Neuroscience out of control

Control-theoretic methods for understanding neural circuits

Guillaume Hennequin

g.hennequin@eng.cam.ac.uk

Lisbon, March 2019

UNIVERSITY OF CAMBRIDGE

wellcome

“ [...] that means teaching them to think about multi-dimensional data, that means they actually should all know linear algebra [...] ”

" [. . .] that means teaching them to think about multi-dimensional data, that means they actually should all know linear algebra [. . .] "

matrices ——— **?** ———→ dynamical system

$\mathbf{A}, \mathbf{B}, \mathbf{C}$

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

“ [...] that means teaching them to think about multi-dimensional data, that means they actually should all know linear algebra [...] ”

matrices ——— ? ———→ dynamical system

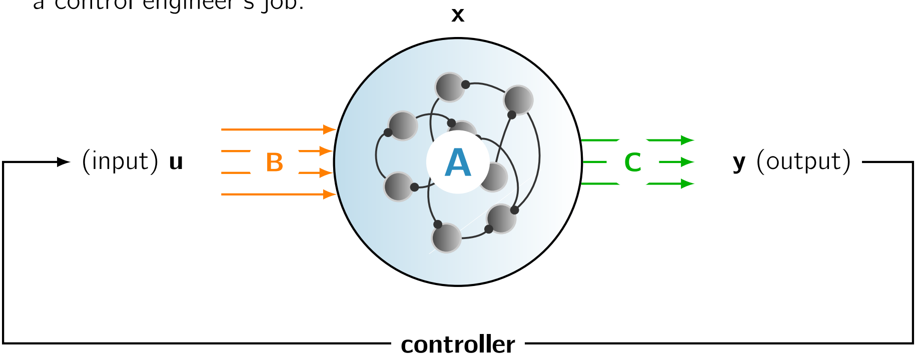
$\mathbf{A}, \mathbf{B}, \mathbf{C}$

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

many decades of insights from control theory, too rarely used in neuro 😞

a control engineer's job:



matrices —— ? ——→ dynamical system

$\mathbf{A}, \mathbf{B}, \mathbf{C}$

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

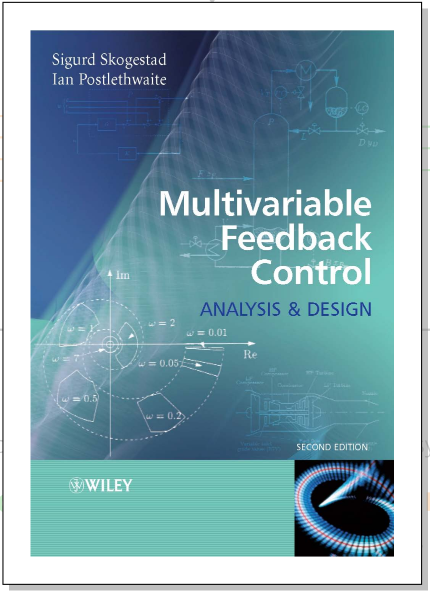
$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

many decades of insights from control theory, too rarely used in neuro 😞

a control engineer's job:

(input) **u**

y (output)



matric … ystem

A, **B**, … **Bu**(t)

$= \mathbf{Cx}$

many decades of insights from control theory, too rarely used in neuro

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory

VOLUME 92, NUMBER 14 PHYSICAL REVIEW LETTERS week ending 9 APRIL 2004

Short-Term Memory in Orthogonal Neural Networks

Olivia L. White,[1] Daniel D. Lee,[2] and Haim Sompolinsky[1,3]
[1]*Harvard University, Cambridge, Massachusetts 02138, USA*
[2]*University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[3]*Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel*
(Received 11 November 2003; published 9 April 2004)

We study the ability of linear recurrent networks obeying discrete time dynamics to store long temporal sequences that are retrievable from the instantaneous state of the network. We calculate this temporal memory capacity for both distributed shift register and random orthogonal connectivity matrices. We show that the memory capacity of these networks scales with system size.

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory

VOLUME 92, NUMBER 14 PHYSICAL REVIEW LETTERS week ending 9 APRIL 2004

[3]Racah Insti

We stu
temporal
temporal
matrices.

PNAS

Memory traces in dynamical systems

Surya Ganguli[a,b,1], Dongsung Huh[c], and Haim Sompolinsky[d,e]

[a]Sloan-Swartz Center for Theoretical Neurobiology, University of California, San Francisco, CA 94143; [b]Cente
University, New York, NY 10032; [c]Computational Neurobiology Program, University of California at San Dieg
for Neural Computation, The Hebrew University, Jerusalem 91904, Israel; and [e]Center for Brain Science, Harv

Communicated by David W. McLaughlin, New York University, New York, NY, October 3, 2008 (received for

To perform nontrivial, real-time computations on a sensory input stream, biological systems must retain a short-term memory trace

these questions in a m
mation to construct a m

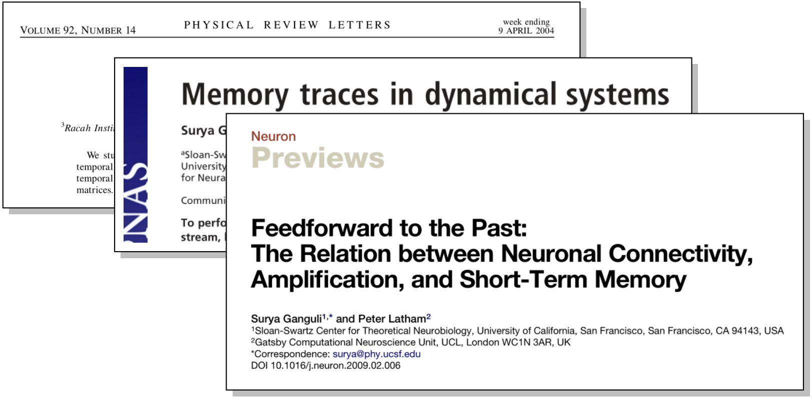
Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory



VOLUME 92, NUMBER 14 PHYSICAL REVIEW LETTERS week ending 9 APRIL 2004

[3]Racah Insti

We stu
temporal
temporal
matrices.

