IEEE Control Systems Letters paper presented at
2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Euclidean Contractivity of Neural Networks with Symmetric Weights

Veronica Centorrino[a], Anand Gokhale[b], Alexander Davydov[b], Giovanni Russo[c], and Francesco Bullo[b]

***Abstract*— This paper investigates stability conditions of continuous-time Hopfield and firing-rate neural networks by leveraging contraction theory. First, we present a number of useful general algebraic results on matrix polytopes and products of symmetric matrices. Then, we give sufficient conditions for strong and weak Euclidean contractivity, i.e., contractivity with respect to the ℓ_2 norm, of both models with symmetric weights and (possibly) non-smooth activation functions. Our contraction analysis leads to contraction rates which are log-optimal in almost all symmetric synaptic matrices. Finally, we use our results to propose a firing-rate neural network model to solve a quadratic optimization problem with box constraints.**

I. Introduction

Continuous-time recurrent neural networks (RNNs) are dynamical models widely studied in computational neuroscience and machine learning. Recent interest has focused on establishing the contractivity properties of RNNs. Contracting dynamics are robustly stable, feature computationally friendly methods for equilibrium computation, and enjoy many other properties. Motivated by optimization [17], [2] and neuroscientific applications [15], [7, Chapter 17], this paper focuses on symmetric synaptic interactions.

While a comprehensive contractivity analysis with respect to ℓ_1 and ℓ_∞ norms was recently presented in [5], the corresponding analysis with respect to weighted Euclidean norms is not complete yet. A recent breakthrough in this direction was obtained by [10]; this work extends and complements these results (a detailed comparison is offered below).

Two common models of RNNs are the *firing-rate neural network (FNN)* and *Hopfield neural network (HNN)*; the main difference being the order by which the activation function acts. Under mild assumptions, FNNs are positive systems and, arguably, more biologically-plausible. HNNs are relevant in optimization and machine learning [17], [2], [15], [20]. For certain synaptic matrices and initial conditions, FNN and HNN are known to be equivalent via an appropriate change of coordinates and input transformation [13]. However, the understanding of this partial correspondence is not complete and, as we will show below, their contractivity properties are not exactly coincident.

This work was in part supported by AFOSR project FA9550-21-1-0203 and NSF Graduate Research Fellowship under Grant No. 2139319. The authors thank Dr. Leo Kozachkov for insightful comments.

[a] Scuola Superiore Meridionale, Naples, Italy. veronica.centorrino@unina.it.

[b] Center for Control, Dynamical Systems, and Computation, UC Santa Barbara, Santa Barbara, CA 93106 USA. anand_gokhale@ucsb.edu, davydov@ucsb.edu, bullo@ucsb.edu.

[c] Department of Information and Electric Engineering and Applied Mathematics, University of Salerno, Italy. giovarusso@unisa.it.

a) Related literature: RNNs naturally emerge when modelling neural processes [7]. Critical questions when studying RNNs are related to finding conditions that guarantee stability and robustness of the network. For example, sufficient conditions for the stability of HNNs are given in [6] based on the use of Lyapunov diagonally stable matrices. Stability and robustness can be simultaneously established using contraction theory. Indeed, contracting systems exhibit highly ordered transient and asymptotic behaviors that appear to be convenient in the context of RNNs. For example: (i) initial conditions are exponentially forgotten [12]; (ii) time-invariant dynamics admits a unique globally exponential stable equilibrium [12]; (iii) contraction ensures entrainment to periodic inputs [16] and (iv) enjoy highly robust behavior, such as input-to-state stability [19]. (v) Moreover, efficient numerical algorithms can be devised for numerical integration and fixed point computation of contracting systems [9]. Recently, non-Euclidean contractivity of RNNs is studied in [5] and in [4], where stability properties of HNN and FNN with dynamic synapses undergoing Hebbian learning are proposed. Euclidean contractivity is studied in [11] to analyze the stability of RNNs with dynamic synapses and in [10], where a number of contractivity conditions are proposed. Finally, the design of norms minimizing the logarithmic norm is reviewed in [3, Section 2.7].

b) Contributions: our main results are a set of sufficient conditions characterizing strong and weak infinitesimal contractivity properties (see Section II for the definitions) of FNNs and HNNs with symmetric weights and possibly non-smooth activation functions. We also establish a lower bound on the contraction rate and, remarkably, demonstrate that the bound is log-optimal in almost all symmetric weight matrices. One of the main benefits of our approach to the study of FNNs and HNNs is that, with just a single condition, it ensures global exponential convergence, along with all the other useful properties of contracting systems. The main results leverage a number of general algebraic results, which are interesting *per se* and are also a contribution of this paper. With these algebraic results, we: (i) determine a weighted ℓ_2 norm for matrix polytopes which is log-optimal for almost all synaptic matrices; (ii) give a lower bound on the spectral abscissa of matrix polytopes; (iii) provide optimal and log-optimal norms for the product of symmetric matrices. Finally, we leverage our sufficient conditions for contractivity to propose a FNN solving certain quadratic optimization problems with box constraints.

Our results for strong infinitesimal contractivity of the FNN and HNN models with symmetric weights are based on and generalize [10, Theorem 2]. Specifically, (i) we provide the explicit expression of the matrix weights for which the models are contracting. The matrices we find are different for the two models, highlighting the importance of choosing the appropriate model based on the properties being studied; (ii) we address the weak contractivity case, i.e., when the contraction rate is 0, making it applicable for, e.g., systems that enjoy conservation or invariance properties; (iii) we handle weakly increasing and (iv) locally Lipschitz activation functions, allowing us to consider common activation functions such as the rectified linear unit (ReLU) and soft thresholding functions.[1]

