Reinforcement Learning for Distributed Transient Frequency Control with Stability and Safety Guarantees

Zhenyi Yuan\textsuperscript{1}  Changhong Zhao\textsuperscript{2}  Jorge Cortés\textsuperscript{1}

Abstract—This paper proposes a reinforcement learning-based approach for optimal transient frequency control in power systems with stability and safety guarantees. Building on Lyapunov stability theory and safety-critical control, we derive sufficient conditions on the distributed controller design that ensure the stability and transient frequency safety of the overall system. These conditions characterize the search space of control policies for our learning approach. We construct neural network controllers that parameterize such control policies and use reinforcement learning to train an optimal one. Simulations on the IEEE 39-bus network illustrate the guaranteed stability and safety properties of the controller along with its significant reduction in cost.

I. INTRODUCTION

With modern power systems shifting from high-inertia traditional generations to low-inertia renewable resources, it is increasingly important to design control mechanisms that allow to operate frequency around its nominal value. To tackle such endeavor, the appeal of learning methods lies in the convenience of incorporating large amounts of data and accounting for optimality considerations in the control design. This paper is a contribution to the growing body of work that seeks to leverage learning in the synthesis of efficient decision-making mechanisms in power systems that have rigorous guarantees on stability and performance.

Literature Review. Transient stability of power systems refers to its ability to regain operating equilibrium after disturbances, while retaining the state within operational margins. The literature has investigated optimal frequency control design for improving transient stability, e.g., load-side control [1], [2], droop coefficient design [3], and PD control [4], to mention a few. These methods either rely on designing optimal linear feedback controllers offline or solving optimization problems in real time to obtain optimal control policies. While these approaches ensure transient stability, they do not strictly guarantee transient safety, as the frequency may enter unsafe regions before convergence. To address this, [5] combines Lyapunov stability analysis and safety-control methods to ensure both stability and transient safety. This approach is further combined with model predictive control in [6], [7] to enhance the cooperation between control signals.

Recent research has employed learning methods to improve frequency control design without the restriction for the controllers to be linear or the need to solve computationally complex optimization problems in real time. Reinforcement learning (RL) has emerged as an attractive method to learn such control policies offline, see e.g., [8]. In general, stability and safety of the closed-loop system are not guaranteed without additional design constraints on the learned policies. This has resulted in a number of recent works that develop stability [9]–[12] or safety [13]–[15] guaranteed RL approaches to learn optimal controllers for frequency [9], [10], [13] and voltage control [11], [12], [14], [15] in power systems. Here, we develop an RL-based approach that jointly addresses guaranteed stability and transient frequency safety.

Statement of Contributions. We study optimal transient frequency control in power systems with dynamics described by the swing equations. We formulate an optimization problem to identify control designs that minimize the frequency deviation from the equilibrium and the control cost over time while ensuring asymptotic stability and transient frequency safety in the presence of disturbances. Leveraging notions of Lyapunov stability and safety-critical control, we identify constraints on the distributed controller design whose satisfaction automatically guarantees that the closed-loop system remains stable and the transient frequency behavior stays within desired safety bounds. These constraints use budgets to break down the requirement of collectively satisfying an inequality to ensure stability into individual stability conditions, one per bus, in a way that is distributed and allows additional design flexibility for certain buses while having others compensate for it. These constraints define the search space of distributed, stable, and safe control policies. We leverage them to enforce appropriate structural constraints on neural networks so that the resulting parameterized controller belongs to the search space and can approximate with arbitrary accuracy any of its elements. Finally, we use a recurrent neural network (RNN)-based RL framework to learn optimal parameters for these neural networks. Simulation results on the IEEE 39-bus power system validate the guarantees on stability and transient safety while significantly reducing the cost.

II. PRELIMINARIES

We introduce here basic notions from algebraic graph theory and the swing dynamics for power systems\textsuperscript{1}.

Graph theory: Here we present some basic notions in graph theory [16]. Let $\mathcal{G} = (\mathcal{I}, \mathcal{E})$ be an undirected graph, where $\mathcal{I} = \{1, \ldots, n\}$ is the node set and $\mathcal{E} = \{e_1, \ldots, e_m\} \subseteq \mathcal{I} \times \mathcal{I}$ is the edge set. Two nodes are neighbors if there exists an

\footnotesize
\begin{itemize}
  \item \textsuperscript{1}Department of Mechanical and Aerospace Engineering, UC San Diego. \{zy\}@ucsd.edu. \textsuperscript{2}Department of Information Engineering, The Chinese University of Hong Kong, chzha@ie.cuhk.edu.hk. The work of ZY and JC was supported by NSF Award 2044900, and the work of CZ was supported by Hong Kong Research Grants Council through ECS Award No. 24210220.
\end{itemize}
edge linking them. We denote by \( N_i \) the set of neighbors of node \( i \). A path is an ordered sequence of nodes such that any pair of consecutive nodes in the sequence is an edge of the graph. The graph \( G \) is connected if there exists a path between any two nodes. The adjacency matrix \( A \) is defined as \(|A|_{ij} > 0 \) if \( i \) and \( j \) are neighbors, 0 otherwise. The Laplacian matrix \( L \) is defined as \(|L|_{ij} = -|A|_{ij} \) for \( i \neq j \), and \(|L|_{ii} = \sum_{j=1,j\neq i}^{n}|A|_{ij} \). The value 0 is an eigenvalue of \( L \) with eigenvector \( \mathbf{1}_n \). This eigenvalue is simple if and only if the graph is connected. For each edge \( e_k \in \mathcal{E} \) with nodes \( i, j \), we assign an arbitrary orientation so that either \( i \) or \( j \) is the source of \( e_k \) and the other node is the target of \( e_k \). Then the incidence matrix \( B = (d_{ik}) \in \mathbb{R}^n \times m \) of graph \( G \) is defined as:

