Lurie Networks with k-contracting Dynamics

Carl R. Richardson[1, 2], Matthew C. Turner[1], Steve R. Gunn[1]

University of Southampton [1] The Alan Turing Institute [2]

Introduction

Motivation: Many interesting dynamical systems exhibit some form of convergence e.g., to equilibrium points or limit cycles. In the central nervous system, convergence is thought to play a crucial role in forming representations, processing information, learning, memory storage, and enhancing robustness.

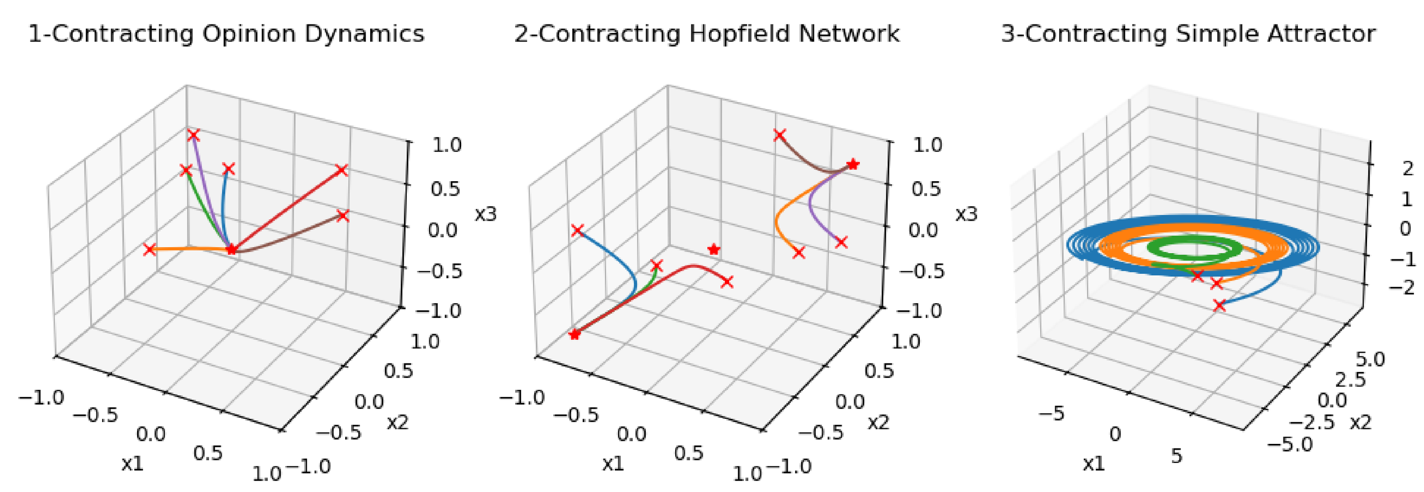

Figure 1. Examples of convergent dynamical systems.

Research questions:

1. Can convergence be harnessed to develop a general and robust ML framework?
2. Does encoding convergence as an inductive bias lead to more robust models of dynamical systems?

Solution: We apply k-contraction analysis to parametrise a novel Neural ODE, the Lurie network, such that convergence to a point, line, or plane in the neural state space is guaranteed. As the parametrisation is unconstrained, the k-contracting Lurie network can still be trained using standard optimisation tools.

Lurie Network

The Lurie network has the functional form below which describes the states evolution by a linear term plus a slope-restricted nonlinearity, $\Phi(\cdot)$. It has trainable weights $A \in \Re^{n\times n}$, $B \in \Re^{n\times m}$, $C \in \Re^{m\times n}$ and biases $b_x \in \Re^n$, $b_y \in \Re^m$.

$$\dot{x} = Ax + B\Phi(y) + b_x \qquad y = Cx + b_y \tag{1}$$

If a Lurie network is k-contracting, it implies that all trajectories in the neural state space converge to a

- set of limit cycles, within the same plane, when $k = 3$.
- set of equilibrium points, connected along a line, when $k = 2$.
- unique equilibrium point, when $k = 1$.

k-contraction Analysis of Lurie Network

Consider (1) with $\Phi(\cdot)$ being slope-restricted with upper bound, g. Fix $k \in \{1, \dots, n\}$ and define $\alpha_k := (2k)^{-1}\sum_{i=1}^{k} \lambda_i(A + A^\top)$. If $\alpha_k < 0$ and