II. Mathematical Preliminaries

Define $(\cdot)_+ \colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ by $(z)_+ = z$ if $z > 0$, $(z)_+ = 0$ if $z \leq 0$. Given $x \in \mathbb{R}^n$, let $[x] \in \mathbb{R}^{n\times n}$ denote the diagonal matrix with diagonal entries equal to x. Vector inequalities of the form $x \leq (\geq)\ y$ are entry-wise. Let $\mathbb{1}_n$, $\mathbb{0}_n \in \mathbb{R}^n$ be the all-ones and all-zeros vectors, respectively, I_n be the $n \times n$ identity matrix, and $\mathbb{S}^n$ be the set of real symmetric $n \times n$ matrices. For $A \in \mathbb{R}^{n\times n}$, let $\mathrm{spec}(A)$, $\rho(A) := \max\{|\lambda| \mid \lambda \in \mathrm{spec}(A)\}$ and $\alpha(A) := \max\{\Re(\lambda) \mid \lambda \in \mathrm{spec}(A)\}$ denote the spectrum, spectral radius and the spectral abscissa of A, respectively; here $\Re(\lambda)$ denotes the real part of λ. For $A \in \mathbb{S}^n$, let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its minimum and maximum eigenvalue, respectively. Given $A, B \in \mathbb{S}^n$, we write $A \preceq B$ (resp. $A \prec B$) if $B - A$ is positive semidefinite (resp. definite). The Moore–Penrose inverse of $A \in \mathbb{R}^{n\times n}$ is the unique matrix $A^\dagger \in \mathbb{R}^{n\times n}$ such that $AA^\dagger A = A$, $A^\dagger A A^\dagger = A^\dagger$, with $AA^\dagger$, $A^\dagger A \in \mathbb{S}^n$. When clear from the context, we omit to specify time dependence of functions.

1) Norms and induced norms: Let $\|\cdot\|$ denote both a norm on $\mathbb{R}^n$ and its corresponding induced matrix norm on $\mathbb{R}^{n\times n}$. Given $A \in \mathbb{R}^{n\times n}$ the *logarithmic norm* (log-norm) induced by $\|\cdot\|$ is $\mu(A) := \lim_{h\to 0^+} \frac{\|I_n + hA\| - 1}{h}$. For an ℓ_p norm, $p \in [1, \infty]$, and for an invertible $Q \in \mathbb{R}^{n\times n}$, the Q-weighted ℓ_p norm is $\|x\|_{p,Q} := \|Qx\|_p$. The corresponding log-norm is $\mu_{p,Q}(A) = \mu_p(QAQ^{-1})$. For two invertible matrices Q_1, $Q_2 \in \mathbb{R}^{n\times n}$, it holds

$$\mu_{p,Q_1Q_2}(A) = \mu_{p,Q_1}(Q_2 A Q_2^{-1}). \tag{1}$$

Given $f\colon \mathbb{R}_{\geq 0} \times C \to \mathbb{R}^n$, with $C \subseteq \mathbb{R}^n$ open and connected, we denote by $\mathsf{osL}(f_t)$ the *one-sided Lipschitz constant* of $f_t := f(t, \cdot)$. For continuously differentiable f_t and convex set C it holds $\mathsf{osL}(f_t) = \sup_{x\in C} \mu(Df(t,x))$, where $Df(t,x) := \partial f(t,x)/\partial x$ is the Jacobian of f with respect to x. We write $\mathsf{osL}_{p,Q}(f_t)$ to specify that the one-sided Lipschitz constant is computed with respect to a Q-weighted ℓ_p norm.

We refer to [3] for a recent review of those tools.

2) Contraction theory for dynamical systems: We start with the following

Definition 1. *Given a norm, a function $f\colon \mathbb{R}_{\geq 0} \times C \to \mathbb{R}^n$, with $C \subseteq \mathbb{R}^n$ f-invariant, open and convex, and a constant $c > 0$ ($c = 0$) referred as* contraction rate*, f is strongly (weakly) infinitesimally contracting on C if*

$$\mathsf{osL}(f_t) \leq -c, \text{ for all } t \in \mathbb{R}_{\geq 0},$$

or, equivalently for differentiable vector fields, if

$$\mu(Df(t,x)) \leq -c, \text{ for all } x \in C \text{ and } t \in \mathbb{R}_{\geq 0}. \tag{2}$$

The next result [5, Theorem 16] allows using condition (2) for locally Lipschitz function, for which, by Rademacher's theorem, $Df(t,x)$ exists almost everywhere (a.e.) in C.

Theorem 1. *Consider a norm, a function $f\colon \mathbb{R}_{\geq 0} \times C \to \mathbb{R}^n$ locally Lipschitz on $C \subset \mathbb{R}^n$ open and convex set. Then for every $c \in \mathbb{R}$ the following statements are equivalent:*

(i) $\mathsf{osL}(f_t) \leq c$, for all $t \in \mathbb{R}_{\geq 0}$,
(ii) $\mu(Df(t,x)) \leq c$, for a.e. $x \in C$ and $t \in \mathbb{R}_{\geq 0}$.

3) Hopfield and firing-rate continuous-time neural networks: We are interested in the following continuous-time FNN and HNN models defined, respectively, as:

$$\dot{x}_{\mathrm{F}} = -x_{\mathrm{F}} + \Phi(Wx_{\mathrm{F}} + u_{\mathrm{F}}) := f_{\mathrm{F}}(x_{\mathrm{F}}, u_{\mathrm{F}}), \tag{3}$$
$$\dot{x}_{\mathrm{H}} = -x_{\mathrm{H}} + W\Phi(x_{\mathrm{H}}) + u_{\mathrm{H}} := f_{\mathrm{H}}(x_{\mathrm{H}}, u_{\mathrm{H}}), \tag{4}$$

where: x_{F}, $x_{\mathrm{H}} \in \mathbb{R}^n$ are neural activation vectors, $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear and diagonal activation function, i.e., for $x \in \mathbb{R}^n$, $(\Phi(x))_i = \phi(x_i)$, where $\phi\colon \mathbb{R} \to \mathbb{R}$. $W \in \mathbb{R}^{n\times n}$ is the synaptic matrix, with $W_{ij} \in \mathbb{R}$ being the synaptic weight from neuron j to neuron i. Finally, u_{F}, $u_{\mathrm{H}} \in \mathbb{R}^n$ are the external stimuli in the FNN and HNN, respectively. The models (3) and (4) assume homogeneous dissipation rates; we leave the heterogeneous case to future work.

Remark 2. *When the activation function is non-negative the positive orthant is forward-invariant for f_{F} in (3) and x_{F} is interpreted as a firing-rate. Instead, in (4) x_{H} is sign indefinite and is interpreted as a membrane potential.*

III. Main Results

First, we give algebraic results on weighted ℓ_2 norms of certain matrix polytopes. Then, we use those results to give sufficient conditions for the strong infinitesimal contractivity of the FNN and the HNN with symmetric weights with respect to weighted Euclidean norms.

