Inference in dynamical systems and the geometry of learning group actions

Geometry and Topology of Data

Sayan Mukherjee

Departments of Statistical Science, Mathematics, Computer Science, Biostatistics & Bioinformatics
Duke University
https://sayanmuk.github.io/

Joint work with:

Dynamical systems — K. McGoff (UNC Ch) | A. Nobel (UNC CH)
Group actions — T. Gao (U Chicago) | J. Brodzki (U Southampton)
Geometry of learning group actions
Modeling variation in shapes

S. J. Gould
From distances to trees

Constructing a Simple Phylogenetic Tree

1. Study groups are compared to outgroup.
2. First node symbolizes common ancestor shared by outgroup and A to D.
   Second node drawn to symbolize common ancestor shared by A to D.
3. Additional branch points are added. Group D has the unique character 1 which is not shared by A to C.
4. Character 5 is shared by all of the study groups but not the outgroup.

Adapted from Campbell "Biology" 4th Edition
50 molars from 5 primate genera
5 primate genera

Spider monkey

Howler Monkey

Squirrel Monkey

Black handed spider monkey

Titi monkey
Parallel transport in a shape space

*Red circle is around Entoconid*
Evolution as broccoli
Synchronization

Given a sequence of objects \( \{o_1, \ldots, o_n\} \) and a group \( G \) learn a collection of group elements \( \rho_{ij} \) that transform object \( o_i \) to \( o_j \).

If it is possible to find a sequence of group elements that allow for an accurate transformation between objects then this set can be synchronized.
Geometry and cohomology of synchronization

- Geometry: Holonomy and fiber bundle structures in synchronization problems;

- Cohomology: Twisted de Rham theory and twisted Laplacian for synchronization;

- Learning group actions algorithm (SynCut);

- Classify primate molars to uncover dietary habits.
Problem setup

\[ \Gamma = (V, E, w): \text{vertex set } V, \text{ edge set } E, \text{ and weights } w_{ij}. \]
Problem setup

$\Gamma = (V, E, w)$: vertex set $V$, edge set $E$, and weights $w_{ij}$.

$G$ is a topological group acting on a normed vector space $F$. 
Problem setup

\[ \Gamma = (V, E, w) \] vertex set \( V \), edge set \( E \), and weights \( w_{ij} \).

\( G \) is a topological group acting on a normed vector space \( F \).

edge potential – \( \rho : E \to G \) with \( \rho_{ij} = \rho_{ji}^{-1} \).

vertex potential – \( f : V \to F \)
Problem setup

$\Gamma = (V, E, w)$: vertex set $V$, edge set $E$, and weights $w_{ij}$.

$G$ is a topological group acting on a normed vector space $F$.

edge potential – $\rho: E \to G$ with $\rho_{ij} = \rho_{ji}^{-1}$.

vertex potential – $f: V \to F$

Goal: given $\rho$ find

$$f_i = \rho_{ij} f_j \quad \forall i \sim j,$$
Problem setup

\( \Gamma = (V, E, w) \): vertex set \( V \), edge set \( E \), and weights \( w_{ij} \).

\( G \) is a topological group acting on a normed vector space \( F \).

Edge potential – \( \rho : E \to G \) with \( \rho_{ij} = \rho_{ji}^{-1} \).

Vertex potential – \( f : V \to F \)

Goal: given \( \rho \) find

\[
f_i = \rho_{ij} f_j \quad \forall i \sim j,
\]

or

\[
\eta = \min_{\substack{f : V \to G \\ ||f|| \neq 0}} \frac{1}{2} \sum_{i,j \in V} w_{ij} \|f_i - \rho_{ij} f_j\|^2_F,
\]

\[
d_i = \sum_{j \sim i} w_{ij}.
\]
Vertex and edge potentials

Let $\Gamma = (V, E)$ be a graph with vertex set $V$ and edge set $E$. Let $G$ be a group.
Vertex and edge potentials

Let $\Gamma = (V, E)$ be a graph with vertex set $V$ and edge set $E$. Let $G$ be a group.

- A vertex potential is a map $f : V \rightarrow G$.
- An edge potential is a function $\rho : E \rightarrow G$ where

  \[ \rho_{ji} = \rho_{ij}^{-1} \quad \forall i \sim j \]
Vertex and edge potentials

$G$-valued 0- and 1-cochains

vertex potentials: $C^0 (\Gamma; G) := \{ f : V \to G \}$

edge potentials: $C^1 (\Gamma; G) := \left\{ \rho : E \to G \mid \rho_{ij} = \rho_{ji}^{-1}, \forall i \sim j \right\}.$
Vertex and edge potentials

$G$-valued 0- and 1-cochains

vertex potentials: $C^0(\Gamma; G) := \{f : V \to G\}$

dge potentials: $C^1(\Gamma; G) := \left\{ \rho : E \to G \mid \rho_{ij} = \rho_{ji}^{-1}, \forall i \sim j \right\}$.

Typically $F = G$ but one can decouple.
Fibre bundle framework
The Geometry of Synchronization Problems

- **Data:**
  - graph $\Gamma = (V, E)$
  - linear algebraic group $G$, equipped with a norm $\|\cdot\|$,
  - **edge potential** $\rho : E \to G$ satisfying $\rho_{ij} = \rho_{ji}^{-1}$, $\forall (i, j) \in E$
The Geometry of Synchronization Problems

- **Data:**
  - graph \( \Gamma = (V, E) \)
  - linear algebraic group \( G \), equipped with a norm \( || \cdot || \)
  - edge potential \( \rho : E \rightarrow G \) satisfying \( \rho_{ij} = \rho_{ji}^{-1}, \forall (i, j) \in E \)

- **Observation:**
  - Let \( \mathcal{U} = \{ U_i \mid 1 \leq i \leq |V| \} \) be an open cover of \( \Gamma \) (viewed as a 1-dimensional simplicial complex), where \( U_i \) is the (open) star neighborhood of vertex \( i \).

- The \( \rho \) defines a flat principal \( G \)-bundle over \( \Gamma \) (denoted as \( B_\rho \)).
Fibre bundles

Fibre Bundle $\mathcal{E} = (E, M, F, \pi)$

- $M$: base manifold
- $F$: fibre manifold
- $E$: total manifold
- $\pi : E \to M$: smooth surjective map (bundle projection)
- local triviality: for “small” open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$
Fibre bundles

Fibre Bundle $\mathcal{E} = (E, M, F, \pi)$

- $M$: base manifold
- $F$: fibre manifold
- $E$: total manifold
- $\pi : E \to M$: smooth surjective map (bundle projection)
- local triviality: for “small” open set $U \subset M$, $\pi^{-1}(U)$ is diffeomorphic to $U \times F$

Commutative diagram

\[
\begin{array}{ccc}
\pi^{-1}(U_i) & \xrightarrow{\phi_i} & U_i \times F \\
\downarrow{\pi} & & \downarrow{\text{Proj}_1} \\
U_i & &
\end{array}
\]
Theorem (Steenrod 1951, §2). If topological group $G$ acts on $Y$ and \{${U_i}$\}, \{${\rho_{ij}}$\} is a system of coordinate transformations in the space $X$ such that

\[
\rho_{ii} = e \in G \quad \text{for all } U_i
\]
\[
\rho_{ij} = \rho_{ji}^{-1} \quad \text{if } U_i \cap U_j \neq \emptyset
\]
\[
\rho_{ij} \rho_{jk} = \rho_{ik} \quad \text{if } U_i \cap U_j \cap U_k \neq \emptyset
\]

then there exists a fibre bundle $\mathcal{B}$ with base space $X$, fibre $Y$, group $G$, and coordinate transforms \{${\rho_{ij}}$\}. 

Theorem (Steenrod 1951, §2). If topological group $G$ acts on $Y$ and \{${U_i}$, \{${\rho_{ij}}$\} is a system of coordinate transformations in the space $X$ satisfying the cycle-consistency conditions then there exists a fibre bundle $B$ with base space $X$, fibre $Y$, group $G$, and coordinate transforms \{${\rho_{ij}}$\}. In other words we have constructed a bundle with fibre $G$ over each vertex of the graph. The graph can be regarded as the nerve of the cover in the above theorem.
Fibre bundles and consistency conditions

**Theorem** (Steenrod 1951, §2). If topological group $G$ acts on $Y$ and $\{U_i\}$, $\{\rho_{ij}\}$ is a system of coordinate transformations in the space $X$ satisfying the cycle-consistency conditions then there exists a fibre bundle $B$ with base space $X$, fibre $Y$, group $G$, and coordinate transforms $\{\rho_{ij}\}$.

