

Toward Bayesian permutation inference for identifying neurons in *C. elegans*.

Gonzalo E. Mena^{1,*}, Scott Linderman^{1,*}, David Belanger², Jasper Snoek², John Cunningham^{1,3}, Liam Paninski^{1,3},
1. Department of Statistics, Columbia University, New York, NY, USA 2. Google Brain, Cambridge, MA. 3. Center for Theoretical Neuroscience and Grossman Center for the Statistics of Mind, Columbia University, New York, NY, USA.,

Summary

Overarching goal

- State and infer Bayesian hierarchical models for the activity in C.elegans combining information (calcium traces) from several worms.
- This is possible as C.elegans nervous system is stereotypical, neurons and connectome don't change across individuals.

Challenge

- If neural identity is known for each trace, one can apply standard Bayesian methodology.
- In practice, laborious human supervision is needed to match recorded traces to canonical neural identities (i.e. names).

Our contribution

- We developed three methods for learning latent matchings. These can be used in variational inference (VI) to jointly estimate a dynamical system and the matching between traces and true neural identities.
- Potentially it may serve to automatize the matching procedure.
- Our method outperforms standard MCMC samplers for inferring permutations.

Future work

- We used real connectome and position information. In the future we plan to use real traces.
- Two new levels of complexity: partially observed brain recordings, more sophisticated dynamical systems.

Model

Simple linear autoregressive model for neural dynamics,

$$\tilde{Y}_t^{(j)} = (W \odot A) \tilde{Y}_{t-1}^{(j)} + \epsilon_t^{(j)}, \quad (1)$$

where $W \in \mathbb{R}^{N \times N}$ is the weight matrix (gaussian prior); $A \in \{0, 1\}^{N \times N}$ is the connectome; $\epsilon_t^{(j)} \sim \mathcal{N}(0, \sigma^2 I)$; and $\tilde{Y}_t^{(j)} \in \mathbb{R}^N$ is the measured neural activity at time t in worm j . The catch is that $\tilde{Y}_t^{(j)}$ is assumed to be in canonical order; i.e. in the same order as the rows and columns of W and A . We actually observe,

$$Y_t^{(j)} = P^{(j)} \tilde{Y}_t^{(j)}.$$

We aim to perform posterior inference of $p(\{W, P^{(j)}\} | A, \{Y^{(j)}\})$.

The permutations are constrained by side information: we use neural position along the worm's body to constrain the possible neural identities for a given recorded neuron. We only allow an observed neuron to be mapped to a known identity if the observed location is within η of the expected location.

