LETTER **Communicated by Lea Duncker**

Probing the Relationship Between Latent Linear Dynamical Systems and Low-Rank Recurrent Neural Network Models

Adrian Valente
adrian.valente@ens.fr
Srdjan Ostojic
srdjan.ostojic@ens.fr
Laboratoire de Neurosciences Cognitives et Computationnelles, INSERM U960, Ecole Normale Superieure–PSL Research University, 75005 Paris, France

Jonathan W. Pillow
pillow@princeton.edu
Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, U.S.A.

A large body of work has suggested that neural populations exhibit low-dimensional dynamics during behavior. However, there are a variety of different approaches for modeling low-dimensional neural population activity. One approach involves latent linear dynamical system (LDS) models, in which population activity is described by a projection of low-dimensional latent variables with linear dynamics. A second approach involves low-rank recurrent neural networks (RNNs), in which population activity arises directly from a low-dimensional projection of past activity. Although these two modeling approaches have strong similarities, they arise in different contexts and tend to have different domains of application. Here we examine the precise relationship between latent LDS models and linear low-rank RNNs. When can one model class be converted to the other, and vice versa? We show that latent LDS models can only be converted to RNNs in specific limit cases, due to the non-Markovian property of latent LDS models. Conversely, we show that linear RNNs can be mapped onto LDS models, with latent dimensionality at most twice the rank of the RNN. A surprising consequence of our results is that a partially observed RNN is better represented by an LDS model than by an RNN consisting of only observed units.

1 Introduction

Recent work on large-scale neural population recordings has suggested that neural activity is often confined to a low-dimensional space, with fewer dimensions than the number of neurons in a population (Churchland, Byron, Sahani, & Shenoy, 2007; Gao & Ganguli, 2015; Gallego, Perich,

Neural Computation **34, 1871–1892** (2022)
https://doi.org/10.1162/neco_a_01522

Miller, & Solla, 2017; Saxena & Cunningham, 2019; Jazayeri & Ostojic, 2021). To describe this activity, modelers have at their disposal a wide array of tools that give rise to different forms of low-dimensional activity (Cunningham & Yu, 2014). Two classes of modeling approaches that have generated a large following in the literature are descriptive statistical models and mechanistic models. Broadly speaking, descriptive statistical models aim to identify a probability distribution that captures the statistical properties of an observed neural dataset, while remaining agnostic about the mechanisms that gave rise to it. Mechanistic models, by contrast, aim to reproduce certain characteristics of observed data using biologically inspired mechanisms, but often with less attention to a full statistical description. Although these two classes of models often have similar mathematical underpinnings, there remain a variety of important gaps between them. Here we focus on reconciling the gaps between two simple but powerful models of low-dimensional neural activity: latent linear dynamical systems (LDS) and linear low-rank recurrent neural networks (RNNs).

The latent LDS model with gaussian noise is a popular statistical model for low-dimensional neural activity in both systems neuroscience (Smith & Brown, 2003; Semedo, Zandvakili, Kohn, Machens, & Byron, 2014) and brain-machine interface settings (Kim, Simeral, Hochberg, Donoghue, & Black, 2008). This model has a long history in electrical engineering, where the problem of inferring latents from past observations has an analytical solution known as the Kalman filter (Kalman, 1960). In neuroscience settings, this model has been used to describe high-dimensional neural population activity in terms of linear projections of low-dimensional latent variables. Although the basic form of the model includes only linear dynamics, recent extensions have produced state-of-the-art models for high-dimensional spike train data (Yu et al., 2005; Petreska et al., 2011; Macke et al., 2011; Pachitariu, Petreska, & Sahani, 2013; Archer, Koster, Pillow, & Macke, 2014; Duncker, Bohner, Boussard, & Sahani, 2019; Zoltowski, Pillow, & Linderman, 2020; Glaser, Whiteway, Cunningham, Paninski, & Linderman, 2020; Kim et al., 2008).

Recurrent neural networks, by contrast, have emerged as a powerful framework for building mechanistic models of neural computations underlying cognitive tasks (Sussillo, 2014; Barak, 2017; Mante, Sussillo, Shenoy, & Newsome, 2013) and have more recently been used to reproduce recorded neural data (Rajan, Harvey, & Tank, 2016; Cohen, DePasquale, Aoi, & Pillow, 2020; Finkelstein et al., 2021; Perich et al., 2021). While randomly connected RNN models typically have high-dimensional activity (Sompolinsky, Crisanti, & Sommers, 1988; Laje & Buonomano, 2013), recent work has shown that RNNs with low-rank connectivity provide a rich theoretical framework for modeling low-dimensional neural dynamics and the resulting computations (Mastrogiuseppe & Ostojic, 2018; Landau & Sompolinsky, 2018; Pereira & Brunel, 2018; Schuessler, Dubreuil, Mastrogiuseppe, Ostojic, & Barak, 2020; Beiran, Dubreuil, Valente, Mastrogiuseppe, &

Ostojic, 2021; Dubreuil, Valente, Beiran, Mastrogiuseppe, & Ostojic, 2022; Bondanelli, Deneux, Bathellier, & Ostojic, 2021; Landau & Sompolinsky, 2021). In these low-rank RNNs, the structure of low-dimensional dynamics bears direct commonalities with latent LDS models, yet the precise relationship between the two classes of models remains to be clarified. Understanding this relationship would open the door to applying to low-rank RNNs probabilistic inference techniques developed for LDS models and conversely could provide mechanistic interpretations of latent LDS models fitted to data.

In this letter, we examine the mathematical relationship between latent LDS and low-rank RNN models. We focus on linear RNNs, which are less expressive but simpler to analyze than their nonlinear counterparts while still leading to rich dynamics (Hennequin, Vogels, & Gerstner, 2014; Kao, Sadabadi, & Hennequin, 2021; Bondanelli et al., 2021). We show that even if both LDS models and linear low-rank RNNs produce gaussian distributed activity patterns with low-dimensional linear dynamics, the two model classes have different statistical structures and are therefore not in general equivalent. More specifically, in latent LDS models, the output sequence has non-Markovian statistics, meaning that the activity in a single time step is not independent of its history given the activity on the previous time step. This stands in contrast to linear RNNs, which are Markovian regardless of the rank of their connectivity. A linear low-rank RNN can nevertheless provide a first-order approximation to the distribution over neural activity generated by a latent LDS model, and we show that this approximation becomes exact in several cases of interest, and in particular, in the limit where the number of neurons is large compared to the latent dimensionality. Conversely, we show that any linear low-rank RNN can be converted to a latent LDS, although the dimensionality of the latent space depends on the overlap between the subspaces spanned by left and right singular vectors of the RNN connectivity matrix and may be as high as twice the rank of this matrix. The two model classes are thus closely related, with linear low-rank RNNs comprising a subset of the broader class of latent LDS models. An interesting implication of our analyses is that the activity of an RNN in which only a subset of neurons are observed is better fit by a latent LDS model than by an RNN consisting only of observed units.

2 Modeling Frameworks

We start with a formal description of the two model classes in question, both of which describe the time-varying activity of a population of n neurons.

2.1 Latent LDS Model. The latent linear dynamical system (LDS) model, also known as a linear gaussian state-space model, describes neural population activity as a noisy linear projection of a low-dimensional latent

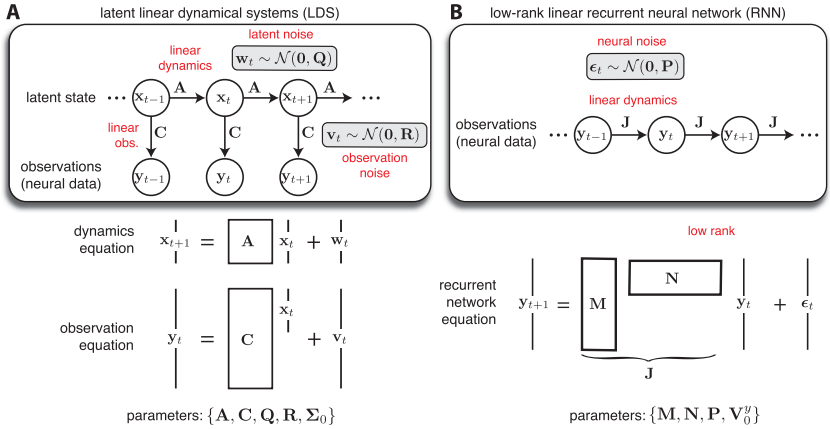

Figure 1: (A) Schematic representation of the latent linear dynamical system model, as defined by equations 2.1 to 2.3. (B) Schematic representation of the low-rank linear RNN, as defined by equations 2.4 and 2.5.

variable governed by linear dynamics with gaussian noise (Kalman, 1960; Roweis & Ghahramani, 1999; see Figure 1A). The model is characterized by the following equations:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \tag{2.1}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{2.2}$$

Here, $\mathbf{x}_t$ is a d-dimensional latent (or "unobserved") vector that follows discrete-time linear dynamics specified by a $d \times d$ matrix $\mathbf{A}$ and is corrupted on each time step by a zero-mean gaussian noise vector $\mathbf{w}_t \in \mathbb{R}^d$ with covariance $\mathbf{Q}$. The vector of neural activity $\mathbf{y}_t$ arises from a linear transformation of $\mathbf{x}_t$ via the $n \times d$ observation (or "emissions") matrix $\mathbf{C}$, corrupted by zero-mean gaussian noise vector $\mathbf{v}_t \in \mathbb{R}^n$ with covariance $\mathbf{R}$. Generally we assume $d < n$, so that the high-dimensional observations $\mathbf{y}_t$ are explained by the lower-dimensional dynamics of the latent vector $\mathbf{x}_t$. For clarity, in the main text, we focus on LDS models without external inputs and study their effect in appendix D.