In other words we have constructed a bundle with fibre $G$ over each vertex of the graph. The graph can be regarded as the nerve of the cover in the above theorem.
**Consistent Shape Maps via Semidefinite Programming**

Qi-Xing Huang and Leonidas Guibas

Computer Science Department, Stanford University, Stanford, CA

**Definition 2.1** Given a shape collection $S = \{S_1, \cdots, S_n\}$ of $n$ shapes where each shape consists of the same number of samples, we say a map collection $\Phi = \{\phi_{ij} : S_i \to S_j | 1 \leq i, j \leq n\}$ of maps between all pairs of shapes is cycle consistent if and only if the following equalities are satisfied:

\[
\phi_{ii} = id_{S_i}, \quad 1 \leq i \leq n, \quad \text{(1-cycle)}
\]
\[
\phi_{ji} \circ \phi_{ij} = id_{S_i}, \quad 1 \leq i < j \leq n, \quad \text{(2-cycle)}
\]
\[
\phi_{ki} \circ \phi_{jk} \circ \phi_{ij} = id_{S_i}, \quad 1 \leq i < j < k \leq n, \quad \text{(3-cycle) (1)}
\]

where $id_{S_i}$ denotes the identity self-map on $S_i$. 
Geometric observations

- Denote

\[
C^0 (\Gamma; G) := \{ f : V \rightarrow G \}
\]
\[
C^1 (\Gamma; G) := \left\{ \rho : E \rightarrow G \mid \rho_{ij} = \rho_{ji}^{-1}, \forall i \sim j \right\}
\]
Geometric observations

- Denote

\[
C^0(\Gamma; G) := \{ f : V \rightarrow G \}
\]

\[
C^1(\Gamma; G) := \left\{ \rho : E \rightarrow G \mid \rho_{ij} = \rho_{ji}^{-1}, \forall i \sim j \right\}
\]

- Consider the right action of \( C^0(\Gamma; G) \) on \( C^1(\Gamma; G) \):

\[
C^1(\Gamma; G) \times C^0(\Gamma; G) \rightarrow C^1(\Gamma; G)
\]

\[
(\rho, f) \mapsto \tau_\rho f
\]

defined as \( (\tau_\rho f)_{ij} := f_i^{-1} \rho_{ij} f_j, \quad \forall i \sim j \).
Geometric observations

- Denote

  \[ C^0 (\Gamma; G) := \{ f : V \to G \} \]
  \[ C^1 (\Gamma; G) := \left\{ \rho : E \to G \mid \rho_{ij} = \rho_{ji}^{-1}, \forall i \sim j \right\} \]

- Consider the right action of \( C^0 (\Gamma; G) \) on \( C^1 (\Gamma; G) \):

  \[ C^1 (\Gamma; G) \times C^0 (\Gamma; G) \to C^1 (\Gamma; G) \]
  \[ (\rho, f) \mapsto \tau_\rho f \]

  defined as \((\tau_\rho f)_{ij} := f_i^{-1} \rho_{ij} f_j, \quad \forall i \sim j\).

- \( \rho \) synchronizable \( \iff \tau_\rho f \) synchronizable for all \( f \in C^0 (\Gamma; G) \), i.e. synchronizability is defined at the level of equivalence classes \( C^1 (\Gamma; G) / C^0 (\Gamma; G) \)
Holonomy and synchronization
Paths

A path $\gamma$ connecting the vertices $v$ and $w$ is a sequence of edges

$$\gamma = (e_1, e_2, \ldots, e_n),$$

$$\gamma^{-1} = (e_n, e_{n-1}, \ldots, e_1).$$
Paths

A path $\gamma$ connecting the vertices $v$ and $w$ is a sequence of edges

$$\gamma = (e_1, e_2, \ldots, e_n),$$

$$\gamma^{-1} = (e_n, e_{n-1}, \ldots, e_1).$$

Maps from edges to group elements

$$\text{hol}_\rho (\gamma) = \rho_{i_0,1} \rho_{i_1,i_2} \cdots \rho_{i_{N-2},i_{N-1}} \rho_{i_{N-1},i_N} \in G.$$ 

$$\text{hol}_\rho (\gamma^{-1}) = \text{hol}_\rho (\gamma)^{-1}, \quad \text{hol}_\rho (\gamma \circ \gamma') = \text{hol}_\rho (\gamma) \text{hol}_\rho (\gamma').$$
Corollary (Gao-Brodzki-M, 2016)

For a connected graph $\Gamma$ and topological group $G$, an edge potential $\rho \in C^1(\Gamma; G)$ is synchronizable if and only if $\text{Hol}_\rho(\Gamma)$ is trivial.
Holonomy and synchronization

Corollary (Gao-Brodzki-M, 2016)

For a connected graph $\Gamma$ and topological group $G$, an edge potential $\rho \in C^1(\Gamma; G)$ is synchronizable if and only if $\text{Hol}_\rho(\Gamma)$ is trivial.

Theorem (Gao-Brodzki-M, 2016)

There exists a one-to-one correspondence between sets

$$C^1(\Gamma; G)/C^0(\Gamma; G) \cong \text{Hom}(\pi_1(\Gamma), G)/G$$

where $G$ acts on $\text{Hom}(\pi_1(\Gamma), G)$ by conjugations. Also, $\text{Hom}(\pi_1(\Gamma), G)/G$ is in one-to-one correspondence with equivalence classes of flat principal $G$-bundles $\mathcal{B}_\rho$ defined by $\rho \in C^1(\Gamma; G)$. 
Obstruction to holonomy
Quantification of obstructions

Let $F$ be a vector space such that $G \subseteq GL(F)$, and $f \in C^0(\Gamma; F)$. 

The frustration of a function $f$ is $
abla(f) = \frac{1}{2} \sum_{i,j \in V} k_{ij} f_i f_j$

where $L_1$ is the Graph Connection Laplacian (GCL) $L_1 f_i = \frac{1}{\deg(i)} \sum_{j \in \text{in}(i)} (f_i \nabla f_j)$. 


Quantification of obstructions

Let $F$ be a vector space such that $G \subset GL(F)$, and $f \in C^0(\Gamma; F)$.

The frustration of a function $f$

$$\eta(f) = \frac{1}{2} \frac{\sum_{i,j \in V} \| f_i - \rho_{ij} f_j \|^2}{\sum_{i \in V} \| f_i \|^2},$$

can be written as a Raleigh quotient

$$\eta(f) = \frac{\langle f, L_1 f \rangle}{\langle f, f \rangle},$$

where $L_1$ is the **Graph Connection Laplacian** (GCL)

$$(L_1 f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - \rho_{ij} f_j)$$
Discrete Hodge theory

\[ \Gamma = (V, E): \]

\[ \Omega^0 (\Gamma) := \{ f : V \to \mathbb{K} \}, \]
\[ \Omega^1 (\Gamma) := \{ \omega : E \to \mathbb{K} \mid \omega_{ij} = -\omega_{ji} \ \forall i \sim j \}, \]
Discrete Hodge theory

\[ \Gamma = (V, E) : \]

\[ \Omega^0 (\Gamma) := \{ f : V \to \mathbb{K} \} , \]
\[ \Omega^1 (\Gamma) := \{ \omega : E \to \mathbb{K} \mid \omega_{ij} = -\omega_{ji} \; \forall i \sim j \} , \]

Define cochain complex

\[ 0 \quad \overset{\Rightarrow}{\leftrightarrow} \quad \Omega^0 (\Gamma) \overset{d}{\underset{\delta}{\leftrightarrow}} \Omega^1 (\Gamma) \overset{\Rightarrow}{\leftrightarrow} 0, \]

where

\[ (df)_{ij} = f_i - f_j , \quad \forall f \in \Omega^0 (\Gamma) , \]
\[ (\delta \omega)_i = \frac{1}{\deg (i)} \sum_{j \sim i} \omega_{ij} , \quad \forall \omega \in \Omega^1 (\Gamma) . \]
Discrete Hodge theory

