-initializing and re-executing this periodically, all processors in the network can be made aware of the largest sensitivity.

**IV. Case Study: IEEE 39 Bus System**

We demonstrate our approach by implementing our algorithm for the \( dV_L/dV_G \) index of Theorem 3.2 on a dynamic model of the reduced New England power grid, containing 9 generators and 30 load buses. A six-state two-axis model is used for the generators consisting of two-state mechanical dynamics, two-state electrical dynamics, a single-state excitation system and a single-state governor with droop [53]; generator and network parameters are drawn from [53]–[55].

In place of a uniform filter time constant \( \tau \), we let each processor implement its \( dV_L/dV_G \) filter with a time constant \( \tau_i \), which we draw from a uniform distribution between 10s and 20s. Synchrophasor measurements are assumed to be corrupted with uncorrelated zero mean Gaussian noise, with standard deviation 0.001 p.u. on voltage magnitudes (arising from quantization and harmonic distortion), and 0.01° on phase angles (due to sampling time discrepancies and inexact synchronization). At a 2σ level, these values are in compliance with the maximum total phasor error of 1% specified by IEEE Standard C37.118-2011 [28]. Beginning from the base load case [55], power demands are ramped along the base case by 15% between \( t = 20s \) and \( t = 40s \), with the newly ramped load being shed abruptly at \( t = 200s \).

The filter states \( x_i(t) \) are plotted in Figure 2 for load buses 3, 12, and 20, and the generator bus 34 whose state is initialized as \( x_{34}(0) = 1 \). The exact steady-state values of the respective stability indices are plotted in dashed black for each bus, as computed by a central processor solving the linear equation (16) at each moment in time. First, we observe that the algorithm is able to accurately track the ramp in load between 20s and 40s. The estimates for buses 3, 12, and 20 have effectively converged to their proper values shortly before \( t = 50s \), but the increase in load has caused the generator at bus 34 to hit its reactive power limits. The respective estimator \( x_{34}(t) \) comes online at roughly \( t = 54s \) and converges rapidly, contributing to a further increase in the index estimates \( x_i(t) \) of all other buses, and in particular at bus 20 which is directly adjacent to bus 34. At \( t = 200s \) the excess load is shed and filter estimates converge back to their original values; the centralized computation displays significant ringing due to transient dynamics, while the filter state converges relatively smoothly due to its natural first-order dynamics, which act as a low-pass filter. The generator falls back below its reactive power limits, and after an anti-chattering delay the estimator for bus 34 is reset.

Figure 3 displays a close-up of the trace of \( x_3(t) \) at bus 20 after the load is shed at \( t = 200s \), along with the resulting dynamics of the corresponding bus voltage \( V_{20}(t) \). As our algorithms use real-time measurements for computing the sensitivity indices, transients experienced by the physical bus variables also impact the filter estimates until convergence occurs. As can be seen from Figure 3 however, transients in physical variables tend to be damped significantly by the filter.

Figure 4 shows the output for all buses of the max-consensus iteration (15). Every 30 seconds, generator states were initialized at \( w_i(t_0) = 1 \), while load bus states were initialized at \( x_i(t_0) \). Iterations were performed once a second, and within four to five iterations each bus converges to the largest bus sensitivity. Comparing Figure 4 to Figure 2, this largest sensitivity can be seen to belong to bus 12, with the trend in Figure 4 accurately tracking the green trace of Figure 2. Each bus therefore quickly obtains knowledge of the worst-case global sensitivity.
power injections. As these additional equilibrium equations state reactive power limits are typically functions of active tor, controller, limiter, and load dynamics. Moreover, steady-equilibrium equations [6, Chapter 6] corresponding to generator, controller, limiter, and load dynamics. Furthermore, steady-state reactive power limits are typically not coupled between buses, designing analogous distributed algorithms should be feasible as inter-bus coupling occurs only through the power flow equations. While we have included a basic load-flow type model of generator reactive power limits, more detailed models could be incorporated which include limits based directly on field and armature currents. Future work will also concern relaxing the assumption of full grid observability to partial observability by combining the algorithms presented herein with distributed estimation. Another key direction is the fusion of monitoring protocols such as the one presented herein with automatic distributed control and protection. It seems plausible that optimal control directions in parameter space [56] can be computed and executed in a distributed way, creating a self-healing grid.

APPENDIX A
TECHNICAL LEMMAS AND PROOFS

Lemma A.1 (Power Flow Jacobian): Let \((\theta, V_L) \in \mathbb{R}^{n+m} \times \mathbb{R}^n_{>0}\) be a solution of the power flow equations (1a)–(1b), and let \(d_{ij}\) and \(D_{ij}\) be as in (9). When evaluated at \((\theta, V_L)\) the partial derivatives of (1a)–(1b) are given (for \(j \neq i\)) by [42, Sec. 3.5.1.1]

\[
\begin{align*}
\frac{\partial P_i}{\partial \theta_i} &= -Q_i + d_{ii}, & \frac{\partial P_i}{\partial \theta_j} &= d_{ij}, \\
V_i \frac{\partial P_i}{\partial V_i} &= P_i + D_{ii}, & V_j \frac{\partial P_i}{\partial V_j} &= D_{ij}, \\
V_i \frac{\partial Q_i}{\partial V_i} &= Q_i + d_{ii}, & V_j \frac{\partial Q_i}{\partial V_j} &= d_{ij}, \\
\frac{\partial Q_i}{\partial \theta_i} &= P_i - D_{ii}, & \frac{\partial Q_i}{\partial \theta_j} &= -D_{ij}.
\end{align*}
\]