\[
d_{ik} = \begin{cases} 
1 & \text{if node } i \text{ is the source of edge } e_k \\
-1 & \text{if node } i \text{ is the target of edge } e_k \\
0 & \text{otherwise}
\end{cases}
\]

Power network dynamics: The power network is modeled by a connected undirected graph \( G = (\mathcal{I}, \mathcal{E}) \), where \( \mathcal{I} = \{1, \ldots, n\} \) is the set of buses and \( \mathcal{E} \subseteq \mathcal{I} \times \mathcal{I} \) is the set of transmission lines. We assume each bus represents an aggregate area consisting of loads and generators. For each bus \( i \in \mathcal{I} \), we use \( \theta_i \in \mathbb{R}, \omega_i \in \mathbb{R}, p_i \in \mathbb{R} \) to represent its voltage angle, frequency deviation (from the nominal value), and active power injection, respectively. The frequency dynamics is described by the swing equations [17]:

\[
\dot{\theta}_i(t) = \omega_i(t), \\
M_i \dot{\omega}_i(t) = -D_i \omega_i(t) - \sum_{j \in N_i} b_{ij} \sin(\theta_i(t) - \theta_j(t)) + u_i(t) + p_i,
\]

for all \( i \in \mathcal{I} \), where \( M_i, D_i \in \mathbb{R}_{\geq 0} \) are the inertia and damping coefficients of bus \( i \), respectively, and \( b_{ij} \in \mathbb{R}_{\geq 0} \) is the susceptance of the transmission line connecting buses \( i \) and \( j \). For simplicity, we assume \( M_1, D_1, b_{ij} \) are all positive.

Define vectors \( \theta \triangleq [\theta_1, \ldots, \theta_n]^\top \in \mathbb{R}^n, \omega \triangleq [\omega_1, \ldots, \omega_n]^\top \in \mathbb{R}^n \) and \( p \triangleq [p_1, \ldots, p_n]^\top \in \mathbb{R}^n \). Let \( B \in \mathbb{R}^{n \times m} \) be the incidence matrix under an arbitrary graph orientation, and define the voltage angle difference vector:

\[
\lambda(t) \triangleq B^\top \omega(t) \in \mathbb{R}^m.
\]

Denote by \( Y_b \in \mathbb{R}^{m \times n} \) the diagonal matrix with \( Y_{b,k,k} = b_{kj} \), for \( k = 1, 2, \ldots, m \), and define \( M \triangleq \text{diag}(M_1, M_2, \ldots, M_n) \in \mathbb{R}^{n \times n}, \ D \triangleq \text{diag}(D_1, D_2, \ldots, D_n) \in \mathbb{R}^{n \times n} \). We rewrite the dynamics (1) in a compact form in terms of \( \lambda(t) \) and \( \omega(t) \) as:

\[
\dot{\lambda}(t) = B^\top \omega(t) \\
M \dot{\omega}(t) = -D \omega(t) - BY_b \sin \lambda(t) + u(t) + p,
\]

where \( u(t) \triangleq [u_1(t), \ldots, u_n(t)]^\top \in \mathbb{R}^n \) and \( \sin \lambda \in \mathbb{R}^m \) is the component-wise sine value of \( \lambda \). Note that the transformation (2) enforces \( \lambda(0) \in \text{Range}(B^\top) \). We refer to any \( \lambda(0) \) satisfying this condition as an admissible initial value. For convenience, we use \( x(t) \triangleq (\lambda(t), \omega(t)) \in \mathbb{R}^{m+n} \) to denote the collection of all state variables, and neglect its dependence on \( t \) if the context is clear.

Let \( L \triangleq BY_b B^\top \), and define \( \omega_{\infty} \triangleq \frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} D_i} \). Then

\[
\|L^\top \tilde{p}\|_{\mathcal{E}, \infty} < 1,
\]

where \( \|y\|_{\mathcal{E}, \infty} \triangleq \max_{(i,j) \in \mathcal{E}} |y_i - y_j| \), then there exists \( \lambda^\infty \in \Gamma \triangleq \{\lambda \mid |\lambda_k| < \pi/2, \forall k = 1, \ldots, m\} \) unique in \( \Gamma \triangleq \{\lambda \mid |\lambda_k| \leq \pi/2, \forall k = 1, \ldots, m\} \) such that

\[
\tilde{p} = BY_b \sin \lambda^\infty \text{ and } \lambda^\infty \in \text{Range}(B^\top).
\]

Provided \( \lambda(0) \in \text{Range}(B^\top) \), (3) with \( u \equiv 0 \) has a unique equilibrium \((\lambda^\infty, \omega^\infty)_{1n}\), which is asymptotically stable.

