Excerpts from *Geometric Bayes*

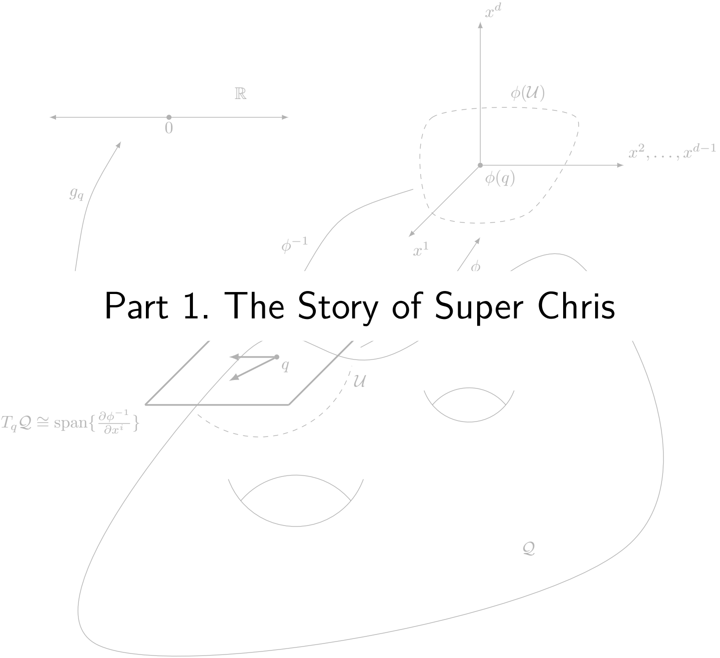
Andrew J. Holbrook

Department of Statistics
University of California, Irvine

Department of Human Genetics
University of California, Los Angeles

JSM 2019

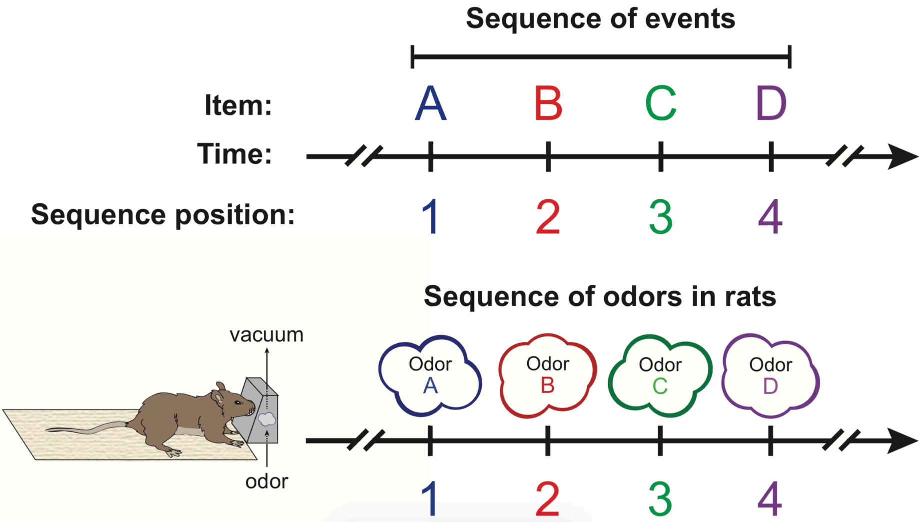
Part 1. The Story of Super Chris

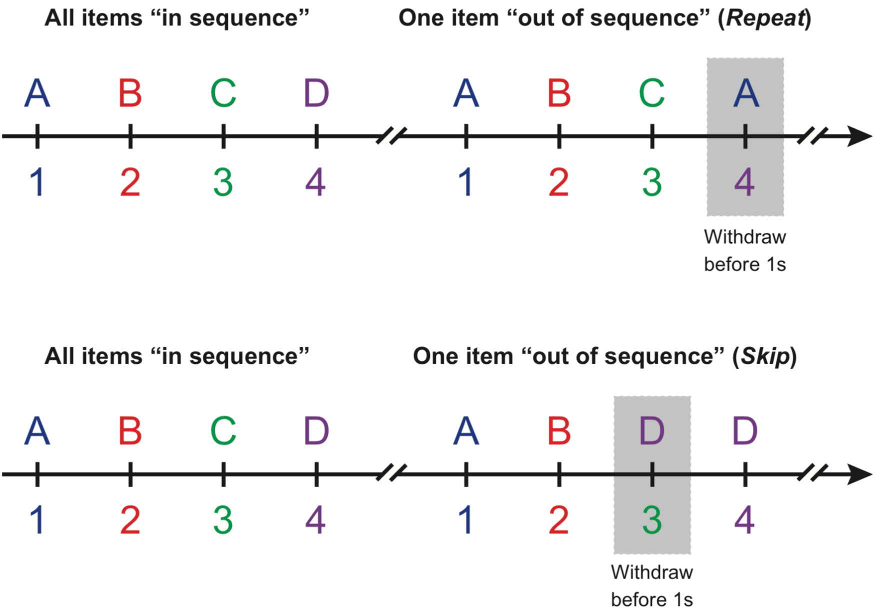


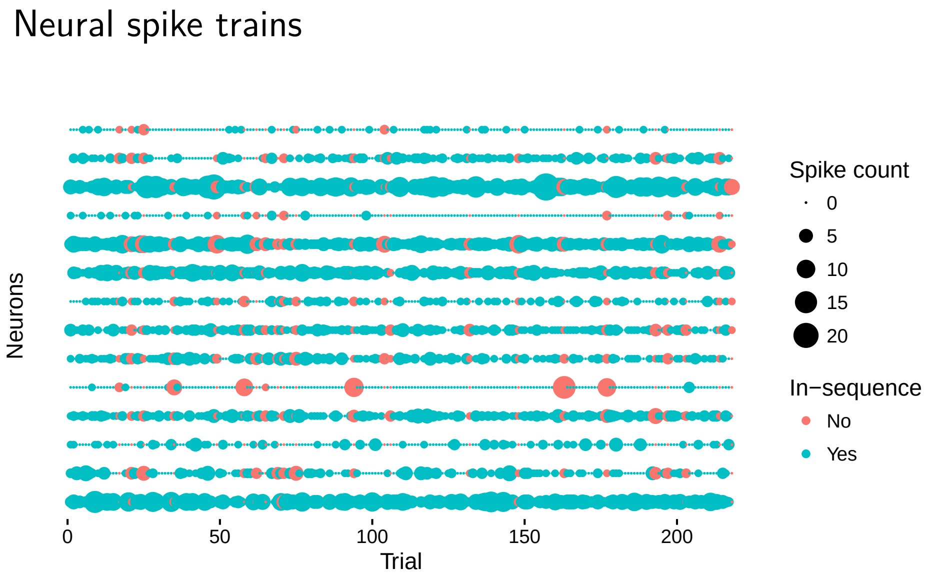
This is Super Chris.

Super Chris is handsome.

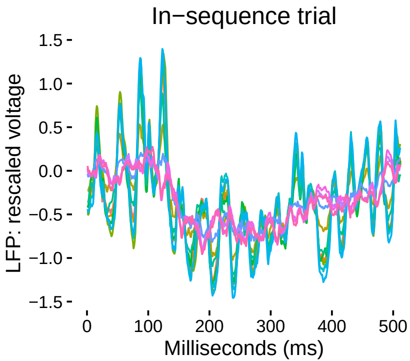
Super Chris is very smart.

