LURIE NETWORKS WITH k-CONTRACTING DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper proposes an approach to enable the weights and biases of a novel neural ODE, the Lurie network, to be trained in such a manner that a generalised concept of stability is guaranteed. This generalised stability measure is derived through the use of k-contraction analysis, which guarantees global convergence to a point, line or plane in the neural state-space. An unconstrained parametrisation of this condition is derived, allowing models to be trained using standard optimisation algorithms, whilst limiting the search space to solutions satisfying the k-contraction constraint. The novel stability result and parametrisation provide a toolset for training over the space of Lurie network's which exhibit the convergent behaviours observed during neural computation in the brain. For example, global convergence to one of multiple equilibrium points or limit cycles are properties observed in associative and working memory.

1 INTRODUCTION

The brain organises its representations of the world and carries out complex functions through collective interactions of simpler modules Kandel et al. (2000). Convergent dynamics in widespread regions of the central nervous system are thought to play a crucial role in: forming some of these representations Khona & Fiete (2022), processing information over extended periods Vogt et al. (2022), learning Kozachkov et al. (2020); Centorrino et al. (2022), memory storage Hopfield (1982; 1984); Kozachkov et al. (2022); Pals et al. (2024) and enhancing the robustness of each of these functions Khona & Fiete (2022). As summarised in Khona & Fiete (2022), convergent dynamics in the brain take several forms. For example, neural circuits with multiple equilibrium points (bistable and multi-stable) have been observed in the anterlor lateral motor cortex of a rat Piet et al. (2017) and are conjectured to appear in the mammalian hippocampus and auditory cortex. Some theories of associative memory also believe reconstruction of a learned pattern is obtained by flow to equilibrium points Krotov & Hopfield (2020); Sharma et al. (2022); Kozachkov et al. (2023). Limit cycles are another form of convergent dynamics for which there are numerous examples within the central nervous system. These include working memory Kozachkov et al. (2022), which is thought to arise from the sustained spiking of neurons Ashwin et al. (2016), and sleep cycle generation Adamantidis et al. (2019).

The convergence and stability analysis of dynamical systems has been well-studied in the control theoretic literature. A Lurie[1] network is a popular class of nonlinear dynamical systems comprising a linear time-invariant (LTI) component interconnected with a, potentially time-varying, nonlinearity. Such systems are ubiquitous throughout the sciences Ofir et al. (2023), engineering Ofir et al. (2024), machine learning (ML) Pauli et al. (2021); Lessard et al. (2016) and neuroscience, where recurrent neural networks (RNNs) are commonly used as computational models Wilson & Cowan (1972). A well-studied problem amongst the control community is the absolute stability problem: where the nonlinearity of the Lurie network is unknown, but assumed to be sector-bounded or slope-restricted. The goal is to find conditions on the model parameters which ensure the trajectories of all Lurie networks, with nonlinearities in the assumed class, uphold a chosen definition of convergence. Approaches to this problem can be classified as Lyapunov analysis Park (1997; 2002); Khalil (2002); Richardson et al. (2023; 2024), Zames-Falb multipliers Zames & Falb (1968); Carrasco et al. (2016); Turner & Drummond (2019); Drummond et al. (2024) or k-contraction analysis Zhang & Cui (2013); Ofir et al. (2023; 2024). Lyapunov analysis and Zames-Falb multipliers are primarily used to analyse the stability of equilibrium points. Indeed, Lyapunov analysis is regularly applied in the study of Hopfield networks Hopfield (1982); Krotov & Hopfield (2020); Ramsauer et al. (2020).

[1] Named after Anatolii Isakovich Lurie and sometimes spelt Lur'e or Lurye.

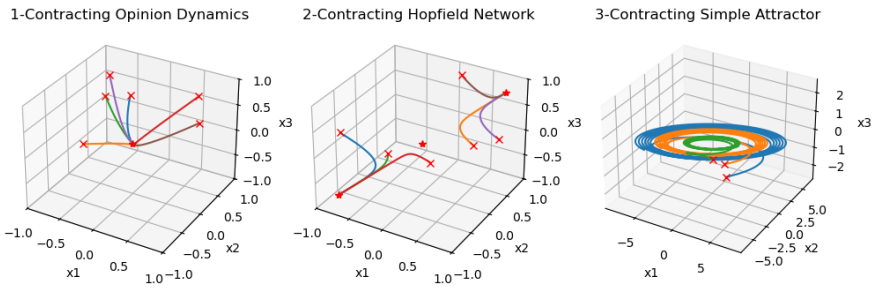

Figure 1: Trajectories from three dynamical systems satisfying the k-contraction property. Crosses denote the initial condition and stars denote equilibirum points.

On the other hand, k-contraction analysis provides a unifying framework to study a variety of global convergence behaviours, including convergence to (unique equilibrium) points, lines (e.g., multiple equilibrium points) and planes (e.g., limit cycles). As a result, the same analysis could be applied to the study of associative and working memory, along with a myriad of other neural functions.

This paper leverages the Lurie network modelling framework and k-contraction analysis tools to simulatneously develop a toolset for: (i) studying the various convergent dynamics observed in the brain and; (ii) developing a machine learning framework inspired by these theoretical properties, such as the ones related to associative memory. The technical steps include: (i) using k-contraction analysis to derive mild constraints on the weights and biases of the Lurie network such that the absolute stability problem is addressed; (ii) establishing unconstrained parametrisations of these conditions which allows the Lurie network to be trained using gradient based optimisation algorithms whilst limiting the search space to weights and biases which satisfy the k-contraction condition.

2 k-CONTRACTION ANALYSIS

k-contraction analysis Wu et al. (2022); Muldowney (1990) is the geometrical generalisation of contraction analysis Lohmiller & Slotine (1998). Intuitively, k-contraction implies the volume of k-dimensional bodies exponentially converges to zero when governed by the system dynamics. Alternatively, this could be thought of as exponential convergence to a $(k-1)$-dimensional subspace. When $k=1$, this reduces to standard contraction Lohmiller & Slotine (1998), which implies that all trajectories exponentially converge to a single trajectory. For a general time-varying dynamical system, satisfying the k-contraction property does not guarantee stability. However, for time-invariant dynamical systems, it has been shown that for every bounded solution: 1-contraction implies global convergence to a unique equilibrium point Lohmiller & Slotine (1998), 2-contraction implies global convergence to an equilibrium point, which is not necessarily unique but must be connected along a line Muldowney (1990), and 3-contraction, under certain assumptions, implies convergence to a non-unique attractor in a 2d subspace Cecilia et al. (2023). *Robustness* is an inherent property of these networks as the trajectories can only converge to a finite number of long term behaviours. This is appealing for both artificial and biological networks. Three examples of k-contracting dynamics are illustrated in Figure 1; these examples are from a variety of domains to highlight the widespread application of the analysis. Next, we present the fundamental Euclidean k-contraction result from Wu et al. (2022). Refer to §A for details on the notation and §B for a summary of compound matrices.

