Random Bits of Noise

Michael W. Mislove

Tulane University
New Orleans, LA

MFPS 25
Oxford, UK

Work sponsored by ONR

Channels

- Mechanism for communication
  - *Overt:* used as intended to exchange information
  - *Covert:* not intended for communication

Channels

- Mechanism for communication
  - *Overt:* used as intended to exchange information
  - *Covert:* not intended for communication
  - *Noise:* converts one symbol into another.

Channels

- Mechanism for communication
  - *Overt:* used as intended to exchange information
  - *Covert:* not intended for communication
  - *Noise:* converts one symbol into another.
- For inputs $X = \{x_0, \ldots, x_{m-1}\}$ and outputs $Y = \{y_0, \ldots, y_{n-1}\}$, a *noise matrix* for a channel $C\colon X \to Y$ is an $m \times n$-matrix:

$$M_C = \begin{pmatrix} P(y_0|x_0) & P(y_1|x_0) & \cdots & P(y_{n-1}|x_0) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{m-1}) & P(y_1|x_{m-1}) & \cdots & P(y_{n-1}|x_{m-1}) \end{pmatrix}$$

Note: We assume C is lossless: $\sum_i P(y_j|x_i) = 1$ $(\forall i)$.

Channels

- Mechanism for communication
  - *Overt:* used as intended to exchange information
  - *Covert:* not intended for communication
  - *Noise:* converts one symbol into another.
- For inputs $X = \{x_0, \ldots, x_{m-1}\}$ and outputs $Y = \{y_0, \ldots, y_{n-1}\}$, a *noise matrix* for a channel $C\colon X \to Y$ is an $m \times n$-matrix:

$$M_C = \begin{pmatrix} P(y_0|x_0) & P(y_1|x_0) & \cdots & P(y_{n-1}|x_0) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{m-1}) & P(y_1|x_{m-1}) & \cdots & P(y_{n-1}|x_{m-1}) \end{pmatrix}$$

  Note: We assume C is lossless: $\sum_i P(y_j|x_i) = 1$ $(\forall i)$.
- For $p = (p_0\ p_1\ \ldots\ p_{m-1})$ probability distribution on X we get

$$p \cdot M_C = \left(\sum_i p_i P(y_0|x_i)\ \ \sum_i p_i P(y_1|x_i)\ \ldots\ \sum_i p_i P(y_{n-1}|x_i)\right)$$

  corresponding distribution on Y.

Stochastic Matrices

- Matrix with non-negative, real entries; each row sums to 1
  - $m \times n$-matrix represents a channel with m inputs and n outputs.

Stochastic Matrices

- Matrix with non-negative, real entries; each row sums to 1
  - $m \times n$-matrix represents a channel with m inputs and n outputs.
- For a binary alphabet, only need first column of noise matrix

$$\begin{array}{c|cc|} & 0 & 1 \\ \hline 0 & P(0|0) & P(1|0) \\ 1 & P(0|1) & P(1|1) \end{array} \quad \leftrightarrow \quad \begin{array}{|c|} 0 \\ \hline P(0|0) \\ P(0|1) \end{array} \quad \simeq \quad [0,1]^2$$

Stochastic Matrices

- Matrix with non-negative, real entries; each row sums to 1
  - $m \times n$-matrix represents a channel with m inputs and n outputs.
- For a binary alphabet, only need first column of noise matrix

$$\begin{array}{c|cc|} & 0 & 1 \\ \hline 0 & P(0|0) & P(1|0) \\ 1 & P(0|1) & P(1|1) \end{array} \quad \leftrightarrow \quad \begin{array}{|c|} 0 \\ \hline P(0|0) \\ P(0|1) \end{array} \quad \simeq \quad [0,1]^2$$

More generally,

$$M_C \;=\; \begin{pmatrix} P(y_0|x_0) & P(y_1|x_1) & \cdots & P(y_{n-1}|x_{m-1}) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{m-1}) & P(y_1|x_{m-1}) & \cdots & P(y_{n-1}|x_{m-1}) \end{pmatrix}$$

$$\leftrightarrow \;\begin{pmatrix} P(y_0|x_0) & P(y_1|x_1) & \cdots & P(y_{n-2}|x_0) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{m-1}) & P(y_1|x_{m-1}) & \cdots & P(y_{n-2}|x_{m-1}) \end{pmatrix} \leftrightarrow ([0,1]^{n-1})^m$$

Information basics

- S sample space

Information basics

- S sample space
- $X: \{x_0, \ldots, x_{m-1}\} \to S$ random variable with probability density p

Information basics

- S sample space
- $X\colon \{x_0, \ldots, x_{m-1}\} \to S$ random variable with probability density p
- *Information* in events:
  - $p(E) = 1 \Rightarrow I(E) = 0$
  - $p(E) \leq p(F) \Rightarrow I(E) \geq I(F)$
  - E, F independent
    $\Rightarrow \; I(E \cap F) = I(E) + I(F)$