Define cochain complex

\[ 0 \leftrightarrow \Omega^0(\Gamma) \xrightarrow{d} \Omega^1(\Gamma) \xrightarrow{\delta} 0, \]

where

\[(df)_{ij} = f_i - f_j, \quad \forall f \in \Omega^0(\Gamma), \]

\[(\delta \omega)_i = \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in \Omega^1(\Gamma). \]

Then

\[(L_0 f)_i := (\delta df)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - f_j) \quad \forall i \in V, \forall f \in \Omega^0(\Gamma). \]
Twisted De Rham-Hodge Theory

\[(L_0f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - f_j), \quad \forall f : V \to \mathbb{K}\]

\[(L_1f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - \rho_{ij} f_j) \quad \forall f : V \to F\]
Twisted De Rham-Hodge Theory

\[(L_0 f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - f_j), \quad \forall f : V \to \mathbb{K}\]

\[(L_1 f)_i = \frac{1}{\deg(i)} \sum_{j \sim i} (f_i - \rho_{ij} f_j), \quad \forall f : V \to F\]

Naïvely:

\[(d_\rho f)_{ij} = f_i - \rho_{ij} f_j, \quad \forall f \in C^0(\Gamma; F)\]

\[(\delta_\rho \omega)_i = \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in C^1(\Gamma; F)\]

then \(L_1 = \delta_\rho d_\rho\). There is a problem
\[(d_\rho f)_{ij} \sim f_i - \rho_{ij} f_j, \quad \forall f \in C^0(\Gamma; F)\]
\[(\delta_\rho \omega)_i \sim \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in C^1(\Gamma; F)\]

**Issue:** \(d_\rho\) does not map into \(C^1(\Gamma; F)\) (no skew-symmetry).

\[f_j - \rho_{ji} f_j = -\rho_{ji} (f_i - \rho_{ij} f_j) \neq -(f_i - \rho_{ij} f_j).\]
\[(d_\rho f)_{ij} \sim f_i - \rho_{ij}f_j, \quad \forall f \in C^0(\Gamma; F)\]

\[(\delta_\rho \omega)_i \sim \frac{1}{\deg(i)} \sum_{j \sim i} \omega_{ij}, \quad \forall \omega \in C^1(\Gamma; F)\]

**Issue:** $d_\rho$ does not map into $C^1(\Gamma; F)$ (no skew-symmetry).

\[f_j - \rho_{ji}f_j = -\rho_{ji}(f_i - \rho_{ij}f_j) \neq -(f_i - \rho_{ij}f_j)\]

**Fix:** Interpret $f_i - \rho_{ij}f_j$ as the “local expression” of $(d_\rho f)_{ij}$ in a local trivialization over $\mathcal{U} = \{U_i \mid 1 \leq i \leq |V|\}$ of the associated $F$-bundle of $\mathcal{B}_\rho$, denoted as $\mathcal{B}_\rho [F]$, such that the extra $\rho_{ji}$ factor encodes a bundle transformation from $U_i$ to $U_j$. 
Twisted De Rham-Hodge Theory

- Combinatorial Hodge Theory:

\[ 0 \leftrightarrow \Omega^0 (\Gamma) \xrightarrow{d} \Omega^1 (\Gamma) \leftrightarrow 0, \]

- Twisted Combinatorial Hodge Theory:

\[ 0 \leftrightarrow C^0 (\Gamma; F) \xrightarrow{d_\rho} \Omega^1 (\Gamma; \mathcal{B}_\rho [F]) \leftrightarrow 0. \]
Twisted De Rham-Hodge Theory

- Combinatorial Hodge Theory:

\[
0 \longleftrightarrow \Omega^0 (\Gamma) \xrightarrow{d} \Omega^1 (\Gamma) \xrightarrow{\delta} 0,
\]

- Twisted Combinatorial Hodge Theory:

\[
0 \longleftrightarrow C^0 (\Gamma; F) \xrightarrow{d_\rho} \Omega^1 (\Gamma; \mathcal{B}_\rho [F]) \xrightarrow{\delta_\rho} 0.
\]

Theorem (Gao, Brodzki, M (2016)))

Define

\[
\Delta^{(0)}_\rho := \delta_\rho d_\rho, \quad \Delta^{(1)}_\rho := d_\rho \delta_\rho
\]

then the following Hodge-type decomposition holds:

\[
C^0 (\Gamma; F) = \ker \Delta^{(0)}_\rho \oplus \im \delta_\rho = \ker d_\rho \oplus \im \delta_\rho,
\]

\[
\Omega^1 (\Gamma; \mathcal{B}_\rho [F]) = \im d_\rho \oplus \ker \Delta^{(1)}_\rho = \im d_\rho \oplus \ker \delta_\rho.
\]
Recovering the GCL

An $O(d)$-valued edge potential $\xi \in C^1(\Gamma; O(d))$ is synchronizable if and only if there exists $g \in C^0(\Gamma; O(d))$ such that $\xi_{ij} = g_i g_j^{-1}$ for all $i \sim j$. 
Recovering the GCL

An $O(d)$-valued edge potential $\xi \in C^1(\Gamma; O(d))$ is synchronizable if and only if there exists $g \in C^0(\Gamma; O(d))$ such that $\xi_{ij} = g_i g_j^{-1}$ for all $i \sim j$.

Define frustration

$$\nu(\Gamma) = \frac{1}{2d \text{ vol}(\Gamma)} \inf_{g \in C^0(\Gamma; O(d))} \sum_{i,j \in V} w_{ij} \left\| g_i g_j^{-1} - \rho_{ij} \right\|_F^2$$

$$= \frac{1}{2d \text{ vol}(\Gamma)} \inf_{\xi \in C^1_{\text{sync}}(\Gamma; O(d))} \sum_{i,j \in V} w_{ij} \left\| \xi_{ij} - \rho_{ij} \right\|_F^2,$$

where we define

$$C^1_{\text{sync}}(\Gamma; O(d)) := \left\{ \xi \in C^1(\Gamma; O(d)) \mid \xi \text{ synchronizable} \right\}$$

$$= \left\{ \xi \in C^1(\Gamma; O(d)) \mid \text{Hol}_\xi(\Gamma) \text{ is trivial} \right\}.$$
Find cochains in \( C^0 (\Gamma; \mathbb{R}^d) \) “closest to” a global frame of \( \mathcal{B}_\rho [\mathbb{R}^d] \)

\[
\eta(f) = \frac{\langle f, \Delta^{(0)} \rho f \rangle}{\|f\|^2} = \frac{1}{2} \sum_{i,j \in V} w_{ij} \| f_i - \rho_{ij} f_j \|^2 \\
= \frac{1}{2 \text{vol}(\Gamma)} [f]^\top L_1 [f], \quad \forall f \in C^0 (\Gamma; \mathbb{S}^{d-1}), \; \|f\| \neq 0.
\]
Learning group actions
Problem (Learning group actions)

Given a group $G$ acting on a set $X$, simultaneously learn actions of $G$ on $X$ and a partition of $X$ into disjoint subsets $X_1, \ldots, X_K$.

Each action is cycle-consistent on each $X_i$ ($1 \leq i \leq K$).
Learning group actions
Problem (Learning group actions by synchronization)

Denote $\mathcal{X}_K$ for all partitions of $\Gamma$ into $K$ nonempty connected subgroups ($K \leq n$) and

$$\nu(S_i) = \min_{f \in C^0(\Gamma;G)} \left\{ \sum_{j,k \in S_i} w_{jk} \text{Cost}_G(f_j, \rho_{jk} f_k) \right\}, \quad \text{vol}(S_i) = \sum_{j \in S_i} d_j,$$

Solve the optimization problem

$$\min_{\{S_1, \ldots, S_K\} \in \mathcal{X}_K} \left\{ \max_{1 \leq i \leq K} \nu(S_i) \right\} \quad \frac{\min_{1 \leq i \leq K} \text{vol}(S_i)}{\min_{1 \leq i \leq K} \text{vol}(S_i)}$$

(1)

and output an optimal partition $\{S_1, \ldots, S_K\}$ together with the minimizing $f \in C^0(\Gamma;G)$. 