Experimental setup

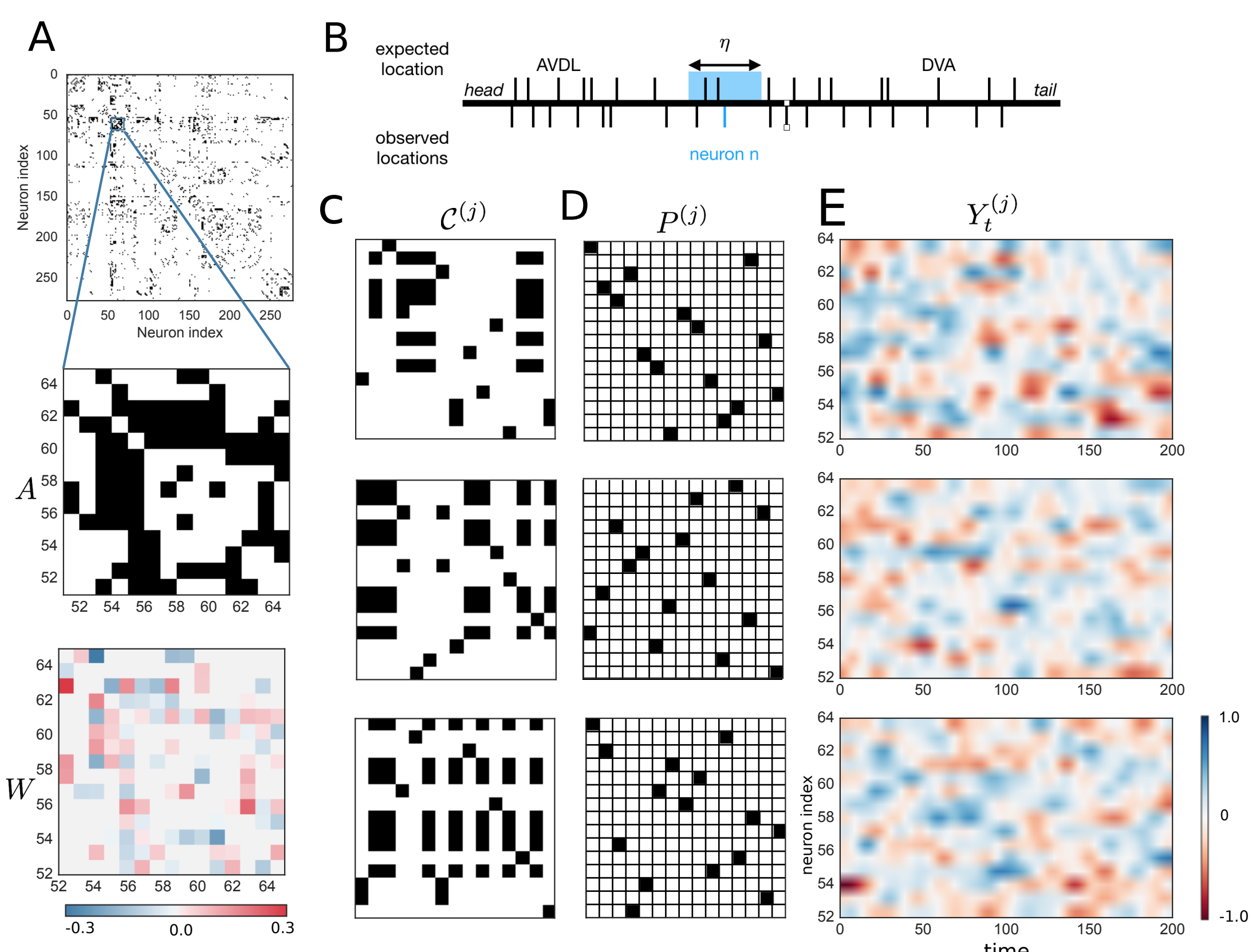


Figure:

Hierarchical Bayesian framework. **A** Adjacency matrix (connectome) A from [11]. We wish to infer the corresponding weight matrix W . **B** We know the typical locations of the neurons [12, 8]. We constrain possible assignments to neuron identities within η of the observed location. **C** These constraints are represented as a matrix $C^{(j)}$ for worm j specifying possible assignments of observed neurons to identities. **D** To infer the weights, we must first infer the permutation $P^{(j)}$ that matching observed neurons to the set of known identities. **E** The observed data is a matrix $Y^{(j)}$ with non-canonical order.

Three reparameterizations for permutations

We extend to permutations the *Concrete* or *Gumbel softmax* relaxations [2, 9] in three different ways. In all relaxations we are concerned with \mathcal{B}_N , the Birkhoff polytope or set of doubly-stochastic matrices.

Stick-Breaking and Rounding

Our stick-breaking construction generalizes from the simplex [6] to the Birkhoff polytope. For the rounding construction, we start with a noise distribution and force it to be close to permutation matrices by rounding them towards the extreme-points of \mathcal{B}_N .