Memory traces in dynamical systems

Surya G

[a]Sloan-Sw
University
for Neura

Communi

To perfo
stream,

Neuron
Previews

Feedforward to the Past:
The Relation between Neuronal Connectivity, Amplification, and Short-Term Memory

Surya Ganguli[1,*] and Peter Latham[2]
[1]Sloan-Swartz Center for Theoretical Neurobiology, University of California, San Francisco, San Francisco, CA 94143, USA
[2]Gatsby Computational Neuroscience Unit, UCL, London WC1N 3AR, UK
*Correspondence: surya@phy.ucsf.edu
DOI 10.1016/j.neuron.2009.02.006

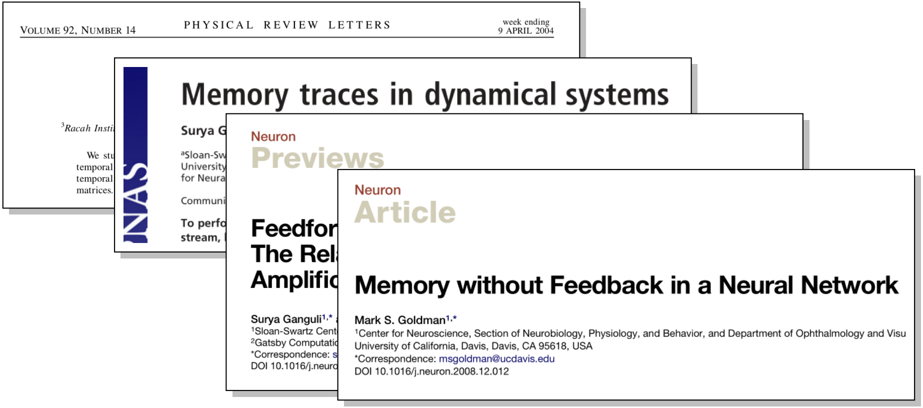
Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory


VOLUME 92, NUMBER 14 PHYSICAL REVIEW LETTERS week ending 9 APRIL 2004

[3]Racah Insti

We stu
temporal
temporal
matrices.

PNAS

Memory traces in dynamical systems

Surya G

[a]Sloan-Sw
University
for Neura

Communi

To perfo
stream,

Neuron
Previews

Feedfor
The Rel
Amplifi

Surya Ganguli[1,*]
[1]Sloan-Swartz Cent
[2]Gatsby Computati
*Correspondence: s
DOI 10.1016/j.neuro

Neuron
Article

Memory without Feedback in a Neural Network

Mark S. Goldman[1,*]
[1]Center for Neuroscience, Section of Neurobiology, Physiology, and Behavior, and Department of Ophthalmology and Visu
University of California, Davis, Davis, CA 95618, USA
*Correspondence: msgoldman@ucdavis.edu
DOI 10.1016/j.neuron.2008.12.012

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory **decision-making**

Neuron

Viewpoint

One-Dimensional Dynamics of Attention and Decision Making in LIP

Surya Ganguli,[1,7,*] James W. Bisley,[2] Jamie D. Roitman,[3] Michael N. Shadlen,[4] Michael E. Goldberg,[5,6] and Kenneth D. Miller[5,7]
[1]Sloan-Swartz Center for Theoretical Neurobiology, University of California, San Francisco, San Francisco, CA 94143, USA
[2]Department of Neurobiology, University of California, Los Angeles, Los Angeles, CA 90025, USA

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory   decision-making   **V1**

Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns

Brendan K. Murphy[1,2]
[1]Graduate Group in Bioph
[2]Center for Theoretical Ne
College of Physicians and
*Correspondence: ken@n
DOI 10.1016/j.neuron.200

Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression

Hirofumi Ozeki,[1] Ian M. Finn,[1] Evan S. Schaffer,[2] Kenneth D. Miller,[2,3,*] and David Ferster[1,3,*]
[1]Department of Neurobiology and Physiology, Northwestern University, Evanston, IL 60208, USA
[2]Center for Theoretical Neuroscience and Department of Neuroscience, Columbia University, College of Physicians and
New York, NY 10032, USA
[3]These authors contributed equally to this work
*Correspondence: ken@neurotheory.columbia.edu (K.D.M.), ferster@northwestern.edu (D.F.)
DOI 10.1016/j.neuron.2009.03.028

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory **decision-making** **V1** **M1**

Optimal Control of Transient Dynamics in Balanced Networks Supports Generation of Complex Movements

Guillaume Hennequin,[1,2,*] Tim P. Vogels,[1,3,4] and Wulfram Gerstner[1,4]
[1]School of Computer and
Lausanne (EPFL), 1015 La
[2]Department of Engineerir
[3]Centre for Neural Circuits
[4]Co-senior author
*Correspondence: gjeh2@
http://dx.doi.org/10.1016/

A neural network that finds a naturalistic solution for the production of muscle activity

David Sussillo[1], Mark M Churchland[2], Matthew T Kaufman[1,4] & Krishna V Shenoy[1,3]

It remains an open question how neural responses in motor cortex relate to movement. We explored the hypothesis that motor cortex reflects dynamics appropriate for generating temporally patterned outgoing commands. To formalize this hypothesis, we trained recurrent neural networks to reproduce the muscle activity of reaching monkeys. Models had to infer dynamics that co transform simple inputs into temporally and spatially complex patterns of muscle activity. Analysis of trained models revealed that the natural dynamical solution was a low-dimensional oscillator that generated the necessary multiphasic commands.

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory decision-making V1 M1 **correlations**

PHYSICAL REVIEW E VOLUME 50, NUMBER 4 OCTOBER 1994

Theory of correlations in stochastic neural networks

OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

How Structure Determines Correlations in Neuronal Networks

Volker Pernice[1,2*]

1 Bernstein Center Freiburg

Insights from a Simple Expression for Linear Fisher Information in a Recurrently Connected Population of Spiking Neurons

Jeffrey Beck

The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability

Guillaume Hennequin,[1,10,*] Yashar Ahmadian,[2,3,4,5,9] Daniel B. Rubin,[2,6,9] Máté Lengyel,[1,7,8] and Kenneth D. Miller[2,3,8]

[1]Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory **decision-making** **V1** **M1** **correlations**

deep learning (!)

