Non-Euclidean Contraction Analysis of Continuous-Time Neural Networks

Alexander Davydov, Graduate Student Member, IEEE, Anton V. Proskurnikov, Senior Member, IEEE, and Francesco Bullo, Fellow, IEEE

Abstract—Critical questions in dynamical neuroscience and machine learning are related to the study of continuous-time neural networks and their stability, robustness, and computational efficiency. These properties can be simultaneously established via a contraction analysis. This article develops a comprehensive non-Euclidean contraction theory for continuous-time neural networks. Specifically, we provide novel sufficient conditions for the contractivity of general classes of continuous-time neural networks including Hopfield, firing rate, Persidskii, Lur'e, and other neural networks with respect to the non-Euclidean $\ell_1/\ell_{\infty}$ norms. These sufficient conditions are based upon linear programming or, in some special cases, establishing the Hurwitzness of a particular Metzler matrix. To prove these sufficient conditions, we develop novel results on non-Euclidean logarithmic norms and a novel necessary and sufficient condition for contractivity of systems with locally Lipschitz dynamics. For each model, we apply our theoretical results to compute the optimal contraction rate and corresponding weighted non-Euclidean norm with respect to which the neural network is contracting.

Index Terms—Contraction theory, neural networks, stability of nonlinear systems.

I. INTRODUCTION

Motivation from dynamical neuroscience and machine learning. Tremendous progress made in neuroscience research has produced new understanding of biological neural processes. Similarly, machine learning has become a key technology in modern society, with remarkable progress in numerous computational tasks. Much ongoing research focuses on artificial learning systems inspired by neuroscience that (i) generalize better, (ii) learn from fewer examples, and (iii) are increasingly energy-efficient. We argue that further progress in these disciplines hinges upon modeling, analysis, and computational challenges, some of which we highlight in what follows.

In dynamical neuroscience, several continuous-time neural network (NN) models are widely studied, including membrane potential models, such as the Hopfield NN [26] and firing-rate models [37]. Clearly, such models are simplifications of complex neural dynamics. For example, if $f(x)$ is an NN model of a neural circuit, the true dynamics may be better described by

$$\dot{x}(t) = f(x(t)) + g(x(t), x(t - \tau(t)))$$

(1)

where $g$ captures model uncertainty and time-delays. In other words, to account for uncertainty in the system, the nominal dynamics $f(x)$ must exhibit robust stability with respect to unmodeled dynamics and delays. In addition, central pattern generators (CPGs) are biological neural circuits that generate periodic signals and are the source of rhythmic motor behaviors, such as walking and swimming. To properly model CPGs in NNs, a computational neuroscientist would need to ensure that, if an NN is interconnected with a CPG, then all trajectories of the NN converge to a unique stable limit cycle.

Machine learning scientists have widely adopted discrete-time NNs for pattern recognition and analysis of sequential data and much recent interest [5], [29], [30], [45] has focused on the closely-related class of implicit NNs. In particular, training implicit networks corresponds to solving fixed-point problems of the form

$$x = \Phi(Ax + Bu + b)$$

(2)

where $x$ is the neural state variable, $\Phi$ is an activation function, $A$ and $B$ are synaptic weights, $u$ is the input stimulus, and $b$ is a bias term. Note that (i) the fixed point in (2) is the equilibrium point of the continuous-time NN $\dot{x} = -x + \Phi(Ax + Bu + b)$, and (ii) the training problem requires the efficient computation of gradients of a given loss function with respect to model parameters; in turn, this computation can be cast again as a fixed-point problem. In other words, in the design of implicit NNs, it is essential to pick model weights in such a way that fixed-point equations have unique solutions for all possible inputs and activation functions, and fixed-points and corresponding gradients can be computed efficiently.

Finally, an additional challenge facing machine learning scientists is robustness to adversarial perturbations. Indeed, it is well-known [53] that artificial deep NNs are sensitive to adversarial perturbations: small input changes may lead to large

Manuscript received 9 September 2023; revised 9 May 2024; accepted 22 June 2024. Date of publication 2 July 2024; date of current version 31 December 2024. This work was supported in part by the National Science Foundation Graduate Research Fellowship under Grant 2139391 and in part by AFOSR under Grant FA9550-22-1-0059. The work of Anton Proskurnikov was supported by the European Union—Next Generation EU within the PRIN 2022 program (D.D. 104 - 02/02/2022 MURI) through project “Higher-order interactions in social dynamics with application to monetary networks,” under Grant 2022K8EZBW. Recommended by Associate Editor A. Rapaport. (Corresponding author: Alexander Davydov.)

Alexander Davydov and Francesco Bullo are with the Department of Mechanical Engineering and the Center for Control, Dynamical Systems, and Computation, University of California, Santa Barbara, CA 93106 USA (e-mail: davydov@ucsb.edu; bullo@ucsb.edu).

Anton V. Proskurnikov is with the Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy (e-mail: anton.p.1982@ieee.org).

Digital Object Identifier 10.1109/TAC.2024.3422217
output changes and loss in pattern recognition accuracy. One proposed remedy is to characterize the Lipschitz constants of these networks and use them as regularizers in the training process. This remedy leads to certifiable robustness bounds with respect to adversarial perturbations [19], [46]. In short, the input/output Lipschitz constants of NNs need to be tightly estimated, e.g., in the context of the fixed-point (2).

A contraction theory for NNs: Motivated by the challenges arising in neuroscience and machine learning, this article aims to perform a robust stability analysis of continuous-time NNs and develop optimization methods for discrete-time NN models. Serendipitously, both these objectives can be simultaneously achieved through a contraction analysis for the NN dynamics.

For concreteness’ sake, we briefly review how the aforementioned challenges are addressed by a contraction analysis. Infinitesimally contracting dynamics enjoy highly ordered transient and asymptotic behaviors as follows:

1. initial conditions are forgotten and a certain distance between trajectories is monotonically vanishing [35];
2. time-invariant systems admit a unique globally exponentially stable equilibrium with two natural Lyapunov functions (distance from the equilibrium and norm of the vector field) [35];
3. periodic systems admit a unique globally exponentially stable periodic solution or, for systems with periodic inputs, each solution entrains to the periodic input [48];
4. contracting vector fields enjoy highly robust behavior, e.g., see [15], [54], including the following:
   a. input-to-state stability;
   b. finite input-state gain;
   c. contraction margin with respect to unmodeled dynamics;
   d. input-to-state stability under delayed dynamics.

Hence, the contraction rate is a natural measure/indicator of robust stability.

Regarding computational efficiency, our recent work [9], [29] shows how to design efficient fixed-point computation schemes for contracting systems (with respect to arbitrary and non-Euclidean $\ell_1$/\$\ell_\infty$ norms) in the style of monotone operator theory [49]. Specifically, for contracting dynamics with respect to a diagonally weighted $\ell_1$/\$\ell_\infty$ norm, optimal step-sizes and convergence factors are given in [29, Th. 2]. These results are directly applicable to the computation of fixed-points in implicit neural networks, as in (2). These step-sizes, however, depend on the contraction rate. Therefore, optimizing the contraction rate of the dynamics directly improves the convergence factor of the corresponding discrete algorithm.

Literature review: The dynamical properties of continuous-time NN models have been studied for several decades. Shortly after Hopfield’s original work [26], control-theoretic ideas were proposed in [36]. Later, the authors in [20], [21], and [31] obtained various versions of the following result: Lyapunov diagonal stability of the synaptic matrix is sufficient, and in some cases necessary, for the existence, uniqueness, and global asymptotic stability of the equilibrium. More recently, Nozari and Cortés [40] studied linear-threshold rate neural dynamics, where activation functions are piecewise-affine; it is shown that the dynamics have a unique equilibrium if and only if the synaptic matrix is a $\mathcal{P}$-matrix, a weaker condition than Lyapunov diagonal stability. Since checking this condition is NP-hard, more conservative conditions are provided as well. Beyond Lyapunov diagonal stability and $\mathcal{P}$-matrices, [18] is the earliest reference on the application of logarithmic norms and contraction-theoretic principles to Hopfield NNs and provides results on $\ell_p$ logarithmic norms of the Jacobian for networks with smooth activation functions. Alternatively, Arik [4] proposed a quasi-domain condition on the synaptic matrix (in lieu of Lyapunov diagonal stability). Finally, similar to non-Euclidean contraction, [42] proposes the notion of the nonlinear measure of a map to study global asymptotic stability; this notion is closely related to the $\ell_1$ one-sided Lipschitz constant of the Hopfield NN vector field. A comprehensive survey on stability criteria for continuous-time NNs is available in [59].

The importance of non-Euclidean log norms in contraction theory is highlighted, for example, in [2] and [48]. In the spirit of these works, the non-Euclidean contractivity of monotone Hopfield NNs is studied in [28]; see also [10] for the non-Euclidean contractivity of Hopfield NNs undergoing Hebbian learning.

Finally, Euclidean contractivity of continuous-time NNs has been studied, e.g., see the early reference [18], the related discussion in [45], and the recent work [33].

Contributions: This article contributes fundamental control-theoretic understanding to the study of artificial NNs in machine learning and neuronal circuits in neuroscience, thereby building a hopefully useful bridge among these three disciplines.

Specifically, this article develops a comprehensive contraction theory for classes of continuous-time NN models. In order to develop this theory, we make several technical contributions on non-Euclidean logarithmic norms and nonsmooth contraction theory. To be specific, first, we obtain novel logarithmic norm results including (i) the quasi-convexity of the $\ell_1$ and $\ell_\infty$ logarithmic norms with respect to diagonal weights and provide novel optimization techniques to compute optimal weights which yield larger contraction rates, (ii) logarithmic norm properties of principal submatrices of a matrix with respect to monotonic norms, and (iii) explicit formulas for the $\ell_1$ and $\ell_\infty$ logarithmic norms under multiplicatively weighted uncertainty, resulting in a maximization of the logarithmic norm over a matrix polytope. The matrix polytopes described in (iii) are of special interest since the Jacobian matrix of the Hopfield or firing-rate NN vector field always lies inside this polytope. The formulas in (iii) generalize previous results [18, Th. 3.8], [25, Lemma 3], and [29, Lemma 8].

Motivated by our non-Euclidean logarithmic norm results, we define $M$-Hurwitz matrices, i.e., matrices whose Metzler majorant is Hurwitz. We compare $M$-Hurwitz matrices with other classes of matrices including quasidominant, totally Hurwitz, and Lyapunov diagonally stable matrices.

Second, we provide a nonsmooth extension to contraction theory. We show that, for locally Lipschitz vector fields, the one-sided Lipschitz constant is equal to the essential supremum of the logarithmic norm of the Jacobian. This equality allows us to use our novel logarithmic norm results and apply them to NNs that have nonsmooth activation functions.
Finally, we apply our theoretical developments as we establish conditions for the non-Euclidean contractivity of multiple classes of recurrent neural circuits and nonlinear dynamical models, including Hopfield, firing rate, Persidskii, Lur’e, and others. We consider locally Lipschitz activation functions that satisfy an inequality of the form \( d_i \leq \frac{\phi(x_i) - \phi(0)}{x_i - 0} \leq d_2 \), for all \( x \neq y \in \mathbb{R} \), where \( d_1 \) may be negative and \( d_2 \) may be infinite. Indeed, the importance of nonmonotonic activation functions is discussed in [38]. This class of activation functions is more general than all of the continuous activation functions mentioned in [59, Sec. II-B]. Thus, our non-Euclidean contraction framework allows for a more systematic framework for the analysis of these classes of NNs with fewer restrictions on the activation functions. For each model, we propose a linear program (LP) to characterize the optimal contraction rate and corresponding weighted non-Euclidean \( \ell_1 \) or \( \ell_\infty \) norm. In some special cases, we show that the LP reduces to checking an \( M \)-Hurwitz condition. Our results simplify the computation of a common Lyapunov function over a polytope with \( 2^n \) vertices to a simple condition involving just 2 of its vertices or, in some cases, all the way to a closed form expression.

For each model, we demonstrate that the dynamics enjoy strong, absolute and total contractivity properties. In the spirit of absolute and connective stability, absolute contractivity means that the dynamics are contracting independently of the choice of activation function and connective stability means that the dynamics remain contracting whenever edges between neurons are removed. Total contractivity means that if the synaptic matrix is replaced by any principal submatrix, the resulting dynamics remain contracting. The process of replacing the nominal NN with a subsystem NN is referred to as “pruning” both in neuroscience and in machine learning.

A preliminary version of this work appeared in [16]. Compared to [16], this version (i) includes proofs of all technical results, (ii) provides closed-form worst-case log norms over a larger class of matrix polytopes in Lemma 8, (iii) studies a more general class of locally Lipschitz activation functions in Section VI, allowing for both nonmonotonic activation functions as well as activations that have unbounded derivative, (iv) has a complete characterization of contractivity of Hopfield and firing-rate NNs with respect to both \( \ell_1 \) and \( \ell_\infty \) norms, (v) provides a novel sufficient (and nearly necessary) condition for the non-Euclidean contractivity of a Lur’e model with multiple nonlinearities in Theorem 31, and (vi) includes additional comparisons to Euclidean contractivity conditions in Remark 19 and to Lyapunov diagonal stability in Section II.