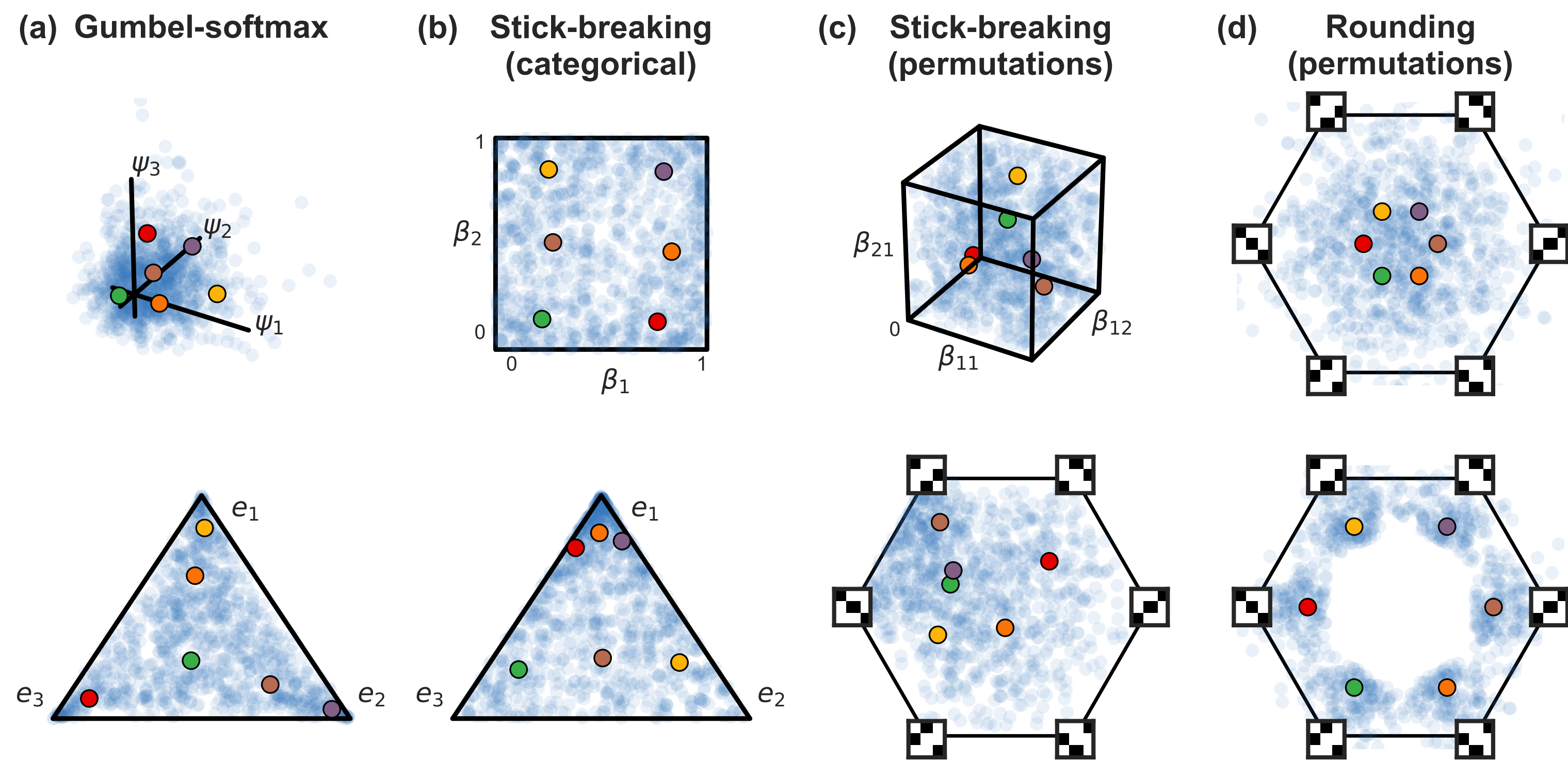


Figure: Rounding and Stick-breaking transformations of noise, and relation to constructions in the simplex

Gumbel-Sinkhorn ($\mathcal{G.S.}$) distribution

We use the *Sinkhorn operator* $S(\cdot)$, the successive row and column normalization of a matrix. This approximates the choice of a permutation $M(X)$; i.e. $M(X) = \lim_{\tau \rightarrow 0} S(X/\tau)$. By adding Gumbel noise we conceive the Gumbel Matching distribution and its approximation, the *G.S. distribution*.