Exact solutions to the nonlinear dynamics of learning in deep linear neural networks

Andrew M
Depart

James L. McCl
D

Surya Gan
De
Stanford U

High-dimensional dynamics of generalization error in neural networks

Madhu S. Advani* MADVANI@FAS.HARVARD.EDU
Center for Brain Science
Harvard University
Cambridge, MA 02138

Andrew M. Saxe* ASAXE@FAS.HARVARD.EDU
Center for Brain Science
Harvard University
Cambridge, MA 02138

Linear systems in neuroscience

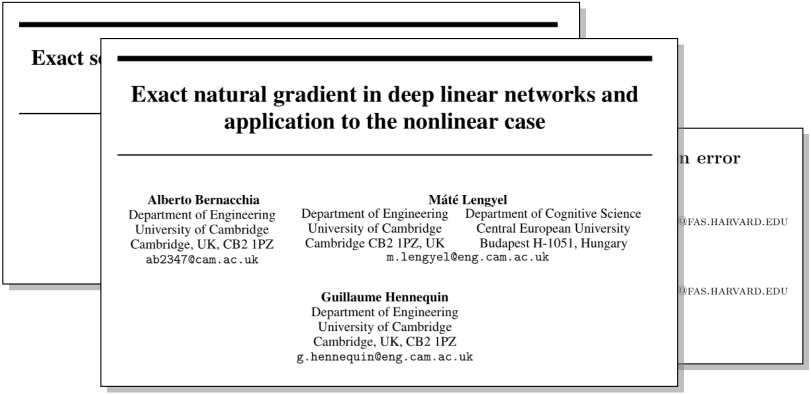
population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory **decision-making** **V1** **M1** **correlations**

deep learning (!)


Exact s

Exact natural gradient in deep linear networks and application to the nonlinear case

Alberto Bernacchia
Department of Engineering
University of Cambridge
Cambridge, UK, CB2 1PZ
ab2347@cam.ac.uk

Máté Lengyel
Department of Engineering
University of Cambridge
Cambridge CB2 1PZ, UK
Department of Cognitive Science
Central European University
Budapest H-1051, Hungary
m.lengyel@eng.cam.ac.uk

Guillaume Hennequin
Department of Engineering
University of Cambridge
Cambridge, UK, CB2 1PZ
g.hennequin@eng.cam.ac.uk

n error

@FAS.HARVARD.EDU

@FAS.HARVARD.EDU

Linear systems in neuroscience

population firing rate vector, or latent variables

$$\tau \frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

short-term memory **decision-making** **V1** **M1** **correlations**

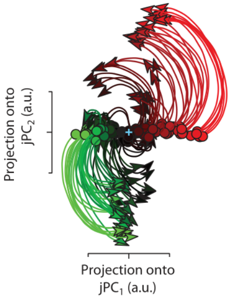
deep learning **data analysis / system identification**

ARTICLE

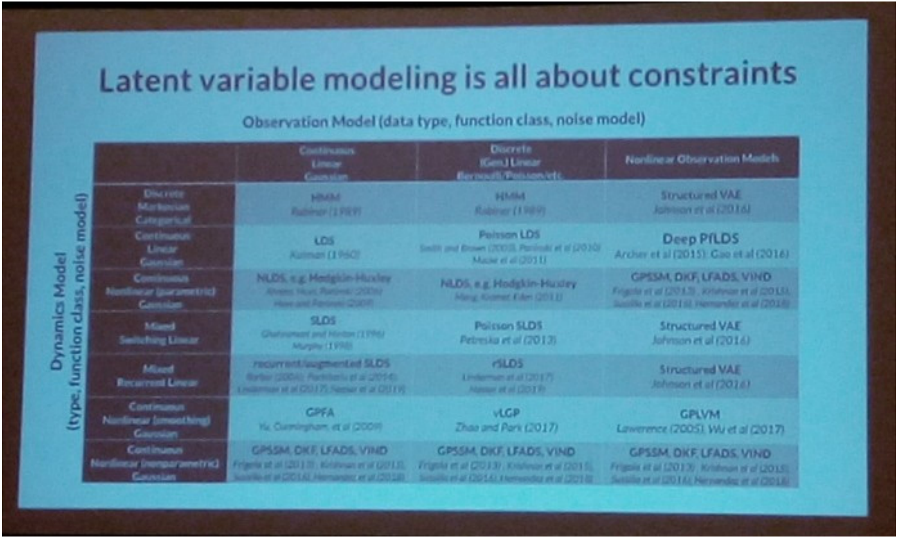
doi:10.1038/nature11129

Neural population dynamics during reaching

Mark M. Churchland[1,2,3]*, John P. Cunningham[4,5]*, Matthew T. Kaufman[2,3], Justin D. Foster[2], Paul Nuyujukian[6,7], Stephen I. Ryu[2,8] & Krishna V. Shenoy[2,3,6,9]



Linear systems in neuroscience



[Scott Linderman, this morning]

Analysis of linear dynamics

population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)

Analysis of linear dynamics

population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]

Analysis of linear dynamics

population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for “normal” architectures [Carandini & Ringach ’97; Dayan & Abbott ’01; Ganguli et al. ’08]
- Schur decomposition – useful description of “nonnormal” architectures, but non-unique [Murphy & Miller, ’09, Goldman ’09, Hennequin et al. ’12]

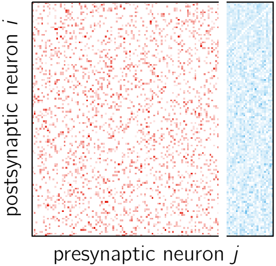
Analysis of linear dynamics

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

population firing rate vector, or latent variables

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]



[Hennequin et al., *Neuron* (2014)]

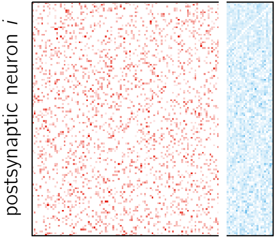
Analysis of linear dynamics

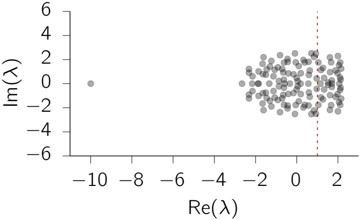
$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

population firing rate vector, or latent variables

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]





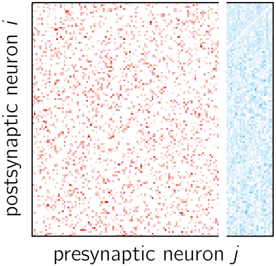
[Hennequin et al., *Neuron* (2014)]

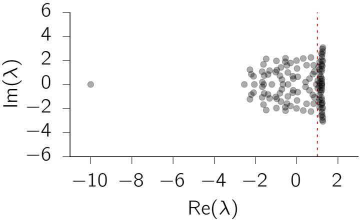
Analysis of linear dynamics

population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \underbrace{\mathbf{W}}_{\text{connectivity matrix}}\mathbf{x} + \text{external input}$$

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]