Algorithm: SynCut

Input: $\Gamma = (V, E, w)$, $\rho \in C^1(\Gamma; G)$, number of partitions $K$
Output: Partitions $\{S_1, \cdots, S_K\}$

1. Solve synchronization problem over $\Gamma$ for $\rho$, obtain $f \in C^0(\Gamma; G)$
Algorithm: SynCut

Input: $\Gamma = (V, E, w)$, $\rho \in C^1 (\Gamma; G)$, number of partitions $K$
Output: Partitions $\{S_1, \cdots, S_K\}$

1. Solve synchronization problem over $\Gamma$ for $\rho$, obtain $f \in C^0 (\Gamma; G)$

2. Compute $d_{ij} = \exp (-w_{ij} \| f_i - \rho_{ij} f_j \|)$ on all edges $(i, j) \in E$
Algorithm: SynCut

Input: $\Gamma = (V, E, w)$, $\rho \in C^1(\Gamma; G)$, number of partitions $K$
Output: Partitions $\{S_1, \cdots, S_K\}$

1. Solve synchronization problem over $\Gamma$ for $\rho$, obtain $f \in C^0(\Gamma; G)$
2. Compute $d_{ij} = \exp(-w_{ij}\|f_i - \rho_{ij}f_j\|)$ on all edges $(i, j) \in E$
3. Spectral clustering on weighted graph $(V, E, d)$ to get $\{S_1, \cdots, S_k\}$
Algorithm: SynCut

Input: $\Gamma = (V, E, w)$, $\rho \in C^1(\Gamma; G)$, number of partitions $K$
Output: Partitions $\{S_1, \cdots, S_K\}$

1. Solve synchronization problem over $\Gamma$ for $\rho$, obtain $f \in C^0(\Gamma; G)$
2. Compute $d_{ij} = \exp(-w_{ij}||f_i - \rho_{ij}f_j||)$ on all edges $(i, j) \in E$
3. Spectral clustering on weighted graph $(V, E, d)$ to get $\{S_1, \cdots, S_k\}$
4. Solve synchronization problem within each partition $S_j$, “glue up” the local solutions to obtain $f_* \in C^0(\Gamma; G)$
Algorithm: SynCut

Input: $\Gamma = (V, E, w)$, $\rho \in C^1(\Gamma; G)$, number of partitions $K$
Output: Partitions $\{S_1, \cdots, S_K\}$

1. Solve synchronization problem over $\Gamma$ for $\rho$, obtain $f \in C^0(\Gamma; G)$
2. Compute $d_{ij} = \exp(-w_{ij} \| f_i - \rho_{ij} f_j \|)$ on all edges $(i, j) \in E$
3. Spectral clustering on weighted graph $(V, E, d)$ to get $\{S_1, \cdots, S_k\}$
4. Solve synchronization problem within each partition $S_j$, “glue up” the local solutions to obtain $f_* \in C^0(\Gamma; G)$
5. $f \leftarrow f_*$, repeat from Step 2
Dietary habits of primates
Geometric morphometrics

second mandibular molar of a Philippine flying lemur

Philippine flying lemur (*Cynocephalus volans*)
Geometric morphometrics

- Manually put $k$ landmarks $p_1, p_2, \ldots, p_k$

second mandibular molar of a Philippine flying lemur
Geometric morphometrics

- Manually put $k$ landmarks $p_1, p_2, \cdots, p_k$

- Use spatial coordinates of the landmarks as features $p_j = (x_j, y_j, z_j), \ j = 1, \cdots, k$

second mandibular molar of a Philippine flying lemur
Geometric morphometrics

- Manually put \( k \) landmarks \( p_1, p_2, \cdots, p_k \)

- Use spatial coordinates of the landmarks as features
  \[ p_j = (x_j, y_j, z_j), \quad j = 1, \cdots, k \]

- Represent a shape in \( \mathbb{R}^{3 \times k} \)

second mandibular molar of a Philippine flying lemur
Shape distances: automated landmarks

\[ d_{cWn}(S_1, S_2): \text{ Conformal Wasserstein Distance (CWD)} \]
\[ d_{cP}(S_1, S_2): \text{ Continuous Procrustes Distance (CPD)} \]
\[ d_{cKP}(S_1, S_2): \text{ Continuous Kantorovich-Procrustes Distance (CKPD)} \]

\[
d_{cP}(S_1, S_2) = \inf_{C \in A(S_1, S_2)} \inf_{R \in \mathbb{E}(3)} \left( \int_{S_1} \| R(x) - C(x) \|^2 \, d\text{vol}_{S_1}(x) \right)^{\frac{1}{2}}
\]
Continuous Procrustes distance

Define $\mathcal{A}(S, S')$ as the set of area preserving diffeomorphisms, maps $a : S \rightarrow S'$ such that for any measurable subset $\Omega \subset S$

$$\int_{\Omega} dA_S = \int_{a(\Omega)} dA_{S'}.$$
Continuous Procrustes distance

Define $\mathcal{A}(S, S')$ as the set of area preserving diffeomorphisms, maps $a : S \rightarrow S'$ such that for any measurable subset $\Omega \subset S$

$$
\int_{\Omega} dA_S = \int_{a(\Omega)} dA_{S'}.
$$

The continuous procrustes distance is

$$
d_p(S, S') = \inf_{a \in \mathcal{A}(S, S')} \min_{R \in \text{rigid motion}} \int_S |R(x) - a(x)|^2 dA_S.
$$
Continuous Procrustes distance

Define $A(S, S')$ as the set of area preserving diffeomorphisms, maps $a : S \rightarrow S'$ such that for any measurable subset $\Omega \subset S$

$$\int_{\Omega} dA_S = \int_{a(\Omega)} dA_{S'}.$$ 

The continuous procrustes distance is

$$d_p(S, S') = \inf_{a \in A(S, S')} \min_{R \in \text{rigid motion}} \int_S |R(x) - a(x)|^2 dA_S.$$ 

Near optimal $a$ are “almost” conformal, so simplify above to searching near conformal maps.
The actions: $G$

$$d_{cP}(S_i, S_j) = \inf_{\mathcal{C} \in \mathcal{A}(S_i, S_j)} \inf_{R \in \mathbb{R}^3} \left( \int_{S_i} \| R(x) - C(x) \|^2 \ d\text{vol}_{S_i}(x) \right)^{\frac{1}{2}}$$
50 molars from 5 primate genera
5 primate genera

- Spider monkey
- Howler Monkey
- Squirrel Monkey
- Black handed spider monkey
- Titi monkey
Folivorous, frugivorous, and insectivorous
Open questions

1. Use Hodge structure to design synchronization algorithms
2. Synchronization beyond the regime of linear algebraic groups
3. Statistical complexity of learning group actions
4. Random walks on fibre bundles
5. Provable algorithms
6. Synchronization on simplicial complexes
7. Bayesian model
Learning dynamical systems
We consider mathematical models of the form:

- $X$ is the “phase space”;
- $X_t$ is the “true state” of bioreactor (your stomach) at time $t$;
- $Y$ is the “observation space”;
- $Y_t$ is our observation at time $t$. 

We only have access to the observations $\{Y_t^k\}_{k=0}^n$. 

General framework
General framework

We consider mathematical models of the form:

- $X$ is the “phase space”;
- $X_t$ is the “true state” of bioreactor (your stomach) at time $t$;
- $Y$ is the “observation space”;
- $Y_t$ is our observation at time $t$.

We only have access to the observations $\{ Y_{t_k} \}_{k=0}^{n}$. 
General questions

Given access to the observations \( \{ Y_{tk} \}_{k=0}^{n} \), we might want to ask

- what is the “true state” of the bioreactor at time \( t \)? (filtering)
General questions

Given access to the observations $\{Y_{tk}\}_{k=0}^{n}$, we might want to ask

- what is the “true state” of the bioreactor at time $t$? (filtering)

- what are we likely to observe at time $t_{n+1}$? (prediction)
General questions

Given access to the observations \( \{ Y_{t_k} \}_{k=0}^n \), we might want to ask

- what is the “true state” of the bioreactor at time \( t \)? (filtering)

- what are we likely to observe at time \( t_{n+1} \)? (prediction)

- what are the rules governing the evolution of the system? (model selection / parameter estimation)

We’ll focus on the last type of question.
Basic assumptions

How are the variables $\{X_{tk}\}_{k=0}^{n}$ and $\{Y_{tk}\}_{k=0}^{n}$ related?

