Is it that simple? The use of linear models in cognitive neuroscience

CCN Generative Adversarial Collaboration Proposal

Scientific question

Do linear models provide an accurate, interpretable, and biologically plausible description of brain activity?

Background

Modern cognitive neuroscience heavily relies on linear models. Such models are used to map between patterns of brain activity and a measure X, where X can be a feature/function of the stimulus (1–10), a behavioral measure (e.g., 11–13), or even brain activity in another species (14,15). The use of linear (as opposed to nonlinear) models is widespread for two main reasons: (a) linear readout is considered to be neurally plausible and thus informative of the underlying neural representations (16–19), and (b) linear models are relatively easy to build and can generalize successfully even in small data regimes (17,20,21).

In recent years, increased availability of large datasets and computational resources has enabled researchers to overcome some of the practical limitations of nonlinear approaches. As a result, many prediction-oriented neuroscience studies have begun to apply nonlinear models to identify neural correlates of brain disorders (22–25) or behavioral traits (26–28). However, basic cognitive neuroscience remains firmly grounded in linear models, resulting in a gap between prediction-oriented and explanation-oriented approaches.

The Controversy

Although the linear readout assumption is widely accepted in cognitive neuroscience, neural computations are, in fact, often nonlinear (29–35). Further, even if we accept the linear readout assumption at the level of individual neurons, it might not hold for signals recorded from outside the skull (36–38) or signals based on indirect measures of neural activity, such as blood flow (39,40). Finally, even if the transmission of signals from one step of a neural computation to another can be approximated with a linear transform, the linearity assumption might break down once we consider multiple successive computations (41,42) or activity aggregated across large neural populations (43–45). As a result, some recent neuroimaging studies have advocated the use of nonlinear models, arguing that they represent a more plausible view of neural interactions within and between brain regions (46–49), at least for higher-level associative cortex (50–52).

The empirical success of linear models seems to speak to the usefulness of the linear readout assumption. However, the persistent focus on linear transformations may be stalling the field, and a shift toward nonlinear models may yield important insights about brain function. We therefore propose to combine

theoretical and experimental approaches in order to examine the validity, benefits, and limitations of linear vs. nonlinear models applied to neuroimaging data.

Competing Hypotheses

Hypothesis 1: Linear models provide an accurate, interpretable, and biologically plausible interpretation of brain activity.

Hypothesis 2: Nonlinear models provide an accurate, interpretable, and biologically plausible interpretation of brain activity, which cannot be achieved with linear models alone.

Approach

We propose an integrated, two-pronged approach for evaluating the use of linear and nonlinear models in cognitive neuroscience. First, we will synthesize existing literature on the linear readout assumption and introduce novel information-theoretic approaches for model evaluation. Then, we will build upon those theoretical insights to evaluate the models' performance on existing datasets.

1. The Theory Branch
    a. Establish the theoretical validity of the linear readout assumptions when applied to neuroimaging data.
    b. Develop information-theoretic criteria for evaluating the use of linear and nonlinear models in neuroimaging research (see 53–55).
2. The Empirical Branch
    a. Evaluate practical limitations of linear vs. nonlinear models, such as the amount of data required for successful performance and the upper limit on feature complexity (see 56–58).
    b. Integrate theory-driven and practical considerations to develop goals and metrics enabling a systematic comparison of linear vs. nonlinear model performance.
    c. Compare the predictive and explanatory power of linear vs. nonlinear models when applied to three different domains:
        i. Mapping from stimulus features to neural activity.
        ii. Mapping from neural activity to behavior.
        iii. Mapping from neural activity in one brain region to neural activity in another brain region.

Concrete outcomes

1. An information-theoretic framework for the use of linear vs. nonlinear models with neuroimaging data.
2. A detailed set of guidelines for the use of linear vs. nonlinear models with neuroimaging data, based on both theoretical and empirical considerations.
3. An online platform enabling researchers to systematically compare linear and nonlinear models according to a predefined set of metrics (see, e.g., 8).

