Hidden Markov Chains Found Again

(Continuous Images of Measures on Shifts of Finite Type).

Karl Petersen

University of North Carolina at Chapel Hill

Setting

I. Basic Setup

Setting

I. Basic Setup

X, Y topologically mixing 1-step shifts of finite type

Setting

I. Basic Setup

X, Y topologically mixing 1-step shifts of finite type

$\pi : X \to Y$ 1-block factor map (continuous, shift-commuting)

Setting

I. Basic Setup

X, Y topologically mixing 1-step shifts of finite type

$\pi : X \to Y$ 1-block factor map (continuous, shift-commuting)

$\mu = \sigma$-invariant Borel probability measure on X

Setting

I. Basic Setup

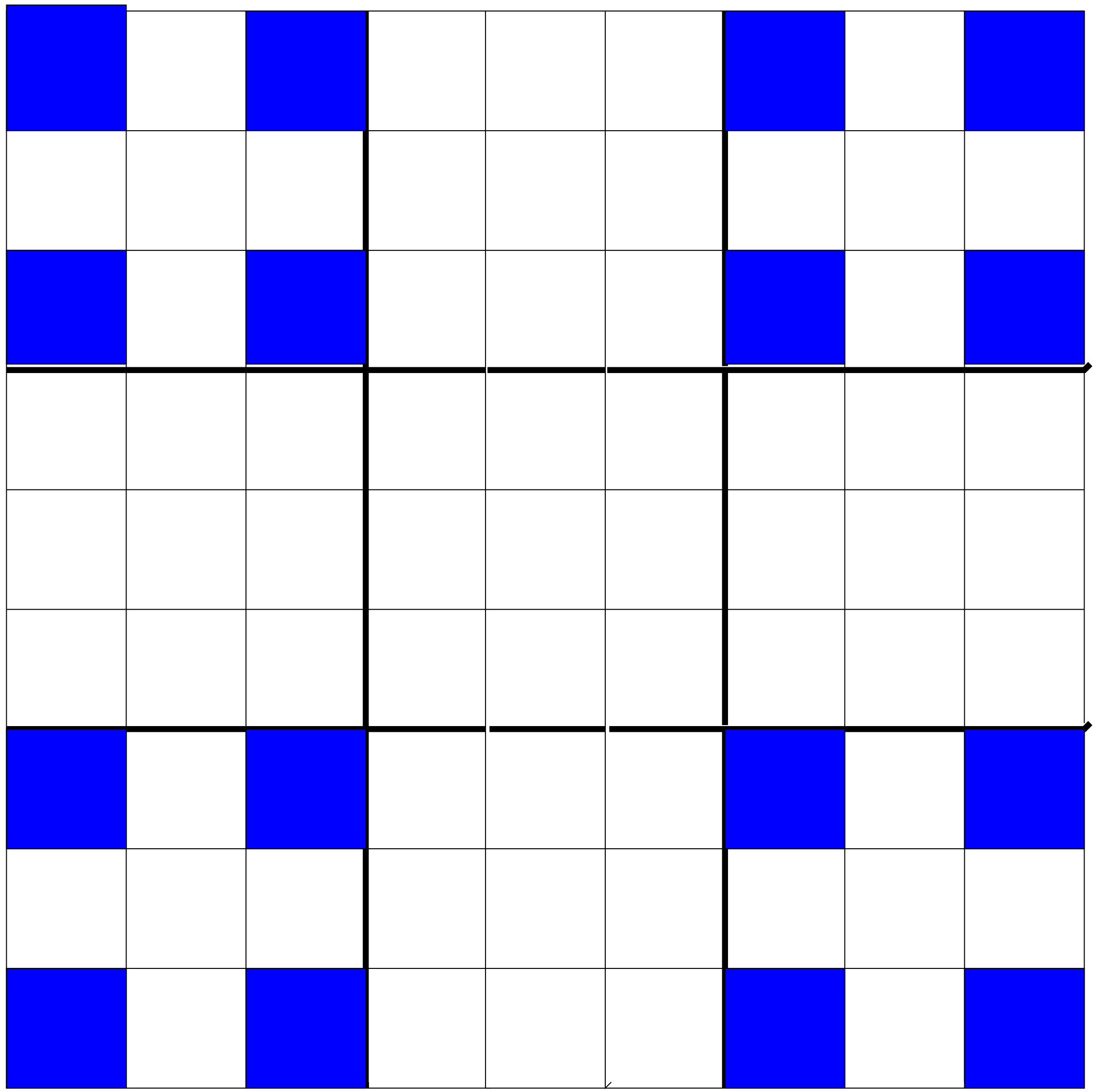
X, Y topologically mixing 1-step shifts of finite type

$\pi : X \to Y$ 1-block factor map (continuous, shift-commuting)

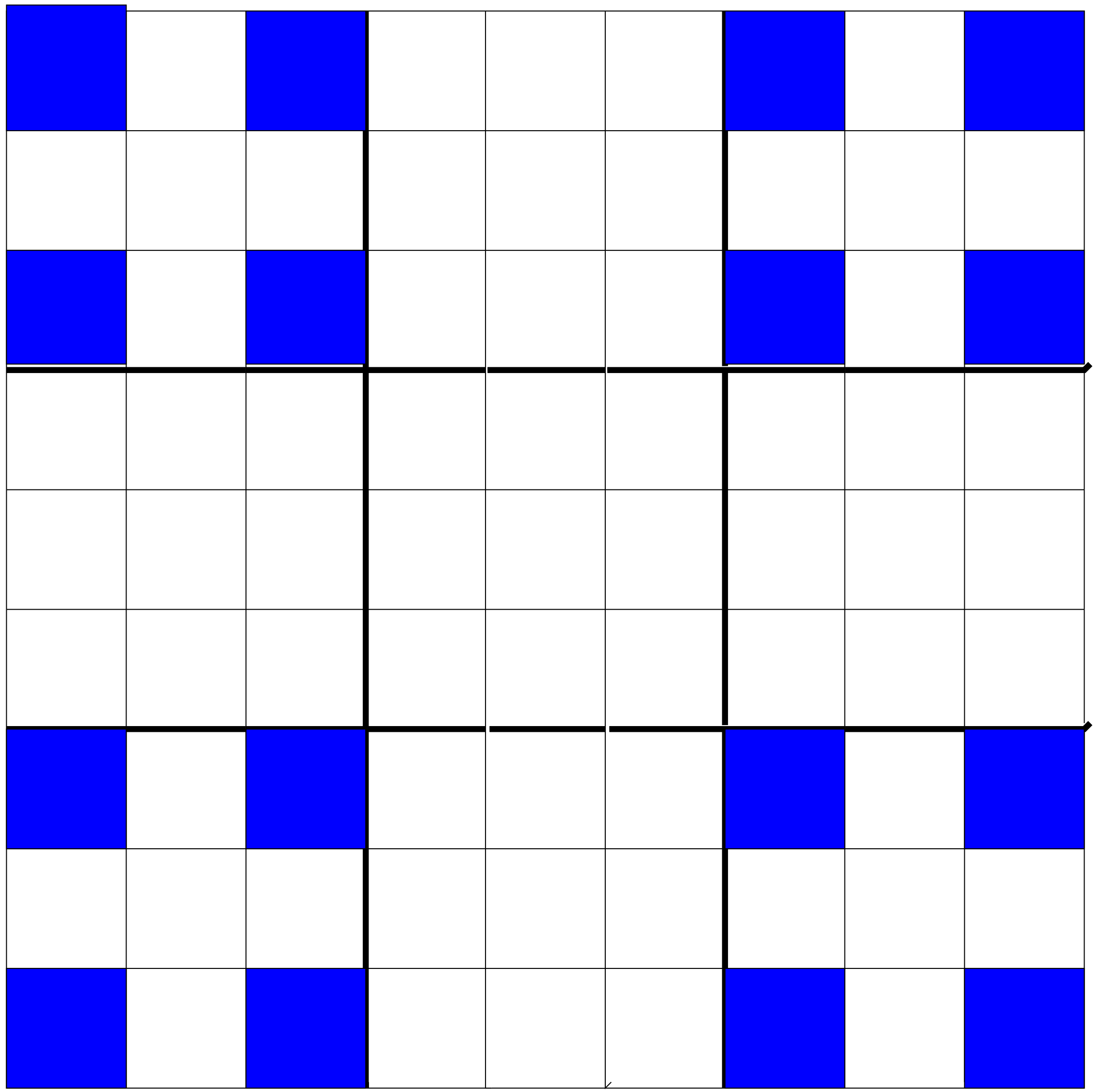
$\mu = \sigma$-invariant Borel probability measure on X