The complete model also contains a specification of the distribution of the initial latent vector $\mathbf{x}_0$, which is commmonly assumed to have a zero-mean gaussian distribution with covariance $\mathbf{\Sigma}_0$:

$$\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_0). \tag{2.3}$$

The complete parameters of the model are thus $\theta_{LDS} = \{\mathbf{A}, \mathbf{C}, \mathbf{Q}, \mathbf{R}, \mathbf{\Sigma}_0\}$. Note that this parameterization of an LDS is not unique: any invertible

linear transformation of the latent space leads to an equivalent model if the appropriate transformations are applied to matrices $\mathbf{A}$, $\mathbf{C}$, $\mathbf{Q}$, and $\mathbf{\Sigma}_0$.

2.2 Low-Rank Linear RNN. A linear RNN, also known as an autoregressive (AR) model, represents observed neural activity as a noisy linear projection of the activity at the previous time step. We can write the model as (see Figure 1B)

$$\mathbf{y}_{t+1} = \mathbf{J}\mathbf{y}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{P}), \tag{2.4}$$

where $\mathbf{J}$ is an $n \times n$ recurrent weight matrix and $\boldsymbol{\epsilon}_t \in \mathbb{R}^n$ is a gaussian noise vector with mean zero and covariance $\mathbf{P}$. Moreover, we assume that the initial condition is drawn from a zero-mean distribution with covariance $\mathbf{V}_0^y$:

$$\mathbf{y}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_0^y). \tag{2.5}$$

A low-rank RNN model is obtained by constraining the rank of the recurrent weight matrix $\mathbf{J}$ to be $r \ll n$. In this case, the recurrence matrix can be factorized as

$$\mathbf{J} = \mathbf{M}\mathbf{N}^\top, \tag{2.6}$$

where $\mathbf{M}$ and $\mathbf{N}$ are both $n \times r$ matrices of rank r.

Note that this factorization is not unique, but a particular factorization can be obtained from a low-rank $\mathbf{J}$ matrix using the truncated singular value decomposition: $\mathbf{J} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, where $\mathbf{U}$ and $\mathbf{V}$ are semiorthogonal $n \times r$ matrices of left and right singular vectors, respectively, and $\mathbf{S}$ is an $r \times r$ diagonal matrix containing the largest singular values. We can then set $\mathbf{M} = \mathbf{U}$ and $\mathbf{N} = \mathbf{S}\mathbf{V}^\top$.

The model parameters of the low-rank linear RNN are therefore given by $\theta_{RNN} = \{\mathbf{M}, \mathbf{N}, \mathbf{P}, \mathbf{V}_0^y\}$.

2.3 Comparing the Two Models. Both models described above exhibit low-dimensional dynamics embedded in a high-dimensional observation space. In the following, we examine the probability distributions $P(\mathbf{y}_1, \ldots, \mathbf{y}_T)$ over time series $(\mathbf{y}_1, \ldots, \mathbf{y}_T)$ generated by the two models. We show that in general, the two models give rise to different distributions, such that the family of probability distributions generated by the LDS model cannot all be captured with low-rank linear RNNs. Specifically, RNN models are constrained to purely Markovian distributions, which is not the case for LDS models. However, the two model classes can be shown to be equivalent when the observations $\mathbf{y}_t$ contain exact information about the latent state $\mathbf{x}_t$, which is in particular the case if the observation noise is orthogonal to the latent subspace or in the limit of a large number of neurons

$n \gg d$. Conversely, a low-rank linear RNN can in general be mapped to a latent LDS with a dimensionality of the latent state at most twice the rank of the RNN.

3 Mapping from LDS Models to Linear Low-Rank RNNs

3.1 Nonequivalence in the General Case. Let us consider a latent LDS described by equations 2.1 to 2.3 and a low-rank linear RNN defined by equations 2.4 and 2.5. We start by comparing the properties of the joint distribution $P(\mathbf{y}_0, \ldots, \mathbf{y}_T)$ for any value of T for the two models. For both models, the joint distribution can be factored under the form

$$P(\mathbf{y}_0, \ldots, \mathbf{y}_T) = P(\mathbf{y}_0) \prod_{t=1}^{T} P(\mathbf{y}_t \mid \mathbf{y}_{t-1}, \ldots, \mathbf{y}_0), \tag{3.1}$$

where each term in the product is the distribution of neural population activity at a single time point given all previous activity (see appendix A for details). More specifically, each of the conditional distributions in equation 3.1 is gaussian, and for the LDS, we can parameterize these distributions as

$$P(\mathbf{x}_t|\mathbf{y}_{t-1}, \ldots, \mathbf{y}_0) := \mathcal{N}(\hat{\mathbf{x}}_t, \mathbf{V}_t), \tag{3.2}$$

$$P(\mathbf{y}_t|\mathbf{y}_{t-1}, \ldots, \mathbf{y}_0) = \mathcal{N}(\mathbf{C}\hat{\mathbf{x}}_t, \mathbf{C}\mathbf{V}_t\mathbf{C}^\top + \mathbf{R}), \tag{3.3}$$

where $\hat{\mathbf{x}}_t$ is the mean of the conditional distribution over the latent at time step t, given observations until time step $t-1$. It obeys the recurrence equation

$$\hat{\mathbf{x}}_{t+1} = \mathbf{A}(\hat{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{y}_t - \mathbf{C}\hat{\mathbf{x}}_t)), \tag{3.4}$$

where $\mathbf{K}_t$ is the Kalman gain given by

$$\mathbf{K}_t = \mathbf{V}_t\mathbf{C}^\top(\mathbf{C}\mathbf{V}_t\mathbf{C}^\top + \mathbf{R})^{-1}, \tag{3.5}$$

and $\mathbf{V}_t$ represents a covariance matrix, which is independent of the observations and follows a recurrence equation detailed in appendix A.

Iterating equation 3.4 over multiple time steps, one can see that $\hat{\mathbf{x}}_{t+1}$ depends not only on the last observation $\mathbf{y}_t$ but on the full history of observations $(\mathbf{y}_0, \ldots, \mathbf{y}_t)$, which therefore affects the distribution at any given time step. The process $(\mathbf{y}_0, \ldots, \mathbf{y}_t)$ generated by the LDS model is hence non-Markovian.

Conversely, for the linear RNN, the observations $(\mathbf{y}_0, \ldots, \mathbf{y}_t)$ instead *do* form a Markov process, meaning that observations are conditionally independent of their history given the activity from the previous time step:

$$P(\mathbf{y}_t \mid \mathbf{y}_{t-1}, \ldots, \mathbf{y}_0) = P(\mathbf{y}_t \mid \mathbf{y}_{t-1}). \tag{3.6}$$

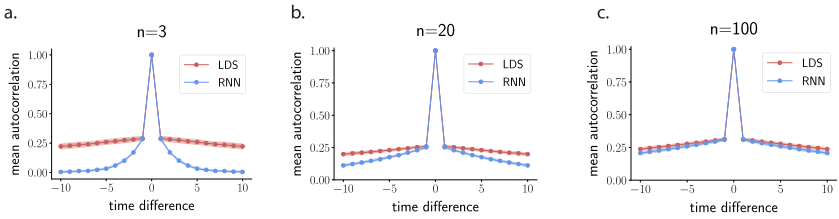

Figure 2: Mean autocorrelation of observations $\mathbf{y}_t$ from latent LDS processes compared with their first-order RNN approximations. The latent space is one-dimensional ($d = 1$), and the dimension n of the observation space is increased from left to right: (a) $n = 3$, (b) $n = 20$, (c) $n = 100$. The parameters of the latent state processes are fixed scalars ($\mathbf{A} = (0.97)$, $\mathbf{Q} = (0.1)$), while the elements of the observation matrices $\mathbf{C}$ are drawn randomly and independently from a centered gaussian distribution of variance 1. The observation noise has covariance $\mathbf{R} = \sigma_v^2 \mathbf{I}_n$ with $\sigma_v^2 = 2$. Note that we have chosen observation noise to largely dominate over latent state noise in order to obtain a large difference between models at low n. Dots and shaded areas indicate, respectively, mean and standard deviation of different estimations of the mean autocorrelation done on 10 independent folds of 100 trials each (where $\mathbf{C}$ was identical across trials).