III. Problem Formulation

Consider a power network modeled as in Section II. Under the condition (4), the unforced system admits a unique locally asymptotically stable equilibrium point \((\lambda^\infty, \omega^\infty)_{1n}\). However, in the presence of disturbances, the transient frequency can enter unsafe regions before convergence to the equilibrium. This can be caused, for instance, by a sudden change in load or generation. If such frequency excursions exceed certain bounds, they might in fact lead to failure of loads or generators. To address this, we need to design feedback controllers \( \{u_i \}_{i \in \mathcal{I}} \) to ensure each nodal frequency \( \omega_i \) stays within its safety bounds \([\overline{\omega}_i, \overline{\omega}_i]\) during transients, while preserving the asymptotic stability of (3). We also seek to minimize the frequency deviation from the equilibrium and the control cost integrated over time. This gives rise to:

\[
\min_u \int_{t=0}^{T} \left\{ \gamma \|\omega(t) - \omega^\infty_{1n}\|^2 + \|u(t)\|^2 \right\} \, dt \tag{5a}
\]

subject to:

\[
\lambda = B^\top \omega \tag{5b}
\]

\[
M \dot{\omega} = -D \omega - BY_b \sin \omega + u + p \tag{5c}
\]

\[
\lim_{t \to \infty} (\lambda, \omega) = (\lambda^\infty, \omega^\infty), \quad \overline{\omega}_i \leq \omega_i \leq \overline{\omega}_i \tag{5d}
\]

where \( T \) is the time horizon of interest and \( \gamma \) is a coefficient balancing control cost and frequency deviation. Here \( p = p_{\text{nom}} + \Delta p \), where \( p_{\text{nom}} \) is the nominal power injection and \( \Delta p \) accounts for the disturbance. We assume \( \Delta p \) vanishes in finite time, and hence only affects the transient behavior. The term \( \gamma \|\omega - \omega^\infty_{1n}\|^2 \) in the objective function penalizes deviation from the equilibrium frequency, and can be interpreted as a soft constraint to provide approximate transient safety. Instead, the safety bound in (5d) is a hard constraint to strictly guarantee transient safety by prohibiting the frequency nadir going outside the safe region. We also require the designed feedback controllers to be distributed, in the sense that each bus can implement \( u_i(x, p) \) using its own information and the information from its neighbors and incident transmission lines.

The infinite-dimensional and nonlinear nature of the optimization (5) makes it complex to solve. Reinforcement learning (RL) is an attractive approach to design by employing data from system executions to train a policy that maps states to input actions. This results in a learned controller with optimized performance for the given data, but does not guarantee the stability and safety of the closed-loop system. Instead, model-based methods leverage knowledge of the dynamics to synthesize feedback controllers that render the system stable and safe, but have trouble dealing with the infinite-dimensional nature of the optimization. The advantages and limitations of RL and model-based approaches motivate us to combine them by identifying conditions on the controller design that ensure stability and safety (cf. Section IV) and incorporating these conditions in the RL policy search (cf. Section V).
IV. Search Space of Control Policies

Here we identify constraints on the control design that ensure asymptotic stability and transient frequency safety. These constraints define later the search space of control policies.

A. Constraint ensuring asymptotic stability

Here we derive a constraint on the control design that ensures asymptotic stability. We approach this by considering an energy function and restricting the input so that its time-derivative along the closed-loop dynamics is nonnegative. Following [5], [19], consider

\[ V(\lambda, \omega) = \frac{1}{2} \sum_{i=1}^{n} M_i (\omega_i - \omega_i^\infty)^2 + \sum_{j=1}^{m} \left[ Y_{0,j} \right] a(\lambda_j), \quad (6) \]

where \( a(\lambda_j) \triangleq \cos \lambda_j^\infty - \cos \lambda_j - \lambda_j \sin \lambda_j^\infty + \lambda_j^\infty \sin \lambda_j \).

The derivative of \( V \) along the dynamics (3) is given by

\[ \dot{V}(\lambda, \omega) = -\sum_{i=1}^{n} D_i (\omega_i - \omega_i^\infty)^2 + \sum_{i=1}^{n} (\omega_i - \omega_i^\infty) u_i(x,p). \quad (7) \]

To ensure \( \dot{V}(\lambda, \omega) \leq 0 \), one can simply ask \( u_i(x,p) \) to satisfy

\[ -D_i (\omega_i - \omega_i^\infty)^2 + (\omega_i - \omega_i^\infty) u_i(x,p) \leq 0, \quad (8) \]

for each \( i \in \mathcal{I} \). This stability condition is convenient, from a network perspective, because it provides an individually decoupled constraint for each bus. Nevertheless, it is over-constraining, as the sum of all terms in (7) is what needs to be nonpositive, not each individual summand. One could envision scenarios where some buses can deal with larger disturbances than others. In such cases, it would be advantageous to allow less capable buses to violate (8) up to a level that can be compensated by more capable buses to still make the overall sum (7) nonpositive. Leveraging this insight, the next result generalizes the stability condition in [5, Lemma 4.1].