$\nu = \pi\mu$ on Y: $\nu(B) = \mu(\pi^{-1}B)$

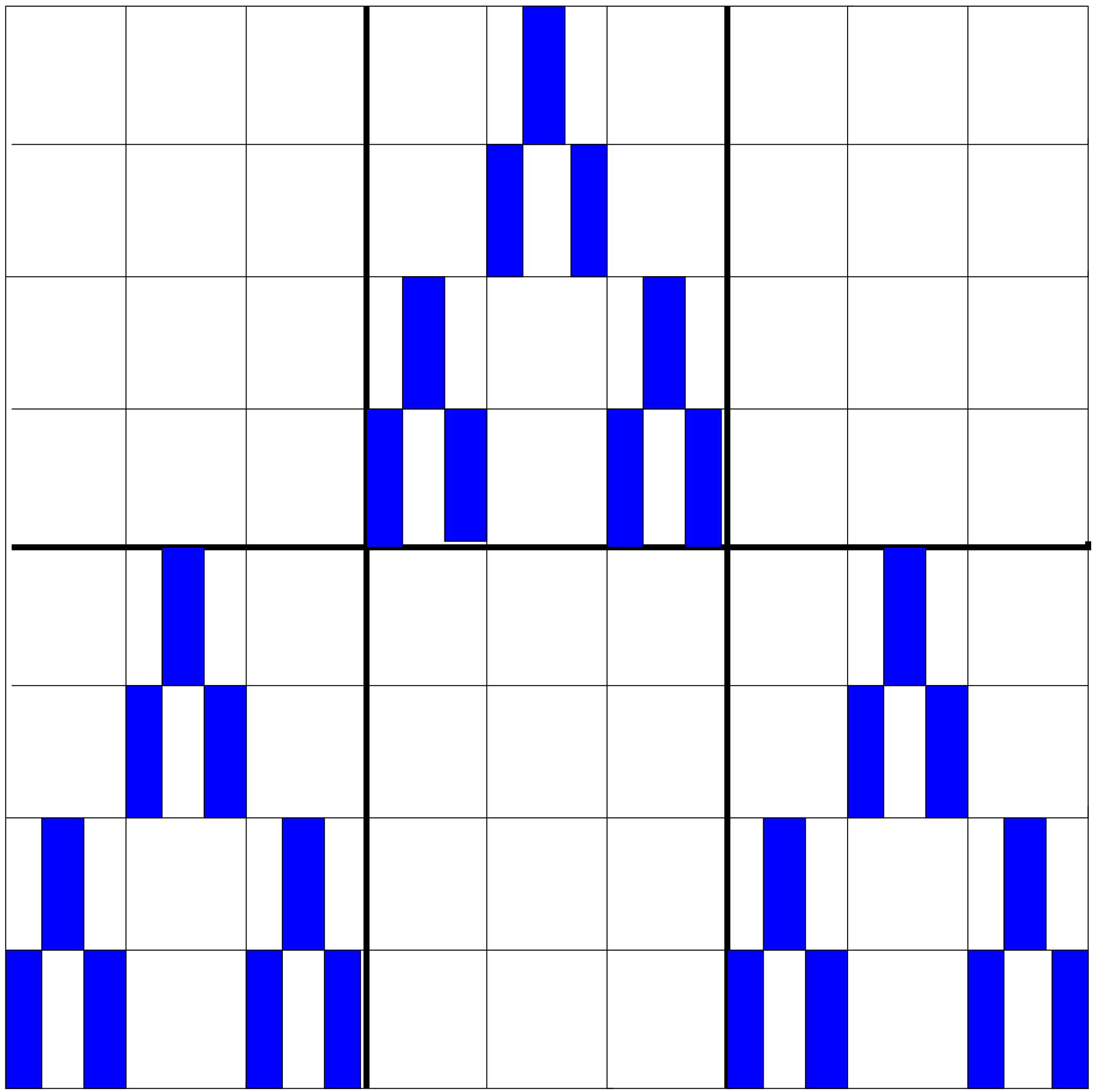
Sierpinski Carpet



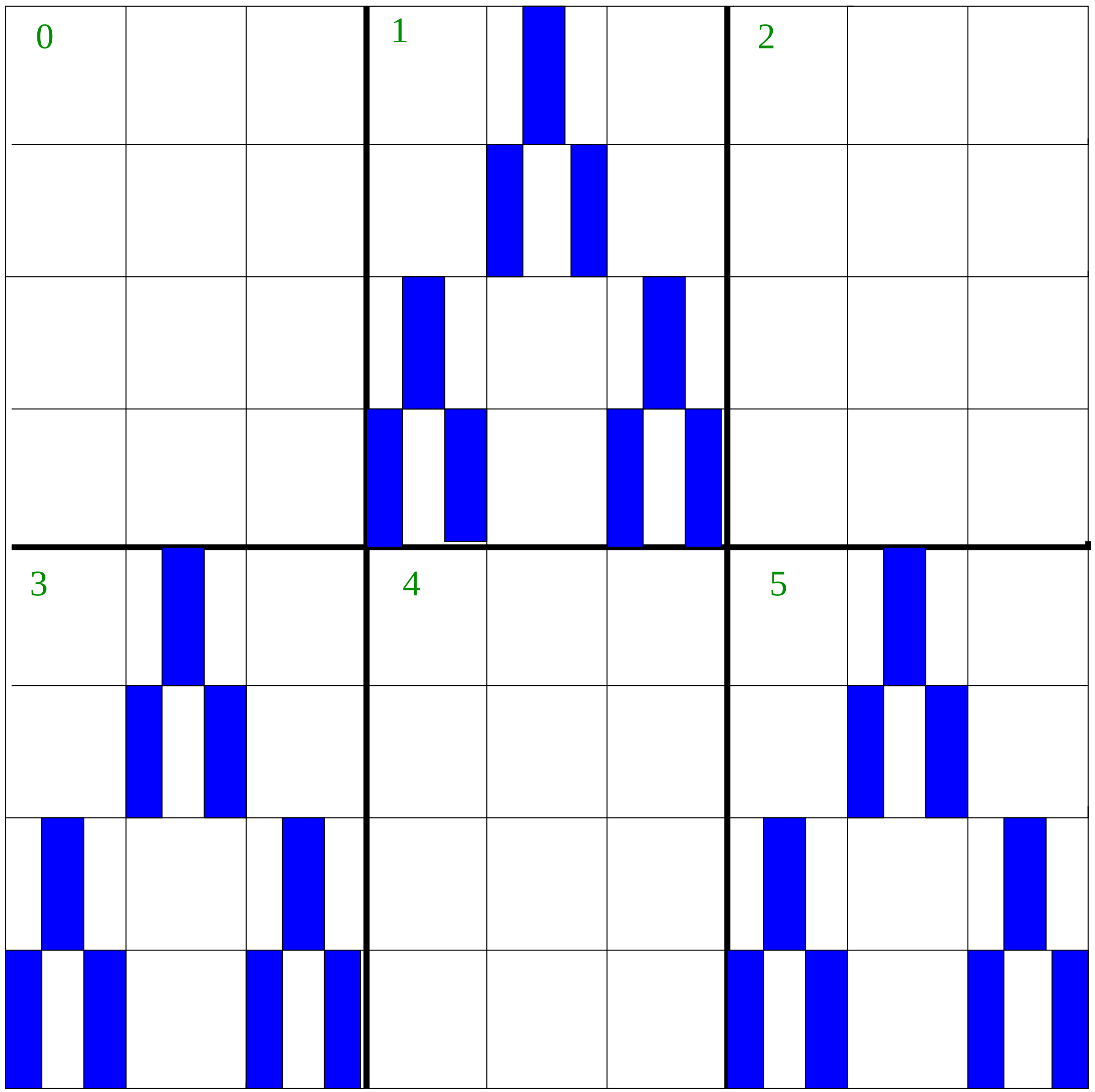
Sierpinski (or Dean Smith) Carpet



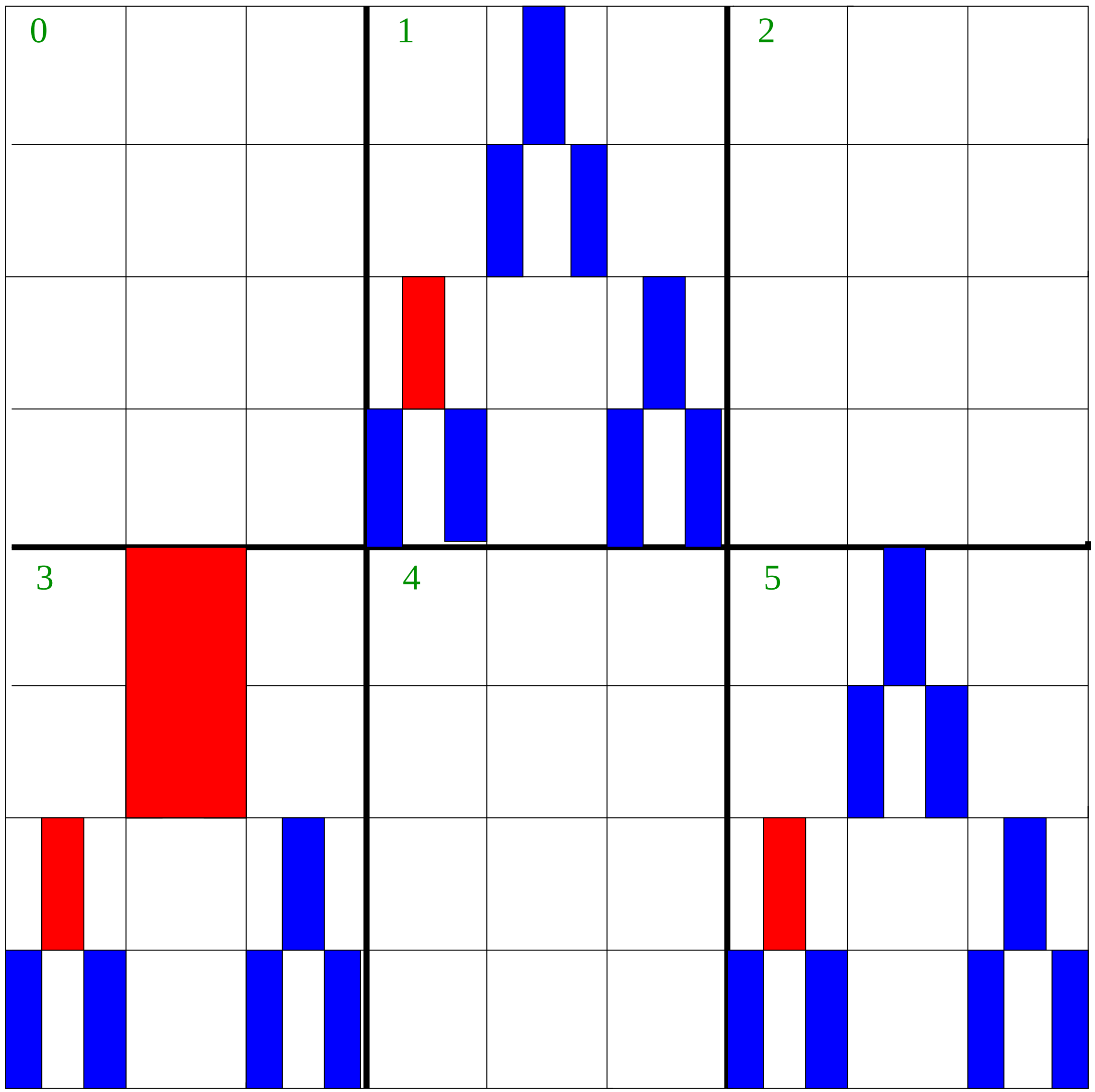
Nonconformal Carpet



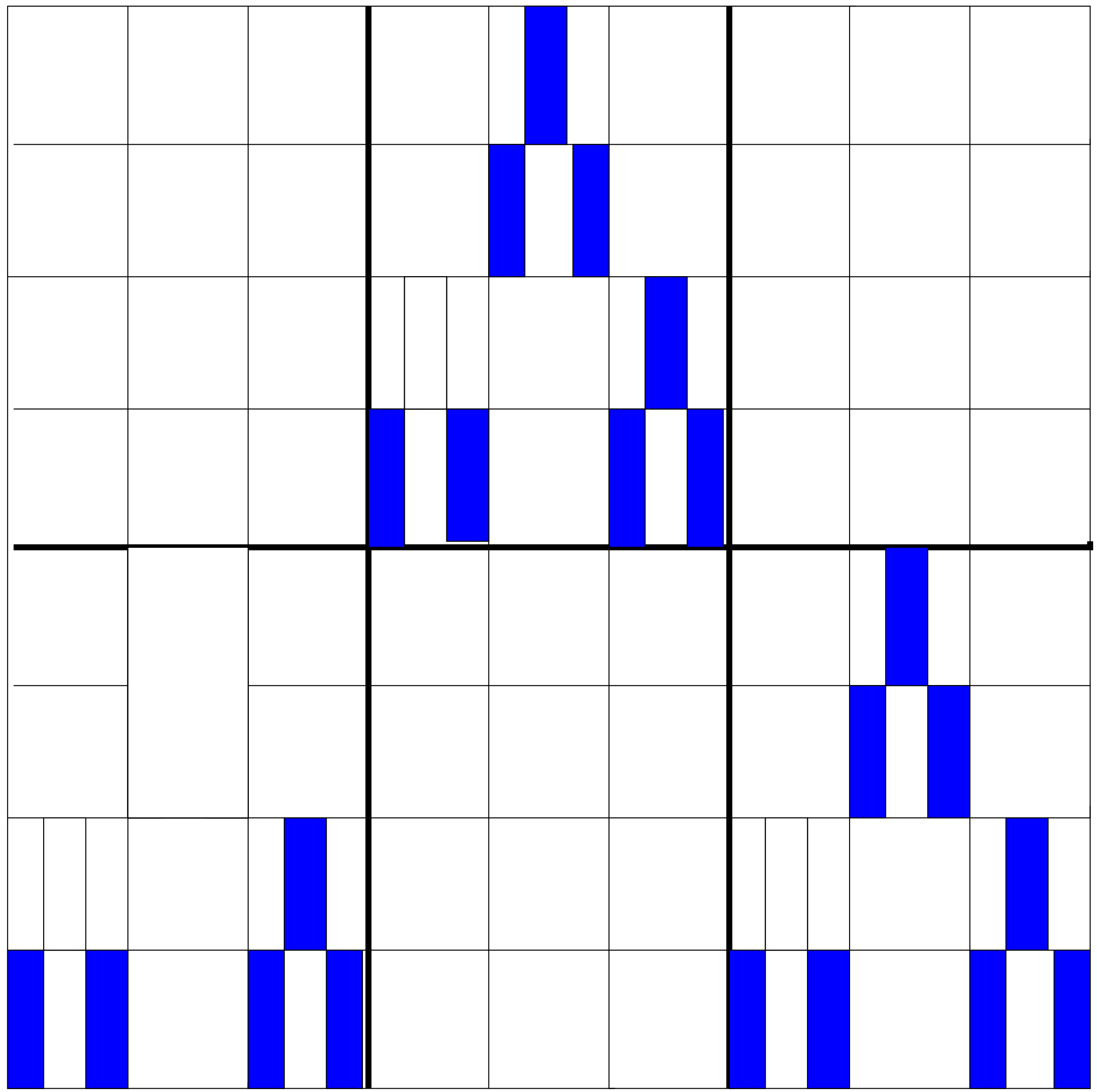
Nonconformal Carpet Coded



Disallow some transitions 31



More worn carpet



Information Loss

Models information loss, “deterministic noise”:

Information Loss

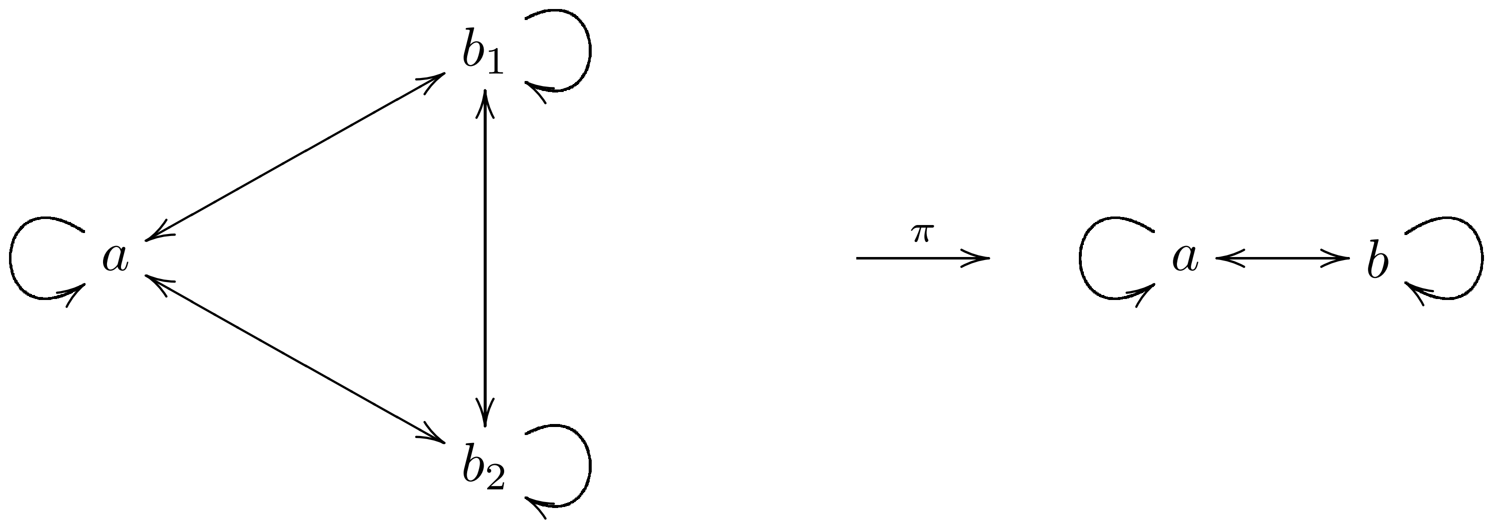
Models information loss, “deterministic noise”:

$h(\pi\mu) = R(\mu) =$information transmission rate

Information Loss

Models information loss, “deterministic noise”:

$h(\pi\mu) = R(\mu) =$information transmission rate

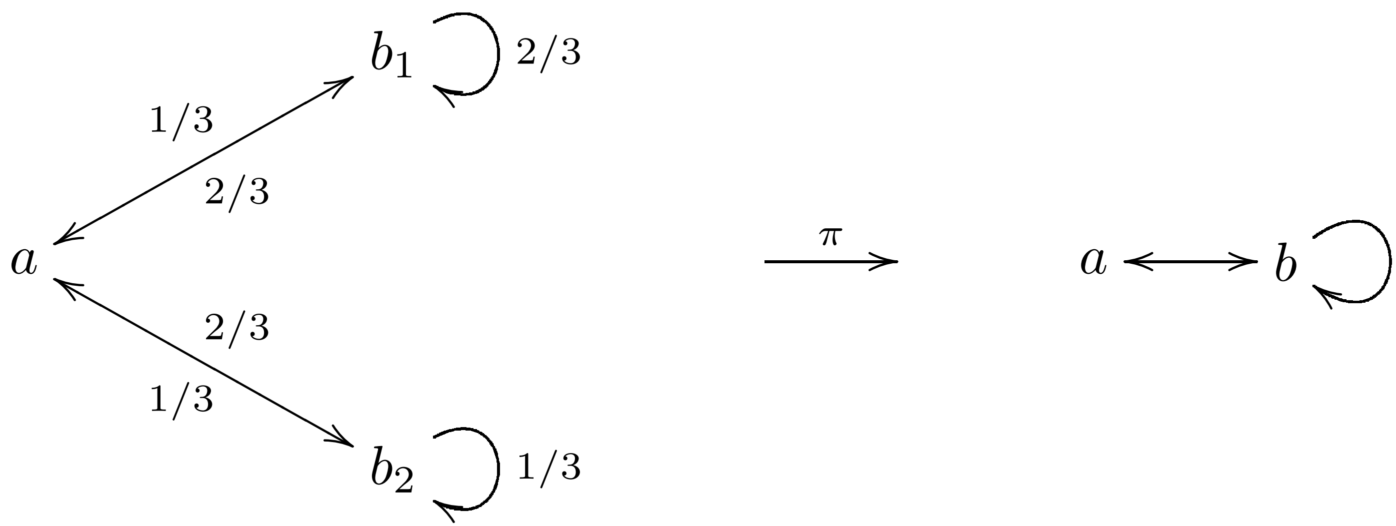


II. Some Bad Examples

Blackwell

II. Some Bad Examples

1. Blackwell



Blackwell

II. Some Bad Examples

1. Blackwell

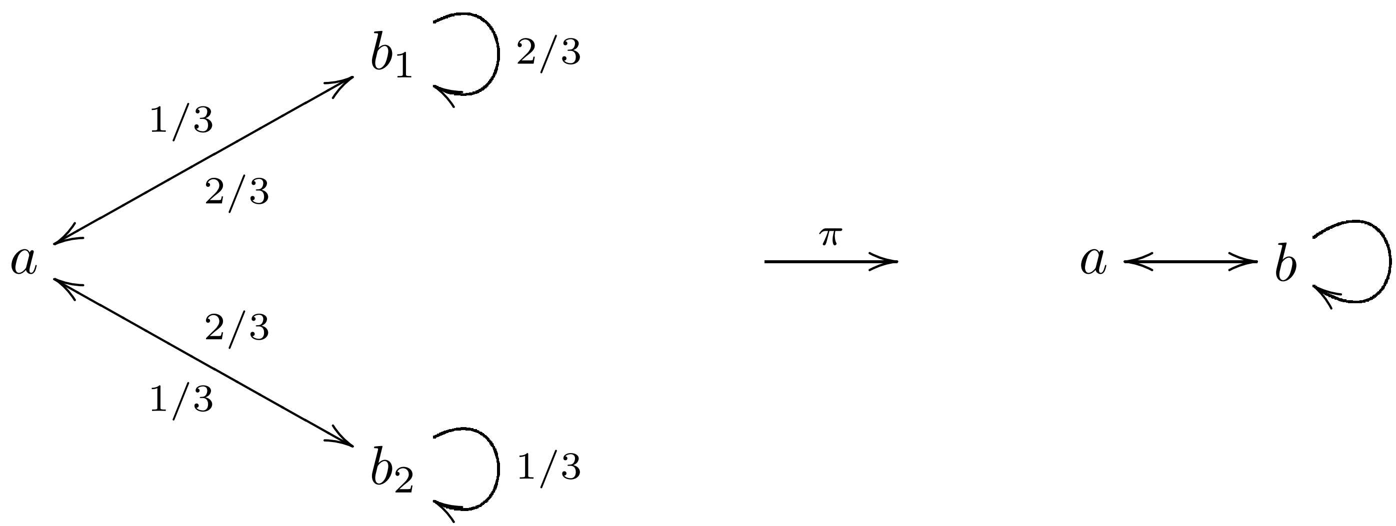


Image measure is not Markov.