Information basics

- S sample space
- $X\colon \{x_0, \ldots, x_{m-1}\} \to S$ random variable with probability density p
- *Information* in events:
  - $p(E) = 1 \Rightarrow I(E) = 0$
  - $p(E) \leq p(F) \Rightarrow I(E) \geq I(F)$
  - E, F independent
    $\Rightarrow\ I(E \cap F) = I(E) + I(F)$

    As a function of $p(E)$,
    $I(p(E)p(F)) =$
    $I(p(E)) + I(p(F))$

Information basics

- S sample space
- $X\colon \{x_0, \ldots, x_{m-1}\} \to S$ random variable with probability density p
- *Information* in events:
  - $p(E) = 1 \Rightarrow I(E) = 0$
  - $p(E) \leq p(F) \Rightarrow I(E) \geq I(F)$
  - E, F independent
    $\Rightarrow \; I(E \cap F) = I(E) + I(F)$

    As a function of $p(E)$,
    $I(p(E)p(F)) =$
    $I(p(E)) + I(p(F))$
- $I(X) = -\log_b(p(X))$. Take $b = 2$

Information basics

- S sample space
- $X\colon \{x_0, \ldots, x_{m-1}\} \to S$ random variable with probability density p
- *Information* in events:
  - $p(E) = 1 \Rightarrow I(E) = 0$
  - $p(E) \leq p(F) \Rightarrow I(E) \geq I(F)$
  - E, F independent $\Rightarrow\ I(E \cap F) = I(E) + I(F)$

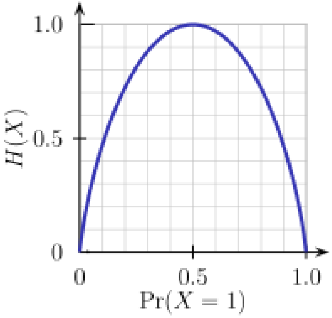
  As a function of $p(E)$, $I(p(E)p(F)) = I(p(E)) + I(p(F))$
- $I(X) = -\log_b(p(X))$. Take $b = 2$

Entropy function

$H(X) = -\sum_i p(s_i) \log_2 p(s_i)$

Average information in X

Binary case:
$H(X) = -\sum_{i=0}^{1} p(s_i) \log_2 p(s_i)$



Joint and Conditional Entropy

Given: two random variables X and Y on the same space.

Joint and Conditional Entropy

Given: two random variables X and Y on the same space.

Joint Entropy

- $H(X,Y) = -\sum_i \sum_j p(s_i,s_j) \log p(s_i,s_j) = -E \log p(X,Y)$

  $p(X,Y)$ – joint probability distribution on $X \times Y$

  $= \frac{p(X|Y)}{p(X)p(Y)}$

Joint and Conditional Entropy

Given: two random variables X and Y on the same space.

Joint Entropy

- $H(X,Y) = -\sum_i \sum_j p(s_i,s_j)\log p(s_i,s_j) = -E\log p(X,Y)$

  $p(X,Y)$ – joint probability distribution on $X \times Y$

  $= \frac{p(X|Y)}{p(X)p(Y)}$

Conditional Entropy

- $H(X|Y) = -\sum_i \sum_j p(s_i,s_j)\log p(s_i|s_j) = -E\log p(X|Y)$

  $= -\sum_j p(s_j)H(X|Y=s_j)$

Mutual Information

- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
$$= -\sum_j p(s_j)\log_2 p(s_j) + \sum_i p(s_i)H(Y|X = s_i)$$

Mutual Information

- $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$
  $= -\sum_j p(s_j)\log_2 p(s_j) + \sum_i p(s_i) H(Y|X = s_i)$

Channel Capacity

- $Cap = \sup_Y I(X;Y)$
  $= \sup_X H(Y) - H(Y|X)$

From Noise Matrices to Random Variables

- $M_C = \left(\begin{smallmatrix} a & 1-a \\ b & 1-b \end{smallmatrix}\right)$, $\quad p = [x \;\; 1-x]$ distribution on inputs

From Noise Matrices to Random Variables

- $M_C = \left(\begin{smallmatrix} a & 1-a \\ b & 1-b \end{smallmatrix}\right), \quad p = [x \;\; 1-x]$ distribution on inputs
- Two distributions:
  - $X = p = [x \;\; 1-x]$ - input distribution
  - $Y = p \cdot M_C = [xa + (1-x)b, \quad x(1-a) + (1-x)(1-b)]$ - output distribution

From Noise Matrices to Random Variables

- $M_C = \begin{pmatrix} a & 1-a \\ b & 1-b \end{pmatrix}$, $\quad p = [x \;\; 1-x]$ distribution on inputs
- Two distributions:
  - $X = p = [x \;\; 1-x]$ - input distribution
  - $Y = p \cdot M_C = [xa + (1-x)b, \quad x(1-a) + (1-x)(1-b)]$ - output distribution