**Lemma IV.1.** (Sufficient condition for local asymptotic stability: Consider system (3) under condition (4). Further suppose that for every \( i \in \mathcal{I} \), \( u_i(x,p) : \mathbb{R}^{m+n} \times \mathbb{R}^n \to \mathbb{R} \) is Lipschitz in \( x \). Let \( c \triangleq \min_{\lambda \in \partial \mathcal{G}_1} V(\lambda, \omega^\infty, 1_n) \) and define

\[ \mathcal{J}_\beta \triangleq \{ (\lambda, \omega) \mid \lambda \in \mathcal{G}_1; V(\lambda, \omega, \omega^\infty, 1_n) \leq c/\beta \} \quad (9) \]

with \( \beta \in \mathbb{R}_{>0} \). If for every \( i \in \mathcal{I} \), \( x \in \mathbb{R}^{m+n} \), and \( p \in \mathbb{R}^n \),

\[ (\omega_i - \omega_i^\infty) u_i(x,p) \leq D_i (\omega_i - \omega_i^\infty)^2 + b_i, \quad \text{if} \quad \omega_i \neq \omega_i^\infty, \]

\[ u_i(x,p) = 0, \quad \text{if} \quad \omega_i = \omega_i^\infty, \]

where \( 0 < D_i < D_i \) and \( \sum_{i=1}^{n} b_i = 0 \). Then, provided \( \lambda(0) \in \text{Range}(B^T) \) and \( (\lambda(0), \omega(0)) \in \mathcal{J}_\beta \) for some \( \beta > 1 \),

1) The solution of the closed-loop system exists and is unique for all \( t \geq 0 \);

2) \( \lambda(t) \in \text{Range}(B^T) \) and \( (\lambda(t), \omega(t)) \in \mathcal{J}_\beta \) for all \( t \geq 0 \);

3) \( (\lambda^\infty, \omega^\infty, 1_n) \) is stable, and \( \lim_{t \to \infty} (\lambda(t), \omega(t)) = (\lambda^\infty, \omega^\infty, 1_n) \).

**Proof.** Note that \( \mathcal{J}_\beta \) is non-empty and compact. Hence if 2) holds, then 1) follows [20, Theorems 3.1 and 3.3]. Therefore we focus on the statements 2)-3). From (7),

\[ \dot{V}(\lambda, \omega) = -\sum_{i=1}^{n} D_i (\omega_i - \omega_i^\infty)^2 + \sum_{i=1}^{n} (\omega_i - \omega_i^\infty) u_i(x,p) \]

Hence, given \( (\lambda(0), \omega(0)) \in \mathcal{J}_\beta \), we have \( V(\lambda, \omega) \leq V(\lambda(0), \omega(0)) \leq c/\beta \), and 2) follows. For 3), note that \( V(\lambda, \omega) > 0 \) for \( (\lambda, \omega) \in \mathcal{J}_\beta \setminus (\lambda^\infty, \omega^\infty, 1_n) \), and \( V(\lambda^\infty, \omega^\infty, 1_n) = 0 \), combined with \( \dot{V}(\lambda, \omega) \leq 0 \), implies that \( (\lambda^\infty, \omega^\infty, 1_n) \) is stable. Furthermore, noticing \( V(\lambda, \omega) = 0 \) implies that \( \omega = \omega^\infty, 1_n \), let \( \Omega = \{ (\lambda, \omega) \in \mathcal{J}_\beta \mid \omega = \omega^\infty, 1_n \} \), it is easy to see from (3) that the largest invariant set in \( \Omega \) is the point \( \{ (\lambda^\infty, \omega^\infty, 1_n) \} \). Then 3) follows the LaSalle Invariance Principle [20, Theorem 4.4].

The quantities \( \{ b_i \}_{i \in \mathcal{I}} \in \mathbb{R}^n \) in Lemma IV.1 correspond to the **budgets** that allow some buses to violate the local condition (8) and instead satisfy

\[ \begin{align*}
    u_i(x,p) &\leq D_i (\omega_i - \omega_i^\infty) + \frac{b_i}{(\omega_i - \omega_i^\infty)} \quad \omega_i > \omega_i^\infty, \\
    u_i(x,p) &\geq D_i (\omega_i - \omega_i^\infty) + \frac{b_i}{(\omega_i - \omega_i^\infty)} \quad \omega_i < \omega_i^\infty,
\end{align*} \]

while ensuring system stability as long as \( \sum_{i=1}^{n} b_i = 0 \). Interestingly, the condition (10) is more general than the stability condition in [5, Lemma 4.1], which requires \( u_i \) to have a different sign from \( (\omega_i - \omega_i^\infty) \), and the stability condition in [9, Theorem 1], which further requires \( u_i \) to be monotonically decreasing with \( (\omega_i - \omega_i^\infty) \).