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

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

Parametrisation of k-contracting Lurie Networks

Given hyperparameters $g > 0$, $k \in \{1, \dots, n\}$ and trainable parameters $U_A, U_B, U_C, V_B, V_C \in \mathcal{O}$, $\Sigma_B, \Sigma_C, G_{A3} \in \mathcal{D}_+$, $Y_A \in \mathrm{Skew}$, $\Sigma_{A1} \in \mathcal{D}$, $G_{A2} > 0$, then (1) constructed according to

$$A := \frac{1}{2}U_A\Sigma_A U_A^\top + \frac{1}{2}Y_A \qquad \Sigma_A := \mathrm{blockdiag}(\Sigma_{A1}, \Sigma_{A2}, \Sigma_{A3})$$

$$B := U_B\Sigma_B V_B^\top \qquad C := U_C\Sigma_C V_C^\top$$

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

is guaranteed to be k-contracting in the 2-norm w.r.t the metric $P = -\alpha_k^{-1} I_n$.

Theoretical Properties

+ The parametrisation does not require symmetry of the Lurie network weights, as is the case in many Hopfield-based models.
+ Neither the k-contraction result or the parametrisation require computation of the troublesome compound matrices.
– Only Lurie networks which are k-contracting in a scalar metric can be verified.

Results: Forming Representations

Table 1. Classification accuracy on FMNIST test set.

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

- k-contraction constraints have minimal impact on expressivity.
- Coupling small CNN compensates for reduced expressivity, whilst k-contraction constraints effectively process info. over time.

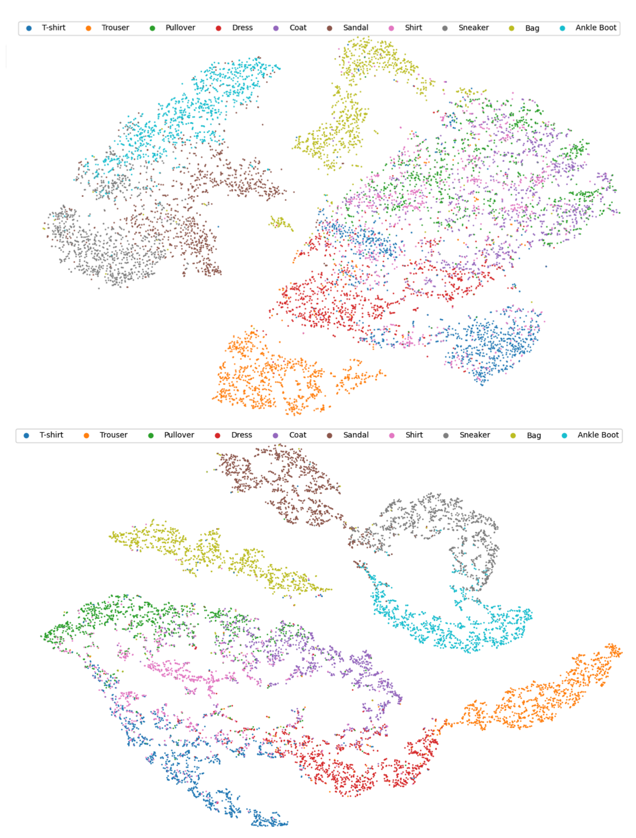

Figure 2. t-SNE plots of FMNIST test set (top) and k-Lurie net.

