Non-Euclidean Contractivity of Recurrent Neural Networks

(Article begins on next page)

Non-Euclidean Contractivity of Recurrent Neural Networks

Alexander Davydov, Anton V. Proskurnikov, and Francesco Bullo

***Abstract*— Critical questions in dynamical neuroscience and machine learning are related to the study of recurrent neural networks and their stability, robustness, and computational efficiency. These properties can be simultaneously established via a contraction analysis.**

This paper develops a comprehensive contraction theory for recurrent neural networks. First, for non-Euclidean ℓ_1/ℓ_∞ logarithmic norms, we establish quasiconvexity with respect to positive diagonal weights and closed-form worst-case expressions over certain matrix polytopes. Second, for locally Lipschitz maps (e.g., arising as activation functions), we show that their one-sided Lipschitz constant equals the essential supremum of the logarithmic norm of their Jacobian. Third and final, we apply these general results to classes of recurrent neural circuits, including Hopfield, firing rate, Persidskii, Lur'e and other models. For each model, we compute the optimal contraction rate and corresponding weighted non-Euclidean norm via a linear program or, in some special cases, via a Hurwitz condition on the Metzler majorant of the synaptic matrix. Our non-Euclidean analysis establishes also absolute, connective, and total contraction properties.

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 via the indicator **(C)** in what follows.

In **dynamical neuroscience**, several recurrent neural network (RNN) models are widely studied, including membrane potential models such as the Hopfield neural network [16] and firing-rate models [26]. Clearly, such models are simplifications of complex neural dynamics. For example, if $f(x)$ is an RNN model of a neural circuit, the true dynamics is better estimated by

$$\dot{x}(t) = f(x(t)) + g(x(t), x(t-\tau)), \tag{1}$$

where g captures model uncertainty and time-delays. In other words, **(C1):** to account for uncertainty in the system, the nominal dynamics $f(x)$ must exhibit robust stability with respect to unmodeled dynamics and delay.

Central pattern generators (CPGs) are biological neural circuits that generate periodic signals and are the source of rhythmic motor behaviors such as walking, swimming, and breathing. To properly model CPGs in RNNs, a computational neuroscientist would need to ensure that, **(C2):** if an RNN is interconnected with a CPG, then entrainment takes place and the trajectories of the RNN converge to a unique stable limit cycle.

Machine learning scientists have widely adopted discrete-time RNNs for pattern recognition and analysis of sequential data and much recent interest [4], [19], [32], [18] has focused on the closely-related class of implicit neural networks. In particular, training implicit networks corresponds to solving fixed-point problems of the form

$$x = \Phi(Ax + Bu + b), \tag{2}$$

where x is the neural state variable, Φ 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 equation (2) is the equilibrium point of a corresponding RNN differential equation, (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 RNNs and implicit neural networks, it is essential to pick model weights in such a way that **(C3):** fixed-point equations have unique solutions for all possible inputs and activation functions, and **(C4):** 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 [39] that artificial deep neural networks are sensitive to adversarial perturbations: small input changes may lead to large 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 [33], [12]. In short, **(C5):** the input/output Lipschitz constants of RNNs need to be tightly estimated, e.g., in the context of the fixed-point equation (2).

A contraction theory for neural networks. Motivated by the challenges arising in neuroscience and machine learning, this paper aims to perform a *robust stability analysis* of continuous-time RNNs and develop *optimization methods*

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2139319 and AFOSR grant FA9550-22-1-0059.

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, 93106-5070, USA. {davydov, bullo}@ucsb.edu.

Anton V. Proskurnikov is with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy anton.p.1982@ieee.org.

for discrete-time RNN models. Serendipitously, both these objectives can be simultaneously achieved through a contraction analysis for the RNN dynamics.

For concreteness' sake, we briefly review how the aforementioned challenges **(C1-C5)** are addressed by a contraction analysis. Infinitesimally contracting dynamics enjoy highly ordered *transient* and *asymptotic* behaviors: **(C1)** initial conditions are forgotten and the distance between trajectories is monotonically vanishing [24], **(C3)** 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) [24], **(C2)** periodic systems admit a unique globally exponentially stable periodic solution or, for systems with periodic inputs, each solution entrains to the periodic input [34], **(C1)** and **(C5)** contracting vector fields enjoy highly robust behavior, e.g., see [40], [9], including (a) input-to-state stability, (b) finite input-state gain, (c) contraction margin with respect to unmodeled dynamics, and (d) input-to-state stability under delayed dynamics. Hence, the contraction rate is a natural measure/indicator of robust stability. Paraphrasing [30], contracting systems are in many ways similar to stable linear systems (but without superposition principle).

With regards to **(C4)**, our recent work [7], [18] shows how to design efficient fixed-point computation schemes for contracting systems (with respect to arbitrary and non-Euclidean ℓ_1/ℓ_∞ norms) in the style of monotone operator theory [35]. Specifically, for contracting dynamics with respect to a diagonally-weighted ℓ_1/ℓ_∞ norm, optimal step-sizes and convergence factors are given in [18, Theorem 2]. These results are directly applicable to the computation of fixed-points in implicit neural networks, as in equation (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 RNN models have been studied for a few decades. Shortly after Hopfield's original work [16], control-theoretic ideas were proposed by is [25]. Later, [20], [13], [14] obtained various version 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. Notably, [11] is the earliest reference on the application of logarithmic norms to Hopfield neural networks and provides results on ℓ_p logarithmic norms of the Jacobian for networks with smooth activation functions. [3] proposes a quasi-dominance condition on the synaptic matrix (in lieu of Lyapunov diagonal stability). [31] proposes the notion of the nonlinear measure of a map to study global asymptotic stability; this notion is closely related to the ℓ_1 one-sided Lipschitz constant of the map. A comprehensive survey on continuous-time RNNs is [42].

Recently, the non-Euclidean contraction of monotone Hopfield neural networks is studied in [17]; see also [8] for the interplay between Metzler matrices and non-Euclidean logarithmic norms. Also recently, [28] studies 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. Since checking this condition is NP-hard, more conservative convex conditions are provided as well. The importance of non-Euclidean log norms in contraction theory is highlighted for example in [34], [2].

Finally, contractivity of RNNs with respect to the ℓ_2 norm has been studied, e.g., see the early reference [11], the related discussion in [32], and the recent work [22].

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

Specifically, the paper develops a comprehensive contraction theory for RNN models through the following contributions. First, we obtain novel logarithmic norm results including (i) the quasiconvexity of the ℓ_1 and ℓ_∞ 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 ℓ_1 and ℓ_∞ logarithmic norms under multiplicatively-weighted uncertainty, resulting in a maximization of the logarithmic norm over a matrix polytope. The formulas in (iii) generalize previous results [11, Theorem 3.8], [15, Lemma 3] and [18, 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 RNNs that have nonsmooth activation functions such as ReLU.

