Disentangling Recurrent and Input-Driven Contributions to Low-Dimensional Neural Dynamics

Akshay Gautam, Angus Chadwick
School of Informatics, University of Edinburgh

Institute for Machine Learning

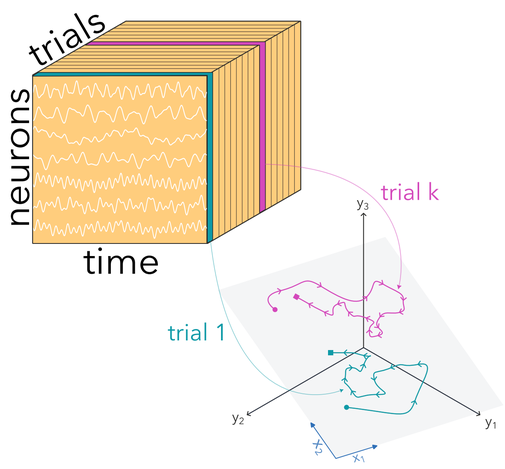
Motivation

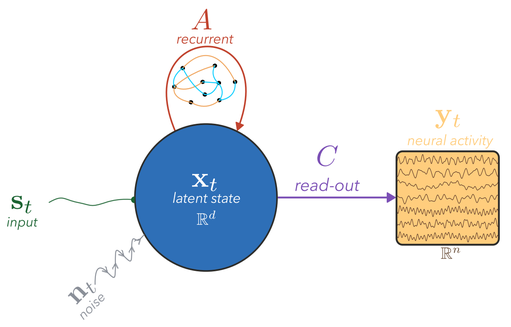
Neural activity $\mathbf{y}_t \in \mathbb{R}^n$ is recorded simultaneously across multiple neurons, over time, and across multiple trials. This activity is dominated by low-dimensional dynamics which are driven by **recurrent connectivity, external inputs, and noise**. The relative contribution of these factors is **unknown and difficult to disentangle in neural data**.

We begin by modelling the neural activity as a latent linear dynamical system:

$$\mathbf{x}_t^{(k)} = A\,\mathbf{x}_{t-1}^{(k)} + \mathbf{s}_t^{(k)} + \mathbf{n}_t^{(k)}, \quad \mathbf{n}_t \sim \mathcal{N}(0, \Sigma_n)$$
$$\mathbf{y}_t^{(k)} = C\,\mathbf{x}_t^{(k)} + \mathbf{d} + \mathbf{e}_t^{(k)}, \qquad \mathbf{e}_t \sim \mathcal{N}(0, \Sigma_e)$$

where $\mathbf{x}_t \in \mathbb{R}^d$ is the latent population state, with $d \ll n$; $\mathbf{s}_t \in \mathbb{R}^d$ is the input acting *within the latent space*.





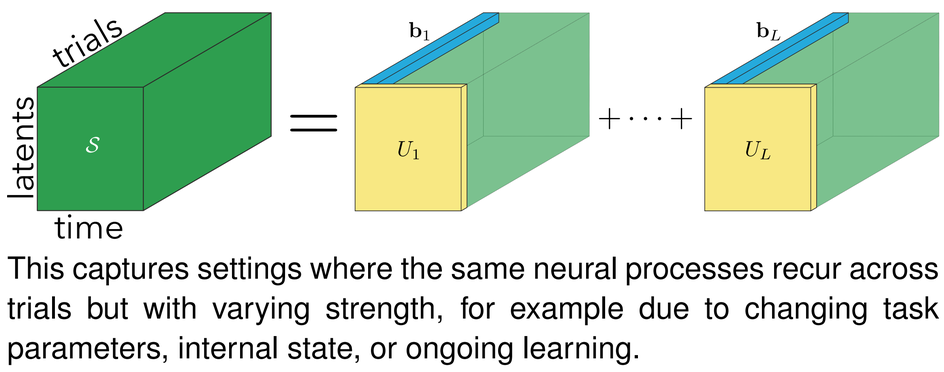
Generally, A, $\mathbf{s}_t$, and n_t are *not jointly identifiable* from the observed data $\mathbf{y}_t$ alone, *without further assumptions*. Trivially, any trajectory can be explained by the inputs alone (set $\mathbf{s}_t = \mathbf{x}_t$; $A, \mathbf{n}_t = 0$).

We ask under what conditions the recurrent dynamics, inputs, and noise are jointly identifiable.