Assumption 1 (Symmetric synaptic weights)**.** *The synaptic matrix $W \in \mathbb{R}^{n\times n}$ is symmetric.*

Under Assumption 1, the eigenvalues of W are real, $\alpha(W) = \lambda_{\max}(W)$ and $W \preceq \alpha(W) I_n$. Moreover, W can be decomposed as

$$W = U\Lambda U^\top, \tag{5}$$

where $U \in \mathbb{R}^{n\times n}$ is the orthogonal matrix whose columns are the eigenvectors of W, and $\Lambda = [\lambda] \in \mathbb{R}^{n\times n}$ is diagonal with $\lambda \in \mathbb{R}^n$ being the vector of the eigenvalues of W.

Given $b > 0$, we define $\theta_b \colon]-\infty, b] \to [2b, +\infty[$ by

$$\theta_b(z) := 2b\big(1 + \sqrt{1 - z/b}\big), \quad \forall z \in]-\infty, b]. \tag{6}$$

[1]See https://arxiv.org/abs/2302.13452 for an extended technical report.

For our derivations, it is useful to denote $\theta_b(\Lambda) := [(\theta_b(\lambda_1), \dots, \theta_b(\lambda_n))]$. Also, we introduce $Q_{\mathrm{F},b} \in \mathbb{R}^{n\times n}$

$$Q_{\mathrm{F},b} := U\theta_b(\Lambda)U^\top \succ 0, \tag{7}$$

and, when W is invertible, $Q_{\mathrm{H},b} \in \mathbb{R}^{n\times n}$ is defined as

$$Q_{\mathrm{H},b} := Q_{\mathrm{F},b}W^{-1} = U\theta_b(\Lambda)\Lambda^{-1}U^\top \succ 0. \tag{8}$$

A. Results on Euclidean log-norms of matrix polytopes

First, we give the following definition for polytopes.

Definition 2 (Log-optimal and log ε-optimal norms for matrix polytopes)**.** *Given $A_1, \dots, A_m \in \mathbb{R}^{n\times n}$, consider the polytope $\mathcal{P} = \{\sum_{j=1}^m \beta_j A_j \mid \beta_j \geq 0, \sum_{j=1}^m \beta_j = 1\}$ and a scalar $\varepsilon > 0$. We say that the norm $\|\cdot\|$ is*

(i) logarithmically optimal (log-optimal) for $\mathcal{P}$ *if*

$$\max_{A\in\mathcal{P}} \alpha(A) = \max_{j\in\{1,\dots,m\}} \mu(A_j),$$

(ii) logarithmically ε-optimal (log ε-optimal) for $\mathcal{P}$ *if*

$$\max_{A\in\mathcal{P}} \alpha(A) \leq \max_{j\in\{1,\dots,m\}} \mu(A_j) \leq \max_{A\in\mathcal{P}} \alpha(A) + \varepsilon.$$

We are specifically interested in the matrix polytopes $\mathcal{P}_\mathrm{F} := \{[d]W \mid d \in [0,1]^n\}$ and $\mathcal{P}_\mathrm{H} := \{W[d] \mid d \in [0,1]^n\}$. Namely, in Theorem 3 we give algebraic results on the Euclidean log-norm of matrices in $\mathcal{P}_\mathrm{F}$ and $\mathcal{P}_\mathrm{H}$ (the proof is in Section IV, together with a number of instrumental results).

Theorem 3 (Euclidean log-norm of matrix polytopes)**.** *Given a symmetric synaptic matrix W (Assumption 1), the following statements holds:*

(i) if $\alpha(W) > 0$, then $\|\cdot\|_{2,Q_{\mathrm{F},\alpha(W)}}$, with $Q_{\mathrm{F},\alpha(W)} \in \mathbb{R}^{n\times n}$ defined in (7), is log-optimal for $\mathcal{P}_\mathrm{F}$, i.e.,

$$\max_{d\in[0,1]^n} \mu_{2,Q_{\mathrm{F},\alpha(W)}}([d]W) = \max_{d\in[0,1]^n} \alpha([d]W) = \alpha(W).$$

In addition, if W is invertible, then $\|\cdot\|_{2,Q_{\mathrm{H},\alpha(W)}}$, with $Q_{\mathrm{H},\alpha(W)} \in \mathbb{R}^{n\times n}$ defined in (8), is log-optimal for $\mathcal{P}_\mathrm{H}$, i.e.,

$$\max_{d\in[0,1]^n} \mu_{2,Q_{\mathrm{H},\alpha(W)}}(W[d]) = \max_{d\in[0,1]^n} \alpha(W[d]) = \alpha(W);$$

(ii) if $\alpha(W) = 0$, then for each $\varepsilon > 0$ the norm $\|\cdot\|_{2,Q_{\mathrm{F},\varepsilon}}$, with $Q_{\mathrm{F},\varepsilon} \in \mathbb{R}^{n\times n}$ defined in (7), is log ε-optimal for $\mathcal{P}_\mathrm{F}$, i.e.,

$$\max_{d\in[0,1]^n} \mu_{2,Q_{\mathrm{F},\varepsilon}}([d]W) \leq \max_{d\in[0,1]^n} \alpha([d]W) + \varepsilon = \varepsilon;$$

(iii) if $\alpha(W) < 0$, then $\|\cdot\|_{2,(-W)^{1/2}}$ is log-optimal for $\mathcal{P}_\mathrm{F}$ and $\mathcal{P}_\mathrm{H}$, i.e.,

$$\max_{d\in[0,1]^n} \mu_{2,(-W)^{1/2}}([d]W) = \max_{d\in[0,1]^n} \alpha([d]W) = 0,$$
$$\max_{d\in[0,1]^n} \mu_{2,(-W)^{1/2}}(W[d]) = \max_{d\in[0,1]^n} \alpha(W[d]) = 0.$$

Remark 4. *Theorem 3 applies to polytopes of the form $aI_n + [d]W$ and $aI_n + W[d]$, for $a \in \mathbb{R}$, via the translation property: for all $A \in \mathbb{R}^{n\times n}$, $\mu(A + aI_n) = \mu(A) + a$.*

B. Contractivity of recurrent neural networks

Next, we consider the neural network dynamics for the FNN in (3) and for the HNN in (4).