Third and finally, we consider multiple models of recurrent neural circuits and nonlinear dynamical models, including Hopfield, firing rate, Persidskii, Lur'e, and others. We consider activation functions that are weakly increasing and Lipschitz (thus more general than the class of piecewise-affine activation functions). For each model, we propose a linear program to characterize the optimal contraction rate and corresponding weighted non-Euclidean ℓ_1 or ℓ_∞ norm. In some special cases, we show that the linear program reduces to checking an M-Hurwitz condition on the synaptic matrix. 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 M-Hurwitz and is replaced by any principal submatrix, the principle submatrix is also M-Hurwitz. The process of replacing the nominal RNN with a subsystem RNN is referred to as "pruning" both in neuroscience and in machine learning.

Paper organization. Section II reviews known preliminary concepts. Section III provides novel logarithmic norms results. Section IV studies nonsmooth contraction theory. Section V establishes conditions for the contractivity of classes of neural dynamics. In the interest of brevity, we refer to the technical report https://arxiv.org/abs/2110.08298 for all omitted proofs.

II. Review of Relevant matrix analysis

For two matrices (or vectors) A, B, we let $A \circ B$ be entrywise multiplication. Vector inequalities of the form $x \leq y$ are entrywise. For a vector $\eta \in \mathbb{R}^n$, we define $[\eta] \in \mathbb{R}^{n\times n}$ to be the diagonal matrix with diagonal entries equal to η. We let $\mathbb{1}_n, \mathbb{0}_n \in \mathbb{R}^n$ be the all-ones and all-zeros vectors, respectively. We say a norm $\|\cdot\|$ on $\mathbb{R}^n$ is *monotonic* if for all $x, y \in \mathbb{R}^n$, $|x| \leq |y| \implies \|x\| \leq \|y\|$, where the absolute value is applied entrywise. 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 $\alpha(A) = \max\{\Re(\lambda) \mid \lambda \in \operatorname{spec}(A)\}$ and its *Metzler majorant* $\lceil A \rceil_{\mathrm{Mzr}} \in \mathbb{R}^{n\times n}$ is defined by $(\lceil A \rceil_{\mathrm{Mzr}})_{ij} = \begin{cases} a_{ii}, & \text{if } i = j \\ |a_{ij}|, & \text{if } i \neq j \end{cases}.$

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

$$\mu(A) := \lim_{h\to 0^+} \frac{\|I_n + hA\| - 1}{h}. \tag{3}$$

We refer to [10] for a list of properties of log norms, which include subadditivity, convexity, and $\alpha(A) \leq \mu(A)$. For an ℓ_p norm, $p \in [1, \infty]$ and for invertible $R \in \mathbb{R}^{n\times n}$, we define the R-weighted ℓ_p norm by $\|x\|_{p,R} = \|Rx\|_p$. It is known that the corresponding log norm is then $\mu_{p,R}(A) = \mu_p(RAR^{-1})$. For diagonally weighted ℓ_1, ℓ_∞, and ℓ_2 norms,

$$\begin{aligned}
\mu_{1,[\eta]}(A) &= \max_{i\in\{1,\dots,n\}} a_{ii} + \sum\nolimits_{j=1, j\neq i}^{n} \frac{\eta_j}{\eta_i}|a_{ji}| \\
&= \min\{b \in \mathbb{R} \mid \lceil A \rceil_{\mathrm{Mzr}}^\top \eta \leq b\eta\}, \\
\mu_{\infty,[\eta]^{-1}}(A) &= \max_{i\in\{1,\dots,n\}} a_{ii} + \sum\nolimits_{j=1, j\neq i}^{n} \frac{\eta_j}{\eta_i}|a_{ij}| \\
&= \min\{b \in \mathbb{R} \mid \lceil A \rceil_{\mathrm{Mzr}} \eta \leq b\eta\}, \\
\mu_{2,[\eta]^{1/2}}(A) &= \min\{b \in \mathbb{R} \mid [\eta]A + A^\top[\eta] \preceq 2b[\eta]\}.
\end{aligned}$$

The following result is due to [38] and [29, Lemma 3].

Lemma 1 (Optimal diagonally-weighted log norms for Metzler matrices)**.** *Given a Metzler matrix $M \in \mathbb{R}^{n\times n}$, $p \in [1, \infty]$, and $\delta > 0$, define $\eta_{M,p,\delta} \in \mathbb{R}^n_{>0}$ by*

$$\eta_{M,p,\delta} = \left(\frac{w_1^{1/p}}{v_1^{1/q}}, \dots, \frac{w_n^{1/p}}{v_n^{1/q}}\right), \tag{4}$$

where $q \in [1, \infty]$ is defined by $1/p + 1/q = 1$ (with the convention $1/\infty = 0$) and where v and $w \in \mathbb{R}^n_{>0}$ are the right and left dominant eigenvectors of the irreducible Metzler matrix $M + \delta\mathbb{1}_n\mathbb{1}_n^\top$ (whose existence is guaranteed by the Perron-Frobenius Theorem). Then for each $\epsilon > 0$ there exists $\delta > 0$ such that

(i) $\alpha(M) \leq \mu_{p,[\eta_{M,p,\delta}]}(M) \leq \alpha(M) + \epsilon$,
(ii) if M is irreducible, then $\alpha(M) = \mu_{p,[\eta_{M,p,0}]}(M)$.

Lemma 1 also ensures that for Metzler matrices $M \in \mathbb{R}^{n\times n}$, $\inf_{\eta\in\mathbb{R}^n_{>0}} \mu_{p,[\eta]}(M) = \alpha(M)$ 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 $A \in \mathcal{H}$, if $\alpha(A) < 0$,
(ii) *totally Hurwitz*, denoted by $A \in \mathcal{TH}$, if all principal submatrices of A are Hurwitz stable,
(iii) *Lyapunov diagonally stable (LDS)*, denoted by $A \in \mathcal{LDS}$, if there exists a $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_{2,[\eta]^{1/2}}(A) < 0$, and
(iv) *M-Hurwitz stable*, denoted by $A \in \mathcal{MH}$, if $\alpha(\lceil A \rceil_{\mathrm{Mzr}}) < 0$.

