Abstract Entropies for Abstract Channels

Annabelle McIver
Carroll Morgan

Motivation

- Aim to synthesise new ways of looking at practical, operationally-motivated techniques based on a complementary approach from programming-language semantics and refinement theory.

- The result is generalisations (from gain functions to Abstract Entropies), and connections to standard mathematics (Giry monads and Kantorovich metrics).

- New results (we think!) based on studying the relationship between the Kantorovich distance and "Abstract Entropy Leakage" of channels.

Old Stuff…

We first recall some details about channels

A *channel* is a triple $(\mathcal{X}, \mathcal{Y}, C)$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite sets (of secret input values and observable output values) and C is an $|\mathcal{X}|\times|\mathcal{Y}|$ *channel matrix* whose entries are between 0 and 1 and whose rows each sum to 1.

For *prior distribution* π on $\mathcal{X}$, the *joint distribution* on $\mathcal{X}\times\mathcal{Y}$ is $p(x, y) = \pi[x]C_{x,y}$, with jointly distributed random variables X, Y whose marginal probabilities are given by $p(x) = \sum_y p(x, y)$ and $p(y) = \sum_x p(x, y)$, and whose conditional probabilities are given by $p(y|x) = {}^{p(x,y)}/_{p(x)}$ (if $p(x)$ is non-zero) and $p(x|y) = {}^{p(x,y)}/_{p(y)}$ (if $p(y)$ is non-zero).

For a given y (such that $p(y)$ is non-zero), the conditional probabilities $p(x|y)$ for each $x\in\mathcal{X}$ form the *posterior distribution* $p_{X|y}$, which is the knowledge that the adversary learns about X by seeing output y.

(From Post-14 submission with Geoffrey Smith.)

Channels

Example 1. Given $\mathcal{X} = \{x_1, x_2, x_3\}$, and $\mathcal{Y} = \{y_1, y_2, y_3, y_4\}$, and (the uniform) prior $\pi = (1/3, 1/3, 1/3)$, consider channel C and its associated joint matrix J as follows:

C	y_1	y_2	y_3	y_4
x_1	1	0	0	0
x_2	0	$1/2$	$1/4$	$1/4$
x_3	$1/2$	$1/3$	$1/6$	0

leads via π to the joint matrix

J	y_1	y_2	y_3	y_4
x_1	$1/3$	0	0	0
x_2	0	$1/6$	$1/12$	$1/12$
x_3	$1/6$	$1/9$	$1/18$	0

.

By summing J's columns we get the (marginal) distribution $p_Y = (1/2, 5/18, 5/36, 1/12)$ and by normalizing the columns we get the posterior distributions $p_{X|y_1} = (2/3, 0, 1/3)$, $p_{X|y_2} = (0, 3/5, 2/5)$, $p_{X|y_3} = (0, 3/5, 2/5)$ and $p_{X|y_4} = (0, 1, 0)$. □

Note for later: these are the same.

Leakage Measures

Shannon leakage : $H(\pi) - H(\pi, C)$

This is based on the Shannon entropy of the prior distribution, $H(\pi) = -\sum_x \pi[x] \log \pi[x]$, and the expected Shannon entropy of the posterior distributions, $H(\pi, C) = \sum_y p(y) H(p_{X|y})$.

Guessing entropy leakage: $G(\pi) - G(\pi, C)$

This is based on the guessing entropy of the prior distribution, $G(\pi) = \sum_i i\, \pi[x_i]$, with $\mathcal{X}$ indexed in non-increasing probability order, and on the expected guessing entropy of the posterior distributions $G(\pi, C) = \sum_y p(y) G(p_{X|y})$.

One-try leakage: $|V(\pi) - V(\pi, C)|$

This is based on *min-entropy*, where the chance that the secret is guessed in one try is measured. $V(\pi) = \max_x \pi[x]$, and $V(\pi, C) = \sum_y p(y) V(p_{X|y})$.

None of the leakage measures depend on the value of the observables.

New Stuff…

(From Post-14 submission with Geoffrey Smith.)

Abstract Channels and Hyper-distributions

Definition 2 (Abstract channel). *The leakage semantics of a channel matrix is the mapping that it gives from priors to hyper-distributions.*

We call such a mapping an abstract channel.

J	y_1	y_2	y_3	y_4
x_1	$1/3$	0	0	0
x_2	0	$1/6$	$1/12$	$1/12$
x_3	$1/6$	$1/9$	$1/18$	0