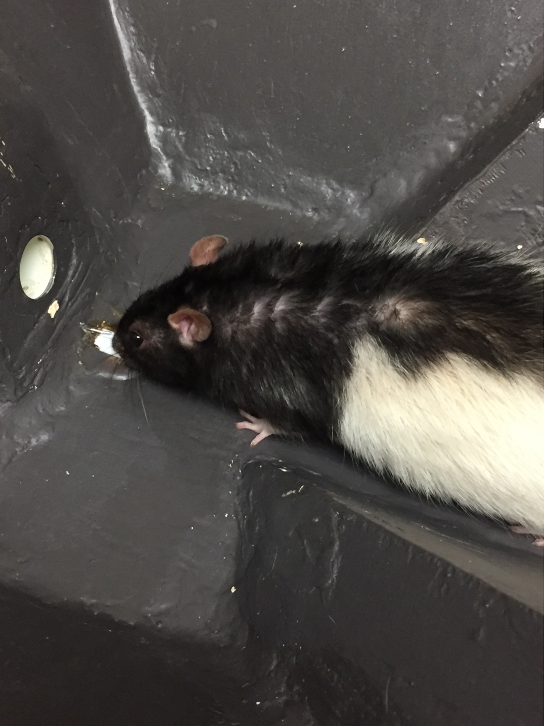

Sequence of events
Item:
A
B
C
D
Time:
Sequence position:
1
2
3
4
Sequence of odors in rats
vacuum
odor
Odor A
Odor B
Odor C
Odor D
1
2
3
4


All items "in sequence"
One item "out of sequence" (*Repeat*)
A B C D
1 2 3 4
A B C A
1 2 3 4
Withdraw before 1s
All items "in sequence"
One item "out of sequence" (*Skip*)
A B C D
1 2 3 4
A B D D
1 2 3 4
Withdraw before 1s


Neural spike trains
Neurons
Trial
0
50
100
150
200
Spike count
0
5
10
15
20
In−sequence
No
Yes