Theorem 1 *Fix $k \in [1, n]$ and consider the nonlinear system $\dot{x} = f(t, x)$ with $f : \Re_+ \times \Re^n \to \Re^n$ continuously differentiable. If there exists $\eta > 0$ and an invertible matrix $\Theta \in \Re^{n \times n}$ such that*

$$\mu_{2,\Theta^{(k)}}\left(J_f^{[k]}(t,x)\right) \leq -\eta \quad \forall\, x \in \Re^n \text{ and } t \in \Re_+ \tag{1}$$

then the nonlinear system is k-contracting in the 2-norm w.r.t the metric $P := \Theta^\top \Theta$.

This result has two features: (i) it requires the existence of an invertible matrix Θ. In the simplest case, one can expect a solution $\Theta = qI_n$ to exist. For other systems, such simple solutions will not exist and more general matrices such as $\Theta \in \mathcal{S}^n$ will be required, making the proofs more

difficult; (ii) it requires the use of compound matrices. For a matrix $W \in \Re^{n\times m}$, the matrix $W^{[k]}$ with $k \in [1, \min(n, m)]$ will have the size $\binom{n}{k} \times \binom{m}{k}$ which is typically much larger and more computationally difficult to work with. A more technical introduction to k-contraction analysis and compound matrices is presented in Bar-Shalom et al. (2023) with the essential results detailed in §B. In §3.1 we derive results which verify (1) for the special case of the Lurie network (2).

3 LURIE NETWORK

A *Lurie network* is defined by (2) with weights $A \in \Re^{n\times n}$, $B \in \Re^{n\times m}$, $C \in \Re^{m\times n}$ and biases $b_x \in \Re^n$, $b_y \in \Re^m$.

$$\dot{x}(t) = Ax(t) + B\Phi\big(y(t)\big) + b_x \qquad y(t) = Cx(t) + b_y \qquad x(0) = x_0 \tag{2}$$

The model has a biased linear component interconnected with a nonlinearity of the form $\Phi(y) := [\phi_1(y_1) \dots \phi_m(y_m)]'$ where $\phi_i(y_i)$ is assumed to be slope-restricted with an upper bound $g > 0$, such that $0 \preceq J_\Phi(y) \preceq gI_m$. This separation of the linear and nonlinear components is useful for analysis. Activation functions which satisfy this slope-restricted assumption include the hyperbolic tangent ($\tanh$) and the rectified linear unit ($ReLU$). For simplicity, we assume the same scalar nonlinearity is applied element-wise and drop the subscript.

Although the brain is subject to various external inputs, it is common to assume that, at least on the timescale of interest, the dynamics evolve in a time-invariant manner Khona & Fiete (2022). As a result, time-invariance of the Lurie network (2) does not prevent it from being a suitable model for neural dynamics. Furthermore, satisfying Theorem 1 ensures the model will inherit the appealing convergence and robustness properties stated in §2.

3.1 k-CONTRACTION ANALYSIS OF LURIE NETWORKS

A sufficient result which satisfies Theorem 1 and guarantees (2) is k-contracting is presented next. Conditions were derived in (Ofir et al., 2024, Theorem 2) which verify Theorem 1 for a Lurie network with $A \in \mathcal{D}^n$ and $b_y = 0$. Theorem 2 extends them to account for $A \in \Re^{n\times n}$ and $b_y \neq 0$. Refer to §C for the proof.

Theorem 2 *Consider the Lurie network (2) with* $\Phi(y) := [\phi_1(y_1) \quad \dots \quad \phi_m(y_m)]'$ *being slope-restricted such that* $0 \preceq J_\Phi(y) \preceq gI_m$. *Fix* $k \in [1, n]$ *and define* $\alpha_k := (2k)^{-1} \sum_{i=1}^{k} \lambda_i(A + A^\top)$. *If* $\alpha_k < 0$ *and*

$$g^2 \sum_{i=1}^{k} \sigma_i^2(B)\sigma_i^2(C) < \alpha_k^2 k \tag{3}$$

then (2) is k-contracting in the 2-norm w.r.t the metric $P = -\alpha_k^{-1} I_n$.

The additional freedom permitted by k-contraction over standard contraction is highlighted by the summation of the eigenvalues and singular values. In 1-contraction, Theorem 2 requires the largest eigenvalue of the symmetric component of A to be negative whereas for $k \in [2, n]$, this condition on A becomes incrementally more relaxed as k is increased. Equation (3) illustrates a similar relaxation of the constraints on B and C. Theorem 2 has several appealing features: (i) it does not require the computation of the troublesome compound matrices; (ii) it provides a way of embedding the k-contraction property into the structure of a Lurie network based on fairly simple unconstrained parametrisations of the weights, as shown in §3.2; (iii) the biases are not present in the condition, so are naturally unconstrained; (iv) the result does not rely on symmetries of the parameters. In many Hopfield-based models of associative memory, symmetry in the parameters is needed to make the existence of a global energy function mathematically tractable; however, this simplification is biologically unrealistic. Furthermore, the symmetry requirement limits expressivity of the model which is a limitation for both artificial and biological systems. The drawbacks of the result are that only Lurie networks which are k-contracting in a scalar metric can be verified and, in the case of 2-contraction, equilibrium points are constrained to a line. Finally, it is important to highlight that Theorem 2 applies to the class of slope-restricted nonlinearities, so this result addresses the absolute stability problem for the k-contraction property.

