Estimating shape distances on neural representations with limited samples

Dean A. Pospisil (dp4846@princeton.edu)
Princeton Neuroscience Institute, Princeton University
Princeton, NJ, 08544
Brett W. Larsen (brettlarsen@flatironinstitute.org)
Center for Computational Neuroscience, Flatiron Institute
New York, NY, 10010
Sarah E. Harvey (sharvey@flatironinstitute.org)
Center for Computational Neuroscience, Flatiron Institute
New York, NY, 10010
Alex H. Williams (awilliams@flatironinstitute.org)
Center for Neural Science, New York University
New York, NY, 10003
Center for Computational Neuroscience, Flatiron Institute
New York, NY, 10010

Abstract

Quantitative comparisons of neural population dynamics across biological systems—e.g. different subjects, animal species, or brain areas—and to artificial network dynamics are of longstanding interest to systems neuroscience. Many metrics of functional, population-level similarity have been proposed including Representational Similarity Analysis (RSA), Centered Kernel Alignment (CKA), and shape distances. However, we still have a poor grasp on fundamental questions: How many neurons, trials, and behavioral conditions do we need to experimentally measure in order to accurately assess the similarity of two neural populations? Here, we mathematically derive concrete answers to these questions for the Procrustes shape distance—a measure of representational distance with desirable theoretical properties (symmetry and triangle inequality) (Williams, Kunz, Kornblith, & Linderman, 2021). We find that the problem is challenging for high-dimensional manifolds—for example, to compensate for a twofold increase in dimensionality, there must be a fourfold increase in the number of sampled conditions. To mitigate these challenges, we introduce a new method-of-moments estimator with a tunable bias-variance tradeoff. We show that this estimator achieves superior performance to standard estimators, particularly in high-dimensional settings. Furthermore, since our approach bounds the bias and variance of the estimate, it naturally produces a confidence interval, which we show to be an accurate reflection of uncertainty in simulation and on semi-synthetic experimental datasets with established ground truth. Finally, we leverage this new estimator to analyze mouse visual cortical responses to 2800 natural images. Thus, we lay the foundation for a rigorous statistical theory for high-dimensional shape analysis, and we contribute a new estimation method well-suited to practical scientific settings.

Keywords: representational geometry, shape metrics, dissimilarity metrics.

We begin by considering a simple setting where each neural network is a deterministic map. A collection of K neural systems can then be viewed as a set of functions, each denoted $h_i : \mathcal{Z} \mapsto \mathbb{R}^N$ for $i \in \{1, \ldots, K\}$. Here, $\mathcal{Z}$ is a feature space and N can be interpreted as the number of neurons in each system (e.g. the hidden layer size or the number of recorded neurons in a biological experiment).

Motivated by the shape theory literature (Kendall et al., 2009; Williams et al., 2021), we consider estimating the ***Procrustes size-and-shape distance***, ρ, and ***Riemannian shape distance***, θ, between neural representations. Let h_i and h_j denote neural systems that are mean-centered and bounded, i.e. $\mathbb{E}[h_i(z)] = \mathbb{E}[h_j(z)] = 0$ and $\|h_i(z)\|_2, \|h_j(z)\|_2 < \sqrt{N}$ almost surely. Throughout, the expectations are taken over $z \sim P$, a distribution over network inputs, and the bounded assumption can be achieved achieved by assuming each neuron has a maximum firing rate. For brevity we focus on the the Procrustes shape distance $\rho(h_i, h_j)$ and which can be defined as $\rho(h_i, h_j) = \min_{Q \in \mathcal{O}(N)} \sqrt{\mathbb{E}\|h_i(z) - Q h_j(z)\|_2^2}$ where $\mathcal{O}(N)$ denotes the set of $N \times N$ orthogonal matrices. It is well-known ρ can be written in terms of the covariance and cross-covariance matrices $\Sigma_{ii} = \mathbb{E}[h_i(z)h_i(z)^\top]$, $\Sigma_{jj} = \mathbb{E}[h_j(z)h_j(z)^\top]$, $\Sigma_{ij} = \mathbb{E}[h_i(z)h_j(z)^\top]$ where $\|\Sigma_{ij}\|_*$ denotes the nuclear norm (or Shatten 1-norm) of the cross-covariance matrix: $\rho^2(h_i, h_j) = \mathrm{Tr}[\Sigma_{ii}] + \mathrm{Tr}[\Sigma_{jj}] - 2\|\Sigma_{ij}\|_*$.