[Hennequin et al., *Neuron* (2014)]

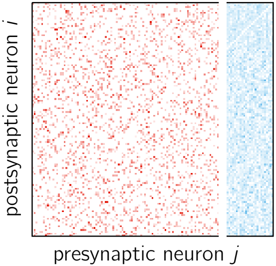
Analysis of linear dynamics

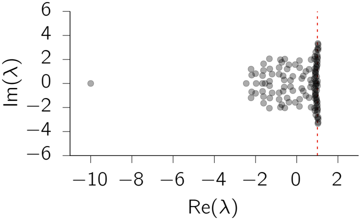
population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]





[Hennequin et al., *Neuron* (2014)]

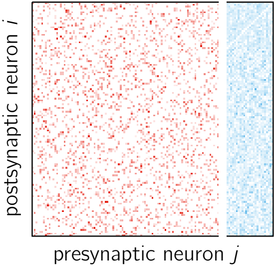
Analysis of linear dynamics

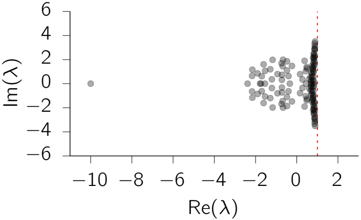
population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for “normal” architectures [Carandini & Ringach ’97; Dayan & Abbott ’01; Ganguli et al. ’08]
- Schur decomposition – useful description of “nonnormal” architectures, but non-unique [Murphy & Miller, ’09, Goldman ’09, Hennequin et al. ’12]





[Hennequin et al., *Neuron* (2014)]

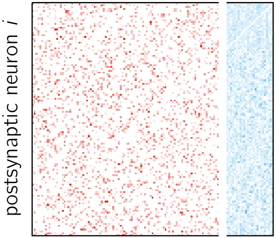
Analysis of linear dynamics

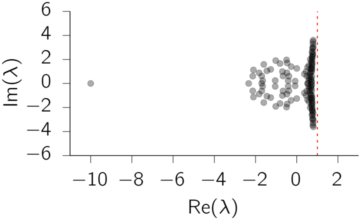
population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]





[Hennequin et al., *Neuron* (2014)]

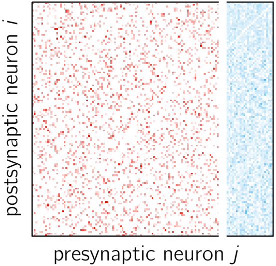
Analysis of linear dynamics

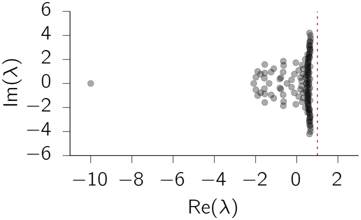
population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]





[Hennequin et al., *Neuron* (2014)]

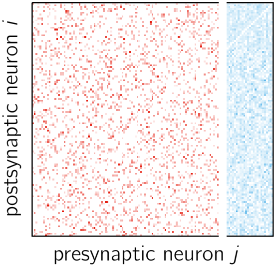
Analysis of linear dynamics

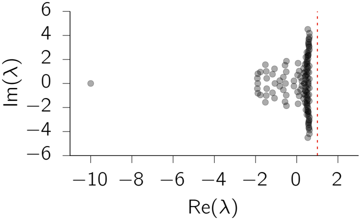
population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]





[Hennequin et al., *Neuron* (2014)]

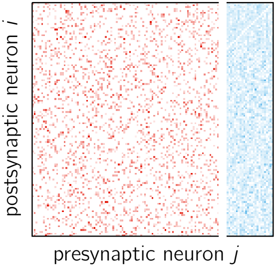
Analysis of linear dynamics

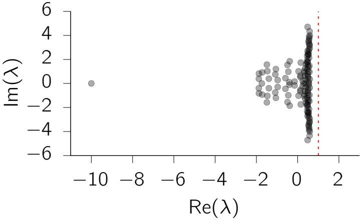
population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]





[Hennequin et al., *Neuron* (2014)]

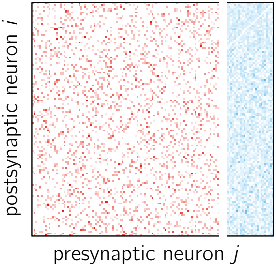
Analysis of linear dynamics

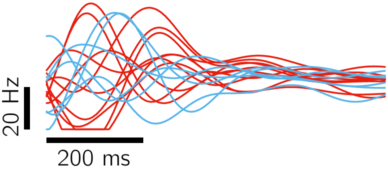
population firing rate vector, or latent variables

$$\mathbf{T}\frac{d\mathbf{x}}{dt} = -\mathbf{x} + \mathbf{W}\mathbf{x} + \text{external input}$$

connectivity matrix

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]



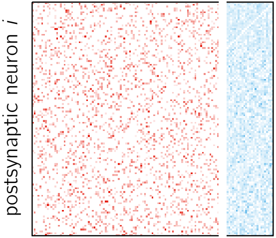


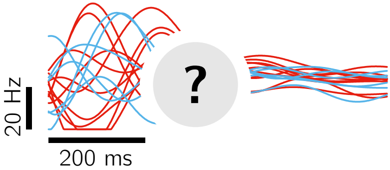
[Hennequin et al., *Neuron* (2014)]

Analysis of linear dynamics

$$\mathbf{T}\frac{d\overset{\text{population firing rate vector, or latent variables}}{\mathbf{x}}}{dt} = -\mathbf{x} + \underset{\text{connectivity matrix}}{\mathbf{W}}\mathbf{x} + \text{external input}$$

- (dynamics analytically solvable)
- eigendecomposition – probably only useful for "normal" architectures [Carandini & Ringach '97; Dayan & Abbott '01; Ganguli et al. '08]
- Schur decomposition – useful description of "nonnormal" architectures, but non-unique [Murphy & Miller, '09, Goldman '09, Hennequin et al. '12]

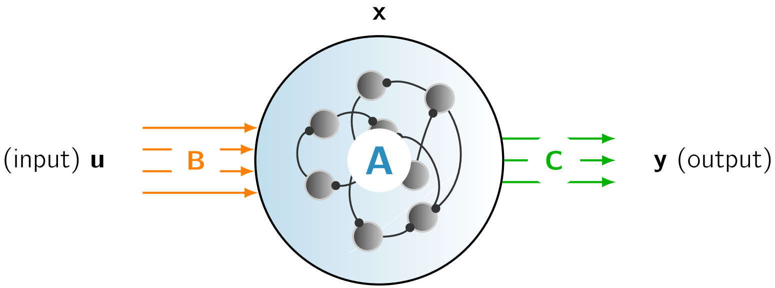




[Hennequin et al., *Neuron* (2014)]