Blackwell

II. Some Bad Examples

1. Blackwell

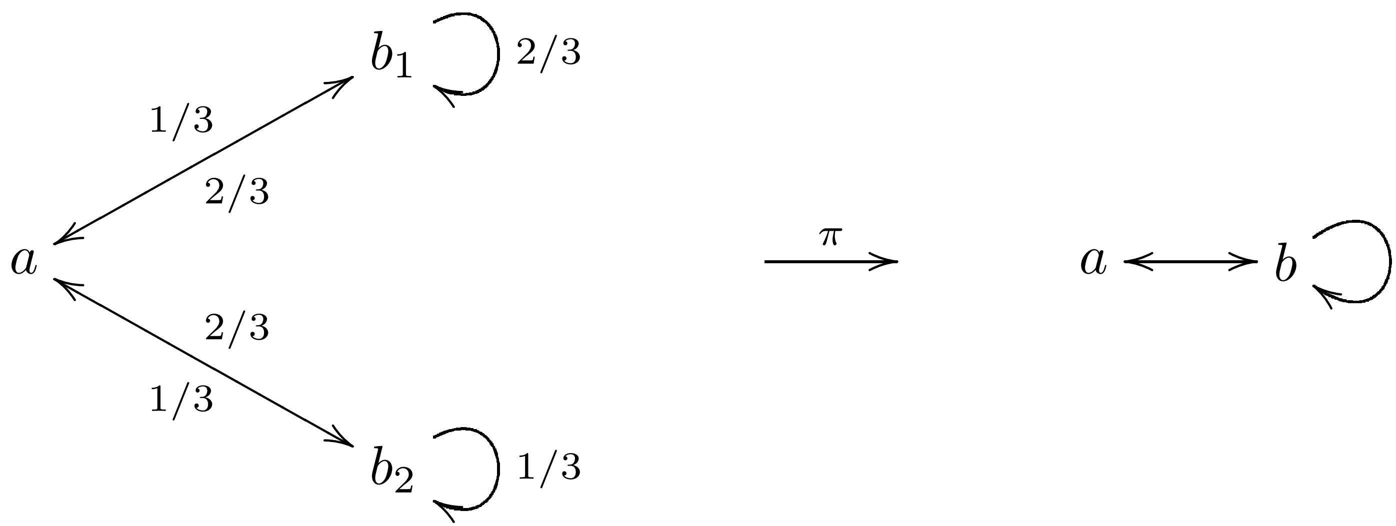
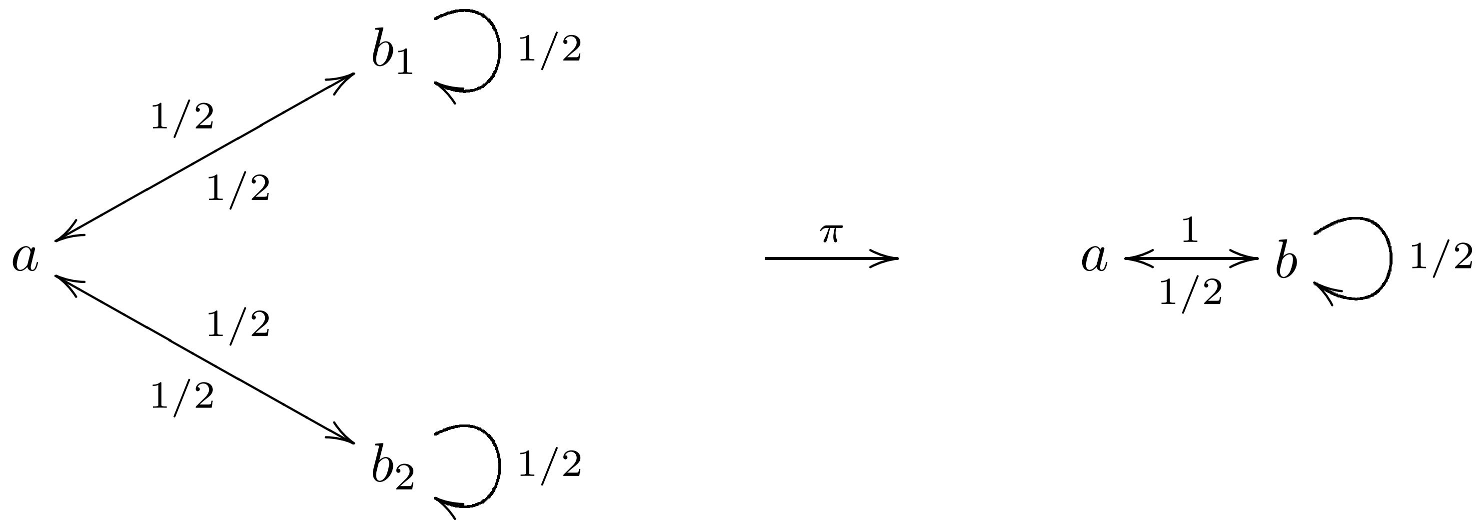


Image measure is not Markov.

Its entropy is hard to compute.

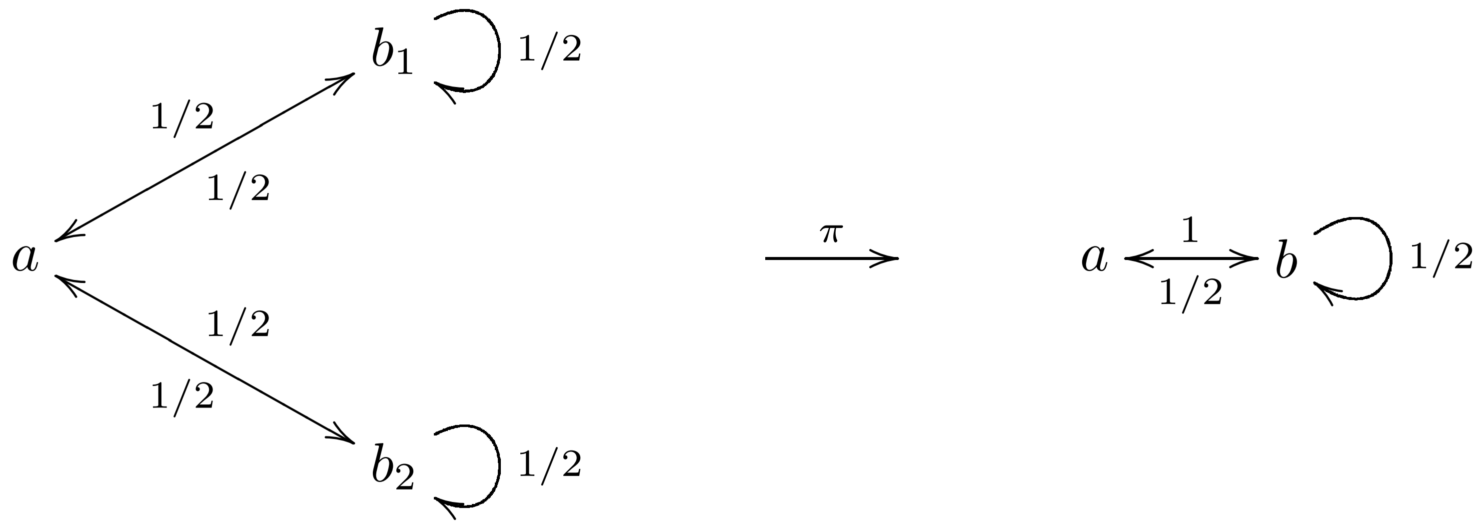
Markovian

But



Markovian

But

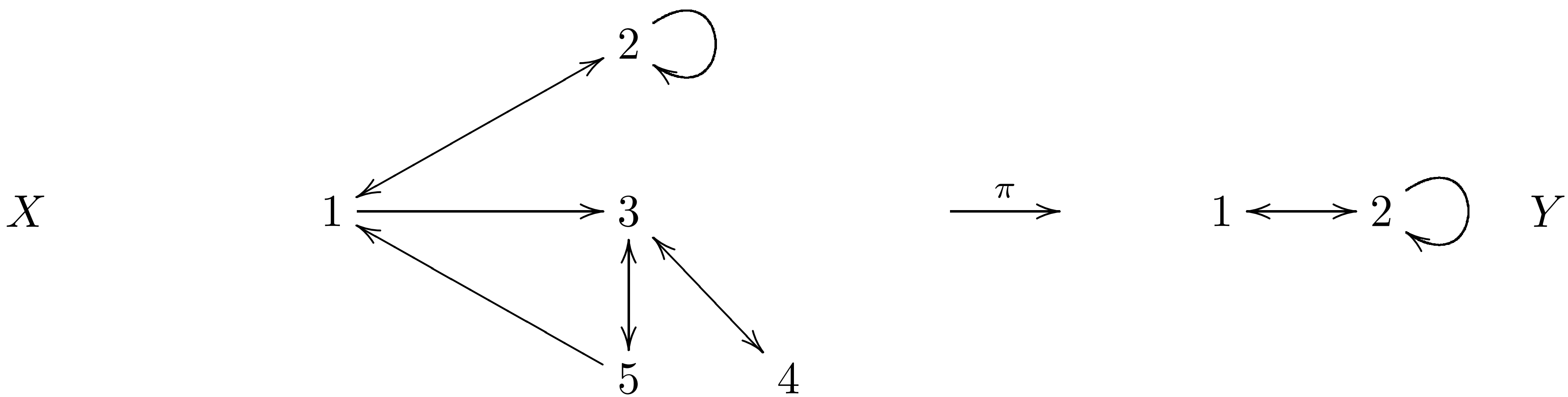


So the code is *Markovian*:

some Markov measure maps to a Markov measure.

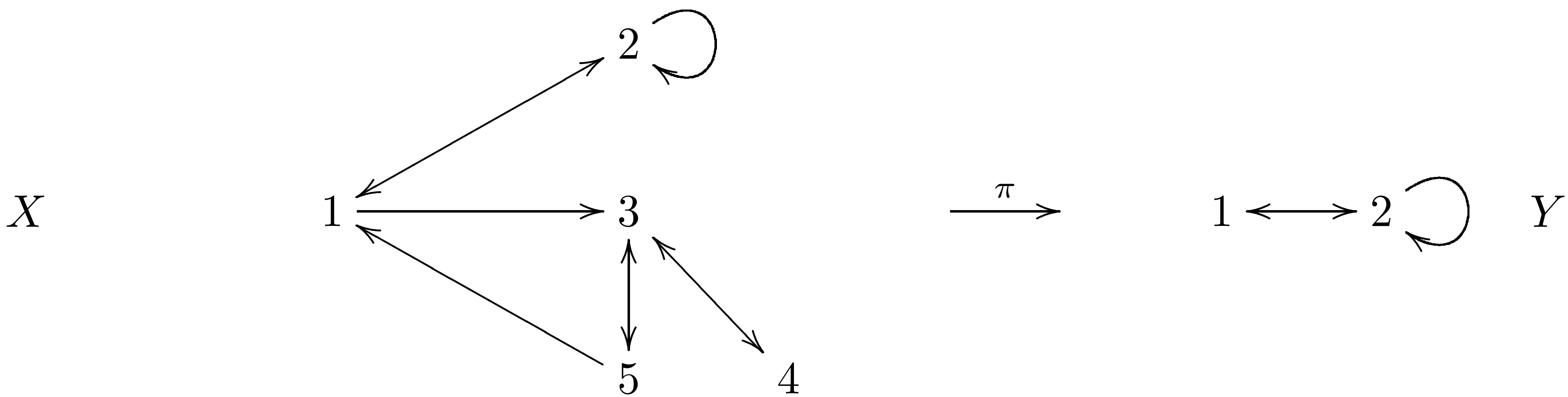
Shin non-Markovian

2. Shin non-Markovian example: some Markov measure does not lift to a Markov measure:



Shin non-Markovian

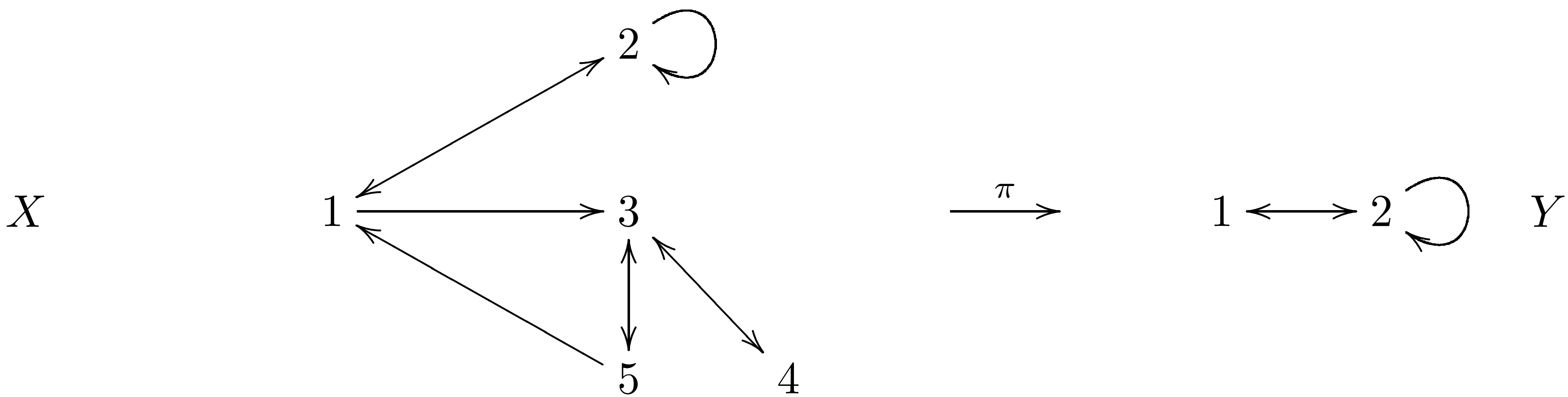
2. Shin non-Markovian example: some Markov measure does not lift to a Markov measure:



Actually *no* Markov lifts to a Markov.

Shin non-Markovian

2. Shin non-Markovian example: some Markov measure does not lift to a Markov measure:



Actually *no* Markov lifts to a Markov.

MPW: Blackwell-type example of a metrically sofic ν on Y that is not the

finite-to-one image of any Markov measure of any order anywhere.