Assumption 2 (Slope-restricted activation function)**.** *The activation function $\phi\colon \mathbb{R} \to \mathbb{R}$ is Lipschitz and slope restricted in $[0,1]$.*

Assumption 2 ensures that $\phi'(x) \in [0,1]$ a.e. in $\mathbb{R}$. Many common activation functions, including ReLUs and sigmoids, satisfy Assumption 2, possibly after rescaling. In fact, if $\phi(\cdot)$ is slope restricted in $[0,\bar{d}]$, with $\bar{d} > 0$, then our next results still hold by redefining $[d] := D\Phi/\bar{d}$ and $W := \bar{d} \cdot W$.

a) Contractivity of firing-rate neural networks: We now provide an upper bound on the ℓ_2 one-sided Lipschitz constant and sufficient conditions for the Euclidean contractivity of FNNs with symmetric weights.

Theorem 5 (Euclidean one-sided Lipschitz constant of the FNN)**.** *Consider the FNN (3) satisfying Assumptions 1, 2,*

(i) if $\alpha(W) > 0$, then $\mathsf{osL}_{2,Q_{\mathrm{F},\alpha(W)}}(f_\mathrm{F}) \leq -1+\alpha(W)$, with $Q_{\mathrm{F},\alpha(W)} \in \mathbb{R}^{n\times n}$ defined in (7);

(ii) if $\alpha(W) = 0$, then for each $\varepsilon > 0$, $\mathsf{osL}_{2,Q_{\mathrm{F},\varepsilon}}(f_\mathrm{F}) \leq -1+\varepsilon$, with $Q_{\mathrm{F},\varepsilon} \in \mathbb{R}^{n\times n}$ defined in (7);

(iii) if $\alpha(W) < 0$, then $\mathsf{osL}_{2,(-W)^{1/2}}(f_\mathrm{F}) \leq -1$.

Proof. Regarding part (i) note that for almost all $x \in \mathbb{R}^n$ we have $Df_\mathrm{F}(x) = -I_n + D\Phi(Wx+u)W$ and

$$\mu_{2,Q_{\mathrm{F},\alpha(W)}}(Df_\mathrm{F}(x)) \leq \max_{d\in[0,1]^n} \mu_{2,Q_{\mathrm{F},\alpha(W)}}(-I_n + [d]W)$$
$$= -1+\alpha(W),$$

where the last equality follows by the log-norm translation property and part (i) in Theorem 3. The proof follows by applying Theorem 1. Parts (ii) and (iii) can be proved similarly, using parts (ii) and (iii) in Theorem 3. □

The next result follows from Theorem 5.

Corollary 6 (Euclidean contractivity of the FNN)**.** *Under the same assumptions and notations from Theorem 5,*

(i) if $\alpha(W) = 1$, then the FNN is weakly infinitesimally contracting with respect to $\|\cdot\|_{2,Q_{\mathrm{F},\alpha(W)}}$;

(ii) if $0 < \alpha(W) < 1$, then the FNN is strongly infinitesimally contracting with rate $1 - \alpha(W) > 0$ with respect to $\|\cdot\|_{2,Q_{\mathrm{F},\alpha(W)}}$;

(iii) if $\alpha(W) = 0$, then for any $0 < \varepsilon < 1$ the FNN is strongly infinitesimally contracting with rate $1 - \varepsilon > 0$ with respect to $\|\cdot\|_{2,Q_{\mathrm{F},\varepsilon}}$;

(iv) if $\alpha(W) < 0$, then the FNN is strongly infinitesimally contracting with rate 1 with respect to $\|\cdot\|_{2,(-W)^{1/2}}$.

b) Contractivity of Hopfield neural networks: We first provide an upper bound on the Euclidean one-sided Lipschitz constant and sufficient conditions for the ℓ_2 contractivity of HNNs with non-singular symmetric synaptic matrices. Then, we give sufficient conditions for the ℓ_2 contractivity with singular symmetric synapses. This latter result is proven in Section IV: differently from our analysis on FNNs, it requires a distinct mathematical approach.

Theorem 7 (Euclidean one-sided Lipschitz constant of the HNN with non-singular symmetric weights)**.** *Consider the HNN (4) satisfying Assumptions 1, 2 with non-singular weight matrix W,*

(i) if $\alpha(W) > 0$, then $\mathsf{osL}_{2,Q_{\mathrm{H},\alpha(W)}}(f_{\mathrm{H}}) \leq -1+\alpha(W)$, with $Q_{\mathrm{H},\alpha(W)} \in \mathbb{R}^{n\times n}$ defined in (8);

(ii) if $\alpha(W) < 0$, then $\mathsf{osL}_{2,(-W)^{1/2}}(f_{\mathrm{H}}) \leq -1$.

Proof. Regarding part (i), for almost all $x \in \mathbb{R}^n$ we have $Df_{\mathrm{H}}(x) = -I_n + WD\Phi(x)$ and $\mu_{2,Q_{\mathrm{H},\alpha(W)}}(Df_{\mathrm{H}}(x)) \leq \max_{d\in[0,1]^n} \mu_{2,Q_{\mathrm{H},\alpha(W)}}(-I_n + W[d]) = -1+\alpha(W)$, where the last equality follows by the log-norm translation property and part (i) in Theorem 3. The proof then follows by applying Theorem 1. Part (ii) can be proved similarly, using part (iii) in Theorem 3. □

Corollary 8 (Euclidean contractivity of the HNN with non-singular symmetric weights)**.** *Under the same assumptions and notations as in Theorem 7,*

(i) if $\alpha(W) = 1$, then the HNN is weakly infinitesimally contracting with respect to $\|\cdot\|_{2,Q_{\mathrm{H},\alpha(W)}}$;

(ii) if $0 < \alpha(W) < 1$, then the HNN is strongly infinitesimally contracting with rate $1-\alpha(W) > 0$ with respect to $\|\cdot\|_{2,Q_{\mathrm{H},\alpha(W)}}$;

(iii) if $\alpha(W) < 0$, then the HNN is strongly infinitesimally contracting with rate 1 with respect to $\|\cdot\|_{2,(-W)^{1/2}}$.

Finally, we give sufficient infinitesimal contractivity conditions of the HNN with singular symmetric synapses (see Section IV for the proof).