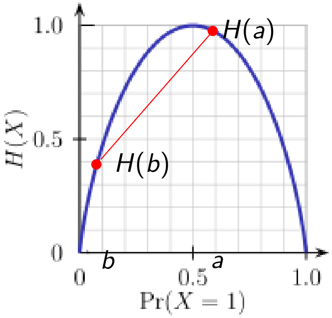
Capacity of Channel with Noise Matrix M_C

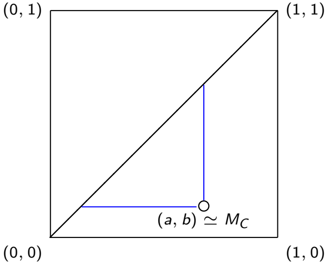
- $$Cap(M) = \sup_p \; H(p \cdot M) - H(p \cdot M \mid p)$$
  $$= \sup_x \; H(xa + (1-x)b) - (xH(a) + (1-x)H(b))$$
  $$(p = [x \quad 1-x]; \;\; H(z) \equiv H([z \;\; 1-z])$$

Capacity of Channel with Noise Matrix M_C

- $Cap(M_C) = \sup_x \; H(xa + (1-x)b) - (xH(a) + (1-x)H(b))$

$$(p = [x \;\; 1-x], \; 0 \leq x \leq 1)$$

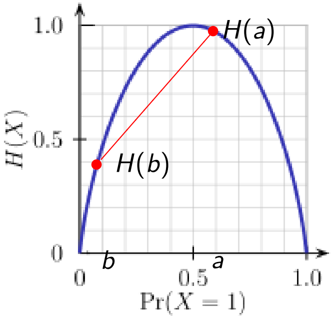


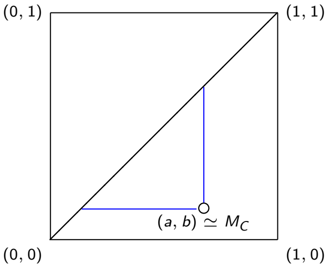


Capacity of Channel with Noise Matrix M_C

- $Cap(M_C) = \sup_x \; H(xa + (1-x)b) - (xH(a) + (1-x)H(b))$

$$(p = [x \;\; 1-x], \; 0 \le x \le 1)$$





Mean Value Theorem $\Longrightarrow$

$$Cap(M_C) = H(x_0 a + (1-x_0)b) - (x_0 H(a) + (1-x_0)H(b))$$

where $H'(x_0) = \frac{H(a)-H(b)}{a-b}$.

$m \times n$-matrices

For

$$M_C \leftrightarrow \begin{pmatrix} P(y_0|x_0) & P(y_1|x_1) & \cdots & P(y_{n-2}|x_0) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{m-1}) & P(y_1|x_{m-1}) & \cdots & P(y_{n-2}|x_{m-1}) \end{pmatrix} \hookrightarrow [0,1]^{m(n-1)}$$

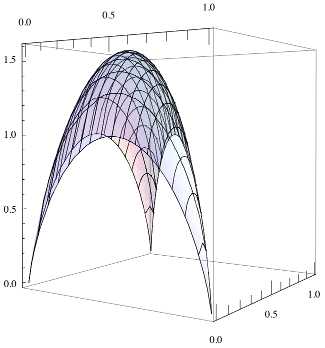
$Cap(M_C) = H(p_0 \cdot M_C) - p_0 \cdot \langle H(M_C) \rangle$, where $\nabla H(p_0 \cdot M_C) = \overrightarrow{n}_P$

$m \times n$-matrices

For

$$M_C \leftrightarrow \begin{pmatrix} P(y_0|x_0) & P(y_1|x_1) & \cdots & P(y_{n-2}|x_0) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{m-1}) & P(y_1|x_{m-1}) & \cdots & P(y_{n-2}|x_{m-1}) \end{pmatrix} \hookrightarrow [0,1]^{m(n-1)}$$

$Cap(M_C) = H(p_0 \cdot M_C) - p_0 \cdot \langle H(M_C) \rangle$, where $\nabla H(p_0 \cdot M_C) = \vec{n}_P$

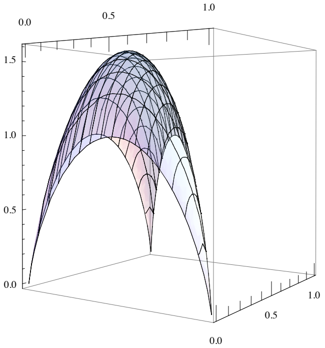


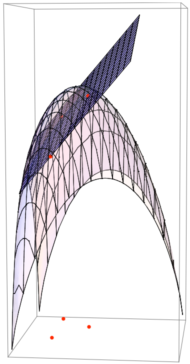
$m \times n$-matrices

For

$$M_C \leftrightarrow \begin{pmatrix} P(y_0|x_0) & P(y_1|x_1) & \cdots & P(y_{n-2}|x_0) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{m-1}) & P(y_1|x_{m-1}) & \cdots & P(y_{n-2}|x_{m-1}) \end{pmatrix} \hookrightarrow [0,1]^{m(n-1)}$$