Walters

3. Walters

$X = Y = \Sigma_2 =$ full 2-shift

Walters

3. Walters

$X = Y = \Sigma_2 =$ full 2-shift

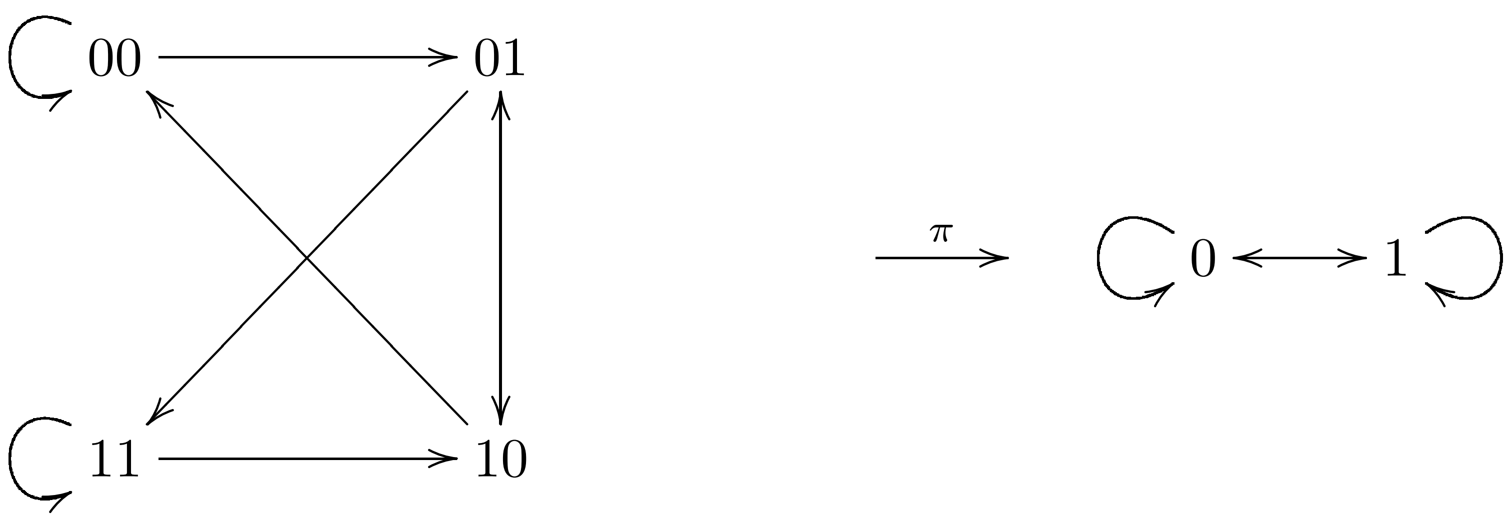
$\pi(x)_0 = x_0 + x_1 \mod 2$

3. Walters

$X = Y = \Sigma_2 =$ full 2-shift

$\pi(x)_0 = x_0 + x_1 \mod 2$

2-block recoding:



Walters–2

- Finite-to-one map, hence Markovian

Walters–2

- Finite-to-one map, hence Markovian
- Bernoulli $1/2, 1/2$ measure on Σ_2 is mapped to itself.

Walters–2

- Finite-to-one map, hence Markovian
- Bernoulli $1/2, 1/2$ measure on Σ_2 is mapped to itself.
- Every Markov ν on Y has a unique relatively maximal lift (in fact unique preimage), which is Markov

Walters–2

- Finite-to-one map, hence Markovian
- Bernoulli $1/2, 1/2$ measure on Σ_2 is mapped to itself.
- Every Markov ν on Y has a unique relatively maximal lift (in fact unique preimage), which is Markov
- For every ergodic ν on Y, *all* of $\pi^{-1}\{\nu\}$ consists of relatively maximal measures over ν, all having the same entropy as ν.

Walters—3

- If $p \neq 1/2$, the two measures on X that correspond to $\mathcal{B}(p, 1-p)$ and $\mathcal{B}(1-p, p)$ both map to ν_p on Y, which is fully supported.

Walters—3

- If $p \neq 1/2$, the two measures on X that correspond to $\mathcal{B}(p, 1-p)$ and $\mathcal{B}(1-p, p)$ both map to ν_p on Y, which is fully supported.
- $\nu_p =$ unique equilibrium state of $V_p \in \mathcal{C}(Y)$ on Y (Phelps).

Walters—3

- If $p \neq 1/2$, the two measures on X that correspond to $\mathcal{B}(p, 1-p)$ and $\mathcal{B}(1-p, p)$ both map to ν_p on Y, which is fully supported.
- $\nu_p =$ unique equilibrium state of $V_p \in \mathcal{C}(Y)$ on Y (Phelps).
- Then $\{$relatively maximal measures over $\nu_p\} = \pi^{-1}\{\nu_p\} =$ equilibrium states of $V_p \circ \pi + G \circ \pi = V_p \circ \pi$ ($G = 0$) (Walters)

Walters—3

- If $p \neq 1/2$, the two measures on X that correspond to $\mathcal{B}(p, 1-p)$ and $\mathcal{B}(1-p, p)$ both map to ν_p on Y, which is fully supported.
- $\nu_p =$ unique equilibrium state of $V_p \in \mathcal{C}(Y)$ on Y (Phelps).
- Then $\{$relatively maximal measures over $\nu_p\} = \pi^{-1}\{\nu_p\} =$ equilibrium states of $V_p \circ \pi + G \circ \pi = V_p \circ \pi$ ($G = 0$) (Walters)
- So this potential function $V_p \circ \pi$ has many equilibrium states.

Entropy increase

4. Marcus-P-Williams

For $\Sigma_3 \longrightarrow \Sigma_2$ as above, there is a 2-step Markov μ that projects to $\pi\mu = \mathcal{B}(1/2, 1/2)$

Entropy increase

4. Marcus-P-Williams

For $\Sigma_3 \to \Sigma_2$ as above, there is a 2-step Markov μ that projects to $\pi\mu = \mathcal{B}(1/2, 1/2)$

while its *1-step Markovization* $\mu^1 \to \pi\mu^1 \neq \mathcal{B}(1/2, 1/2)$.

Entropy increase

4. Marcus-P-Williams

For $\Sigma_3 \to \Sigma_2$ as above, there is a 2-step Markov μ that projects to $\pi\mu = \mathcal{B}(1/2, 1/2)$

while its *1-step Markovization* $\mu^1 \to \pi\mu^1 \neq \mathcal{B}(1/2, 1/2)$.

Thus $h(\mu^1) > h(\mu)$, while $h(\pi\mu^1) < h(\pi\mu)$.

III. Background Concepts

1. Relative pressure 1

III. Background Concepts

1. Relative pressure 1

III. Background Concepts

1. Relative pressure 1

Let $V \in \mathcal{C}(X)$ (a *potential function*)

III. Background Concepts

1. Relative pressure 1

Let $V \in \mathcal{C}(X)$ (a *potential function*)

For each $n = 1, 2, \cdots$ and $y \in Y$, let $D_n(y)$ be a set consisting of exactly one point from each nonempty set $[x_0 \cdots x_{n-1}] \cap \pi^{-1}(y)$.

III. Background Concepts

1. Relative pressure 1

Let $V \in \mathcal{C}(X)$ (a *potential function*)

For each $n = 1, 2, \cdots$ and $y \in Y$, let $D_n(y)$ be a set consisting of exactly one point from each nonempty set $[x_0 \cdots x_{n-1}] \cap \pi^{-1}(y)$.

$$P(\pi, V)(y) = \limsup_{n \to \infty} \frac{1}{n} \log \Bigg[\sum_{x \in D_n(y)} \exp \Big(\sum_{i=0}^{n-1} V(\sigma^i x) \Big) \Bigg].$$

Relative Entropy

2. Relative entropy 1 (Ledrappier-Walters)

Relative Entropy

2. Relative entropy 1 (Ledrappier-Walters)

For all $y \in Y$,

$$P(\pi, 0)(y) = \limsup_{n \to \infty} \frac{1}{n} \log \left| D_n(y) \right|$$

(with $V \equiv 0$).

Relative Variational Principle

3. Relative Variational Principle (Ledrappier-Walters)

Relative Variational Principle

3. Relative Variational Principle (Ledrappier-Walters)

For each $\nu \in M(Y)$,

$$\int P(\pi, V)\, d\nu = \sup \left\{ h(\mu) + \int V\, d\mu \,\middle|\, \pi\mu = \nu \right\} - h(\nu).$$

Relative Variational Principle

3. Relative Variational Principle (Ledrappier-Walters)

For each $\nu \in M(Y)$,

$$\int P(\pi, V)\, d\nu = \sup\left\{ h(\mu) + \int V\, d\mu \,\middle|\, \pi\mu = \nu \right\} - h(\nu).$$

In particular, for a fixed $\nu \in M(Y)$,

$$\sup\{h_\mu(X|Y) : \pi\mu = \nu\} = \sup\{h(\mu) - h(\nu) : \pi\mu = \nu\} = \int_Y P(\pi, 0)\, d\nu.$$

Relative Pressure—2

4. Relative pressure 2 (P-Shin)

Relative Pressure—2

4. Relative pressure 2 (P-Shin)

Theorem. *For each* $n = 1, 2, \cdots$ *and* $y \in Y$ *let* $E_n(y)$ *be a set consisting of exactly one point from each nonempty cylinder* $[x_0 \cdots x_{n-1}] \subset \pi^{-1}[y_0 \cdots y_{n-1}]$.

Then for each $V \in C(Y)$,

$$P(\pi, V)(y) = \limsup_{n \to \infty} \frac{1}{n} \log \Big[\sum_{x \in E_n(y)} \exp \Big(\sum_{i=0}^{n-1} V(\sigma^i x) \Big) \Big]$$

a.e. with respect to every invariant measure on Y.

Relative Pressure—2

4. Relative pressure 2 (P-Shin)

Theorem. *For each* $n = 1, 2, \cdots$ *and* $y \in Y$ *let* $E_n(y)$ *be a set consisting of exactly one point from each nonempty cylinder* $[x_0 \cdots x_{n-1}] \subset \pi^{-1}[y_0 \cdots y_{n-1}]$.

Then for each $V \in C(Y)$,

$$P(\pi, V)(y) = \limsup_{n \to \infty} \frac{1}{n} \log \Big[\sum_{x \in E_n(y)} \exp \Big(\sum_{i=0}^{n-1} V(\sigma^i x) \Big) \Big]$$

a.e. with respect to every invariant measure on Y.

Thus, we obtain the value of $P(\pi, V)(y)$ a.e. with respect to every invariant measure on Y if we delete from the definition of $D_n(y)$ the requirement that $x \in \pi^{-1}(y)$.

Relative Entropy—2

1.(P-Shin)

Relative Entropy—2

1.(P-Shin)

A finite-range, combinatorial approach to computing relative entropy:

Relative Entropy—2

1.(P-Shin)

A finite-range, combinatorial approach to computing relative entropy:
For μ relatively maximal over ν,

$$h_\mu(X|Y) = \int_Y \lim_{n\to\infty} \frac{1}{n} \log |\pi^{-1}[y_0 \dots y_{n-1}]| \, d\nu(y).$$

Relative Entropy—2

1.(P-Shin)

A finite-range, combinatorial approach to computing relative entropy: For μ relatively maximal over ν,

$$h_\mu(X|Y) = \int_Y \lim_{n\to\infty} \frac{1}{n} \log |\pi^{-1}[y_0 \dots y_{n-1}]| \, d\nu(y).$$

$$P(\pi, 0)(y) = \limsup_{n\to\infty} \frac{1}{n} \log |\pi^{-1}[y_0 \dots y_{n-1}]|$$

a.e. with respect to every invariant measure on Y.

Compensation Functions

2. Compensation function (Boyle-Tuncel, Walters)

Compensation Functions

2. Compensation function (Boyle-Tuncel, Walters)

A continuous function $F : X \to \mathbb{R}$ such that

$$P_Y(V) = P_X(V \circ \pi + F) \quad \text{for all } V \in \mathcal{C}(Y).$$

Compensation Functions

2. Compensation function (Boyle-Tuncel, Walters)

A continuous function $F : X \to \mathbb{R}$ such that

$$P_Y(V) = P_X(V \circ \pi + F) \quad \text{for all } V \in \mathcal{C}(Y).$$

Idea: Because $\pi : \mathcal{M}(X) \to \mathcal{M}(Y)$ is many-to-one, we always have

$$\begin{aligned} P_Y(V) &= \sup\{h_\nu(\sigma) + \int_Y V \, d\nu : \nu \in \mathcal{M}(Y)\} \\ &\leq \sup\{h_\mu(\sigma) + \int_X V \circ \pi \, d\mu : \mu \in \mathcal{M}(X)\}. \end{aligned}$$

Compensation Functions

2. Compensation function (Boyle-Tuncel, Walters)

A continuous function $F : X \to \mathbb{R}$ such that

$$P_Y(V) = P_X(V \circ \pi + F) \quad \text{for all } V \in \mathcal{C}(Y).$$

Idea: Because $\pi : \mathcal{M}(X) \to \mathcal{M}(Y)$ is many-to-one, we always have

$$\begin{aligned} P_Y(V) &= \sup\{h_\nu(\sigma) + \int_Y V\, d\nu : \nu \in \mathcal{M}(Y)\} \\ &\leq \sup\{h_\mu(\sigma) + \int_X V \circ \pi\, d\mu : \mu \in \mathcal{M}(X)\}. \end{aligned}$$

F takes into account, for all potential functions V on Y at once, the extra freedom, information, or free energy available in X as compared to Y because of the ability to move around in fibers over points of Y.

Properties of Compensation Functions

- For SFT's X and Y, there is always a compensation function.

Properties of Compensation Functions

- For SFT's X and Y, there is always a compensation function.
- The following are equivalent:

Properties of Compensation Functions

- For SFT's X and Y, there is always a compensation function.
- The following are equivalent:
  1. There are Markov μ, ν with $\pi\mu = \nu$.

Properties of Compensation Functions

- For SFT's X and Y, there is always a compensation function.
- The following are equivalent:
  1. There are Markov μ, ν with $\pi\mu = \nu$.
  2. For every Markov ν on Y there are uncountably many Markov μ on X with $\pi\mu = \nu$.

Properties of Compensation Functions

- For SFT's X and Y, there is always a compensation function.
- The following are equivalent:
  1. There are Markov μ, ν with $\pi\mu = \nu$.
  2. For every Markov ν on Y there are uncountably many Markov μ on X with $\pi\mu = \nu$.
  3. There is a *locally constant* compensation function.