The fact that this property does not in general hold for the latent LDS shows that the two model classes are not equivalent. Due to this fundamental constraint, the RNN can only approximate the complex distribution (see equation 3.1) parameterized by an LDS, as detailed in the following section and illustrated in Figure 2.

3.2 Matching the First-Order Marginals of an LDS Model. We can obtain a Markovian approximation of the LDS-generated sequence of observations $(\mathbf{y}_0, \ldots, \mathbf{y}_t)$ by deriving the conditional distribution $P(\mathbf{y}_{t+1} \mid \mathbf{y}_t)$ under the LDS model and matching it with a low-rank RNN (Pachitariu et al., 2013). This type of first-order approximation will preserve exactly the one-time-step-difference marginal distributions $P(\mathbf{y}_{t+1}, \mathbf{y}_t)$ although structure across longer timescales might not be captured correctly.

First, we note that we can express both $\mathbf{y}_t$ and $\mathbf{y}_{t+1}$ as noisy linear projections of $\mathbf{x}_t$:

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t, \tag{3.7}$$

$$\mathbf{y}_{t+1} = \mathbf{C}(\mathbf{A}\mathbf{x}_t + \mathbf{w}_t) + \mathbf{v}_{t+1}, \tag{3.8}$$

which follows from equation 2.1.

Let $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t)$ denote the gaussian marginal distribution over the latent vector $\mathbf{x}_t$ at time t. Then we can use standard identities for linear

transformations of gaussian variables to derive the joint distribution over $\mathbf{y}_t$ and $\mathbf{y}_{t+1}$:

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t+1} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{C}\boldsymbol{\Sigma}_t\mathbf{C}^\top + \mathbf{R} & \mathbf{C}\boldsymbol{\Sigma}_t\mathbf{A}^\top\mathbf{C}^\top \\ \mathbf{C}\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{C}^\top & \mathbf{C}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^\top + \mathbf{Q})\mathbf{C}^\top + \mathbf{R} \end{bmatrix} \right). \quad (3.9)$$

We can then apply the formula for conditioning of multivariate gaussians (see Bishop, 2006, equations 2.81 and 2.82) to obtain

$$\mathbf{y}_{t+1} \mid \mathbf{y}_t \sim \mathcal{N}\left(\mathbf{J}_t\mathbf{y}_t, \mathbf{P}_t\right), \quad (3.10)$$

where

$$\mathbf{J}_t = \mathbf{C}\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{C}^\top(\mathbf{C}\boldsymbol{\Sigma}_t\mathbf{C}^\top + \mathbf{R})^{-1}, \quad (3.11)$$

$$\mathbf{P}_t = \mathbf{C}(\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^\top + \mathbf{Q})\mathbf{C}^\top + \mathbf{R} - \mathbf{C}\mathbf{A}\boldsymbol{\Sigma}_t\mathbf{C}^\top(\mathbf{C}\boldsymbol{\Sigma}_t\mathbf{C}^\top + \mathbf{R})^{-1}\mathbf{C}\boldsymbol{\Sigma}_t\mathbf{A}^\top\mathbf{C}^\top. \quad (3.12)$$

In contrast, from equation 2.4, for a low-rank RNN, the first-order marginal is given by

$$\mathbf{y}_{t+1} \mid \mathbf{y}_t \sim \mathcal{N}\left(\mathbf{J}\mathbf{y}_t, \mathbf{P}\right). \quad (3.13)$$

Comparing equations 3.10 and 3.13, we see for the LDS model that the effective weights $\mathbf{J}_t$ and the covariance $\mathbf{P}_t$ depend on time through $\boldsymbol{\Sigma}_t$, the marginal covariance of the latent at time t, while for the RNN, they do not. Note, however, that $\boldsymbol{\Sigma}_t$ follows the recurrence relation

$$\boldsymbol{\Sigma}_{t+1} = \mathbf{A}\boldsymbol{\Sigma}_t\mathbf{A}^\top + \mathbf{Q}, \quad (3.14)$$

which converges toward a fixed point $\boldsymbol{\Sigma}_\infty$ that obeys the discrete Lyapunov equation,

$$\boldsymbol{\Sigma}_\infty = \mathbf{A}\boldsymbol{\Sigma}_\infty\mathbf{A}^\top + \mathbf{Q}, \quad (3.15)$$

provided all eigenvalues of $\mathbf{A}$ have absolute value less than 1.

The LDS can therefore be approximated by an RNN with constant weights when the initial covariance $\boldsymbol{\Sigma}_0$ is equal to the asymptotic covariance $\boldsymbol{\Sigma}_\infty$, as noted previously (Pachitariu et al., 2013). Even if this condition does not hold at time 0, $\boldsymbol{\Sigma}_\infty$ will in general be a good approximation of the latent covariance after an initial transient. In this case, we obtain the fixed recurrence weights

$$\mathbf{J} = \mathbf{C}\mathbf{A}\boldsymbol{\Sigma}_\infty\mathbf{C}^\top(\mathbf{C}\boldsymbol{\Sigma}_\infty\mathbf{C}^\top + \mathbf{R})^{-1} := \mathbf{M}\mathbf{N}^\top, \quad (3.16)$$

where we define $\mathbf{M} = \mathbf{C}$, which has shape $n \times d$, and $\mathbf{N}^\top = \mathbf{A}\boldsymbol{\Sigma}_\infty \mathbf{C}^\top (\mathbf{C}\boldsymbol{\Sigma}_\infty \mathbf{C}^\top + \mathbf{R})^{-1}$, which has shape $d \times n$, so that $\mathbf{J}$ is a rank r matrix with $r = d$.

3.3 Cases of Equivalence between LDS and RNN Models. Although latent LDS and low-rank linear RNN models are not equivalent in general, we can show that the first-order Markovian approximation introduced above becomes exact in two limit cases of interest: for observation noise orthogonal to the latent subspace and in the limit $n \gg d$, with coefficients of the observation matrix generated randomly and independently.

Our key observation is that if $\mathbf{K}_t \mathbf{C} = \mathbf{I}$ in equation 3.4 with $\mathbf{I}$ the identity matrix, we have $\hat{\mathbf{x}}_{t+1} = \mathbf{A}\mathbf{K}_t \mathbf{y}_t$, so that the dependence on the observations before time step t disappears and the LDS therefore becomes Markovian. Interestingly, this condition $\mathbf{K}_t \mathbf{C} = \mathbf{I}$ also implies that the latent state can be inferred from the current observation $\mathbf{y}_t$ alone (see equation A.7 in appendix A) and that this inference is exact, since the variance of the distribution $p(\mathbf{x}_t | \mathbf{y}_t)$ is then equal to 0 as seen from equation A.8. We next examine two cases where this condition is satisfied.

We first consider the situation where the observation noise vanishes: $\mathbf{R} = \mathbf{0}$. Then, as shown in appendix A, the Kalman gain is $\mathbf{K}_t = (\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{C}^\top$, so that $\mathbf{K}_t \mathbf{C} = \mathbf{I}$. In that case, the approximation of the LDS by the RNN defined in section 3.2 is exact, with equations 3.11 and 3.12 becoming:

$$\mathbf{J} = \mathbf{C}\mathbf{A}(\mathbf{C}^\top \mathbf{C})^{-1} \mathbf{C}^\top, \tag{3.17}$$

$$\mathbf{P} = \mathbf{C}\mathbf{Q}\mathbf{C}^\top. \tag{3.18}$$

More generally, this result remains valid when the observation noise is orthogonal to the latent subspace spanned by the columns of the observation matrix $\mathbf{C}$ (in which case the recurrence noise given by equation 3.18 becomes $\mathbf{P} = \mathbf{C}\mathbf{Q}\mathbf{C}^\top + \mathbf{R}$).