Controllability & Observability

(two-sided "extension of PCA" to *dynamical systems*)

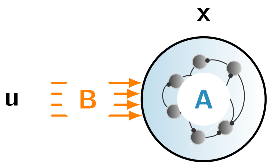


$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$$

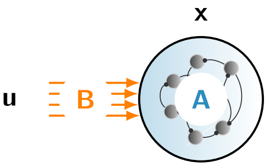
$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

controllability & observability are core concepts in control-theory that afford useful re-interpretations for the analysis of circuits

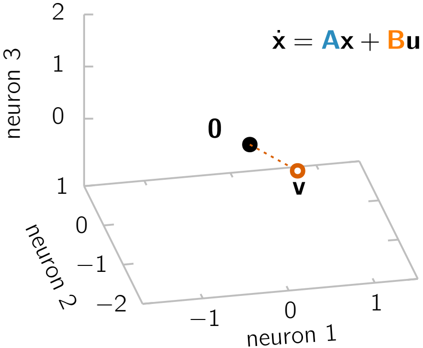
Controllability & Observability



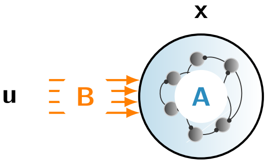
Controllability & Observability



using these input channels,
can I steer **x** from **0** to any **v**?



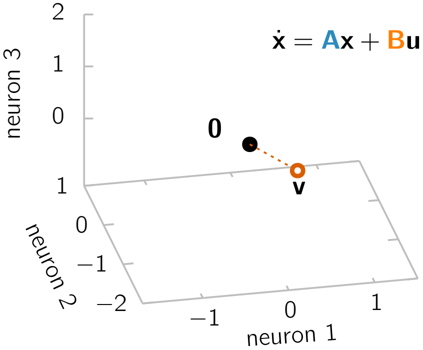
Controllability & Observability



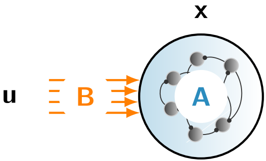
using these input channels,
can I steer $\mathbf{x}$ from $\mathbf{0}$ to any $\mathbf{v}$?

controllability Gramian $\mathbf{P} \succeq 0$

$$\mathbf{P} \equiv \int_0^\infty \exp(t\mathbf{A})\mathbf{B}\mathbf{B}^T \exp(t\mathbf{A}^T)dt$$



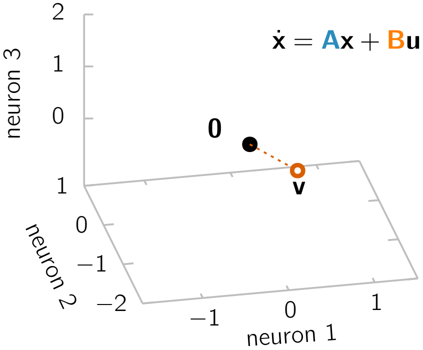
Controllability & Observability



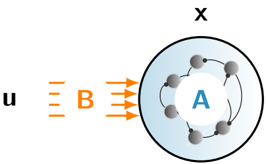
using these input channels,
can I steer **x** from **0** to any **v**?

controllability Gramian $\mathbf{P} \succeq 0$

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0$$



Controllability & Observability



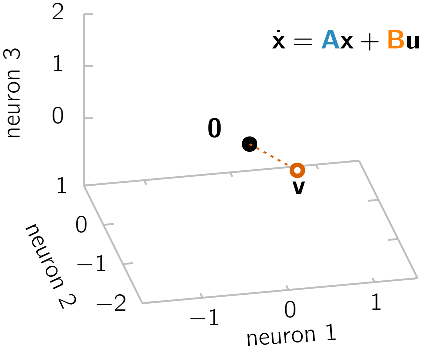
using these input channels,
can I steer $\mathbf{x}$ from $\mathbf{0}$ to any $\mathbf{v}$?

controllability Gramian $\mathbf{P} \succeq 0$

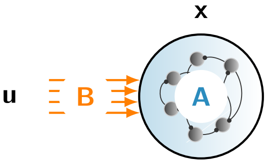
$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0$$

answer is yes,
provided $\mathbf{P}$ is non-singular

more generally, $\mathbf{v}^T\mathbf{P}^{-1}\mathbf{v}$ is
how much input energy is required



Controllability & Observability



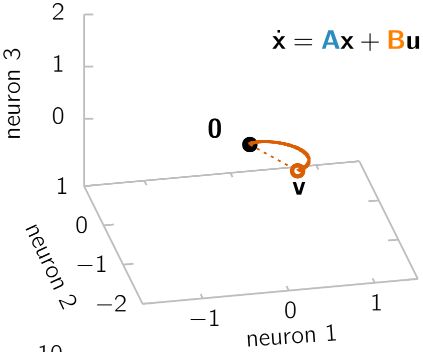
using these input channels,
can I steer $\mathbf{x}$ from $\mathbf{0}$ to any $\mathbf{v}$?

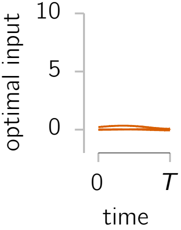
controllability Gramian $\mathbf{P} \succeq 0$

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0$$

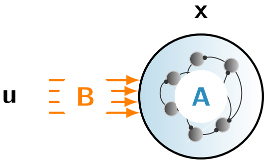
answer is yes,
provided $\mathbf{P}$ is non-singular

more generally, $\mathbf{v}^T\mathbf{P}^{-1}\mathbf{v}$ is
how much input energy is required





Controllability & Observability



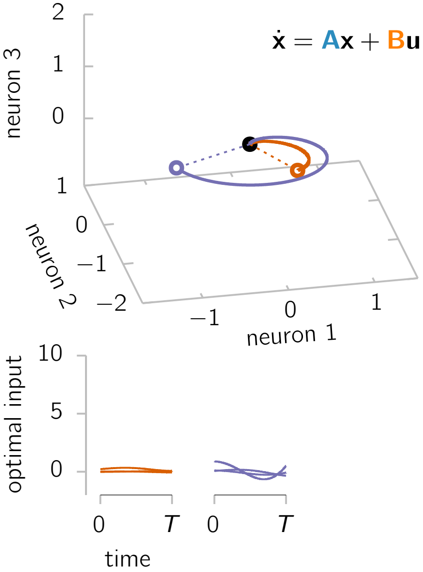
using these input channels,
can I steer $\mathbf{x}$ from $\mathbf{0}$ to any $\mathbf{v}$?