Properties of Compensation Functions

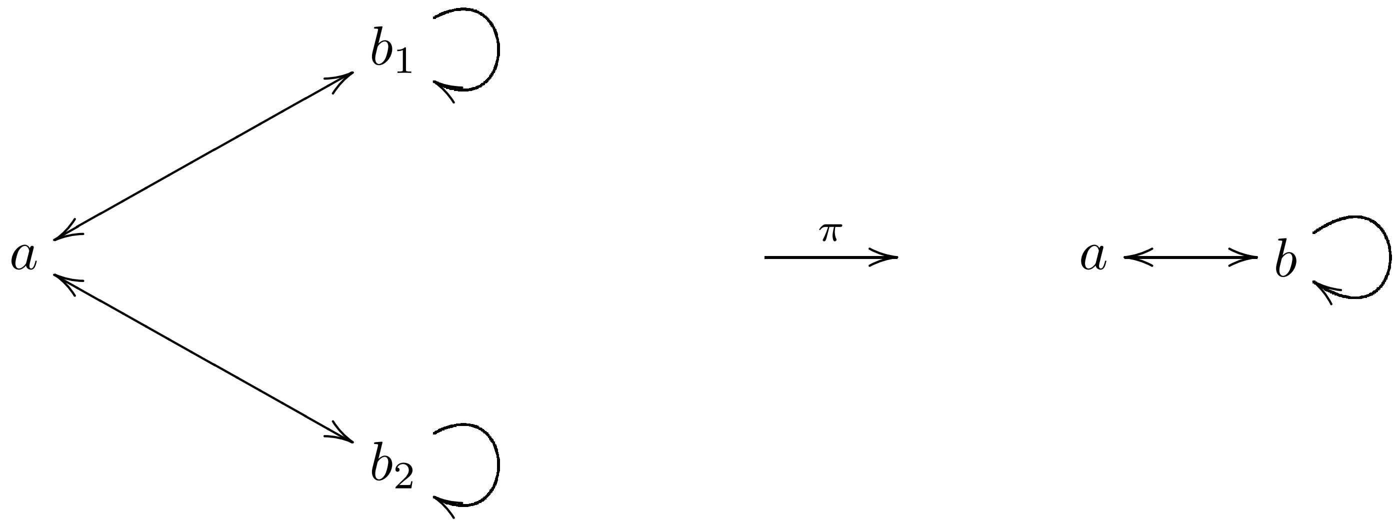
- For SFT's X and Y, there is always a compensation function.
- The following are equivalent:
  1. There are Markov μ, ν with $\pi\mu = \nu$.
  2. For every Markov ν on Y there are uncountably many Markov μ on X with $\pi\mu = \nu$.
  3. There is a *locally constant* compensation function.
- $\pi(\max_X) = \max_Y$ if and only if there is a *constant* compensation function.

Properties of Compensation Functions

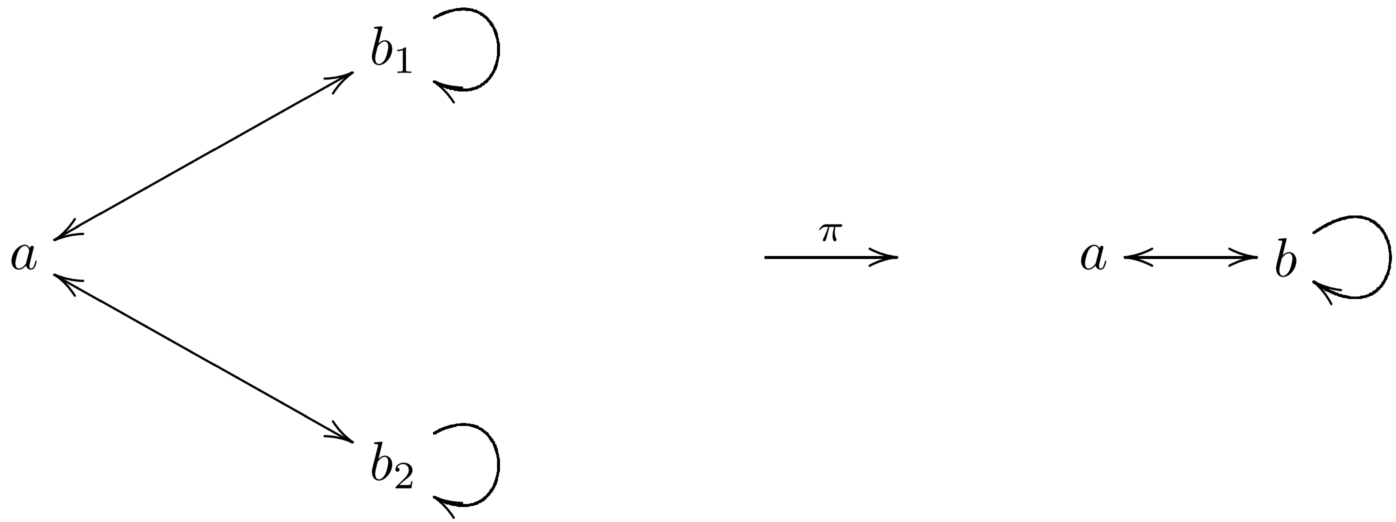
- For SFT's X and Y, there is always a compensation function.
- The following are equivalent:
  1. There are Markov μ, ν with $\pi\mu = \nu$.
  2. For every Markov ν on Y there are uncountably many Markov μ on X with $\pi\mu = \nu$.
  3. There is a *locally constant* compensation function.
- $\pi(\max_X) = \max_Y$ if and only if there is a *constant* compensation function.
- If $G \in \mathcal{F}(Y)$ (Walters class), then $G \circ \pi$ is a (*saturated*) compensation function if and only if there is $c > 0$ such that

$$\frac{1}{c} \leq e^{S_n G(y)} \left|\pi^{-1}[y_0 \dots y_{n-1}]\right| \leq c \text{ for all } y, n.$$

3. Example of a Compensation Function

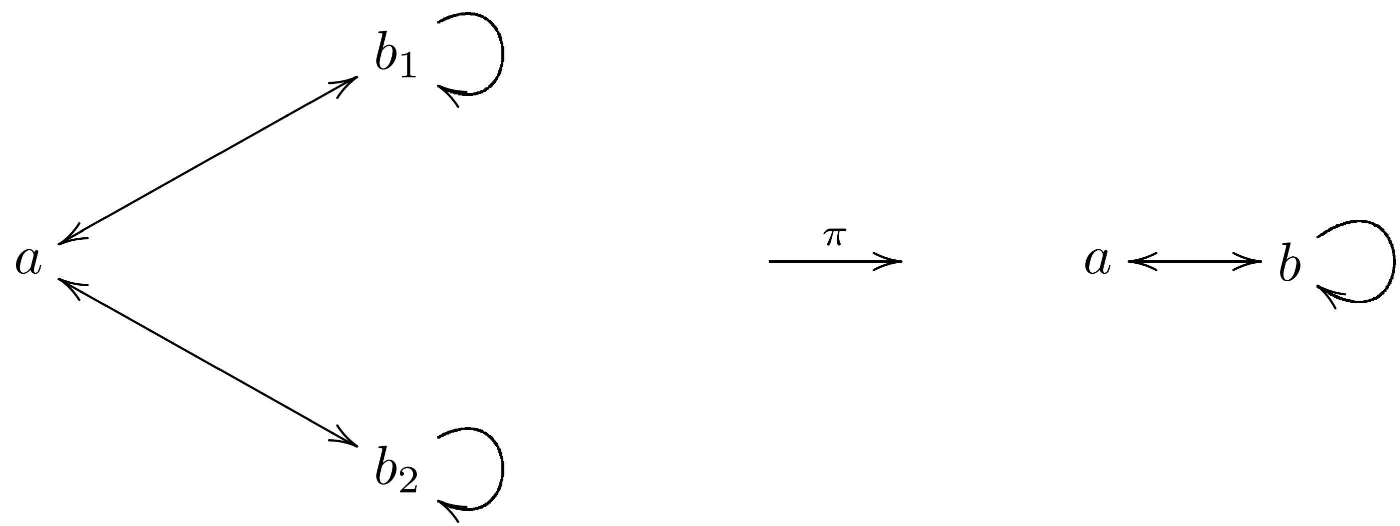


3. Example of a Compensation Function



$$G(y) = \begin{cases} -\log 2 & \text{if } y = .a\ldots \\ 0 & \text{if } y = .b\ldots \end{cases}$$

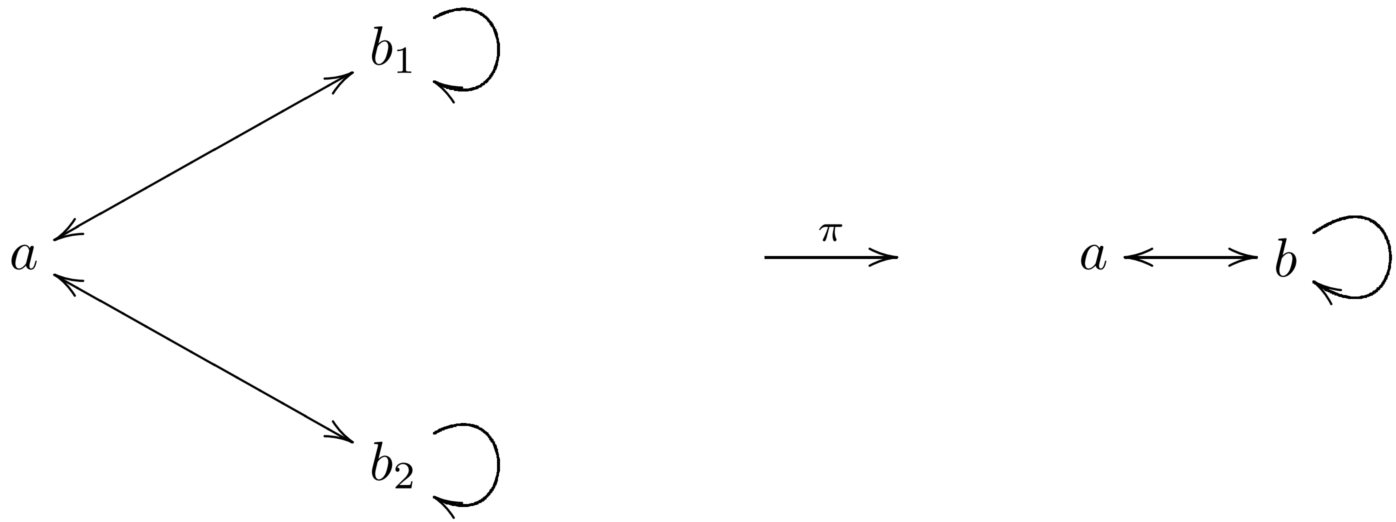
3. Example of a Compensation Function



$$G(y) = \begin{cases} -\log 2 & \text{if } y = .a\ldots \\ 0 & \text{if } y = .b\ldots \end{cases}$$

$e^{-G(y)}$ measures *branching freedom* at y (or x).

3. Example of a Compensation Function

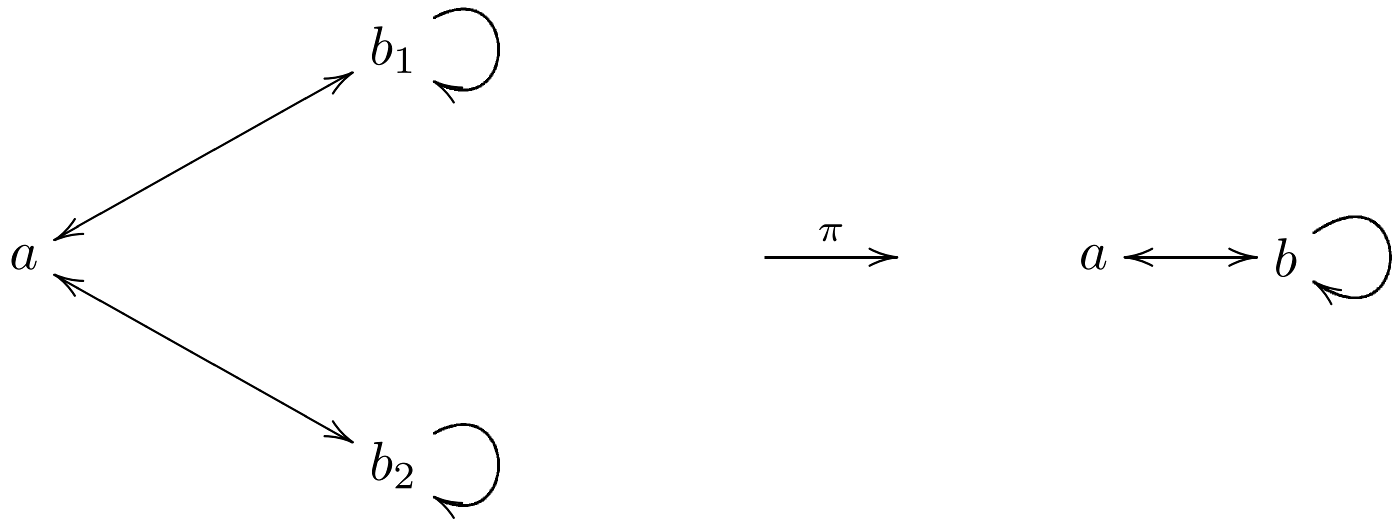


$$G(y) = \begin{cases} -\log 2 & \text{if } y = .a\ldots \\ 0 & \text{if } y = .b\ldots \end{cases}$$

$e^{-G(y)}$ measures *branching freedom* at y (or x).

$\pi^{-1}[y_0 \ldots y_{n-1}] \sim 2^{\#a} \sim e^{-S_n G(y)}$:

3. Example of a Compensation Function



$$G(y) = \begin{cases} -\log 2 & \text{if } y = .a\ldots \\ 0 & \text{if } y = .b\ldots \end{cases}$$

$e^{-G(y)}$ measures *branching freedom* at y (or x).

$\pi^{-1}[y_0 \ldots y_{n-1}| \sim 2^{\#a} \sim e^{-S_n G(y)}$:

When in y we see $ab^{k_1}ab^{k_2}a\ldots ab^{k_r}a$, multiply in: 1 at each b, 2 at each a .

IV. Relatively Maximal Measures

1. The problem:
For fixed ν on Y, look for the $\mu \in \pi^{-1}\nu$ which have maximal entropy in that fiber.

IV. Relatively Maximal Measures

1. The problem:
For fixed ν on Y, look for the $\mu \in \pi^{-1}\nu$ which have maximal entropy in that fiber.

Unique? (Shannon-Parry in case $Y = \{1\}$.)

IV. Relatively Maximal Measures

1. The problem:
For fixed ν on Y, look for the $\mu \in \pi^{-1}\nu$ which have maximal entropy in that fiber.

Unique? (Shannon-Parry in case $Y = \{1\}$.)

No—plenty of examples, including with fully supported ν ($\mathcal{B}(p, 1-p), \mathcal{B}(1-p, p)$).

IV. Relatively Maximal Measures

1. The problem:
For fixed ν on Y, look for the $\mu \in \pi^{-1}\nu$ which have maximal entropy in that fiber.

Unique? (Shannon-Parry in case $Y = \{1\}$.)

No—plenty of examples, including with fully supported ν ($\mathcal{B}(p, 1-p), \mathcal{B}(1-p, p)$).

Theorem (Shin)**.** *Suppose that* $\nu \in \mathcal{E}(Y)$ *and* $\pi\mu = \nu$*. Then* μ *is relatively maximal over* ν *if and only if there is* $V \in \mathcal{C}(Y)$ *such that* μ *is an equilibrium state of* $V \circ \pi$*.*

Lifting Markov Measures

- If there is a *locally constant* saturated compensation function $G \circ \pi$, then every Markov measure on Y has a unique relatively maximal lift, which is Markov, because then the relatively maximal measures over an equilibrium state of $V \in \mathcal{C}(Y)$ are the equilibrium states of $V \circ \pi + G \circ \pi$ (Walters).