Notations: For a set \( S \), we let \( S^n \) be the Cartesian product of \( n \) copies of \( S \), \( |S| \) be its cardinality, and, if \( S \subseteq \mathbb{R} \), \( \text{conv} S \) be the convex hull of \( S \). For two matrices \( A, B \), we let \( AB \) denote the product of \( A \) by the \( j \)th row and \( i \)th column of \( B \), and \( |AB| \) be the entrywise multiplication and \( |A| \) be the entrywise absolute value. For \( p \in [1, \infty) \), we let \( \| \cdot \|_p \) denote the \( \ell_p \) norm, i.e., for a vector \( x \in \mathbb{R}^n \), \( \| x \|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p} \) if \( p \in [1, \infty) \) and \( \| x \|_\infty = \max_{i \in \{1, \ldots, n\}} |x_i| \). For an invertible matrix \( R \in \mathbb{R}^{n \times n} \), we define the \( R \)-weighted \( \ell_p \) norm by \( \| x \|_{p, R} = \| Rx \|_p \). Vector inequalities of the form \( x \leq y \) are entrywise. For a vector \( \eta \in \mathbb{R}^n \), we define \( \| \eta \|_R = \max_{i \in \{1, \ldots, n\}} \eta_i \) to be the diagonal matrix with diagonal entries equal to \( \eta \). We let \( \mathbb{R}^n_{\infty}, \mathbb{R}^n_{\infty,0} \) be the all-ones and all-zeros vectors, respectively.

We say a norm \( \| \cdot \| \) on \( \mathbb{R}^n \) is monotonic if for all \( x, y \in \mathbb{R}^n \), \( |x| \leq |y| \Rightarrow \|x\| \leq \|y\| \). A matrix \( M \in \mathbb{R}^{n \times n} \) is Metzler if \( M_{ij} \geq 0 \) for all \( i \neq j \). For a matrix \( A \in \mathbb{R}^{n \times n} \), its spectral abscissa is \( \alpha(A) = \max\{\Re(\lambda) \mid \lambda \in \text{spec}(A)\} \), where \( \Re(\lambda) \) denotes the real part of \( \lambda \), and its Metzler majorant \( |A|_{Mzr} \in \mathbb{R}^{n \times n} \) is defined by \( (|A|_{Mzr})_{ij} = \max\{|A_{ij}|, \text{if } i = j\} \).

II. PREVIEW OF MAIN CONTRACTIVITY RESULTS AND ADVANTAGES OF A NON-EUCLIDEAN ANALYSIS

To motivate the mathematical tools and analysis in Sections IV–VI, we will showcase the main contractivity results for Hopfield and firing-rate NNs under simplifying assumptions to provide a baseline for comparison to other standard stability conditions for these classes of NNs.

The continuous-time Hopfield and firing-rate NNs are the following two dynamical systems:

\[
\dot{x} = -Cx + Af(x) + u =: f_H(x)
\]
\[
\dot{x} = -Cx + \Phi(Ax + u) =: f_{FR}(x)
\]

where \( x \in \mathbb{R}^n \) is the state of the NN (either a vector of membrane potentials or firing rates), \( C \in \mathbb{R}^{n \times n} \) is a positive semidefinite diagonal matrix of dissipation rates, \( A \in \mathbb{R}^{n \times n} \) is the synaptic matrix, \( u \in \mathbb{R}^n \) is a constant external stimulus, and \( \Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n \) is an activation function which satisfies \( \Phi(x) = (\phi_1(x_1), \ldots, \phi_n(x_n)) \). In the machine learning literature, such continuous-time NNs have been given the name neural ODES [11].

For expository’s sake, we make the following standing assumptions throughout the rest of this section.

Assumption 1:  
(i) \( C = I_n \),  
(ii) the matrix \( |A|_{Mzr} \) is irreducible, and  
(iii) each \( \phi_i \) is continuously differentiable and satisfies  
\[ 0 \leq \phi_i'(x) \leq 1 \quad \text{for all } x \in \mathbb{R}^n. \]

Under these assumptions, we can state our main results compactly:

Proposition 1: Consider the Hopfield and firing-rate NNs (3) and (4) satisfying Assumption 1, suppose \( \alpha(|A|_{Mzr}) < 1 \), and define \( c = 1 - \max\{\alpha(|A|_{Mzr}, 0)\} \). Then

(i) the Hopfield NN is contracting with rate \( c > 0 \), i.e., any two trajectories \( x_1(\cdot), x_2(\cdot) \) of (3) satisfy  
\[ \|x_1(t) - x_2(t)\|_{1,0} \leq e^{-ct}\|x_1(0) - x_2(0)\|_{1,0} \]

for all \( t \geq 0 \), where \( \eta \in \mathbb{R}^n_{\infty,0} \) is the dominant left eigenvector of the Metzler matrix \( |A|_{Mzr} \).

(ii) the firing-rate NN is contracting with rate \( c > 0 \), i.e., any two trajectories \( x_1(\cdot), x_2(\cdot) \) of (4) satisfy  
\[ \|x_1(t) - x_2(t)\|_{\infty,0} \leq e^{-ct}\|x_1(0) - x_2(0)\|_{\infty,0} \]

for all \( t \geq 0 \) where \( \xi \in \mathbb{R}^n_{\infty,0} \) is the dominant right eigenvector of the Metzler matrix \( |A|_{Mzr} \).

In particular, under Assumption 1 and \( \alpha(|A|_{Mzr}) < 1 \) (or equivalently \( \alpha(-I_n + A|_{Mzr}) < 0 \)), for each \( u \in \mathbb{R}^n \), both the Hopfield and firing-rate NNs have unique globally exponentially stable equilibria and thus the condition \( \alpha(-I_n + A|_{Mzr}) < 0 \) provides a novel sufficient condition for the existence of a unique globally exponential stable equilibrium along with many
additional robustness properties offered by contracting systems, such as robustness to uncertainties and entrainment to periodic inputs.

Although in this article we primarily study the continuous-time NNs (3) and (4), we remark that many results apply to classes of discrete-time NNs as well. Specifically, given a continuous-time NN, \( x = f_{\text{NN}}(x) \), which is contracting, the forward Euler discretization of the continuous-time NN with stepsize \( h > 0 \) yields a residual NN

\[
x_{k+1} = x_k + hf_{\text{NN}}(x_k)
\]

which is contracting in the sense of the Banach fixed point theorem for sufficiently small \( h \) (see, e.g., [9, Th. 8]). For recent results on contraction for a different class of discrete-time NNs, we refer to [44].

The condition \( \alpha([-I_n + A]_{\text{Mat}}) < 0 \) is different from the well-known result that Lyapunov diagonal stability (LDS) of \(-I_n + A\), i.e., existence of a vector \( \eta \in \mathbb{R}_{\geq 0}^n \) satisfying

\[
[\eta]([-I_n + A] + ([I_n + A]^T[\eta]) < 0
\]

implies the existence of a unique globally asymptotically stable equilibrium point for the Hopfield NN [21]. Moreover, the condition \( \alpha([-I_n + A]_{\text{Mat}}) < 0 \) is stronger than LDS of \(-I_n + A\), which we prove in Lemma 3, yet it implies the stronger property of contractivity.

Beyond LDS, an alternative way to establish the stability of the NNs (3) and (4) is via absolute stability analysis of Lur’e systems and methods via quadratic Lyapunov functions. These methods are typically based upon linear matrix inequalities (LMIs), see, e.g., [14], [22] and the discussion in [59, Sec. I.V.]. Compared to these classical approaches, establishing contractivity with respect to diagonally-weighted \( \ell_1 \) or \( \ell_\infty \) norms provides both computational and practical advantages, which we highlight in the following paragraphs.

**Computational benefits:** In the non-Euclidean contraction analysis of many classes of NNs, contractivity is checked either via linear programming or, in some simpler instances, the stability of appropriate Metzler matrices. As argued in [43], from a computational point of view, both of these tests are more scalable than LMIs are. Indeed, there exist efficient algorithms for computing Perron eigenvalues and eigenvectors for irreducible Metzler matrices [55].

**Practical benefits:** Compared to stability with respect to a quadratic Lyapunov function, there are also practical advantages to establishing contractivity with respect to diagonally-weighted \( \ell_1 \) and \( \ell_\infty \) norms. These benefits include

(i) the \( \ell_1 \) norm (respectively, the \( \ell_\infty \) norm) is well suited for systems with conserved quantities (respectively, systems with translation invariance), e.g., see the theory of weakly contracting and monotone systems in [8, Ch. 4],

(ii) in machine learning, analysis of the adversarial robustness of a NN often needs to be performed in a non-Euclidean norm, because NNs are known to be vulnerable to small disturbances as measured in the \( \ell_\infty \) norm [23], and

(iii) contractivity with respect to non-Euclidean norms ensures robustness with respect to edge removals and structural perturbations, e.g., see the notion of connective stability in [50].

To elaborate on point (iii) in the previous paragraph, in continuous-time NNs, such as the Hopfield and firing rate NNs (3) and (4), the synaptic matrix \( A \) defines a graph structure whereby there is an outgoing synapse from neuron \( j \) to neuron \( i \) provided that \( A_{ij} \neq 0 \). As we will show in Corollary 13, if the NN is contracting with respect to a diagonally-weighted \( \ell_1 \) or \( \ell_\infty \), it is connectively contracting. Specifically, the removal of any edge\(^1\) or neuron from the graph\(^2\) yields a new NN that remains contracting with a rate greater than or equal to the rate of contraction of the original NN. Note that this property is not enjoyed by stability conditions requiring a matrix to be LDS. Indeed, for the Hopfield NN (3), consider

\[
A = \begin{bmatrix}
0 & -1 & 1 \\
1 & 0 & 15 \\
-1 & -15 & 0
\end{bmatrix}, \quad \tilde{A} = \begin{bmatrix}
0 & -1 & 1 \\
1 & 0 & 15 \\
-1 & 0 & 0
\end{bmatrix}.
\]

Note that \( A \) satisfies (6) with \( \eta = 1 \), so \(-I_n + A\) is LDS and thus the Hopfield NN is stable. However, zeroing out \( A_{32} \) yields \( \tilde{A} \) which verifies \( \alpha([-I_n + \tilde{A}]) > 0 \), so the resulting NN is not absolutely stable.

In the following sections, we introduce additional mathematical tools to prove Proposition 1 under assumptions weaker than those listed in Assumption 1. Specifically, we (1) relax item (i) to \( C \), which is diagonal and positive semidefinite, (2) relax item (ii) to also study \( |A|_{\text{Mat}} \) which may be reducible, and (3) relax item (iii) to study nonsmooth activation functions, which may be nonmonotonic and may have unbounded slope. See Theorems 18, 21, and 24 for these results. Beyond the proof of a more general version of Proposition 1, we also establish \( \ell_\infty \) contractivity of the Hopfield NN in Theorem 20, the \( \ell_1 \) contractivity of the firing-rate NN in Theorem 22, and study the contractivity of other classes of NNs in Section VI-D. In the interest of readability, we postpone proofs of most technical results to Appendix A and include proofs regarding contractivity of classes of NNs in the main body of the text.

**III. REVIEW OF RELEVANT MATRIX ANALYSIS**

**A. Log Norms**

Let \( \| \cdot \| \) be a norm on \( \mathbb{R}^n \) and its corresponding induced norm on \( \mathbb{R}^{n \times n} \). The logarithmic norm (also called log norm or matrix measure) of a matrix \( A \in \mathbb{R}^{n \times n} \) is

\[
\mu(A) := \lim_{h \to 0^+} \left| \frac{\|I_n + hA\| - 1}{h} \right|.
\]

We refer to [17] for a list of properties of log norms, which include subadditivity, convexity, and \( \alpha(A) \leq \mu(A) \). It is known that the log norm corresponding to an \( R \)-weighted \( \ell_p \) norm is \( \mu_{p,R}(A) = \mu_p(RAR^{-1}) \). For diagonally weighted \( \ell_1 \), \( \ell_\infty \), and \( \ell_2 \) norms with \( \eta \in \mathbb{R}_{\geq 0}^n \)

\[
\mu_{1,\eta}(A) = \max_{i \in \{1, \ldots, n\}} A_{ii} + \sum_{j=1, j \neq i}^n \eta_j |A_{ji}|
\]

\(^1\)Removing an edge corresponds to zeroing a non-diagonal entry of \( A \).

\(^2\)Removing the \( i \)-th neuron corresponds to removing the \( i \)-th row and column of \( A \).
\[ \mu_{∞,[η]−1}(A) = \max_{i \in \{1, \ldots, n\}} A_{ii} + \sum_{j=1, j \neq i}^{n} \frac{η_j}{η_i} |A_{ij}| \]
\[ = \min \{ b \in \mathbb{R} \mid \begin{bmatrix} 0 \\ A^\top \end{bmatrix} \eta \leq b \eta \} \]
\[ \mu_{2,[η]^{1/2}}(A) = \min \{ b \in \mathbb{R} \mid \begin{bmatrix} 0 \\ A \end{bmatrix} η \leq b \eta \} \]

The following result is due to [41, Lemma 3] and [52].

**Lemma 2 (Optimal diagonally-weighted log norms for Metzler matrices):** Given a Metzler matrix \( M \in \mathbb{R}^{n \times n} \), let \( η ∈ \mathbb{R}^n \) and \( δ \geq 0 \), define \( η_{M,p,δ} ∈ \mathbb{R}^n \) by
\[ η_{M,p,δ} = \left( \frac{1/p}{w_1}, \ldots, \frac{1/p}{w_n} \right) \]
where \( q \in [1, \infty] \) is defined by \( 1/p + 1/q = 1 \) (with the convention \( 1/∞ = 0 \) and where \( v \) and \( w \) are nonnegative for which the Perron–Frobenius theorem holds). Then, for each \( ε \geq 0 \) there exists \( δ > 0 \) such that
\( (1) \) \( \alpha(M) \leq μ_{[η_{M,p,δ}]}(M) \leq \alpha(M) + ε \) and
\( (2) \) if \( M \) is irreducible, then \( \alpha(M) = μ_{[η_{M,p,δ}]}(M) \).