Benefit to the community

Given the overwhelming use of linear models in the field, we believe that a thorough examination of the linear readout assumption is required to ensure that researchers do not overlook important insights about the brain by unnecessarily restricting the set of models they consider. On the other hand, given the potentially unbounded complexity of nonlinear models, neuroscientists must be careful in their application and interpretation. If we demonstrate the benefit of nonlinear models, at least in some cases, our work may catalyze a new line of inquiry in cognitive computational neuroscience. If we demonstrate that linear models satisfy the field's criteria of being accurate, interpretable, and biologically plausible, our work will allow future researchers to continue relying on linear rather than nonlinear approaches, thus saving money, time, and computational resources.

Thus, we expect that our findings will be relevant to any neuroscientist who uses multivariate methods to analyze neuroimaging data. They also have the potential to benefit other researchers investigating complex information processing systems (e.g. artificial neural networks).

Core members

Team linear models:	Martin Schrimpf (graduate student, MIT)
	Leyla Isik (assistant professor, Johns Hopkins University)
Team nonlinear models:	Anna Ivanova (graduate student, MIT)
	Stefano Anzellotti (assistant professor, Boston College)
The advisory team:	Noga Zaslavsky (postdoctoral fellow, MIT)
	Evelina Fedorenko (associate professor, MIT)

Member roles

All members will contribute to organizing the workshop and writing the paper. In addition, we will perform the following tasks:

Anna Ivanova:	examine the theoretical assumptions underlying linear models applied to neuroimaging data.
Martin Schrimpf:	evaluate linear and nonlinear model performance on existing neuroimaging datasets.
Leyla Isik:	develop goals/metrics to systematically evaluate linear vs. nonlinear model performance.
Stefano Anzellotti:	examine methodological benefits and limitations of linear vs. nonlinear models.
Noga Zaslavsky:	develop information-theoretic methods for studying and evaluating linear and non-linear models; guide the theory branch of the project.
Evelina Fedorenko:	guide the empirical branch of the project and the integration of final results.

Signed: Anna Ivanova, Martin Schrimpf, Leyla Isik, Stefano Anzellotti, Noga Zaslavsky, and Evelina Fedorenko

References

1. Hebart MN, Bankson BB, Harel A, Baker CI, Cichy RM. The representational dynamics of task and object processing in humans. Culham JC, editor. eLife. 2018 Jan 31;7:e32816.
2. Isik L, Meyers EM, Leibo JZ, Poggio T. The dynamics of invariant object recognition in the human visual system. Journal of Neurophysiology. 2013 Oct 2;111(1):91–102.
3. Isik L, Tacchetti A, Poggio T. A fast, invariant representation for human action in the visual system. Journal of Neurophysiology. 2017 Nov 14;119(2):631–40.
4. Isik L, Mynick A, Pantazis D, Kanwisher N. The speed of human social interaction perception. NeuroImage. 2020 Jul 15;215:116844.
5. Kubilius J, Schrimpf M, Kar K, Rajalingham R, Hong H, Majaj N, et al. Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc F, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32. Curran Associates, Inc.; 2019 [cited 2020 Jul 16]. p. 12805–12816.
6. Mitchell TM, Shinkareva SV, Carlson A, Chang K-M, Malave VL, Mason RA, et al. Predicting human brain activity associated with the meanings of nouns. Science. 2008 May 30;320(5880):1191–5.
7. Pereira F, Lou B, Pritchett B, Ritter S, Gershman SJ, Kanwisher N, et al. Toward a universal decoder of linguistic meaning from brain activation. Nat Commun. 2018 Mar 6;9(1):1–13.
8. Schrimpf M, Kubilius J, Hong H, Majaj NJ, Rajalingham R, Issa EB, et al. Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like? bioRxiv. 2018 Sep 5;407007.
9. Schrimpf M, Blank I, Tuckute G, Kauf C, Hosseini EA, Kanwisher N, et al. Artificial Neural Networks Accurately Predict Language Processing in the Brain. bioRxiv. 2020 Jun 27;2020.06.26.174482.
10. Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci USA. 2014 Jun 10;111(23):8619–24.
11. Majaj NJ, Hong H, Solomon EA, DiCarlo JJ. Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. J Neurosci. 2015 Sep 30;35(39):13402–18.
12. Mathis A, Rokni D, Kapoor V, Bethge M, Murthy VN. Reading Out Olfactory Receptors: Feedforward Circuits Detect Odors in Mixtures without Demixing. Neuron. 2016 Sep 7;91(5):1110–23.
13. Mruczek REB, Sheinberg DL. Activity of inferior temporal cortical neurons predicts recognition choice behavior and recognition time during visual search. J Neurosci. 2007 Mar 14;27(11):2825–36.
14. Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, et al. Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron. 2008 Dec 26;60(6):1126–41.
15. Mantini D, Hasson U, Betti V, Perrucci MG, Romani GL, Corbetta M, et al. Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. Nature Methods. 2012 Mar;9(3):277–82.
16. Diedrichsen J, Kriegeskorte N. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLOS Computational Biology. 2017 Apr 24;13(4):e1005508.