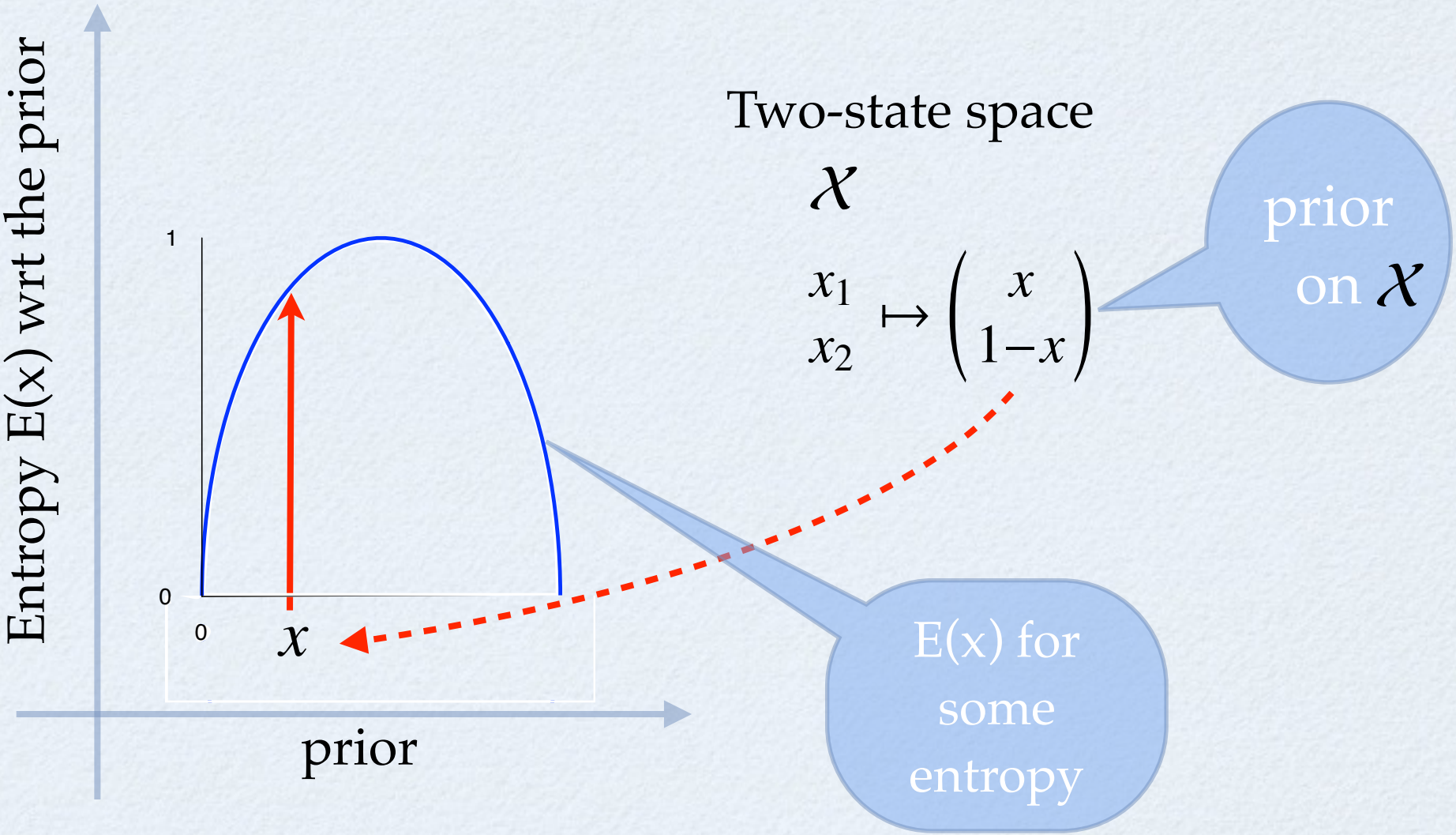
	$1/2$	$15/36$	$1/12$
x_1	$2/3$	0	0
x_2	0	$3/5$	1
x_3	$1/3$	$2/5$	0

Joint distribution y_2, y_3 normalise to the same posterior distribution, but different marginals (10/36 & 5/36). Just add to form a single column.

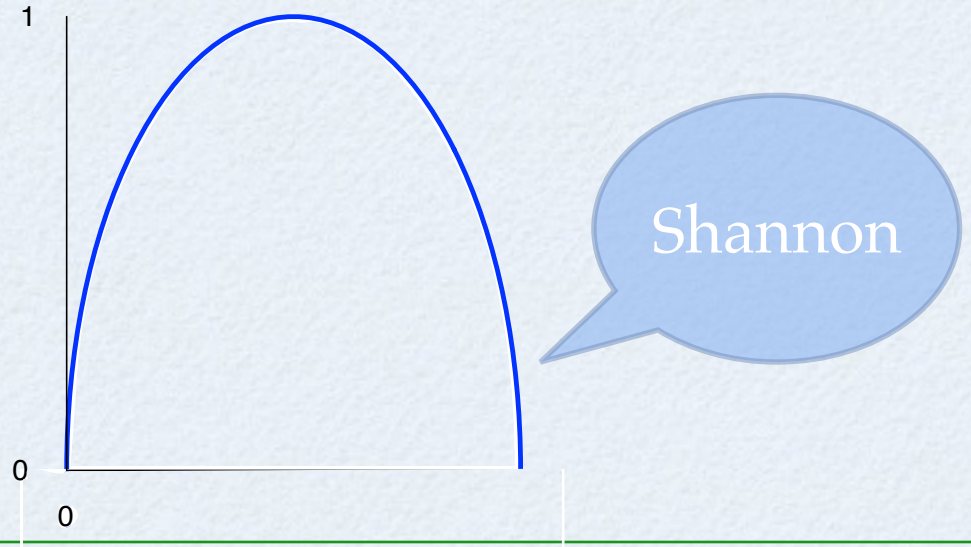
The resulting "hyper-distribution" with middle columns merged and observable names replaced with their marginal probabilities. This is all we need to study leakage.