**Lemma 2** also ensures that for Metzler matrices \( M \in \mathbb{R}^{n \times n} \), inf \( η_{M,p,δ} μ_{[η]}(M) = \alpha(M) \) for every \( p \in [1, \infty] \).

**B. Classes of Matrices**

We say a matrix \( A \in \mathbb{R}^{n \times n} \) is

(i) **Hurwitz stable**, denoted by \( A ∈ H \), if \( \alpha(A) < 0 \),
(ii) **totally Hurwitz**, denoted by \( A ∈ T H \), if all principal submatrices of \( A \) are Hurwitz stable,
(iii) **Lyapunov diagonally stable (LDS)**, denoted by \( A ∈ \mathbb{LDS} \), if there exists a \( η ∈ \mathbb{R}^n \) such that \( \mu_{2,[η]^{1/2}}(A) < 0 \), and
(iv) **M-Hurwitz stable**, denoted by \( A ∈ MH \), if
\[ \alpha([A]_{\mathbb{M}}) < 0. \]

A matrix \( A \in \mathbb{R}^{n \times n} \) is **quasi-dominant** [39] if there exists a vector \( η ∈ \mathbb{R}^n \) such that
\[ η_i A_{ii} > \sum_{j=1, j \neq i}^{n} η_j |A_{ij}|, \quad \text{for all } i ∈ \{1, \ldots, n\}. \]

This is equivalent to \( [−A]_{\mathbb{M}} η < 0_n \), which, in turn, is equivalent (see, for example, [7, Th. 15.17]) to the inequality \( \alpha([-A]_{\mathbb{M}}) < 0 \), i.e., \( −A ∈ MH \).

The following results are essentially known in the literature, but not collected in a unified manner.

**Lemma 3 (Inclusions for classes of matrices):**

\( A ∈ MH \) implies \( A ∈ \mathbb{LDS} \), \( A ∈ \mathbb{LDS} \) implies \( A ∈ T H \), and \( A ∈ T H \) implies \( A ∈ H \).

We show that the counter-implications in Lemma 3 do not hold.

---

**IV. NOVEL LOG NORM RESULTS**

**A. Optimizing Non-Euclidean Log Norms**

First, we provide novel results on optimizing diagonal weights for \( ℓ_1 \) and \( ℓ_∞ \) log norms and provide computational methods to compute these weights.

**Theorem 5 (Quasiconvexity of \( μ \) with respect to diagonal weights):** For fixed \( A \in \mathbb{R}^{n \times n} \), consider the maps from \( \mathbb{R}^n \) to \( \mathbb{R} \) defined by
\[ η \mapsto μ_{1,[η]}(A), \quad η \mapsto μ_{∞,[η]−1}(A). \]

Then

(i) the maps in (9) are continuous, quasiconvex, and their sublevel sets are polytopes and
(ii) minimizing the maps in (9) may be executed via the minimization problems
\[ \inf_{b ∈ \mathbb{R}, η ∈ \mathbb{R}^n} b \]
\[ \text{s.t.} \quad \begin{bmatrix} 0 \\ A \end{bmatrix} η \leq b \eta \]
for \( μ_{1,[η]}(A) \) and
\[ \inf_{b ∈ \mathbb{R}, η ∈ \mathbb{R}^n} b \]
\[ \text{s.t.} \quad \begin{bmatrix} 0 \\ A \end{bmatrix} η \leq b \eta \]
for \( μ_{∞,[η]−1}(A) \).

**Remark 6:** If \( [A]_{\mathbb{M}} \) is irreducible, by Lemma 2 the optimization problems in (10) and (11) attain their minima so that the inf may be replaced by min. Then the problems may be solved by a bisection on \( b ∈ [−\|A\|, \|A\|] \), where each step of the algorithm is a LP in \( η \).

Moreover, the minima in (10) and (11) exist for many types of reducible matrices, e.g., when \( [A]_{\mathbb{M}} \) is a block-diagonal matrix whose diagonal blocks are irreducible.

In the event that the minimum does not exist, let \( b^* \) be the infimum value of either (10) or (11). Then, for any \( ε > 0 \), one can still apply the bisection algorithm to find a choice of \( η \) such that \( μ_{[η]}(A) ≤ b^* + ε \), where \( μ_{[η]}(\cdot) \) denotes either \( μ_{1,[η]}(\cdot) \) or \( μ_{∞,[η]−1}(\cdot) \).

**Remark 7:** Notice that the sets of feasible vectors \( η \) in (10) and (11) are polyhedral cones, that is, if \( η \) is feasible, then \( θη \) is also feasible for all \( θ > 0 \). Hence, the constraint \( η ∈ \mathbb{R}^n \) can be replaced by an equivalent constraint \( η ∈ [ε, ∞[^{n},\]

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where \( \varepsilon > 0 \) is an arbitrary constant. This can be useful, because LP solvers usually handle problems with nonstrict inequalities.

Next, we provide closed-form expressions for \( \ell_1 \) and \( \ell_\infty \) log norms over a certain polytope of matrices. Polytopes of interest are defined by a nominal matrix multiplied by a diagonally weighted uncertainty and shifted by an additive diagonal matrix. Such matrix polytopes arise in tests verifying the contractivity of Hopfield and firing-rate NNs and will play a critical role in our analysis.

**Lemma 8 (Max value of \( \ell_1 / \ell_\infty \) log norms under multiplicative scalings):** Any \( A \in \mathbb{R}^{n \times n} \), \( c \in \mathbb{R}^n \), \( d_1 \leq d_2 \in \mathbb{R} \), and \( \eta \in \mathbb{R}_{\geq 0}^n \) satisfy formulas (12)–(15) shown at the bottom of this page, where \( \tilde{d} = \max\{|d_1|, |d_2|\} \).

Recall that the log norm is a convex function and that the maximum value of a convex function over a polytope is achieved at one of the vertices of the polytope. In the special case in Lemma 8, formulas (12)–(15) ensure that one needs to check only 2 vertices of the polytope, rather than \( 2^n \).

Finally, we show how the optimal diagonal weights that minimize the maximum value of the log norm of a matrix polytope as in Lemma 8 can be easily computed.

**Corollary 9:** Let \( A, c, d_1, d_2 \) be as in Lemma 8. Then, for \( \mu_{\eta}(\cdot) \) denoting either \( \mu_{\ell_1,\eta}(\cdot) \) or \( \mu_{\ell_\infty,\eta}(\cdot) \) the minimax problems

\[
\inf_{\eta \in \mathbb{R}_{\geq 0}^n} \max_{c, d} \mu_{\eta}(\|c + [d]A\|) \\
\inf_{\eta \in \mathbb{R}_{\geq 0}^n} \max_{c, d} \mu_{\eta}(\|c + A[d]\|)
\]

may each be solved by a bisection algorithm, each step of which is an LP.

**Proof:** The proof is an immediate consequence of the formulas (12)–(15) shown at the bottom of this page as well as the fact that a max of quasiconvex functions is quasiconvex. Therefore, a bisection algorithm similar to the one in Theorem 5 (ii) may be used to compute the optimal \( \eta \).

\[\square\]

### B. Monotonicity of Diagonally-Weighted Log Norms

**Theorem 10 (Monotonicity of \( \alpha \) and \( \mu \)):** For any \( A \in \mathbb{R}^{n \times n} \)

1. \( \alpha(A) \leq \alpha([A]_{\text{Max}}) \),
2. for all \( p \in [1, \infty) \) and \( \eta \in \mathbb{R}_{\geq 0}^n \), we have \( \mu_{p,\eta}(A) \leq \mu_{p,\eta}([A]_{\text{Max}}) \), with equality holding for \( p \in \{1, \infty\} \), and
3. \( \mu_{p,\eta}(A) \) is monotonic, let \( \alpha_p(A) \) is an arbitrary constant. This can be useful, because an LP at every step of the bisection, while optimizing weights for the \( \ell_2 \) norm is a semidefinite program at every step, which is more computationally challenging than an LP of similar dimension.

**Example 11:** The matrix \( A_1 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \) has eigenvalues \( \{1 + i, 1 - i\} \) whereas \( [A_1]_{\text{Max}} \) has eigenvalues \( \{2, 0\} \). Therefore, \( \alpha(A_1) = 1 < 2 = \alpha([A_1]_{\text{Max}}) \). In addition, \( (A_1 + A_1^\top)/2 = I_2 \Rightarrow \mu_2(A_1) = 1 \) and \( \mu_2([A_1]_{\text{Max}}) = 2 \).

### C. Log Norms of Principal Submatrices

Given a matrix \( A \in \mathbb{R}^n \) and a nonempty index set \( I \subseteq \{1, \ldots, n\} \), let \( A_I \in \mathbb{R}^{\|I\| \times \|I\|} \) denote the principal submatrix obtained by removing the rows and columns of \( A \) which are not in \( I \). Next, given a nonempty \( I \subseteq \{1, \ldots, n\} \), define the zero-padding map \( \text{pad}_I : \mathbb{R}^{\|I\| \times \|I\|} \to \mathbb{R}^n \) as follows: \( \text{pad}_I(y) \) is obtained by inserting zeros among the entries of \( y \) corresponding to the indices in \( \{1, \ldots, n\} \setminus I \). For example, with \( n = 3 \) and \( I = \{1, 3\} \), we define \( \text{pad}_{\{1,3\}}((y_1, y_2, y_3) = (y_1, 0, y_2). \) Then, it is easy to see that given a norm \( \| \cdot \| \) on \( \mathbb{R}^n \) and nonempty \( I \subseteq \{1, \ldots, n\} \), the map \( \| \cdot \|_I : \mathbb{R}^{\|I\| \times \|I\|} \to \mathbb{R}_{\geq 0} \) defined by \( \|y\|_I = \|\text{pad}_I(y)\| \) is a norm on \( \mathbb{R}^n \).

**Lemma 12 (Norm and log norm of principal submatrices):** Assume \( \| \cdot \| \) is monotonic, let \( \mu \) and \( \mu_\eta \) denote the log norms associated to \( \| \cdot \| \) and \( \| \cdot \|_I \), respectively. Any matrix \( A \in \mathbb{R}^{n \times n} \) satisfies

1. \( \|A_I\|_I \leq \|A\| \),
2. \( \mu_\eta(A_I) \leq \mu(A) \), and
3. \( \text{if } \mu(A) < 0, \text{ then } A \in TH \).

**Corollary 13:** Suppose \( A \in \mathcal{MH} \subset \mathbb{R}^{n \times n} \). Then

i. \( A_I \in \mathcal{MH} \subset \mathbb{R}^{\|I\| \times \|I\|} \) for every nonempty \( I \subseteq \{1, \ldots, n\} \) and
ii. \( A - A_{ij}e_i e_j \in \mathcal{MH} \) for all \( i, j \in \{1, \ldots, n\}, i \neq j \), where \( e_i \) is a matrix with all zeros and unity in its \( i \)th entry.

In the context of Proposition 1, since our sufficient condition for the contractivity of the Hopfield and firing-rate NNs is \(-I_n +\)
A ∈ M H, Corollary 13 implies that this sufficient condition implies total and structural contractivity, i.e., the removal of any neuron or edge from the NN yields a new NN that remains contracting.

V. ONE-SIDED LIPSCHITZ MAPS AND NONSMOOTH CONTRACTION THEORY

A. Review of One-Sided Lipschitz Maps

We review weak pairings and one-sided Lipschitz maps as introduced in [15]; see also the earlier works [1], [51].

Definition 14 (Weak pairing): A weak pairing on R^n is a map [[ , ]]: R^n × R^n → R satisfying:
(i) Subadditivity and continuity in its first argument: 
\[ [x_1 + x_2, y] ≤ [x_1, y] + [x_2, y], \] for all x_1, x_2, y ∈ R^n and [ , ] is continuous in its first argument,
(ii) Weak homogeneity: 
\[ \alpha [x, y] = [\alpha x, y] = [x, \alpha y] \] and 
\[-[x, -y] = [x, y] \] for all x, y ∈ R^n, α ≥ 0,
(iii) Positive definiteness: 
\[ [x, x] > 0 \] for all x ≠ 0,
(iv) Cauchy–Schwarz: 
\[ [x, y] ≤ [x, x]^{1/2} [y, y]^{1/2} \] for all x, y ∈ R^n.

In addition, we say a weak pairing satisfies Deimling’s inequality if for all x, y ∈ R^n, it holds: 
\[ [x, y] ≤ \|y\| \lim_{h \to 0^+} h^{-1}(\|y + hx\| - \|y\|) \] for all x, y ∈ R^n, where \( \|\cdot\| = \left[\cdot, \cdot\right]^{1/2} \).

Deimling’s inequality is well-defined since \[\left[\cdot, \cdot\right]\] is continuous in both arguments. Then if a weak pairing \([\cdot, \cdot]\) satisfies the above inequality, it is also a one-sided Lipschitz constant of \( f \).

The relationship between weak pairings and log norms is given by Lumer’s equality.