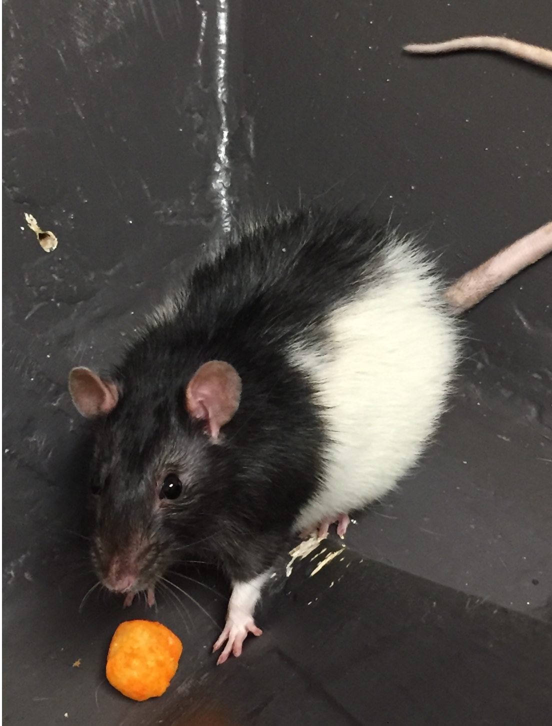
Local field potentials



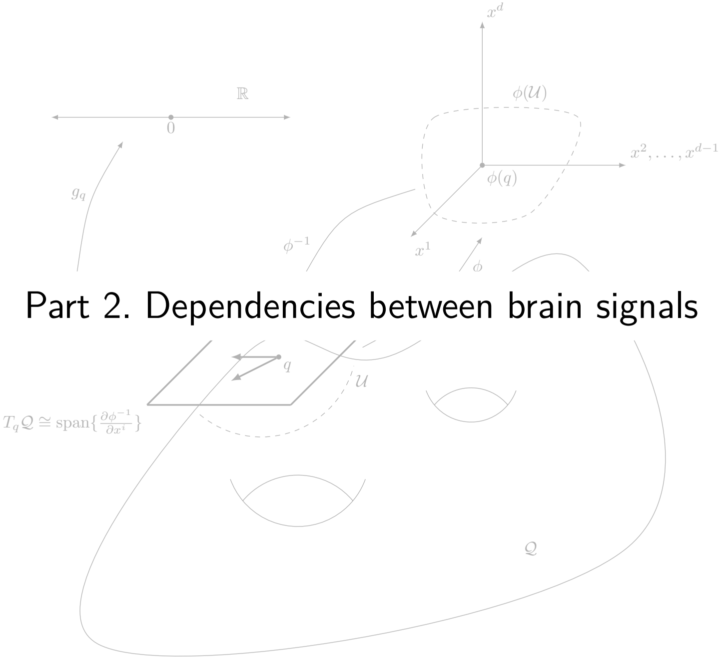


Super Chris gets it right!

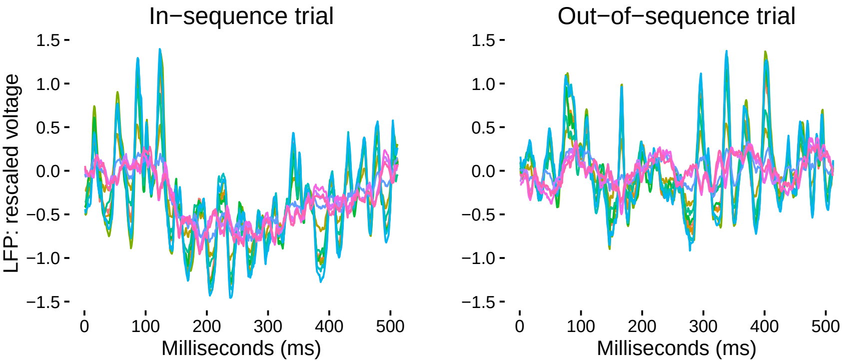
Well done, Super Chris!



Part 2. Dependencies between brain signals



Local field potentials



Spectral density and coherence

Auto-covariance matrix at lag ℓ:

$$\Gamma_\ell = \mathrm{Cov}\big(y(t),\, y(t-\ell)\big) = \mathsf{E}\Big(\big(y(t)-\mu\big)\big(y(t-\ell)-\mu\big)^T\Big)$$

Power spectral density matrix at frequency ω:

$$\Sigma(\omega) = \sum_{\ell=-\infty}^{\infty} \Gamma_\ell \exp(-i2\pi\omega\ell)$$

The squared coherence matrix at frequency ω:

$$\rho_{ij}^2(\omega) = \frac{|\Sigma_{ij}(\omega)|^2}{\Sigma_{ii}(\omega)\,\Sigma_{jj}(\omega)}$$

The Whittle likelihood

For $\omega_k = \frac{k}{T}$ and $k = -(\frac{T}{2} - 1), \dots, \frac{T}{2}$, define

$$Y(\omega_k) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y(t) \exp(-i2\pi\omega_k t)\,.$$

In many situations, we have

$$Y(\omega_k) \overset{ind}{\sim} \mathrm{CN}_d(0, \Sigma(\omega_k))\,,$$

and it is common to assume

$$Y(\omega_k) \overset{iid}{\sim} \mathrm{CN}_d(0, \Sigma_\alpha)$$

for α a band of frequencies.

Convenient and inconvenient priors

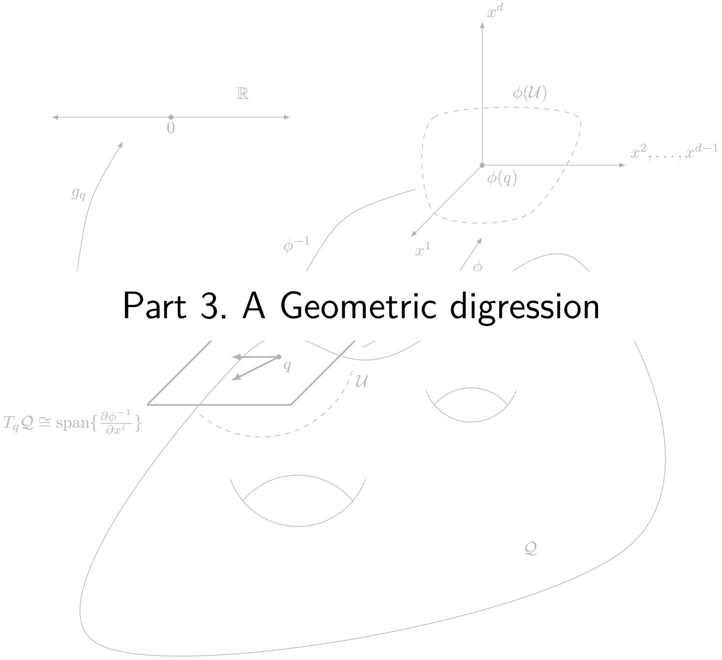
Table 1: Priors for postive definite matrices