Lifting Markov Measures

- If there is a *locally constant* saturated compensation function $G \circ \pi$, then every Markov measure on Y has a unique relatively maximal lift, which is Markov, because then the relatively maximal measures over an equilibrium state of $V \in \mathcal{C}(Y)$ are the equilibrium states of $V \circ \pi + G \circ \pi$ (Walters).
- Further, μ_X is the unique equilibrium state of the potential function 0 on X; and the relatively maximal measures over μ_Y are the equilibrium states of $G \circ \pi$.

Only Finitely Many

2. An answer

Only Finitely Many

2. An answer

Theorem (P-Quas-Shin)**.** *For each ergodic* ν *on* Y*, there are only a finite number of relatively maximal measures over* ν*.*

In fact, the number of ergodic invariant measures of maximal entropy in the fiber $\pi^{-1}\{\nu\}$ *is at most*

$$N_\nu(\pi) = \min\{|\pi^{-1}\{b\}| : b \in \mathcal{A}(Y), \nu[b] > 0\}.$$

Only Finitely Many

2. An answer

Theorem (P-Quas-Shin). *For each ergodic ν on Y, there are only a finite number of relatively maximal measures over ν.*

In fact, the number of ergodic invariant measures of maximal entropy in the fiber $\pi^{-1}\{\nu\}$ is at most

$$N_\nu(\pi) = \min\{|\pi^{-1}\{b\}| : b \in \mathcal{A}(Y), \nu[b] > 0\}.$$

This follows from

Only Finitely Many

2. An answer

Theorem (P-Quas-Shin)**.** *For each ergodic* ν *on* Y*, there are only a finite number of relatively maximal measures over* ν*.*

In fact, the number of ergodic invariant measures of maximal entropy in the fiber $\pi^{-1}\{\nu\}$ *is at most*

$$N_\nu(\pi) = \min\{|\pi^{-1}\{b\}| : b \in \mathcal{A}(Y), \nu[b] > 0\}.$$

This follows from

Theorem (P-Quas-Shin)**.** *For each ergodic* ν *on* Y*, any two distinct ergodic measures on* X *of maximal entropy in the fiber* $\pi^{-1}\{\nu\}$ *are* relatively orthogonal*.*

Relatively Independent Joining

For $\mu_1, \ldots, \mu_n \in \mathcal{M}(X)$ with $\pi\mu_i = \nu$ for all i, their *relatively independent joining* $\hat{\mu}$ over ν is defined by:

Relatively Independent Joining

For $\mu_1, \ldots, \mu_n \in \mathcal{M}(X)$ with $\pi\mu_i = \nu$ for all i, their *relatively independent joining* $\hat{\mu}$ over ν is defined by:

if $A_1, \ldots, A_n$ are measurable subsets of X and $\mathcal{F}$ is the σ-algebra of Y, then

$$\hat{\mu}(A_1 \times \ldots \times A_n) = \int_Y \prod_{i=1}^{n} \mathbb{E}_{\mu_i}(\mathbf{1}_{A_i} | \pi^{-1}\mathcal{F}) \circ \pi^{-1} \, d\nu.$$

Relatively Independent Joining

For $\mu_1, \ldots, \mu_n \in \mathcal{M}(X)$ with $\pi\mu_i = \nu$ for all i, their *relatively independent joining* $\hat{\mu}$ over ν is defined by:

if $A_1, \ldots, A_n$ are measurable subsets of X and $\mathcal{F}$ is the σ-algebra of Y, then

$$\hat{\mu}(A_1 \times \ldots \times A_n) = \int_Y \prod_{i=1}^{n} \mathbb{E}_{\mu_i}(\mathbf{1}_{A_i} | \pi^{-1}\mathcal{F}) \circ \pi^{-1} \, d\nu.$$

Two measures $\mu_1, \mu_2 \in \mathcal{E}(X)$ with $\pi\mu_1 = \pi\mu_2 = \nu$ are *relatively orthogonal* (over ν), $\mu_1 \perp_\nu \mu_2$, if

$$(\mu_1 \otimes_\nu \mu_2)\{(u,v) \in X \times X : u_0 = v_0\} = 0.$$

Relatively Independent Joining

For $\mu_1, \ldots, \mu_n \in \mathcal{M}(X)$ with $\pi\mu_i = \nu$ for all i, their *relatively independent joining* $\hat{\mu}$ over ν is defined by:

if $A_1, \ldots, A_n$ are measurable subsets of X and $\mathcal{F}$ is the σ-algebra of Y, then

$$\hat{\mu}(A_1 \times \ldots \times A_n) = \int_Y \prod_{i=1}^{n} \mathbb{E}_{\mu_i}(\mathbf{1}_{A_i} | \pi^{-1}\mathcal{F}) \circ \pi^{-1} \, d\nu.$$

Two measures $\mu_1, \mu_2 \in \mathcal{E}(X)$ with $\pi\mu_1 = \pi\mu_2 = \nu$ are *relatively orthogonal* (over ν), $\mu_1 \perp_\nu \mu_2$, if

$$(\mu_1 \otimes_\nu \mu_2)\{(u, v) \in X \times X : u_0 = v_0\} = 0.$$

There is zero probability of coincidence.

Proof of First Theorem

3. The second theorem implies the first.

Proof of First Theorem

3. The second theorem implies the first.

Suppose that we have $n > N_\nu(\pi)$ ergodic measures $\mu_1, \ldots, \mu_n$ on X, each projecting to ν and each of maximal entropy in the fiber $\pi^{-1}\{\nu\}$.

Proof of First Theorem

3. The second theorem implies the first.

Suppose that we have $n > N_\nu(\pi)$ ergodic measures $\mu_1, \ldots, \mu_n$ on X, each projecting to ν and each of maximal entropy in the fiber $\pi^{-1}\{\nu\}$.

Form the relatively independent joining $\hat{\mu}$ on X^n of the measures μ_i as above.

Proof of First Theorem

3. The second theorem implies the first.

Suppose that we have $n > N_\nu(\pi)$ ergodic measures $\mu_1, \ldots, \mu_n$ on X, each projecting to ν and each of maximal entropy in the fiber $\pi^{-1}\{\nu\}$.

Form the relatively independent joining $\hat{\mu}$ on X^n of the measures μ_i as above.

Let b be a symbol in the alphabet of Y such that b has $N_\nu(\pi)$ preimages

$a_1, \ldots, a_{N_\nu(\pi)}$ under the block map π.

Pigeonholing

Since $n > N_\nu(\pi)$, for every $\hat{x} \in \phi^{-1}[b]$ there are $i \neq j$ with $(p_i \hat{x})_0 = (p_j \hat{x})_0$.

Pigeonholing

Since $n > N_\nu(\pi)$, for every $\hat{x} \in \phi^{-1}[b]$ there are $i \neq j$ with $(p_i \hat{x})_0 = (p_j \hat{x})_0$.

At least one of the sets $S_{i,j} = \{\hat{x} \in X^n : (p_i \hat{x})_0 = (p_j \hat{x})_0\}$ must have positive $\hat{\mu}$-measure,

Pigeonholing

Since $n > N_\nu(\pi)$, for every $\hat{x} \in \phi^{-1}[b]$ there are $i \neq j$ with $(p_i\hat{x})_0 = (p_j\hat{x})_0$.

At least one of the sets $S_{i,j} = \{\hat{x} \in X^n : (p_i\hat{x})_0 = (p_j\hat{x})_0\}$ must have positive $\hat{\mu}$-measure,

and then also

$$(\mu_i \otimes_\nu \mu_j)\{(u,v) \in X \times X : \pi u = \pi v, u_0 = v_0\} > 0,$$

contradicting relative orthogonality.

Pigeonholing

Since $n > N_\nu(\pi)$, for every $\hat{x} \in \phi^{-1}[b]$ there are $i \neq j$ with $(p_i\hat{x})_0 = (p_j\hat{x})_0$.

At least one of the sets $S_{i,j} = \{\hat{x} \in X^n : (p_i\hat{x})_0 = (p_j\hat{x})_0\}$ must have positive $\hat{\mu}$-measure,

and then also

$$(\mu_i \otimes_\nu \mu_j)\{(u,v) \in X \times X : \pi u = \pi v, u_0 = v_0\} > 0,$$

contradicting relative orthogonality.

(If you have more measures than preimage symbols, two of those measures have to coincide on one of the symbols: with respect to each measure, that symbol a.s. appears infinitely many times in the same place.)

Interleaving Sequences

Idea of the proof of the second theorem.

Interleaving Sequences

Idea of the proof of the second theorem.

Writing p_i for the projection $X^n \to X$ onto the i'th coordinate, we note that for $\hat{\mu}$-almost every $\hat{x}$ in X^n, $\pi(p_i(\hat{x}))$ is independent of i; denote it by $\phi(\hat{x})$.

Interleaving Sequences

Idea of the proof of the second theorem.

Writing p_i for the projection $X^n \to X$ onto the i'th coordinate, we note that for $\hat{\mu}$-almost every $\hat{x}$ in X^n, $\pi(p_i(\hat{x}))$ is independent of i; denote it by $\phi(\hat{x})$.

If there two relatively maximal measures over ν which are not relatively orthogonal, then the measures can be 'mixed' to give a measure with greater entropy.

Interleaving Sequences

Idea of the proof of the second theorem.

Writing p_i for the projection $X^n \to X$ onto the i'th coordinate, we note that for $\hat{\mu}$-almost every $\hat{x}$ in X^n, $\pi(p_i(\hat{x}))$ is independent of i; denote it by $\phi(\hat{x})$.

If there two relatively maximal measures over ν which are not relatively orthogonal, then the measures can be 'mixed' to give a measure with greater entropy.

We concatenate words from the two processes, using the the fact that the two measures are supported on sequences that agree infinitely often. Since X is a 1-step SFT, we can switch over whenever a coincidence occurs.

Weaving In More Entropy

Let $w \in \mathcal{B}(1/2, 1/2)$, symbols 1 and 2.

Weaving In More Entropy

Let $w \in \mathcal{B}(1/2, 1/2)$, symbols 1 and 2.

Suppose $u_s = v_s, u_t = v_t, u_r = v_r, \ldots$.

$$u = \ldots u_s \ldots u_{t-1} u_t \ldots u_{r-1} u_r \ldots$$

$$v = \ldots v_s \ldots\ v_{t-1} v_t \ldots\ v_{r-1} v_r \ldots$$

$$w = \ldots 1? \ldots\ldots ??2? \ldots\ldots ??1? \ldots$$

Weaving In More Entropy

Let $w \in \mathcal{B}(1/2, 1/2)$, symbols 1 and 2.

Suppose $u_s = v_s, u_t = v_t, u_r = v_r, \ldots$.

$$u = \ldots u_s \ldots u_{t-1} u_t \ldots u_{r-1} u_r \ldots$$

$$v = \ldots v_s \ldots \; v_{t-1} v_t \ldots \; v_{r-1} v_r \ldots$$

$$w = \ldots 1? \ldots \ldots ??2? \ldots \ldots ??1? \ldots$$

$\pi_3 : X \times X \times \mathcal{B}(1/2, 1/2) \to X$,

Weaving In More Entropy

Let $w \in \mathcal{B}(1/2, 1/2)$, symbols 1 and 2.

Suppose $u_s = v_s, u_t = v_t, u_r = v_r, \ldots$.

$$u = \ldots u_s \ldots u_{t-1} u_t \ldots u_{r-1} u_r \ldots$$
$$v = \ldots v_s \ldots \; v_{t-1} v_t \ldots \; v_{r-1} v_r \ldots$$
$$w = \ldots 1? \ldots \ldots ??2? \ldots \ldots ??1? \ldots$$

$\pi_3 : X \times X \times \mathcal{B}(1/2, 1/2) \to X$,

$$\pi_3(u, v, w) = \ldots (u_s u_{s+1} \ldots u_{t-1})(v_t v_{t+1} \ldots v_{r-1})(u_r u_{r+1} \ldots) \ldots$$

Why Does It Go Up?

The switching increases entropy.

Why Does It Go Up?

The switching increases entropy.

The argument uses

- strict concavity of $-t \log t$
- lots of calculations with conditional expectations .

V. Recognizing the hidden Markov measures

1. Identify images of Markov measures (metrically sofic, hidden Markov). Heller, Robertson, Furstenberg, Binkowska-Kaminski.

V. Recognizing the hidden Markov measures

1. Identify images of Markov measures (metrically sofic, hidden Markov). Heller, Robertson, Furstenberg, Binkowska-Kaminski.

$\mathcal{A}$=free associative algebra over $\mathbb{R}$ generated by the alphabet A of Y

V. Recognizing the hidden Markov measures

1. Identify images of Markov measures (metrically sofic, hidden Markov). Heller, Robertson, Furstenberg, Binkowska-Kaminski.

$\mathcal{A}$=free associative algebra over $\mathbb{R}$ generated by the alphabet A of Y

$\phi(\epsilon) = 1, \phi(y_1 \dots y_n) = \nu[y_1 \dots y_n]$ extends to linear functional on $\mathcal{A}$.

Metrically Sofic vs. Finitary

$\mathcal{N}$=largest left ideal in kernel$(\phi) = \{a \in \mathcal{A} : \phi(wa) = 0$ for all $w \in A^*\}$

Metrically Sofic vs. Finitary

$\mathcal{N}$=largest left ideal in kernel$(\phi) = \{a \in \mathcal{A} : \phi(wa) = 0$ for all $w \in A^*\}$

$(A^{\mathbb{Z}}, \nu)$ is *finitary* iff the vector space $\mathcal{A}/\mathcal{N}$ is finite dimensional.

Metrically Sofic vs. Finitary

$\mathcal{N}$=largest left ideal in kernel$(\phi) = \{a \in \mathcal{A} : \phi(wa) = 0$ for all $w \in A^*\}$

$(A^{\mathbb{Z}}, \nu)$ is *finitary* iff the vector space $\mathcal{A}/\mathcal{N}$ is finite dimensional.

Heller: Metrically sofic implies finitary, but *not conversely*.

Metrically Sofic vs. Finitary

$\mathcal{N}$=largest left ideal in kernel$(\phi) = \{a \in \mathcal{A} : \phi(wa) = 0$ for all $w \in A^*\}$

$(A^{\mathbb{Z}}, \nu)$ is *finitary* iff the vector space $\mathcal{A}/\mathcal{N}$ is finite dimensional.

Heller: Metrically sofic implies finitary, but *not conversely*.

Robertson: Mixing and finitary implies K.

Metrically Sofic vs. Finitary

$\mathcal{N}$=largest left ideal in kernel$(\phi) = \{a \in \mathcal{A} : \phi(wa) = 0$ for all $w \in A^*\}$

$(A^{\mathbb{Z}}, \nu)$ is *finitary* iff the vector space $\mathcal{A}/\mathcal{N}$ is finite dimensional.

Heller: Metrically sofic implies finitary, but *not conversely*.

Robertson: Mixing and finitary implies K.

Furstenberg: Characterization of metrically sofic in terms of finite-dimensionality of a related algebra by a different left ideal.

2. Formal languages characterization

Kleene, Schützenberger, Hansel-Perrin, etc.

2. Formal languages characterization

Kleene, Schützenberger, Hansel-Perrin, etc.