17. Kriegeskorte N. Pattern-information analysis: From stimulus decoding to computational-model testing. NeuroImage. 2011 May 15;56(2):411–21.
18. Yamins DLK, DiCarlo JJ. Eight open questions in the computational modeling of higher sensory cortex. Current Opinion in Neurobiology. 2016 Apr 1;37:114–20.
19. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 2005 May;8(5):679–85.
20. Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage. 2003 Jun 1;19(2):261–70.
21. Misaki M, Kim Y, Bandettini PA, Kriegeskorte N. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. NeuroImage. 2010 Oct 15;53(1):103–18.
22. Hasanzadeh F, Mohebbi M, Rostami R. Prediction of rTMS treatment response in major depressive disorder using machine learning techniques and nonlinear features of EEG signal. Journal of Affective Disorders. 2019 Sep 1;256:132–42.
23. Kazemi Y, Houghten S. A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data. In: 2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). 2018. p. 1–8.
24. Kim J, Calhoun VD, Shim E, Lee J-H. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. NeuroImage. 2016 Jan 1;124:127–46.
25. Leming M, Górriz JM, Suckling J. Ensemble Deep Learning on Large, Mixed-Site fMRI Datasets in Autism and Other Tasks. Int J Neur Syst. 2020 Jul;30(07):2050012.
26. Kumar S, Yoo K, Rosenberg MD, Scheinost D, Constable RT, Zhang S, et al. An information network flow approach for measuring functional connectivity and predicting behavior. Brain and Behavior. 2019;9(8):e01346.
27. Morioka H, Calhoun V, Hyvärinen A. Nonlinear ICA of fMRI reveals primitive temporal structures linked to rest, task, and behavioral traits. NeuroImage. 2020 Sep 1;218:116989.
28. Xiao L, Stephen JM, Wilson TW, Calhoun VD, Wang Y-P. Alternating Diffusion Map Based Fusion of Multimodal Brain Connectivity Networks for IQ Prediction. IEEE Transactions on Biomedical Engineering. 2019 Aug;66(8):2140–51.
29. Ghazanfar AA, Nicolelis MA. Nonlinear processing of tactile information in the thalamocortical loop. J Neurophysiol. 1997 Jul;78(1):506–10.
30. Gidon A, Zolnik TA, Fidzinski P, Bolduan F, Papoutsi A, Poirazi P, et al. Dendritic action potentials and computation in human layer 2/3 cortical neurons. Science. 2020 Jan 3;367(6473):83–7.
31. Kouh M, Poggio T. A canonical neural circuit for cortical nonlinear operations. Neural Comput. 2008 Jun;20(6):1427–51.
32. Pagan M, Simoncelli EP, Rust NC. Neural Quadratic Discriminant Analysis: Nonlinear Decoding with V1-Like Computation. Neural Comput. 2016 Nov;28(11):2291–319.
33. Ukita J, Yoshida T, Ohki K. Characterisation of nonlinear receptive fields of visual neurons by convolutional neural network. Scientific Reports. 2019 Mar 7;9(1):3791.
34. Walker EY, Sinz FH, Cobos E, Muhammad T, Froudarakis E, Fahey PG, et al. Inception loops discover what excites neurons most using deep predictive models. Nature Neuroscience. 2019 Dec;22(12):2060–5.
35. Shamir M, Sompolinsky H. Nonlinear Population Codes. Neural Computation. 2004 Jun 1;16(6):1105–36.