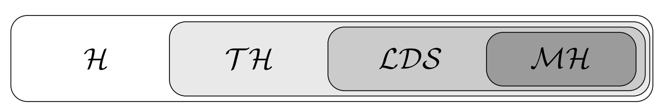
A matrix $A \in \mathbb{R}^{n\times n}$ is *quasidominant* [27] if there exists a vector $\eta \in \mathbb{R}^n_{>0}$ such that

$$\eta_i a_{ii} > \sum\nolimits_{j=1, j\neq i}^{n} \eta_j |a_{ij}|, \quad \text{for all } i \in \{1, \dots, n\}.$$

This is equivalent to $\lceil -A \rceil_{\mathrm{Mzr}}\eta < \mathbb{0}_n$, which, in turn, is equivalent (see, for example, [6, Theorem 15.17]) to the inequality $\alpha(\lceil -A \rceil_{\mathrm{Mzr}}) < 0$, i.e., $-A \in \mathcal{MH}$.

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

Lemma 2 (Inclusions for classes of matrices)**.** *$(A \in \mathcal{MH})$ implies $(A \in \mathcal{LDS})$, $(A \in \mathcal{LDS})$ implies $(A \in \mathcal{TH})$, and $(A \in \mathcal{TH})$ implies $(A \in \mathcal{H})$.*



We show that the counter-implications in Lemma 2 do not hold.

Example 3. *(i) $(A \in \mathcal{LDS} \not\Rightarrow A \in \mathcal{MH})$ The matrix $A = \begin{bmatrix} -1 & -1 \\ 2 & -1 \end{bmatrix}$ satisfies $\mu_2(A) = -0.5$, so $A \in \mathcal{LDS}$. However, $\alpha(\lceil A \rceil_{\mathrm{Mzr}}) = \sqrt{2} - 1 > 0$, so $A \notin \mathcal{MH}$.*
(ii) $(A \in \mathcal{TH} \not\Rightarrow A \in \mathcal{LDS})$ is proved in [5, Remark 4].
(iii) $(A \in \mathcal{H} \not\Rightarrow A \in \mathcal{TH})$ The matrix $A = \begin{bmatrix} 1 & 1 \\ -4 & -3 \end{bmatrix}$ satisfies $\alpha(A) = -1$, so $A \in \mathcal{H}$. However, $A \notin \mathcal{TH}$ since it has a positive diagonal entry.

This insert corresponds to Lemma 6. For $A \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$, $0 \le d_{\min} \le d_{\max} \in \mathbb{R}$, and $\eta \in \mathbb{R}^n_{>0}$,

$$\max_{d\in[d_{\min},d_{\max}]^n} \mu_{\infty,[\eta]}([c]+[d]A) = \max\left\{\mu_{\infty,[\eta]}([c]+d_{\min}A), \mu_{\infty,[\eta]}([c]+d_{\max}A)\right\}, \tag{5}$$

$$\max_{d\in[d_{\min},d_{\max}]^n} \mu_{1,[\eta]}([c]+A[d]) = \max\left\{\mu_{1,[\eta]}([c]+d_{\min}A), \mu_{1,[\eta]}([c]+d_{\max}A)\right\}, \tag{6}$$

$$\max_{d\in[d_{\min},d_{\max}]^n} \mu_{\infty,[\eta]}([c]+A[d]) = \max\{\mu_{\infty,[\eta]}([c]+d_{\max}A), \mu_{\infty,[\eta]}([c]+d_{\max}A-(d_{\max}-d_{\min})(I_n\circ A))\}, \tag{7}$$

$$\max_{d\in[d_{\min},d_{\max}]^n} \mu_{1,[\eta]}([c]+[d]A) = \max\{\mu_{1,[\eta]}([c]+d_{\max}A), \mu_{1,[\eta]}([c]+d_{\max}A-(d_{\max}-d_{\min})(I_n\circ A))\}. \tag{8}$$

III. 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 4 (Quasiconvexity of μ with respect to diagonal weights)**.** *For fixed $A \in \mathbb{R}^{n\times n}$, consider the maps from $\mathbb{R}^n_{>0}$ to $\mathbb{R}$ defined by*

$$\begin{aligned} &\eta \mapsto \mu_{1,[\eta]}(A), \\ &\eta \mapsto \mu_{\infty,[\eta]^{-1}}(A). \end{aligned} \tag{9}$$

Then

(i) The maps in (9) are quasiconvex and their sublevel sets are polytopes.

(ii) Minimizing the maps in (9) may be executed via the optimization problems

$$\begin{aligned} \inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\ \text{s.t.} \quad & \lceil A\rceil_{\mathrm{Mzr}}^\top \eta \le b\eta, \end{aligned} \tag{10}$$

for $\mu_{1,[\eta]}(A)$ and

$$\begin{aligned} \inf_{b\in\mathbb{R},\eta\in\mathbb{R}^n_{>0}} \quad & b \\ \text{s.t.} \quad & \lceil A\rceil_{\mathrm{Mzr}} \eta \le b\eta, \end{aligned} \tag{11}$$

for $\mu_{\infty,[\eta]^{-1}}(A)$.

Remark 5. *The optimization problems in (10) and (11) may be modified such that $\eta \in [\varepsilon,\infty[^n$ for $\varepsilon > 0$ sufficiently small so that the $\inf$ becomes a $\min$. Then the problems may be solved by a bisection on $b \in [-\|A\|,\|A\|]$, where each step of the algorithm is a linear program (LP) in η.*

Next, we provide closed-form expressions for ℓ_1 and ℓ_∞ log norms over a certain polytopes 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 several classes of RNNs.

Lemma 6 (Worst-case ℓ_1/ℓ_∞ log norms under multiplicative scalings)**.** *Any $A \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$, $0 \le d_{\min} \le d_{\max} \in \mathbb{R}$, and $\eta \in \mathbb{R}^n_{>0}$ satisfy formulas (5)-(8).*

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 6, formulas (5)-(8) 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 worst-case log norm of a matrix polytope as in Lemma 6 can be easily computed.

Corollary 7. *Let $A, c, d_{\min}$, and $d_{\max}$ be as in Lemma 6. Then for $\mu_{[\eta]}(\cdot)$ denoting either $\mu_{1,[\eta]}(\cdot)$ or $\mu_{\infty,[\eta]^{-1}}(\cdot)$ the minimax problems*

$$\min_{\eta\in[\varepsilon,\infty[^n} \max_{d\in[d_{\min},d_{\max}]^n} \mu_{[\eta]}([c]+[d]A),$$
$$\min_{\eta\in[\varepsilon,\infty[^n} \max_{d\in[d_{\min},d_{\max}]^n} \mu_{[\eta]}([c]+A[d]),$$

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

B. Monotonicity of diagonally-weighted log norms

Theorem 8 (Monotonicity of α and μ)**.** *For any $A \in \mathbb{R}^{n\times n}$*

(i) $\alpha(A) \le \alpha(\lceil A\rceil_{\mathrm{Mzr}})$,

(ii) for all $p \in [1,\infty]$ and $\eta \in \mathbb{R}^n_{>0}$, we have $\mu_{p,[\eta]}(A) \le \mu_{p,[\eta]}(\lceil A\rceil_{\mathrm{Mzr}})$, with equality holding for $p \in \{1,\infty\}$.