Theorem 9 (Contractivity of the HNN with singular symmetric weights)**.** *Consider the HNN (4) satisfying Assumptions 1, 2 with W having kernel $\mathcal{K} \neq \{\mathbb{0}_n\}$, and such that $\alpha(W) < 1$. Then, for each $\varepsilon \in\,]0, 1-\alpha(W)[$ the HNN is strongly infinitesimally contracting with rate $|1-\alpha(W)-\varepsilon|$.*

IV. Proofs and Additional Results

We now present the proofs of Theorems 3 and 9 and additional algebraic results on matrix polytopes and symmetric matrices. First, note that since the spectral abscissa is a continuous function, the maximum value over a polytope is greater than or equal to the value at one of its vertices. That is, for any $W \in \mathbb{R}^{n\times n}$,

$$\max_{d\in[0,1]^n} \alpha([d]W) \geq \alpha(W)_+, \quad \max_{d\in[0,1]^n} \alpha(W[d]) \geq \alpha(W)_+. \quad (9)$$

We now give the proof of Theorem 3. To enhance clarity we prove its parts one by one. Lemma 10 and parts (i) and (ii) in Theorem 3, are based upon and extend the treatment in [10, Theorem 2] – see our statement of contributions.

Lemma 10 (Splitting upper-bounded symmetric matrices)**.** *Consider W satisfying Assumptions 1. Assume $W \preceq bI_n$, for some $b > 0$ and let $\theta_b(\cdot)$ and $Q_{\mathrm{F},b}$ be defined in (6) and (7), respectively. Then,*

$$W = Q_{\mathrm{F},b} - \frac{1}{4b} Q_{\mathrm{F},b}^2. \quad (10)$$

Proof. By definition of the function $\theta_b(\cdot)$, for all $\lambda_i \leq b$, $i \in \{1,\dots,n\}$, it holds that

$$\lambda_i = \theta_b(\lambda_i) - \frac{1}{4b}\theta_b(\lambda_i)^2. \quad (11)$$

Equation (11) implies $\Lambda = \theta_b(\Lambda) - \frac{1}{4b}\theta_b(\Lambda)^2$. Equality (10) then follows by multiplying by U and $U^\top$ to the left and to the right, respectively, with U defined as in (5). □

First, we prove part (i).

Proof of part (i). First, we prove that $\|\cdot\|_{2,Q_{\mathrm{F},\alpha(W)}}$ is log-optimal for $\mathcal{P}_{\mathrm{F}}$ and $\max_{d\in[0,1]^n} \alpha([d]W) = \alpha(W)$. To this purpose, define $P := \frac{1}{4\alpha(W)} Q_{\mathrm{F},\alpha(W)}^2 \succ 0$. Lemma 10 implies $W = Q_{\mathrm{F},\alpha(W)} - P$. Next, pick $d \in \mathbb{R}^n$ satisfying $\mathbb{0}_n < d \leq \mathbb{1}_n$, so that $[d]$ is diagonal and invertible. Then

$$2\alpha(W)P - \frac{1}{2}Q_{\mathrm{F},\alpha(W)}^2 \succeq 0 \quad (12)$$

$$\implies 2\alpha(W)P - \frac{1}{2}Q_{\mathrm{F},\alpha(W)}[d]Q_{\mathrm{F},\alpha(W)} \succeq 0$$

$$\iff 2\alpha(W)P - Q_{\mathrm{F},\alpha(W)}[d]P(2P[d]P)^{-1}P[d]Q_{\mathrm{F},\alpha(W)} \succeq 0.$$

Since $P[d]P \succ 0$, we can apply the Schur complement to this LMI to conclude that

$$y^\top \begin{bmatrix} 2\alpha(W)P & -Q_{\mathrm{F},\alpha(W)}[d]P \\ -P[d]Q_{\mathrm{F},\alpha(W)} & 2P[d]P \end{bmatrix} y \geq 0, \ \forall y \in \mathbb{R}^{2n}. \quad (13)$$

Setting $y = (y_1, y_1)$ for arbitrary $y_1 \in \mathbb{R}^n$, (13) implies

$$2\alpha(W)P - Q_{\mathrm{F},\alpha(W)}[d]P - P[d]Q_{\mathrm{F},\alpha(W)} + 2P[d]P \succeq 0$$

$$\overset{W=Q_{\mathrm{F},\alpha(W)}-P}{\iff} W[d]P + P[d]W \preceq 2\alpha(W)P$$

$$\iff Q_{\mathrm{F},\alpha(W)}^2[d]W + W[d]Q_{\mathrm{F},\alpha(W)}^2 \preceq 2\alpha(W)Q_{\mathrm{F},\alpha(W)}^2. \quad (14)$$

Hence, we showed that the weak LMI (12) (independent of d) implies the weak LMI (14) for all $0 < d \leq \mathbb{1}_n$. By weak LMI, we mean that the inequality is not strict. It is known [8, Theorem 6.3.5] that the eigenvalues of a symmetric matrix are continuous functions of the matrix entries. Therefore, the LMI (14) holds also for $\mathbb{0}_n \leq d \leq \mathbb{1}_n$. Finally, note that the LMI (14) is equivalent to $\mu_{2,Q_{\mathrm{F},\alpha(W)}}([d]W) \leq \alpha(W)$ for all $d \in [0,1]^n$. The proof follows from (9).

Next, assume that W is invertible. We need to prove that $\|\cdot\|_{2,Q_{\mathrm{H},\alpha(W)}}$ is log-optimal for $\mathcal{P}_{\mathrm{H}}$ and that it holds $\max_{d\in[0,1]^n} \alpha(W[d]) = \alpha(W)$. We have $\max_{d\in[0,1]^n} \mu_{2,Q_{\mathrm{H},\alpha(W)}}(W[d]) = \max_{d\in[0,1]^n} \mu_{2,Q_{\mathrm{F},\alpha(W)}W^{-1}}(W[d]) \overset{(1)}{=} \max_{d\in[0,1]^n} \mu_{2,Q_{\mathrm{F},\alpha(W)}}([d]W) = \alpha(W)$, where the last equality follows from the log-optimality of $\|\cdot\|_{2,Q_{\mathrm{F},\alpha(W)}}$ for $\mathcal{P}_{\mathrm{F}}$. The proof again follows from (9). □

The proof of part (ii) of Theorem 3, follows the same reasoning as that of part (i) by considering $\varepsilon > 0$ instead of $\alpha(W)$. Hence, we omit it here for brevity.

Finally, we prove part (iii). To do so, we give the following algebraic result.

