Overview of Markovian maps

Mike Boyle
University of Maryland

BIRS October 2007

---

*these slides are online, via
Google: Mike Boyle

Outline of the talk

I. Subshift background

1. Subshifts

2. (Sliding) block codes

3. Measures

4. Why use subshifts to consider measures?

5. Hidden Markov measures

II. Markovian maps and thermodynamics

III. Markovian maps and resolving maps

I. Subshift background

1. Subshifts

Notations:

- $\mathcal{A} = \{0, 1, \ldots, N-1\}$, the alphabet
- $X_N = \mathcal{A}^{\mathbb{Z}}$
  x in X_N is a bisequence $\ldots x_{-1}x_0x_1\ldots$,
  with all x_i in $\mathcal{A}$
- X_N is a metric space,
  $\text{dist}(x, y) = \frac{1}{k+1}$, if $k = \min\{|j| : x_j \neq y_j\}$
- $\sigma : X_N \to X_N$ is the shift map $(\sigma x)_i = x_{i+1}$
- σ and σ^{-1} are 1-1, onto, continuous.
- (X_N, σ) is the *full shift on N symbols*

- If Y is a closed σ-invariant subset of X_N, then (Y, σ) is a *subshift*.

- For such a Y, there exists a set $\mathcal{F}$ of words on $\mathcal{A}$ such that
$Y = \{x \in X_N : \ \forall i \leq j, \ x_i x_{i+1} \cdots x_j \notin \mathcal{F}\}$.
If it is possible to choose $\mathcal{F}$ a finite set, then Y is a *subshift of finite type* (SFT).

- Example: let A be an $N \times N$ zero-one matrix with rows and columns indexed by $\mathcal{A}$. Define
$X_A = \{x \in \mathcal{A}^{\mathbb{Z}} : \forall i, A(x_i, x_{i+1} = 1\}$.
(X_A, σ) is a *topological Markov chain* [Parry 1964]. It is SFT.

In this talk: we require the matrix defining a TMC to be irreducible
(for all i, j, $\exists m$ such that $A^m(i, j) > 0$).

2. (Sliding) block codes

Let Y be a subshift on alphabet $\mathcal{A}$.
Let Y' be a subshift on alphabet $\mathcal{A}'$.

ABUSE OF NOTATION: by a k-block code I will mean a function $\phi : Y \to Y'$ for which there is a function $\Phi : \mathcal{A}^k \to \mathcal{A}'$ such that $(\phi x)_i = \Phi(x[i, \ldots, i + k - 1])$ for all $i \in \mathbb{Z}$.

I'll say ϕ is a block code if for some k it is a k-block code.

Curtis-Hedlund-Lyndon: For subshifts Y, Y', a map $\psi : Y \to Y'$ is continuous with $\psi\sigma = \sigma\psi$ if and only if ψ is a block code composed with a power of the shift.

From here: ϕ denotes a block code.

ϕ is a *factor map* if ϕ is onto ($\phi : Y \twoheadrightarrow Y'$).
ϕ is an *isomorphism*, or *topological conjugacy*, if it is bijective.

EXAMPLE.
Given a subshift (X, σ), define $X^{[k]}$ as the image of X under ϕ where

$$(\phi x)_i = x[i, i+1, \ldots, i+k-1].$$

The subshift $(X^{[k]}, \sigma)$ is the *k-block presentation* of (X, σ), and is topologically conjugate to (X, σ).

3. **Measures.**

Given a subshift (X, σ):

• $\mathcal{M}(X)$ denotes the space of σ-invariant Borel probabilities on X (these are the measures for which the coordinate projections on X give a 2-sided stationary stochastic process).

• $\mathcal{M}_k(X)$ denotes the k-(step)Markov measures in $\mathcal{M}(X)$ which have full support (all allowed words in X have strictly positive probability).

• $x[i, j]$ may denote either the word $x_i \cdots x_j$ or the set $\{y \in X : \; x_k = y_k, \; i \leq k \leq j\}$.