Lemma 15 (Lumer’s equality [15, Th. 18]): Let \( \|\cdot\| \) be a norm on R^n with compatible weak pairing \([\cdot, \cdot]\). Then
\[ \mu(A) = \sup_{x \neq 0} \frac{[Ax, x]}{\|x\|^2}, \] for all A ∈ R^n×n.

Definition 16 (One-sided Lipschitz maps [15, Def. 26]): Consider \( f : U \to \mathbb{R}^n \) where U ⊆ R^n is open and connected. We say f is one-sided Lipschitz with respect to a weak pairing \([\cdot, \cdot]\) if there exists \( b \in \mathbb{R} \) such that
\[ [f(x) - f(y), x - y] ≤ b \|x - y\|^2, \] for all x, y ∈ U.

We say b is a one-sided Lipschitz constant of f. Moreover, the minimal one-sided Lipschitz constant of f is
\[ \text{osL}(f) := \sup_{x, y \in U, x \neq y} \frac{[f(x) - f(y), x - y]}{\|x - y\|^2}. \]

If f is continuously differentiable and U is convex, it can be shown that \( \text{osL}(f) = \sup_{x \in U} \mu(Df(x)) \), where \( Df := \frac{\partial f}{\partial x} \) is the Jacobian matrix of f.

A vector field \( f : \mathbb{R}^n \to \mathbb{R}^n \) satisfying \( \text{osL}(f) \leq -c < 0 \) is said to be strongly infinitesimally contracting with rate c. A consequence of \( \text{osL}(f) \leq -c < 0 \) is that the function \( V(x, y) = \|x - y\| \) serves as an incremental Lyapunov function [3] establishing incremental exponential stability, i.e., any two trajectories \( x(t), y(t) \) satisfying \( \dot{x} = f(x) \) and \( \dot{y} = f(y) \) additionally satisfy \( \|x(t) - y(t)\| ≤ e^{-ct}\|x(0) - y(0)\| \) for all \( t ≥ 0 \) [15, Th. 31]. Moreover, if f is continuous, then all solutions converge to a unique equilibrium. Thus, in order to establish contractivity of NN dynamics, it is equivalent to establish that their one-sided Lipschitz constants are negative.

B. Nonsmooth Contraction Theory

In this section, we consider locally Lipschitz f and show that in this case, the definition of osL \( f \) does not depend on the weak pairing and instead depends only on the norm through the log norm.

Theorem 17 (osL simplification for locally Lipschitz f): Let \( f : U \to \mathbb{R}^n \) be locally Lipschitz on an open convex set, \( U ⊆ \mathbb{R}^n \). Then, for every c ∈ R, the following statements are equivalent:

(i) \( \text{osL}(f) ≤ c \), and
(ii) \( \mu(Df(x)) ≤ c \) for almost every \( x ∈ U \).

Specifically, \( \text{osL}(f) = \text{ess sup}_{x ∈ U} \mu(Df(x)) \), where ess sup denotes the essential supremum.

Recall that \( Df(x) \) exists for almost every \( x ∈ U \) by Rademacher’s theorem and thus the essential supremum ignores the Lebesgue measure zero set where \( Df \) does not exist.

Theorem 17 demonstrates that locally Lipschitz f enjoy a similar simplification in the osL definition as do continuously differentiable functions.

In NN models, nonsmooth activation functions, such as ReLU, LeakyReLU, and nonsmooth saturation functions, are prevalent; Theorem 17 allows us to use standard log norm results to analyze these models. In other words, for a given continuous-time NN dynamics \( \dot{x} = f_{\text{NN}}(x) \), with locally Lipschitz \( f_{\text{NN}} \), to establish contractivity, it suffices to verify that \( \mu(Df_{\text{NN}}(x)) ≤ -c \) for almost every x.

VI. CONTRACTING NN DYNAMICS

In this section, we prove Proposition 1 in greater generality. Specifically, we establish tight estimates for the one-sided Lipschitz constant for both the Hopfield and firing-rate NNs with respect to both diagonally-weighted \( \ell_1 \) and \( \ell_∞ \) norms. In instances where the one-sided Lipschitz constant is negative, we conclude that the NN is strongly infinitesimally contracting. Beyond the Hopfield and firing-rate NNs, we also study the non-Euclidean contractivity of other classes of NNs.

A. One-Sided Lipschitz Characterization of Hopfield NNs

Recall the Hopfield NN dynamics, first introduced in [26]
\[ \dot{x} = -C x + A \Phi(x) + u =: f_H(x) \]
where \( C ∈ \mathbb{R}^{n×n} \) is a positive semidefinite diagonal matrix, \( A ∈ \mathbb{R}^{n×n} \) is arbitrary, \( u ∈ \mathbb{R}^n \) is a constant input, and \( \Phi : \mathbb{R}^n → \mathbb{R}^n \) is an activation function. We make the following assumption on our activation functions.

Assumption 2 (Activation functions): Activation functions are locally Lipschitz and diagonal, i.e., \( \Phi(x) = (\phi_1(x_1), \ldots, \phi_n(x_n)) \) where each \( \phi_i : \mathbb{R} → \mathbb{R} \) satisfies
\[ d_1 := \inf_{x, y ∈ R, x \neq y} \frac{\phi_i(x) - \phi_i(y)}{x - y} > -∞ \]
\[
d_2 := \sup_{x, y \in \mathbb{R}, x \neq y} \frac{\phi_i(x) - \phi_i(y)}{x - y},
\]
(19)

When \( \phi_i \) satisfies (19) with finite \( d_2 \), we write \( \phi_i \in \text{slope}\{d_1, d_2\} \). If \( d_2 = \infty \), we write \( \phi_i \in \text{slope}\{d_1, \infty\} \).

Assumption 2 with finite \( d_2 \) implies that activation functions are (globally) Lipschitz and that \( \phi_i'(x) \in \{d_1, d_2\} \) for almost every \( x \in \mathbb{R} \). Many common activation functions satisfy these assumptions including ReLU, tanh, and sigmoids. If, instead, \( d_2 = \infty \), we can consider locally Lipschitz activation functions with unbounded slope including rectified polynomials \( \phi(x) = \max\{x, 0\}^r \) for \( r \in \mathbb{Z}_{\geq 0} \) which have been studied in [34]. Note that compared to Assumption 1, our activation functions do not need to be differentiable and are permitted more arbitrary bounds on their slopes.

The following theorem is the counterpart to Proposition 1 (i) under more general assumptions.

**Theorem 18 (\( \ell_1 \) one-sided Lipschitzness of Hopfield NN):** Consider the Hopfield NN (18) with each \( \phi_i \in \text{slope}\{d_1, d_2\} \). Then, the following holds.

(i) For arbitrary \( \eta \in \mathbb{R}_{\geq 0} \), \( \text{osL}_{1,|\eta|}(f_H) = \max \{ \mu_{1,|\eta|}(\cdot - C + d_1A), \mu_{1,|\eta|}(\cdot - C + d_2A) \} \).

(ii) The vector \( \eta \) minimizing \( \text{osL}_{1,|\eta|}(f_H) \) is the solution to

\[
\inf_{b \in \mathbb{R}, \eta \in \mathbb{R}_{\geq 0}} T
\]

subject to

\[
\begin{align*}
-C + [d_1A]_{\text{Mtr}}^\top \eta & \leq b \\
-C + [d_2A]_{\text{Mtr}}^\top \eta & \leq b
\end{align*}
\]

and if the infimum value is attained at parameter values \( b^*, \eta^* \), then \( \text{osL}_{1,|\eta^*|}(f_H) = b^* \).

Further, suppose \([A]_{\text{Mtr}}\) is irreducible. Then, the following holds.

(iii) If \( C = cI_n \) and \( d_1 \geq 0 \), then, with \( w_A \in \mathbb{R}_{\geq 0} \) being the left dominant eigenvector of \([A]_{\text{Mtr}}\)
\[
\inf_{\eta \in \mathbb{R}_{\geq 0}} \text{osL}_{1,|\eta|}(f_H) = \text{osL}_{1,|\eta|}(f_H) = -c + \max \{ d_1 \alpha([A]_{\text{Mtr}}), d_2 \alpha([A]_{\text{Mtr}}) \},
\]
(20)

(iv) If \( d_1 = 0 \) and \( C \gtrsim 0 \), then, with \( w_A \in \mathbb{R}_{\geq 0} \) being the left dominant eigenvector of \([A]_{\text{Mtr}}\)
\[
\inf_{\eta \in \mathbb{R}_{\geq 0}} \text{osL}_{1,|\eta|}(f_H) = \text{osL}_{1,|\eta|}(f_H) = \max \{ \alpha(-C), \alpha(-C + d_2[A]_{\text{Mtr}}) \}.
\]

In particular, Theorem 18 provides exact values for the minimal one-sided Lipschitz constant of the Hopfield NN with respect to diagonally-weighted \( \ell_1 \) norms.

As a consequence of this theorem, suppose the inf in statement (ii) is attained and let \( b^*, \eta^* \) be the optimal parameters for the LP. If \( b^* < 0 \), then the Hopfield NN (18) is strongly infinitesimally contracting with rate \( |b^*| \) with respect to \( ||\cdot||_{1,|\eta^*|} \). Note, in particular, that if \( d_1 = 0, d_2 = 1, C = I_n \), and \( \alpha([A]_{\text{Mtr}}) < 1 \), statement (iii) is equivalent to the statement in Proposition 1 (i).

**Proof of Theorem 18:** Regarding statement (i), for any \( \eta \in \mathbb{R}_{\geq 0} \)
\[
\text{osL}_{1,|\eta|}(f_H) = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \mu_{1,|\eta|}(Df_H(x))
\]
\[
= \sup_{x \in \mathbb{R}^n \setminus \{0\}} \mu_{1,|\eta|}(-C + A D\Phi(x))
\]
\[
= \max_{d \in \{d_1, d_2\}^n} \mu_{1,|\eta|}(-C + A[d])
\]
\[
= \max \{ \mu_{1,|\eta|}(-C + d_1A), \mu_{1,|\eta|}(-C + d_2A) \}
\]
where the second-to-last equality holds by Assumption 2 and the last equality holds by Lemma 8.

Statement (ii) holds by Corollary 9. Regarding statement (iii), if \( C = cI_n \) and \( d_1 \geq 0 \), then
\[
\text{osL}_{1,|\eta|}(f_H) = -c + \max \{ \mu_{1,|\eta|}(d_1A), \mu_{1,|\eta|}(d_2A) \}
\]
\[
= -c + \max \{ d_1 \alpha([A]_{\text{Mtr}}), d_2 \alpha([A]_{\text{Mtr}}) \}.
\]

In addition, recall that \( \eta = w_A \) is the optimal weight from Lemma 2 for the irreducible Metzler matrix \([A]_{\text{Mtr}}\) with respect to \( p = 1 \). Therefore
\[
\inf_{\eta \in \mathbb{R}_{\geq 0}} \text{osL}_{1,|\eta|}(f_H) = \text{osL}_{1,|\eta|}(f_H)
\]
\[
= -c + \max \{ d_1 \alpha([A]_{\text{Mtr}}), d_2 \alpha([A]_{\text{Mtr}}) \}
\]

Regarding statement (iv), if \( d_1 = 0 \) and \( C \gtrsim 0 \), we compute
\[
\text{osL}_{1,|\eta|}(f_H) = \max \{ \mu_{1,|\eta|}(-C), \mu_{1,|\eta|}(-C + d_2A) \}
\]
\[
= \max \{ \alpha(-C), \mu_{1,|\eta|}(-C + d_2A) \}
\]
which holds because \( \mu_{1,|\eta|}(-C) = \max_{i \in \{1, \ldots, n\}} -c_{ii} = \alpha(-C) \) for every \( \eta \in \mathbb{R}_{\geq 0} \). In addition, we have that \( \eta = w_A \) is the optimal weight for the irreducible Metzler matrix \(-C + d_2A\) by Lemma 2. Thus
\[
\inf_{\eta \in \mathbb{R}_{\geq 0}} \text{osL}_{1,|\eta|}(f_H) = \max \{ \alpha(-C), \alpha(-C + d_2A) \}
\]

which proves the result. \( \square \)

**Remark 19 (Comparison to [33, Th. 1], [8, Exercise 2.22]):** For \( A \in \mathbb{R}^{n \times n} \), define its nonnegative Metzler majorant \([A]_{\text{Mtr}}^+ \) by
\[
([A]_{\text{Mtr}}^+)^{ij} = \begin{cases} 
\max\{A_{ij}, 0\}, & \text{if } i = j, \\
[A_{ij}], & \text{if } i \neq j.
\end{cases}
\]

In [33, Th. 1], for \( d_1 = 0 \) and \( C = I_n \) it is shown that if \( \alpha(-I_n + d_2[A]_{\text{Mtr}}^+) < 0 \), then the Hopfield NN (18) is contracting with respect to a diagonally weighted \( \ell_2 \) norm which is given in Lemma 2. Compared to the condition \( \alpha(-I_n + [A]_{\text{Mtr}}^+) < 0 \), the condition in Theorem 18 (iv) replaces \([A]_{\text{Mtr}}\) with \([A]_{\text{Mtr}}^+ \) and thus guarantees that a larger class of synaptic matrices still guarantee contractivity of the Hopfield NN.

In addition, beyond Proposition 1 (i), we characterize the \( \ell_1 \) one-sided Lipschitz constant of the Hopfield NN in the following theorem.