(iii) For $p \in \{1,\infty\}$,

$$\inf_{\eta\in\mathbb{R}^n_{>0}} \mu_{p,[\eta]}(A) = \alpha(\lceil A\rceil_{\mathrm{Mzr}}) \ge \alpha(A).$$

Theorem 8(iii) demonstrates that using diagonally-weighted ℓ_1 and ℓ_∞ log norms, the best bound one can achieve on $\alpha(A)$ is $\alpha(\lceil A\rceil_{\mathrm{Mzr}})$, which may be conservative. In the following example, we show that the ℓ_2 norm does not have the same conservatism. Despite the conservatism, Theorem 4 demonstrates that optimizing diagonal weights is computationally efficient, being an LP at every step of the bisection, while optimizing weights for the ℓ_2 norm is an LMI at every step, which is more computationally challenging than an LP of similar dimension.

Example 9. *The matrix $A_* = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ has eigenvalues $\{1+\mathrm{i}, 1-\mathrm{i}\}$ whereas $\lceil A_*\rceil_{\mathrm{Mzr}}$ has eigenvalues $\{2,0\}$. Therefore, $\alpha(A_*) = 1 < 2 = \alpha(\lceil A_*\rceil_{\mathrm{Mzr}})$. Additionally, $(A_* + A_*^\top)/2 = I_2 \implies \mu_2(A_*) = 1$ and $\mu_2(\lceil A_*\rceil_{\mathrm{Mzr}}) = 2$.*

C. Log norms of principal submatrices

Given a matrix $A \in \mathbb{R}^n$ and a non-empty index set $\mathcal{I} \subset \{1,\dots,n\}$, let $A_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$ denote the *principal submatrix* obtained by removing the rows and columns of A which are not in $\mathcal{I}$. Next, given a non-empty $\mathcal{I} \subset \{1,\dots,n\}$, define the *zero-padding map* $\mathrm{pad}_{\mathcal{I}} : \mathbb{R}^{|\mathcal{I}|} \to \mathbb{R}^n$ as follows:

$\mathrm{pad}_{\mathcal{I}}(y)$ is obtained by inserting zeros among the entries of y corresponding to the indices in $\{1,\dots,n\}\setminus\mathcal{I}$. For example, with $n=3$ and $\mathcal{I}=\{1,3\}$, we define $\mathrm{pad}_{\{1,3\}}(y_1,y_2)=(y_1,0,y_2)$. Then it is easy to see that given a norm $\|\cdot\|$ on $\mathbb{R}^n$ and non-empty $\mathcal{I}\subset\{1,\dots,n\}$, the map $\|\cdot\|_{\mathcal{I}}:\mathbb{R}^{|\mathcal{I}|}\to\mathbb{R}_{\geq 0}$ defined by $\|y\|_{\mathcal{I}}=\|\mathrm{pad}_{\mathcal{I}}(y)\|$ is a norm on $\mathbb{R}^{|\mathcal{I}|}$.

Lemma 10 (Norm and log norm of principal submatrices)**.** *Assume $\|\cdot\|$ is monotonic, let μ and $\mu_{\mathcal{I}}$ denote the log norms associated to $\|\cdot\|$ and $\|\cdot\|_{\mathcal{I}}$ respectively. Any matrix $A\in\mathbb{R}^{n\times n}$ satisfies*

(i) $\|A_{\mathcal{I}}\|_{\mathcal{I}}\leq\|A\|$,
(ii) $\mu_{\mathcal{I}}(A_{\mathcal{I}})\leq\mu(A)$,
(iii) if $\mu(A)<0$, *then* $A\in\mathcal{TH}$.

Corollary 11. *If $A\in\mathcal{MH}\subset\mathbb{R}^{n\times n}$, then $A_{\mathcal{I}}\in\mathcal{MH}$ for every non-empty $\mathcal{I}\subset\{1,\dots,n\}$.*

IV. One-sided Lipschitz maps and nonsmooth contraction theory

A. Review of one-sided Lipschitz functions

We review weak pairings and one-sided Lipschitz maps as introduced in [9]; see also the earlier works [37], [1].

Definition 12 (Weak pairing)**.** *A* weak pairing *on $\mathbb{R}^n$ is a map $[\![\cdot,\cdot]\!]:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ satisfying:*

(i) (Subadditivity and continuity in its first argument) $[\![x_1+x_2,y]\!]\leq[\![x_1,y]\!]+[\![x_2,y]\!]$, *for all* $x_1,x_2,y\in\mathbb{R}^n$ *and* $[\![\cdot,\cdot]\!]$ *is continuous in its first argument,*
(ii) (Weak homogeneity) $[\![\alpha x,y]\!]=[\![x,\alpha y]\!]=\alpha\,[\![x,y]\!]$ *and* $[\![-x,-y]\!]=[\![x,y]\!]$ *for all* $x,y\in\mathbb{R}^n$, $\alpha\geq 0$,
(iii) (Positive definiteness) $[\![x,x]\!]>0$ *for all* $x\neq\mathbb{0}_n$,
(iv) (Cauchy-Schwarz) $|\,[\![x,y]\!]\,|\leq[\![x,x]\!]^{1/2}\,[\![y,y]\!]^{1/2}$ *for all* $x,y\in\mathbb{R}^n$.

Additionally, we say a weak pairing satisfies Deimling's inequality *if* $[\![x,y]\!]\leq\|y\|\lim_{h\to 0^+}h^{-1}(\|y+hx\|-\|y\|)$ *for all $x,y\in\mathbb{R}^n$, where $\|\cdot\|=[\![\cdot,\cdot]\!]^{1/2}$.*

Deimling's inequality is well-defined since $[\![\cdot,\cdot]\!]^{1/2}$ defines a norm on $\mathbb{R}^n$. Conversely, if $\mathbb{R}^n$ is equipped with a norm $\|\cdot\|$ then there exists a (possibly non-unique) weak pairing $[\![\cdot,\cdot]\!]$ such that $\|\cdot\|=[\![\cdot,\cdot]\!]^{1/2}$; see [9, Theorem 16]. Henceforth, we assume that weak pairings satisfy Deimling's inequality.

We establish the relationship between weak pairings and log norms in the following lemma.