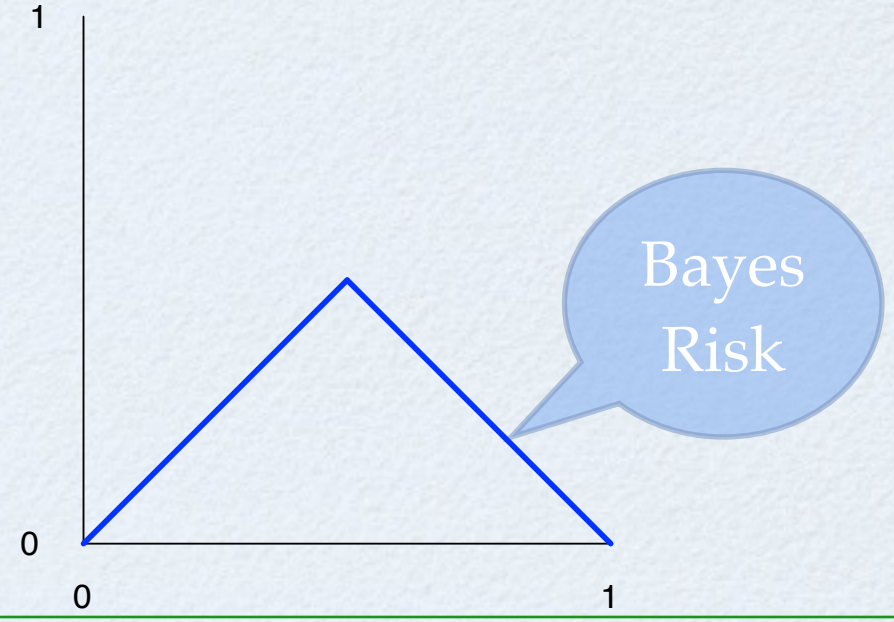
Even newer Stuff…

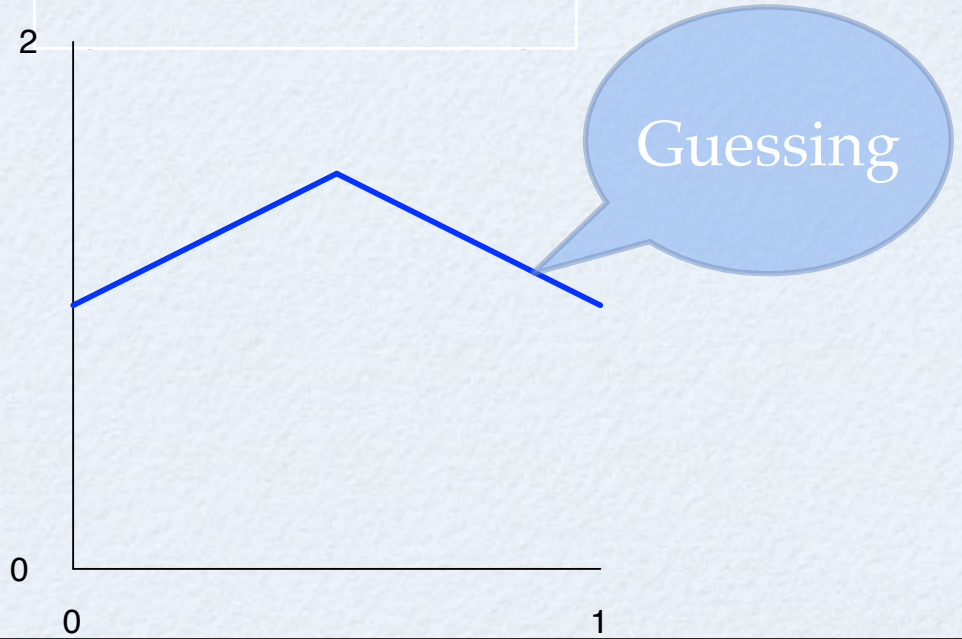
What do entropies have in common?

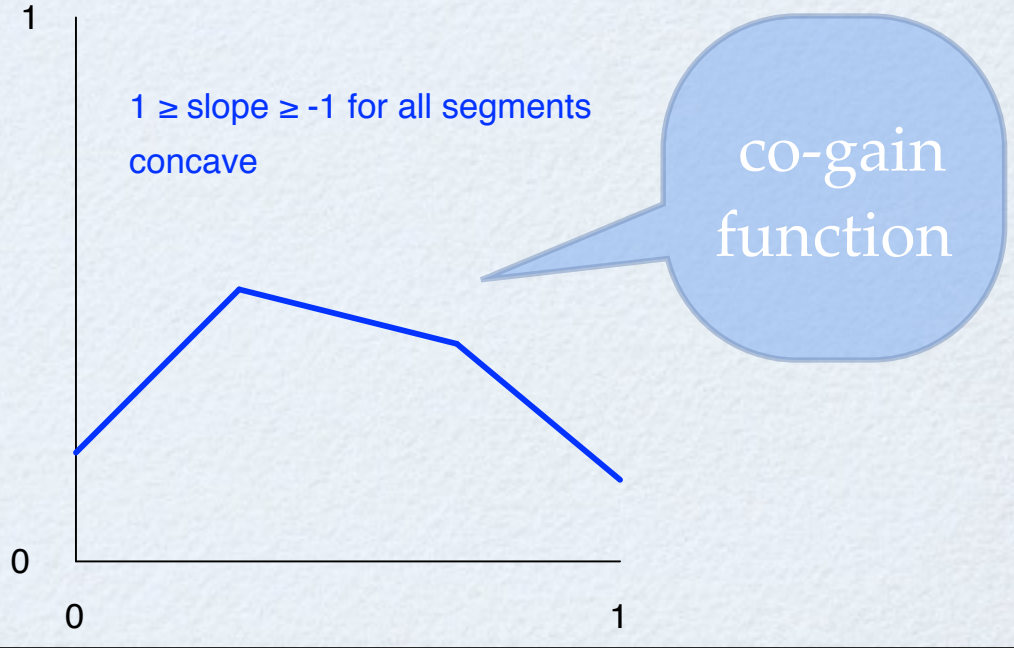


They are all concave, continuous functions of the prior.
In fact we're going to restrict to 1-Lipshitz functions (for the moment excluding Shannon…)