Trial mode

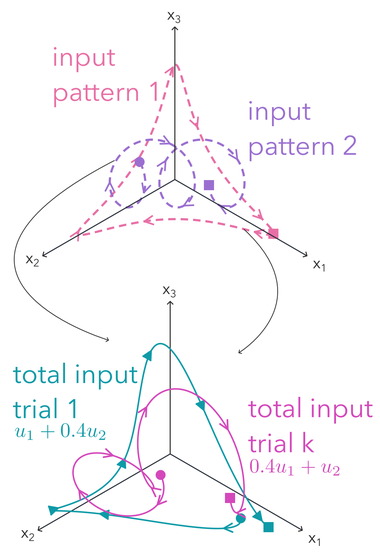
Decomposing the input across the trial dimension gives a small set of L input patterns that drive every trial, re-weighted trial by trial, based on the trial specific weights $\mathbf{b}^{(k)}$:

$$\mathbf{s}_t^{(k)} = U_t\,\mathbf{b}^{(k)} \iff \mathcal{S} = \sum_{l=1}^{L} U_l \otimes \mathbf{b}_l, \qquad \mathcal{U} \in \mathbb{R}^{d\times T\times L}, \quad B \in \mathbb{R}^{L\times K}.$$



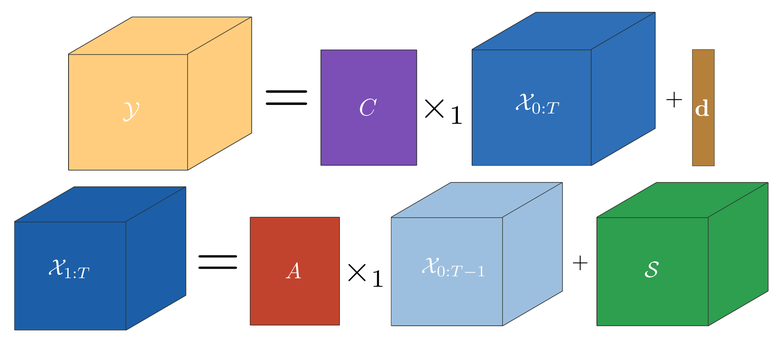
This captures settings where the same neural processes recur across trials but with varying strength, for example due to changing task parameters, internal state, or ongoing learning.

Related decompositions have appeared before (Khan et al., 2018), where d-dimensional inputs are locked to experimental conditions: in effect a one-hot $\mathbf{b}^{(k)}$ that assigns each trial to a single condition. Here $\mathbf{b}^{(k)}$ is freely learnt, so the same templates can be shared and re-weighted both within and across conditions, giving a more general account of trial-to-trial variability.



Low rank inputs and identifiability

Viewing the linear dynamical system in tensor form, we explore whether **constraining the input tensor $\mathcal{S}$ to be low rank improves identifiability.**

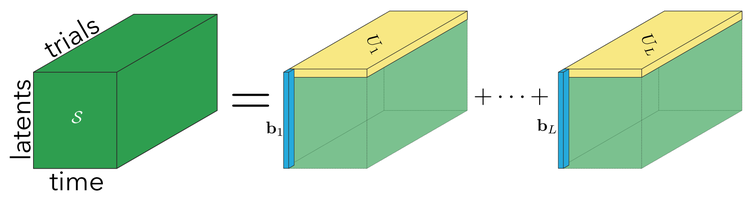


We find that it does for **trial mode**, and only *partially* for neuron mode.

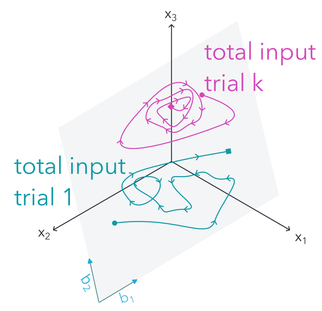
Neuron mode

Decomposing the input across the latent dimension constrains the inputs to lie in a fixed low-dimensional subspace but allows them to freely vary across trials:

$$\mathbf{s}_t^{(k)} = B\,\mathbf{u}_t^{(k)} \iff \mathcal{S} = \sum_l \mathbf{b}_l \otimes U_l, \qquad B \in \mathbb{R}^{d\times L}, \quad \mathcal{U} \in \mathbb{R}^{L\times T\times K}.$$



This is the standard model of low-dimensional inputs used in neuroscience and control theory (Soldado-Magraner et al., 2024, Vyas et al., 2020), here viewed as a tensor decomposition. Typically the input patterns $\mathbf{u}_t$ are taken as known and only the input subspace is inferred, or there is a penalty on input strength; we instead infer B and $\mathbf{u}_t$ jointly, without constraints, making identifiability the central question.



Model fitting

The model is fit with a standard **EM algorithm** for linear dynamical systems (Ghahramani & Hinton, 1996).

- **E-step**: the standard Kalman filter and RTS smoother give the posterior over the latent states $\mathbf{x}_t^{(k)}$ given the observations $\mathbf{y}_t^{(k)}$ and the current parameters $\Theta = \{A, C, \mathbf{d}, \Sigma_n, \Sigma_e, B, \mathcal{U}\}$.
- **M-step**: standard closed-form updates for $A, C, \mathbf{d}, \Sigma_n, \Sigma_e$. The input factors admit an *exact* update, given by a rank-L SVD of the dynamics residuals $(\mathcal{U}, B) = \mathrm{SVD}_L(\mathcal{X}_+ - A\,\mathcal{X}_-)$.

The model order (d, L) is chosen by **nested cross-validation**, with outer folds over trials and inner folds over neurons.

Are the decompositions identifiable?