Lemma 13 (Lumer's equality [9, Theorem 18])**.** *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ with compatible weak pairing $[\![\cdot,\cdot]\!]$. Then*

$$\mu(A)=\sup_{x\in\mathbb{R}^n,x\neq\mathbb{0}_n}\frac{[\![Ax,x]\!]}{\|x\|^2},\qquad\text{for all } A\in\mathbb{R}^{n\times n}.\quad(12)$$

Definition 14 (One-sided Lipschitz functions [9, Definition 26])**.** *Consider $f:U\to\mathbb{R}^n$ where $U\subseteq\mathbb{R}^n$ 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*

$$[\![f(x)-f(y),x-y]\!]\leq b\|x-y\|^2,\quad\text{for all } x,y\in U.$$

We say b is a one-sided Lipschitz constant *of f. Moreover, the minimal one-sided Lipschitz constant of f is*

$$\mathrm{osL}(f):=\sup_{x,y\in U,x\neq y}\frac{[\![f(x)-f(y),x-y]\!]}{\|x-y\|^2}.\quad(13)$$

If f is continuously differentiable and U is convex, it can be shown that $\mathrm{osL}(f)=\sup_{x\in U}\mu(Df(x))$, 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 $\mathrm{osL}(f)\leq -c<0$ is said to be *strongly infinitesimally contracting* with rate c. Any two trajectories $x(\cdot),y(\cdot)$ satisfying $\dot{x}=f(x)$ additionally satisfy $\|x(t)-y(t)\|\leq e^{-ct}\|x(0)-y(0)\|$ for all $t\geq 0$. Moreover, if f is continuous, then all solutions converge to a unique equilibrium.

B. Nonsmooth contraction theory

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

Theorem 15 (osL simplification for locally Lipschitz f)**.** *For $f:U\to\mathbb{R}^n$ locally Lipschitz on an open convex set, $U\subseteq\mathbb{R}^n$. Then for every $c\in\mathbb{R}$ the following statements are equivalent:*

(i) $\mathrm{osL}(f)\leq c$,
(ii) $\mu(Df(x))\leq c$ *for almost every* $x\in U$.

Note that $Df(x)$ exists for almost every $x\in U$ by Rademacher's theorem. Theorem 15 demonstrates that locally Lipschitz f enjoy a similar simplification in the osL definition as do continuously differentiable functions.

In neural network models, nonsmooth activation functions such as ReLU, LeakyReLU, and nonsmooth saturation functions are prevalent; Theorem 15 allows us to use standard log norm results to analyze these models.

V. Contracting neural dynamics

We consider several models of neural circuits and characterize their one-sided Lipschitz constants and therefore their strong infinitesimal contractivity.

A. Hopfield neural network

We start with the continuous-time Hopfield neural network model, first introduced in [16]:

$$\dot{x}=-Cx+A\Phi(x)+u=:f_{\mathrm{H}}(x),\quad(14)$$

where $C\in\mathbb{R}^{n\times n}$ is a positive semi-definite diagonal matrix, $A\in\mathbb{R}^{n\times n}$ is arbitrary, $u\in\mathbb{R}^n$ is a (possibly time-varying) input, and Φ is a diagonal activation function. In other words, $\Phi(x)=[\phi_1(x_1),\dots,\phi_n(x_n)]$, where each $\phi_i:\mathbb{R}\to\mathbb{R}$ is Lipschitz and satisfies the slope-restricted constraints

$$\begin{aligned}d_{\min}&:=\inf_{x,y\in\mathbb{R},x\neq y}\frac{\phi_i(x)-\phi_i(y)}{x-y}\geq 0,\\ d_{\max}&:=\sup_{x,y\in\mathbb{R},x\neq y}\frac{\phi_i(x)-\phi_i(y)}{x-y}<\infty.\end{aligned}\quad(15)$$

In other words, this ensures that $\phi_i'(x) \in [d_{\min}, d_{\max}]$ for almost every $x \in \mathbb{R}$. Many common activation functions satisfy these constraints including ReLU, $\tanh$ and sigmoids.

Theorem 16 (One-sided Lipschitzness of Hopfield neural network)**.** *Consider the Hopfield neural network model (14) with irreducible $\lceil A \rceil_{\mathrm{Mzr}}$ and constant $u \in \mathbb{R}^n$. Then*

(i) $\mathrm{osL}_{1,[\eta]}(f_{\mathrm{H}}) = \max\left\{\mu_{1,[\eta]}(-C + d_{\min}A), \mu_{1,[\eta]}(-C + d_{\max}A)\right\}$*, for arbitrary $\eta \in \mathbb{R}^n_{>0}$.*

(ii) The vector η minimizing $\mathrm{osL}_{1,[\eta]}(f_{\mathrm{H}})$ is the solution to

$$\begin{aligned} \inf_{b \in \mathbb{R}, \eta \in \mathbb{R}^n_{>0}} \quad & b \\ \textit{s.t.} \quad & (-C + d_{\min}\lceil A \rceil_{\mathrm{Mzr}}^\top)\eta \leq b\eta, \\ & (-C + d_{\max}\lceil A \rceil_{\mathrm{Mzr}}^\top)\eta \leq b\eta. \end{aligned}$$

(iii) if $C = cI_n$, then, with $w_A \in \mathbb{R}^n_{>0}$ being the left dominant eigenvector of $\lceil A \rceil_{\mathrm{Mzr}}$,

$$\begin{aligned} \inf_{\eta \in \mathbb{R}^n_{>0}} \mathrm{osL}_{1,[\eta]}(f_{\mathrm{H}}) &= \mathrm{osL}_{1,[w_A]}(f_{\mathrm{H}}) \\ &= -c + \max\{d_{\min}\alpha(\lceil A \rceil_{\mathrm{Mzr}}), d_{\max}\alpha(\lceil A \rceil_{\mathrm{Mzr}})\}. \end{aligned} \tag{16}$$

(iv) if $d_{\min} = 0$ and $C \succ 0$, then, with $w_* \in \mathbb{R}^n_{>0}$ being the left dominant eigenvector of $-C + d_{\max}\lceil A \rceil_{\mathrm{Mzr}}$,

$$\begin{aligned} \inf_{\eta \in \mathbb{R}^n_{>0}} \mathrm{osL}_{1,[\eta]}(f_{\mathrm{H}}) &= \mathrm{osL}_{1,[w_*]}(f_{\mathrm{H}}) \\ &= \max\left\{\alpha(-C), \alpha(-C + d_{\max}\lceil A \rceil_{\mathrm{Mzr}})\right\}. \end{aligned} \tag{17}$$

In particular, Theorem 16 provides *exact values* for the minimal one-sided Lipschitz constant of the Hopfield neural network with respect to diagonally-weighted ℓ_1 norms.

As a consequence of this theorem, let $b^\star, \eta^\star$ be the optimal solution for the LP in statement (ii). If $b^\star < 0$, then the Hopfield neural network (14) is strongly infinitesimally contracting with rate $|b^\star|$ with respect to $\|\cdot\|_{1,[\eta^\star]}$.