3.2 Parametrisation of k-contracting Lurie networks

To train a k-contracting Lurie network using gradient based optimisers, parametrisations which express the constrained weights in terms of unconstrained variables must be found. To formalise this idea, we define the set $\Omega_2(g,k)$ which contains all combinations of the Lurie network's weights which satisfy Theorem 2. As the biases do not appear in this set, they are naturally unconstrained.

$$\Omega_2(g,k) := \left\{ (\bar{A},\bar{B},\bar{C}) \;\middle|\; \alpha_k < 0,\; z := g^2 \sum_{i=1}^{k} \sigma_i^2(\bar{B})\sigma_i^2(\bar{C}) < \alpha_k^2 k \right\} \tag{4a}$$

The next results presents a parametrisation of the set $\Omega_2(g,k)$. See §D for the proof which leverages the eigenvalue and singular value decompositions.

Theorem 3 *Given $g > 0$, $k \in [1,n]$, $U_A, U_B, V_C \in \mathcal{SO}(n)$, $V_B, U_C \in \mathcal{SO}(m)$, $\Sigma_B \in \mathcal{D}_+^{nm}$, $\Sigma_C \in \mathcal{D}_+^{mn}$, $Y_A \in so(n)$, $\Sigma_{A1} \in \mathcal{D}^{k-1}$, $G_{A2} > 0$, $G_{A3} \in \mathcal{D}_+^{n-k}$ and define*

$$A := \frac{1}{2}U_A\Sigma_A U_A^\top + \frac{1}{2}Y_A \qquad B := U_B\Sigma_B V_B^\top \qquad C := U_C\Sigma_C V_C^\top \tag{5a}$$

$$\Sigma_A := \mathrm{blockdiag}(\Sigma_{A1}, \Sigma_{A2}, \Sigma_{A3}) \qquad \Sigma_{A1} \in \mathcal{D}^{k-1} \tag{5b}$$

$$\Sigma_{A2} := -\sqrt{4kz} - \sum_{i}^{k-1}(\Sigma_{A1})_{ii} - G_{A2} \qquad \Sigma_{A3} := \min(\Sigma_{A1}, \Sigma_{A2})I_{n-k} - G_{A3} \tag{5c}$$

then $(A,B,C) \in \Omega_2(g,k)$.

Remark 1 *The mapping between $so(\cdot)$ and $\mathcal{SO}(\cdot)$ was exploited to express the orthogonal variables as unconstrained skew-symmetric variables Lezcano-Casado & Martınez-Rubio (2019). The remaining variables are also unconstrained or simply require positive elements, which can be obtained by taking the absolute value of unconstrained elements.*

In Theorem 3, the B and C matrices are unconstrained since they are simply expressed by their singular value decomposition. The variables representing the singular values of B and C can then directly upper bound α_k. The only source of conservatism in the parametrisation is introduced through the definition of Σ_A. The definition of Σ_A is split into one unconstrained block for the largest $(k-1)$ eigenvalues, Σ_{A1}, a block to ensure Theorem 2 is satisfied, Σ_{A2}, and another block, Σ_{A3}, for ensuring the remaining eigenvalues are less than the other k. As k is a hyperparameter of the model, it is useful to note that $\Omega_2(g,1) \subseteq \Omega_2(g,2) \subseteq \Omega_2(g,3) \cdots \subseteq \Omega_2(g,n)$. Hence, if the best value of k is unknown for a given application, then setting $k = 3$ allows the model to search over all the stable k-contracting systems parametrised by Theorem 3.

Finally, a Lurie network (2) has a total parameter count $N_L = n^2 + 2nm + n + m$; whereas a k-contracting Lurie network, constructed according to Theorem 3, has $N_K = 2n^2 + m^2 + 2\min(n,m)$ parameters. The number of parameters only increases significantly when $m >> n$. As RNNs typically set $n = m$, this is not a concern.

4 Experiments

In Section 1, it was highlighted that convergent dynamics are thought to play a crucial role in forming representations and enhancing robustness. In the following section, we begin to study the impact of the k-contraction constraints in terms of these properties. Herein, we refer to (2) as the Lurie network and Theorem 3 as the k-contracting Lurie network.

4.1 Forming representations

To test the models ability to form representations, the k-contracting Lurie network was benchmarked on the fashion MNIST classification task Xiao et al. (2017). The k-contraction parameter was set to its highest value, $k = 3$, to permit training over the three convergence behaviours illustrated in Figure 1. Due to the superior performance of CNN's on computer vision tasks, and their biological inspiration, the Lurie network's were also coupled with a small CNN ($55,744$ parameters) which pre-processed the images before feeding the output into the Lurie network via the initial condition. Further details regarding the data, architectures and training details were included in Appendix E.

Table 1: Classification accuracy on Fashion-MNIST test set. Other models were either the highest ranked or biologically-inspired from https://paperswithcode.com/sota/fashion-mnist.

Model	**# Params.**	**Acc. (%)**
LR-Net	$1,028,234$	95.03
Inception v3	$23,851,784$	94.44
CNN + Wilson-Cowan RNN	$5,179,521$	91.35
Wilson-Cowan RNN	-	88.39
CNN + k-Lurie Network	$1,057,994$	91.33
CNN + Lurie Network	$1,057,994$	90.95
Lurie Network	$1,853,386$	89.64
k-Lurie Network	$1,853,386$	85.78

Table 2: MSE ($\times 10^{-1}$) on test and noisy out of distribution data from simulated Hopfield network and simple attractor.

Model	$\textbf{MSE}^{Hop}_{test}$	$\textbf{MSE}^{Hop}_{ood}$	$\textbf{MSE}^{Att}_{test}$	$\textbf{MSE}^{Att}_{ood}$
k-Lurie Network	0.150	3.200	0.035	12.80
Lurie Network	3.600	4.400	5.100	67.90
Neural ODE	0.250	16.30	0.200	47.60

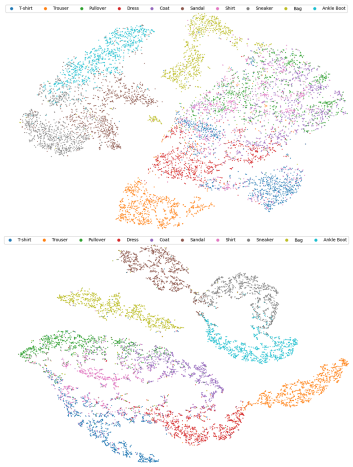

Figure 2: t-SNE plots of F-MNIST test set (top) and output of k-contracting Lurie network with test set as initial conditions (bottom).