Lemma 11 (Optimal norms for products of symmetric matrices)**.** *Let $A_1 = SQ \in \mathbb{R}^{n\times n}$ and $A_2 = QS \in \mathbb{R}^{n\times n}$ where S, $Q \in \mathbb{S}^n$, with $Q \succ 0$. Then, for each $i \in \{1,2\}$,* *(i)* $\mathrm{spec}(A_i)$ *is real and has the same number of negative, zero, and positive eigenvalues as* S*;*

(ii) the norm $\|\cdot\|_{2,Q^{1/2}}$ *is optimal for the matrix* A_i*, i.e.,* $\|A_i\|_{2,Q^{1/2}} = \rho(A_i)$*;*

(iii) the norm $\|\cdot\|_{2,Q^{1/2}}$ *is log-optimal for* A_i*, i.e.,* $\mu_{2,Q^{1/2}}(A_i) = \alpha(A_i)$*.*

Proof. Let $i = 1$. A_1 is similar to $Q^{1/2}SQ^{1/2} \in \mathbb{S}^n$, hence $\mathrm{spec}(A_1)$ is real. Part (i) then follows from Sylvester's law of inertia. Regarding part (ii), we compute

$$\|A_1\|^2_{2,Q^{1/2}} = \lambda_{\max}(Q^{-1}A_1^\top QA_1) = \lambda_{\max}((SQ)^2) = \rho(SQ)^2$$

where the last equality follows from the fact that $(SQ)^2$ has the same eigenvectors as SQ and real eigenvalues equal to the square of the real eigenvalues of SQ. Finally, to prove part (iii) we compute $\mu_{2,Q^{1/2}}(A_1) = \lambda_{\max}\Big(\frac{QA_1Q^{-1}+A_1^\top}{2}\Big) = \lambda_{\max}(QS) = \lambda_{\max}(QA_1Q^{-1}) = \lambda_{\max}(A_1) = \alpha(A_1)$. The proof for $i = 2$ follows the same reasoning. ☐

Proof of part (iii). Pick $d \in \mathbb{R}^n$ satisfying $\mathbb{0}_n \leq d \leq \mathbb{1}_n$ and consider the matrices $[d]W$ and $W[d]$. Lemma 11 with $S := [-d]$ and $Q := -W \succ 0$, implies that the spectrum of the product matrices $[d]W = ([-d])(-W)$ and $W[d] = (-W)([-d])$ is real and has the same number of negative, zero, positive eigenvalues as $[-d]$. Therefore,

$$\mu_{2,(-W)^{1/2}}([d]W) = \alpha([d]W) \begin{cases} < 0 & \text{if } d > \mathbb{0}_n, \\ \leq 0 & \text{otherwise}, \end{cases} \quad (15)$$

$$\mu_{2,(-W)^{1/2}}(W[d]) = \alpha(W[d]) \begin{cases} < 0 & \text{if } d > \mathbb{0}_n, \\ \leq 0 & \text{otherwise}. \end{cases} \quad (16)$$

Maximizing over $d \in [0,1]^n$ we get part (iii). ☐

Proof of Theorem 9. Let r be the number of non-zero eigenvalues of W. Without loss of generality, we reorder the elements in $\lambda \in \mathbb{R}^n$ and $U \in \mathbb{R}^{n\times n}$, so that $\lambda = (\lambda_1,\dots,\lambda_r,0,\dots,0)$ and $U = [u_1,\dots,u_r,u_{r+1},\dots,u_n]$, where $u_i \in \mathbb{R}^n$ is the eigenvector of $\lambda_i \in \mathbb{R}$. Next, let $\mathcal{K}^* := \mathrm{span}\{u_1,\dots,u_r\}$, $n_\parallel := \dim(\mathcal{K}^*)$, $\mathcal{K} := \mathrm{span}\{u_{r+1},\dots,u_n\}$, $n_\perp := \dim(\mathcal{K})$, and define $U_\parallel := [u_1,\dots,u_r] \in \mathbb{R}^{n\times n_\parallel}$, $U_\perp := [u_{r+1},\dots,u_n] \in \mathbb{R}^{n\times n_\perp}$, so that $U = [U_\parallel \;\; U_\perp]$. We have $\mathbb{R}^n = \{x \in \mathbb{R}^n \mid x \in \mathcal{K}^*\} \oplus \{x \in \mathbb{R}^n \mid x \in \mathcal{K}\}$. Therefore, given $x \in \mathbb{R}^n$ we can always define $x_\parallel = U_\parallel^\top x \in \mathcal{K}^*$ and $x_\perp = U_\perp^\top x \in \mathcal{K}$. We note that $U^\top U = I_n$ implies $U_\parallel^\top U_\parallel = I_{n_\parallel}$, $U_\perp^\top U_\perp = I_{n_\perp}$, $U_\perp^\top U_\parallel = \mathbb{0}_{n_\perp\times n_\parallel}$ and $U_\parallel^\top U_\perp = \mathbb{0}_{n_\parallel\times n_\perp}$. Also, $W = U_\parallel\Lambda_\parallel U_\parallel^\top$ and $Q_{\mathrm{F},\alpha(W)} := U\theta_{\alpha(W)}(\Lambda)U^\top = U_\parallel\theta_\parallel U_\parallel^\top + U_\perp\theta_\perp U_\perp^\top$, thus $Q_{\mathrm{F}\parallel} := U_\parallel^\top Q_{\mathrm{F},\alpha(W)}U_\parallel = \theta_\parallel$. From Corollary 6 we have

$$\max_{d\in[0,1]^n} \mu_{2,\theta_\parallel}(-I_{n_\parallel} + U_\parallel^\top[d]U_\parallel\Lambda_\parallel) \leq -1+\alpha(W). \quad (17)$$

Moreover, we get:

$$\begin{cases} \dot{x}_{\mathrm{H}}^\perp = -x_{\mathrm{H}}^\perp + u_{\mathrm{H}}^\perp := f_{\mathrm{H}}^\perp(x_{\mathrm{H}}^\perp, u_{\mathrm{H}}^\perp), & (18) \\ \dot{x}_{\mathrm{H}}^\parallel = -x_{\mathrm{H}}^\parallel + \Lambda_\parallel U_\parallel^\top\Phi(x_{\mathrm{H}}) + u_{\mathrm{H}}^\parallel := f_{\mathrm{H}}^\parallel(x_{\mathrm{H}}, u_{\mathrm{H}}^\parallel), & (19) \end{cases}$$