Remark 17. *In the event that $\lceil A \rceil_{\mathrm{Mzr}}$ is reducible, the results from Theorem 16 still provide tests for contraction of the Hopfield model. Consider, for example, case (iii) above. The model is strongly infinitesimally contracting provided that $\mathrm{osL}(f_{\mathrm{H}}) < 0$, and if $\lceil A \rceil_{\mathrm{Mzr}}$ is reducible, by Lemma 1 for every $\epsilon > 0$, there exists $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_{1,[\eta]}(\lceil A \rceil_{\mathrm{Mzr}}) \leq \alpha(\lceil A \rceil_{\mathrm{Mzr}}) + \epsilon$. Thus, if (16) is negative, then $-c + \max\{d_{\max}(\alpha(\lceil A \rceil_{\mathrm{Mzr}}) + \epsilon), d_{\min}(\alpha(\lceil A \rceil_{\mathrm{Mzr}}) + \epsilon)\}$ may be made negative as well by taking ϵ small enough.*

B. Firing-rate neural network model

A related model, which is frequently used in the machine learning literature and is closely-related to the Hopfield neural network model is the model

$$\dot{x} = -Cx + \Phi(Ax + u) =: f_{\mathrm{FR}}(x), \tag{18}$$

which we refer to as the *firing-rate model*. 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 is interpreted as a vector of firing-rates, while in the Hopfield model, x can be negative and is thus interpreted as a vector of membrane potentials.

In what follows, we show that while the Hopfield model is naturally one-sided Lipschitz with respect to a diagonally-weighted ℓ_1 norm, the firing-rate model is naturally one-sided Lipschitz with respect to a diagonally-weighted ℓ_∞ norm.

Theorem 18 (One-sided Lipschitzness of firing-rate model)**.** *Consider the firing-rate model (18) with invertible A and irreducible $\lceil A \rceil_{\mathrm{Mzr}}$ and constant $u \in \mathbb{R}^n$. Then*

(i) $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}}) = \max\{\mu_{\infty,[\eta]^{-1}}(-C + d_{\min}A), \mu_{\infty,[\eta]^{-1}}(-C + d_{\max}A)\}$*, for arbitrary $\eta \in \mathbb{R}^n_{>0}$*

(ii) The choice of η minimizing $\mathrm{osL}_{\infty,[\eta]^{-1}}(f_{\mathrm{FR}})$ is the solution to

$$\begin{aligned} \inf_{b \in \mathbb{R}, \eta \in \mathbb{R}^n_{>0}} \quad & b \\ \textit{s.t.} \quad & (-C + d_{\min}\lceil A \rceil_{\mathrm{Mzr}})\eta \leq b\eta, \\ & (-C + d_{\max}\lceil A \rceil_{\mathrm{Mzr}})\eta \leq b\eta. \end{aligned}$$

(iii) if $C = cI_n$, then, with $v_A \in \mathbb{R}^n_{>0}$ being the right dominant eigenvector of $\lceil A \rceil_{\mathrm{Mzr}}$,

$$\begin{aligned} \inf_{\eta \in \mathbb{R}^n_{>0}} \mathrm{osL}_{\infty,[\eta]}(f_{\mathrm{FR}}) &= \mathrm{osL}_{\infty,[v_A]^{-1}}(f_{\mathrm{FR}}) \\ &= -c + \max\{d_{\min}\alpha(\lceil A \rceil_{\mathrm{Mzr}}), d_{\max}\alpha(\lceil A \rceil_{\mathrm{Mzr}})\}. \end{aligned} \tag{19}$$

(iv) if $d_{\min} = 0$ and $C \succ 0$, then, with $v_* \in \mathbb{R}^n_{>0}$ being the right dominant eigenvector of $-C + d_{\max}\lceil A \rceil_{\mathrm{Mzr}}$,

$$\begin{aligned} \inf_{\eta \in \mathbb{R}^n_{>0}} \mathrm{osL}_{\infty,[\eta]}(f_{\mathrm{FR}}) &= \mathrm{osL}_{\infty,[v_*]^{-1}}(f_{\mathrm{FR}}) \\ &= \max\left\{\alpha(-C), \alpha(-C + d_{\max}\lceil A \rceil_{\mathrm{Mzr}})\right\}. \end{aligned} \tag{20}$$

C. Other related models

We apply Theorem 16 and the log norm results in Lemma 6 to the following related neural circuit models, all of which are studied in the classic book [21]. In the following Theorems, we assume all Metzler matrices are irreducible.

Theorem 19 (Contractivity of special Hopfield models)**.** *(i) If $A \in \mathcal{MH}$, and $d_{\min} > 0$, the Persidskii-type[1] model*

$$\dot{x} = A\Phi(x)$$

is strongly infinitesimally contracting with respect to norm $\|\cdot\|_{1,[w_A]}$ with rate $d_{\min}|\alpha(\lceil A \rceil_{\mathrm{Mzr}})|$.

(ii) If $-C + d_{\max}A \in \mathcal{MH}$, the Hopfield neural network f_{H} with $d_{\min} = 0$ and positive diagonal C is strongly infinitesimally contracting with respect to $\|\cdot\|_{1,[w_*]}$ with rate $-\max\left\{\alpha(-C), \alpha(-C + d_{\max}\lceil A \rceil_{\mathrm{Mzr}})\right\} > 0$.

Theorem 20. *From [21, Theorem 3.2.4], consider*

$$\dot{x} = Ax - C\Phi(x),$$

with diagonal $C \succeq 0$. If $A - d_{\min}C \in \mathcal{MH}$ with corresponding dominant left eigenvector w_{**}, then this model is strongly infinitesimally contracting with respect to $\|\cdot\|_{1,[w_{**}]}$ with rate $-\alpha(\lceil A \rceil_{\mathrm{Mzr}} - d_{\min}C) > 0$.

Theorem 21. *From [21, Theorem 3.2.10], consider*

$$\dot{x}_i = \sum\nolimits_{j=1}^n a_{ij}\phi_{ij}(x_j)$$

[1] See[21, Definition 3.2.1]

for each $i \in \{1, \dots, n\}$ and with each ϕ_{ij} Lipschitz and slope-restricted in $[d_{\min}, d_{\max}]$. If $d_{\min} > 0$ and

$$B := d_{\max} A - (d_{\max} - d_{\min})(I_n \circ A) \in \mathcal{MH},$$

with corresponding dominant left and right eigenvectors w_B, v_B, respectively, then this model is strongly infinitesimally contracting with rate $-\alpha(\lceil B \rceil_{\text{Mzr}}) > 0$ with respect to both $\|\cdot\|_{1,[w_B]}$ and $\|\cdot\|_{\infty,[v_B]^{-1}}$.

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 [41].