EXAMPLE. Let P be an $N \times N$ irreducible stochastic matrix, and p the stochastic row vector such that $pP = P$.

Define an $N \times N$ zero-one matrix A by $A(i, j) = 0$ if $P(i, j) = 0$, and $A(i, j) = 1$ otherwise.

Then P determines a μ in $\mathcal{M}_1(X_A)$:

$\mu(x, [i, j]) =$
$p(x_i)P(x_i, x_{i+1})P(x_{i+1}, x_{i+2}) \cdots P(x_{j-1}, x_j)$.

DEFINITION:
μ in $\mathcal{M}(X)$ is k-Markov if for all $i \geq 0$ and $j \geq k-1$,

$$\mu\Big(x[0,i] \;\Big|\; x[-j,0]\Big) = \mu\Big(x[0,i] \;\Big|\; x[-(k-1),0]\Big) \ .$$

A measure is 1-Markov iff it is defined from a stochastic matrix, as on the last slide.

A measure μ is k-Markov iff the topological conjugacy taking X to its k block presentation takes μ to a 1-Markov measure.

A *Markov measure* is a measure which is k-Markov for some k.

From here, "Markov" always means "Markov with full support".

4. Why use subshifts to consider measures?

- We can consider many measures in a common setting. we can study those measures by relating them to continuous functions ("thermodynamics"). We may find distinguished measures (e.g. solving some variational problem involving functions).

- Modulo topological conjugacy (topologically invariant properties), we might conceptually simplify a presentation (e.g., using a higher block presentation, we can reduce many block-code problems to problems involving just one-block codes).

- With topological ideas we might see some structure behind the complications of a particular example.

5. Hidden Markov measures.

Suppose $\mu \in \mathcal{M}(X_A)$, and $\phi : X_A \to Y$.
Then the measure $\phi\mu \in \mathcal{M}(Y)$ is defined by $(\phi\mu)(E) = \mu(\phi^{-1}(E))$.
If $\mu \in \mathcal{M}_k(X_A)$, then $\phi\mu$ is called a *hidden Markov measure* (and various other names).

PROBLEM [Burke Rosenblatt 1958] For ϕ a 1-block code and μ 1-Markov, when is $\phi\mu$ 1-Markov?

- The problem was solved (several times).
- Via the higher block presentation, we likewise can decide whether $\phi\mu$ is k-Markov.
- Given ϕ and μ Markov, we know k such that either $\phi\mu$ is k-Markov or $\phi\mu$ is not Markov.

ABOVE: given μ, consider $\phi\mu$.
NEXT: given ν, consider $\{\mu : \phi\mu = \nu\}$.

II. Markovian maps and thermodynamics

6. Markovian maps

EXAMPLE [MPW 1984] There exists $\phi : X_A \to X_B$ such that if ν is a supported Markov measure on X_B and $\phi\mu = \nu$, then μ is not a supported Markov measure on X_A.

DEFNITION [BT 1983] $\phi : X_A \to X_B$ is *Markovian* if for every supported Markov measure ν on X_B, $\exists$ a supported Markov measure on X_A such that $\phi\mu = \nu$.

THEOREM [BT 1983] For $\phi : X_A \twoheadrightarrow X_B$, if there exists any supported Markov μ and ν with $\phi\mu = \nu$, then ϕ is Markovian.

(We will see a little better later how the Markovian property is a kind of uniform finiteness property.)

A SIMPLE EXAMPLE.

This is to suggest that by being able to lift one Markov measure to a Markov measure, we may be able to lift other Markov measures to Markov measures. (From here, "Markov" means "Markov with full support".)

Consider the one-block ϕ from $X_3 = \{0, 1, 2\}^{\mathbb{Z}}$ to $X_2 = \{0, 1\}^{\mathbb{Z}}$, via $0 \mapsto 0$ and $1, 2 \mapsto 1$.