Prior	Symmetric	Hermitian
Wishart	$\lvert\Sigma\rvert^{(n-d-1)/2} \exp\left(-\mathrm{tr}\{\Psi^{-1}\Sigma\}/2\right)$	$\lvert\Sigma\rvert^{n-d} \exp\left(-\mathrm{tr}\{\Psi^{-1}\Sigma\}\right)$
inverse-Wishart	$\lvert\Sigma\rvert^{-(n+d+1)/2} \exp\left(-\mathrm{tr}\{\Psi\Sigma^{-1}\}/2\right)$	$\lvert\Sigma\rvert^{-(n+d)} \exp\left(-\mathrm{tr}\{\Psi\Sigma^{-1}\}\right)$
uniform	1	1
Jeffreys	$\lvert\Sigma\rvert^{-(d+1)/2}$	$\lvert\Sigma\rvert^{-d}$
reference	$\left(\lvert\Sigma\rvert \prod_{i<j}(d_i - d_j)\right)^{-1}$	$\left(\lvert\Sigma\rvert \prod_{i<j}(d_i - d_j)^2\right)^{-1}$

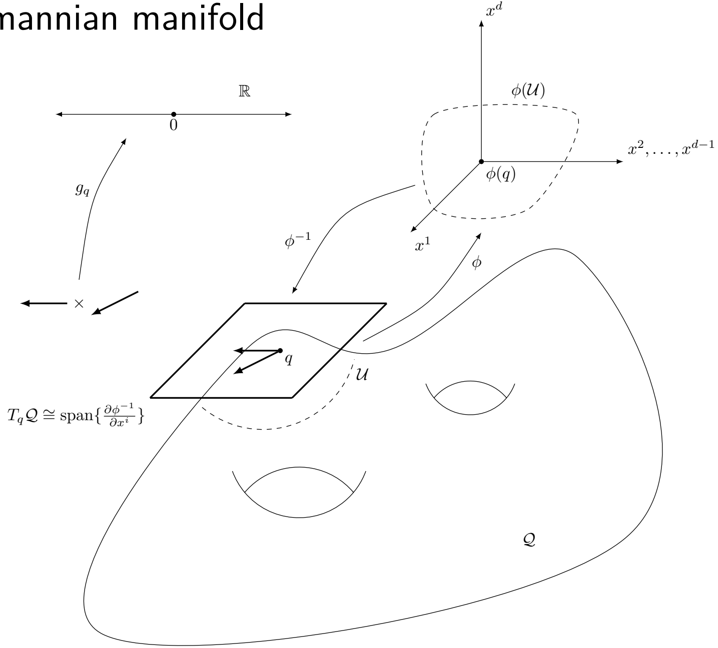
Inference technique should:

- give freedom to choose
- not require complicated parameterizations
- be exact

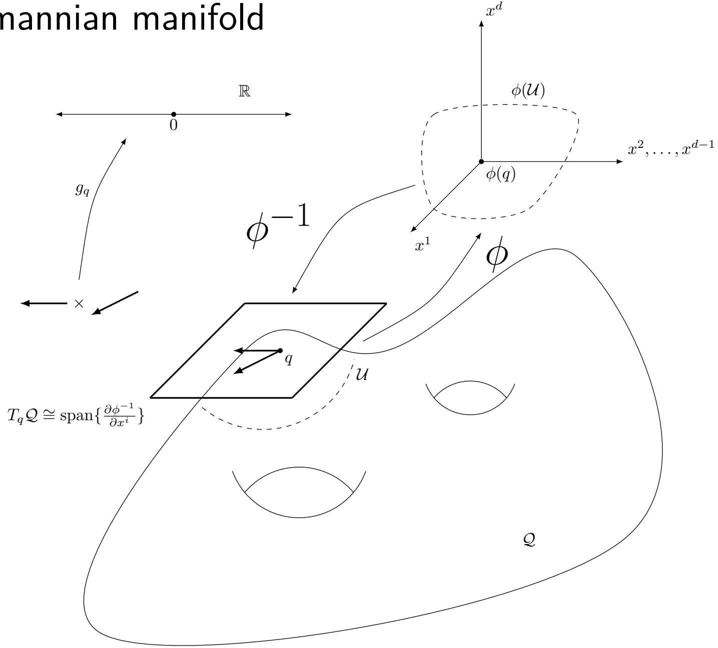
Part 3. A Geometric digression



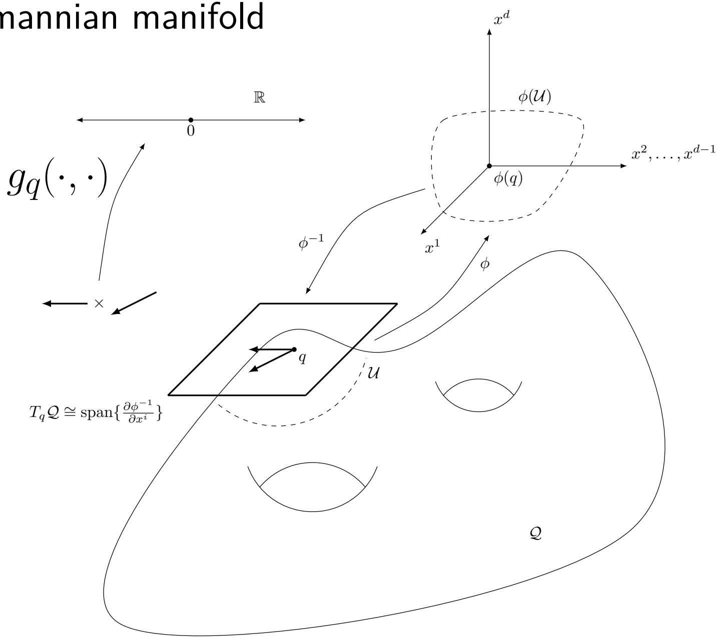
A Riemannian manifold



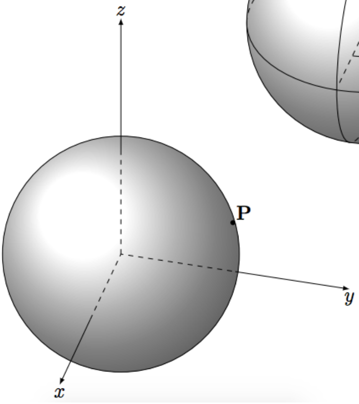
A Riemannian manifold

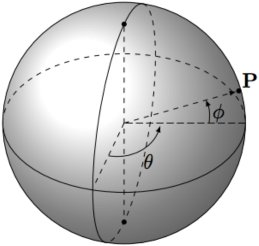


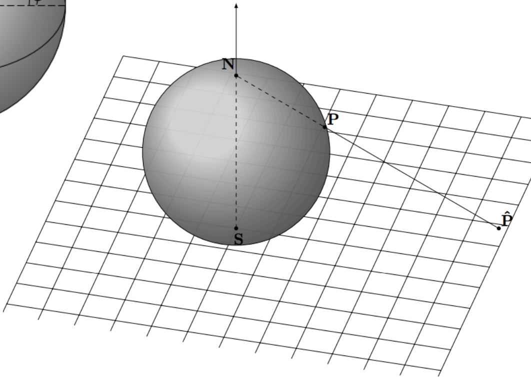
A Riemannian manifold



Pick your parameterization





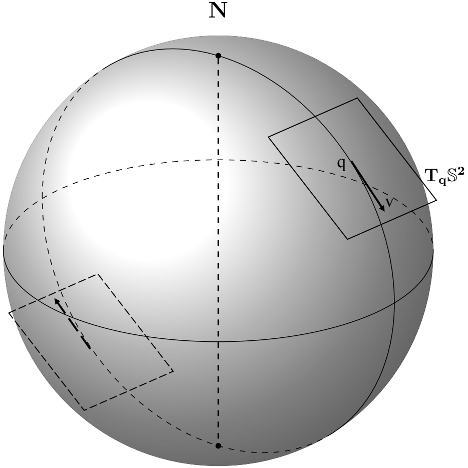


Non-Euclidean Hamiltonian Monte Carlo

Byrne and Girolami (2013) extend HMC to Riemannian manifolds using *isometric embeddings*.

Gaussian velocities are generated on tangent space at current position.

Geodesics traverse the state space.



Cartan's canonical metric

One may endow the smooth manifold of positive definite Hermitian matrices with the Riemannian metric

$$g_\Sigma(V_1, V_2) = \mathrm{tr}\left(\Sigma^{-1} V_1 \Sigma^{-1} V_2\right) ,$$

for V_1 and V_2 any Hermitian matrices. Under this metric, the geodesic equations are

$$\Sigma(t) = \Sigma(0)^{1/2} \exp\left(t\,\Sigma(0)^{-1/2} V(0) \Sigma(0)^{-1/2}\right) \Sigma(0)^{1/2}$$
$$V(t) = V(0)\Sigma(0)^{-1/2} \exp\left(t\Sigma(0)^{-1/2} V(0) \Sigma(0)^{-1/2}\right) \Sigma(0)^{1/2} ,$$

but the *isometric embedding* is unknown.

Algorithm 1 Geodesic Lagrangian Monte Carlo

Let $q = q^{(k)}$ be the kth state of the Markov chain. The next sample is generated according to the following procedure.