B. Constraint ensuring frequency invariance

We next turn our attention to the identification of conditions on the controller design that ensure the transient safety requirement, i.e., \( \omega_i(t) \) staying in \( [\omega_{j}, \omega_{j+1}] \) for all \( i \in \mathcal{I} \) and all \( t \geq 0 \). For convenience, let \( \mathcal{Q}_i \triangleq \{ x \mid \omega_i \leq \omega_i \leq \omega_{j+1}, \forall i \in \mathcal{I} \} \). To make this set forward invariant, one simply needs to ensure that the time-derivative of the frequency is negative when \( \omega_i = \omega_{j+1} \), positive when \( \omega_i = \omega_{j} \), and anything when \( \omega_j \in (\omega_{j}, \omega_{j+1}) \). However, such specification may result in discontinuous controllers. Instead, we seek a specification that gradually kicks in as the frequency reaches certain thresholds, while retaining the stability properties of (3) in the absence of input when the frequency is inside the thresholds. Built on this idea, the next result identifies a sufficient condition for a continuous controller design to ensure invariance of \( \mathcal{Q}_i \).

**Lemma IV.2.** (Sufficient condition for frequency invariance [5, Lemma 4.4]): Assume the solution of (3) exists and is unique for every admissible initial condition. For each \( i \in \mathcal{I} \), let \( \omega_{j}^{th}, \omega_{j+1}^{th} \in \mathbb{R} \) be such that \( \omega_{j}^{th} < \omega_{j}^{th} < \omega_{j}^{th} < \omega_{j+1}^{th} \) and \( \omega_{j}^{th} \) and \( \omega_{j+1}^{th} \) be class-K functions. If for all \( x \in \mathbb{R}^{m+n} \) and \( p \in \mathbb{R}^n \),

\[ \begin{align*}
    u_i(x,p) &\leq \frac{\alpha_{i}^{th}(\omega_i - \omega_i^{th})}{\omega_i - \omega_i^{th}} + q_i(x,p), \quad \omega_i > \omega_i^{th}, \\
    u_i(x,p) &\geq \frac{\alpha_{i}^{th}(\omega_i - \omega_i^{th})}{\omega_i - \omega_i^{th}} + q_i(x,p), \quad \omega_i < \omega_i^{th},
\end{align*} \]

where \( q_i(x,p) \triangleq D_i \omega_i + [B^T] \), then \( \mathcal{Q}_i \) is invariant.

C. Distributed, stable and safe control policies

Here, we combine the results of the previous sections to identify the search space of distributed policies. Note that both sufficient conditions (10)-(11) are naturally distributed,
\[
\begin{align*}
\left\{ \begin{array}{l}
u_i(x, p) \leq \min \left\{ \widetilde{D}_i(\omega_i - \omega^\infty) + \frac{b_i}{(\omega_i - \omega^\infty)^2}, \frac{\pi_i(x, p)}{(\omega_i - \omega^\infty)^{\alpha_i}} + q_i(x, p) \right\} \\
u_i(x, p) = 0 \\
u_i(x, p) \geq \max \left\{ \widetilde{D}_i(\omega_i - \omega^\infty) + \frac{b_i}{(\omega_i - \omega^\infty)^2}, \frac{\alpha_i(x, p)}{(\omega_i - \omega^\infty)^{\alpha_i}} + q_i(x, p) \right\}
\end{array} \right. \\
\omega_i > \omega^\infty, \\
\omega^\infty \leq \omega_i \leq \omega^\infty, \\
\omega_i < \omega^\infty. 
\end{align*}
\]

except for the requirement that \( \sum_{i=1}^{n} b_i = 0 \). The satisfaction of this equality requires coordination across the buses. Note that a static, a priori budget assignment will in general not work. This is because, if \( b_i \neq 0 \), then (10) might require the control input to be infinitely large (instead of vanishing) when \( \omega_i \) approaches \( \omega^\infty \). Instead, we need a dynamic budget assignment which makes \( b_i \) approach zero as \( \omega_i \) approaches \( \omega^\infty \).

Motivated by this, we propose next a policy design with a distributed dynamic budget assignment to satisfy (10)-(11).

**Theorem IV.3.** (Distributed control policies with asymptotic stability and transient safety guarantees): Given thresholds \( \omega^\infty \), \( \omega^\infty \) such that \( \omega^\infty \in (\omega^\infty, \omega^\infty) \), let \( \xi \in \mathbb{R}^n \) satisfy \( \|\mathcal{L}_i\|_{\|\cdot\|_\infty} \leq \min\{\widetilde{D}_i(\mathcal{E}_{\xi} - \omega^\infty)^2, \widetilde{D}_i(\mathcal{L}_i - \omega^\infty)^2\} \) for all \( i \in \mathcal{I} \). For \( x \in \mathcal{R}^m \), let \( \mathcal{F} \subseteq \mathcal{I} \) denote the set of buses satisfying \( \omega_i \notin (\omega^\infty, \omega^\infty) \) and \( \mathcal{E}_{\xi} \subseteq \mathcal{E} \) the set of edges between any two nodes in \( \mathcal{T}_{\xi} \). Define the (possibly unconnected) state-dependent subgraph \( \mathcal{G} = (\mathcal{I}, \mathcal{E}_{\xi}) \) of \( \mathcal{G} = (\mathcal{I}, \mathcal{E}) \) and let \( \mathcal{L} \) be its Laplacian matrix. Under condition (4), consider the system (3) with a Lipschitz control policy satisfying (12) with budgets \( b_i = |\mathcal{L}||\xi| \) for each \( i \in \mathcal{I} \). If \( \lambda(t) \in \text{Range}(B^1) \) and \( (\lambda(0), \omega(0)) \in \mathcal{F}(\beta, \omega(0)) \) for some \( \beta > 0 \), then the following holds:

1) The solution of the closed-loop system exists and is unique for all \( t \geq 0 \);
2) \( \lambda(t) \in \text{Range}(B^1) \) and \( (\lambda(t), \omega(t)) \in \mathcal{F}(\beta, \omega(t)) \) for all \( t \geq 0 \); 3) \((\lambda(\infty), \omega(\infty)) \) is stable, and \( \lim_{t \to \infty} (\lambda(t), \omega(t)) = (\lambda(\infty), \omega(\infty)) \); 4) For each \( i \in \mathcal{I} \), if \( \omega_i(0) \in [\omega^\infty, \omega^\infty] \), then \( \omega_i(t) \in [\omega^\infty, \omega^\infty] \) for all \( t > 0 \); 5) For each \( i \in \mathcal{I} \), \( b_i = 0 \) whenever \( \omega_i(t) \in [\omega^\infty, \omega^\infty] \).

**Proof.** Statements 1)-4) readily follow Lemmas IV.1 and IV.2 if (12) (i) ensures that (10)-(11) hold and (ii) defines a specification that can be satisfied by a Lipschitz controller. For (i), note that \( \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} \mathcal{L}_i|\xi| = \sum_{i=1}^{n} \mathcal{E}_{\xi} = 0 \), and therefore (12) implies both (10) and (11). For (ii), notice that the only problem is that (12) requires \( u_i = 0 \) when \( \omega_i \in [\omega^\infty, \omega^\infty] \). Hence, to guarantee it admits a Lipschitz controller, we need to show that \( u_i \) can be chosen as 0 right after \( \omega_i \) passes the thresholds \( \omega^\infty \) and \( \omega^\infty \). Hence, it suffices to show that \( \lim_{\omega_i \to (\omega^\infty)^+} \min \{ \tilde{D}_i(\omega_i - \omega^\infty) + \frac{|\mathcal{L}_i|\xi}{(\omega_i - \omega^\infty)^2}, \frac{\pi_i(x, p)}{(\omega_i - \omega^\infty)^{\alpha_i}} + q_i(x, p) \} \geq 0 \) and \( \lim_{\omega_i \to (\omega^\infty)^-} \max \{ \tilde{D}_i(\omega_i - \omega^\infty) + \frac{|\mathcal{E}_{\xi}(\omega_{\infty} - \omega_i)^2}{} \xi + q_i(x, p) \} \leq 0 \). Now we only show the case when \( \omega_i \to (\omega^\infty)^+ \), since the other case can be proved similarly. Note that \( \lim_{\omega_i \to (\omega^\infty)^+} \min \{ \tilde{D}_i(\omega_i - \omega^\infty) + \frac{|\mathcal{L}_i|\xi}{(\omega_i - \omega^\infty)^2}, \frac{\pi_i(x, p)}{(\omega_i - \omega^\infty)^{\alpha_i}} + q_i(x, p) \} = +\infty \), and hence the minimum is attained by the first term. We end the proof by noting that \( \lim_{\omega_i \to (\omega^\infty)^+} \min \{ \tilde{D}_i(\omega_i - \omega^\infty) + \frac{|\mathcal{L}_i|\xi}{(\omega_i - \omega^\infty)^2}, \frac{\pi_i(x, p)}{(\omega_i - \omega^\infty)^{\alpha_i}} + q_i(x, p) \} = +\infty \), where we have employed the fact that \( ||\mathcal{L}_i|\xi| \leq |||\mathcal{L}_i|||_{\|\cdot\|_\infty} \). Finally, for (5), note that according to the definition of \( \mathcal{L} \), if \( i \notin \mathcal{T}_{\xi} \), then \( |\mathcal{L}_i| = 0 \), and hence \( b_i = |\mathcal{L}_i|\xi = 0 \).

Theorem IV.3 provides a characterization of the search space of distributed control policies that guarantee stable and safe closed-loop systems. Fig. 1 illustrates the search space.

![Fig. 1: The colored region shows the search space for the controller defined by (12), cf. Theorem IV.3, which ensures asymptotic stability and transient safety.](image)

**V. SYNTHESIS OF DISTRIBUTED NEURAL NETWORK CONTROLLERS**

We construct neural networks that parameterize control policies satisfying the requirements in Section IV-C, and then apply an RNN-based FR framework to train an optimal one.

**A. Selection of class-K functions and frequency thresholds**

The condition (12) obtained in Theorem IV.3 depends upon the functions \( \pi_i, \alpha_i \) and the parameters \( \omega^\infty, \omega^\infty \). The choice of class-K functions \( \pi_i \) and \( \alpha_i \) affects the performance of the control policies, including control magnitudes and robustness to disturbances. The choice of \( \omega^\infty \) and \( \omega^\infty \) determines the thresholds for activation of the control action. One can make specific choices for these design variables according to practical considerations. Alternatively, one can use neural networks to parameterize and train them along with the control policy. For the latter, we next provide the details of such parameterizations using single hidden layer neural networks.