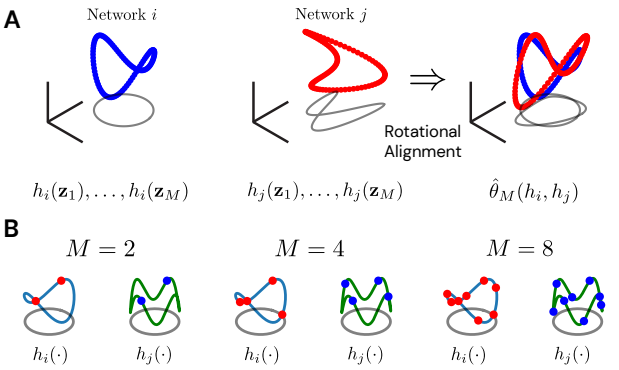

Figure 1: (A) Classical shape distances (Kendall et al., 2009) can be used to provide a rotation-invariant distance between neural representations (Williams et al., 2021). Given two labelled points clouds in N-dimensional space (*left* and *middle*), the distance is computed after an optimal orthogonal transformation is chosen to align the point clouds (*right*). (B) Our ability to estimate the shape distance is related to M, the number of stimuli. As M increases (*left* to *right*) the number of sampled points along the underlying manifold increases, and we are better able to resolve shape differences.

Nonasymptotic Bounds on Plug-in Estimator Performance. Suppose we are given M independent and identically distributed network inputs $z_1, \ldots, z_M \sim P$. How well can we approximate the shape distances between two networks, as a function of M? The standard approach (Williams et al., 2021) is to use a *plug-in estimator* by substituting the empirical covariances $\hat{\Sigma}_{ii} = \frac{1}{M}\sum_{m=1}^M h_i(z_m)h_i(z_m)^\top$, $\hat{\Sigma}_{ij} = \frac{1}{M}\sum_{m=1}^M h_i(z_m)h_j(z_m)^\top$ to approximate the true covariances: $\hat{\rho}^2(h_i, h_j) = \mathrm{Tr}[\hat{\Sigma}_{ii}] + \mathrm{Tr}[\hat{\Sigma}_{jj}] - 2\|\hat{\Sigma}_{ij}\|_*$

We showed under our mean zero and bounded assumptions that with at least at least 95% probability, the following upper bound holds:

$$\frac{|\hat{\rho}^2 - \rho^2|}{N} \leq \frac{2N\log(2N)}{3M} + \frac{2N\sqrt{\log(2N)}}{M^{1/2}} + \left(\frac{1}{M^{1/2}} + \frac{2}{N^{1/2}M^{1/2}}\right)\sqrt{2\log 120} \tag{1}$$

We can gain intuition for eq. 1 by ignoring logarithmic factors and noticing that the second term dominates. Then, roughly speaking, eq. 1 says that we can guarantee the plug-in error decreases as a function of $NM^{-1/2}$. Thus, for any fixed N,

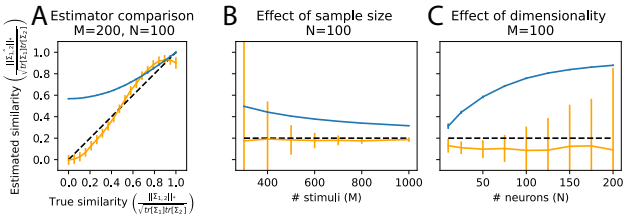

Figure 2: Validation of estimator on synthetic data. (A) The moment based estimator (orange) compared to plug-in estimator (blue) in simulation with standard deviation bars calculated across simulations. (B) Effect of increasing sample size when moment estimator is constrained to have a bias less than 5%. (C) Effect of increasing dimensionality.

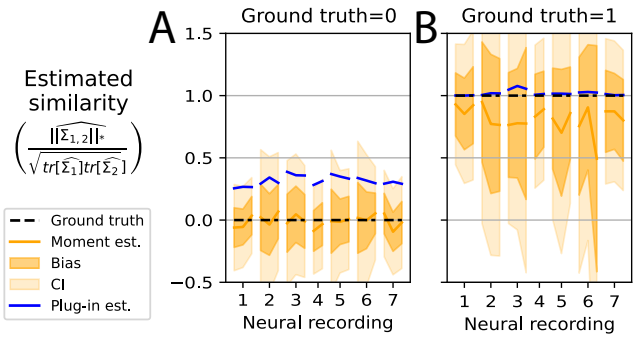

Figure 3: Validation of estimator on neural data (Stringer et al., 2019) (A) Comparison of estimators when ground truth similarity of neural data is set to 0. The estimator is applied to three disjoint sets of random stimuli for each recording ($n = 7$). The estimated maximal bias is plotted in dark orange area; the confidence interval in light orange. (B) Same simulation as (A) except ground truth is 1.

we need to increase M by a factor of 4 to decrease estimation error by a factor of 2. Furthermore, when comparing higher-dimensional neural representations ($\uparrow N$) we need to sample more landmarks—if N increases by a factor of 2, then M must increase by a factor of 4 to compensate.

We also sought to understand if the bound is tight and prove under our assumptions, the rate cannot be improved beyond $M^{-1/2}$ but could potentially improved by a factor of $N^{1/2}$.