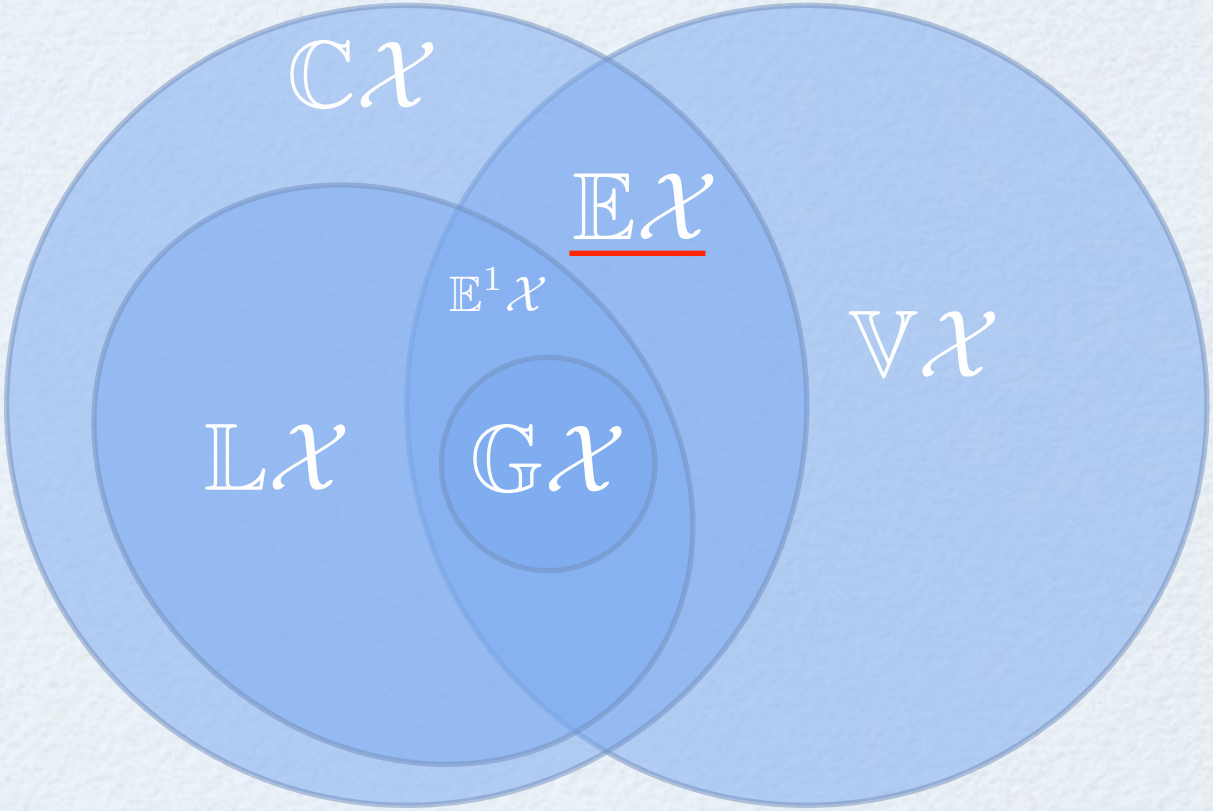






Abstract entropies

Part of a family of functions on discrete distributions

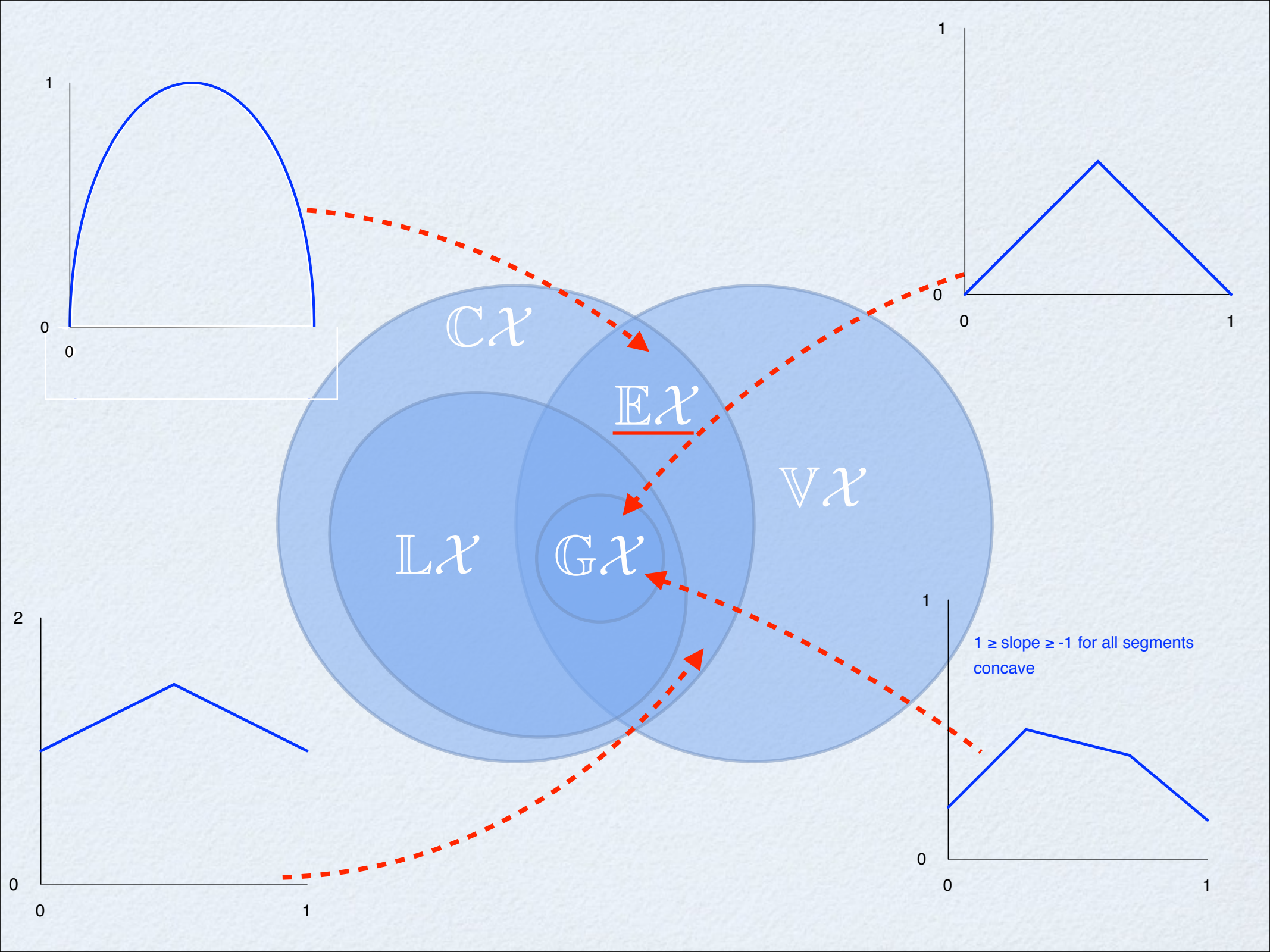


$\mathbb{D}\mathcal{X}$	Discrete distribution on (finite) set $\mathcal{X}$
$\mathbb{C}\mathcal{X}$	Continuous functions in $\mathbb{D}\mathcal{X}\to\mathbb{R}^+$
$\mathbb{V}\mathcal{X}$	ConcaVe functions in $\mathbb{D}\mathcal{X}\to\mathbb{R}^+$
$\mathbb{L}\mathcal{X}$	1-Lipschitz functions $\mathcal{L}^1(\mathbb{D}\mathcal{X})\to\mathbb{R}^+$

Standard constructions

Novel constructions

$\mathbb{G}\mathcal{X}$	*co-Gain Functions* in $\mathbb{D}\mathcal{X}\to[0,1]$
$\mathbb{E}\mathcal{X}$	*Abstract Entropies* in $\mathbb{D}\mathcal{X}\to\mathbb{R}^+$


1
0
0
1
0
1
0
1
$\mathbb{C}\mathcal{X}$
$\mathbb{E}\mathcal{X}$
$\mathbb{V}\mathcal{X}$
$\mathbb{L}\mathcal{X}$
$\mathbb{G}\mathcal{X}$