Let ν be the 1-Markov measure on X_2 given by the transition matrix $\begin{pmatrix} (1/2) & (1/2) \\ (1/2) & (1/2) \end{pmatrix}$.

Given positive numbers α, β, γ less than 1, the stochastic matrix

$$\begin{pmatrix} (1/2) & \alpha(1/2) & (1-\alpha)(1/2) \\ (1/2) & \beta(1/2) & (1-\beta)(1/2) \\ (1/2) & \gamma(1/2) & (1-\gamma)(1/2) \end{pmatrix}$$

defines a measure on X_2 which maps to ν.

But now, if ν' is any other 1-Markov measure on X_2, given by a stochastic matrix $\begin{pmatrix} p & q \\ r & s \end{pmatrix}$, then ν' will lift to the 1-Markov measure defined by the stochastic matrix

$$\begin{pmatrix} p & \alpha q & (1-\alpha)q \\ r & \beta s & (1-\beta)s \\ r & \gamma s & (1-\gamma)s \end{pmatrix} .$$

(This example map ϕ is "e-resolving", and all e-resolving maps are Markovian.)

OPEN PROBLEM Give a procedure to decide, given $\phi : X_A \twoheadrightarrow X_B$, whether ϕ is Markovian.

THE 1- MARKOVIAN COMPUTATION.

Suppose for a 1-block $\phi : X_A \to X_B$, with $\phi(i)$ denoted $\bar{i}$, that $\phi\mu = \nu$ where μ, ν are 1-Markov defined by stochastic matrices P, Q.

Suppose $\nu' \in \mathcal{M}_1(X_B)$, defined by a stochastic matrix Q'. We will define a stochastic matrix P' defining μ' in $\mathcal{M}_1(X_A)$ so that $\phi\mu' = \nu'$.

First define a matrix M of size matching P by
$M(i,j) = 0$ if $P(i,j) = 0$, and otherwise
$M(i,j) = Q'(\bar{i},\bar{j})P(i,j)/Q(\bar{i},\bar{j})$.

This matrix M will have spectral radius 1 but might not have row sums 1. Let r be a positive right eigenvalue for M. Then P' is the matrix defined by

$$P'(i,j) = r(i)^{-1}M(i,j)\ r(j)\ .$$

This is the germ of a more general thermodynamic result.

7. Thermodynamics on subshifts 001.

ENTROPY

Given: subshift (X, σ), $\mu \in \mathcal{M}(X)$.

- $h(X) = \lim_n \frac{1}{n} \log |\{x[0, n-1] : x \in X\}|$
  is the topological entropy of the map $\sigma|_X$.

- $h_\mu(X) = \lim_n \frac{1}{n} \sum -\mu[W] \log \mu[W]$,
  with the sum over W in $\{x[0, n-1] : x \in X\}$,
  is the measure theoretic entropy of μ
  (with respect to σ).

PRESSURE is a refinement of entropy which takes into account not only the map $\sigma : X \to X$ but also weights coming from a given function.

Given $f \in C(X, \mathbb{R})$,
$P(f, \sigma) = \lim_n \frac{1}{n} \log \sum_W \exp[S_n(f, W)]$
where $S_n(f, W) = \sum_{i=0}^{n-1} f(\sigma^i x)$,
for some $x \in X$ such that $x[0, n-1] = W$
(in the limit the choice of x doesn't matter).

So, $P(f, \sigma) = h(X)$ if $f \equiv 0$.
VARIATIONAL PRINCIPLE FOR PRESSURE:
$P(f, \sigma) = \sup\{h_\mu + \int f \; d\mu : \mu \in \mathcal{M}(X)\}$.

An *equilibrium state* for f (w.r.t. σ) is a measure $\mu = \mu_f$ such that $P(f, \sigma) = h_\mu + \int f \; d\mu$.
Often: μ_f is a *Gibbs measure* for f:
with $P(f, \sigma) = \log(\rho)$,

$$\mu_f(x[0, n-1]) \sim \rho^{-n} \exp S_n f(x)$$

("$\sim$" means the ratio of the two sides is bounded above and away from zero, uniformly in x, n.)