Table 1 highlights that the k-contraction constraints result in only a 3.86% drop in classification accuracy compared to the Lurie network. This suggests the k-contraction constraints do not limit the expressivity of the Lurie network too significantly, even with the scalar metric limitation. Figure 2 shows how the k-contracting Lurie network is able to untangle the classes with a degree of interpretability by locating similar classes, such as footwear, closer together in the latent space. The Lurie network also achieves the highest accuracy on this task for models which do not involve a CNN. When leveraging a small CNN to pre-process the images, the k-contracting Lurie network achieves superior performance. This accuracy is similar to the CNN + Wilson-Cowan RNN Marino et al. (2024) which has nearly $5\times$ the number of parameters, only 12% of which are attributed to the Wilson-Cowan RNN with the other 88% contributing to the CNN.

4.2 Robustness

For the second experiment, two regression problems were considered. The dynamics of a Hopfield network (Hop) and simple attractor (Att) were simulated for 2000 time steps, corresponding to a 20s period. For each dataset, 1000 trajectories were generated and 100 were held out for testing. To study the robustness of the representations, a new dataset was simulated over a 30s period with initial conditions sampled outside the training distribution. The initial conditions were subject to additive noise sampled from the standard normal distribution. See Appendix F for more details.

Table 2 presents the mean squared error (MSE) between the actual and predicted trajectories on the test and noisy out of distribution (OOD) datasets. The k-contracting Lurie network outperforms the Lurie network on each example, indicating that the k-contraction constraints play an important role in reducing the search space to a more generalisable region. Furthermore, the Neural ODE is competitive on the test set, particularly for the Hopfield network; but, is less so on the noisy OOD data. Furthermore, Figure 3 shows the k-contracting Lurie network has the most accurate approximation of the correct long-term behaviour.

5 Conclusion

The Lurie network was presented as a novel neural ODE, along with a stability result and parametrisation which ensured only models satisfying a generalised form of stability could be optimised over. The stability constraints include networks which converge to both multiple equilibrium points and limit cycles, theoretical properties observed in associative and working memory. Future theoretical work will try to overcome the scalar metric limitation and develop k-contraction constraints for graph-coupled networks. Future empirical work will study the performance and robustness more rigorously, by considering a wider range of machine learning applications.

REFERENCES

Antoine R Adamantidis, Carolina Gutierrez Herrera, and Thomas C Gent. Oscillating circuitries in the sleeping brain. *Nature Reviews Neuroscience*, 20(12):746–762, 2019.

Peter Ashwin, Stephen Coombes, and Rachel Nicks. Mathematical frameworks for oscillatory network dynamics in neuroscience. *The Journal of Mathematical Neuroscience*, 6:1–92, 2016.

Eyal Bar-Shalom, Omri Dalin, and Michael Margaliot. Compound matrices in systems and control theory: a tutorial. *Mathematics of Control, Signals, and Systems*, pp. 1–55, 2023.

Joaquin Carrasco, Matthew C Turner, and William P Heath. Zames-Falb multipliers for absolute stability: From O'Shea's contribution to convex searches. *European Journal of Control*, 28:1–19, 2016.

Andreu Cecilia, Samuele Zoboli, Daniele Astolfi, Ulysse Serres, and Vincent Andrieu. Generalized Lyapunov conditions for k-contraction: analysis and feedback design. 2023.

Veronica Centorrino, Francesco Bullo, and Giovanni Russo. Contraction analysis of Hopfield neural networks with Hebbian learning. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pp. 622–627. IEEE, 2022.

Ross Drummond, Chris Guiver, and Matthew C Turner. Exponential input-to-state stability for Lur'e systems via integral quadratic constraints and Zames–Falb multipliers. *IMA Journal of Mathematical Control and Information*, pp. dnae003, 2024.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

John 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.

Roger A Horn and Charles R Johnson. *Topics in matrix analysis*. Cambridge university press, 1994.

Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven Siegelbaum, A James Hudspeth, Sarah Mack, et al. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

Hassan K Khalil. Nonlinear systems. *Patience Hall*, 115, 2002.

Mikail Khona and Ila R Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23(12):744–766, 2022.

Leo Kozachkov, Mikael Lundqvist, Jean-Jacques Slotine, and Earl K Miller. Achieving stable dynamics in neural circuits. *PLoS computational biology*, 16(8):e1007659, 2020.

Leo Kozachkov, John Tauber, Mikael Lundqvist, Scott L Brincat, Jean-Jacques Slotine, and Earl K Miller. Robust and brain-like working memory through short-term synaptic plasticity. *PLOS Computational Biology*, 18(12):e1010776, 2022.

Leo Kozachkov, Jean-Jacques Slotine, and Dmitry Krotov. Neuron-astrocyte associative memory. *arXiv preprint arXiv:2311.08135*, 2023.

Dmitry Krotov and John Hopfield. Large associative memory problem in neurobiology and machine learning. *arXiv preprint arXiv:2008.06996*, 2020.

Laurent Lessard, Benjamin Recht, and Andrew Packard. Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM Journal on Optimization*, 26(1):57–95, 2016.

Mario Lezcano-Casado and David Martınez-Rubio. Cheap orthogonal constraints in neural networks: A simple parametrization of the orthogonal and unitary group. In *International Conference on Machine Learning*, pp. 3794–3803. PMLR, 2019.

Winfried Lohmiller and Jean-Jacques E Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998.

Raffaele Marino, Lorenzo Buffoni, Lorenzo Chicchi, Francesca Di Patti, Diego Febbe, Lorenzo Giambagli, and Duccio Fanelli. Learning in wilson-cowan model for metapopulation. *arXiv preprint arXiv:2406.16453*, 2024.

James S Muldowney. Compound matrices and ordinary differential equations. *The Rocky Mountain Journal of Mathematics*, pp. 857–872, 1990.

Ron Ofir, Jean-Jacques Slotine, and Michael Margaliot. k-contraction in a generalized Lurie system. *arXiv preprint arXiv:2309.07514*, 2023.

Ron Ofir, Alexander Ovseevich, and Michael Margaliot. Contraction and k-contraction in Lurie systems with applications to networked systems. *Automatica*, 159:111341, 2024.

Matthijs Pals, Jakob H Macke, and Omri Barak. Trained recurrent neural networks develop phase-locked limit cycles in a working memory task. *PLOS Computational Biology*, 20(2):e1011852, 2024.