2
0
0
1
1
0
0
1
1 ≥ slope ≥ -1 for all segments
concave

Abstract channels ... abstractly

$\mathbb{D}\mathcal{X}$ — The set of distributions over finite state space $\mathcal{X}$

$\mathbb{D}^2\mathcal{X}$ — The set of hyper-distributions over $\mathbb{D}\mathcal{X}$

$[\pi]$ — the point hyper-distribution based on the prior π

$[\pi, C]$ — The hyper-distribution output by C w.r.t. prior $\pi \in \mathbb{D}\mathcal{X}$

$V_g(\Delta)$ — The expected value of gain function g w.r.t. $\Delta \in \mathbb{D}^2\mathcal{X}$.

An abstract channel has type $\mathbb{D}\mathcal{X} \to \mathbb{D}^2\mathcal{X}$

$V_{br}[\pi, C]$, where $br(\pi)$ gives the "Bayes Risk"

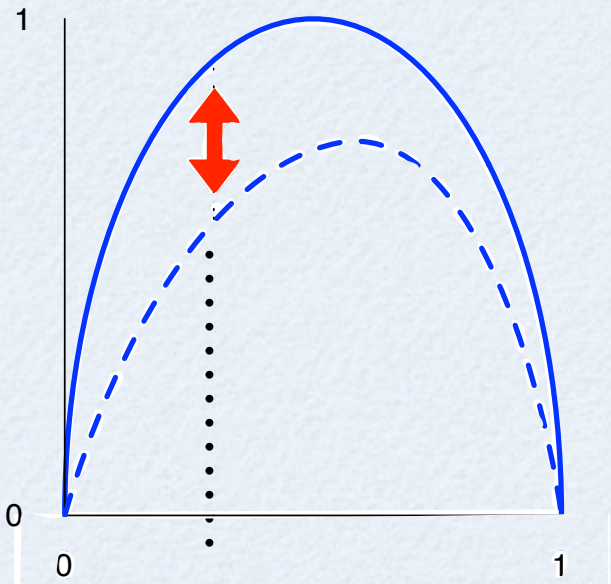
Leakage Measures on Abstract Channels

Channel over space {x_1, x_2} ⋯⋯▶ $\begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix}$