Results: Robustness

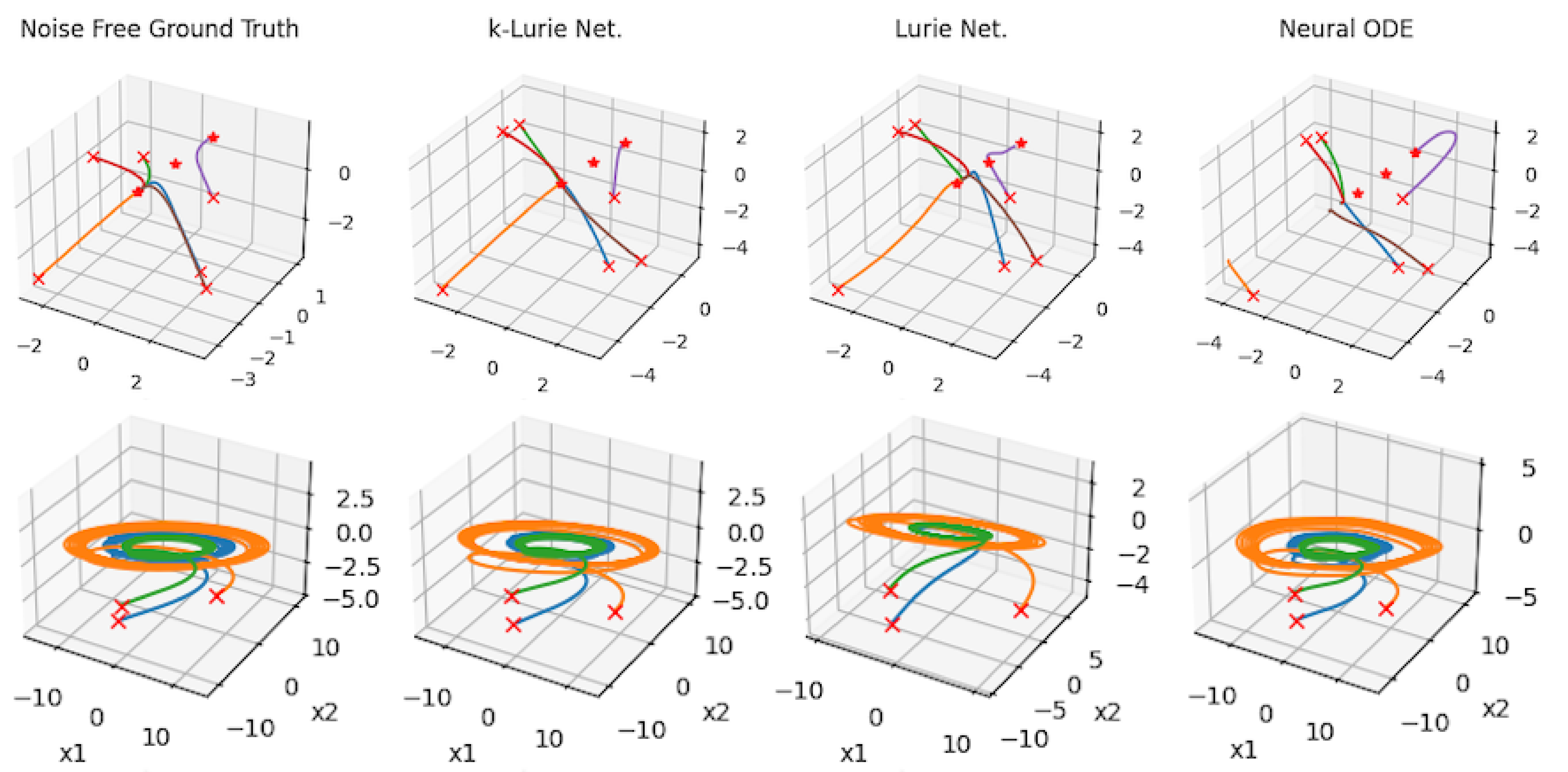

Figure 3. Predictions given noisy and OOD initial conditions, ×. Equilib. points denoted by *.

- k-Lurie network predicts the long-term behaviour most accurately, on both examples, even when initial conditions are OOD and subject to noise. This implies building convergence into the model leads to more robust solutions.

Future Work

- Generalise results to systems which are k-contracting in a diagonal metric.
- Extend to k-contracting graph-coupled Lurie networks.
- Apply to a wider range of applications including PDEs, graphs with temporally evolving features, and associative / working memory.
- Use as the recurrent module in larger ML frameworks e.g., RL agent.

Notation and Code

$\lambda_i(H)$	i^{th} largest eigenvalue of $H \in \Re^{n\times n}$
$\sigma_i(H)$	i^{th} largest singular value of $H \in \Re^{n\times m}$
$\mathcal{O}$	set of orthogonal matrices
$\mathcal{D}$	set of diagonal matrices
$\mathcal{D}_+$	set of positive diagonal matrices
Skew	set of skew-symmetric matrices

University of Southampton

The Alan Turing Institute

UK Research and Innovation

cr2g16@soton.ac.uk