controllability Gramian $\mathbf{P} \succeq 0$

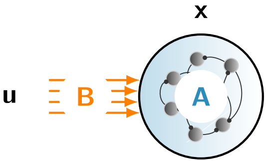
$$\mathbf{AP} + \mathbf{PA}^T + \mathbf{BB}^T = 0$$

answer is yes,
provided $\mathbf{P}$ is non-singular

more generally, $\mathbf{v}^T\mathbf{P}^{-1}\mathbf{v}$ is
how much input energy is required



Controllability & Observability



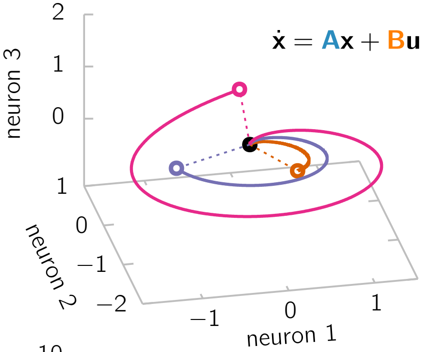
using these input channels,
can I steer $\mathbf{x}$ from $\mathbf{0}$ to any $\mathbf{v}$?

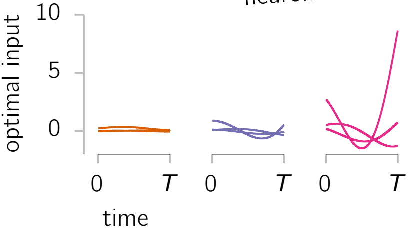
controllability Gramian $\mathbf{P} \succeq 0$

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0$$

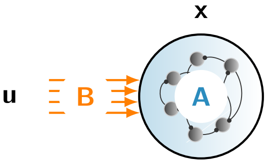
answer is yes,
provided $\mathbf{P}$ is non-singular

more generally, $\mathbf{v}^T\mathbf{P}^{-1}\mathbf{v}$ is
how much input energy is required





Controllability & Observability



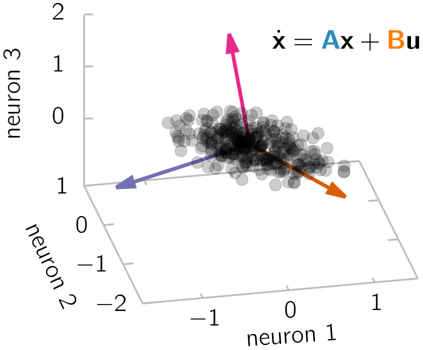
using these input channels,
can I steer $\mathbf{x}$ from $\mathbf{0}$ to any $\mathbf{v}$?

controllability Gramian $\mathbf{P} \succeq 0$

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0$$

answer is yes,
provided $\mathbf{P}$ is non-singular

more generally, $\mathbf{v}^T\mathbf{P}^{-1}\mathbf{v}$ is
how much input energy is required

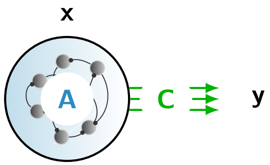


if $\mathbf{u}(t) =$ white noise

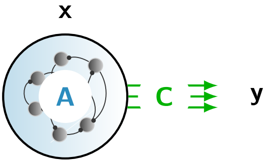
$$\mathbf{P} = \left\langle \mathbf{x}(t)\mathbf{x}(t)^T \right\rangle_t$$

cf. "intrinsic manifold"
[Sadtler et al., *Nature* (2010)]

Controllability & Observability

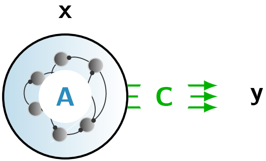


Controllability & Observability



if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$,
can I reconstruct
the initial state $\mathbf{x}(0)$?

Controllability & Observability

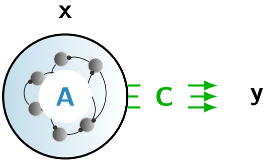


if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$,
can I reconstruct
the initial state $\mathbf{x}(0)$?

observability Gramian $\mathbf{Q} \succeq 0$

$$\mathbf{Q} \equiv \int_0^\infty \exp(t\mathbf{A}^T)\mathbf{C}^T\mathbf{C}\exp(t\mathbf{A})dt$$

Controllability & Observability

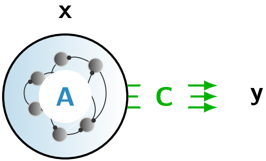


if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$,
can I reconstruct
the initial state $\mathbf{x}(0)$?

observability Gramian $\mathbf{Q} \succeq 0$

$$\mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = 0$$

Controllability & Observability



if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$,
can I reconstruct
the initial state $\mathbf{x}(0)$?

observability Gramian $\mathbf{Q} \succeq 0$

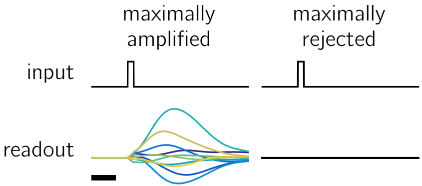
$$\mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = 0$$

answer is yes,
provided $\mathbf{Q}$ is non-singular

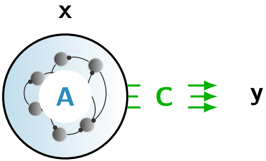
Controllability & Observability

alternative, useful interpretation:

$\mathbf{x}_0^T \mathbf{Q} \mathbf{x}_0$ is the output energy
evoked by initial condition $\mathbf{x}_0$



[Hennequin et al., *Neuron* (2014)]



if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$,
can I reconstruct
the initial state $\mathbf{x}(0)$?

observability Gramian $\mathbf{Q} \succeq 0$

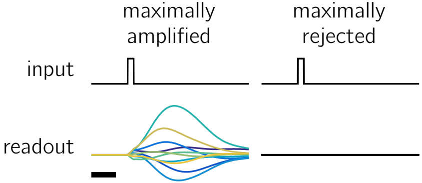
$$\mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = 0$$

answer is yes,
provided $\mathbf{Q}$ is non-singular

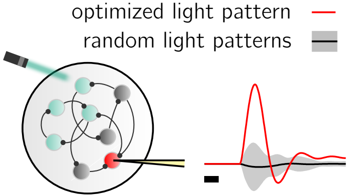
Controllability & **Observability**

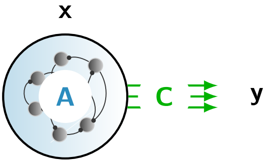
alternative, useful interpretation:

$\mathbf{x}_0^T \mathbf{Q} \mathbf{x}_0$ is the output energy evoked by initial condition $\mathbf{x}_0$



[Hennequin et al., *Neuron* (2014)]





if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$,
can I reconstruct
the initial state $\mathbf{x}(0)$?

observability Gramian $\mathbf{Q} \succeq 0$

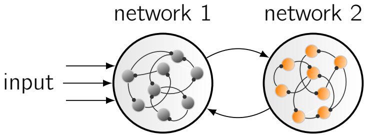
$$\mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = 0$$

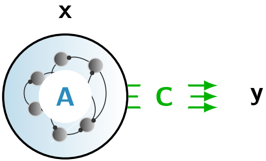
answer is yes,
provided $\mathbf{Q}$ is non-singular

Controllability & **Observability**

alternative, useful interpretation:

$\mathbf{x}_0^T \mathbf{Q} \mathbf{x}_0$ is the output energy evoked by initial condition $\mathbf{x}_0$





if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$, can I reconstruct the initial state $\mathbf{x}(0)$?

observability Gramian $\mathbf{Q} \succeq 0$

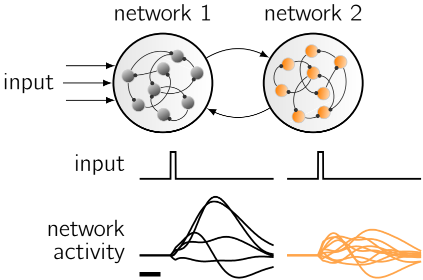
$$\mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = 0$$

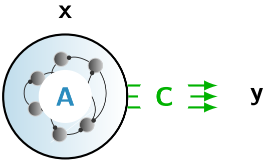
answer is yes, provided $\mathbf{Q}$ is non-singular

Controllability & Observability

alternative, useful interpretation:

$\mathbf{x}_0^T \mathbf{Q} \mathbf{x}_0$ is the output energy evoked by initial condition $\mathbf{x}_0$





if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$, can I reconstruct the initial state $\mathbf{x}(0)$?

observability Gramian $\mathbf{Q} \succeq 0$

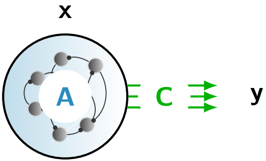
$$\mathbf{A}^T \mathbf{Q} + \mathbf{Q} \mathbf{A} + \mathbf{C}^T \mathbf{C} = 0$$

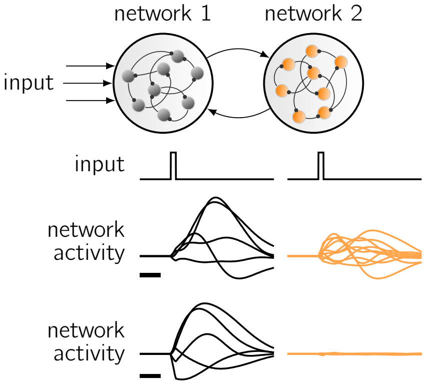
answer is yes, provided $\mathbf{Q}$ is non-singular

Controllability & Observability

alternative, useful interpretation:

$\mathbf{x}_0^T \mathbf{Q} \mathbf{x}_0$ is the output energy
evoked by initial condition $\mathbf{x}_0$





if I observe $\mathbf{u}(t)$ and $\mathbf{y}(t)$,
can I reconstruct
the initial state $\mathbf{x}(0)$?

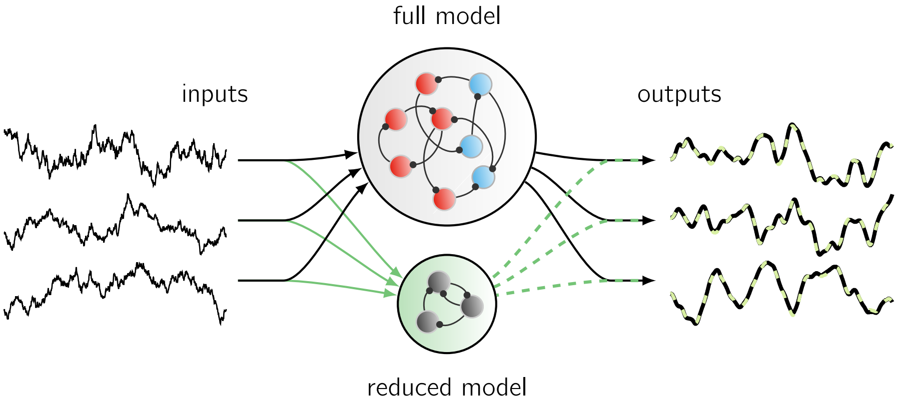
observability Gramian $\mathbf{Q} \succeq 0$

$$\mathbf{A}^T \mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T \mathbf{C} = 0$$

answer is yes,
provided $\mathbf{Q}$ is non-singular

"dynamical nullspace" [Duncker et al., Cosyne '17]

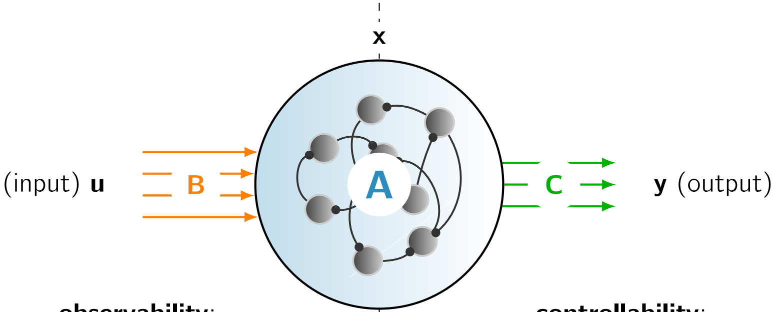
Model reduction



classic methods try to capture the top subspace of both **P** and **Q**

fun fact: for our ISN model of M1, we recover rotational dynamics
(also revealed by jPCA)

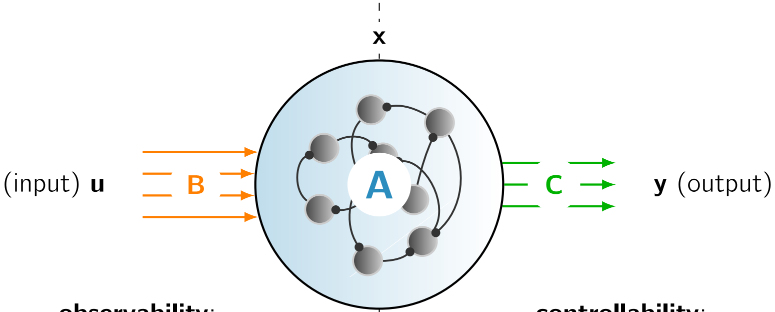
Observability & Controllability



observability:
"what input patterns is
my network most/least sensitive to?"

controllability:
"what activity patterns does
my network like/hate to produce?"