A second case in which we can obtain $\mathbf{K}_t \mathbf{C} \approx \mathbf{I}$ is in the limit of many neurons, $n \gg d$, assuming that coefficients of the observation matrix are generated randomly and independently. Indeed, under these hypotheses, the Kalman gain given by equation 3.5 is dominated by the term $\mathbf{C}\mathbf{V}_t \mathbf{C}^\top$, so that the observation covariance $\mathbf{R}$ becomes negligible, as shown formally in appendix B. Intuitively this means that the information about the latent state $\hat{\mathbf{x}}_t$ is distributed over a large enough population of neurons for the Kalman filter to average out the observation noise and estimate it optimally without making use of previous observations. Ultimately this makes the LDS asymptotically Markovian in the case where we have an arbitrarily large neural population relative to the number of latent dimensions.

To illustrate the convergence of the low-rank RNN approximation to the target latent LDS in the large n limit, in Figure 2, we consider a simple example with a one-dimensional latent space and observation spaces of

increasing dimensionality. To visualize the difference between the LDS and its low-rank RNN approximation, we plot the trace of the autocorrelation matrix of observations $\mathbf{y}_t$ in the stationary regime, $\rho(\delta) = \mathrm{Tr}(\mathbb{E}[\mathbf{y}_t \mathbf{y}_{t+\delta}^T])$. Since the RNNs are constructed to capture the marginal distributions of observations separated by at most one time step, the two curves match exactly for a lag $\delta \in \{-1, 0, 1\}$, but dependencies at longer timescales cannot be accurately captured by an RNN due to its Markov property (see Figure 2a). However, these differences vanish as the dimensionality of the observation space becomes much larger than that of the latent space (see Figures 2b and 2c), which illustrates that the latent LDS converges to a process equivalent to a low-rank RNN.

4 Mapping Low-Rank Linear RNNs onto Latent LDS Models

We now turn to the reverse question: Under what conditions can a low-rank linear RNN be expressed as a latent LDS model? We start with an intuitive mapping for the deterministic case (when noise covariance $\mathbf{P} = \mathbf{0}$) and then extend it to a more general mapping valid in the presence of noise.

We first consider a deterministic linear low-rank RNN obeying

$$\mathbf{y}_{t+1} = \mathbf{M}\mathbf{N}^\top \mathbf{y}_t, \tag{4.1}$$

Since $\mathbf{M}$ is an $n \times r$ matrix, it is immediately apparent that for all t, $\mathbf{y}_t$ is confined to a linear subspace of dimension r, spanned by the columns of $\mathbf{M}$. Hence, we can define the r-dimensional latent state as

$$\mathbf{x}_t = \mathbf{M}^+ \mathbf{y}_t, \tag{4.2}$$

where $\mathbf{M}^+$ is the pseudoinverse of $\mathbf{M}$ defined as $\mathbf{M}^+ = (\mathbf{M}^\top \mathbf{M})^{-1} \mathbf{M}^\top$ (well defined since $\mathbf{M}$ is of rank r), so that we retrieve $\mathbf{y}_t$ as

$$\mathbf{y}_t = \mathbf{M}\mathbf{x}_t. \tag{4.3}$$

We then obtain a recurrence equation for the latent state:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{M}^+ \mathbf{y}_{t+1} \\ &= \mathbf{M}^+ \mathbf{M}\mathbf{N}^\top \mathbf{y}_t \\ &= \mathbf{N}^\top \mathbf{M}\mathbf{x}_t \\ &:= \mathbf{A}\mathbf{x}_t, \end{aligned} \tag{4.4}$$

which with $\mathbf{A} = \mathbf{N}^\top \mathbf{M}$ describes the dynamics of a latent LDS with $d = r$. A key insight from equation 4.4 is that the overlap between the columns of $\mathbf{N}$ and $\mathbf{M}$ determines the part of the activity that is integrated by the recurrent

dynamics (Mastrogiuseppe & Ostojic, 2018; Schuessler et al., 2020; Beiran et al., 2021; Dubreuil et al., 2022).

In presence of noise $\boldsymbol{\epsilon}_t$, $\mathbf{y}_t$ is no longer confined to the column space of $\mathbf{M}$. Part of this noise is integrated into the recurrent dynamics and can contribute to the activity accross many time steps. This integration of noise can occur in an LDS at the level of latent dynamics through $\mathbf{w}_t$, but not at the level of observation noise $\mathbf{v}_t$, which is independent accross time steps. As noted above, recurrent dynamics only integrate the activity present in the column space of $\mathbf{N}$. In the presence of noise, this part of state space therefore needs to be included into the latent variables. More important, a similar observation can be made about external inputs when they are added to the RNN dynamics (see appendix D).

A full mapping from a noisy low-rank RNN to an LDS model can therefore be built by extending the latent space to the linear subspace $\mathcal{F}$ of $\mathbb{R}^n$ spanned by the columns of $\mathbf{M}$ and $\mathbf{N}$ (see appendix C), which has dimension d with $r \leq d \leq 2r$. Let $\mathbf{C}$ be a matrix whose columns form an orthogonal basis for this subspace (which can be obtained via the Gram-Schmidt algorithm). In that case, we can define the latent vector as

$$\mathbf{x}_t = \mathbf{C}^\top \mathbf{y}_t, \tag{4.5}$$

and the latent dynamics are given by

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}_t, \tag{4.6}$$

where the recurrence matrix is $\mathbf{A} = \mathbf{C}^\top \mathbf{J} \mathbf{C}$, and the latent dynamics noise is $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ with $\mathbf{Q} = \mathbf{C}^\top \mathbf{P} \mathbf{C}$. Introducing $\mathbf{v}_t = \mathbf{y}_t - \mathbf{C}\mathbf{x}_t$, under a specific condition on the noise covariance $\mathbf{P}$, we obtain a normal random variable independent of the other sources of noise in the process (appendix C), so that $\mathbf{y}_t$ can be described as a noisy observation of the latent state $\mathbf{x}_t$ as in the LDS model:

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t. \tag{4.7}$$

4.1 Subsampled RNNs. Experimental recordings typically access only the activity of a small fraction of neurons in the local network. An important question for interpreting neural data concerns the statistics of activity when only a random subset of k neurons in an RNN is observed. This situation can be formalized by introducing the set of observed activities $\mathbf{o}_t$:

$$\begin{aligned} \mathbf{y}_{t+1} &= \mathbf{J}\mathbf{y}_t + \boldsymbol{\epsilon}_t, \\ \mathbf{o}_t &= \mathbf{y}_t[:k] = \mathbf{D}\mathbf{y}_t. \end{aligned} \tag{4.8}$$

Here $[:k]$ symbolizes the selection of the first k values of a vector and $\mathbf{D}$ is the corresponding projection matrix on the subspace spanned by the first k

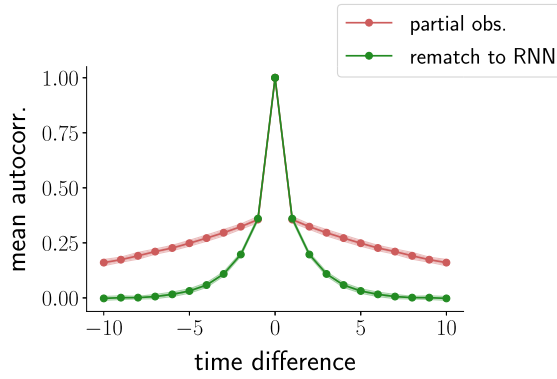

Figure 3: Mean autocorrelation of k neurons subsampled from an n-dimesional rank-one RNN, compared with a k-dimensional RNN built to match the first-order marginals of partial observations. Formally, we first built an LDS equivalent to the partial observations as in equation 4.9, and then the corresponding RNN as in section 3.2. The rank-one RNN contains $n = 20$ neurons, of which only $k = 3$ are observed. The mismatch occurs because the long-term correlations present in the partial observations are caused by the larger size of the original RNN with 20 neurons and cannot be reproduced by an RNN with only 3 neurons.

neurons. The system described by equation 4.8 is exactly an LDS but with latent state $\mathbf{y}_t$ and observations $\mathbf{o}_t$. In contrast to the regime considered in the previous sections, the latents have a higher dimensionality than observations. However, assuming as before that $\mathbf{J}$ is low-rank, this model can be mapped onto an equivalent LDS following the steps in appendix C:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t + \mathbf{w}_t, \\ \mathbf{o}_t &= \mathbf{D}\mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{v}_t. \end{aligned} \tag{4.9}$$

This LDS is equivalent to equation 4.8, but with latent dynamics $\mathbf{x}_t$ of dimension $r < d < 2r$ where r is the rank of $\mathbf{J}$. The dynamics of the latent state $\mathbf{x}_t$ are identical to those of the fully observed low-rank RNN (see equation 4.6), but the observations are generated from a subsampled observation matrix $\mathbf{DC}$.