**Lemma V.1.** (Neural network parameterization of class-K function): Let \( \sigma(x) = \max(0, x) \) be the ReLU function. Let 
\[
\sigma_i(s) = \sigma_i^+ \alpha_{1m}(s) + \sigma_i^- \pi_{1m}(s),
\]
for \( i \in \mathcal{I} \), where \( \sigma_i^+ \), \( \sigma_i^- \) are bias vectors with \( m \) hidden units satisfying \( \sigma_i^+ = \sigma_i^- = 0 \), \( \sigma_i^+ \), \( \sigma_i^- \) are vectors satisfying \( \sum_{j=1}^{m} \sigma_i^+ \leq \sigma_i^- \), and \( \sigma_i^+ \), \( \sigma_i^- \) are weight vectors satisfying \( \sum_{j=1}^{m} \sigma_i^+ \leq \sigma_i^- \), for \( \ell \in [1, m] \). Then, \( \sigma_i \) is of class-K. Furthermore, for any class-K function \( \kappa \) and given any compact domain \( K \subset \mathbb{R} \), there exist \( \sigma_i^+, \sigma_i^- \) and \( m \) such that \( |\kappa(s) - \sigma_i(s)| \leq \epsilon \) for all \( s \in K \).
We omit its proof due to space limitations, but note that it is analogous to the proof of [9, Theorem 2]. Also note that \( \alpha_i \) can be constructed in the same way with weight vectors \( \tilde{z}_i^+, \tilde{z}_i^- \) and the bias vectors \( \tilde{z}_i^+, \tilde{z}_i^- \).

**Lemma V.2. (Neural network parameterization of frequency threshold):** Let \( \varsigma(x) = \frac{1}{1+e^{-x}} \) be the sigmoid function. Let
\[
\omega_i^{th} = (\omega_\infty - \omega_i) \varsigma(v_i^+) + \omega_i, \quad \omega_i^{-th} = (\omega_\infty - \omega_i) \varsigma(v_i^-) + \omega_i,
\]
for \( i \in I \), where \( v_i^+, v_i^- \in \mathbb{R} \) are biases. Then \( \omega_i^{th} \) and \( \omega_i^{-th} \) approximate any values in \( (\omega_\infty, \omega_{\infty}^{-}) \), and \( (\omega_\infty, \omega_{\infty}^{+}) \), respectively.

The proof readily follows the definition of sigmoid function.

**B. Neural network controller design**

We first give the final ingredient to parameterize control policies that satisfy condition (12) using neural networks. The next result provides a parameterization of any function \( \omega_i \mapsto f_i(\omega_i) \) satisfying \( f_i(\omega_i) = 0 \) for \( \omega_i \in [\omega_i^{th}, \omega_i^{-th}] \).

**Lemma V.3. (Neural network parameterization of \( f_i \)):** Let
\[
f_i(\omega_i) = q_i^+ \sigma(1_m(\omega_i - \omega_i^{th}) + r_i^+) + q_i^- \sigma(-1_m(\omega_i - \omega_i^{-th}) + r_i^-),
\]
for \( i \in I \), where \( r_i^+, r_i^- \in \mathbb{R}^m \) are bias vectors with \( m \) hidden units satisfying \( |r_i^+|^2 \leq 0 \) and \( |r_i^-|^2 \leq 0 \) for all \( j \in [1, m] \), and \( q_i^+, q_i^- \in \mathbb{R}^{1 \times m} \) are weight vectors. Then, \( f_i(\omega_i) = 0 \) for \( \omega_i \in [\omega_i^{th}, \omega_i^{-th}] \). Moreover, for any Lipschitz function \( g_i : \mathbb{R} \rightarrow \mathbb{R} \) satisfying \( g_i(\omega_i) = 0 \) for \( \omega_i \in [\omega_i^{th}, \omega_i^{-th}] \) and given any compact domain \( K \subset \mathbb{R} \) and \( \epsilon > 0 \), there exists \( q_i^+, q_i^- \in \mathbb{R}^{1 \times m} \) and \( m \) such that \( |f_i(\omega_i) - g_i(\omega_i)| < \epsilon \) for all \( \omega_i \in K \).

The proof of this result uses the definition of RELU function and exploits a piece-wise linear approximation similar to that in Lemma V.1. Let \( z := \{ \xi_i^+, \xi_i^-, \tilde{z}_i^+, \tilde{z}_i^- \} \in \mathbb{E} \) and \( c := \{ \tilde{r}_i^+, \tilde{r}_i^-, \tilde{z}_i^+, \tilde{z}_i^- \} \in \mathbb{E} \). Then, \( u \) and \( \tilde{r} \) are constructed using the distributed network controllers.