Neuron mode — degenerate. The input $\mathbf{u}_t^{(k)}$ is free at every trial and time, so it can absorb any noise lying in $\mathrm{span}(B)$:

$$B\mathbf{u}_t^{(k)} + \mathbf{n}_t^{(k)} = \underbrace{B\big(\mathbf{u}_t^{(k)} + B^{+}\mathbf{n}_t^{(k)}\big)}_{B\tilde{\mathbf{u}}_t^{(k)}} + \underbrace{(I - BB^{+})\,\mathbf{n}_t^{(k)}}_{\tilde{\mathbf{n}}_t^{(k)}}$$

shrinking the residual noise covariance onto the complement of $\mathrm{span}(B)$:

$$\Sigma_n \mapsto (I - BB^{+})\,\Sigma_n\,(I - BB^{+})^{\top}, \quad \mathrm{rank} \le d - L.$$

$\tilde{\Sigma}_n$ is singular. Without a prior on the input the model *prefers* this, as it raises the likelihood: **input and noise are confounded** in $\mathrm{span}(B)$.

This in turn corrupts the dynamics. Any added recurrence in $\mathrm{span}(B)$ can be reabsorbed,

$$(A + BM)\,\mathbf{x}_{t-1}^{(k)} + B\big(\mathbf{u}_t^{(k)} - M\mathbf{x}_{t-1}^{(k)}\big) = A\,\mathbf{x}_{t-1}^{(k)} + B\mathbf{u}_t^{(k)},$$

and with the noise in $\mathrm{span}(B)$ largely absorbed, little remains to pin A there. **Recurrence into $\mathrm{span}(B)$ is poorly identified.**

Trial mode — identifiable. The templates U_t are shared across all K trials and the weights $\mathbf{b}^{(k)}$ are constant in time, so noise can only be absorbed by one correction common to every trial:

$$U_t\mathbf{b}^{(k)} + \mathbf{n}_t^{(k)} = \underbrace{\big(U_t + N_tB^{+}\big)\mathbf{b}^{(k)}}_{\tilde{U}_t\,\mathbf{b}^{(k)}} + \underbrace{\mathbf{n}_t^{(k)} - N_tB^{+}\mathbf{b}^{(k)}}_{\tilde{\mathbf{n}}_t^{(k)}}$$

where $N_t = [\,\mathbf{n}_t^{(1)} \cdots \mathbf{n}_t^{(K)}]$. This shrinks but does not collapse the noise covariance:

$$\Sigma_n \;\mapsto\; \left(1 - \tfrac{L}{K}\right)\Sigma_n, \quad \text{full rank.}$$

Generally, as $K \gg L$ the effect vanishes: **input and noise separate** and **the dynamics A are recovered**.

Confirmed on synthetic data: from a known ground-truth model, trial mode recovers the true model order, dynamics and inputs. Neuron mode fails to recover model order; with (d, L) fixed to ground-truth, it fits the data with distorted dynamics and noisier inputs.

Real data: motor cortex

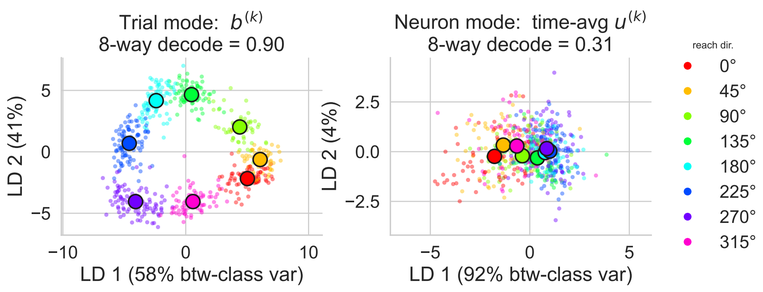
We fit both decompositions to monkey M1/PMd population activity during a centre-out reaching task (Perich et al., 2018; $n \approx 245$, $K \approx 618$, $T \approx 16$) and asked whether the inferred per-trial input separates by reach direction.

8-way reach-direction decoding:

Trial mode	$\mathbf{b}^{(k)}$	**0.90**
Neuron mode	time-averaged $\mathbf{u}^{(k)}$	0.31
Neuron mode	time-resolved $\mathbf{u}^{(k)}$	0.48

Trial mode's per-trial weight $\mathbf{b}^{(k)}$ separates cleanly by reach direction; **neuron mode**'s per-trial time-averaged input collapses to the centre.

Per-trial input loadings in reach-direction discriminant space



Conclusions

- Recurrence, inputs, and noise are **not jointly identifiable** in a latent LDS without constraints on the input.
- Constraining the input to be **low rank** resolves this, but *how* it is decomposed matters:
  - **Trial mode** (templates shared across trials) makes inputs, noise, and dynamics **identifiable**: the input cannot absorb noise or mimic recurrence.
  - **Neuron mode** (the standard fixed-subspace input) remains **degenerate**: the input confounds with noise and recurrence within $\mathrm{span}(B)$.
- On motor cortex (Perich et al.), only trial mode recovers a per-trial input that **separates by reach direction** (0.90 vs 0.31), confirming the analysis on real data.
- **Takeaway:** dynamics, inputs, and noise *can* be jointly recovered when the inputs are **shared across trials** (fixed templates, re-weighted per trial). It is this structure, not low rank alone, that breaks the degeneracy.

agautam2@ed.ac.uk