Poogyeon Park. A revisited Popov criterion for nonlinear Lur'e systems with sector-restrictions. *International Journal of Control*, 68(3):461–470, 1997.

PooGyeon Park. Stability criteria of sector-and slope-restricted Lur'e systems. *IEEE Transactions on Automatic Control*, 47(2):308–313, 2002.

Patricia Pauli, Dennis Gramlich, Julian Berberich, and Frank Allgöwer. Linear systems with neural network nonlinearities: Improved stability analysis via acausal Zames-Falb multipliers. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 3611–3618. IEEE, 2021.

Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.

Alex T Piet, Jeffrey C Erlich, Charles D Kopec, and Carlos D Brody. Rat prefrontal cortex inactivations during decision making are explained by bistable attractor dynamics. *Neural computation*, 29(11):2861–2886, 2017.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, et al. Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*, 2020.

Carl Richardson, Matthew Turner, Steve Gunn, and Ross Drummond. Strengthened stability analysis of discrete-time Lurie systems involving ReLU neural networks. In *6th Annual Learning for Dynamics & Control Conference*, pp. 209–221. PMLR, 2024.

Carl R Richardson, Matthew C Turner, and Steve R Gunn. Strengthened Circle and Popov Criteria for the stability analysis of feedback systems with ReLU neural networks. *IEEE Control Systems Letters*, 2023.

Sugandha Sharma, Sarthak Chandra, and Ila Fiete. Content addressable memory without catastrophic forgetting by heteroassociation with a fixed scaffold. In *International Conference on Machine Learning*, pp. 19658–19682. PMLR, 2022.

Matthew C Turner and Ross Drummond. Analysis of MIMO Lurie systems with slope restricted nonlinearities using concepts of external positivity. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 163–168. IEEE, 2019.

Ryan Vogt, Maximilian Puelma Touzel, Eli Shlizerman, and Guillaume Lajoie. On Lyapunov exponents for RNNs: Understanding information propagation using dynamical systems tools. *Frontiers in Applied Mathematics and Statistics*, 8:818799, 2022.

Hugh R Wilson and Jack D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1–24, 1972.

Chengshuai Wu, Ilya Kanevskiy, and Michael Margaliot. k-contraction: Theory and applications. *Automatica*, 136:110048, 2022.

Hedi Xia, Vai Suliafu, Hangjie Ji, Tan Nguyen, Andrea Bertozzi, Stanley Osher, and Bao Wang. Heavy ball neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 34:18646–18659, 2021.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

George Zames and PL Falb. Stability conditions for systems with monotone and slope-restricted nonlinearities. *SIAM Journal on Control*, 6(1):89–108, 1968.

Xiaojiao Zhang and Baotong Cui. Synchronization of Lurie system based on contraction analysis. *Applied Mathematics and Computation*, 223:180–190, 2013.

A NOTATION

For two integers $i < j$, we define $[i, j] := \{i, i+1, \dots, j\}$. The set of non-negative real numbers is denoted by $\Re_+$. Symmetric matrices of dimension n are denoted by $\mathcal{S}^n$ with the positive definite subset denoted by $\mathcal{S}^n_+$. All other positive definite subsets are denoted by a $+$ subscript. Square diagonal matrices are denoted by $\mathcal{D}^n$ and $n \times m$ diagonal matrices are symbolised by $\mathcal{D}^{nm}$. A positive definite (semi-definite) matrix P is sometimes indicated by $P \succ 0$ ($P \succeq 0$). Negative definite (semi-definite) matrices are indicated analogously. The set of $n \times n$ orthogonal and skew-symmetric matrices are respectively denoted by $\mathcal{SO}(n)$ and $so(n)$. For $W \in \Re^{n\times m}$, the ordered singular values are represented by $\sigma_1(W) \geq \cdots \geq \sigma_{\min(n,m)}(W) \geq 0$ and for $W \in \Re^{n\times n}$, the ordered eigenvalues are denoted by $\lambda_1(W) \geq \cdots \geq \lambda_n(W)$. The k-multiplicative and k-additive compound matrices of W are respectively denoted by $W^{(k)}$ and $W^{[k]}$. The Jacobian of a function $f(t, x)$ is denoted by $J_f(t, x)$. The scaled 2-norm of a vector $x \in \Re^n$ with respect to (w.r.t) an invertible scaling matrix $\Theta \in \Re^{n\times n}$ is defined by $|x|_{2,\Theta} := |\Theta x|_2$, and the matrix measure induced by the scaled 2-norm is

$$\mu_{2,\Theta}(W) := \mu_2(\Theta W \Theta^{-1}) = \lambda_1\Big(\frac{\Theta W \Theta^{-1} + (\Theta W \Theta^{-1})^\top}{2}\Big)$$

B COMPOUND MATRICES

In this section, we document several known definitions and algebraic results related to compound matrices. The results are included without proof; the interested reader should refer to Bar-Shalom et al. (2023) for a more detailed tutorial on the topic.

Let n be a positive integer and fix $k \in [1, n]$. The ordered set of *increasing* sequences of k integers from $[1, n]$ is denoted by $Q(k, n)$. For example: $Q(3, 4) = \{(1, 2, 3), (1, 2, 4), (1, 3, 4), (2, 3, 4)\}$.

Now consider a matrix $W \in \Re^{n\times m}$. For $\alpha \in Q(k, n)$ and $\beta \in Q(k, m)$, the matrix $W[\alpha|\beta]$ denotes the $k \times k$ sub-matrix obtained by taking the entries of W along the rows indexed by α and columns indexed by β. As an example, if $k = 2$ and $n = m = 4$, then $Q(2, 4) = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$. The sub-matrix $W[(1, 2)|(3, 4)]$ would then be given by

$$W[(1,2)|(3,4)] = \begin{bmatrix} w_{13} & w_{14} \\ w_{23} & w_{24} \end{bmatrix}$$

The k-minors of the matrix W are defined as $W(\alpha|\beta) := \det(W[\alpha|\beta])$.