$Cap(M_C) = H(p_0 \cdot M_C) - p_0 \cdot \langle H(M_C) \rangle$, where $\nabla H(p_0 \cdot M_C) = \overrightarrow{n}_P$



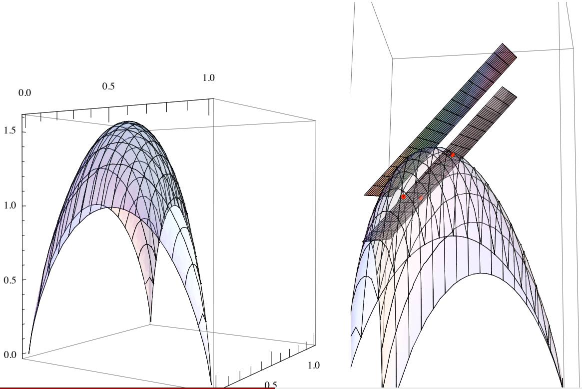


$m \times n$-matrices

For

$$M_C \leftrightarrow \begin{pmatrix} P(y_0|x_0) & P(y_1|x_1) & \cdots & P(y_{n-2}|x_0) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{m-1}) & P(y_1|x_{m-1}) & \cdots & P(y_{n-2}|x_{m-1}) \end{pmatrix} \hookrightarrow [0,1]^{m(n-1)}$$

$$Cap(M_C) = H(p_0 \cdot M_C) - p_0 \cdot \langle H(M_C) \rangle, \text{ where } \nabla H(p_0 \cdot M_C) = \vec{n}_P$$



$m \times n$-matrices

The Moral of the Story

- $\mathbb{ST}_{m,n}$ – $m \times n$ stochastic matrices

  $\mathcal{C}([0,1]^{m(n-1)})$ – compact, convex subsets of $[0,1]^{m(n-1)}$

$m \times n$-matrices

The Moral of the Story

- $\mathbb{ST}_{m,n}$ – $m \times n$ stochastic matrices

  $\mathcal{C}([0,1]^{m(n-1)})$ – compact, convex subsets of $[0,1]^{m(n-1)}$
  - $(\mathcal{C}([0,1]^{m(n-1)}), \supseteq)$ is a domain: $A \ll A' \Leftrightarrow A' \subseteq A^\circ$

$m \times n$-matrices

The Moral of the Story

- $\mathbb{ST}_{m,n}$ – $m \times n$ stochastic matrices

  $\mathcal{C}([0,1]^{m(n-1)})$ – compact, convex subsets of $[0,1]^{m(n-1)}$
  - $(\mathcal{C}([0,1]^{m(n-1)}), \supseteq)$ is a domain: $A \ll A' \Leftrightarrow A' \subseteq A^\circ$
  - $M \stackrel{\phi}{\mapsto} \langle M(1), M(2), \ldots, M(m)\rangle \colon \mathbb{ST}_{m,n} \to \mathcal{C}([0,1]^{m(n-1)})$ is continuous.

$m \times n$-matrices

The Moral of the Story

- $\mathbb{ST}_{m,n}$ – $m \times n$ stochastic matrices

  $\mathcal{C}([0,1]^{m(n-1)})$ – compact, convex subsets of $[0,1]^{m(n-1)}$
  - $(\mathcal{C}([0,1]^{m(n-1)}), \supseteq)$ is a domain: $A \ll A' \Leftrightarrow A' \subseteq A^\circ$
  - $M \overset{\phi}{\mapsto} \langle M(1), M(2), \ldots, M(m)\rangle : \mathbb{ST}_{m,n} \to \mathcal{C}([0,1]^{m(n-1)})$ is continuous.
  - $\phi(M_C) = \phi(M_{C'}) \ \Rightarrow \ Cap(M_C) = Cap(M_{C'})$

    Induces $Cap' : (\mathcal{C}([0,1]^{m(n-1)}), \supseteq) \to \mathbb{R}^{\mathrm{op}}_{\geq 0}$.