We’ll assume the process $(X_t, Y_t)_t$ has:

- **stationarity**: the rules governing both the state space and our observations don’t change over time.
Basic assumptions

How are the variables \( \{ X_{tk} \}_{k=0}^n \) and \( \{ Y_{tk} \}_{k=0}^n \) related?

We’ll assume the process \((X_t, Y_t)_t\) has:

- **stationarity**: the rules governing both the state space and our observations don’t change over time.

- **Markov property**: given the microbial population today, the microbial population tomorrow is independent of the population yesterday.
Basic assumptions

How are the variables \( \{ X_{tk} \}_{k=0}^n \) and \( \{ Y_{tk} \}_{k=0}^n \) related?

We’ll assume the process \((X_t, Y_t)_t\) has:

- **stationarity**: the rules governing both the state space and our observations don’t change over time.

- **Markov property**: given the microbial population today, the microbial population tomorrow is independent of the population yesterday.

- **conditionally independent observations**: given the state of the population today, today’s observation is independent of any other variables.

Such systems are called “hidden Markov models” (HMMs).
HMMs

\[
p(X_{t+1}|X_t) \]

\[
p(y_{t+1}|X_{t+1}) \]
Dynamic linear models

\[ x_{t+1} = A_{t+1} x_t \]
\[ y_t = B_t x_t + v_t, \]

Here:

- \( y_t \) is an observation in \( \mathbb{R}^p \);
- \( x_t \) is a hidden state in \( \mathbb{R}^q \);
- \( A_t \) is a \( p \times p \) state transition matrix;
- \( B_t \) is a \( q \times p \) observation matrix;
- \( v_t \) is a zero-mean vector in \( \mathbb{R}^q \).
Stochastic versus deterministic systems

Should the process \((X_t)_t\) be stochastic or deterministic?
Should the process \((X_t)_t\) be stochastic or deterministic?

- If the conditional distribution of \(X_{t_{k+1}}\) given \(X_{t_k}\) has positive variance, then we’ll say the process \((X_t)_t\) is stochastic.
Stochastic versus deterministic systems

Should the process \((X_t)_t\) be stochastic or deterministic?

- If the conditional distribution of \(X_{t_{k+1}}\) given \(X_{t_k}\) has positive variance, then we’ll say the process \((X_t)_t\) is stochastic.
- Otherwise, we’ll say the process \((X_t)_t\) is deterministic.

In ecology both types of systems are commonly used.
Deterministic dynamics: for each $\theta$, there is a map $T_\theta : X \to X$ such that $X_{t+1} = T_\theta(X_t)$. 
Deterministic dynamics: for each \( \theta \), there is a map \( T_\theta : X \rightarrow X \) such that \( X_{t+1} = T_\theta(X_t) \).

The Markov transition kernel is degenerate:

\[
Q_\theta(x, y) = \delta_{T_\theta(x)}(y).
\]
Deterministic dynamics: for each $\theta$, there is a map $T_\theta : X \to X$ such that $X_{t+1} = T_\theta(X_t)$.

The Markov transition kernel is degenerate:

$$Q_\theta(x, y) = \delta_{T_\theta(x)}(y).$$

Such systems do not satisfy the strong stochastic mixing conditions used in previous work for HMMs.
Setting for deterministic dynamics

Suppose that for each $\theta$ in $\Theta$ (parameter space), we have $(X, \mathcal{X}, T_\theta, \mu_\theta)$, where

- $X$ is a complete separable metric space with Borel $\sigma$-algebra $\mathcal{X}$
Setting for deterministic dynamics

Suppose that for each $\theta$ in $\Theta$ (parameter space), we have $(X, \mathcal{X}, T_\theta, \mu_\theta)$, where

- $X$ is a complete separable metric space with Borel $\sigma$-algebra $\mathcal{X}$
- $T_\theta : X \to X$ is a measurable map,
Setting for deterministic dynamics

Suppose that for each $\theta$ in $\Theta$ (parameter space), we have $(X, \mathcal{X}, T_\theta, \mu_\theta)$, where

- $X$ is a complete separable metric space with Borel $\sigma$-algebra $\mathcal{X}$
- $T_\theta : X \to X$ is a measurable map,
- $\mu_\theta$ is a probability measure on $(X, \mathcal{X})$ is $T_\theta$-invariant if $\mu_\theta(T_\theta^{-1}A) = \mu_\theta(A)$, $\forall A \in \mathcal{X}$
Setting for deterministic dynamics

Suppose that for each \( \theta \) in \( \Theta \) (parameter space), we have \((X, \mathcal{X}, T_\theta, \mu_\theta)\), where

- \( X \) is a complete separable metric space with Borel \( \sigma \)-algebra \( \mathcal{X} \)
- \( T_\theta : X \to X \) is a measurable map,
- \( \mu_\theta \) is a probability measure on \((X, \mathcal{X})\) is \( T_\theta \)-invariant if \( \mu_\theta(T_\theta^{-1}A) = \mu_\theta(A), \ \forall A \in \mathcal{X} \)
- the measure preserving system \((X, \mathcal{X}, T_\theta, \mu_\theta)\) is ergodic if \( T_\theta^{-1}A = A \) implies \( \mu(A) = \{0, 1\} \).
Setting for deterministic dynamics

Suppose that for each \( \theta \) in \( \Theta \) (parameter space), we have \((X, \mathcal{X}, T_\theta, \mu_\theta)\), where

- \( X \) is a complete separable metric space with Borel \( \sigma \)-algebra \( \mathcal{X} \),
- \( T_\theta : X \to X \) is a measurable map,
- \( \mu_\theta \) is a probability measure on \((X, \mathcal{X})\) is \( T_\theta \)-invariant if \( \mu_\theta(T_\theta^{-1}A) = \mu_\theta(A) \), \( \forall A \in \mathcal{X} \),
- the measure preserving system \((X, \mathcal{X}, T_\theta, \mu_\theta)\) is ergodic if \( T_\theta^{-1}A = A \) implies \( \mu(A) = \{0, 1\} \).

Family of systems \((X, \mathcal{X}, T_\theta, \mu_\theta)_{\theta \in \Theta} \equiv (T_\theta, \mu_\theta)_{\theta \in \Theta} \).
Stochastic mixing: Let \((X_t)\) be a stochastic process. Consider the function \(\alpha(s)\)

\[\alpha(s) = \sup\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \} \]

such that \(A \in X_{t-\infty}, B \in X_{t+s}\) the process is strongly mixing if \(\alpha(s) \to 0\) as \(s \to \infty\)
Mixing

Stochastic mixing: Let $(X_t)$ be a stochastic process. Consider the function $\alpha(s)$

$$\alpha(s) = \sup \{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\}$$

such that $A \in X_{-\infty}^t$, $B \in X_{t+s}^\infty$ the process is strongly mixing if $\alpha(s) \to 0$ as $s \to \infty$

Dynamical mixing: $T$ is strongly mixing if for all $A, B \in \mathcal{X}$

$$\lim_{n \to \infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B).$$
An example

- $X_0 \sim U[0, 1]$;
- $X_{k+1} = \theta X_k (1 - X_k)$;
An example

- $X_0 \sim U[0, 1];$
- $X_{k+1} = \theta X_k (1 - X_k);$
- $Y_k \sim N(X_k, \sigma_\theta^2).$
Classical Bayesian inference

Likelihood: $\text{Lik}(\text{data} \mid \theta)$
Classical Bayesian inference

Likelihood: $\text{Lik}(\text{data} \mid \theta)$

Prior: $\pi(\theta)$
Classical Bayesian inference

Likelihood: $\text{Lik}(\text{data} \mid \theta)$

Prior: $\pi(\theta)$

Marginal likelihood: $\int_{\theta} \text{Lik}(\text{data} \mid \theta) \times \pi(\theta)d\theta = \Pr(\text{data})$
Classical Bayesian inference

Likelihood: $\text{Lik}(\text{data} \mid \theta)$

Prior: $\pi(\theta)$

Marginal likelihood: $\int_{\theta} \text{Lik}(\text{data} \mid \theta) \times \pi(\theta) d\theta = \Pr(\text{data})$

Bayes rule:

$$\sigma_n(\theta \mid \text{data}) = \frac{\text{Lik}(\text{data} \mid \theta) \times \pi(\theta)}{\Pr(\text{data})}.$$
Example