If $f \in C(X_A, \mathbb{R})$, with $f(x) = f(x_0 x_1)$, then f has a unique equilibrium state μ_f, and $\mu_f \in \mathcal{M}_1(\sigma_A)$. This μ_f is defined by the stochastic matrix $P = \text{stoch}(Q)$, where

$$\begin{aligned} Q(i,j) &= 0 \text{ if } A(i,j) = 0, \\ &= \exp[f(i,j)] \text{ otherwise } \quad . \end{aligned}$$

and the *stochasticization* of Q is

$$\text{stoch}(Q) = (1/\rho) D^{-1} Q D,$$

where
ρ is the spectral radius of Q,
D is diagonal with $D(i,i) = r(i)$, and
$r > 0$ and $Qr = \lambda r$.

The pressure of f is $\log \rho$.

Likewise: if $f(x) = f(x_0, x_1, \ldots, x_k)$, then f has a unique equilibrium state μ, and μ is a k-step Markov measure.

Let $C_k(X, \mathbb{R}) = \{f : f(x) = f(x[0, k-1])$. Then [Parry-Tuncel] for f, g in $C_k(X, \mathbb{R})$, T.F.A.E.

- $\mu_f = \mu_g$
- $\exists h \in C(X, \mathbb{R})$ such that $f = g + (h - h \circ \sigma) +$ constant
- $\exists h \in C_{k-1}(X, \mathbb{R})$ such that $f = g + (h - h \circ \sigma) +$ constant

Let W denote the vector space of functions $h - h \circ \sigma +$ constant, with h locally constant. Then the map $C_k(X, \mathbb{R})/W \ \rightarrow \ \mathcal{M}_k(\sigma_A)$, $[f] \mapsto \mu_f$, is a bijection.

So, the Markov measures are identified with the locally constant functions (modulo W).

8. Compensation functions

Let $\phi : X_A \to X_B$. Suppose

- $\mu \in \mathcal{M}(X_A),\ \nu \in \mathcal{M}(X_B)$
- μ and ν are ergodic
- $\nu = \nu_f$
  (i.e. ν is an eq. state for $f \in C(X_B, \mathbb{R})$)
- $\mu = \mu_F$ (write F as $(f \circ \phi) + c$)

Then for any g in $C(X_B, \mathbb{R})$ with unique eq. state ν_g we have:

- if $\mu = \mu_{(g \circ \phi)+c}$, then $\phi\mu = \nu_g$.

Such a function c is called a compensation function [Walters 1986]. It turns out, a compensation function is a function $c \in C(X_A, \mathbb{R})$ such that for all $g \in C(X_B, \mathbb{R})$
• $P(g) = P\Big((g \circ \phi) + c\Big)$, $\forall g \in C(X_B, \mathbb{R})$.
For such a c:

- There is a lift of measures matching an affine embedding of continuous functions:
  $C(X_B) \hookrightarrow C(X_A)$, via $g \to (g \circ \phi) + c$
  $\mathcal{M}(X_B) \hookrightarrow \mathcal{M}(X_A)$, via $\mu_g \to \mu_{(g\circ\phi)+c}$

A compensation function is a kind of oracle which gives a relation on functions that must be respected by sufficiently closely related measures (eq. states).
FACT: ϕ is Markovian iff
ϕ has a compensation function which is locally constant.

In our 1-Markovian computation:
an associated compensation function is
$c(x) = \log P(i,j) - \log Q(\bar{i}, \bar{j})$ when $x_0 x_1 = ij$.

Markovian maps and Resolving Maps

9. Resolving Maps. Let $\phi : X_A \twoheadrightarrow Y$ be a 1-block code, denoting a symbol $(\phi x)_0$ as $\overline{x_0}$. (Y is not necessarily SFT.)