**Theorem 20 (\( \ell_1 \) one-sided Lipschitzness of Hopfield NN):** Consider the Hopfield NN (18) with each \( \phi_i \in \text{slope}\{d_1, d_2\} \). Let \( \tilde{d} = \max\{|d_1|, d_2\} \). Then

(i) For arbitrary \( \eta \in \mathbb{R}_{\geq 0} \), \( \text{osL}_{\infty,|\eta|^{-1}}(f_H) = \max \{ \mu_{\infty,|\eta|^{-1}}(-C + \tilde{d}A - (\tilde{d} - d_1)(I_n \circ A)), \mu_{\infty,|\eta|^{-1}}(-C + \tilde{d}A - (\tilde{d} - d_2)(I_n \circ A)) \} \)

(ii) the vector \( \eta \) minimizing \( \text{osL}_{\infty,|\eta|^{-1}}(f_H) \) is the solution to
\[
\inf_{b \in \mathbb{R}, \eta \in \mathbb{R}^n_{>0}} \ b \\
\text{s.t.} \\
(C + [\bar{d}A - (\bar{d} - d_1)(I_n \circ A)]_{\text{Mtr}}) \eta \leq b \eta \\
(C + [\bar{d}A - (\bar{d} - d_2)(I_n \circ A)]_{\text{Mtr}}) \eta \leq b \eta \\
\] 
and if the infimum value is attained at parameter values \(b^*, \eta^*\), then \(\text{osL}_{\infty,[\eta^*]}^{-1}(f_{\text{FR}}) = b^*\).

**Proof:** Regarding statement (i), in analogy to the proof of Theorem 18 (i), we have
\[
\text{osL}_{\infty,[\eta]}^{-1}(f_{\text{FR}}) = \max_{d \in [d_1, d_2]^n} \mu_{\infty,[\eta]}^{-1}(-C + A[d]) \\
= \max\{\mu_{\infty,[\eta]}^{-1}(-C + \bar{d}A - (\bar{d} - d_1)(I_n \circ A)), \mu_{\infty,[\eta]}^{-1}(-C + \bar{d}A - (\bar{d} - d_2)(I_n \circ A))\}
\]
where the final equality is by Lemma 8. Statement (ii) is then a consequence of Corollary 9. \(\square\)

**B. One-Sided Lipschitz Characterization of Firing-Rate NNs**

Recall the firing-rate NN dynamics
\[
\dot{x} = -Cx + \Phi(Ax + u) =: f_{\text{FR}}(x).
\] (22)
The interpretation for this name is that if \(\Phi(x)\) is nonnegative for all \(x \in \mathbb{R}^n\) (as is ReLU), then the positive orthant is forward-invariant and \(x\), interpreted as a vector of firing-rates, while in the Hopfield NN, \(x\) can be negative and is thus interpreted as a vector of membrane potentials.

The following two theorems are generalizations of Proposition 1 (ii) under more general assumptions. Specifically, Theorem 21 characterizes one-sided Lipschitzness of the firing-rate NN with respect to diagonally, weighted \(\ell_1\) norms, while Theorem 22 does the same with respect to diagonally-weighted \(\ell_1\) norms.

**Theorem 21 \((\ell_\infty \text{ one-sided Lipschitzness of firing-rate NN}):**
Consider the firing-rate NN (22) with each \(\phi_i \in \text{slope}[d_1, d_2]\) and invertible \(A\). Then, the following holds:

(i) for arbitrary \(\eta \in \mathbb{R}^n_{0,0}\), \(\text{osL}_{\infty,[\eta]}^{-1}(f_{\text{FR}}) = \max\{\mu_{\infty,[\eta]}^{-1}(-C + d_1 A), \mu_{\infty,[\eta]}^{-1}(-C + d_2 A)\};
(ii) the choice of \(\eta\) minimizing \(\text{osL}_{\infty,[\eta]}^{-1}(f_{\text{FR}})\) is the solution to
\[
\inf_{b \in \mathbb{R}, \eta \in \mathbb{R}^n_{>0}} \ b \\
\text{s.t.} \\
(-C + [d_1 A])_{\text{Mtr}} \eta \leq b \eta \\
(-C + [d_2 A])_{\text{Mtr}} \eta \leq b \eta \\
\] 
and if the infimum value is attained at parameter values \(b^*, \eta^*\), then \(\text{osL}_{\infty,[\eta^*]}^{-1}(f_{\text{FR}}) = b^*\).

Further, suppose that \([A]_{\text{Mtr}}\) is irreducible. Then, the following holds:

(iii) if \(C = cI_n\) and \(d_1 \geq 0\), then, with \(v_A \in \mathbb{R}^n_{\geq 0}\) being the right dominant eigenvector of \([A]_{\text{Mtr}}\)
\[
\inf_{\eta \in \mathbb{R}^n_{>0}} \text{osL}_{\infty,[\eta]}(f_{\text{FR}}) = \text{osL}_{\infty,[v_A]}^{-1}(f_{\text{FR}}) \\
= -c + \max\{d_1 \alpha([A]_{\text{Mtr}}), d_2 \alpha([A]_{\text{Mtr}})\};
\] (23)
(iv) if \(d_1 = 0\) and \(C \succ 0\), then, with \(v_A \in \mathbb{R}^n_{\geq 0}\) being the right dominant eigenvector of \(-C + d_2 [A]_{\text{Mtr}}\)
\[
\inf_{\eta \in \mathbb{R}^n_{>0}} \text{osL}_{\infty,[\eta]}(f_{\text{FR}}) = \text{osL}_{\infty,[v_A]}^{-1}(f_{\text{FR}}) \\
= \max\{\alpha(-C), \alpha(-C + d_2 [A]_{\text{Mtr}})\}. \] (24)

**Proof:** Regarding statement (i), for any \(\eta \in \mathbb{R}^n_{0,0}\) we compute
\[
\text{osL}_{\infty,[\eta]}^{-1}(f_{\text{FR}}) = \sup_{x \in \mathbb{R}^n \setminus \Omega_{\text{FR}}} \mu_{\infty,[\eta]}^{-1}(-C + D\Phi(Ax + u)A) \\
= \sup_{x \in \mathbb{R}^n \setminus \Omega_{\text{FR}}} \mu_{\infty,[\eta]}^{-1}(-C + [d]A) \\
= \max\{\mu_{\infty,[\eta]}^{-1}(-C + d_1 A), \mu_{\infty,[\eta]}^{-1}(-C + d_2 A)\}
\]
where the second-to-last equality holds by Assumption 2 and because \(A\) is invertible. The last equality holds by Lemma 8.

Statement (ii) is a consequence of Corollary 9. Regarding statement (iii), if \(C = cI_n\) and \(d_1 \geq 0\), then
\[
\text{osL}_{\infty,[\eta]}^{-1}(f_{\text{FR}}) = -c + \max\{\mu_{\infty,[\eta]}^{-1}(d_1 A), \mu_{\infty,[\eta]}^{-1}(d_2 A)\} \\
= -c + \max\{d_1 \mu_{\infty,[\eta]}^{-1}(A), d_2 \mu_{\infty,[\eta]}^{-1}(A)\}.
\]
In addition, recall that \(\eta = \nu_A\) is the optimal weight for Lemma 2 for the irreducible Metzler matrix \([A]_{\text{Mtr}}\) with respect to \(p = \infty\). Therefore
\[
\inf_{\eta \in \mathbb{R}^n_{>0}} \text{osL}_{\infty,[\eta]}^{-1}(f_{\text{FR}}) = \text{osL}_{\infty,[v_A]}^{-1}(f_{\text{FR}}) \\
= -c + \max\{d_1 \mu_{\infty,[\eta]}^{-1}(A), d_2 \mu_{\infty,[\eta]}^{-1}(A)\} \\
= -c + \max\{d_1 \alpha([A]_{\text{Mtr}}), d_2 \alpha([A]_{\text{Mtr}})\}.
\]

Regarding statement (iv), if \(d_1 = 0\) and \(C \succ 0\), we compute
\[
\text{osL}_{\infty,[\eta]}^{-1}(f_{\text{FR}}) = \max\{\mu_{\infty,[\eta]}^{-1}(-C), \mu_{\infty,[\eta]}^{-1}(-C + d_2 A)\} \\
= \max\{\alpha(-C), \mu_{\infty,[\eta]}^{-1}(-C + d_2 A)\}
\]
which holds because \(\mu_{\infty,[\eta]}^{-1}(-C) = \max_{i \in \{1, \ldots, n\}} -c_i = \alpha(-C)\) for every \(\eta \in \mathbb{R}^n_{0,0}\). In addition, we have that \(\eta = v_A\) is the optimal weight for the irreducible Metzler matrix \(-C + d_2 [A]_{\text{Mtr}}\) by Lemma 2. Thus
\[
\inf_{\eta \in \mathbb{R}^n_{>0}} \text{osL}_{\infty,[\eta]}^{-1}(f_{\text{FR}}) = \text{osL}_{\infty,[v_A]}^{-1}(f_{\text{FR}}) \\
= \max\{\alpha(-C), \alpha(-C + d_2 [A]_{\text{Mtr}})\}
\]
which proves the result. \(\square\)

**Theorem 22 \((\ell_1 \text{ one-sided Lipschitzness of firing-rate NN}):**
Consider the firing-rate NN (22) with each \(\phi_i \in \text{slope}[d_1, d_2]\), and invertible \(A\). Let \(\bar{d} = \max\{[d_1], [d_2]\}\). Then

(i) for arbitrary \(\eta \in \mathbb{R}^n_{0,0}\), \(\text{osL}_{1,[\eta]}(f_{\text{FR}}) = \max\{\mu_{1,[\eta]}(-C + \bar{d}A - (\bar{d} - d_1)(I_n \circ A)), \mu_{1,[\eta]}(-C + \bar{d}A - (\bar{d} - d_2)(I_n \circ A))\}.
(ii) the vector \(\eta\) minimizing \(\text{osL}_{1,[\eta]}(f_{\text{FR}})\) is the solution to
\[
\inf_{b \in \mathbb{R}, \eta \in \mathbb{R}^n_{0,0}} \ b \\
\text{s.t.} \\
(C + [\bar{d}A - (\bar{d} - d_1)(I_n \circ A)]_{\text{Mtr}})^\top \eta \leq b \eta \\
(C + [\bar{d}A - (\bar{d} - d_2)(I_n \circ A)]_{\text{Mtr}})^\top \eta \leq b \eta
\]
and if the infimum value is attained at parameter values \( b^*, \eta^* \), then \( \text{osL}_{1, \eta^*}(f_{\text{FR}}) = b^* \).

**Proof:** Regarding statement (i), in analogy to the proof of Theorem 21 (i), we have

\[
\text{osL}_{1, \eta}(f_{\text{FR}}) = \max_{d \in [d_1, d_2]^n} \mu_{1, \eta}(C + [d] A) = \max_{i \in \{1, \ldots, n\}} (-C + d_1) (I_n \circ A)
\]

where the final equality is by Lemma 8. Statement (ii) is then a consequence of Corollary 9. □

**Remark 23:** For invertible \( A \in \mathbb{R}^{n \times n} \), Theorems 21 (i) and 22 (i) provide an exact value for the minimal one-sided Lipschitz constant of the firing rate model with respect to a given norm. If \( A \) is not invertible, then the closure of the image of the map \( x \mapsto D\Phi(Ax + u) \) may not contain all the vertices of the set \([d_1, d_2]^n\). For noninvertible \( A \) and arbitrary \( \eta \in \mathbb{R}^n_{\geq 0} \), the values presented in these theorems are instead upper bounds on the minimal one-sided Lipschitz constant.

In Fig. 1, we plot the phase portrait of a 2-D firing-rate NN (22), along with level sets of the corresponding Lyapunov function. We highlight the utility of optimizing the weight of the \( \ell_\infty \) norm. Namely, although the firing-rate NN example is not contracting with respect to the \( \ell_\infty \) norm, it is contracting with respect to a weighted \( \ell_\infty \) norm, where the optimal diagonal weight is \( [\eta^*]^{-1} \), where \( \eta^* \) is the right dominant eigenvector of \( [A]_{\text{Mfr}} \).

**C. Contractivity of Hopfield and Firing-Rate NNs With Unbounded Slope**

In the spirit of the classic work [21], which studies Hopfield NNs, which have monotone activation functions with unbounded slope, we present the following result on the contractivity of Hopfield and firing-rate NNs when \( \phi_1 \) is in the set \( [d_1, \infty) \).

**Theorem 24** (Contractivity under unbounded slope): Consider the Hopfield NN (18) and firing-rate NN (22) with \( \phi_1 \) in the set \( [d_1, \infty) \) and irreducible \( [A]_{\text{Mfr}} \) with dominant left and right eigenvectors \( w_A, v_A \), respectively, and suppose that

- (A1) \( A \in \mathcal{M}H \) and
- (A2) \( A \in \mathbb{R}^{n \times n}, C \geq 0 \), and \( d_1 \in \mathbb{R} \) satisfy

\[
-\alpha(C) + \max_{i} \{d_i, 0\} \alpha([A]_{\text{Mfr}}) > -|d_1| - d_1 \min_{i \in \{1, \ldots, n\}} A_{ii}
\]

Then

(i) the Hopfield NN (18) is strongly infinitesimally contracting with respect to \( \| \cdot \|_{1, [w_A]} \) with rate \( -\alpha(C) + \max_{i} \{d_i, 0\} \alpha([A]_{\text{Mfr}}) + |d_1| - d_1 \min_{i \in \{1, \ldots, n\}} A_{ii} > 0 \) and

(ii) the firing-rate NN (22) is strongly infinitesimally contracting with respect to \( \| \cdot \|_{\infty, [v_A]} \) with rate \( -\alpha(C) + \max_{i} \{d_i, 0\} \alpha([A]_{\text{Mfr}}) + |d_1| - d_1 \min_{i \in \{1, \ldots, n\}} A_{ii} > 0 \).