Definition 1 (k-multiplicative compound) *Let $W \in \Re^{n\times m}$ and fix $k \in [1, \min(n, m)]$. The k-multiplicative compound of W, denoted $W^{(k)}$, is the $\binom{n}{k} \times \binom{m}{k}$ matrix containing all the k-minors of W ordered lexicographically.*

For example, if we have $n = m = 3$ and $k = 2$ then $\alpha, \beta \in Q(2, 3) = \{(1, 2), (1, 3), (2, 3)\}$ and

$$W^{(2)} = \begin{bmatrix} W\big((1,2)|(1,2)\big) & W\big((1,2)|(1,3)\big) & W\big((1,2)|(2,3)\big) \\ W\big((1,3)|(1,2)\big) & W\big((1,3)|(1,3)\big) & W\big((1,3)|(2,3)\big) \\ W\big((2,3)|(1,2)\big) & W\big((2,3)|(1,3)\big) & W\big((2,3)|(2,3)\big) \end{bmatrix}$$

Some important special cases include

$$W^{(1)} = W \qquad W^{(n)} = \det(W) \qquad (pI_n)^{(k)} = p^k I_s \qquad W \in \mathcal{D}^n \to W^{(k)} \in \mathcal{D}^s \tag{6}$$

with $s := \binom{n}{k}$. Next, we present a series of algebraic results concerned with the k-multiplicative compound.

Fact 1 (Cauchy-Binet Formula) *If $U \in \Re^{n\times m}$, $V \in \Re^{m\times p}$ and $k \in [1, \min(n, m, p)]$, then*

$$(UV)^{(k)} = U^{(k)} V^{(k)}$$

Fact 2 *Fix $k \in [1, \min(n, m)]$. As a consequence of Definition 1, if $W \in \Re^{n\times m}$ then*

$$(W^\top)^{(k)} = (W^{(k)})^\top$$

Fact 3 *Fix $k \in [1, n]$. If $W \in \Re^{n \times n}$ is non-singular, then by Theorem 1*

$$(W^{-1})^{(k)} = (W^{(k)})^{-1}$$

Fact 4 *Fix $k \in [1, \min(n, m, p)]$. If $W \in \Re^{n \times n}$, $U \in \Re^{p \times n}$ and $V \in \Re^{n \times p}$, then by Theorem 1*

$$(UWV)^{(k)} = U^{(k)} W^{(k)} V^{(k)}$$

Fact 5 *Fix $k \in [1, n]$. An implication of Theorem 1 is that if $W \in \Re^{n \times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$, then the eigenvalues of $W^{(k)}$ are the $\binom{n}{k}$ products*

$$\Big\{ \prod_{l=1}^{k} \lambda_{i_l} : 1 \leq i_1 < \cdots < i_k \leq n \Big\}$$

We now introduce the definition of a second compound matrix, the k-additive compound, and a set of algebraic results related to it.

Definition 2 (k-additive compound) *Let $W \in \Re^{n \times n}$ and $k \in [1, n]$. The k-additive compound of W is the $\binom{n}{k} \times \binom{n}{k}$ matrix defined by*

$$W^{[k]} := \frac{d}{d\epsilon} \big(I_n + \epsilon W \big)^{(k)} |_{\epsilon = 0}$$

Special cases include

$$W^{[1]} = W \qquad W^{[n]} = tr(W) \qquad (pI_n)^{[k]} = kpI_s \qquad W \in \mathcal{D}^n \rightarrow W^{[k]} \in \mathcal{D}^s \tag{7}$$

with $s := \binom{n}{k}$. Like before, we now present some useful algebraic results related to the k-additive compound.

Fact 6 *If $W \in \Re^{n \times n}$ and $k \in [1, n]$, then as a consequence of Definition 2*

$$(W^\top)^{[k]} = (W^{[k]})^\top$$

Fact 7 *Fix $k \in [1, n]$. For $W \in \Re^{n \times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$, the eigenvalues of $W^{[k]}$ are the $\binom{n}{k}$ sums*

$$\{ \sum_{l=1}^{k} \lambda_{i_l} : 1 \leq i_1 < \cdots < i_k \leq n \}$$

An important consequence of Fact 7 is that if W is positive definite (semi-definite), then this property is upheld by $W^{[k]}$. Opposite conclusions can be drawn if W is negative definite (semi-definite).

Fact 8 *Fix $k \in [1, n]$. If $U, V \in \Re^{n \times n}$, then*

$$(U + V)^{[k]} = U^{[k]} + V^{[k]}$$

Fact 9 *Fix $k \in [1, \min(n, p)]$. If $W \in \Re^{n \times n}$, $U \in \Re^{p \times n}$, $V \in \Re^{n \times p}$ and $UV = I_p$, then*

$$(UWV)^{[k]} = U^{(k)} W^{[k]} (U^{(k)})^{-1}$$

C Proof of Theorem 2

We aim to verify Theorem 1 for the particular case where the nonlinear system is described by the Lurie network (2). Our proof begins with Theorem 4, which restates (Ofir et al., 2024, Theorem 1). This result is sufficient to satisfy Theorem 1 for systems of the form (8).

Theorem 4 *Fix $k \in [1, n]$ and consider the system below.*

$$\dot{x} = \bar{A}x(t) - \bar{B}\Psi(t, y) \qquad y = \bar{C}x \tag{8}$$

If there exists $\eta_1, \eta_2 > 0$ and an invertible $\Theta \in \Re^{n\times n}$ such that

$$P^{(k)}\bar{A}^{[k]} + (\bar{A}^{[k]})^\top P^{(k)} + \Theta^{(k)}\Big((\Theta\bar{B}\bar{B}^\top\Theta)^{[k]} + (\Theta^{-1}\bar{C}^\top\bar{C}\Theta^{-1})^{[k]}\Big)\Theta^{(k)} \preceq -\eta_1 P^{(k)} \tag{9}$$

and

$$\Big(\Theta^{-1}\bar{C}^\top(J_\Psi^\top(t,y)J_\Psi(t,y) - I_m)\bar{C}\Theta^{-1}\Big)^{[k]} \preceq -\eta_2 I_s \quad \forall\, t \in \Re_+ \text{ and } y \in \Re^m \tag{10}$$

where $s = \binom{n}{k}$, then (8) is k-contracting in the 2-norm w.r.t the metric $P := \Theta^\top\Theta$.