DEFINITION ϕ is right resolving if for all symbols $i, \bar{i}, k$ such that $\bar{i}k$ occurs in Y, there is at most one j such that ij occurs in X_A and $\bar{j} = k$. In other words, for any diagram

$$\begin{array}{ccc} i & & \\ \big\downarrow & & \\ \bar{i} & \longrightarrow & k \end{array}$$

there is at most one j such that

$$\begin{array}{ccc} i & \longrightarrow & j \\ \big\downarrow & & \big\downarrow \\ \bar{i} & \longrightarrow & k \end{array}$$

DEFINITION ϕ is right e-resolving if it satisfies the definition above, with "at most one" is replaced by "at least one".

Reverse the roles of i and j above to define left resolving and left e-resolving. A map ϕ is resolving (e-resolving) if it is left or right resolving (e-resolving).

FACTS:

- If ϕ is resolving, then $h(X_A) = h(Y)$
- If $Y = X_B$ and $h(X_A) = h(X_B)$, then ϕ is e-resolving iff ϕ is resolving.
- If ϕ is e-resolving, then $Y = X_B$.
- If ϕ is e-resolving, then ϕ is (transparently) Markovian.

10. A Putnam diagram

THEOREM [B 2005] Suppose $\phi : X_A \twoheadrightarrow Y$. Then (canonically) there is a commuting diagram of factor maps

$$\begin{array}{ccc} X_F & \xrightarrow{\widetilde{\phi}} & X_B \\ \pi \downarrow & & \downarrow \phi_+ \\ X_A & \xrightarrow[\phi]{} & Y \end{array}$$

with properties

- $$\begin{array}{ccc} & \xrightarrow{\text{left e-resolving}} & \\ \text{right resolving} \downarrow & & \downarrow \text{right resolving} \\ & \xrightarrow[\phi]{} & \end{array}$$

- ϕ_+ and π are 1-1 a.e. (bijective μ-a.e. for every ergodic μ with full support).

- ϕ_+ is a canonical "futures" cover of Y determined by ϕ.

- The maps $\pi, \widetilde{\phi}$ are restrictions of the fibered product of ϕ and π to a canonical irreducible component.

Given a 1-1 a.e. factor map from one irreducible Smale space to another, Putnam [2005] constructed a diagram with the indicated resolving and 1-1 a.e. properties (i.e., a "Putnam diagram").

[B2005] is a restriction to the zero dimensional case; but, there a more concrete construction is feasible, and we need not assume Y is Smale (SFT) or even that $h(X_A) = h(Y)$.

The construction draws on work of Nasu; Kitchens; and Kitchens-Marcus-Trow.

NOTE: e-resolving maps are Markovian.

So, the diagram shows that in some sense, all block codes are close to being Markovian.

Is the diagram of more specific use?

11. Some References

Burke, C. J.; Rosenblatt, M. A Markovian function of a Markov chain. Ann. Math. Statist. 29 (1958) 1112–1122.

Parry, William; Tuncel, Selim. On the stochastic and topological structure of Markov chains. Bull. London Math. Soc. 14 (1982), no. 1, 16–27.

Marcus, Brian; Petersen, Karl; Williams, Susan Transmission rates and factors of Markov chains. Conference in modern analysis and probability (New Haven, Conn., 1982), 279–293, Contemp. Math., 26, Amer. Math. Soc., Providence, RI, 1984.

Boyle,M. and Tuncel, S. Infinite-to-one codes and Markov measures. Trans. Amer. Math. Soc. 285 (1984), 657-684.

Walters, Peter. Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts. Trans. Amer. Math. Soc. 296 (1986), no. 1, 1–31.

Putnam, Ian F. Lifting factor maps to resolving maps. Israel J. Math. 146 (2005), 253–280.

Boyle, Mike. Putnam's resolving maps in dimension zero. Ergodic Theory Dynam. Systems 25 (2005), 1485-1502.

For a hierarchy of conditions on compensation functions and related problems, see Walters' paper above, and the papers of Sujin Shin; Karl Petersen and Sujin Shin; and Petersen, Quas and Shin.