Likelihood:  \( X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, 1) \)
Example

Likelihood: \( X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 1) \)
Prior: \( \theta \sim N(0, 1) \)
Example

Likelihood: $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, 1)$
Prior: $\theta \sim \mathcal{N}(0, 1)$

Bayes rule

$$
\sigma_n(\theta \mid X_1, \ldots, X_n) = \frac{(2\pi)^{-(n+1)/2} e^{-\sum_i (x_i - \theta)^2 / 2} \times e^{-\theta^2 / 2}}{\int_{-\infty}^{\infty} (2\pi)^{-(n+1)/2} e^{-\sum_i (x_i - \theta)^2 / 2} \times e^{-\theta^2 / 2} d\theta}.
$$

$$
\theta \mid X_1, \ldots, X_n \sim \mathcal{N} \left( \frac{n}{n+1} \bar{X}, \frac{1}{n+1} \right), \quad \bar{X} = \frac{1}{n} \sum_i X_i.
$$
Observation system \((\mathcal{X}, T, \nu)\) with \(T : \mathcal{X} \rightarrow \mathcal{X}\)
Preliminaries

Observation system \((\mathcal{Y}, T, \nu)\) with \(T : \mathcal{Y} \rightarrow \mathcal{Y}\)

Tracking systems:
Compact metrizable space \(\mathcal{X} := X \times \Theta\) with map \(S : \mathcal{X} \rightarrow \mathcal{X}\).

\[
S : \Theta \times X \rightarrow X, \quad S_\theta : X \rightarrow X.
\]
Observation system \((\mathcal{Y}, T, \nu)\) with \(T : \mathcal{Y} \to \mathcal{Y}\)

Tracking systems:
Compact metrizable space \(\mathcal{X} := X \times \Theta\) with map \(S : \mathcal{X} \to \mathcal{X}\). 

\[
S : \Theta \times X \to X, \quad S_\theta : X \to X.
\]

Loss or regret: \(c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+\). Cost of

\[
c_n(x, y) := c_n(x_0^{n-1}, y_0^{n-1}) = \sum_{k=0}^{n-1} c(x_k, y_k),
\]

\(x_0^{n-1} = (x, Sx, \ldots, S^{n-1}x)\) and \(y_0^{n-1} = (y, Ty, \ldots, T^{n-1}y)\).
Gibbs posterior

Given observations \((y, Ty, \ldots, T^{n-1}y) \in \mathcal{Y}^n\) and a prior \(\pi\) on \(\mathcal{X}\).
Gibbs posterior

Given observations \((y, Ty, \ldots, T^{n-1}y) \in \mathcal{Y}^n\) and a prior \(\pi\) on \(\mathcal{X}\).

Consider the probability measure over Borel sets \(A \subset \mathcal{X}\)

\[
\sigma_n(A | y) = \frac{\int_A \exp(-c_n(x, y)) \, d\pi(x)}{Z_n(y)}, \quad A \subset \Theta \times X
\]

\[
Z_n(y) = \int_{\mathcal{X}} \exp(-c_n(x, y)) \, d\pi(x).
\]
Gibbs posterior

Given observations \((y, Ty, \ldots, T^{n-1}y) \in \mathcal{Y}^n\) and a prior \(\pi\) on \(\mathcal{X}\).

Consider the probability measure over Borel sets \(A \subset \mathcal{X}\)

\[
\sigma_n(A \mid y) = \frac{\int_A \exp(-c_n(x, y)) \, d\pi(x)}{Z_n(y)}, \quad A \subset \Theta \times X
\]

\[
Z_n(y) = \int_{\mathcal{X}} \exp(-c_n(x, y)) \, d\pi(x).
\]

Two questions

(1) Is \(\lim_{n \to \infty} \sigma_n(\cdot \mid y)\) unique.
Gibbs posterior

Given observations \((y, Ty, \ldots, T^{n-1}y) \in \mathcal{Y}^n\) and a prior \(\pi\) on \(\mathcal{X}\).

Consider the probability measure over Borel sets \(A \subset \mathcal{X}\)

\[
\sigma_n(A \mid y) = \frac{\int_A \exp(-c_n(x, y)) \, d\pi(x)}{Z_n(y)}, \quad A \subset \Theta \times X
\]

\[
Z_n(y) = \int_\mathcal{X} \exp(-c_n(x, y)) \, d\pi(x).
\]

Two questions

(1) Is \(\lim_{n \to \infty} \sigma_n(\cdot \mid y)\) unique.
(2) Does \(\lim_{n \to \infty} \sigma_n(\cdot \mid y)\) concentrate around \(T\).
Gibbs posterior

(1) Decision theoretic perspective of Bayesian inference, coherent inference with respect to a utility.
Gibbs posterior

(1) Decision theoretic perspective of Bayesian inference, coherent inference with respect to a utility.

(2) If $c_n$ is the negative log likelihood then recover standard posterior.
Gibbs posterior

(1) Decision theoretic perspective of Bayesian inference, coherent inference with respect to a utility.

(2) If $c_n$ is the negative log likelihood then recover standard posterior.

(3) Robust to misspecification, robust statistics.
Gibbs posterior

(1) Decision theoretic perspective of Bayesian inference, coherent inference with respect to a utility.

(2) If $c_n$ is the negative log likelihood then recover standard posterior.

(3) Robust to misspecification, robust statistics.

(4) Calibration/violation of likelihood principle

$$
\sigma_n(A \mid y) = \frac{\int_A \exp(-\psi c_n(x,y)) \, d\pi(x)}{Z_n(y)}.
$$
Gibbs measures

Given $\mathcal{X}$, the map $S$, a potential function $h$, and a measure $\mu_0$

\[ \sigma_n(x; \mu_0, h) = \frac{\exp\left(\sum_{k=0}^{m} h(S^k x)\right)}{\int_{\mathcal{X}} \exp\left(\sum_{k=0}^{m} h(S^k x)\right) d\mu_0}. \]

The Gibbs measure is the limit point of the sequence $\sigma_n(x; \mu_0, h)$ and the Gibbs measure is denoted as $\mu_0(h)$. 
Given $\mathcal{X}$, the map $S$, a potential function $h$, and a measure $\mu_0$

$$\sigma_n(x; \mu_0, h) = \frac{\exp\left(\sum_{k=0}^{m} h(S^k x)\right)}{\int_{\mathcal{X}} \exp\left(\sum_{k=0}^{m} h(S^k x)\right) d\mu_0}.$$ 

The Gibbs measure is the limit point of the sequence $\sigma_n(x; \mu_0, h)$ and the Gibbs measure is denoted as $\mu_0(h)$.

Recall

$$\sigma_n(x \mid y) = \frac{\exp\left(- \sum_{k=0}^{m} c(S^k x, T^k y)\right)}{\int_{\mathcal{X}} \exp\left(- \sum_{k=1}^{m} c(S^k x, T^k y)\right) d\pi(x)}.$$
Definition (Joining)

Let $(X, A, \mu, T)$ and $(Y, B, \nu, S)$ be two dynamical systems. A joining of $T$ and $S$ is a probability measure $\lambda$ on $X \times Y$, with marginals $\mu$ and $\nu$ respectively, and invariant to the product map $T \times S$. 
Joinings and couplings

Definition (Joining)
Let \((X, A, \mu, T)\) and \((Y, B, \nu, S)\) be two dynamical systems. A joining of \(T\) and \(S\) is a probability measure \(\lambda\) on \(X \times Y\), with marginals \(\mu\) and \(\nu\) respectively, and invariant to the product map \(T \times S\).

Definition (Coupling)
A coupling of two random variable \(X\) and \(X'\) taking values in \((E, \mathcal{E})\) is any pair of random variables \((Y, Y')\) taking values in \((E \times E, \mathcal{E} \times \mathcal{E})\) whose marginals have the same distribution as \(X\) and \(X'\), \(X \overset{D}{=} Y\) and \(X' \overset{D}{=} Y'\).
$\mathcal{J}(\mu, \nu)$ is the set of all joinings of $(\mathcal{X}, S, \mu)$ and $(\mathcal{Y}, T, \nu)$.