$m \times n$-matrices

The Moral of the Story

- $\mathbb{ST}_{m,n}$ – $m \times n$ stochastic matrices

  $\mathcal{C}([0,1]^{m(n-1)})$ – compact, convex subsets of $[0,1]^{m(n-1)}$
  - $(\mathcal{C}([0,1]^{m(n-1)}), \supseteq)$ is a domain: $A \ll A' \Leftrightarrow A' \subseteq A^\circ$
  - $M \stackrel{\phi}{\mapsto} \langle M(1), M(2), \ldots, M(m) \rangle \colon \mathbb{ST}_{m,n} \to \mathcal{C}([0,1]^{m(n-1)})$ is continuous.
  - $\phi(M_C) = \phi(M_{C'}) \;\Rightarrow\; Cap(M_C) = Cap(M_{C'})$

    Induces $Cap' \colon (\mathcal{C}([0,1]^{m(n-1)}), \supseteq) \to \mathbb{R}^{\mathrm{op}}_{\geq 0}$.
  - [Chatzikokolakis & Martin] $Cap' \colon (\mathcal{C}([0,1]^{m(n-1)}), \supseteq) \to \mathbb{R}^{\mathrm{op}}_{\geq 0}$ is measurement. It also is monotone and Lawson continuous.

$m \times n$-matrices

The Moral of the Story

- $\mathbb{ST}_{m,n}$ – $m \times n$ stochastic matrices

  $\mathcal{C}([0,1]^{m(n-1)})$ – compact, convex subsets of $[0,1]^{m(n-1)}$
  - $(\mathcal{C}([0,1]^{m(n-1)}), \supseteq)$ is a domain: $A \ll A' \Leftrightarrow A' \subseteq A^\circ$
  - $M \overset{\phi}{\mapsto} \langle M(1), M(2), \ldots, M(m)\rangle \colon \mathbb{ST}_{m,n} \to \mathcal{C}([0,1]^{m(n-1)})$ is continuous.
  - $\phi(M_C) = \phi(M_{C'}) \Rightarrow Cap(M_C) = Cap(M_{C'})$

    Induces $Cap' \colon (\mathcal{C}([0,1]^{m(n-1)}), \supseteq) \to \mathbb{R}^{\mathrm{op}}_{\geq 0}$.
  - [Chatzikokolakis & Martin] $Cap' \colon (\mathcal{C}([0,1]^{m(n-1)}), \supseteq) \to \mathbb{R}^{\mathrm{op}}_{\geq 0}$ is measurement. It also is monotone and Lawson continuous.
- *Question:* What is the *optimal* map $F \colon (\mathcal{C}([0,1]^{m(n-1)}), \supseteq) \to \mathbb{R}^{\mathrm{op}}_{\geq 0}$?

  $F(A) \leq F(A') \Leftrightarrow Cap'(A) \leq Cap'(A')$

m-dimensional stochastic matrices

- A monoid S is *affine* if S is a subset of a vector space and $x \mapsto sx, xs\colon S \to S$ are affine.
- *Example:* $\mathbb{ST}_m$ – $m \times m$ stochastic matrices.

m-dimensional stochastic matrices

- A monoid S is *affine* if S is a subset of a vector space and $x \mapsto sx, xs \colon S \to S$ are affine.
- *Example:* $\mathbb{ST}_m$ – $m \times m$ stochastic matrices.
- S compact implies S has a smallest two-sided, closed ideal, $M(S)$.

  Also affine if S affine.

m-dimensional stochastic matrices

- A monoid S is *affine* if S is a subset of a vector space and $x \mapsto sx, xs \colon S \to S$ are affine.
- *Example:* $\mathbb{ST}_m$ – $m \times m$ stochastic matrices.
- S compact implies S has a smallest two-sided, closed ideal, $M(S)$.

  Also affine if S affine.
- $$M(\mathbb{ST}_m) = \left\{ \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ \vdots & \vdots & & \vdots \\ x_{11} & x_{12} & \cdots & x_{1m} \end{pmatrix} \mid x_{1i} \in [0,1], \sum_i x_{1i} = 1 \right\}$$

  consists of *right zeroes:* $MN = N$ $(\forall N \in M(\mathbb{ST}_m))$.

m-dimensional stochastic matrices

- A monoid S is *affine* if S is a subset of a vector space and $x \mapsto sx, xs \colon S \to S$ are affine.
- *Example:* $\mathbb{ST}_m$ – $m \times m$ stochastic matrices.
- S compact implies S has a smallest two-sided, closed ideal, $M(S)$.

  Also affine if S affine.
-
$$M(\mathbb{ST}_m) = \left\{ \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ \vdots & \vdots & & \vdots \\ x_{11} & x_{12} & \cdots & x_{1m} \end{pmatrix} \mid x_{1i} \in [0,1], \sum_i x_{1i} = 1 \right\}$$

  consists of *right zeroes:* $MN = N$ $(\forall N \in M(\mathbb{ST}_m))$.
- $\phi \colon \mathbb{ST}_m \to \mathcal{C}([0,1]^{m(m-1)})$ is a closed congruence, so $\phi(\mathbb{ST}_m)$ is a compact monoid.

m-dimensional stochastic matrices

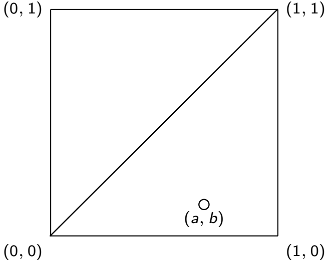
- A monoid S is *affine* if S is a subset of a vector space and $x \mapsto sx, xs \colon S \to S$ are affine.
- *Example:* $\mathbb{ST}_m$ – $m \times m$ stochastic matrices.
- S compact implies S has a smallest two-sided, closed ideal, $M(S)$.