We first need to express the Lurie network in the form (8). By (3) there exists $\gamma < 0$ satisfying

$$0 < \gamma^2 < \alpha_k^2 \quad \text{and} \quad g^2\sum_{i=1}^{k}\sigma_i^2(B)\sigma_i^2(C) < \gamma^2 k \tag{11}$$

Using γ, we can express (2) in the form (8) through the definitions below, where the dependence on t has been dropped from Ψ.

$$\bar{A} := A \qquad \bar{B} := \gamma I_n \qquad \bar{C} := I_n \qquad \Psi(x) := -\gamma^{-1}B\Phi(Cx + b_y) - \gamma^{-1}b_x \tag{12}$$

The next step is to verify (9). Subbing (12) into the left hand side of (9) and assuming $\Theta = \Theta^\top$ results in the first equality. Setting $P := pI_n$ with $p > 0$ results in the second. Now we must leverage some of the facts presented in §B. Using the relevant special cases from (6) and (7) leads to equality three and consequently applying Fact 6 and Fact 8 results in equality four. Re-applying (7) and Fact 8 results in the final equality.

$$\begin{aligned}
&= P^{(k)}A^{[k]} + (A^{[k]})^\top P^{(k)} + \Theta^{(k)}\Big((\gamma^2 P)^{[k]} + (P^{-1})^{[k]}\Big)\Theta^{(k)} \\
&= (pI_n)^{(k)}A^{[k]} + (A^{[k]})^\top(pI_n)^{(k)} + (p^{\frac{1}{2}}I_n)^{(k)}\Big((\gamma^2 pI_n)^{[k]} + (p^{-1}I_n)^{[k]}\Big)(p^{\frac{1}{2}}I_n)^{(k)} \\
&= p^k\big(A^{[k]} + (A^{[k]})^\top\big) + k(\gamma^2 p + p^{-1})p^k I_s \\
&= p^k\big((A + A^\top)^{[k]} + k(\gamma^2 p + p^{-1})I_s\big) \\
&= p^k\big(A + A^\top + (\gamma^2 p + p^{-1})I_n\big)^{[k]}
\end{aligned}$$

If the matrix above is negative definite, then (9) is satisfied for some suitably chosen $\eta_1 > 0$. This is true when

$$\big(A + A^\top + (\gamma^2 p + p^{-1})I_n\big)^{[k]} \prec 0$$

By Fact 7, the inequality above can be equivalently expressed as a condition on the sum of the k largest eigenvalues of the matrix inside the k-compound operator. Leveraging (Petersen et al., 2008, Eq. 285) allows us to separate p from the eigenvalues of the symmetric component of A, resulting in the equality below.

$$\sum_{i=1}^{k}\lambda_i\big(A + A^\top + (\gamma^2 p + p^{-1})I_n\big) = k(\gamma^2 p + p^{-1}) + \sum_{i=1}^{k}\lambda_i(A + A^\top) < 0$$

By the definition of α_k in Theorem 2, this simplifies to

$$\gamma^2 p^2 + 2\alpha_k p + 1 < 0$$

For γ satisfying (11), the quadratic inequality always emits at least one solution $p = -\alpha_k^{-1}$.

The final step is to verify (10). The Jacobian of Ψ, as defined in (12), is

$$J_\Psi(x) = -\gamma^{-1} B J_\Phi C$$

For $\Theta = p^{\frac{1}{2}} I_n$ and the definitions from (12), the left hand side of (10) reduces to

$$= \left(p^{-1} J_\Psi^\top J_\Psi - p^{-1} I_n\right)^{[k]}$$

If the matrix above is negative definite, then (10) is satisfied for some suitably chosen $\eta_2 > 0$. Repeating the same steps as above, this negative definite requirement reduces to the inequality below.

$$\sum_{i=1}^{k} \lambda_i(p^{-1} J_\Psi^\top J_\Psi) - kp^{-1} = p^{-1} \sum_{i=1}^{k} \sigma_i^2(J_\Psi) - kp^{-1} < 0$$

Subbing in the definition of J_Ψ and applying the well-known property of singular values (Horn & Johnson, 1994, Theorem 3.3.14), then (10) is true if

$$\gamma^{-2} \sum_{i=1}^{k} \sigma_i^2(B)\sigma_i^2(J_\Phi)\sigma_i^2(C) < k$$

By the assumption made on the slope of Φ, this inequality will always be satisfied if (3) holds. □

D Proof of Theorem 3

The aim of this proof is to express the weights of the Lurie network (2) such that Theorem 2 is always satisfied. More formally, this requires $A, B, C \in \Omega_2(g, k)$ to always hold.

To expose the singular values of B, C, we leverage the singular value decomposition. This requires the matrices U_B, U_C, V_B, V_C to be orthogonal. We can immediately use the unconstrained parametrisation of the orthogonal class from Lezcano-Casado & Martınez-Rubio (2019) to express these matrices as unconstrained symmetric matrices. The matrices Σ_B, Σ_C contain the singular values of the respective matrix, hence $\Sigma_B \in \mathcal{D}_+^{nm}$ and $\Sigma_C \in \mathcal{D}_+^{mn}$. We also treat these as unconstrained sets since any element can be obtained by taking the absolute value of an unconstrained diagonal matrix with the same shape.

To verify Theorem 2, we can combine $\alpha_k < 0$ and (3) into one inequality representing the intersection of the two sets.

$$\sum_{i=1}^{k} \lambda_i(A + A^T) < -2k\sqrt{\frac{z}{k}} \quad \text{where } z := g^2 \sum_{i=1}^{k} \sigma_i^2(B)\sigma_i^2(C) \tag{13}$$

Thanks to the definition of B and C, the right hand side is a function of the hyperparameters g, k and elements of the parameters Σ_B, Σ_C, so can be easily computed using sort and sum functions.

To impose this constraint directly on the eigenvalues of the symmetric component of A, we express A as a sum of symmetric and skew-symmetric matrices. The skew-symmetric matrix is unconstrained, so this can be left alone. Finally, we apply the eigenvalue decomposition of a symmetric matrix to obtain (5a). This expression allows us to directly place the constraint above on the diagonal matrix Σ_A.

Defining Σ_A as in (5b) guarantees both conditions of Theorem 2 will be satisfied via (13). The definition of Σ_A is split into one unconstrained block for the first $(k-1)$-eigenvalues, a block for $\lambda_k(A + A^\top)$ which is defined to ensure (13) holds, and finally a block for the remaining eigenvalues which must be defined to ensure the k eigenvalues involved in (13) are the largest. □

E EXPERIMENTS: FORMING REPRESENTATIONS

E.1 DATA

The Fashion-MNIST dataset contains 28×28 grayscale images of $70,000$ fashion products, from 10 different categories. For all experiments, the images were normalised and zero-centred.