**Proof:** Regarding statement (i), we adopt the shorthand \( r_i := A_{ii} + \sum_{j \neq i} |A_{ij}|/(w_A)_{ij} \). Then, we observe that \( A \in \mathcal{M}H \) implies that for every \( i \in \{1, \ldots, n\}, r_i \leq \alpha([A]_{\text{Mfr}}) < 0 \). Then, for every \( x \in \mathbb{R}^n \setminus \Omega_{fr} \)

\[
\mu_{1, [w_A]}(Df_{\text{H}}(x)) = \mu_{1, [w_A]}(-C + AD\Phi(x)) = \max_{i \in \{1, \ldots, n\}} (-C + d_1) (I_n \circ A)
\]

where inequality \( \star \) holds because \( r_i < 0 \) and \( A_{ii} < 0 \) for all \( i \). This inequality holds for all \( x \in \mathbb{R}^n \setminus \Omega_{fr} \), we conclude that \( \text{osL}_{1, [w_A]}(f_{\text{FR}}) \leq \alpha(C) + \max_{i} \{d_i, 0\} \alpha([A]_{\text{Mfr}}) - |d_1| - d_1 \min_{i \in \{1, \ldots, n\}} A_{ii} \), which implies the result. The proof of statement (ii) is essentially identical and thus omitted. □

**Remark 25:** Note that in Theorem 24

(i) if \( d_1 \geq 0 \), then condition (A1) immediately implies condition (A2). Hence, \( A \in \mathcal{M}H \) is a sufficient condition for the strong infinitesimal contractivity of (18) and (22) with unbounded-slope monotonic activation functions and

(ii) alternatively if \( \alpha([A]_{\text{Mfr}}) = 0 \) and \( C > 0 \), then the condition \( -\alpha(C) > -|d_1| - d_1 \min_{i \in \{1, \ldots, n\}} A_{ii} \) is a sufficient condition for the strong infinitesimal contractivity of (18) and (22). Note that in this case, we only need \( [A]_{\text{Mfr}} \) to be marginally stable.
D. Contractivity of Other Continuous-Time NNs

We apply Theorem 18 and the log norm results in Section IV to the following related neural circuit models, all of which are studied in the classic book [32]. In the following theorems, we assume all Metzler matrices are irreducible.

Theorem 26 (Contractivity of special Hopfield models): (i) If $A \in \mathcal{MH}$, and $d_1 > 0$, the Persidskii-type$^3$ model
\[ \dot{x} = A \Phi(x) \]
with each $\phi_i \in \text{slope}[d_1, d_2]$ is strongly infinitesimally contracting with respect to norm $\| \cdot \|_{1,\{w_i\}}$ with rate $d_1 \alpha([A]_{\text{Mzr}})$.

(ii) If $-C + d_2 A \in \mathcal{MH}$, the Hopfield NN (18) with $d_1 = 0$ and positive diagonal $C$ is strongly infinitesimally contracting with respect to $\| \cdot \|_{1,\{w_i\}}$ with rate $-\max\{\alpha(-C), \alpha(-C + d_2 [A]_{\text{Mzr}})\} > 0$.

Proof: Regarding statement (i), let $f_\Phi(x) := A \Phi(x)$. By Theorem 18 (iii) with $c = 0$, $\text{osL}_{1,\{w_i\}}(f_\Phi) = \max\{d_1 \alpha([A]_{\text{Mzr}}), d_2 \alpha([A]_{\text{Mzr}})\}$. However, since $A \in \mathcal{MH}$, $\alpha([A]_{\text{Mzr}}) < 0$, so $\text{osL}_{1,\{w_i\}}(f_\Phi) = d_1 \alpha([A]_{\text{Mzr}})$. Thus, the Persidskii-type model is strongly infinitesimally contracting with respect to norm $\| \cdot \|_{1,\{w_i\}}$ with rate $d_1 \alpha([A]_{\text{Mzr}})$.

Regarding statement (ii), by Theorem 18 (iv)
\[ \text{osL}_{1,\{w_i\}}(f_{\text{HH}}) = \max\{\alpha(-C), \alpha(-C + d_2 [A]_{\text{Mzr}})\}. \]
In particular, since $-C + d_2 A \in \mathcal{MH}$, $\alpha(-C + d_2 [A]_{\text{Mzr}}) < 0$ and since $C$ is positive diagonal, we have $\text{osL}_{1,\{w_i\}}(f_{\text{HH}}) < 0$ so that the Hopfield NN is strongly infinitesimally contracting with respect to $\| \cdot \|_{1,\{w_i\}}$ with rate $-\max\{\alpha(-C), \alpha(-C + d_2 [A]_{\text{Mzr}})\} > 0$. \[ \square \]

Theorem 27: From [32, Th. 3.2.4], consider
\[ \dot{x} = Ax - C \Phi(x) \]
with diagonal $C \succeq 0$ and each $\phi_i \in \text{slope}[d_1, d_2]$. If $A - d_1 C \in \mathcal{MH}$ with corresponding dominant left eigenvector $w_{\text{B}},$ then this model is strongly infinitesimally contracting with respect to $\| \cdot \|_{1,\{w_i\}}$ with rate $-\alpha([A]_{\text{Mzr}} - d_1 C)$.

Proof: We compute the one-sided Lipschitz constant of $f(x) := Ax - C \Phi(x)$ with respect to norm $\| \cdot \|_{1,\{w_i\}}$
\[ \text{osL}_{1,\{w_i\}}(f) = \sup_{x \in \mathbb{R}^n} \mu_{1,\{w_i\}}(Df(x)) = \sup_{x \in \mathbb{R}^n} \mu_{1,\{w_i\}}(A - CD \Phi(x)) \]
\[ = \max_{d \in [d_1, d_2]} \mu_{1,\{w_i\}}(A - C[d]) = \mu_{1,\{w_i\}}(A - d_1 C) \]
\[ = \mu_{1,\{w_i\}}([A]_{\text{Mzr}} - d_1 C) \]
where the equality $\ast$ is due to Assumption 2, equality $\ast$ is because $C \succeq 0$, equality $\ast$ is by Theorem 10 (ii), and $\ast$ is by Lemma 2. Moreover, since $A - d_1 C \in \mathcal{MH}$, $\alpha([A]_{\text{Mzr}} - d_1 C) < 0$ so $f$ is strongly infinitesimally contracting with respect to $\| \cdot \|_{1,\{w_i\}}$ with rate $-\alpha([A]_{\text{Mzr}} - d_1 C) > 0$. \[ \square \]

Theorem 28: From [32, Th. 3.2.10], consider
\[ \dot{x}_i = \sum_{j=1}^n A_{ij} \phi_{ij}(x_j) \]
for each $i \in \{1, \ldots, n\}$ and with each $\phi_{ij} \in \text{slope}[d_1, d_2]$. If $d_1 > 0$ and
\[ B := d_2 A - (d_2 - d_1)(I_n \circ A) \in \mathcal{MH} \]
with corresponding dominant left and right eigenvectors $w_{\text{B}}, v_{\text{B}},$ respectively, then this model is strongly infinitesimally contracting with rate $-\alpha([B]_{\text{Mzr}}) > 0$ with respect to both $\| \cdot \|_{1,\{w_i\}}$ and $\| \cdot \|_{\infty,\{w_i\}}^{-1}$. \[ \square \]

Proof: First note that the assumption $B \in \mathcal{MH}$ implies that $A_{i_0} < 0$ for every $i_0 \in \{1, \ldots, n\}$ since the diagonal elements of $B$ are $d_1 A_{i_0}$ and a necessary condition for $B \in \mathcal{MH}$ is $B_{ii} < 0$ since $\mathcal{MH} \subset \mathcal{T}_H$. Let $f$ denote the vector field given by $f(x) = \sum_{j=1}^n A_{ij} \phi_{ij}(x_j)$.

We compute $(Df(x))_{ij} = \frac{\partial}{\partial x_j} \sum_{j=1}^n A_{ij} \phi_{ij}(x_j) = A_{ij} \phi_{ij}'(x_j)$ for almost every $x \in \mathbb{R}^n$. In other words, $Df(x) = A \circ D \Phi(x)$, for almost every $x \in \mathbb{R}^n$, where $(D \Phi(x))_{ij} = \phi_{ij}'(x_j)$. We now proceed to element-wise upper bound $(Df(x))_{\text{Mzr}}$. Observe that for every $i \neq j \in \{1, \ldots, n\}$
\[ (Df(x))_{\text{Mzr}} = \mu_{1,\{w_i\}}(Df(x)) = \mu_{1,\{w_i\}}(A_{ij} \phi_{ij}'(x_j)) \]
where the second inequality holds because $A_{ij} < 0$ for every $i \in \{1, \ldots, n\}$. Now observe that for any matrix $A \in \mathbb{R}^{n \times n}$, if $[A]_{\text{Mzr}} \leq A'$ elementwise, then both $\mu_{1,\{w_i\}}(A) \leq \mu_{1,\{w_i\}}(A')$ and $\mu_{\infty,\{w_i\}}(A) \leq \mu_{\infty,\{w_i\}}(A')$ hold for any $\eta \in \mathbb{R}_+^n$. Then, we can observe that
\[ \text{osL}_{1,\{w_i\}}(f) = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \mu_{1,\{w_i\}}(Df(x)) = \mu_{1,\{w_i\}}([Df(x)]_{\text{Mzr}}) \leq \mu_{1,\{w_i\}}([B]_{\text{Mzr}}) = \mu_{1,\{w_i\}}([B]_{\text{Mzr}}) = \alpha([B]_{\text{Mzr}}) \]
where the final equality holds by Lemma 2. An analogous computation shows that $\text{osL}_{\infty,\{w_i\}}(f) \leq \alpha([B]_{\text{Mzr}})$. As a consequence, since $B \in \mathcal{MH}$, this model is strongly infinitesimally contracting with respect to both $\| \cdot \|_{1,\{w_i\}}$ and $\| \cdot \|_{\infty,\{w_i\}}$ with rate $-\alpha([B]_{\text{Mzr}}) > 0$. \[ \square \]

The next two theorems serve as non-Euclidean versions of early results on contractivity of Lur’e systems (in application to the entrainment problem) established first in [57].

Theorem 29 (Contractivity of Lur’e system): From [32, Th. 3.2.7], consider the Lur’e system
\[ \dot{x} = Ax + v(x) \]
\[ y = w^\top x \]
where $A \in \mathbb{R}^{n \times n}$, $v, w \in \mathbb{R}^n$ and $\phi \in \text{slope}[d_1, d_2]$. Consider the following two minimization problems:
\[ \inf_{b \in \mathbb{R}^n, \eta \in \mathbb{R}_+^n} [A + d_1 w^\top w_{\text{Mzr}}] \eta \leq b \eta \]
\[ [A + d_2 w^\top w_{\text{Mzr}}] \eta \leq b \eta \] (25)

---

3See [32, Definition 3.2.1]
Let $b^*$ and $c^*$ be infimum values for (25) and (26), respectively. Then

\begin{enumerate}
\item if $b^* < 0$, then for every $\varepsilon \in [0,|b^*|]$, there exists $\eta \in \mathbb{R}_{>0}$ such that the closed-loop dynamics are strongly infinitesimally contracting with rate $|b^*| - \varepsilon > 0$ with respect to $\| \cdot \|_{1,[\varepsilon]}$ and
\item if $c^* < 0$, then for every $\varepsilon \in [0,|c^*|]$, there exists $\xi \in \mathbb{R}_{>0}$ such that the closed-loop dynamics are strongly infinitesimally contracting with rate $|c^*| - \varepsilon > 0$ with respect to $\| \cdot \|_{\infty,[\varepsilon]}$.
\end{enumerate}

**Proof:** Let $f_L(x) := Ax + \nu \phi(w^Tx)$. Regarding statement (i), computing the one-sided Lipschitz constant of $f_L$ with respect to $\| \cdot \|_{1,[\varepsilon]}$ for arbitrary $\eta \in \mathbb{R}_{>0}$ yields

$$
\inf_{\varepsilon \in \mathbb{R}_{>0}} \max \{ \mu_{1,\varepsilon}(A + d_1 w^T), \mu_{1,\varepsilon}(A + d_2 w^T) \}
$$

where $\sharp$ holds by Assumption 2 on $\phi$ and $\#$ holds because the maximum of a convex function ($\mu$ in this case) over a compact interval occurs at one of the endpoints of the interval. As a consequence, $\inf_{\varepsilon \in \mathbb{R}_{>0}} (f_L)$ is less than 0 if and only if

$$
\inf_{\varepsilon \in \mathbb{R}_{>0}} \mu_{1,\varepsilon}(A + d_1 w^T), \mu_{1,\varepsilon}(A + d_2 w^T) 0.
$$

Therefore, if $b^* < 0$ for problem (25), then, by a continuity argument, for every $\varepsilon \in [0,|b^*|]$, there exists $\eta \in \mathbb{R}_{>0}$ such that $\mu_{1,\varepsilon}(A + d_1 w^T) \leq b^* + \varepsilon$ and $\mu_{1,\varepsilon}(A + d_2 w^T) \leq b^* + \varepsilon$. Therefore, if $b^* < 0$, then we conclude that the Lur’e system is strongly infinitesimally contracting with respect to $\| \cdot \|_{1,[\varepsilon]}$ with rate $|b^*| - \varepsilon$. The proof of statement (ii) is essentially identical, replacing $\| \cdot \|_{1,[\varepsilon]}$ with $\| \cdot \|_{\infty,[\varepsilon]}$.