  Also affine if S affine.
-

$$M(\mathbb{ST}_m) = \left\{ \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ \vdots & \vdots & & \vdots \\ x_{11} & x_{12} & \cdots & x_{1m} \end{pmatrix} \mid x_{1i} \in [0,1], \sum_i x_{1i} = 1 \right\}$$

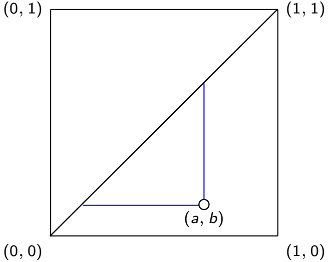
  consists of *right zeroes:* $MN = N$ $(\forall N \in M(\mathbb{ST}_m))$.
- $\phi \colon \mathbb{ST}_m \to \mathcal{C}([0,1]^{m(m-1)})$ is a closed congruence, so $\phi(\mathbb{ST}_m)$ is a compact monoid.
- $Cap(M) = Cap(N)$ if $M(\mathbb{ST}_m)M = M(\mathbb{ST}_m)N \;\Leftrightarrow\; \phi(M) = \phi(N)$.

Binary Noise Matrices - Martin, Moskowitz & Allwein



$\mathcal{N} = \{\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \mid a \geq b\}$ – non-negative noise matrices $(\det \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \geq 0)$

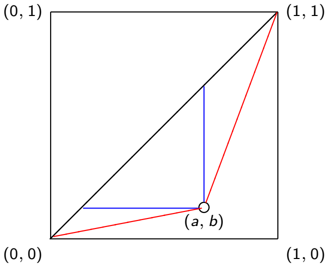
Binary Noise Matrices - Martin, Moskowitz & Allwein



$\mathcal{N} = \{(\begin{smallmatrix} a \\ b \end{smallmatrix}) \mid a \geq b\}$ – non-negative noise matrices $(\det(\begin{smallmatrix} a \\ b \end{smallmatrix}) \geq 0)$

$- - \equiv (\begin{smallmatrix} a \\ b \end{smallmatrix}) \leq (\begin{smallmatrix} c \\ d \end{smallmatrix}) \;\Leftrightarrow\; (\begin{smallmatrix} c \\ d \end{smallmatrix}) \in N(\begin{smallmatrix} a \\ b \end{smallmatrix})$

Binary Noise Matrices - Martin, Moskowitz & Allwein

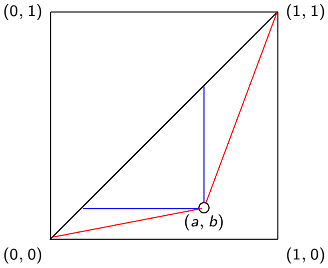


$\mathcal{N} = \{\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \mid a \geq b\}$ – non-negative noise matrices ($\det \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \geq 0$)

– – $\equiv \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \leq \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \Leftrightarrow \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \in N \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$

– – $\equiv \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \leq \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \Leftrightarrow \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \in \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) N$

Binary Noise Matrices - Martin, Moskowitz & Allwein



$\mathcal{N} = \{\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \mid a \geq b\}$ – non-negative noise matrices $(\det \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \geq 0)$

$$- - \equiv \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \leq \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \;\Leftrightarrow\; \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \in N \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right)$$

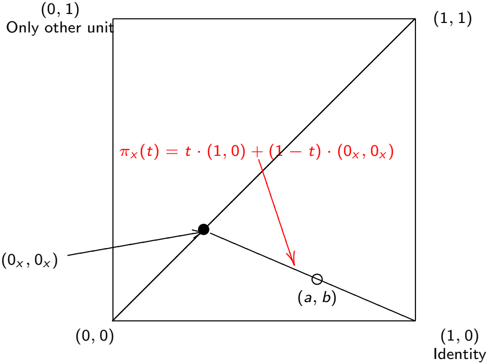
$$- - \equiv \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \leq \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \;\Leftrightarrow\; \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \in \left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) N$$

$$\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} c \\ c \end{smallmatrix}\right) = \begin{pmatrix} a(c-c)+c \\ b(c-c)+c \end{pmatrix} = \left(\begin{smallmatrix} c \\ c \end{smallmatrix}\right)$$

$$(\forall M)(\exists M' \in \mathcal{N}) \; Cap(M) = Cap(M').$$

Binary Noise Matrices - Martin, Moskowitz & Allwein

$$x = \binom{a}{b} \quad 0_x = \frac{b}{1-\det x} \quad \binom{0_x}{0_x} = \lim_n x^n$$



One-parameter Semigroup

A *one-parameter semigroup* is a monoid homomorphism $\phi\colon ([0,1),+) \to S$ into a monoid S.