(a) Generate proposal state q^*:

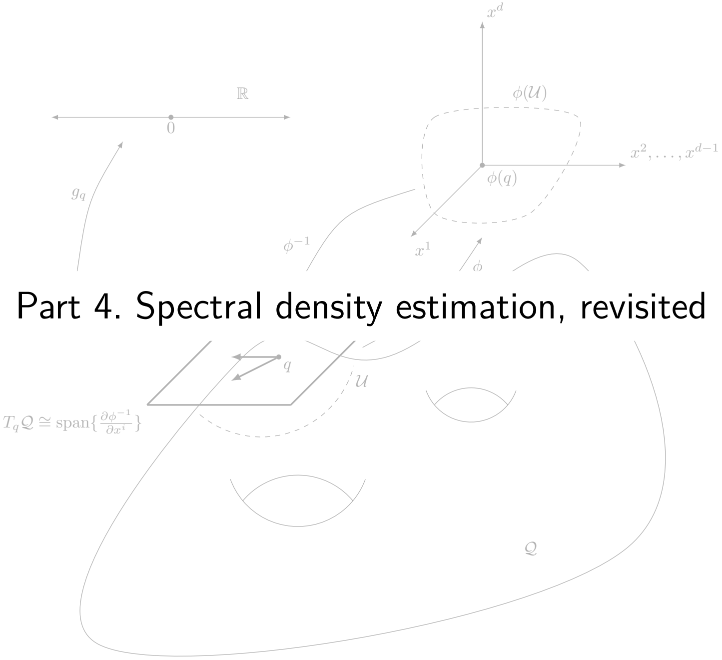
$v \sim N(0, G^{-1}(q))$
$e \leftarrow -\log \pi(q) - \frac{1}{2}\log|G(q)| + \frac{1}{2}v^\top \boxed{G(q)} v$
$q^* \leftarrow q$
for $\tau = 1, \ldots, T$ **do**
  $v \leftarrow v + \frac{\epsilon}{2} G(q^*)^{-1} \nabla_q \left(\log \pi(q^*) + \frac{1}{2}\log|G(q^*)|\right)$
  $\boxed{\text{Progress } (q^*, v) \text{ along the geodesic flow for time } \epsilon.}$
  $v \leftarrow v + \frac{\epsilon}{2} G(q^*)^{-1} \nabla_q \left(\log \pi(q^*) + \frac{1}{2}\log|G(q^*)|\right)$
end for
$e^* \leftarrow -\log \pi(q^*) - \frac{1}{2}\log|G(q^*)| + \frac{1}{2}v^\top G(q^*) v$

(b) Accept proposal with probability $\min\{1, \exp(e)/\exp(e^*)\}$:

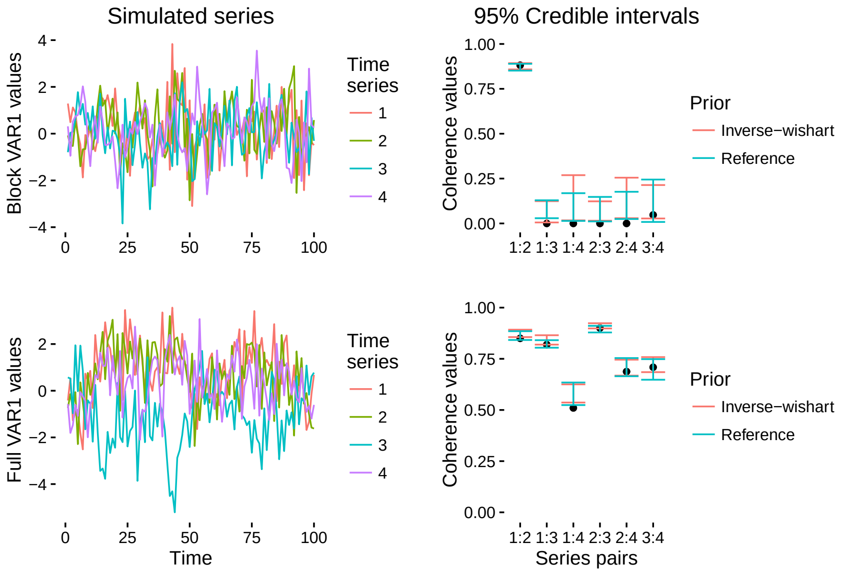
$u \sim U(0, 1)$
if $u < \exp(e - e^*)$ **then**
  $q \leftarrow q^*$
end if

(c) Assign value q to $q^{(k+1)}$, the $(k+1)$th state of the Markov chain.

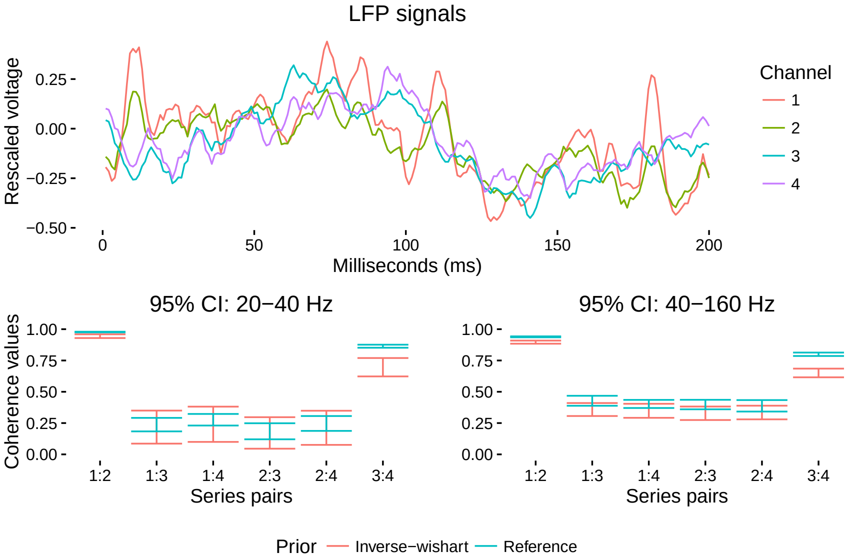
Part 4. Spectral density estimation, revisited