Shift-invariant μ on $A^{\mathbb{N}}$ is a function $A^* \to \mathbb{R}_+$ or a *formal series*

$$\sum_{w \in A^*} s(w)\, w.$$

2. Formal languages characterization

Kleene, Schützenberger, Hansel-Perrin, etc.
Shift-invariant μ on $A^{\mathbb{N}}$ is a function $A^* \to \mathbb{R}_+$ or a *formal series*

$$\sum_{w \in A^*} s(w)\, w.$$

Language $\sim 0, 1$-valued formal series

2. Formal languages characterization

Kleene, Schützenberger, Hansel-Perrin, etc.
Shift-invariant μ on $A^{\mathbb{N}}$ is a function $A^* \to \mathbb{R}_+$ or a *formal series*

$$\sum_{w \in A^*} s(w)\, w.$$

Language $\sim 0, 1$-valued formal series
$\mathcal{F}(A)$ = set of formal series is a *semiring*

2. Formal languages characterization

Kleene, Schützenberger, Hansel-Perrin, etc.
Shift-invariant μ on $A^{\mathbb{N}}$ is a function $A^* \to \mathbb{R}_+$ or a *formal series*

$$\sum_{w \in A^*} s(w)\, w.$$

Language $\sim 0, 1$-valued formal series
$\mathcal{F}(A)$ = set of formal series is a *semiring*

$$(s_1 s_2)(w) = \sum_{u,v \in A^*, uv=w} s_1(u)\, s_2(v).$$

Module structure

$\mathcal{F}(A)$ is an $\mathbb{R}_+$-module

Module structure

$\mathcal{F}(A)$ is an $\mathbb{R}_+$-module

and A^* acts on $\mathcal{F}(A)$:

Module structure

$\mathcal{F}(A)$ is an $\mathbb{R}_+$-module

and A^* acts on $\mathcal{F}(A)$:

$(w, F) \to w^{-1}F = \sum_{v \in A^*} F(wv)\, v$

Module structure

$\mathcal{F}(A)$ is an $\mathbb{R}_+$-module

and A^* acts on $\mathcal{F}(A)$:

$(w, F) \longrightarrow w^{-1}F = \sum_{v \in A^*} F(wv)\, v$

$(w^{-1}F)(v) = F(wv)$ for all $v \in A^*$.

Module structure

$\mathcal{F}(A)$ is an $\mathbb{R}_+$-module

and A^* acts on $\mathcal{F}(A)$:

$(w, F) \to w^{-1}F = \sum_{v \in A^*} F(wv)\, v$

$(w^{-1}F)(v) = F(wv)$ for all $v \in A^*$.

A submodule $M \subset \mathcal{F}(A)$ is *stable* if $w^{-1}M \subset M$ for all $w \in A^*$.

Equivalent conditions for metrically sofic

For any $F \in \mathcal{F}(A)$ that corresponds to a shift-invariant probability measure μ on $A^{\mathbb{N}}$ the following are equivalent:

Equivalent conditions for metrically sofic

For any $F \in \mathcal{F}(A)$ that corresponds to a shift-invariant probability measure μ on $A^{\mathbb{N}}$ the following are equivalent:

1. F is *linearly representable*: there are $n \geq 1, x \in \mathbb{R}_+^n, y \in (\mathbb{R}_+^n)^*$, and a morphism $\phi : A^* \to \mathbb{R}_+^{n \times n}$ such that

$$F(w) = x\,\phi(w)\,y \quad \text{for all } w \in A^*.$$

Equivalent conditions for metrically sofic

For any $F \in \mathcal{F}(A)$ that corresponds to a shift-invariant probability measure μ on $A^{\mathbb{N}}$ the following are equivalent:

1. F is *linearly representable*: there are $n \geq 1, x \in \mathbb{R}_+^n, y \in (\mathbb{R}_+^n)^*$, and a morphism $\phi : A^* \to \mathbb{R}_+^{n \times n}$ such that

$$F(w) = x\,\phi(w)\,y \quad \text{for all } w \in A^*.$$

2. F is a member of a finitely-generated stable submodule of $\mathcal{F}(A)$.

Equivalent conditions for metrically sofic

For any $F \in \mathcal{F}(A)$ that corresponds to a shift-invariant probability measure μ on $A^{\mathbb{N}}$ the following are equivalent:

1. F is *linearly representable*: there are $n \geq 1, x \in \mathbb{R}_+^n, y \in (\mathbb{R}_+^n)^*$, and a morphism $\phi : A^* \to \mathbb{R}_+^{n \times n}$ such that

$$F(w) = x\,\phi(w)\,y \quad \text{for all } w \in A^*.$$

2. F is a member of a finitely-generated stable submodule of $\mathcal{F}(A)$.

3. F is *rational*–can be obtained by starting with a finite set of polynomials (finitely-supported series) and applying finitely many *rational operations*: sum, product, multiplication by $\mathbb{R}_+$, and $f \to f^* = \sum_{n=0}^{\infty} f^n$ for $f(\epsilon) = 0$.

Equivalent conditions for metrically sofic

For any $F \in \mathcal{F}(A)$ that corresponds to a shift-invariant probability measure μ on $A^{\mathbb{N}}$ the following are equivalent:

1. F is *linearly representable*: there are $n \geq 1, x \in \mathbb{R}_+^n, y \in (\mathbb{R}_+^n)^*$, and a morphism $\phi : A^* \to \mathbb{R}_+^{n \times n}$ such that

$$F(w) = x\,\phi(w)\,y \quad \text{for all } w \in A^*.$$

2. F is a member of a finitely-generated stable submodule of $\mathcal{F}(A)$.

3. F is *rational*–can be obtained by starting with a finite set of polynomials (finitely-supported series) and applying finitely many *rational operations*: sum, product, multiplication by $\mathbb{R}_+$, and $f \to f^* = \sum_{n=0}^{\infty} f^n$ for $f(\epsilon) = 0$.

4. μ is the image under a 1-block map of a 1-step Markov measure.

VI. Measures of Maximal Hausdorff Dimension

Find measures of maximal Hausdorff dimension for expanding (not necessarily conformal) maps on manifolds restricted to compact invariant sets.

VI. Measures of Maximal Hausdorff Dimension

Find measures of maximal Hausdorff dimension for expanding (not necessarily conformal) maps on manifolds restricted to compact invariant sets.

E.g., Sierpinski carpet type sets studied by McMullen, Bedford, Kenyon, Gatzouras, Peres.

VI. Measures of Maximal Hausdorff Dimension

Find measures of maximal Hausdorff dimension for expanding (not necessarily conformal) maps on manifolds restricted to compact invariant sets.

E.g., Sierpinski carpet type sets studied by McMullen, Bedford, Kenyon, Gatzouras, Peres.

Theorem (Shin)**.** *If there is a saturated compensation function $G \circ \pi$ with $G \in \mathcal{C}(Y)$, then the measures which maximize the weighted entropy functional*

$$\phi_\alpha(\mu) = \frac{1}{\alpha + 1}[h(\mu) + \alpha h(\pi\mu)]$$

are the equilibrium states for $\frac{\alpha}{\alpha+1}G \circ \pi$.

VI. Measures of Maximal Hausdorff Dimension

Find measures of maximal Hausdorff dimension for expanding (not necessarily conformal) maps on manifolds restricted to compact invariant sets.

E.g., Sierpinski carpet type sets studied by McMullen, Bedford, Kenyon, Gatzouras, Peres.

Theorem (Shin)**.** *If there is a saturated compensation function* $G \circ \pi$ *with* $G \in \mathcal{C}(Y)$*, then the measures which maximize the weighted entropy functional*

$$\phi_\alpha(\mu) = \frac{1}{\alpha + 1}[h(\mu) + \alpha h(\pi\mu)]$$

are the equilibrium states for $\frac{\alpha}{\alpha+1} G \circ \pi$*.*

So in some cases they are unique, Bernoulli, etc.

Carpets

Ledrappier-Young:

$$\mathrm{HD}(\mu) = \frac{h_\mu(f)}{\lambda_\mu^1(f)} + \left[\frac{1}{\lambda_\mu^2(f)} - \frac{1}{\lambda_\mu^1(f)}\right] h_{\pi\mu}(f_*)$$

$f_* =$action of f on leaves of strong unstable foliation

Carpets

Ledrappier-Young:

$$\mathsf{HD}(\mu) = \frac{h_\mu(f)}{\lambda^1_\mu(f)} + \left[\frac{1}{\lambda^2_\mu(f)} - \frac{1}{\lambda^1_\mu(f)}\right] h_{\pi\mu}(f_*)$$

$f_* =$action of f on leaves of strong unstable foliation

E.g., $T(x,y) = (3x, 2y) \mod 1$ on 2-torus.

Carpets

Ledrappier-Young:

$$\mathrm{HD}(\mu) = \frac{h_\mu(f)}{\lambda^1_\mu(f)} + \left[\frac{1}{\lambda^2_\mu(f)} - \frac{1}{\lambda^1_\mu(f)}\right] h_{\pi\mu}(f_*)$$

$f_* =$action of f on leaves of strong unstable foliation

E.g., $T(x,y) = (3x, 2y) \mod 1$ on 2-torus.

Have $\Sigma_3 \times \Sigma_2$, but restrict to subset of 6-element alphabet

Carpets

Ledrappier-Young:

$$\text{HD}(\mu) = \frac{h_\mu(f)}{\lambda_\mu^1(f)} + \left[\frac{1}{\lambda_\mu^2(f)} - \frac{1}{\lambda_\mu^1(f)}\right] h_{\pi\mu}(f_*)$$

$f_* =$action of f on leaves of strong unstable foliation

E.g., $T(x, y) = (3x, 2y) \mod 1$ on 2-torus.

Have $\Sigma_3 \times \Sigma_2$, but restrict to subset of 6-element alphabet

Or to SFT—results by Yuki Yayama on existence, uniqueness, properties of measures of maximal Hausdorff dimension (Gibbs, Bernoulli natural extension, not Gibbs but strongly mixing, computes HD, etc.)

Uses set-up of Gatzouras-Peres and results of Hofbauer, Markley-Paul (on grid functions), Walters (Bowen class), Coelho-Quas, ...

Carpets

Ledrappier-Young:

$$\text{HD}(\mu) = \frac{h_\mu(f)}{\lambda^1_\mu(f)} + \left[\frac{1}{\lambda^2_\mu(f)} - \frac{1}{\lambda^1_\mu(f)}\right] h_{\pi\mu}(f_*)$$

$f_* =$action of f on leaves of strong unstable foliation

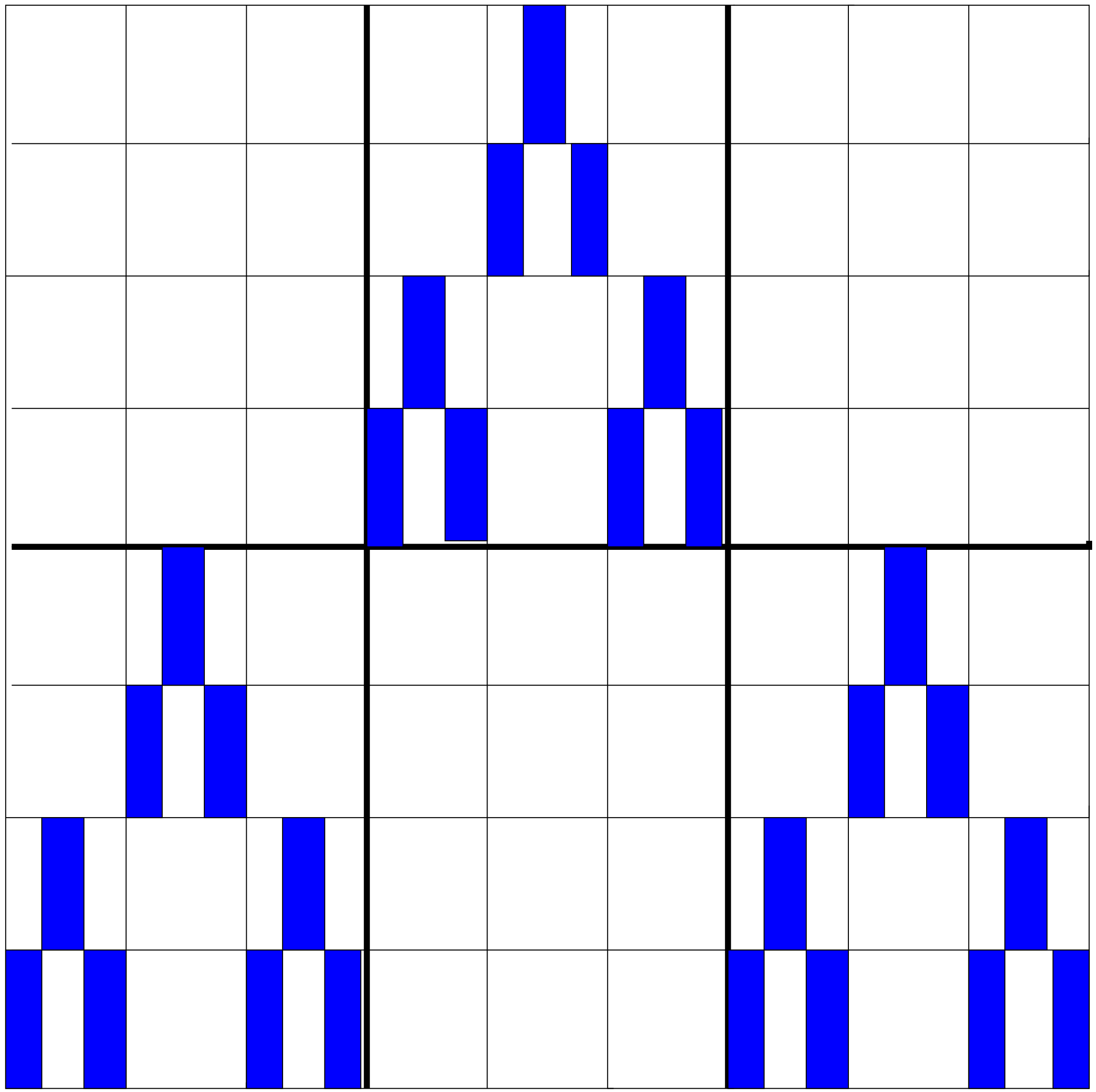
E.g., $T(x,y) = (3x, 2y) \mod 1$ on 2-torus.