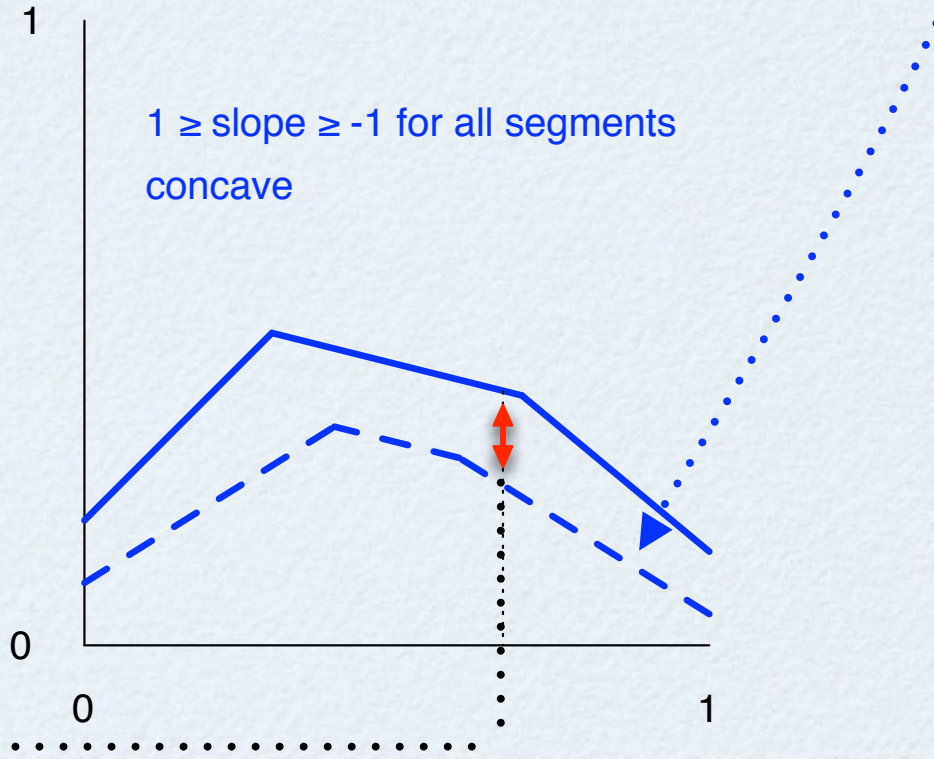
Resulting hyperdistribution w.r.t. prior (x, 1-x) ⋯⋯▶

	$x/2$	$1-x/2$
x_1	1	$x/(2-x)$
x_2	0	$2(1-x)/(2-x)$

Expected value of the "entropy" wrt the hyper-distribution



Respective leakages at a prior x are given by the difference.



New results…

Leakage formulation now has (almost) the same form as the formulation for the Kantorovich distance between hyper-distributions.

Abstract Leakage. $$|\Delta_1 - \Delta_2|_{\mathbb{E}^1} := \max_{h \in \mathbb{E}^1 \mathcal{X}} |E_h(\Delta_1) - E_h(\Delta_2)|$$

Kantorovich distance. $$|\Delta_1 - \Delta_2|_{\mathbb{L}} := \max_{h \in \mathbb{L} \mathcal{X}} |E_h(\Delta_1) - E_h(\Delta_2)|$$

$$\mathbb{E}^1 \mathcal{X} \subseteq \mathbb{L} \mathcal{X} \quad \Rightarrow \quad |\Delta_1 - \Delta_2|_{\mathbb{E}^1} \leq |\Delta_1 - \Delta_2|_{\mathbb{L}}$$

Metric on $\mathbb{D}\mathcal{X}$ is the Manhattan distance: $$|\pi_1 - \pi_2|_M := 1/2 \times \sum_{x \in \mathcal{X}} |\pi_1(x) - \pi_2(x)|$$

Co-gain functions are 1-Lipshitz; this implies we have some immediate bounds given by the Kantorovich distance, and measuring between the point prior and the hyper-distribution made as a result of the channel leakage semantics.

Gain-function leakage

$$|[\pi] - [\pi, C]|_{\mathbb{G}} \quad := \quad \max_{g \in \mathbb{G}\mathcal{X}} |\overline{V}_g([\pi]) - \overline{V}_g[\pi, C]|$$

$$|[\pi] - [\pi, C]|_{\mathbb{G}} \quad \leq \quad |[\pi] - [\pi, C]|_{\mathbb{E}^1} \quad \leq \quad |[\pi] - [\pi, C]|_{\mathbb{L}}$$

In the special case that one of the hyperdistributions is a point hyperdistribution, the Kantorovich-Rubenstein theorem implies that it is optimised by a co-gain (style) function. This makes the Kantorovich distance equal to the "abstract entropy additive leakage".