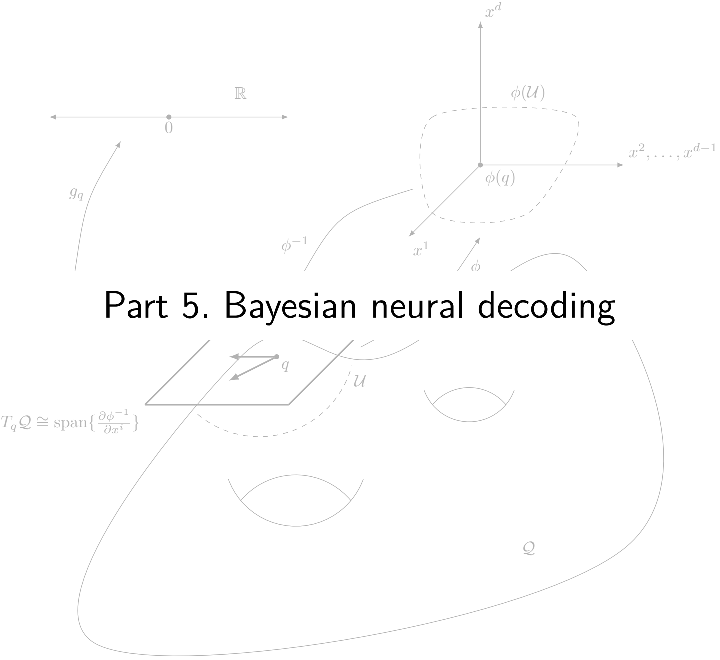
Simulated processes



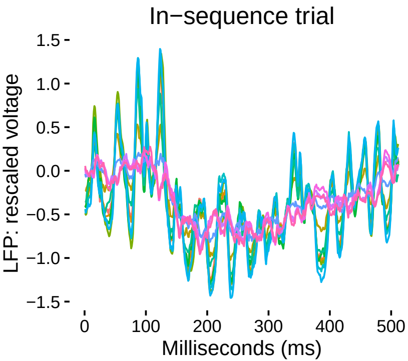
Local field potentials



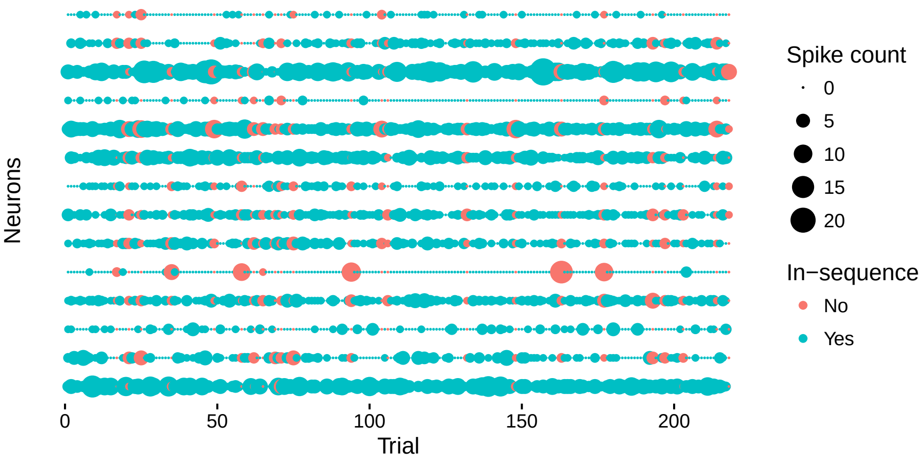
Part 5. Bayesian neural decoding



Local field potentials



Neural spike trains



sDDR: highly structured inference

LFP module:

$$\tilde{x}_i^L = U^L \Lambda^L z_i^L + \mu^L + \epsilon_i^L,$$
$$\epsilon_i^L \sim N_d(0, \sigma_L^2 I), \qquad \mu^L \sim N_d(0, \tau_L^2 I),$$
$$z_i^L \sim N_k(0, I), \qquad \sigma_L^2, \tau_L^2, \lambda_j^L \sim \text{Cauchy}^+(0, 5),$$
$$j = 1, \ldots, k, \quad \lambda_j^L > \lambda_{j'}^L, \quad j > j'.$$

Neural spikes module:

$$x_i^S \sim \text{Pois}_{\otimes}\big(\exp\{ U^S \Lambda^S z_i^S + \mu^S \}\big),$$
$$\mu^S \sim N_d(0, \tau_S^2 I), \qquad z_i^S \sim N_k(0, I)$$
$$\tau_S^2, \lambda_j^S \sim \text{Cauchy}^+(0, 5),$$
$$j = 1, \ldots, k, \quad \lambda_j^S > \lambda_{j'}^S, \quad j > j'.$$

Sequential classification module:

$$y_i \sim \text{Bernoulli}\big(\text{logit}^{-1}(\beta + \beta_S^T z_i^S + \beta_L^T z_i^L)\big) \ ,$$
$$\beta \sim N(0, 10^2), \quad \beta_S, \beta_L \sim N_k(0, 10^2 I) \ .$$

Two loading matrix models

$$L_{d\times k} = U_{d\times k}\,\Lambda_{k\times k} = U_{d\times k}\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_k \end{pmatrix}$$