**Theorem V.4. (Distributed neural network controllers):** For each \( i \in I \), let \( \tilde{r}_i, \alpha_i, \tilde{z}_i^{th}, \tilde{z}_i^{-th}, \) and \( f_i \) be constructed according to the assumptions of Lemmas V.1, V.2 and V.3, respectively. Under the assumptions of Theorem IV.3, let \( \pi_{i,\phi}(x, p) \) be the minimizer of \( D_i(\omega_i - \omega_\infty) + \frac{\|\tilde{z}_i^{th} - \tilde{z}_i^{-th}\|^2}{\omega_i - \omega_\infty} + q_i(x, p) \) and \( u_i,\phi(x, p) \) be the maximizer of \( D_i(\omega_i - \omega_\infty) + \frac{\|\tilde{z}_i^{th} - \tilde{z}_i^{-th}\|^2}{\omega_i - \omega_\infty} + q_i(x, p) \).

**C. Learning optimal control policy using RNN**

Having parameterized in Theorem V.4 the search space identified in Section IV-C, here we describe an approach to train an optimal control policy adopting the RNN-based RL framework proposed in [9]. To simulate the trajectories for training the neural network controller (13), we use a first-order Euler discretization with stepsize \( \Delta t \) for problem (5). Let \( K \) be the total number of timesteps. The discrete-time optimization problem is
\[
\min_{\phi} \frac{1}{K} \sum_{k=0}^{K-1} \gamma R_k + \|u_k\|^2 \quad \text{s.t.} \quad L_k = (k+1) + B^T \omega(k-1)
\]

where \( u_k = [u_{1, \phi}, \ldots, u_{n, \phi}]^T \). The learning algorithm works as follows. At the beginning of the training process, all parameters in \( \phi \) are randomly generated. Training is implemented in a batch updating style, where the initial states of \( \omega(0) \) and \( L(0) \) in each batch are randomly generated. In each episode, we use the current control policy \( u_\phi \) to generate states trajectories of length \( K \) for all batches through dynamics (14b),(14c), and compute the loss function (14a). The trainable parameters \( \phi \) are updated by gradient descent on the loss function (14a) and converge to a local optimum.

**VI. CASE STUDY**

We conduct a case study to illustrate the performance of the proposed approach. We consider the IEEE 39-bus power network and assume each bus represents an area containing loads and generators. The system parameters are from [5], [21]. We consider the time horizon of interest to be \( T = 50 \) seconds and bus 38 encountering a sudden change in power injection during the time interval (0, 2) seconds, with \( \Delta L_{38} = \Delta p_{38} \). The nominal frequency is 60 Hz, and the safe region is set at \( [59.8, 60.2] \) Hz for every bus.

**Simulation Setup.** We build the RL environment using TensorFlow 2.7.0 and conduct the training process in Google Colab with a single TPU with 32 GB memory. We consider the discretization stepsize \( \Delta t \) to be 0.0008 seconds. To facilitate the training process, we only evaluate the first 10 seconds in each episode, meaning the total number of stages \( K \) in each episode is 12500. For each \( i \in I, \omega_i(0) \) is randomly generated in \( [59.9, 60.1] \) Hz, and \( \lambda(0) \) is calculated using power injections randomly generated over \( [0.9 p_{\text{nom}}(0), 1.1 p_{\text{nom}}(0)] \). The balancing coefficient in the objective function is \( \gamma = 40 \), and the number of episodes, batch size, and the number of neurons \( m \) are 150, 50, 20, respectively. We use Adam algorithm [22] to update \( \phi \) in each episode with learning rate 0.05.

**Baseline for Comparison.** We compare our approach to the non-optimal method proposed in [5], where the controller was designed as \( u_i(x, p) = \min(0, \tilde{r}_i^{th} + q_i(x, p)) \) for \( \omega_i > \omega_i^{th}, u_i(x, p) = \max(0, \tilde{r}_i^{-th} + q_i(x, p)) \) for \( \omega_i < \omega_i^{th}, \) and \( u_i(x, p) = 0 \) otherwise, with \( \pi_i(s) = \alpha_i(s) = 2s \) and \( \tilde{z}_i^{th} = 0.1, \tilde{z}_i^{-th} = -0.1 \) for all \( i \in I \).

**Simulation Results.** Fig. 2 illustrates the performance of the RL-based method and the method in [5] with the same randomly generated initial states. Both of them guarantee asymptotic stability and transient safety, while RL-based method reduces the cost by 49.4%. Interestingly, in the RL-based method, more nodes contribute to the transient frequency.
regulation, utilizing the network to cooperatively minimize the cost. This method also achieves faster convergence and smaller transient fluctuations.

VII. CONCLUSIONS

We have presented a reinforcement learning approach to the synthesis of optimal controllers that are distributed and guarantee the stability and transient safety of power networks. Leveraging notions of Lyapunov stability and safety-critical control, we have identified conditions on the controller design that ensure stability and transient frequency safety. These constraints incorporate the idea of endowing some buses with additional design flexibility through budgets in a way that collectively ensures the stability of the overall system. We have constructed neural networks to parameterize the control policies within the identified search space and employed an RL framework to learn an optimal controller. Simulations illustrate the guaranteed stability and transient frequency safety of the resulting closed-loop system while showing a significant reduction in the cost. Future work will investigate additional schemes for dynamic budget allocation in the buses and extensions to higher-order power system dynamical models.

REFERENCES