Observability & Controllability



observability:
"what input patterns is
my network most/least sensitive to?"

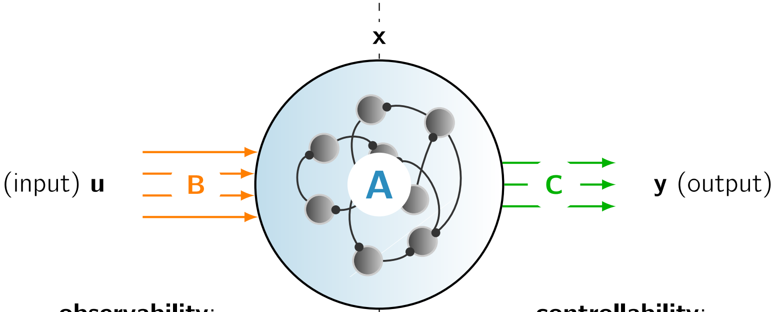
controllability:
"what activity patterns does
my network like/hate to produce?"

P and **Q** are all you need to know about **A**

for biologically realistic (nonnormal) networks, $\mathbf{P} \neq \mathbf{Q}$

[Kao, Sadabadi & Hennequin, Cosyne '18]

Observability & Controllability



observability:
"what input patterns is
my network most/least sensitive to?"

controllability:
"what activity patterns does
my network like/hate to produce?"

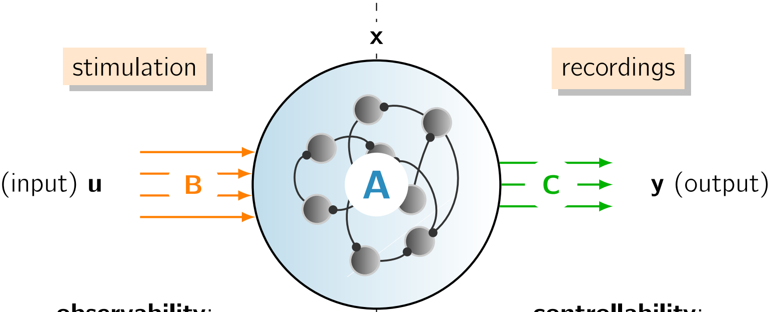
P and **Q** are all you need to know about **A**

for biologically realistic (nonnormal) networks, $\mathbf{P} \neq \mathbf{Q}$

[Kao, Sadabadi & Hennequin, Cosyne '18]

implications for **system identification**?

Observability & Controllability



observability:
"what input patterns is
my network most/least sensitive to?"

controllability:
"what activity patterns does
my network like/hate to produce?"

P and **Q** are all you need to know about **A**

for biologically realistic (nonnormal) networks, $\mathbf{P} \neq \mathbf{Q}$

[Kao, Sadabadi & Hennequin, Cosyne '18]

implications for **system identification**?

System identification

Calvin Kao

simplified scenario ($N = 5$):

$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \text{noise}$

$\mathbf{y}_t = \mathbf{x}_t + \text{noise}$

maximum likelihood parameter estimation
(new method, regularised, stable)

$\hat{\mathbf{A}} = \text{argmax}_{\mathsf{A}} \log p(\mathbf{y}_{0:T} | \mathbf{A})$

System identification

Calvin Kao

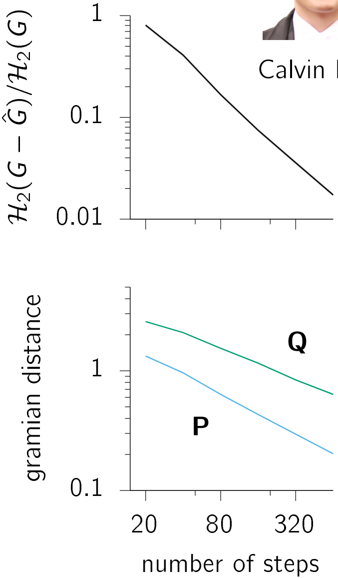
simplified scenario ($N = 5$):

$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \text{noise}$

$\mathbf{y}_t = \mathbf{x}_t + \text{noise}$

maximum likelihood parameter estimation (new method, regularised, stable)

$\hat{\mathbf{A}} = \text{argmax}_{\mathbf{A}} \log p(\mathbf{y}_{0:T} | \mathbf{A})$



System identification

Calvin Kao

simplified scenario ($N = 5$):

$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \text{noise}$

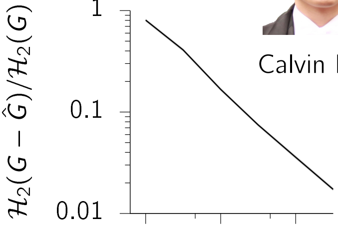
$\mathbf{y}_t = \mathbf{x}_t + \text{noise}$

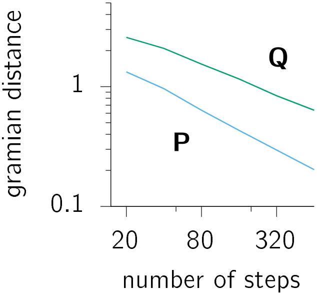
maximum likelihood parameter estimation
(new method, regularised, stable)

$\hat{\mathbf{A}} = \text{argmax}_{\mathbf{A}} \log p(\mathbf{y}_{0:T} | \mathbf{A})$

Q recovered less accurately than **P**

learned model not very predictive of
the effect of stimulation





Outlook

control-theoretic perspectives on RNNs offers new ways of understanding them (here, we've barely scratched the surface)

use of optogenetic access to constrain **I/O models** of neural circuits

motor control, BCIs, reinforcement learning, behavioural modelling, . . .