Interestingly, this mapping highlights the fact that the activity statistics of the k subsampled neurons are in general not Markovian, in contrast to the full activity $\mathbf{y}_t$ of the n neurons in the underlying RNN. In particular, for that reason, the statistics of $\mathbf{o}_t$ cannot be exactly reproduced by a smaller RNN consisting of k units (see Figure 3). Remarkably, when considering the subsampled activity of an RNN, a latent LDS is therefore a more accurate model than a smaller RNN containing only the observed units.

5 Discussion

In this letter, we have examined the relationship between two simple yet powerful classes of models of low-dimensional activity: latent linear dynamical systems (LDS) and low-rank linear recurrent neural networks (RNN). We have focused on these tractable linear models with additive gaussian noise to highlight their mathematical similarities and differences. Although both models induce a jointly gaussian distribution over neural population activity, generic latent LDS models can exhibit long-range, non-Markovian temporal dependencies that cannot be captured by low-rank linear RNNs, which describe neural population activity with a first-order Markov process. Conversely, we showed that generic low-rank linear RNNs can be captured by an equivalent latent LDS model. However, we have shown that the two classes of models are effectively equivalent in limit cases of practical interest for neuroscience, in particular when the number of sampled neurons is much higher than the latent dimensionality.

Although these two model classes can generate similar sets of neural trajectories, different approaches are typically used for fitting them to neural data: parameters of LDS models are in general inferred by variants of the expectation-maximization algorithm (Yu et al., 2005; Pachitariu et al., 2013; Nonnenmacher, Turaga, & Macke, 2017; Durstewitz, 2017), which include the Kalman smoothing equations (Roweis & Ghahramani, 1999), while RNNs are often fitted with variants of linear regression (Rajan et al., 2016; Eliasmith & Anderson, 2003; Pollock & Jazayeri, 2020; Bondanelli et al., 2021) or backpropagation through time (Dubreuil et al., 2022). The relationship uncovered here therefore opens the door to comparing different fitting approaches more directly, and in particular to developing probabilistic methods for inferring RNN parameters from data.

We have considered here only linear RNN and latent LDS models. Nonlinear low-rank RNNs without noise can be directly reduced to nonlinear latent dynamics with linear observations following the same mapping as in section 4 (Mastrogiuseppe & Ostojic, 2018; Schuessler et al., 2020; Beiran et al., 2021; Dubreuil et al., 2022) and therefore define a natural class of nonlinear LDS models. A variety of other nonlinear generalizations of LDS models have been considered in the literature. One line of work has examined linear latent dynamics with a nonlinear observation model (Yu et al., 2005) or nonlinear latent dynamics (Yu et al., 2005; Durstewitz, 2017; Duncker et al., 2019; Pandarinath et al., 2018; Kim et al., 2008). Another line of work has focused on switching LDS models (Linderman et al., 2017; Glaser et al., 2020) for which the system undergoes different linear dynamics depending on a hidden discrete state, thus combining elements of latent LDS and hidden Markov models. Both nonlinear low-rank RNNs and switching LDS models are universal approximators of low-dimensional dynamical systems (Funahashi & Nakamura, 1993; Chow & Li, 2000; Beiran et al., 2021). Relating switching LDS models to local linear approximations

of nonlinear low-rank RNNs (Beiran et al., 2021; Dubreuil et al., 2022) is therefore an interesting avenue for future investigations.

Appendix A: Kalman Filtering Equations

We reproduce in this appendix the recurrence equations followed by the conditional distributions in equation 3.1 for both the latent LDS and the linear RNN models.

For the latent LDS model, the conditional distributions are gaussians, and their form is given by the Kalman filter equations (Kalman, 1960; Yu et al., 2004; Welling, 2010). Following Yu et al. (2004), we observe that for any two time steps $\tau \leq t$ the conditional distributions $P(\mathbf{y}_{t+1}|\mathbf{y}_\tau, \ldots, \mathbf{y}_0)$ and $P(\mathbf{x}_{t+1}|\mathbf{y}_\tau, \ldots, \mathbf{y}_0)$ are gaussian, and we introduce the notations

$$P(\mathbf{y}_t|\mathbf{y}_\tau, \ldots, \mathbf{y}_0) := \mathcal{N}(\hat{\mathbf{y}}_t^\tau, \mathbf{W}_t^\tau), \tag{A.1}$$

$$P(\mathbf{x}_t|\mathbf{y}_\tau, \ldots, \mathbf{y}_0) := \mathcal{N}(\hat{\mathbf{x}}_t^\tau, \mathbf{V}_t^\tau). \tag{A.2}$$

In particular, we are interested in expressing $\hat{\mathbf{y}}_{t+1}^t$ and $\hat{\mathbf{x}}_{t+1}^t$, which are the predicted future observation and latent state, but also in $\hat{\mathbf{x}}_t^t$ which represents the latent state inferred from the history of observations until time step t included. To lighten notations, in the main text, we remove the exponent when it has one time step difference with the index, by writing $\hat{\mathbf{x}}_{t+1}$, $\hat{\mathbf{y}}_{t+1}$, $\mathbf{W}_{t+1}$ and $\mathbf{V}_{t+1}$ instead of, respectively, $\hat{\mathbf{x}}_{t+1}^t$, $\hat{\mathbf{y}}_{t+1}^t$, $\mathbf{W}_{t+1}^t$ and $\mathbf{V}_{t+1}^t$.

First, note that we have the natural relationships

$$\hat{\mathbf{x}}_{t+1}^t = \mathbf{A}\hat{\mathbf{x}}_t^t, \tag{A.3}$$

$$\hat{\mathbf{y}}_{t+1}^t = \mathbf{C}\hat{\mathbf{x}}_{t+1}^t, \tag{A.4}$$

$$\mathbf{V}_{t+1}^t = \mathbf{A}\mathbf{V}_t^t\mathbf{A}^\top + \mathbf{Q}, \tag{A.5}$$

$$\mathbf{W}_{t+1}^t = \mathbf{C}\mathbf{V}_{t+1}^t\mathbf{C}^\top + \mathbf{R}, \tag{A.6}$$

so that it is sufficient to find expressions for $\hat{\mathbf{x}}_t^t$ and $\mathbf{V}_t^t$. After calculations detailed in Yu et al. (2004) or Welling (2010), we obtain

$$\hat{\mathbf{x}}_t^t = \hat{\mathbf{x}}_t^{t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{C}\hat{\mathbf{x}}_t^{t-1}), \tag{A.7}$$

$$\mathbf{V}_t^t = (\mathbf{I} - \mathbf{K}_t\mathbf{C})\mathbf{V}_t^{t-1}, \tag{A.8}$$

where $\mathbf{K}_t$ is the Kalman gain given by

$$\mathbf{K}_t = \mathbf{V}_t^{t-1}\mathbf{C}^\top(\mathbf{C}\mathbf{V}_t^{t-1}\mathbf{C}^\top + \mathbf{R})^{-1}. \tag{A.9}$$

These equations form a closed recurrent system, as can be seen by combining equations A.3 and A.7 and equations A.5 and A.8 to obtain a self-consistent set of recurrence equations for the predicted latent state and

its variance:

$$\hat{\mathbf{x}}_{t+1}^{t} = \mathbf{A}(\hat{\mathbf{x}}_{t}^{t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{C}\hat{\mathbf{x}}_{t}^{t-1})), \tag{A.10}$$

$$\begin{aligned}\mathbf{V}_{t+1}^{t} &= \mathbf{A}(\mathbf{I} - \mathbf{K}_t\mathbf{C})\mathbf{V}_{t}^{t-1}\mathbf{A}^{\top} + \mathbf{Q}, \\ &= \mathbf{A}(\mathbf{I} - \mathbf{V}_{t}^{t-1}\mathbf{C}^{\top}(\mathbf{C}\mathbf{V}_{t}^{t-1}\mathbf{C}^{\top} + \mathbf{R})^{-1}\mathbf{C})\mathbf{V}_{t}^{t-1}\mathbf{A}^{\top} + \mathbf{Q}.\end{aligned} \tag{A.11}$$

From equation A.10, we see that the predicted state at time $t+1$, and thus the predicted observation, depends on observations at time steps $\tau \leq t-1$ through the term $\hat{\mathbf{x}}_t$, making the system non-Markovian. Also note that equations for the variances do not involve any of the observations $\mathbf{y}_t$, showing these are exact values and not estimations.