Generalizing to higher dimensions

One parameter semigroups

S compact, affine, $e = e^2, f = f^2 \in S$, $ef = fe = e$ implies $\{(1-\lambda)f + \lambda e \mid 0 \leq \lambda \leq 1\}$ is a one-parameter semigroup from f to e.

Generalizing to higher dimensions

One parameter semigroups

S compact, affine, $e = e^2, f = f^2 \in S$, $ef = fe = e$ implies $\{(1-\lambda)f + \lambda e \mid 0 \leq \lambda \leq 1\}$ is a one-parameter semigroup from f to e.

One "running through" each $M \in \langle Id_m, M(\mathbb{ST}_m)\rangle$.

Generalizing to higher dimensions

One parameter semigroups

S compact, affine, $e = e^2, f = f^2 \in S$, $ef = fe = e$ implies $\{(1-\lambda)f + \lambda e \mid 0 \leq \lambda \leq 1\}$ is a one-parameter semigroup from f to e.

One "running through" each $M \in \langle Id_m, M(\mathbb{ST}_m)\rangle$.

Many elements of $\mathbb{ST}_m$ lie on a translate of such a semigroup.

Generalizing to higher dimensions

One parameter semigroups

S compact, affine, $e = e^2, f = f^2 \in S$, $ef = fe = e$ implies $\{(1-\lambda)f + \lambda e \mid 0 \leq \lambda \leq 1\}$ is a one-parameter semigroup from f to e.

One "running through" each $M \in \langle Id_m, M(\mathbb{ST}_m)\rangle$.

Many elements of $\mathbb{ST}_m$ lie on a translate of such a semigroup.

Example. But:

$$\left\{\lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (1-\lambda)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \mid 0 \leq \lambda \leq 1\right\}$$

goes through

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

Generating $\mathbb{ST}_m$

The Group of Units

- A monoid S has an identity – associated to it is the *group of units* $G(S)$ of the monoid.
- $G(S)$ - compact group if S compact.
- $G(\mathbb{ST}_m) = S_m$ – symmetric group on m letters.

Generating $\mathbb{ST}_m$

The Group of Units

- A monoid S has an identity – associated to it is the *group of units* $G(S)$ of the monoid.
- $G(S)$ - compact group if S compact.
- $G(\mathbb{ST}_m) = S_m$ – symmetric group on m letters.

First approximate:

$\langle G(\mathbb{ST}_m), M(\mathbb{ST}_m)\rangle$ monoid generated by group of units and minimal ideal.

- Compact affine monoid.
- Every element in $\langle Id_m, M(\mathbb{ST}_m)\rangle$ lies on a unique one-parameter semigroup – all straight lines.
- $\cup_{g\in S_m}\langle g, M(\mathbb{ST}_m)\rangle / \equiv_{Cap}$ identifies each element with some element in $M(\mathbb{ST}_m)$; i.e, with some compact, convex subset of $[0,1]^{m(m-1)}$.

A smaller submonoid:

Doubly Stochastic Matrices

$M \in \mathbb{ST}_m$ is *doubly stochastic* if each column also sums to 1.

$\mathbb{DST}_m$ – doubly stochastic $m \times m$ matrices.

- Again a monoid with same group of units.
- $\langle \mathbb{DST}_m \rangle$ compact monoid.
- *NOT* a group!!

A smaller submonoid:

Doubly Stochastic Matrices

$M \in \mathbb{ST}_m$ is *doubly stochastic* if each column also sums to 1.

$\mathbb{DST}_m$ – doubly stochastic $m \times m$ matrices.

- Again a monoid with same group of units.
- $\langle \mathbb{DST}_m \rangle$ compact monoid.
- *NOT* a group!!

Why not?

Theorem: Any compact affine group is a point.

A smaller submonoid:

Doubly Stochastic Matrices

$M \in \mathbb{ST}_m$ is *doubly stochastic* if each column also sums to 1.

$\mathbb{DST}_m$ – doubly stochastic $m \times m$ matrices.

- Again a monoid with same group of units.
- $\langle \mathbb{DST}_m \rangle$ compact monoid.
- *NOT* a group!!

Why not?

Theorem: Any compact affine group is a point.

Proof: [Hofmis] Suppose G is such a group. Let $x \in G$ have order n.

Then $\sum_{i \leq n} \frac{x^i}{n} = (\sum_{i \leq n} \frac{x^i}{n})^2$ by Fubini.

So, $\sum_{i \leq n} \frac{x^1}{n} = e$; then $x^i = e$ as e is extreme. □

A smaller submonoid:

Doubly Stochastic Matrices

$M \in \mathbb{ST}_m$ is *doubly stochastic* if each column also sums to 1.