**Theorem 30 (Multivariable Lur’e system):** Consider the multivariable Lur’e system

$$
\begin{aligned}
\dot{x} &= Ax + B\Phi(y) \\
y &= Cx
\end{aligned}
$$

where $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, C \in \mathbb{R}^{m \times n}$, $\phi_i \in \text{slope}[d_1, d_2]$ with $d_1 \geq 0$ for all $i \in \{1, \ldots, m\}$. Define $\cdot_+$ and $\cdot_-$ by $(x)_+ = \max\{x, 0\}$ and $(x)_- = \min\{x, 0\}$. Define $F \in \mathbb{R}^{n \times n}$ componentwise by

$$
F_{ij} = A_{ij} + d_2 \sum_{j=1}^{m} (B_{ij} C_{kj})_+ + d_1 \sum_{j=1}^{m} (B_{ij} C_{kj})_-
$$

$$
F_{ij} = |A_{ij}| + \max \left\{ d_2 \sum_{k=1}^{m} (B_{ik} C_{kj})_+, d_1 \sum_{k=1}^{m} (B_{ik} C_{kj})_- \right\}
$$

for $i \neq j$. Then, if $F \in \mathcal{M}H$ with corresponding dominant left and right eigenvectors $w_F, v_F$, the closed-loop dynamics are strongly infinitesimally contracting with rate $-\alpha(|F|_{\mathcal{M}H}) > 0$ with respect to both $\| \cdot \|_{1,[\varepsilon]}$ and $\| \cdot \|_{\infty,[\varepsilon]}$.

**Proof:** Let $f_{\text{ML}}(x) := Ax + B\Phi(Cx)$ and note $Df_{\text{ML}}(x) = A + B[d]C$ for some $d \in [d_1, d_2]^m$. Also note $(B[d]C)_{ij} = \sum_{k=1}^{m} B_{ik} d_k C_{kj}$. The proof follows from noting that the matrix $F$ is an entrywise upper bound on $|Df_{\text{ML}}(x)|_{\mathcal{M}H}$, for all $x$, in analogy with the proof of Theorem 28.

Finally, we present a sharper condition for the non-Euclidean contractivity of the multivariable Lur’e system with $d_1$ that can be negative. For $\eta \in \mathbb{R}_{>0}$, $\delta = \max\{|d_1|, |d_2|\}, M := |A + \delta B| C|_F$, and $g = |A|_{\infty,|\cdot|-1} + |\delta| B C|_F|_{\infty,|\cdot|-1}$, consider the following mixed-integer linear program (MILP):

$$
\begin{aligned}
\max_{y \in \mathbb{R}, z \in \mathbb{R}^{m \times n}, c \in [d_1, d_2]^m, w \in (0,1)^n} & y \\
\text{subject to} & \\
& Z \leq A + B[d]C + 2M \circ (W - (I_n \circ W)) \\
& Z \leq -A - B[d]C + 2M \circ (I_n - W - (I_n \circ W)) \\
& y \leq (A + B[d]C)_{ij} + \sum_{j \neq i} Z_{ij} \eta_j + 2gW_{ii}, \forall i \in \{1, \ldots, n\}
\end{aligned}
$$

$$
\text{Trace}(W) = n - 1.
$$

**Theorem 31 (One-sided Lipschitzness of multi-variable Lur’e system):** Consider the multivariable Lur’e system (27), let $f_{\text{ML}}(x) = Ax + B\Phi(Cx)$ be the closed-loop dynamics with each $\phi_i \in \text{slope}[d_1, d_2]$ and let $y^*$ be the optimal value for the MILP (28). Then, the following statements hold:

\begin{enumerate}
\item $\text{osL}_{\infty,|\cdot|-1}(f_{\text{ML}}) \leq y^*$;
\item if $C$ is full row rank, then $\text{osL}_{\infty,|\cdot|-1}(f_{\text{ML}}) = y^*$.
\end{enumerate}

**Proof:** Note that

$$
\text{osL}_{\infty,|\cdot|-1}(f_{\text{ML}}) \leq \max_{d \in [d_1, d_2]^m} \mu_{\infty,|\cdot|-1}(A + B[d]C)
$$

with equality holding if $C$ is full row rank. Therefore, all that remains is to show that $\max_{d \in [d_1, d_2]^m} \mu_{\infty,|\cdot|-1}(A + B[d]C) = y^*$. The proof of this result is a consequence of the formula for $\mu_{\infty,|\cdot|-1}$ using a so-called “big-M” formulation (see, e.g., [47, Section III-C]) with $Z_{ij} \leq \|A + B[d]C\|_{ij}$ for $i \neq j$ and $y \leq \max_{i \in \{1, \ldots, n\}} (A + B[d]C)_{ii} + \sum_{j \neq i} Z_{ij}$.

The challenge of additionally optimizing $\eta \in \mathbb{R}_{>0}$ so that $\text{osL}_{\infty,|\cdot|-1}(f_{\text{ML}})$ is minimized remains an open problem.

**VII. CONCLUSION**

In this article, we present novel non-Euclidean log norm results and a nonsmooth contraction theory simplification and we apply these results to study the contractivity of continuous-time NN models, primarily focusing on the Hopfield and firing-rate...
models. We provide efficient algorithms for computing optimal non-Euclidean contraction rates and corresponding norms. Our approach is robust with respect to activation function and additional unmodeled dynamics and, more generally, establishes the strong contractivity property which, in turn, implies strong robustness properties.

As a first direction of future research, we plan to investigate the multistability of continuous-time NNs via generalizations of contraction theory. Contraction theory ensures the uniqueness of a globally exponentially stable equilibrium, but several classes of NNs exhibit multiple equilibria [12]. As a second direction, we plan to investigate the role of non-Euclidean contractivity in NNs for controller design and system identification in the spirit of the works [24], [56], [58]. As a third line of research, we aim to implement non-Euclidean contracting NNs in machine learning problems akin to methods from [33]. More broadly, we believe that our non-Euclidean contraction framework for continuous-time NNs serves as a first step to analyzing robustness and convergence properties of other classes of neural circuits and other machine learning architectures.

APPENDIX

A. Proof of Theorem 5

Proof: Regarding statement (i), we provide the proof for $p = 1$ since $p = \infty$ is essentially identical. Continuity is a straightforward consequence of the formula for $\mu_1[\cdot]$. Regarding quasiconvexity, we will show that sublevel sets of the map $\eta \mapsto \mu_1[\cdot][A]$ are convex. For fixed $b \in \mathbb{R}$, the set \{\( \eta \in \mathbb{R}^n \mid \mu_1[\cdot][A] \leq b \}\) is characterized by $\eta$ satisfying

\[ \eta A_{ii} + \sum_{j=1, j \neq i}^{n} \eta_j A_{ij} \leq \eta_i b, \text{ for all } i \in \{1, \ldots, n\}. \]

Since each of these inequalities is linear in $\eta_i$, for fixed $b$, the above set is a polytope, proving quasiconvexity. Statement (ii) follows from the definitions of $\mu_1[\cdot][A]$ and $\mu_{\infty}[\cdot][A]$.

B. Proof of Lemma 8

To prove Lemma 8, we first need a technical result.

Lemma 32: For any $\gamma \in \mathbb{R}$, $A \in \mathbb{R}^{n \times n}$, the following holds:

\[ [\gamma A]_{Mn} = [\gamma |A| - (\gamma | - \gamma)(I_n \circ A)]_{Mn}. \]

Proof: The proof follows by checking that the corresponding entries of each matrix are equal.

Proof of Lemma 8: First we show (12). Use the short-hand $r_i := A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i$ and $D := \{d_1, d_2\}$. Then

\[
\max_{d \in \{d_1, d_2\}} \mu_{\infty}[\cdot][A] \leq \max_{i \in \{1, \ldots, n\}} \max_{d \in \{d_1, d_2\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

\[
\geq \max_{i \in \{1, \ldots, n\}} \max_{d \in \{d_1, d_2\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

where the equality $\star$ holds because the function $d_i \mapsto c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i$ is convex. Since the maximum value of a convex function over an interval $d_i \in [d_1, d_2]$ occurs at one of the endpoints, the equality $\star$ is justified.

In addition, note that for any $\gamma \in \mathbb{R}$

\[
\gamma A_{ii} + |\gamma| \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i = |\gamma| r_i - (|\gamma| - \gamma) A_{ii}.
\]

Therefore

\[
\max_{d \in \{d_1, d_2\}} |\mu_{\infty}[\cdot][A] + A[d]|
\]

\[
= \max_{i \in \{1, \ldots, n\}} \max_{d \in \{d_1, d_2\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

\[
\leq \max_{i \in \{1, \ldots, n\}} \max_{d \in \{d_1, d_2\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

where the equality $\star$ holds because the function $d_i \mapsto c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i$ is convex. Since the maximum value of a convex function over an interval $d_i \in [d_1, d_2]$ occurs at one of the endpoints, the equality $\star$ holds.

In addition, note that for any $\gamma \in \mathbb{R}$

\[
\gamma A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i = d_i - (d_i - \gamma) A_{ii}.
\]

Therefore

\[
\max_{d \in \{d_1, d_2\}} |\mu_{\infty}[\cdot][A] + A[d]|
\]

\[
= \max_{i \in \{1, \ldots, n\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

\[
= \max_{i \in \{1, \ldots, n\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

where the equality $\star$ holds because the function $d_i \mapsto c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i$ is convex. Since the maximum value of a convex function over an interval $d_i \in [d_1, d_2]$ occurs at one of the endpoints, the equality $\star$ holds.

In addition, note that for any $\gamma \in \mathbb{R}$

\[
\gamma A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i = d_i - (d_i - \gamma) A_{ii}.
\]

Therefore

\[
\max_{d \in \{d_1, d_2\}} |\mu_{\infty}[\cdot][A] + A[d]|
\]

\[
= \max_{i \in \{1, \ldots, n\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

\[
= \max_{i \in \{1, \ldots, n\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

where the equality $\star$ holds because the function $d_i \mapsto c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i$ is convex. Since the maximum value of a convex function over an interval $d_i \in [d_1, d_2]$ occurs at one of the endpoints, the equality $\star$ holds.

In addition, note that for any $\gamma \in \mathbb{R}$

\[
\gamma A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i = d_i - (d_i - \gamma) A_{ii}.
\]

Therefore

\[
\max_{d \in \{d_1, d_2\}} |\mu_{\infty}[\cdot][A] + A[d]|
\]

\[
= \max_{i \in \{1, \ldots, n\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

\[
= \max_{i \in \{1, \ldots, n\}} c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i
\]

where the equality $\star$ holds because the function $d_i \mapsto c_i + d_i A_{ii} + \sum_{j \neq i} |A_{ij}| \eta_j / \eta_i$ is convex. Since the maximum value of a convex function over an interval $d_i \in [d_1, d_2]$ occurs at one of the endpoints, the equality $\star$ holds.
Let \( e_k \) and \( e_m \) be unit vectors with 1 in their \( k \)th and \( m \)th entry, respectively, and define
\[
d_k = \gamma 1_n - (\gamma - d_1) e_k, \quad d_m = \gamma 1_n - (\gamma - d_2) e_m.
\]
Then, by construction
\[
\mu_{\infty, \eta}([c] + A[d_k]) = c_k + d_1 A kk + \sum_{j \neq k} |A_{kj}| \frac{\eta_j}{\eta_i}
\]
\[
= c_k + (d_1 - \bar{d}) A kk + \bar{d} A kk + \sum_{j \neq k} |A_{kj}| \frac{\eta_j}{\eta_i}
\]
\[
= c_k + \bar{d} r_k - (\bar{d} - d_1) A kk
\]
\[
= \mu_{\infty, \eta}([c] + A[d_m]) = \mu_{\infty, \eta}([c] + A[1_n - (\bar{d} - d_2) I_n]).
\]
(29)

Analogously, we have that \( \mu_{\infty, \eta}([c] + A[d_m]) = \mu_{\infty, \eta}([c] + \bar{d} A - (\bar{d} - d_2) (I_n \circ A)). \)

In addition, we see
\[
\max_{d \in [d_k, d_m]} \mu_{\infty, \eta}([c] + A[d]) 
= \max \{ \mu_{\infty, \eta}([c] + A[d]) \}
\]
\[
= \max \{ \mu_{\infty, \eta}([c] + A[d]) \}
\]
\[
= \mu_{\infty, \eta}([c] + A[d])
\]
(29)

The proofs for (13) and (15) are straightforward applications of the fact that \( \mu_{1, \eta} (B) = \mu_{\infty, \eta} (B^\top) \) and by applying (12) and (14), respectively.