This derivation, however, is not valid in the limit case $\mathbf{R} = \mathbf{0}$, since $\mathbf{K}_t$ is then undefined. In that case, however, we can observe that $\mathbf{y}_t$ lies in the linear subspace spanned by the columns of $\mathbf{C}$, so that one can simply replace equation A.7 by

$$\hat{\mathbf{x}}_{t}^{t} = \mathbf{C}^{+}\mathbf{y}_t = \mathbf{x}_t, \tag{A.12}$$

where $\mathbf{C}^{+} = (\mathbf{C}^{\top}\mathbf{C})^{-1}\mathbf{C}^{\top}$ is the pseudoinverse of $\mathbf{C}$. Since this equation is deterministic, the variance of the estimated latent state is equal to $\mathbf{0}$, so that equation A.8 becomes $\mathbf{V}_{t}^{t} = \mathbf{0}$. This case can be encompassed by equations A.3 to A.8 if we rewrite the Kalman gain as

$$\mathbf{K}_t = \mathbf{C}^{+} = (\mathbf{C}^{\top}\mathbf{C})^{-1}\mathbf{C}^{\top}. \tag{A.13}$$

Finally, for the linear RNN, the conditional distribution of equation A.1 is directly given by

$$P(\mathbf{y}_{t+1}|\mathbf{y}_t, \ldots, \mathbf{y}_0) = \mathcal{N}(\mathbf{J}\mathbf{y}_t, \mathbf{P}), \tag{A.14}$$

which shows that the predicted observation depends only on the last one, making the system Markovian.

Appendix B: Equivalence in the Large Network Limit

Here we make the assumption that the coefficients of the observation matrix are generated randomly and independently. We show that in the limit of large n with d fixed, one obtains $\mathbf{K}_t\mathbf{C} \to \mathbf{I}$ so that the LDS is asymptotically Markovian and can therefore be exactly mapped to an RNN.

We start by considering a latent LDS whose conditional distributions obey equations A.1 to A.11, with the Kalman gain obeying equation A.9. To simplify equation A.9, we focus on the steady state where variance $\mathbf{V}_t$ has reached its stationary limit $\mathbf{V}$ in equation A.11.

Without loss of generality, we reparameterize the LDS by applying a change of basis to the latent states such that $\mathbf{V} = \mathbf{I}$. We also apply a change of basis to the observation space such that $\mathbf{R} = \mathbf{I}$ in the new basis (this transformation does not have an impact on the conditional dependencies between the $\mathbf{y}_t$ at different time steps, and it can also be shown that it cancels out in the expression $\mathbf{K}_t\mathbf{C}$). Equation A.9 then becomes

$$\mathbf{K}_t\mathbf{C} = \mathbf{C}^\top(\mathbf{I} + \mathbf{C}\mathbf{C}^\top)^{-1}\mathbf{C}. \tag{B.1}$$

Applying the matrix inversion lemma gives $(\mathbf{I} + \mathbf{C}\mathbf{C}^\top)^{-1} = \mathbf{I} - \mathbf{C}(\mathbf{I} + \mathbf{C}^\top\mathbf{C})^{-1}\mathbf{C}^\top$, from which we get

$$\mathbf{K}_t\mathbf{C} = \mathbf{C}^\top\mathbf{C} - \mathbf{C}^\top\mathbf{C}(\mathbf{I} + \mathbf{C}^\top\mathbf{C})^{-1}\mathbf{C}^\top\mathbf{C}.$$

Using a Taylor expansion, we then write

$$\begin{aligned}(\mathbf{I} + \mathbf{C}^\top\mathbf{C})^{-1} &= (\mathbf{I} + (\mathbf{C}^\top\mathbf{C})^{-1})^{-1}(\mathbf{C}^\top\mathbf{C})^{-1} \\ &= \left(\sum_{k=0}^{\infty}(-(\mathbf{C}^\top\mathbf{C})^{-1})^k\right)(\mathbf{C}^\top\mathbf{C})^{-1} \\ &\approx (\mathbf{C}^\top\mathbf{C})^{-1} - ((\mathbf{C}^\top\mathbf{C})^{-1})^2 + ((\mathbf{C}^\top\mathbf{C})^{-1})^3,\end{aligned}$$

which gives

$$\begin{aligned}\mathbf{K}_t\mathbf{C} &\approx \mathbf{C}^\top\mathbf{C} - \mathbf{C}^\top\mathbf{C}(\mathbf{C}^\top\mathbf{C})^{-1}\mathbf{C}^\top\mathbf{C} + \mathbf{C}^\top\mathbf{C}((\mathbf{C}^\top\mathbf{C})^{-1})^2\mathbf{C}^\top\mathbf{C} \\ &\quad - \mathbf{C}^\top\mathbf{C}((\mathbf{C}^\top\mathbf{C})^{-1})^3\mathbf{C}^\top\mathbf{C} \\ &\approx \mathbf{C}^\top\mathbf{C} - \mathbf{C}^\top\mathbf{C} + \mathbf{I} - (\mathbf{C}^\top\mathbf{C})^{-1}.\end{aligned}$$

Assuming the coefficients of the observation matrix are independent and identically distributed (i.i.d.) with zero mean and unit variance, for n large, we obtain $\mathbf{C}^\top\mathbf{C} = n\mathbf{I} + \mathcal{O}(\sqrt{n})$ from the central limit theorem so that $(\mathbf{C}^\top\mathbf{C})^{-1} = \mathcal{O}(1/n)$ (which can again be proven with a Taylor expansion). This finally leads to $\mathbf{K}_t\mathbf{C} = \mathbf{I} + \mathcal{O}(1/n)$.

An alternative proof takes advantage of the spectral theorem applied to $\mathbf{C}^\top\mathbf{C}$. Indeed, since it is a symmetric matrix, it can be decomposed as $\mathbf{C}^\top\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ where $\mathbf{U}$ is an orthonormal matrix and $\mathbf{D}$ the diagonal matrix of eigenvalues. Starting from equation B.1 we derive

$$\begin{aligned}\mathbf{K}_t\mathbf{C} &= \mathbf{C}^\top\mathbf{C} - \mathbf{C}^\top\mathbf{C}(\mathbf{I} + \mathbf{C}^\top\mathbf{C})^{-1}\mathbf{C}^\top\mathbf{C} \\ &= \mathbf{U}\mathbf{D}\mathbf{U}^\top - \mathbf{U}\mathbf{D}\mathbf{U}^\top(\mathbf{I} + \mathbf{U}\mathbf{D}\mathbf{U}^\top)^{-1}\mathbf{U}\mathbf{D}\mathbf{U}^\top \\ &= \mathbf{U}\mathbf{D}\mathbf{U}^\top - \mathbf{U}\mathbf{D}\mathbf{U}^\top(\mathbf{U}(\mathbf{D} + \mathbf{I})\mathbf{U}^\top)^{-1}\mathbf{U}\mathbf{D}\mathbf{U}^\top \\ &= \mathbf{U}\mathbf{D}\mathbf{U}^\top - \mathbf{U}\mathbf{D}\mathbf{U}^\top\mathbf{U}(\mathbf{D} + \mathbf{I})^{-1}\mathbf{U}^\top\mathbf{U}\mathbf{D}\mathbf{U}^\top\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{U}\mathbf{D}\mathbf{U}^\top - \mathbf{U}\mathbf{D}^2(\mathbf{D}+\mathbf{I})^{-1}\mathbf{U}^\top \\
&= \mathbf{U}(\mathbf{D} - \mathbf{I}/(\mathbf{D}+\mathbf{I}))\mathbf{U}^\top \\
&= \mathbf{U}(\mathbf{D}/(\mathbf{D}+\mathbf{I}))\mathbf{U}^\top \\
&= \mathbf{U}(\mathbf{I} - \mathbf{I}/(\mathbf{D}+\mathbf{I}))\mathbf{U}^\top \\
&= \mathbf{I} - \mathbf{U}(\mathbf{I}/(\mathbf{D}+\mathbf{I}))\mathbf{U}^\top.
\end{aligned}
$$

Assuming as before that the coefficients of $\mathbf{C}$ are i.i.d. gaussian with zero mean and unit variance, $\mathbf{C}^\top\mathbf{C}$ is then the empirical covariance of i.i.d. samples of a gaussian ensemble with identity matrix covariance. The matrix $\mathbf{C}^\top\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}^\top$ then follows the $(\mathbf{I}, n)$-Wishart distribution, and for n large, its eigenvalues are all greater than $\sqrt{n}$ (using e.g., the tail bounds of Wainwright, 2019, theorem 6.1). This shows that $(\mathbf{I}/(\mathbf{D}+\mathbf{I})) = \mathcal{O}(1/\sqrt{n})\mathbf{I}$, completing the proof.