Proof of Theorem 3.2: To begin, note that by setting \(\delta P = 0_{n+m}\) and \(\delta Q_L = 0_n\) in (6) and eliminating \(\delta \theta\) from the first two blocks of equations, one obtains

\[
0_n = \frac{\delta Q_L}{\delta V_L} \delta V_L + \frac{\delta Q_L}{\delta V_G} \delta V_G,
\]
where $\delta Q_L/\delta V_L$ is as in (7) and

$$
\frac{\delta Q_L}{\delta V_G} = \frac{\partial Q_L}{\partial V_G} - \frac{\partial Q_L}{\partial \theta} \left( \frac{\partial P}{\partial \theta} \right)^T \frac{\partial P}{\partial V_G}.
$$

It follows then from the definition (3) of the index $\mathcal{J}$ that

$$
\mathcal{J} = - \left( \frac{\delta Q_L}{\delta V_L} \right)^{-1} \frac{\delta Q_L}{\delta V_G} \delta m,
$$

where $\mathcal{J} = (\mathcal{J}_1, \ldots, \mathcal{J}_n)$. Using Lemma A.1 one may deduce that the distributed algorithm in Theorem 3.2 may be written in vector form as

$$
\tau \left( \begin{array}{c} \dot{x} \\ \dot{y} \\ \dot{z} \end{array} \right) = - \begin{bmatrix} \frac{\partial Q_L}{\partial V_G} & \frac{\partial Q_L}{\partial \theta} & \frac{\partial P}{\partial \theta} \\ \frac{\partial Q_L}{\partial V_G} & \frac{\partial Q_L}{\partial \theta} & 0 \\ 1 & 0 & \frac{\partial \theta}{\partial \theta} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} \frac{\partial Q_L}{\partial V_G} \delta m \\ 0 \\ -\delta m \frac{\partial Q_L}{\partial V_G} \end{bmatrix}.
$$

(17)

Setting the left-hand side of (17) to zero and eliminating the auxiliary state $y$, one finds that the unique $x$-component of any equilibrium is given uniquely by $x = \mathcal{J}$. Convergence of $x(t)$ to $\mathcal{J}$ follows from arguments similar to those in the proof of Theorem 3.1.

Proof of Theorem 3.3: To begin, set $\delta P = 0_{m+1}$ and $\delta V_G = 0_{m}$ in (6) and eliminate $\delta \theta$ from the second and third blocks of equations to obtain

$$
\delta Q_L = \frac{\delta Q_L}{\delta V_L} \delta V_L, \quad \delta Q_G = \frac{\delta Q_G}{\delta V_L} \delta V_L,
$$

where $\delta Q_L/\delta V_L$ is as in (7) and

$$
\frac{\delta Q_G}{\delta V_L} = \frac{\partial Q_G}{\partial V_L} - \frac{\partial Q_G}{\partial \theta} \left( \frac{\partial P}{\partial \theta} \right)^{-1} \frac{\partial P}{\partial V_L}.
$$

Eliminating $\delta V_L$ from this pair, we find that

$$
\frac{\delta Q_G}{\delta Q_L} = \left( \frac{\delta Q_G}{\delta V_L} \right)^{-1} \frac{\delta Q_L}{\delta V_L}.
$$

Comparing to the definition of the index $\mathcal{K}$ in (4), we find that in vector form

$$
\mathcal{K} = \begin{bmatrix} \frac{\delta Q_G}{\delta Q_L} \\ \frac{\delta Q_G}{\delta V_L} \end{bmatrix}^T 1_m = \left( \frac{\delta Q_G}{\delta V_L} \right)^{-T} \left( \frac{\delta Q_G}{\delta V_L} \right)^T 1_m,
$$

where $\mathcal{K} = (\mathcal{K}_1, \ldots, \mathcal{K}_n)$. Using Lemma A.1 one may deduce that the distributed algorithm in Theorem 3.2 may be written in vector form as

$$
\tau \left( \begin{array}{c} \dot{x} \\ \dot{y} \\ \dot{z} \end{array} \right) = - \begin{bmatrix} \frac{\delta Q_G}{\delta V_L}^T & \frac{\delta Q_G}{\delta V_L}^T & \frac{\partial \theta}{\partial \theta}^T \\ \frac{\delta Q_G}{\partial \theta}^T & \frac{\delta Q_G}{\partial \theta}^T & 0 \\ 1 & 0 & \frac{\partial \theta}{\partial \theta}^T \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} \frac{\delta Q_G}{\delta V_L} \delta m \\ 0 \\ -\delta m \frac{\partial Q_G}{\partial V_G} \end{bmatrix}.
$$

(18)

By setting the left-hand side of (18) to zero and eliminating $y$ and $z$, one may verify that the $x$-component of any equilibrium is given uniquely by $x = \mathcal{K}$. Convergence of $x(t)$ to the index $\mathcal{K}$ follows from arguments similar to those in the proof of Theorem 3.1. □