**Corollary 33 (Some simplifications)**: Using the same notation as in Lemma 8, suppose

(i) \( \bar{d} = d_2 \) (note that this implies \( d_2 \geq 0 \)). Then
\[
\max_{d \in [d_k, d_m]} \mu_{\infty, \eta}([c] + A[d]) = \max \{ \mu_{\infty, \eta}([c] + A[d]) \}
\]
\[
= \mu_{\infty, \eta}([c] + A[d])
\]

(ii) \( \bar{d} = d_1 \) (note that this implies \( d_1 \leq 0 \)). Then
\[
\max_{d \in [d_k, d_m]} \mu_{\infty, \eta}([c] + A[d]) = \max \{ \mu_{\infty, \eta}([c] + A[d]) \}
\]
\[
= \mu_{\infty, \eta}([c] + A[d])
\]

C. Proof of Theorem 10

**Proof**: From [27, Th. 8.1.18], for all \( A \in \mathbb{R}^{n \times n} \), we have
\[
\rho(A) \leq \rho(|A|)
\]
(30)

where \( \rho \) denotes the spectral radius of a matrix. Regarding statement (i), pick \( \gamma > \max_{i \neq j} |A_{ij}| \) and define \( A = A + \gamma I_n \) so that \( |A| = |A|_{\operatorname{Max}} + \gamma I_n \). We note that \( \alpha(A) \leq \rho(A) \) (which is true for any matrix) and, from inequality (30), we know
\[
\alpha(A) + \gamma = \alpha(A) \leq \rho(A) \leq \rho(|A|) = \alpha(|A|)
\]
(31)

Here, \( \rho(|A|) = \alpha(|A|) \) follows from the Perron–Frobenius theorem for nonnegative matrices. This proves statement (i).

Regarding statement (ii), note that the norm \( \| \cdot \|_{\rho, \eta} \) is monotonic, it is easy to see that, for all matrices \( B \), we have \( \| B \|_{p, \eta} \leq \| B \|_{p, \eta} \).

*Proof of Theorem 12 and Corollary 13*

**Proof of Lemma 12**: Regarding statement (i), let \( D_2 \) denote the diagonal matrix with entries \( (D_2)_{ii} = 1 \) if \( i \in I \) and \( (D_2)_{ij} = 0 \) if \( i \notin I \).

With this notation, we are ready to compute
\[
\| A_T x \| = \max_{y \in \mathbb{R}^{n_{\eta, |y|}} = 1} \| A_T y \|
\]
(32)
\[
= \max_{y \in \mathbb{R}^{n_{\eta, |y|}} = 1} \| y \|
\]
(33)
\[
= \max_{y \in \mathbb{R}^{n_{\eta, |y|}} = 1} \| (D_2)_{A_T} y \|
\]
(34)
\[
\leq \max_{x \in \mathbb{R}^{n_{\eta, |x|}} = 1} \| (D_2)_{A_T} x \|
\]
(35)
\[
= \| (D_2)_{A_T} x \| \leq \| D_2 \| \| A \| \| D_2 \| = \| A \|
\]
(36)

The last equality holds because the monotonicity of \( \| \cdot \| \) implies \( \| D_2 \| = 1 \). This concludes the proof of (i).

Statement (ii) follows from the definition of log norm and applying statement (i) to the matrix \( I_1 + h A_T \) as a principal submatrix of \( I_n + h A \)
\[
\mu_{\eta}(A_T) := \lim_{h \to 0^+} \frac{\| I_1 + h A_T \| - 1}{h}
\]
(37)
\[
\leq \lim_{h \to 0^+} \frac{I_n + h A - 1}{h} = \mu(A).
\]

Finally, statement (iii) is an immediate consequence of (ii).

**Proof of Corollary 13**: Regarding item (i), since \( A \in \mathbb{M}(H, \alpha(|A|_{\operatorname{Max}}) < 0 \). By Lemma 2 and Theorem 10 (ii), for sufficiently small \( \varepsilon > 0 \), there exists \( \eta \in \mathbb{R}^n \) such that \( \mu_{1, \eta} (A) = \mu_{1, \eta} (|A|_{\operatorname{Max}}) \leq \alpha(|A|_{\operatorname{Max}}) + \varepsilon < 0 \). Then, by Lemma 12 (ii), for nonempty \( I \subset \{1, \ldots, n\} \), \( \mu_{1, \eta} (A_{ii}) \leq \mu_{1, \eta} (|A|_{\operatorname{Max}}) \leq 0 \). Moreover, by Theorem 10 (ii), \( \alpha(|A|_{\operatorname{Max}}) \leq \mu_{1, \eta} (A_{ii}) \leq 0 \). We conclude that \( A_{\leq} \) is \( \mathbb{M}(H) \). Regarding item (ii), note that \( A = A_{ij} \) is the matrix \( A \) with its \( ij \)th entry zeroed out. Then, since \( A \in \mathbb{M}(H) \), for sufficiently small \( \varepsilon > 0 \), there exists \( \eta \in \mathbb{R}^n \) such that \( \mu_{1, \eta} (A) < 0 \). The result is then a consequence of the fact that \( \alpha(|A - A_{ij} e_i e_j|_{\operatorname{Max}}) \leq \mu_{1, \eta} (|A - A_{ij} e_i e_j|_{\operatorname{Max}}) \leq \mu_{1, \eta} (|A - A_{ij} e_i e_j|) < 0 \).

E. Proof of Theorem 17

To prove Theorem 17, we first recall Clarke’s generalized Jacobian from nonsmooth analysis.

**Definition 34 ([13, Def. 2.6.1])**: Let \( f : U \to \mathbb{R}^m \) be locally Lipschitz on an open set \( U \subset \mathbb{R}^n \) and let \( \Omega_f \subset U \) be the set of
points where $f$ is not differentiable. Then Clarke’s generalized Jacobian at $x$ is
\[
\partial f(x) = \text{conv} \{ \lim_{t \to \infty} Df(x_i) \mid x_i \to x \text{ and } x_i \notin \Omega_f \}. \tag{37}
\]

The mean-value theorem has the following generalization for locally Lipschitz functions. For any two points $x, y \in \mathbb{R}^n$, denote $[x, y] := \{ tx + (1 - t)y \mid t \in [0, 1]\}$.

**Lemma 35:** For $f : U \to \mathbb{R}^m$ locally Lipschitz on an open convex set $U \subseteq \mathbb{R}^n$, let $[x, y] \subseteq U$. Then, there exists matrix $A \in \mathbb{R}^{m \times n}$ such that
\[
f(x) - f(y) = A(x - y) \quad \text{and} \quad A \in \text{conv} \bigcup_{u \in [x, y]} \partial f(u). \tag{38}
\]

**Proof:** Since $[x, y]$ is a compact subset of $\mathbb{R}^n$, it can be easily seen that there exists a convex open set $U_0$ such that $[x, y] \subseteq U_0 \subseteq U$ (where $U_0$ is the closure of $U_0$). The statement now follows from [13, Prop. 2.6.5].

**Proof of Theorem 17:** Regarding (ii) $\Rightarrow$ (i), Let $x, y \in U$. Since $U$ is convex, $[x, y] \subseteq U$. Then, by Lemma 35, there exists $A$ satisfying the conditions (38). Condition (ii) in implies that $\mu(\cdot)$ does not exceed $C$ on each set $\partial f(u)$ by continuity and convexity of $\mu$, entailing that $\mu(A) \leq C$. Therefore
\[
|f(x) - f(y)| \leq \mu(A)|x - y|^2 \leq C|x - y|^2
\]
where the second line is due to Lumer’s equality, Lemma 15. Regarding (i) $\Rightarrow$ (ii), let $x \in U$ such that $Df(x)$ exists and let $v \in \mathbb{R}^n$ and $h > 0$. Then, by assumption
\[
\|f(x + hv) - f(x, hv)\| \leq C\|hv\|^2
\]
which holds by the weak homogeneity of the weak pairing. Dividing by $h^2 > 0$ and taking the limit as $h \to 0$ implies
\[
\lim_{h \to 0^+} \frac{f(x + hv) - f(x)}{h} \leq C
\]
where the final implication holds by taking the supremum over all $v \in \mathbb{R}^n$ with $|v| = 1$ together with Lumer’s equality, Lemma 15. Therefore, statement (ii) holds. \hfill \square

### References

[1] Z. Aminzare and E. D. Sontag, “Contraction methods for nonlinear systems: A brief introduction and some open problems,” in *Proc. IEEE Conf. Decis. Control*, Dec. 2014, pp. 3835–3847.

[2] Z. Aminzare and E. D. Sontag, “Synchronization of diffusively-connected nonlinear systems: Results based on contractions with respect to general norms,” *IEEE Trans. Netw. Sci. Eng.*, vol. 1, no. 2, pp. 91–106, Jul.–Dec. 2014.

[3] D. Angeli, “A Lyapunov approach to incremental stability properties,” *IEEE Trans. Autom. Control*, vol. 47, no. 3, pp. 410–421, Mar. 2002.

[4] S. Arik, “A note on the global stability of dynamical neural networks,” *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 4, pp. 502–504, Apr. 2002.

[5] S. Bai, J. Z. Kolter, and V. Koltun, “Deep equilibrium models,” in *Proc. Ann. Conf. Neural Inf. Process. Syst.*, 2019. [Online]. Available: https://arxiv.org/abs/1909.01377

[6] G. P. Barker, A. Berman, and R. J. Plemmons, “Positive diagonal solutions to the Lyapunov equations,” *Linear Multilinear Algebra*, vol. 5, no. 4, pp. 249–256, 1978.

[7] F. Bullo, *Lectures on Netw. Syst.*., Seattle, WA, USA: Kindle Direct Publishing, Jan. 2022. [Online]. Available: https://bullo.github.io/lns

[8] F. Bullo, *Contraction Theory for Dynamical Syst.*, Seattle, WA, USA: Kindle Direct Publishing, 2023. [Online]. Available: https://bullo.github.io/lns

[9] F. Bullo, P. Cisneros-Velarde, A. Davydov, and S. Jafarpour, “From contraction theory to fixed point algorithms on Riemannian and non-Euclidean spaces,” in *Proc. IEEE Conf. Decis. Control*, Dec. 2021, pp. 2923–2928.

[10] V. Centorrino, F. Bullo, and G. Russo, “Contraction analysis of hopfield neural networks with Hebbian learning,” in *Proc. IEEE Conf. Decis. Control*, Cancún, México, Dec. 2022, pp. 622–627.

[11] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, “Neural ordinary differential equations,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2018, pp. 6571–6583. [Online]. Available: https://arxiv.org/abs/1806.07366

[12] C.-Y. Cheng, K.-H. Lin, and C.-W. Shih, “Multistability in recurrent neural networks,” *SIAM J. Appl. Math.*, vol. 66, no. 4, pp. 1301–1320, 2006.

[13] F. H. Clarke, *Optimization and Nonsmooth Analysis*. Hoboken, NJ, USA: Wiley, 1983.

[14] L. D’Alotto and M. Corless, “Incremental quadratic stability,” *Numer. Algebra, Control Optim.*, vol. 3, pp. 175–201, 2013.

[15] A. Davydov, S. Jafarpour, and F. Bullo, “Non-Euclidean contraction theory for robust nonlinear stability,” *IEEE Trans. Autom. Control*, vol. 67, no. 12, pp. 6667–6681, Dec. 2022.

[16] A. Davydov, A. V. Proskurnikov, and F. Bullo, “Non-Euclidean contrac-tivity of recurrent neural networks,” in *Proc. Amer. Control Conf.*, Atlanta, USA, May 2022, pp. 1527–1534.

[17] C. A. Desoer and H. Hanaoka, “The measure of a matrix as a tool to analyze computer algorithms for circuit analysis,” *IEEE Circuits Theory Circuits.*, vol. 19, no. 5, pp. 480–486, Sep. 1972.

[18] Y. Fang and T. G. Kincaid, “Stability analysis of dynamical neural networks,” *IEEE Trans. Neural Netw.*, vol. 7, no. 4, pp. 996–1006, Jul. 1996.

[19] M. Fazlyab, M. Morari, and G. J. Pappas, “Safety verification and robustness analysis of neural networks via quadratic constraints and semidefinite programming,” *IEEE Trans. Autom. Control*, vol. 67, no. 1, pp. 1–15, Jan. 2022.

[20] M. Forti, S. Manetti, and M. Marin, “Necessary and sufficient condition for absolute stability of neural networks,” *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 41, no. 7, pp. 491–494, Jul. 1994.

[21] M. Forti and A. Tesi, “New conditions for global stability of neural networks with application to linear and quadratic programming problems,” *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 42, no. 7, pp. 354–366, Jul. 1995.

[22] M. Giaccagi, V. Andrieu, S. Tarbouriech, and D. Astolfi, “Infinite gain margin, contraction and optimality: An LMI-based design,” *Eur. J. Control*, vol. 63, no. 2022, Art. no. 100685.

[23] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015. [Online]. Available: https://arxiv.org/abs/1412.6572

[24] A. Harapanahalli, S. Jafarpour, and S. Coogan, “Contraction-guided adaptive partitioning for reachability analysis of neural network controlled systems,” in *Proc. IEEE Conf. Decis. Control*, 2023, pp. 6044–6051.

[25] W. He and J. Cao, “Exponential synchronization of chaotic neural networks: A matrix measure approach,” *Nonlinear Dyn.*, vol. 55, pp. 55–65, 2009.

[26] J. J. Hopfield, “Neurons with graded response have collective computational properties like those of two-state neurons,” *Proc. Nat. Acad. Sci.*, USA, May 2018, pp. 6571–6583. [Online]. Available: https://arxiv.org/abs/1412.6572

[27] F. Bullo, “Contraction theory to fixed point algorithms on Riemannian and non-Euclidean spaces,” in *Proc. Conf. on Contr. Theory and Prac.*. Cancún, México, Dec. 2022, pp. 622–627.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