Appendix C: Derivation of the RNN to LDS Mapping

As mentioned in section 4, we consider an RNN defined by equation 2.4 with $\mathbf{J} = \mathbf{M}\mathbf{N}^\top$ and note $\mathbf{C}$ an orthonormal matrix whose columns form a basis of $\mathcal{F}$, the linear subspace spanned by the columns of $\mathbf{M}$ and $\mathbf{N}$. Note that $\mathbf{C}\mathbf{C}^\top$ is an orthogonal projector onto the subspace $\mathcal{F}$ and that since all columns of $\mathbf{M}$ and $\mathbf{N}$ belong to this subspace, we have $\mathbf{C}\mathbf{C}^\top\mathbf{M} = \mathbf{M}$ and $\mathbf{C}\mathbf{C}^\top\mathbf{N} = \mathbf{N}$. Hence, we have

$$\mathbf{C}\mathbf{C}^\top\mathbf{J}\mathbf{C}\mathbf{C}^\top = \mathbf{J}. \tag{C.1}$$

We thus define the latent vector as $\mathbf{x}_t = \mathbf{C}^\top\mathbf{y}_t$, and we can then write

$$
\begin{aligned}
\mathbf{x}_{t+1} &= \mathbf{C}^\top\mathbf{y}_{t+1} \\
&= \mathbf{C}^\top\mathbf{J}\mathbf{y}_t + \mathbf{C}^\top\boldsymbol{\epsilon}_t \\
&= \mathbf{C}^\top\mathbf{C}\mathbf{C}^\top\mathbf{J}\mathbf{C}\mathbf{C}^\top\mathbf{y}_t + \mathbf{C}^\top\boldsymbol{\epsilon}_t \quad \text{(by equation C.1)} \\
&= \mathbf{C}^\top\mathbf{J}\mathbf{C}\mathbf{C}^\top\mathbf{y}_t + \mathbf{C}^\top\boldsymbol{\epsilon}_t \quad \text{(because } \mathbf{C}^\top\mathbf{C} = \mathbf{I}\text{)} \\
&= \mathbf{A}\mathbf{x}_t + \mathbf{w}_t,
\end{aligned}
$$

where we have defined the recurrence matrix $\mathbf{A} = \mathbf{C}^\top\mathbf{J}\mathbf{C}$ and the latent dynamics noise $\mathbf{w}_t = \mathbf{C}^\top\boldsymbol{\epsilon}_t$, which follows $\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ with $\mathbf{Q} = \mathbf{C}^\top\mathbf{P}\mathbf{C}$.

Let us define $\mathbf{v}_t = \mathbf{y}_t - \mathbf{C}\mathbf{x}_t = (\mathbf{I} - \mathbf{C}\mathbf{C}^\top)\mathbf{y}_t$. We need to determine the conditions under which $\mathbf{v}_t$ is normally distributed and independent of $\mathbf{y}_{t-1}$ and $\mathbf{x}_t$. For this, we write

$$\mathbf{C}\mathbf{x}_t = \mathbf{C}\mathbf{A}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{w}_{t-1}$$

$$= \mathbf{C}\mathbf{C}^\top \mathbf{J}\mathbf{C}\mathbf{x}_{t-1} + \mathbf{C}\mathbf{w}_{t-1}$$
$$= \mathbf{C}\mathbf{C}^\top \mathbf{J}\mathbf{C}\mathbf{C}^\top \mathbf{y}_{t-1} + \mathbf{C}\mathbf{w}_{t-1}$$
$$= \mathbf{J}\mathbf{y}_{t-1} + \mathbf{C}\mathbf{w}_{t-1},$$

and hence,

$$\mathbf{v}_t = \boldsymbol{\epsilon}_{t-1} - \mathbf{C}\mathbf{w}_{t-1}$$
$$= (\mathbf{I} - \mathbf{C}\mathbf{C}^\top)\boldsymbol{\epsilon}_{t-1},$$

which is independent of $\mathbf{y}_{t-1}$ and has a marginal distribution $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ with $\mathbf{R} = \mathbf{P} - \mathbf{C}\mathbf{C}^\top \mathbf{P}\mathbf{C}\mathbf{C}^\top$ but is not in general independent of $\mathbf{w}_{t-1}$. A sufficient and necessary condition for the independence of $\mathbf{w}_{t-1}$ and $\mathbf{v}_t$ is that the RNN noise covariance $\mathbf{P}$ has all its eigenvectors either aligned with or orthogonal to the subspace $\mathcal{F}$ (in this case, the covariance $\mathbf{R}$ is degenerate and has $\mathcal{F}$ as a null space, which implies that observation noise is completely orthogonal to $\mathcal{F}$). If that is not the case, the reparameterization stays valid up to the fact that the observation noise $\mathbf{v}_t$ and the latent dynamics noise $\mathbf{w}_t$ can be correlated.

Appendix D: Addition of Input Terms

Let us consider an extension of both the latent LDS and the linear RNN models to take into account inputs. More specifically, we consider adding to both model classes an input under the form of a time-varying signal u_t fed to the network through a constant set of input weights. In the latent LDS model, the input is fed directly to the latent variable and equations 2.1 and 2.2 become

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}u_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \tag{D.1}$$
$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t \qquad\quad + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \tag{D.2}$$

The linear RNN equation 2.4 becomes

$$\mathbf{y}_t = \mathbf{J}\mathbf{y}_{t-1} + \mathbf{W}_{in}u_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{P}), \tag{D.3}$$

so that we will represent by $\mathbf{B}$ a low-dimensional input projection and $\mathbf{W}_{in}$ a high-dimensional one.

For the LDS-to-RNN mapping, we can directly adapt the derivations of section 3.2, which lead to

$$\mathbf{y}_{t+1} \mid \mathbf{y}_t \sim \mathcal{N}\left(\mathbf{C}\mathbf{B}u_t + \mathbf{J}_t\mathbf{y}_t, \mathbf{P}_t\right) \tag{D.4}$$

with the same expressions for $\mathbf{J}_t$ and $\mathbf{P}_t$, given in equations 3.11 and 3.12.

For the RNN-to-LDS mapping, assuming again that $\mathbf{J}$ is low-rank and written as $\mathbf{J} = \mathbf{M}\mathbf{N}^\top$, we can define

$$\mathbf{x}_t = \mathbf{C}^\top \mathbf{y}_t,$$

where $\mathbf{C}$ is a matrix whose columns form an orthonormal basis for the subspace $\mathcal{F}$ spanned by the columns of $\mathbf{M}$, $\mathbf{N}$, and $\mathbf{W}_{in}$. This latent vector then follows the dynamics

$$\mathbf{x}_{t+1} = \mathbf{C}\mathbf{J}\mathbf{C}^\top \mathbf{x}_t + \mathbf{C}^\top \mathbf{W}_{in} u_t + \mathbf{C}^\top \boldsymbol{\epsilon}_t, \tag{D.5}$$

which corresponds to equation D.1, and it is straightforward to show that it leads to equation D.2, with the technical condition that the covariance of $\boldsymbol{\epsilon}_t$ should have its eigenvectors aligned with the subspace $\mathcal{F}$ to avoid correlations between observation and recurrent noises.

Acknowledgments

We thank both reviewers for constructive suggestions that have significantly improved this letter. In particular, we thank Scott Linderman for the alternative proof in appendix B. A.V. and S.O. were supported by the program Ecoles Universitaires de Recherche (ANR-17-EURE-0017), the CRCNS program through French Agence Nationale de la Recherche (ANR-19-NEUC-0001-01), and the NIH BRAIN initiative (U01NS122123). J.W.P. was supported by grants from the Simons Collaboration on the Global Brain (SCGB AWD543027), the NIH BRAIN initiative (R01EB026946), and a visiting professorship grant from the Ecole Normale Superieure.

References

Archer, E., Koster, U., Pillow, J., & Macke, J. (2014). Low-dimensional models of neural population activity in sensory cortical circuits. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems, 27* (pp. 343–351). Red Hook, NY: Curran.

Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. *Current Opinion in Neurobiology*, *46*, 1–6. 10.1016/j.conb.2017.06.003, PubMed: 28668365

Beiran, M., Dubreuil, A., Valente, A., Mastrogiuseppe, F., & Ostojic, S. (2021). Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Computation*, *33*, 1572–1615. 10.1162/neco_a_01381, PubMed: 34496384

Bishop, C. (2006). *Pattern recognition and machine learning*. Berlin: Springer.

Bondanelli, G., Deneux, T., Bathellier, B., & Ostojic, S. (2021). Network dynamics underlying OFF responses in the auditory cortex. *eLife*, *10*, e53151. 10.7554/eLife.53151, PubMed: 33759763

Chow, T., & Li, X. (2000). Modeling of continuous time dynamical systems with input by recurrent neural networks. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *47*, 575–578. 10.1109/81.841860

Churchland, M., Byron, M., Sahani, M., & Shenoy, K. (2007). Techniques for extracting single-trial activity patterns from large-scale neural recordings. *Current Opinion in Neurobiology*, *17*, 609–618. 10.1016/j.conb.2007.11.001, PubMed: 18093826

Cohen, Z., DePasquale, B., Aoi, M., & Pillow, J. (2020). *Recurrent dynamics of prefrontal cortex during context-dependent decision-making*. bioRxiv.