where (18) was obtained by multiplying (4) by $U_\perp^\top$, while (19) was obtained by multiplying (4) by $U_\parallel^\top$. Equation (18) is contracting with respect to any norm in the subspace $\mathcal{K}$ with $\mathsf{osL}(f_{\mathrm{H}}^\perp) = -1$, being $\mu(Df_{\mathrm{H}}^\perp) = -1$. For (19) we first note that $Q_{\mathrm{H}\parallel} := U_\parallel^\top Q_{\mathrm{F},\alpha(W)}W^\dagger U_\parallel = \theta_\parallel\Lambda_\parallel^{-1}$ and $Df_{\mathrm{H}}^\parallel = -I_{n_\parallel} + \Lambda_\parallel U_\parallel^\top[d]U_\parallel$. Thus, we have $\mathsf{osL}_{2,Q_{\mathrm{H}\parallel}}(f_{\mathrm{H}}^\parallel) \leq \max_{d\in[0,1]^n} \mu_{2,\theta_\parallel\Lambda_\parallel^{-1}}(-I_{n_\parallel} + \Lambda_\parallel U_\parallel^\top[d]U_\parallel) \overset{(17)}{=} \max_{d\in[0,1]^n} \mu_{2,\theta_\parallel}(-I_{n_\parallel} + U_\parallel^\top[d]U_\parallel\Lambda_\parallel) \leq -1+\alpha(W)$. Hence (19) is strongly infinitesimally contracting in $\mathcal{K}^*$ with respect to $\|\cdot\|_{Q_{\mathrm{H}\parallel}}$ with rate $1-\alpha(W)$.

Finally, we note that at fixed $x_\parallel$ and t, the map $x_\perp \to f_\parallel$ is Lipschitz with constant $\mathsf{L}_{\parallel\perp} := \alpha(W)$. We can now construct the gain matrix (see [3, Theorem 3.23])

$$\Gamma = \begin{bmatrix} -1 & 0 \\ \alpha(W) & -1+\alpha(W) \end{bmatrix} \in \mathbb{R}^{2\times 2}, \quad (20)$$

whose eigenvalues are $\lambda_1 = -1, \lambda_2 = -1+\alpha(W)$. The fact that $\mathcal{K} \neq \{\mathbb{0}_n\}$ implies $\alpha(W) \geq 0$. In turn, since by assumptions $\alpha(W) < 1$, we have $\lambda_2 \in [-1, 0[$. Thus, Γ is Hurwitz and $\alpha(\Gamma) = -1+\alpha(W)$. Now, from [3, Theorem 3.23] it follows that, for each $\varepsilon \in]0, 1-\alpha(W)[$, the HNN is strongly infinitesimally contracting with rate $|\alpha(\Gamma)+\varepsilon|$. This concludes the proof. ☐

V. Using Euclidean Contractivity to Solve Quadratic Optimization Problems

We now apply the results to propose a FNN solving certain quadratic optimization problems with box constraints, inspired by the widely cited work [2].

Given $A = A^\top \succ 0$, an input $u \in \mathbb{R}^n$, and $\mu \leq \nu \in \mathbb{R}^n$ the *quadratic optimization problem with box constraints* is

$$\min_{y\in\mathbb{R}^n} \Big(J_{A,u}(y) := \frac{1}{2}y^\top Ay - u^\top y\Big), \quad \text{s.t. } \mu \leq y \leq \nu. \quad (21)$$

Note that $J_{A,u}(\cdot)$ is strongly convex and the constraints are convex, thus (21) admits a unique global optimal solution.

We propose the following FNN model to solve (21). Given a single-layered neural network of n neurons, the state $x \in \mathbb{R}^n$ evolves according to

$$\dot{x} = -x + \mathrm{sat}_{\mu,\nu}((I_n - A)x + u), \quad (22)$$

with output $y = x$. The activation function $\mathrm{sat}_{\mu,\nu}(\cdot)\colon \mathbb{R}^n \to [\mu,\nu] := [\mu_1,\nu_1]\times\cdots\times[\mu_n,\nu_n]$ is defined as $(\mathrm{sat}_{\mu,\nu}(x))_i = \mathrm{sat}_{\mu_i,\nu_i}(x_i)$, where $\mathrm{sat}_{\mu_i,\nu_i}(\cdot)\colon \mathbb{R} \to [\mu_i,\nu_i]$ is defined by $\mathrm{sat}_{\mu_i,\nu_i}(x_i) = \min\{\max\{\mu_i, x_i\}, \nu_i\}$. Note that $\mathrm{sat}_{\mu_i,\nu_i}(\cdot)$ satisfies Assumption 2. To simplify the notation we use the same symbol for both the scalar and vector forms. Next, we use Corollary 6 to give sufficient conditions for the strong infinitesimal contractivity of (22). Then, we show that the equilibrium of (22) is the optimal solution of (21).

Lemma 12 (Strong infinitesimal contractivity)**.** *Let* $A = A^\top \succ 0$ *in (22). The FNN (22) is strongly infinitesimally contracting with rate* $c > 0$ *with respect to* $\|\cdot\|_{2,P}$*, where*

(i) if $\lambda_{\min}(A) < 1$*, then* $c = \lambda_{\min}(A)$ *and* $P = Q_{\mathrm{F},1-\lambda_{\min}(A)}$*, with* $Q_{\mathrm{F},1-\lambda_{\min}(A)}$ *defined in (7);*

(ii) if $\lambda_{\min}(A) = 1$*, then for any* $0 < \varepsilon < 1$*,* $c = 1-\varepsilon > 0$ *and* $P = Q_{\mathrm{F},\varepsilon}$*, with* $Q_{\mathrm{F},\varepsilon}$ *defined in (7);*

(iii) if $\lambda_{\min}(A) > 1$, then $c = 1$ and $P = (A - I_n)^{1/2}$.

Proof. The result follows by applying Corollary 6 noticing that $A \succ 0$ implies $W = I_n - A \prec I_n$, thus $\alpha(W) = 1 - \lambda_{\min}(A) < 1$, and $\mathrm{sat}_{\mu,\nu}(\cdot)$ satisfies Assumption 2. □

An immediate consequence of Lemma 12 is that (22) admits a unique equilibrium point. Next, we prove that this equilibrium point is the optimal solution of (21).