Define $\mathcal{J}(S : \nu) = \bigcup_\mu \mathcal{J}(\mu, \nu)$, where the union is over all $S$-invariant Borel probability measures $\mu$, $\mu \in M(\mathcal{X}, S)$.
Variational formulation of $Z_n(y)$ – average cost

Recall $\nu$ is the measure for $T$ and $\lambda \in \mathcal{J}(S : \nu)$
Recall $\nu$ is the measure for $T$ and $\lambda \in \mathcal{J}(S : \nu)$

Define $\lambda_y \in M(\mathcal{X})$ (\(\lambda\) “projected" onto $d\nu_y$)

$$\lambda = \int_Y \lambda_y \otimes \delta_y \, d\nu(y).$$
Variational formulation of $Z_n(y) - \text{average cost}$

Recall $\nu$ is the measure for $T$ and $\lambda \in \mathcal{J}(S : \nu)$

Define $\lambda_y \in \mathcal{M}(\mathcal{X})$ ($\lambda$ “projected" onto $d\nu_y$)

$$\lambda = \int_\mathcal{Y} \lambda_y \otimes \delta_y \, d\nu(y).$$

Limiting average cost

$$\lim_{n \to \infty} \frac{1}{n} \int_\mathcal{X} c_n(x, y) \, d\lambda_y(x) = \int c \, d\lambda.$$
Variational formulation of $Z_n(y)$ – entropy term

Given two Borel probability measures $\pi$ and $\mu$ on $\mathcal{X}$ and a finite measurable partition $\xi$ of $\mathcal{X}$.
Denote $\mu \prec_\xi \pi$ as $\pi(C) = 0 \Rightarrow \mu(C) = 0$ for $C \in \xi$. 

Variational formulation of $Z_n(y)$ – entropy term

Given two Borel probability measures $\pi$ and $\mu$ on $\mathcal{X}$ and a finite measurable partition $\xi$ of $\mathcal{X}$.

Denote $\mu \prec_\xi \pi$ as $\pi(C) = 0 \Rightarrow \mu(C) = 0$ for $C \in \xi$.

Define

$$L(\sigma \parallel \pi, \xi) = \begin{cases} \sum_{C \in \xi} \sigma(C) \log \pi(C), & \text{if } \sigma \prec_\xi \pi \\ -\infty, & \text{otherwise,} \end{cases}$$

with $0 \cdot \log 0 = 0$. 

Variational formulation of $Z_n(y)$ – entropy term

Given two Borel probability measures $\pi$ and $\mu$ on $\mathcal{X}$ and a finite measurable partition $\xi$ of $\mathcal{X}$.

Denote $\mu \prec_\xi \pi$ as $\pi(C) = 0 \Rightarrow \mu(C) = 0$ for $C \in \xi$.

Define

$$L(\sigma \parallel \pi, \xi) = \begin{cases} \sum_{C \in \xi} \sigma(C) \log \pi(C), & \text{if } \sigma \prec_\xi \pi \\ -\infty, & \text{otherwise} \end{cases}$$

with $0 \cdot \log 0 = 0$.

In spirit consider all finite measurable partitions $\xi$

$$F(\sigma, \pi) = \sup_{\xi} L(\sigma \parallel \pi, \xi).$$
Variational formulation of $Z_n(y)$ – weak Gibbs prior

A prior $\pi$ satisfies the weak Gibbs property if there exists a continuous function $\phi : \mathcal{X} \rightarrow \mathbb{R}$ and refining sequence of partitions $\{\xi^m\}$ such that $\text{diam}(\xi^m) \rightarrow 0$ where $\xi^m$ has zero boundary for all $S$-invariant measures and for each $m$, there exists $\alpha = \alpha(m) > 0$ and $K = K(m)$ such that for all $n \in \mathbb{N}$ and $x \in \mathcal{X}$

$$K^{-1} e^{-\alpha n} \leq \frac{\pi(\xi^m_n(x))}{e^{\phi_n(x)}} \leq Ke^{\alpha n}.$$ 

and $\alpha(m) \rightarrow 0$. 
Convergence

Theorem (McGoff-M.-Nobel)

Suppose a weak Gibbs prior, then for $\nu$ almost every $y$,

$$\lim_{n \to \infty} -\frac{1}{n} \log Z_n(y) = \inf_{\lambda \in \mathcal{J}(S; \nu)} \left\{ \int c \, d\lambda + F(\lambda, \pi) \right\},$$

and the infimum in the above expression is attained.
Convergence

Theorem (McGoff-M.-Nobel)

Suppose a weak Gibbs prior, then for $\nu$ almost every $y$,

$$\lim_{n \to \infty} \frac{1}{n} \log Z_n(y) = \inf_{\lambda \in \mathcal{J}(S:\nu)} \left\{ \int c \, d\lambda + F(\lambda, \pi) \right\},$$

and the infimum in the above expression is attained.

Furthermore,

$$\frac{1}{n} \sum_{k=0}^{n-1} \hat{\lambda}_n(S^{-k} \mid Y_0^{n-1}) \to \mathcal{P}_{eq}(\mathcal{X}, S).$$

$\mathcal{P}_{eq}(\mathcal{X}, S)$ are the set of processes that minimize the variational expression above.
Proposition (McGoff-M.-Nobel)

Suppose a weak Gibbs prior and consider the pressure

\[ P(\mu, \nu) = \inf_{\lambda \in \mathcal{J}(S:\nu)} \left\{ \int c \, d\lambda + F(\lambda, \pi) \right\} \]

\[ P(\theta : \nu) = \inf_{\mu \in \mathcal{M}(\mathcal{X}_\theta, S_\theta)} P(\mu, \nu), \]

\[ \theta_* = \arg \min_{\theta \in \Theta} P(\theta : \nu). \]
Convergence

Proposition (McGoff-M.-Nobel)

Suppose a weak Glbbs prior and consider the pressure

\[
P(\mu, \nu) = \inf_{\lambda \in \mathcal{J}(S;\nu)} \left\{ \int c \, d\lambda + F(\lambda, \pi) \right\}
\]

\[
P(\theta : \nu) = \inf_{\mu \in M(X_\theta, S_\theta)} P(\mu, \nu),
\]

\[
\theta_* = \arg \min_{\theta \in \Theta} P(\theta : \nu).
\]

For all \( \varepsilon > 0 \)

\[
P(d(S_{\theta_*}, T) < \varepsilon) \to 1 \text{ a.s as } n \to \infty.
\]
Toy example: Markov model

- \{\mu_\theta : \theta \in \Theta\} is a collection of Gibbs measures on a common finite state space;
- there exists \( \theta^* \in \Theta \) such that \( \hat{\lambda} = \mu_{\theta^*} \);
- \( \ell(\theta; y_0^{n-1}) = -\log \mu_\theta(y_0^{n-1}) \).

The standard Variational Principle for Gibbs measures yields that the posterior distribution converges almost surely to \( \theta^* \).
Toy example: Markov model

- $\{\mu_{\theta} : \theta \in \Theta\}$ is a collection of Gibbs measures on a common finite state space;
- there exists $\theta^* \in \Theta$ such that $\hat{\lambda} = \mu_{\theta^*}$;
- $\ell(\theta; y_{0}^{n-1}) = -\log \mu_{\theta}(y_{0}^{n-1})$.

The standard Variational Principle for Gibbs measures yields that the posterior distribution converges almost surely to $\theta^*$.

More generally: convergence analysis for Gibbs posteriors under dependence.
For each \( t \in \mathbb{N} \) and \( x, y \in \mathcal{X} \), let

\[
d_t(x, y) = \max\{d(S^k x, S^k y) : 0 \leq k < t\},
\]

where \( d \) is some metric on \( \mathcal{X} \).
Objective Bayesian inference

For each $t \in \mathbb{N}$ and $x, y \in \mathcal{X}$, let

$$d_t(x, y) = \max \{ d(S^k x, S^k y) : 0 \leq k < t \},$$

where $d$ is some metric on $\mathcal{X}$.

$N_{sep}(\epsilon, t)$ is packing number and $U_{n,\epsilon}$ is the packing set. Consider the distribution

$$\frac{1}{N_{sep}(\epsilon, t)} \sum_{x \in U_{n,\epsilon}} \delta_x.$$
Objective Bayesian inference

For each \( t \in \mathbb{N} \) and \( x, y \in \mathcal{X} \), let

\[
d_t(x, y) = \max \{ d(S^k x, S^k y) : 0 \leq k < t \},
\]

where \( d \) is some metric on \( \mathcal{X} \).