$\mathbb{DST}_m$ – doubly stochastic $m \times m$ matrices.

- Again a monoid with same group of units.
- $\langle \mathbb{DST}_m \rangle$ compact monoid.
- *NOT* a group!!

Why not?

Theorem: Any compact affine group is a point.

Theorem: [Brown] Any compact group of non-negative matrices is finite.

A smaller submonoid:

Doubly Stochastic Matrices

$M \in \mathbb{ST}_m$ is *doubly stochastic* if each column also sums to 1.

$\mathbb{DST}_m$ – doubly stochastic $m \times m$ matrices.

- Again a monoid with same group of units.
- $\langle \mathbb{DST}_m \rangle$ compact monoid.
- *NOT* a group!!

Why not?

Theorem: Any compact affine group is a point.

Theorem: [Brown] Any compact group of non-negative matrices is finite.

Structure of $\langle \mathbb{DST}_m \rangle$: Compact monoid; $G(\langle \mathbb{DST}_m \rangle) = S_m$;
$M(\langle \mathbb{DST}_m \rangle) = \left\{ \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \right\}$. Each element of S_3 lies on a line to $\frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$.

Universal affine semigroups:

Generalizing to Stochastic Relations

$M_C \colon X \to Y$ generalizes to $f \colon X \to \mathrm{Prob}(Y)$.

For Y compact, T_2, so is $\mathrm{Prob}(Y)$ in weak*-topology.

Universal affine semigroups:

Generalizing to Stochastic Relations

$M_C \colon X \to Y$ generalizes to $f \colon X \to \mathrm{Prob}(Y)$.

For Y compact, T_2, so is $\mathrm{Prob}(Y)$ in weak*-topology.

Here's an explanation:

$$X \mapsto C(X,\mathbb{R}) \colon \mathsf{Comp} \to \mathsf{Ban}$$

is contravariant: $f \colon X \to Y \;\Rightarrow\; C(f) \colon C(Y,\mathbb{R}) \to C(X,\mathbb{R})$

$$C(X,\mathbb{R}) \mapsto C(C(X,\mathbb{R}),\mathbb{R}) \colon \mathsf{Ban} \to \mathsf{Ban}.$$

But $C^2(X,\mathbb{R}) \simeq \mathsf{Meas}(X,\mathbb{R})$. Extract $\mathrm{Prob}(X)$ by restriction.

Universal affine semigroups:

Generalizing to Stochastic Relations

$M_C : X \to Y$ generalizes to $f : X \to \mathrm{Prob}(Y)$.

For Y compact, T_2, so is $\mathrm{Prob}(Y)$ in weak*-topology.

If $(S, \cdot)$ is a compact semigroup, then

$$\cdot : S \times S \to S \;\Rightarrow\; * : \mathrm{Prob}(S) \times \mathrm{Prob}(S) \to \mathrm{Prob}(S)$$

by $(\mu * \nu)(A) = (\mu \times \nu)(\{(x, y) \in S \times S \mid x \cdot y \in A\})$.

Then $(\mathrm{Prob}(S), *)$ is a compact semigroup.

Universal affine semigroups:

Generalizing to Stochastic Relations

$M_C \colon X \to Y$ generalizes to $f \colon X \to \mathrm{Prob}(Y)$.

For Y compact, T_2, so is $\mathrm{Prob}(Y)$ in weak*-topology.

Theorem $\mathrm{Prob}(S)$ is the universal compact affine semigroup over S.

Universal affine semigroups:

Generalizing to Stochastic Relations

$M_C\colon X \to Y$ generalizes to $f\colon X \to \mathrm{Prob}(Y)$.

For Y compact, T_2, so is $\mathrm{Prob}(Y)$ in weak*-topology.

Theorem $\mathrm{Prob}(S)$ is the universal compact affine semigroup over S.

Moreover, $x \mapsto \delta_x\colon S \to \mathrm{Prob}(S)$ sends S into the set of extreme points of $\mathrm{Prob}(S)$.

Universal affine semigroups:

Generalizing to Stochastic Relations

$M_C\colon X \to Y$ generalizes to $f\colon X \to \mathrm{Prob}(Y)$.

For Y compact, T_2, so is $\mathrm{Prob}(Y)$ in weak*-topology.

Theorem $\mathrm{Prob}(S)$ is the universal compact affine semigroup over S.

Moreover, $x \mapsto \delta_x\colon S \to \mathrm{Prob}(S)$ sends S into the set of extreme points of $\mathrm{Prob}(S)$.

If S is a compact monoid, then $\langle\{\delta_g \mid g \in G(S)\}\rangle$ corresponds to the doubly stochastic matrices.

For G a compact group, $M(\langle\{\delta_g \mid g \in G\}\rangle) = \{\mu_G\}$ – Haar measure on G.