Table 3: Fashion-MNIST settings.

Parameter	**Value**
Batch size	250
Test split	$\frac{1}{7}$

E.2 MODELS

For the Lurie network and k-contracting Lurie network, the transformed images were flattened and directly passed into the models through the initial condition. Euler integration was used to approximate the trajectory of (2) and the final state was mapped through a linear layer and followed by a softmax layer to obtain the categorical predictions.

Table 4: k-Lurie network and Lurie network settings.

Parameter	**Description**	**Value**
Step size	Euler integration step size	1×10^{-2}
Steps	Number of Euler integration steps	100
Activation function	-	$\tanh$
g	Upper bound on slope of activation	1
n (without CNN / with CNN)	dimension of x	$784/576$
m (without CNN / with CNN)	dimension of y	$784/576$
k	k-contraction parameter	3

Table 5: CNN settings.

Layer	**Convolution**	**Activation**	**Downsampling / Reshaping**
1	2D convolution (1 input, 32 outputs, 3×3 kernel)	$ReLU$	2D max pool (2×2 kernel)
2	2D convolution (32 inputs, 64 outputs, 3×3 kernel)	$ReLU$	2D max pool (2×2 kernel)
3	2D convolution (64 inputs, 64 outputs, 3×3 kernel)	$ReLU$	Flatten (1D output)

E.3 TRAINING

Table 6: Training settings.

Parameter	**Value**
Loss	Cross Entropy
Optimiser	Adam
Weight decay	1×10^{-5}
Epochs	50
Learning rate (Lurie net.)	1×10^{-3}
Learning rate (k-Lurie net.)	1×10^{-2}
Learning rate cut	0.5 at epoch 35

F EXPERIMENTS: ROBUSTNESS

F.1 DATA

The data was synthetically generated by numerically integrating over the analytical models of each dynamical system. For each dataset, 1000 trajectories were sampled every 0.01 seconds over a 20 second interval. The test sets were formed by holding out 100 trajectories.

Table 7: Hopfield network and simple attractor settings.

Parameter	**Value**
Batch size	100
Test split	0.1

F.1.1 HOPFIELD NETWORK

This model is a variation of the Hopfield network presented in Ofir et al. (2024). It has the following state space equations

$$\dot{x} = -2.5I_3 + B\Phi(x)$$

where

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

and the $\tanh$ function is also applied element-wise. The model is 2-contracting and has two stable equilibrium points: $e_1 = [0.79 \quad 0.79 \quad 0.79]^\top$, $e_2 = -e_1$; and an unstable equilibrium point $e_3 = 0$. Initial conditions were sampled from a uniform distribution with domain $(-1, +1)^3$.

F.1.2 SIMPLE ATTRACTOR

This model was presented in Cecilia et al. (2023). It has the following state space equations

$$\dot{x} = Ax + B\Phi(Cx)$$

where

$$A = \begin{bmatrix} 0 & 1 & -2 \\ -1 & 0 & -1 \\ 0.5 & 0 & -0.5 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -0.5 & 0 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

and $\phi(z) = z^3$ is the nonlinearity applied element-wise. This function is not slope-restricted, so the simple attractor does not satisfy the assumptions of the k-contracting Lurie network. The model is 3-contracting and has several attractor states. Initial conditions were sampled from a uniform distribution with domain $(-3, +3)^3$.

F.2 MODELS

Table 8: k-Lurie network and Lurie network settings.

Parameter	**Description**	**Value**
Step size	Euler integration step size	1×10^{-2}
Steps	Number of Euler integration steps	2000
Activation function	-	$\tanh$
g	Upper bound on slope of activation	1
n	dimension of x	3
m	dimension of y	3
k	k-contraction parameter	3

Table 9: Neural ODE settings based on Xia et al. (2021).

Layer	Output dim.	Activation
1	20	ReLU
2	20	ReLU
3	3	Linear

F.3 Training

Table 10: Training settings.

Parameter	Value
Loss	Mean squared error
Optimiser	Adam
Weight decay	1×10^{-5}
Epochs	100
Learning rate (k-Lurie net.)	1×10^{-2}
Learning rate (Lurie net.)	1×10^{-3}
Learning rate (Neural ODE)	1×10^{-3}

F.4 Extended results

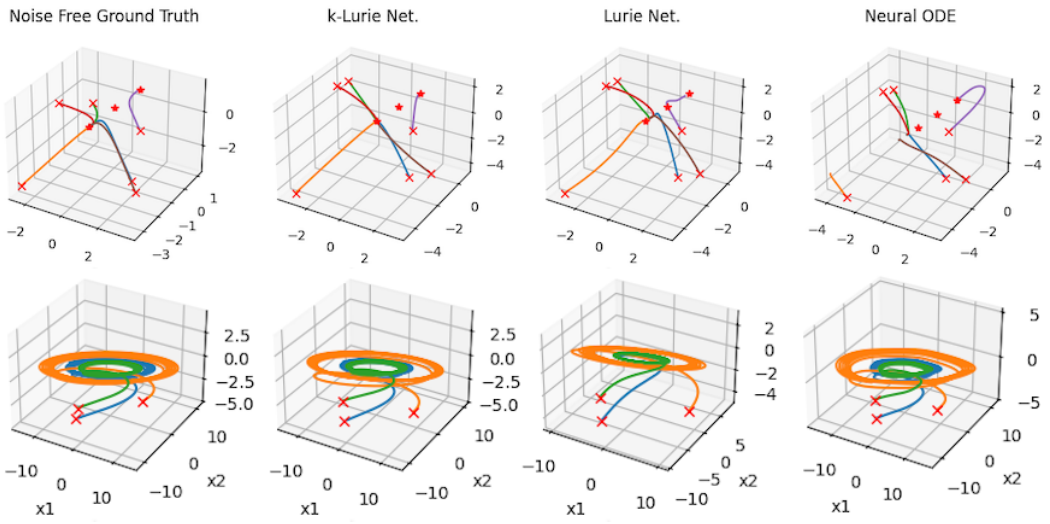

Figure 3: Trajectories from Hopfield network (top) and simple attractor (bottom). Crosses denote the initial condition and stars denote equilibirum points.