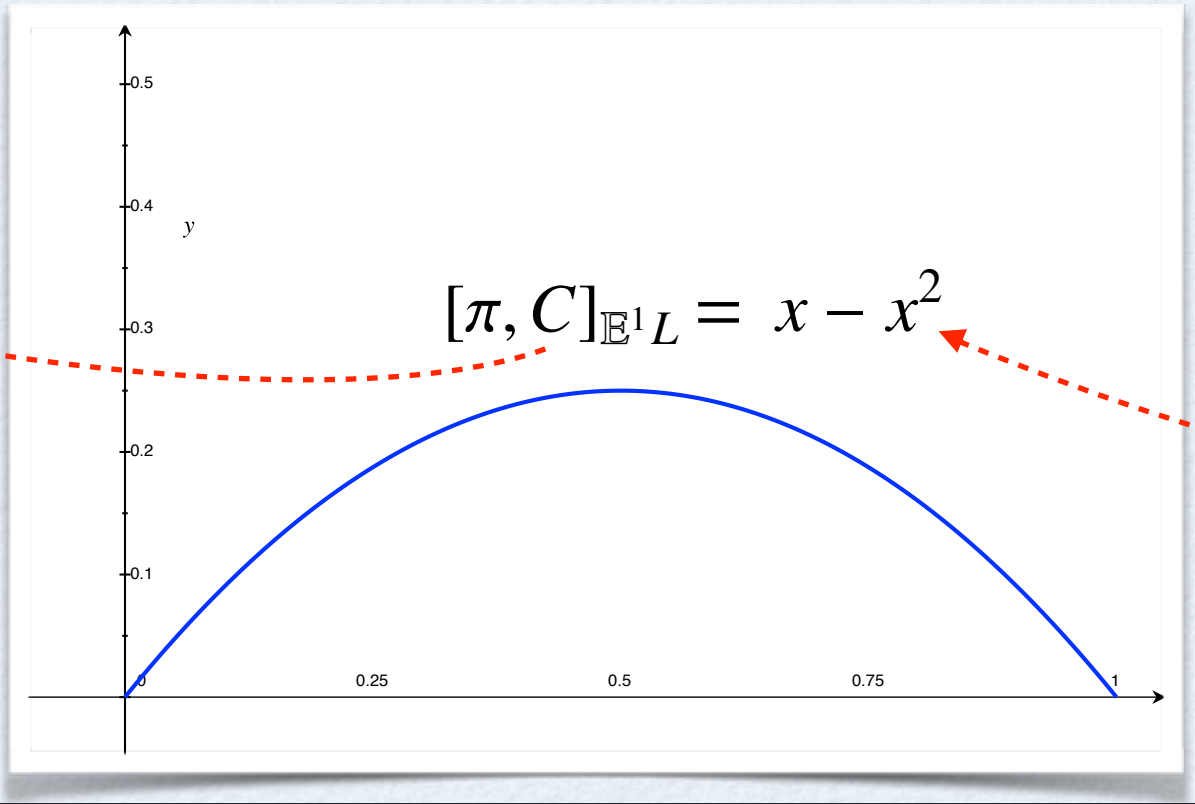
$$|[\pi] - [\pi, C]|_{\mathbb{E}^1} \quad = \quad |[\pi] - [\pi, C]|_{\mathbb{L}}$$

The Leakage coincides with the Kantorovich distance:

$$|[\pi] - [\pi, C]|_{\mathbb{E}^1} \quad = \quad |[\pi] - [\pi, C]|_{\mathbb{L}}$$

This is equivalent to the "Earth Moving Distance" between $[\pi]$ and $[\pi, C]$ for which there are straightforward algorithms.