A New Estimator with Controllable Bias. The results in the previous section show the plug-in estimator of $\|\Sigma_{ij}\|_*$ has low variance but large and slowly decaying bias. Hence we developed an alternative estimator that is nearly unbiased. First, note that the eigenvalues of $\Sigma_{ij}\Sigma_{ij}^\top$ correspond to the squared singular values of Σ_{ij}. Thus, $\mathrm{Tr}[(\Sigma_{ij}\Sigma_{ij}^\top)^{1/2}] = \|\Sigma_{ij}\|_*$, and so we can reduce our problem to estimating the trace of $(\Sigma_{ij}\Sigma_{ij}^\top)^{1/2}$, which is symmetric. Leveraging ideas from a well-developed literature (Lin, Saad, & Yang, 2016; Kong & Valiant, 2017; Adams et al., 2018), we proceed to define the p^{th} moment of this matrix as $W_p = \mathrm{Tr}[(\Sigma_{ij}\Sigma_{ij}^\top)^p] = \sum_{n=1}^N \lambda_n^p$ where $\lambda_1, \ldots, \lambda_N$ denote the eigenvalues of $\Sigma_{ij}\Sigma_{ij}^\top$. Now, for any function $f : \mathbb{R} \mapsto \mathbb{R}$ and symmetric matrix S with eigenvalues $\lambda_1, \ldots, \lambda_N$, we define $\mathrm{Tr}[f(S)] = \sum_i f(\lambda_i)$. So long as f is reasonably well-behaved, we can approximate it using a truncated power series with P terms where $\gamma_0, \ldots, \gamma_P$ are scalar coefficients for the terms in the power series. In our case, we approximate $f(x) = \sqrt{x}$.

We thus estimate $\|\Sigma_{ij}\|_*$ by (a) specifying an estimator of the top eigenmoments, $W_1, \ldots, W_P$, and (b) specifying a desired set of scalar coefficients $\gamma_0, \ldots, \gamma_P$. To select the scalar coefficients, we propose an optimization procedure that trades off between an upper bound on bias and variance in the estimate of $\|\Sigma_{ij}\|_*$. The estimator includes a user-defined parameter that controls this trade-off in terms of a bias upper bound.

Validation on Synthetic and Neural Data. We first compared the bias of the plug-in estimator to that of the moment-based estimator on synthetic data across a range of ground truth shape similarities (Fig 2A) and as we increase the number of measured stimuli (Fig 2B) or number of neurons (Fig 2C). Next we study our estimator applied to neural data: calcium recordings from mouse primary visual cortex in responses to a set of 2,800 natural images repeated twice (Stringer et al., 2019). We constructed populations of neurons ($N = 40$ each) where the ground truth was 0 (Fig. 3A) and 1 (Fig. 3B). As seen in Fig. 3A, the moment based estimator can accurately determine when the similarity is low in noisy neural data whereas the plug-in estimator cannot.

Discussion. Here, we theoretically characterized "plug-in" estimates of shape distance in high-dimensional, noisy, and sample-limited regimes. We found that these estimates tend to over-estimate representational similarity when the true similarity is small. Further, they require a large number of samples, M, to overcome this bias in high-dimensional regimes. Eq. 1 provides precise guarantees on the worst-case performance of plug-in estimators. These bounds can guide the design of biological experiments, including pre-registered statistical power analysis. However, in terms of the number of samples, the bounds definitively establish that the plug-in error decays at a rate proportional to $M^{-1/2}$.

An equally important contribution of our work is to provide a practical method to *(a)* reduce the bias of plug-in estimators of shape distance and *(b)* enable practicioners to explicitly trade off estimator bias and variance. When employed on a biological dataset published by (Stringer et al., 2019), we find that shape similarity estimates are highly uncertain, revealing the challenging nature of the problem in high dimensions and with noisy data. Importantly, this degree of uncertainty is not obvious from the procedures and plug-in estimates advertised by existing work on this subject.

In summary, our work characterizes the challenges of estimating shape distances in high-dimensional spaces. While shape distances can be well-behaved in certain settings (e.g. in noiseless artificial networks with many sampled conditions), our results suggest the need for carefully designed experiments and estimation procedures in sample-limited regimes.

References

Adams, R. P., Pennington, J., Johnson, M. J., Smith, J., Ovadia, Y., Patton, B., & Saunderson, J. (2018). *Estimating the spectral density of large implicit matrices.*

Kendall et al. (2009). *Shape and shape theory*. John Wiley & Sons.

Kong, W., & Valiant, G. (2017). Spectrum estimation from samples. *The Annals of Statistics*, *45*(5), 2218 – 2247. Retrieved from https://doi.org/10.1214/16-AOS1525 doi: 10.1214/16-AOS1525

Lin, L., Saad, Y., & Yang, C. (2016). Approximating spectral densities of large matrices. *SIAM Review*, *58*(1), 34-65. doi: 10.1137/130934283

Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., & Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature*, *571*(7765), 361–365.

Williams, A. H., Kunz, E., Kornblith, S., & Linderman, S. (2021). Generalized shape metrics on neural representations. In *Neurips* (Vol. 34).