\( N_{sep}(\epsilon, t) \) is packing number and \( U_{n, \epsilon} \) is the packing set. Consider the distribution

\[
\frac{1}{N_{sep}(\epsilon, t)} \sum_{x \in U_{n, \epsilon}} \delta_x.
\]

The Gibbs posterior distribution is

\[
\sigma_{n, \epsilon}(\cdot \mid y) = \frac{1}{Z_{n, \epsilon}(y)} \sum_{x \in U_{n, \epsilon}} \exp(-c_n(x, y)) \delta_x.
\]
Contributions

Reframes posterior consistency as two-stage process: first find the limiting variational problem, and then analyze this problem to address consistency.

Provides general framework and suite of tools from the thermodynamic formalism for analyzing asymptotic behavior of Gibbs posteriors.
Questions

Statistics questions.

- What types of observations and models are amenable to this analysis?
Statistics questions.

- What types of observations and models are amenable to this analysis?
- For which combinations of observations and models can one establish posterior consistency?
Questions

Statistics questions.

- What types of observations and models are amenable to this analysis?
- For which combinations of observations and models can one establish posterior consistency?

Dynamics questions.

- How far can the thermodynamic formalism be pushed?
Questions

Statistics questions.
- What types of observations and models are amenable to this analysis?
- For which combinations of observations and models can one establish posterior consistency?

Dynamics questions.
- How far can the thermodynamic formalism be pushed?
- Under what conditions is there a limiting variational characterization?
Questions

Statistics questions.

- What types of observations and models are amenable to this analysis?
- For which combinations of observations and models can one establish posterior consistency?

Dynamics questions.

- How far can the thermodynamic formalism be pushed?
- Under what conditions is there a limiting variational characterization?
- Under what conditions is there a unique equilibrium joining?
Open problems

(1) Rates of convergence for a family of dynamical systems $\mathcal{F}$. 
Open problems

(1) Rates of convergence for a family of dynamical systems $\mathcal{F}$.
(2) General conditions for learnability in dynamical systems.
Open problems

(1) Rates of convergence for a family of dynamical systems $\mathcal{F}$.
(2) General conditions for learnability in dynamical systems.
(3) Extension to continuous time dynamics, differential equations.
Open problems

(1) Rates of convergence for a family of dynamical systems $\mathcal{F}$.
(2) General conditions for learnability in dynamical systems.
(3) Extension to continuous time dynamics, differential equations.
(4) Computational issues.
Open problems

(1) Rates of convergence for a family of dynamical systems $\mathcal{F}$.
(2) General conditions for learnability in dynamical systems.
(3) Extension to continuous time dynamics, differential equations.
(4) Computational issues.
(5) Integration of ideas from statistical models of time series and dynamical systems theory.
Acknowledgements

Thanks:
Part I: Ingrid Debauchies, Misha Belkin, Lek-Heng Lim, Leo Guibas, Doug Boyer, Shmuel Weinberger

Part II: Konstantin Mischaikow, Ramon van Handel, Steve Lalley, Jonathan Mattingly, Karl Petersen, Ioanna Manolopoulou, Jim Berger.
Acknowledgements

Thanks:
Part I: Ingrid Debauchies, Misha Belkin, Lek-Heng Lim, Leo Guibas, Doug Boyer, Shmuel Weinberger

Part II: Konstantin Mischaikow, Ramon van Handel, Steve Lalley, Jonathan Mattingly, Karl Petersen, Ioanna Manolopoulou, Jim Berger.

Funding:
- Center for Systems Biology at Duke
- NSF DMS, CCF, CISE
- AFOSR
- DARPA
- NIH
New journal

http://www.siam.org/journals/siaga.php

SIAM Journal on Applied Algebra and Geometry
Evolution of cooperation in mammals
Evolution of cooperation in mammals

**Figure 1: Schematic representation of the overall study design.**

**Figure 2:** Weight measurement and blood collection from meerkats. Study subjects at Clutton-Brock's field site are individually recognized and highly habituated. (A) All study subjects are trained to climb onto electronic balances, making it possible to repeatedly measure their weight to calculate daily resource intake (based on weight change between the start of the day, midday, and departure of the observers). (B) A member of the study population under brief and reversible anesthesia with isoflurane, which enables (C) blood sample collection and X-ray imaging (see Figure 3). It is possible to schedule specific study subjects for sample/image collection on a given day.

**AIM 1**
- NFkB
- Gene expression
- DNA methylation
- Chromatin accessibility

**AIM 2 A**
- X-ray images
- 3D model reconstruction

**AIM 2 B**
- anti-CD3
- Heat killed Mycobacterium bovis
- Gardiquimod
- LPS

**AIM 3**
- Immunity
- Growth
- Reproduction

Field site in Kalahari Desert

1. **STUDY POPULATION**
   - 50 helpers
   - 50 breeders
   - ~20 transitioning

2. **DATA COLLECTION**
   - Gene expression
   - DNA methylation
   - Chromatin accessibility
   - PBMCs steroid hormone control (vehicle)
   - RNA-seq
   - X-ray images
   - 3D model reconstruction

3. **50 dominant breeders**

4. **50 age- and sex-matched helpers (~50% of each sex)**

5. **15-20 animals followed across the helper-breeder transition**

To collect longitudinal samples, we will oversample helpers at the beginning of the study, and then resample the subset that transition to dominant breeders during the first two years of the study.

Data collection and analysis:
- In **Aim 1**, we will compare breeder and helper gene expression (RNA-seq), chromatin accessibility (ATAC-seq), and DNA methylation data (RRBS).

Tracks show a an example of complementary RNA-seq read pile-ups (at gene exons), DNA methylation levels, and accessible chromatin peaks near a key immune gene, NFKB1 (as in [47]). We will also test whether steroid hormone administration to PBMCs from helpers recapitulate breeder-like gene regulatory states.

- In **Aim 2**, we will develop computational geometric/topological approaches for reconstructing 3D skeletal shape from 2D X-ray images (2A) and use experimental immunogenomic methods to quantitatively assess the response to multiple types of pathogens (2B).

The resulting data will allow us to test how growth and immune defense differ between helpers and breeders.

- In **Aim 3**, the full team will integrate data on growth, immune response and reproductive effort to test the hypothesis that energetic trade-offs between these tasks create "competition" at the transcriptional level, which is resolved differently by helpers and breeders.

Each axis represents investment in one of the three dimensions; size of the triangle represents resource intake, which we will also measure for each individual (see Figure 2A).
Evolution of cooperation in mammals

Figure 3: brief and reversible anesthesia with isoflurane, which enables making are individually recognized and highly habituate by helpers and breeders.

Figure 6: PROPOSED RESEARCH

HFSP Form RGP

Aim 1: Measuring resource intake

- We will track growth at local and whole skeleton scales for helpers and breeders.
- Each axis represents investment in one of the three dimensions; size of the triangle will allow us to test how growth will oversample breeders during the first two years of the study.
- We will also measure for each individual (see Figure 2A).
- We will collect sets of X-ray images of a specific reference mesh (obtained through CT scan). We will then perform individual bone from non-resident animals.
- We will use experimental immunogenomic methods to quantitatively assess the response to LPS.
- We will test whether the immune response and reproductive effort tradeoffs between helpers and breeders.
- We will track growth at local and whole skeleton scales for helpers and breeders.
- We will collect sets of X-ray images of a specific reference mesh (obtained through CT scan). We will then perform individual bone from non-resident animals.
- We will use experimental immunogenomic methods to quantitatively assess the response to LPS.
- We will test whether the immune response and reproductive effort tradeoffs between helpers and breeders.
- We will track growth at local and whole skeleton scales for helpers and breeders.
- We will collect sets of X-ray images of a specific reference mesh (obtained through CT scan). We will then perform individual bone from non-resident animals.
- We will use experimental immunogenomic methods to quantitatively assess the response to LPS.
- We will test whether the immune response and reproductive effort tradeoffs between helpers and breeders.
- We will track growth at local and whole skeleton scales for helpers and breeders.
- We will collect sets of X-ray images of a specific reference mesh (obtained through CT scan). We will then perform individual bone from non-resident animals.
- We will use experimental immunogenomic methods to quantitatively assess the response to LPS.
- We will test whether the immune response and reproductive effort tradeoffs between helpers and breeders.