36. Robinson PA, Rennie CJ, Wright JJ, Bahramali H, Gordon E, Rowe DL. Prediction of electroencephalographic spectra from neurophysiology. Phys Rev E. 2001 Jan 18;63(2):021903.
37. Stam CJ. Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. Clinical Neurophysiology. 2005 Oct 1;116(10):2266–301.
38. David O, Friston KJ. A neural mass model for MEG/EEG:: coupling and neuronal dynamics. NeuroImage. 2003 Nov 1;20(3):1743–55.
39. Heeger DJ, Ress D. What does fMRI tell us about neuronal activity? Nature Reviews Neuroscience. 2002 Feb;3(2):142–51.
40. Bao P, Purington CJ, Tjan BS. Using an achiasmic human visual system to quantify the relationship between the fMRI BOLD signal and neural response. Culham JC, editor. eLife. 2015 Nov 27;4:e09600.
41. DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? Neuron. 2012 Feb 9;73(3):415–34.
42. Pagan M, Urban LS, Wohl MP, Rust NC. Signals in inferotemporal and perirhinal cortex suggest an untangling of visual target information. Nature Neuroscience. 2013 Aug;16(8):1132–9.
43. Gallego JA, Perich MG, Miller LE, Solla SA. Neural Manifolds for the Control of Movement. Neuron. 2017 Jun 7;94(5):978–84.
44. Sohn H, Narain D, Meirhaeghe N, Jazayeri M. Bayesian Computation through Cortical Latent Dynamics. Neuron. 2019 Sep 4;103(5):934-947.e5.
45. Maass W, Joshi P, Sontag ED. Principles of real-time computing with feedback applied to cortical microcircuit models. In: Weiss Y, Schölkopf B, Platt JC, editors. Advances in Neural Information Processing Systems 18. MIT Press; 2006 [cited 2020 Jul 17]. p. 835–842.
46. Güçlü U, van Gerven MAJ. Modeling the Dynamics of Human Brain Activity with Recurrent Neural Networks. Front Comput Neurosci. 2017 [cited 2020 Jun 21];11.
47. Plis SM, Hjelm DR, Salakhutdinov R, Allen EA, Bockholt HJ, Long JD, et al. Deep learning for neuroimaging: a validation study. Front Neurosci. 2014 [cited 2020 Jul 17];8.
48. Thomas AW, Heekeren HR, Müller K-R, Samek W. Analyzing Neuroimaging Data Through Recurrent Deep Learning Models. Front Neurosci. 2019 [cited 2020 Jul 17];13.
49. Fang M, Aglinskas A, Li Y, Anzellotti S. Identifying hubs that integrate responses across multiple category-selective regions. PsyArXiv. 2019 Nov 14 [cited 2020 Jul 17];
50. Anzellotti S, Fedorenko E, Kell AJE, Caramazza A, Saxe R. Measuring and Modeling Nonlinear Interactions Between Brain Regions with fMRI. bioRxiv. 2017 Sep 12;074856.
51. Bertolero MA, Bassett DS. Deep Neural Networks Carve the Brain at its Joints. arXiv:200208891 [physics, q-bio]. 2020 Feb 20 [cited 2020 Jun 21].
52. Dezfouli A, Morris R, Ramos FT, Dayan P, Balleine B. Integrated accounts of behavioral and neuroimaging data using flexible recurrent neural network models. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. Advances in Neural Information Processing Systems 31. Curran Associates, Inc.; 2018 [cited 2020 Jul 17]. p. 4228–4237.
53. Palmer SE, Marre O, Berry MJ, Bialek W. Predictive information in a sensory population. PNAS. 2015 Jun 2;112(22):6908–13.
54. Panzeri S, Magri C, Logothetis NK. On the use of information theory for the analysis of the relationship between neural and imaging signals. Magn Reson Imaging. 2008 Sep;26(7):1015–25.
55. Xu Y, Zhao S, Song J, Stewart R, Ermon S. A Theory of Usable Information Under

Computational Constraints. 2020 Feb 25 [cited 2020 Jul 16]; Available from: https://arxiv.org/abs/2002.10689v1
56. He T, Kong R, Holmes AJ, Sabuncu MR, Eickhoff SB, Bzdok D, et al. Is deep learning better than kernel regression for functional connectivity prediction of fluid intelligence? In: 2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI). 2018. p. 1–4.
57. Schulz M-A, Yeo BTT, Vogelstein JT, Mourao-Miranada J, Kather JN, Kording K, et al. Deep learning for brains?: Different linear and nonlinear scaling in UK Biobank brain images vs. machine-learning datasets. bioRxiv. 2019 Sep 6;757054.
58. Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. Neuroimage. 2018 15;180(Pt A):68–77.