Model 1:

$$U \sim Uni_{\mathcal{H}}\left(\mathcal{O}_{d\times k}\right) \quad \Longrightarrow \quad U^T U = I_k$$

Model 2:

$$U_{ij} \overset{iid}{\sim} N(0,1) \quad \Longrightarrow \quad E_{p(U)}\left(U^T U\right) \propto I_k$$

Predictive fit

Table 2: 10-Fold cross-validation results

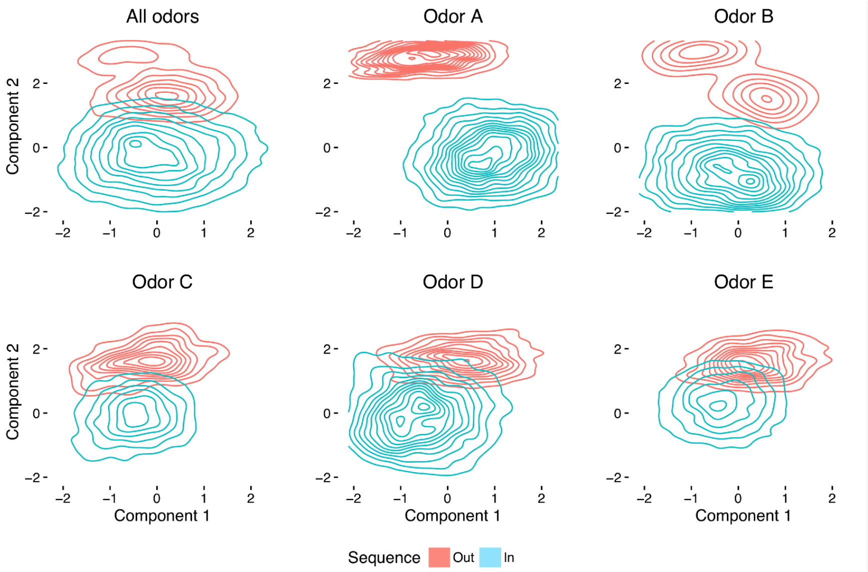
Method	0-1 Error		
	LFP	Spikes	Joint
sDDR, Gaussian	0.110	0.064	0.060
sDDR, Stiefel	0.106	0.069	0.064
Logistic lasso	0.106	0.092	0.087
Random forest	0.106	0.096	0.106
PLS-DA	0.106	0.073	0.096

Predictive fit

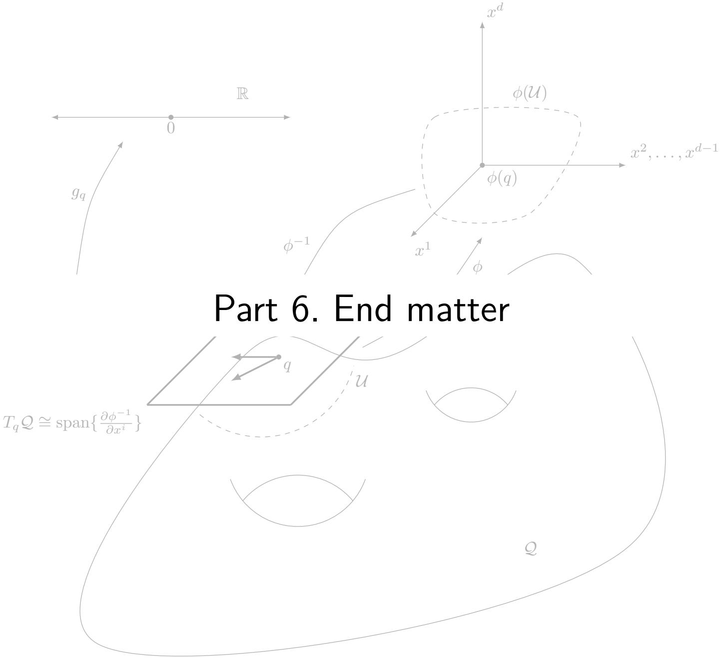
Table 2: 10-Fold cross-validation results

Method	0-1 Error		
	LFP	Spikes	Joint
sDDR, Gaussian	0.110	0.064	0.060
sDDR, Stiefel	0.106	0.069	0.064
Logistic lasso	0.106	0.092	0.087
Random forest	0.106	0.096	0.106
PLS-DA	0.106	0.073	0.096

The Information bottleneck



Part 6. End matter



Takeaways

- Theory *directly* contributes to applied science.

- Differential geometry enables basic/fundamental/categorical improvements in Bayesian inference.

- Geometric components seamlessly integrate into advanced/hierarchical models.

- *Much* work to do in Bayesian spectral analysis.

There's more!

- details
- information geometry and nonparametric density estimation
- $\nabla \mathrm{Det}(A) = A^{+} \mathrm{Det}(A)$
- geodesic Monte Carlo with non-trivial mass matrix
- inference on infinite dimensional manifolds

Many thanks to ...

- my thesis advisor, Prof. Shahbaba; my non-thesis advisor, Prof. Gillen; my mentor, Prof. Ombao;
- Alexander Vandenberg-Rodes; Shiwei Lan; Jeffrey Streets; Norbert Fortin;
- the Savage Award committee;
- Prof. Daniels;
- and ...

Thank you, Super Chris.