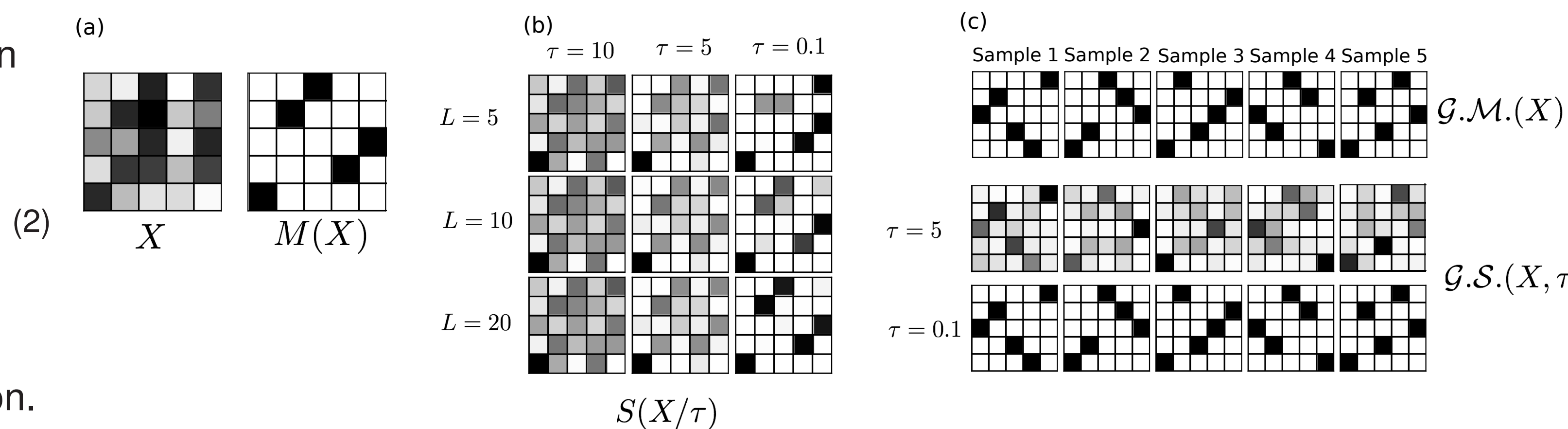


Figure: Matching and Sinkhorn operators, and the Gumbel-Matching and Gumbel-Sinkhorn distributions.

Results

We compared against: (i) naïve variational inference, where we do not enforce the constraint that $P^{(j)}$ be a permutation; (ii) MCMC, where we alternate between sampling from the conditionals of W (Gaussian) and $P^{(j)}$, from which one can sample by proposing local swaps, as described in [1], and (iii) MAP estimation.

Table: Accuracy in the C.elegans neural identification problem, for varying mean number of candidate neurons (10, 30, 45, 60) and number of worms.

	10		30		45		60	
	1 worm	4 worms	1 Worm	4 worms	1 worm	4 worms	1 worms	4 worms
NAIVE VI	.34	.32	.16	.16	.13	.12	.11	.12
MAP	.34	.32	.17	.17	.14	.13	.13	.12
MCMC	.34	.65	.18	.28	.14	.17	.13	.15
VI	.79	.94	.4	.69	.25	.51	.21	.44

Table: Accuracy in inferring true neural identity for different of proportion of known neurons and η .

	40.%		30.%		20.%		10.%	
	$\eta = 0.1$	$\eta = 0.2$	$\eta = 0.1$	$\eta = 0.2$	$\eta = 0.1$	$\eta = 0.2$	$\eta = 0.1$	$\eta = 0.2$
Naive VI	.43	.41	.33	.31	.23	.22	.12	.1
MAP	.42	.41	.33	.32	.23	.22	.12	.11
MCMC	.85	.80	.52	.46	.3	.26	.15	.12
VI	.97	.96	.92	.84	.74	.58	.44	.23

References

- [1] P. Diaconis. The Markov chain Monte Carlo revolution. *Bulletin of the American Mathematical Society*, 46(2):179–205, 2009.
- [2] E. Jang, S. Gu, and B. Poole. Categorical reparameterization with Gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.
- [3] S. Kato, H. Kaplan, T. Schröder, S. Skora, T. Lindsay, E. Yemini, S. Lockery, and M. Zimmer. Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell*, 163(3):656 – 669, 2015. ISSN 0092-8674.
- [4] D. P. Kingma and M. Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations*, 2014.
- [5] R. G. Krishnan, U. Shalit, and D. Sontag. Deep Kalman filters. *arXiv preprint arXiv:1511.05121*, 2015.
- [6] S. Linderman, M. Johnson, and R. P. Adams. Dependent multinomial models made easy: Stick-breaking with the Polya-gamma augmentation. In *Advances in Neural Information Processing Systems*, pages 3456–3464, 2015.
- [7] S. Linderman, M. Johnson, A. Miller, R. Adams, D. Blei, and L. Paninski. Bayesian learning and inference in recurrent switching linear dynamical systems. In *Artificial Intelligence and Statistics*, pages 914–922, 2017.
- [8] R. Lints, Z. F. Altun, H. Weng, T. Stepheyne, G. Stepheyne, M. Volaski, and D. H. Hall. WormAtlas Update. 2005.
- [9] C. J. Maddison, A. Mnih, and Y. W. Teh. The Concrete distribution: A continuous relaxation of discrete random variables. *arXiv preprint arXiv:1611.00712*, 2016.
- [10] D. Soudry, S. Keshri, P. Stinson, M.-h. Oh, G. Iyengar, and L. Paninski. Efficient "shotgun" inference of neural connectivity from highly sub-sampled activity data. *PLoS computational biology*, 11(10):e1004464, 2015.
- [11] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Computational Biology*, 7(2):e1001066, 2011.
- [12] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The structure of the nervous system of the nematode *Caenorhabditis elegans*: the mind of a worm. *Phil. Trans. R. Soc. Lond.*, 314:1–340, 1986.
- [13] R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3–4):229–256, 1992.