Theorem 22 (Contractivity of Lur'e system)**.** *From [21, Theorem 3.2.7], consider the Lur'e system*

$$\begin{aligned} \dot{x} &= Ax + v\phi(y), \\ y &= w^\top x, \end{aligned}$$

where $A \in \mathbb{R}^{n \times n}, v, w \in \mathbb{R}^n$ and $\phi : \mathbb{R} \to \mathbb{R}$ is Lipschitz and slope-restricted in $[d_{\min}, d_{\max}]$. Consider the following two optimization problems:

$$\begin{aligned} \min_{b \in \mathbb{R}, \eta \in [\varepsilon, \infty[^n} \quad & b \\ \text{s.t.} \quad & \lceil A + d_{\min} v w^\top \rceil_{\text{Mzr}}^\top \eta \leq b\eta, \\ & \lceil A + d_{\max} v w^\top \rceil_{\text{Mzr}}^\top \eta \leq b\eta, \end{aligned} \tag{21}$$

and

$$\begin{aligned} \min_{c \in \mathbb{R}, \xi \in [\varepsilon, \infty[^n} \quad & c \\ \text{s.t.} \quad & \lceil A + d_{\min} v w^\top \rceil_{\text{Mzr}} \xi \leq c\xi, \\ & \lceil A + d_{\max} v w^\top \rceil_{\text{Mzr}} \xi \leq c\xi. \end{aligned} \tag{22}$$

Let $b^\star, \eta^\star$ be optimal parameters for (21) and $c^\star, \xi^\star$ be optimal parameters for (22). Then

(i) if $b^\star < 0$, then the closed-loop dynamics are strongly infinitesimally contracting with rate $|b^\star|$ with respect to $\|\cdot\|_{1,[\eta^\star]}$.

(ii) if $c^\star < 0$, then the closed-loop dynamics are strongly infinitesimally contracting with rate $|c^\star|$ with respect to $\|\cdot\|_{\infty,[\xi^\star]^{-1}}$.

Theorem 23 (Multivariable Lur'e system)**.** *Consider the multivariable Lur'e system*

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

where $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, C \in \mathbb{R}^{m \times n}$, and Φ is diagonal and is slope-restricted in $[d_{\min}, d_{\max}]$ with $d_{\min} \geq 0$. Define $(\cdot)_+$ and $(\cdot)_-$ by $(x)_+ = \max\{x, 0\}$ and $(x)_- = \min\{x, 0\}$. Define $F \in \mathbb{R}^{n \times n}$ componentwise by

$$\begin{aligned} F_{ii} &= A_{ii} + d_{\max} \sum_{j=1}^m (B_{ij} C_{ji})_+ + d_{\min} \sum_{j=1}^m (B_{ij} C_{ji})_-, \\ F_{ij} &= |A_{ij}| \\ &\quad + \max \Big\{ d_{\max} \sum_{k=1}^m (B_{ik} C_{kj})_+ + d_{\min} \sum_{k=1}^m (B_{ik} C_{kj})_-, \\ &\qquad -d_{\min} \sum_{k=1}^m (B_{ik} C_{kj})_+ - d_{\max} \sum_{k=1}^m (B_{ik} C_{kj})_- \Big\}, \end{aligned}$$

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

D. Remarks on absolute, connective, and total contractivity

In this section we clarify that our results in Theorems 16–23 indeed establish absolute, connective, and total contraction, in the following senses.

First, in the spirit of the classic work on absolute stability [41], [13], by *absolutely contracting* we mean dynamical systems that are contracting for all choices of activation functions in a given class. (The class of activation function in this paper is all weakly increasing Lipschitz functions.)

Second, in the spirit of the classic work on connective stability [36], by *connectively contracting* we mean dynamical networks that remain contracting under the removal of any possible subset of edges (other than self-loops). It is easy to see that the action of removing any edge from a synaptic matrix leads to an equal or larger contraction rate.

Third and final, if each component of the state x corresponds to a single neuron, the removal of some neurons corresponds to *pruning* the neural network. By Corollary 11, if $A \in \mathcal{MH}$, then any principal submatrix of A is also in $\mathcal{MH}$. In other words, if any neurons are removed from the neural network, the resulting neural network is guaranteed to remain contracting with an equal or larger contraction rate. We refer to this property as *total contraction*, because of the analogy with the property of totally Hurwitz matrices.

VI. Discussion

In this paper, we present novel non-Euclidean log norm results and a non-smooth contraction theory simplification and we apply these results to study the contractivity of RNN models, primarily focusing on the Hopfield and firing-rate models. We provide efficient algorithms for computing the optimal non-Euclidean contraction rate and corresponding norm. 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 contractivity under conditions such as Lyapunov diagonal stability (LDS) of the synaptic matrix. LDS is known to imply asymptotic stability of Hopfield neural networks, however, it is not known to imply contractivity (with respect to a constant norm). More broadly, we believe that our non-Euclidean contraction framework for RNNs serves as a first step to analyzing robustness and convergence properties of other classes of neural circuits including central pattern generators and other machine learning architectures including modern Hopfield networks [23].

Acknowledgments. The authors wish to thank Dr. Saber Jafarpour for stimulating conversations and valuable insights about contraction theory and neural networks.

References

[1] Z. Aminzare and E. D. Sontag. Contraction methods for nonlinear systems: A brief introduction and some open problems. In *IEEE Conf. on Decision and Control*, pages 3835–3847, December 2014. doi:10.1109/CDC.2014.7039986.

[2] Z. Aminzare and E. D. Sontag. Synchronization of diffusively-connected nonlinear systems: Results based on contractions with respect to general norms. *IEEE Transactions on Network Science and Engineering*, 1(2):91–106, 2014. doi:10.1109/TNSE.2015.2395075.

[3] S. Arik. A note on the global stability of dynamical neural networks. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 49(4):502–504, 2002. doi:10.1109/81.995665.

[4] S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, 2019. URL: https://arxiv.org/abs/1909.01377.

[5] G. P. Barker, A. Berman, and R. J. Plemmons. Positive diagonal solutions to the Lyapunov equations. *Linear and Multilinear Algebra*, 5(4):249–256, 1978. doi:10.1080/03081087808817203.

[6] F. Bullo. *Lectures on Network Systems*. Kindle Direct Publishing, 1.6 edition, January 2022, ISBN 978-1986425643. URL: http://motion.me.ucsb.edu/book-lns.

[7] F. Bullo, P. Cisneros-Velarde, A. Davydov, and S. Jafarpour. From contraction theory to fixed point algorithms on Riemannian and non-Euclidean spaces. In *IEEE Conf. on Decision and Control*, December 2021. doi:10.1109/CDC45484.2021.9682883.

[8] S. Coogan. A contractive approach to separable Lyapunov functions for monotone systems. *Automatica*, 106:349–357, 2019. doi:10.1016/j.automatica.2019.05.001.