Have $\Sigma_3 \times \Sigma_2$, but restrict to subset of 6-element alphabet

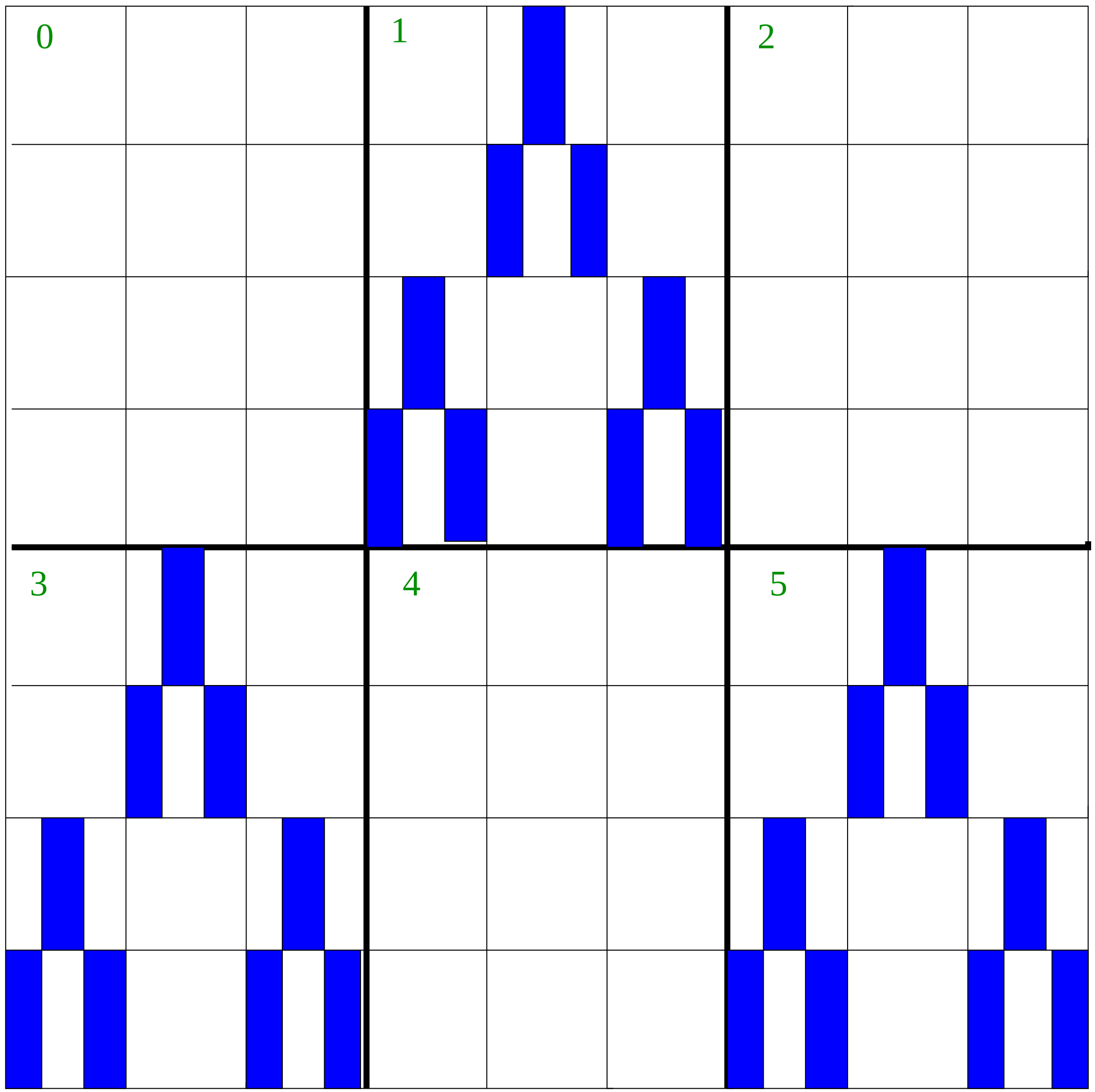
Or to SFT—results by Yuki Yayama on existence, uniqueness, properties of measures of maximal Hausdorff dimension (Gibbs, Bernoulli natural extension, not Gibbs but strongly mixing, computes HD, etc.)

Uses set-up of Gatzouras-Peres and results of Hofbauer, Markley-Paul (on grid functions), Walters (Bowen class), Coelho-Quas, ...

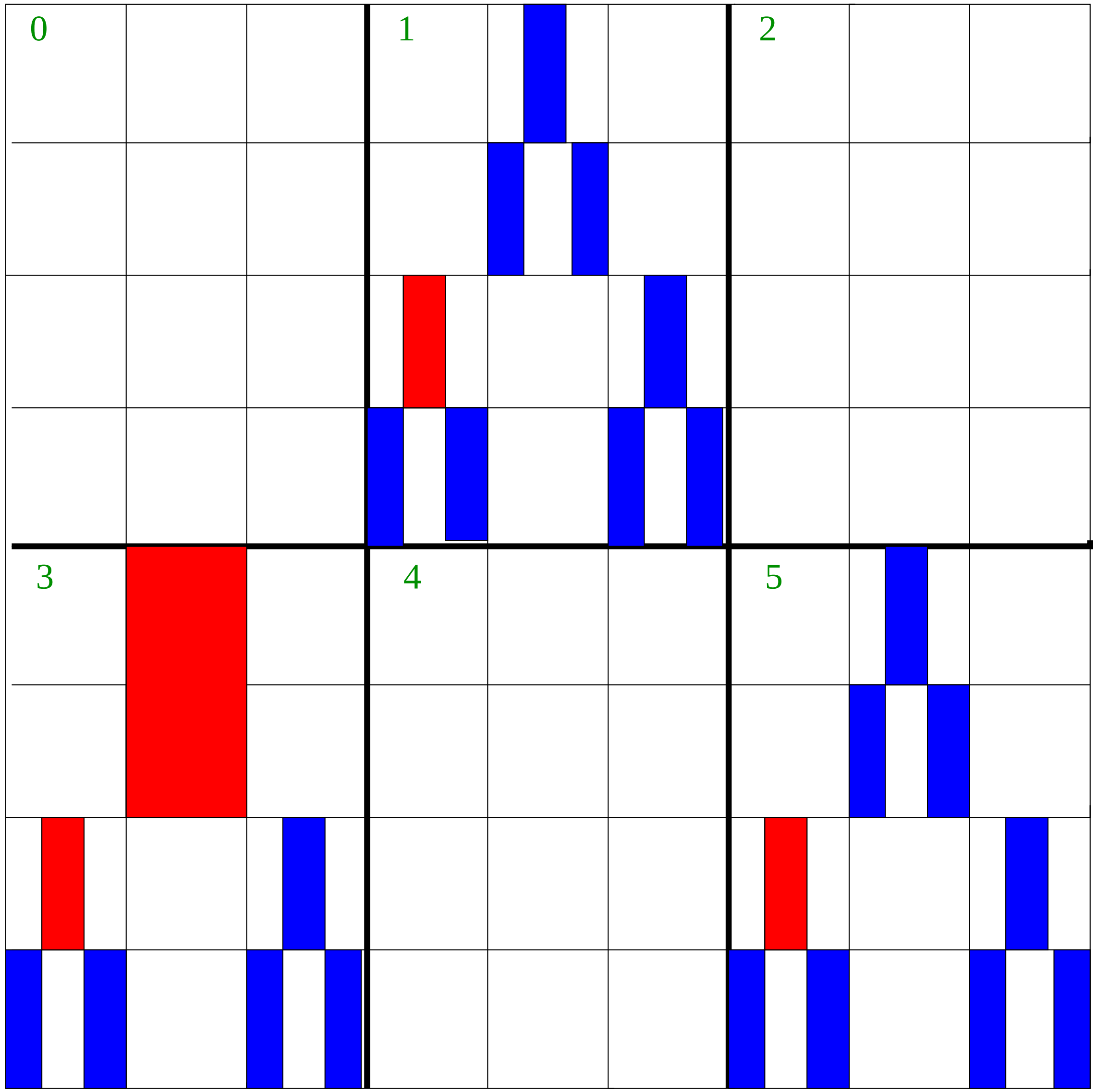
Nonconformal Carpet



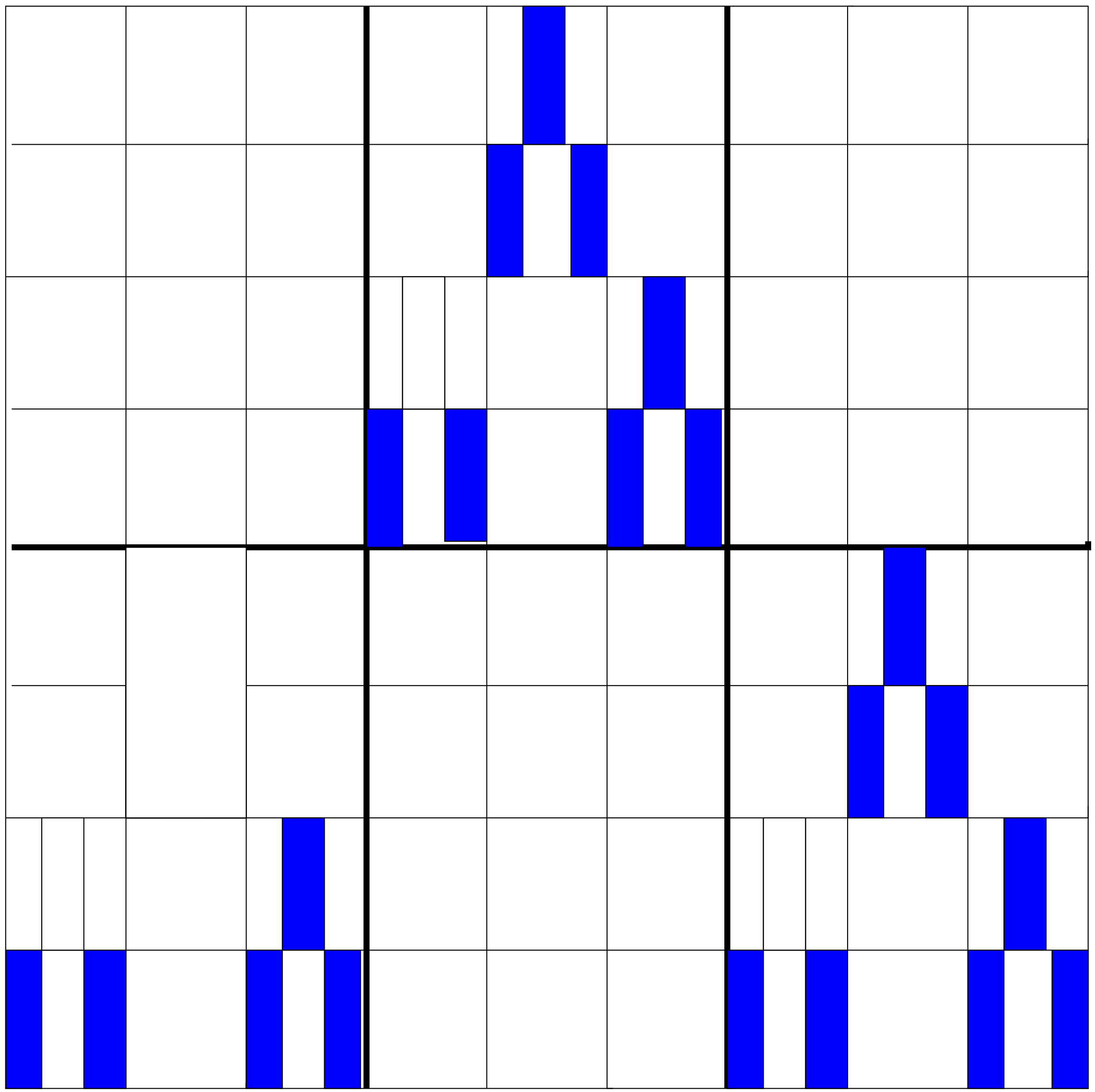
Nonconformal Carpet Coded



Disallow some transitions 31

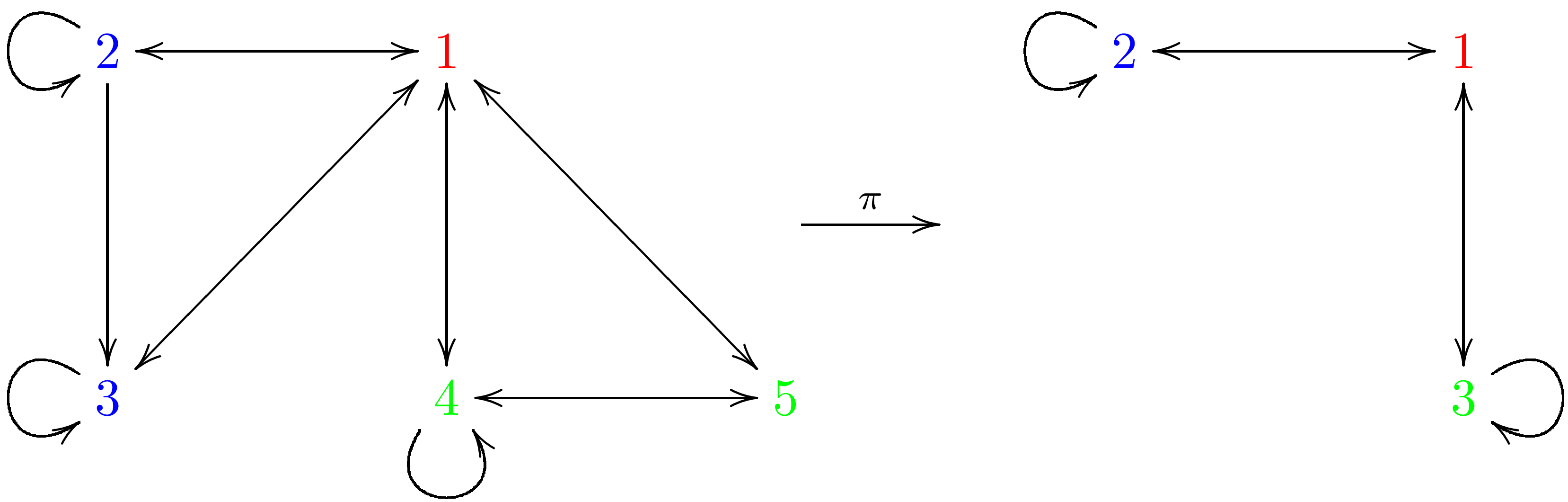


More worn carpet



A candidate for nonuniqueness

$\pi(1) = 1, \pi(2) = \pi(3) = 2, \pi(4) = \pi(5) = 3.$



VII. Some Questions

VII. Some Questions

1. Decide when π takes Markov to Markov, Gibbs to Gibbs.

VII. Some Questions

1. Decide when π takes Markov to Markov, Gibbs to Gibbs.

Chazottes-Ugalde in certain cases find that the image is Gibbs and identify the potential function.

VII. Some Questions

1. Decide when π takes Markov to Markov, Gibbs to Gibbs.

Chazottes-Ugalde in certain cases find that the image is Gibbs and identify the potential function.

2. Construction of relatively maximal measures. Our proof uses *relative g-functions* and shows that the measures are *relatively Markov*:

$$\alpha \perp_{\sigma^{-1}\alpha \vee \pi^{-1}\mathcal{B}(Y)} \alpha_2^\infty, \qquad H_\mu(\alpha | \alpha_1^\infty \vee \pi^{-1}\mathcal{B}_Y) = H_\mu(\alpha | \sigma^{-1}\alpha \vee \pi^{-1}\mathcal{B}_Y).$$

VII. Some Questions

1. Decide when π takes Markov to Markov, Gibbs to Gibbs.

Chazottes-Ugalde in certain cases find that the image is Gibbs and identify the potential function.

2. Construction of relatively maximal measures. Our proof uses *relative g-functions* and shows that the measures are *relatively Markov*:

$$\alpha \perp_{\sigma^{-1}\alpha \vee \pi^{-1}\mathcal{B}(Y)} \alpha_2^{\infty}, \qquad H_\mu(\alpha | \alpha_1^{\infty} \vee \pi^{-1}\mathcal{B}_Y) = H_\mu(\alpha | \sigma^{-1}\alpha \vee \pi^{-1}\mathcal{B}_Y).$$

Construct them as weak* limits of well-distributed measures on periodic orbits?