Lemma 13. *The vector $x^* \in \mathbb{R}^n$ is the global minimum for (21) if and only if x^* is the equilibrium point of (22).*

Proof. Let $x^* \in \mathbb{R}^n$ be a global minimum for (21), thus $x^* \in [\mu, \nu]$. Then it follows from the KKT conditions that, for all $i \in \{1, \dots, n\}$,

$$\frac{\partial J_{A,u}}{\partial x_i}(x^*) = (Ax^*)_i - u_i \begin{cases} \geq 0 & \text{if } x_i^* = \mu_i, \\ = 0 & \text{if } \mu_i < x_i^* < \nu_i, \\ \leq 0 & \text{if } x_i^* = \nu_i. \end{cases} \tag{23}$$

Note that x^* is an equilibrium of (22) if, for all i, we have

$$-x_i^* + \mathrm{sat}_{\mu_i,\nu_i}(x_i^* - (Ax^*)_i + u_i) = 0. \tag{24}$$

If $x_i^* = \mu_i$, let $z^\star := (Ax^*)_i|_{x_i^* = \mu_i} - u_i$. By definition of $\mathrm{sat}_{\mu_i,\nu_i}(\cdot)$ it holds $-\mu_i + \mathrm{sat}_{\mu_i,\nu_i}(\mu_i - z^\star) \geq 0$. Moreover, from the KKT conditions (23), and being $\mathrm{sat}_{\mu_i,\nu_i}(\cdot)$ monotonically non-decreasing, we get the reverse inequality. Thus $x_i^* = \mu_i$ verifies (24). Similarly it can be proved that (24) holds for $\mu_i < x_i^* < \nu_i$, and $x_i^* = \nu_i$.

Vice versa, let $x^* \in \mathbb{R}^n$ be an equilibrium of (21), i.e., (24) holds. If $x_i^* \leq \mu_i$, then (24) implies $x_i^* = \mathrm{sat}_{\mu_i,\nu_i}(\mu_i - z^\star)$. By definition of $\mathrm{sat}_{\mu_i,\nu_i}(\cdot)$ we get that $x_i^* \in [\mu_i, \nu_i]$, thus $x_i^* = \mu_i$, and $\mu_i - z^\star \leq \mu_i$, which implies $z^\star \geq 0$. Similarly, if $\mu_i < x_i^* < \nu_i$, then $z^\star = 0$, while if $x_i^* \geq \nu_i$, then $x_i^* = \nu_i$ and $z^\star \leq 0$. This ends the proof since we have shown that the KKT conditions (23) hold for all i. □

VI. Conclusion

We presented sharp conditions for strong and weak Euclidean contractivity of Hopfield and firing-rate neural networks with symmetric weights together with a number of general algebraic results. Specifically, we analyzed the Euclidean log-norm of matrix polytopes, proposing norms that are log-optimal for almost all matrices, and provided optimal and log-optimal norms for the product of symmetric matrices. We considered networks with (possibly) non-smooth activation functions, which allows us to consider common activation functions such as ReLU and the soft thresholding function. Finally, to demonstrate the practical implications of our results, we proposed a FNN to solve quadratic optimization problems with box constraints.

As future work, it would be useful to (i) extend our results to arbitrary synaptic matrices (as opposed to only symmetric) and heterogeneous dissipation matrices, (ii) establish higher-order contractivity properties [18] and consider stochastic models [1], and (iii) apply these results to neuroscience and machine learning problems. For example, we plan to study sparse reconstruction networks (inspired by [15]) and implicit learning models (e.g., see [14]).

References

[1] Z. Aminzare. Stochastic logarithmic Lipschitz constants: A tool to analyze contractivity of stochastic differential equations. *IEEE Control Systems Letters*, 6:2311–2316, 2022. doi:10.1109/LCSYS.2022.3148945.

[2] A. Bouzerdoum and T. R. Pattison. Neural network for quadratic optimization with bound constraints. *IEEE Transactions on Neural Networks*, 4(2):293–304, 1993. doi:10.1109/72.207617.

[3] F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.1 edition, 2023, ISBN 979-8836646806. URL: http://motion.me.ucsb.edu/book-ctds.

[4] V. Centorrino, F. Bullo, and G. Russo. Modelling and contractivity of neural-synaptic networks with Hebbian learning. 2022. Submitted. arXiv:2204.05382, doi:10.48550/arXiv.2204.05382.

[5] A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contraction analysis of continuous-time neural networks. 2022. Submitted. doi:10.48550/arXiv.2110.08298.

[6] 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.

[7] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski. *Neuronal Dynamics: From Single Neurons To Networks and Models of Cognition*. Cambridge University Press, 2014, ISBN 9781107635197. URL: https://neuronaldynamics.epfl.ch.

[8] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2nd edition, 2012, ISBN 0521548233.

[9] 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. doi:10.48550/arXiv.2106.03194.

[10] L. Kozachkov, M. Ennis, and J.-J. E. Slotine. RNNs of RNNs: Recursive construction of stable assemblies of recurrent neural networks. In *Advances in Neural Information Processing Systems*, December 2022. doi:10.48550/arXiv.2106.08928.

[11] L. Kozachkov, M. Lundqvist, J.-J. E. Slotine, and E. K. Miller. Achieving stable dynamics in neural circuits. *PLoS Computational Biology*, 16(8):1–15, 2020. doi:10.1371/journal.pcbi.1007659.

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

[13] 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.

[14] M. Revay and I. Manchester. Contracting implicit recurrent neural networks: Stable models with improved trainability. In *Conference on Learning for Dynamics and Control*, volume 120, pages 393–403, 2020. URL: https://proceedings.mlr.press/v120/revay20a.html.

[15] C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural Computation*, 20(10):2526–2563, 2008. doi:10.1162/neco.2008.03-07-486.

[16] 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.

[17] D. W. Tank and J. J. Hopfield. Simple "neural" optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit. *IEEE Transactions on Circuits and Systems*, 33(5):533–541, 1986. doi:10.1109/TCS.1986.1085953.

[18] C. Wu, I. Kanevskiy, and M. Margaliot. k-contraction: Theory and applications. *Automatica*, 136:110048, 2022. doi:10.1016/j.automatica.2021.110048.

[19] S. Xie, G. Russo, and R. H. Middleton. Scalability in nonlinear network systems affected by delays and disturbances. *IEEE Transactions on Control of Network Systems*, 8(3):1128–1138, 2021. doi:10.1109/TCNS.2021.3058934.

[20] A. Yang, J. Xiong, M. Raginsky, and E. Rosenbaum. Input-to-state stable neural ordinary differential equations with applications to transient modeling of circuits. In *Conference on Learning for Dynamics and Control*, volume 168, pages 663–675, 2022. URL: https://proceedings.mlr.press/v168/yang22b.html.