[9] A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 2022. doi:10.1109/TAC.2022.3183966.

[10] C. A. Desoer and H. Haneda. The measure of a matrix as a tool to analyze computer algorithms for circuit analysis. *IEEE Transactions on Circuit Theory*, 19(5):480–486, 1972. doi:10.1109/TCT.1972.1083507.

[11] Y. Fang and T. G. Kincaid. Stability analysis of dynamical neural networks. *IEEE Transactions on Neural Networks*, 7(4):996–1006, 1996. doi:10.1109/72.508941.

[12] M. Fazlyab, M. Morari, and G. J. Pappas. Safety verification and robustness analysis of neural networks via quadratic constraints and semidefinite programming. *IEEE Transactions on Automatic Control*, 2020. doi:10.1109/TAC.2020.3046193.

[13] M. Forti, S. Manetti, and M. Marini. Necessary and sufficient condition for absolute stability of neural networks. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 41(7):491–494, 1994. doi:10.1109/81.298364.

[14] M. Forti and A. Tesi. New conditions for global stability of neural networks with application to linear and quadratic programming problems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 42(7):354–366, 1995. doi:10.1109/81.401145.

[15] W. He and J. Cao. Exponential synchronization of chaotic neural networks: a matrix measure approach. *Nonlinear Dynamics*, 55:55–65, 2009. doi:10.1007/s11071-008-9344-4.

[16] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10):3088–3092, 1984. doi:10.1073/pnas.81.10.3088.

[17] S. Jafarpour, A. Davydov, and F. Bullo. Non-Euclidean contraction theory for monotone and positive systems. *IEEE Transactions on Automatic Control*, September 2021. URL: https://arxiv.org/abs/2104.01321.

[18] S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, December 2021. URL: http://arxiv.org/abs/2106.03194.

[19] A. Kag, Z. Zhang, and V. Saligrama. RNNs incrementally evolving on an equilibrium manifold: A panacea for vanishing and exploding gradients? In *International Conference on Learning Representations*, 2020. URL: https://openreview.net/forum?id=HylpqA4FwS.

[20] E. Kaszkurewicz and A. Bhaya. On a class of globally stable neural circuits. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 41(2):171–174, 1994. doi:10.1109/81.269055.

[21] E. Kaszkurewicz and A. Bhaya. *Matrix Diagonal Stability in Systems and Computation*. Springer, 2000, ISBN 978-0-8176-4088-0.

[22] L. Kozachkov, M. Ennis, and J.-J. E Slotine. Recursive construction of stable assemblies of recurrent neural networks, 2021. URL: https://arxiv.org/abs/2106.08928.

[23] D. Krotov and J. J. Hopfield. Large associative memory problem in neurobiology and machine learning. In *International Conference on Learning Representations*, 2021. URL: https://openreview.net/forum?id=X4y_10OX-hX.

[24] W. Lohmiller and J.-J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. doi:10.1016/S0005-1098(98)00019-3.

[25] A. N. Michel, J. A. Farrell, and W. Porod. Qualitative analysis of neural networks. *IEEE Transactions on Circuits and Systems*, 36(2):229–243, 1989. doi:10.1109/31.20200.

[26] K. D. Miller and F. Fumarola. Mathematical equivalence of two common forms of firing rate models of neural networks. *Neural Computation*, 24(1):25–31, 2012. doi:10.1162/NECO_a_00221.

[27] P. J. Moylan. Matrices with positive principal minors. *Linear Algebra and its Applications*, 17(1):53–58, 1977. doi:10.1016/0024-3795(77)90040-4.

[28] E. Nozari and J. Cortés. Hierarchical selective recruitment in linear-threshold brain networks—part I: Single-layer dynamics and selective inhibition. *IEEE Transactions on Automatic Control*, 66(3):949–964, 2021. doi:10.1109/TAC.2020.3004801.

[29] O. Pastravanu and M. Voicu. Generalized matrix diagonal stability and linear dynamical systems. *Linear Algebra and its Applications*, 419(2):299–310, 2006. doi:10.1016/j.laa.2006.04.021.

[30] A. Pavlov and N. Van de Wouw. Convergent systems: Nonlinear simplicity. In N. van de Wouw, E. Lefeber, and A. I. Lopez, editors, *Nonlinear Systems*, pages 51–77. Springer, 2017. doi:10.1007/978-3-319-30357-4_3.

[31] H. Qiao, J. Peng, and Z.-B. Xu. Nonlinear measures: A new approach to exponential stability analysis for Hopfield-type neural networks. *IEEE Transactions on Neural Networks*, 12(2):360–370, 2001. doi:10.1109/72.914530.

[32] M. Revay, R. Wang, and I. R. Manchester. Lipschitz bounded equilibrium networks. 2020. URL: https://arxiv.org/abs/2010.01732.

[33] M. Revay, R. Wang, and I. R. Manchester. A convex parameterization of robust recurrent neural networks. *IEEE Control Systems Letters*, 5(4):1363–1368, 2021. doi:10.1109/LCSYS.2020.3038221.

[34] G. Russo, M. Di Bernardo, and E. D. Sontag. Global entrainment of transcriptional systems to periodic inputs. *PLoS Computational Biology*, 6(4):e1000739, 2010. doi:10.1371/journal.pcbi.1000739.

[35] E. K. Ryu and W. Yin. *Large-Scale Convex Optimization via Monotone Operators*. Cambridge, 2022.

[36] D. D. Šiljak. *Large-Scale Dynamic Systems Stability & Structure*. North-Holland, 1978, ISBN 0486462854.

[37] G. Söderlind. The logarithmic norm. History and modern theory. *BIT Numerical Mathematics*, 46(3):631–652, 2006. doi:10.1007/s10543-006-0069-9.

[38] J. Stoer and C. Witzgall. Transformations by diagonal matrices in a normed space. *Numerische Mathematik*, 4:158–171, 1962. doi:10.1007/BF01386309.

[39] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. In *International Conference on Learning Representations*, 2014. URL: https://arxiv.org/abs/1312.6199.

[40] H. Tsukamoto, S.-J. Chung, and J.-J. E Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview. *Annual Reviews in Control*, 52:135–169, 2021. doi:10.1016/j.arcontrol.2021.10.001.

[41] V. A. Yakubovich. Method of matrix inequalities in theory of nonlinear control systems stability. I. Forced oscillations absolute stability. *Avtomatika i Telemekhanika*, 25(7):1017–1029, 1964. (In Russian). URL: http://mi.mathnet.ru/eng/at11685.

[42] H. Zhang, Z. Wang, and D. Liu. A comprehensive review of stability analysis of continuous-time recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 25(7):1229–1262, 2014. doi:10.1109/TNNLS.2014.2317880.