Cunningham, J., & Yu, B. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*, 1500–1509. 10.1038/nn.3776, PubMed: 25151264

Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., & Ostojic, S. (2022). The role of population structure in computations through neural dynamics. *Nature Neuroscience*, *25*, 783–794. 10.1038/s41593-022-01088-4, PubMed: 35668174

Duncker, L., Bohner, G., Boussard, J., & Sahani, M. (2019). Learning interpretable continuous-time models of latent stochastic dynamical systems. In *Proceedings of the International Conference on Machine Learning* (pp. 1726–1734).

Durstewitz, D. (2017). A state space approach for piecewise-linear recurrent neural networks for identifying computational dynamics from neural measurements. *PLOS Computational Biology*, *13*, e1005542. 10.1371/journal.pcbi.1005542, PubMed: 28574992

Eliasmith, C., & Anderson, C. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Finkelstein, A., Fontolan, L., Economo, M., Li, N., Romani, S., & Svoboda, K. (2021). Attractor dynamics gate cortical information flow during decision-making. *Nature Neuroscience*, *24*, 843–850. 10.1038/s41593-021-00840-6, PubMed: 33875892

Funahashi, K., & Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Networks*, *6*, 801–806. 10.1016/S0893-6080(05)80125-X

Gallego, J., Perich, M., Miller, L, & Solla, S. (2017). Neural manifolds for the control of movement. *Neuron*, *94*, 978–984. 10.1016/j.neuron.2017.05.025, PubMed: 28595054

Gao, P., & Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, *32*, 148–155. 10.1016/j.conb.2015.04.003, PubMed: 25932978

Glaser, J., Whiteway, M., Cunningham, J., Paninski, L., & Linderman, S. (2020). Recurrent switching dynamical systems models for multiple interacting neural populations. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems*, *33* (pp. 14867–14878). Red Hook, NY: Curran.

Hennequin, G., Vogels, T., & Gerstner, W. (2014). Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron*, *82*, 1394–1406. 10.1016/j.neuron.2014.04.045, PubMed: 24945778

Jazayeri, M., & Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Current Opinion in Neurobiology*, *70*, 113–120. 10.1016/j.conb.2021.08.002, PubMed: 34537579

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*, 35–45. 10.1115/1.3662552

Kao, T., Sadabadi, M., & Hennequin, G. (2021). Optimal anticipatory control as a theory of motor preparation: A thalamo-cortical circuit model. *Neuron*, *109*, 1567–1581. 10.1016/j.neuron.2021.03.009, PubMed: 33789082

Kim, S., Simeral, J., Hochberg, L., Donoghue, J., & Black, M. (2008). Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. *Journal of Neural Engineering*, *5*, 455. 10.1088/1741-2560/5/4/010

Laje, R., & Buonomano, D. (2013). Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nature Neuroscience*, *16*, 925–933. 10.1038/nn.3405, PubMed: 23708144

Landau, I., & Sompolinsky, H. (2018). Coherent chaos in a recurrent neural network with structured connectivity. *PLOS Computational Biology*, *14*, e1006309. 10.1371/journal.pcbi.1006309, PubMed: 30543634

Landau, I., & Sompolinsky, H. (2021). Macroscopic fluctuations emerge in balanced networks with incomplete recurrent alignment. *Phys. Rev. Research.*, *3*, 023171. 10.1103/PhysRevResearch.3.023171

Linderman, S., Johnson, M., Miller, A., Adams, R., Blei, D., & Paninski, L. (2017). Bayesian learning and inference in recurrent switching linear dynamical systems. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (pp. 914–922).

Macke, J., Buesing, L., Cunningham, J., Byron M., Shenoy, K., & Sahani, M. (2011). Empirical models of spiking in neural populations. In S. Solla, T. Leen, & K. R. Müller (Eds.), *Advances in neural information processing systems* (pp. 1350–1358). Cambridge, MA: MIT Press.

Mante, V., Sussillo, D., Shenoy, K., & Newsome, W. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*, 78–84. 10.1038/nature12742, PubMed: 24201281

Mastrogiuseppe, F., & Ostojic, S. (2018). Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, *99*, 609–623. 10.1016/j.neuron.2018.07.003, PubMed: 30057201

Nonnenmacher, M., Turaga, S., & Macke, J. (2017). Extracting low-dimensional dynamics from multiple large-scale neural population recordings by learning to predict correlations. In I. Guyon, Y. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems, 30*. Red Hook, NY: Curran.

Pachitariu, M., Petreska, B., & Sahani, M. (2013), Recurrent linear models of simultaneously-recorded neural populations. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems*, *26* (pp. 3138–3146). Red Hook, NY: Curran.

Pandarinath, C., O'Shea, D., Collins, J., Jozefowicz, R., Stavisky, S., Kao, J., . . . Sussilllo, D. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, *15*, 805–815. 10.1038/s41592-018-0109-9, PubMed: 30224673

Pereira, U., & Brunel, N. (2018). Attractor dynamics in networks with learning rules inferred from in vivo data. *Neuron*, *99*, 227–238. 10.1016/j.neuron.2018.05.038, PubMed: 29909997

Perich, M., Arlt, C., Soares, S., Young, M., Mosher, C., Minxha, J., . . . Rajan, K. (2021). *Inferring brain-wide interactions using data-constrained recurrent neural network models.* bioRxiv:2020-12.

Petreska, B., Byron, M., Cunningham, J., Santhanam, G., Ryu, S., Shenoy, K., & Sahani, M. (2011). Dynamical segmentation of single trials from population neural data. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems, 24* (pp. 756–764). Red Hook, NY: Curran.

Pollock, E., & Jazayeri, M. (2020). Engineering recurrent neural networks from task-relevant manifolds and dynamics. *PLOS Computational Biology*, *16*, e1008128. 10.1371/journal.pcbi.1008128

Rajan, K., Harvey, C., & Tank, D. (2016). Recurrent network models of sequence generation and memory. *Neuron*, *90*, 128–142. 10.1016/j.neuron.2016.02.009, PubMed: 26971945

Roweis, S., & Ghahramani, Z. (1999). A unifying review of linear gaussian models. *Neural Computation*, *11*, 305–345. 10.1162/089976699300016674, PubMed: 9950734

Saxena, S., & Cunningham, J. (2019). Towards the neural population doctrine. *Current Opinion in Neurobiology*, *55*, 103–111. 10.1016/j.conb.2019.02.002, PubMed: 30877963

Schuessler, F., Dubreuil, A., Mastrogiuseppe, F., Ostojic, S., & Barak, O. (2020). Dynamics of random recurrent networks with correlated low-rank structure. *Physical Review Research*, *2*, 013111. 10.1103/PhysRevResearch.2.013111

Semedo, J., Zandvakili, A., Kohn, A., Machens, C., & Byron, M. (2014). Extracting latent structure from multiple interacting neural populations. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems, 27* (pp. 2942–2950). Red Hook, NY: Curran.

Smith, A., & Brown, E. (2003). Estimating a state-space model from point process observations. *Neural Computation*, *15*, 965–991. 10.1162/089976603765202622, PubMed: 12803953

Sompolinsky, H., Crisanti, A., & Sommers, H. (1988). Chaos in random neural networks *Phys. Rev. Lett.*, *61*, 259–262. 10.1103/PhysRevLett.61.259, PubMed: 10039285

Sussillo, D. (2014). Neural circuits as computational dynamical systems. *Current Opinion in Neurobiology*, *25*, 156–163. 10.1016/j.conb.2014.01.008, PubMed: 24509098

Wainwright, M. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge: Cambridge University Press.

Welling, M. (2010). *The Kalman filter* (Caltech lecture note 136-93).

Yu, B., Afshar, A., Santhanam, G., Ryu, S., Shenoy, K., & Sahani, M. (2005). Extracting dynamical structure embedded in neural activity. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in neural information processing systems, 18*. Cambridge, MA: MIT Press.

Yu, B., Shenoy, K., & Sahani, M. (2004). *Derivation of Kalman filtering and smoothing equations*. Stanford, CA: Stanford University.

Zoltowski, D., Pillow, J., & Linderman, S. (2020). A general recurrent state space framework for modeling neural dynamics during decision-making. In *Proceedings of the International Conference on Machine Learning* (pp. 11680–11691).

Received October 26, 2021; accepted April 15, 2022.