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

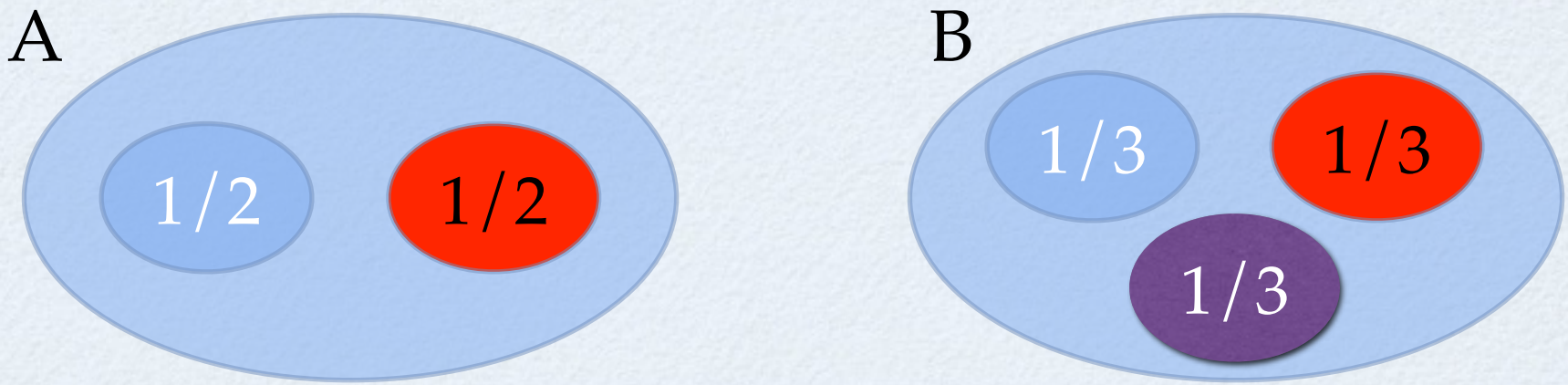


$$[\pi, C]_{\mathbb{E}^1 L} = x - x^2$$

$$\begin{pmatrix} x \\ 1-x \end{pmatrix}$$

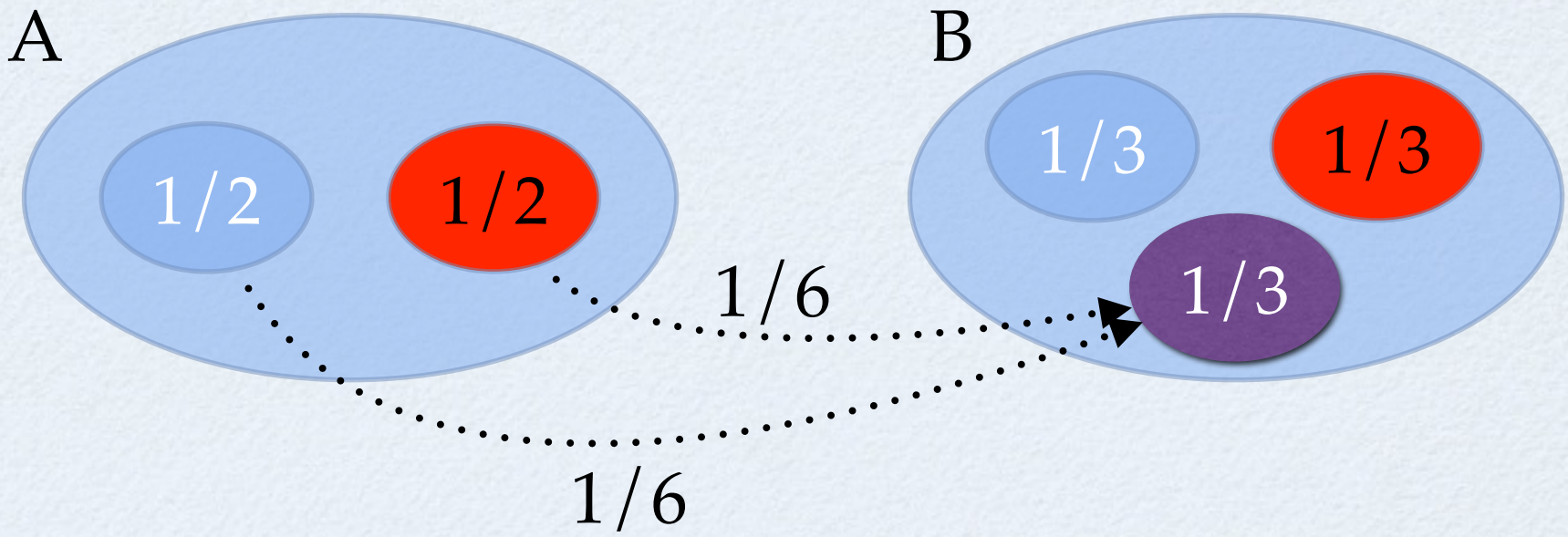
Earth Moving "at a Glance"

A and B are distributions over blobs.



Think of the cost of transforming A into B as "moving" weight from the A-blobs to the B-blobs. The average cost is "distance" between blobs (for a move) multiplied by the "amount" needed to be moved.

Earth Moving "at a Glance"



The cost of this move is 1/6 * d(,) + 1/6* d(,)

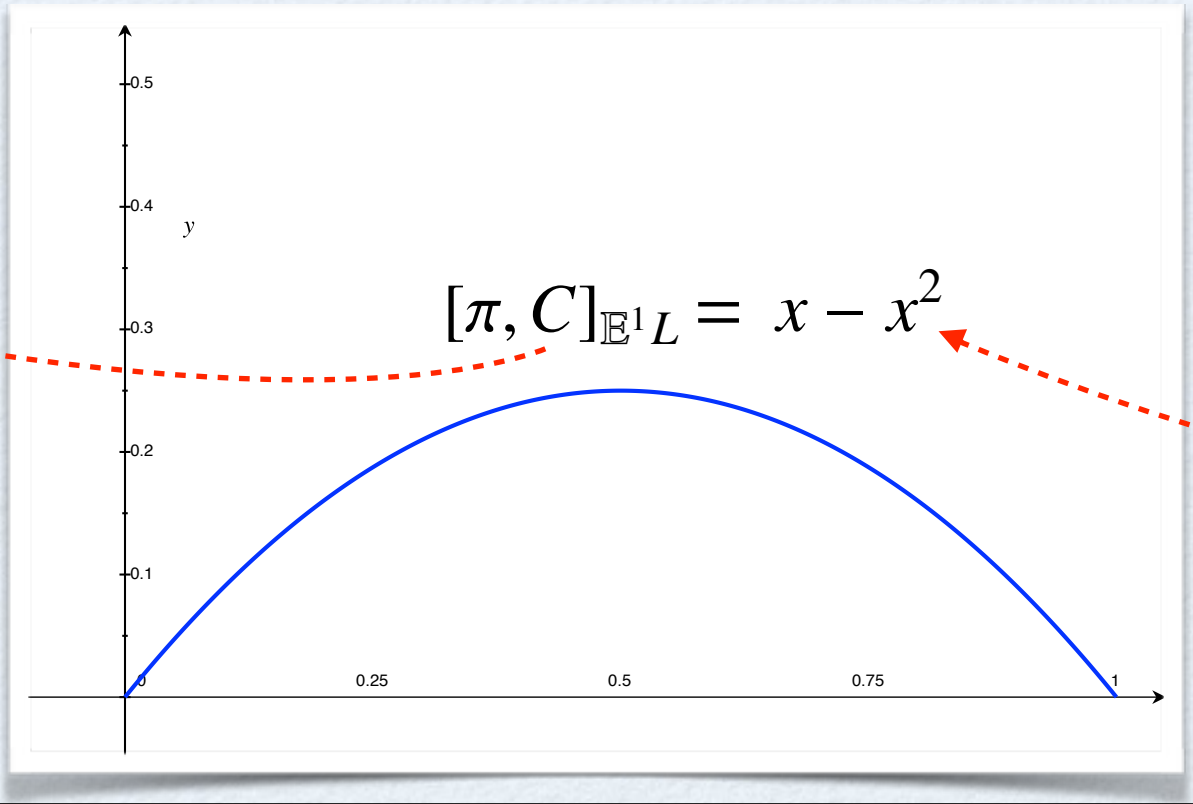
The Earth Moving distance is the smallest transformation cost.

The Leakage coincides with the Kantorovich distance:

$$|[\pi] - [\pi, C]|_{\mathbb{E}^1} \quad = \quad |[\pi] - [\pi, C]|_{\mathbb{L}}$$

This is equivalent to the "Earth Moving Distance" between $[\pi]$ and $[\pi, C]$ for which there are straightforward algorithms to compute it.

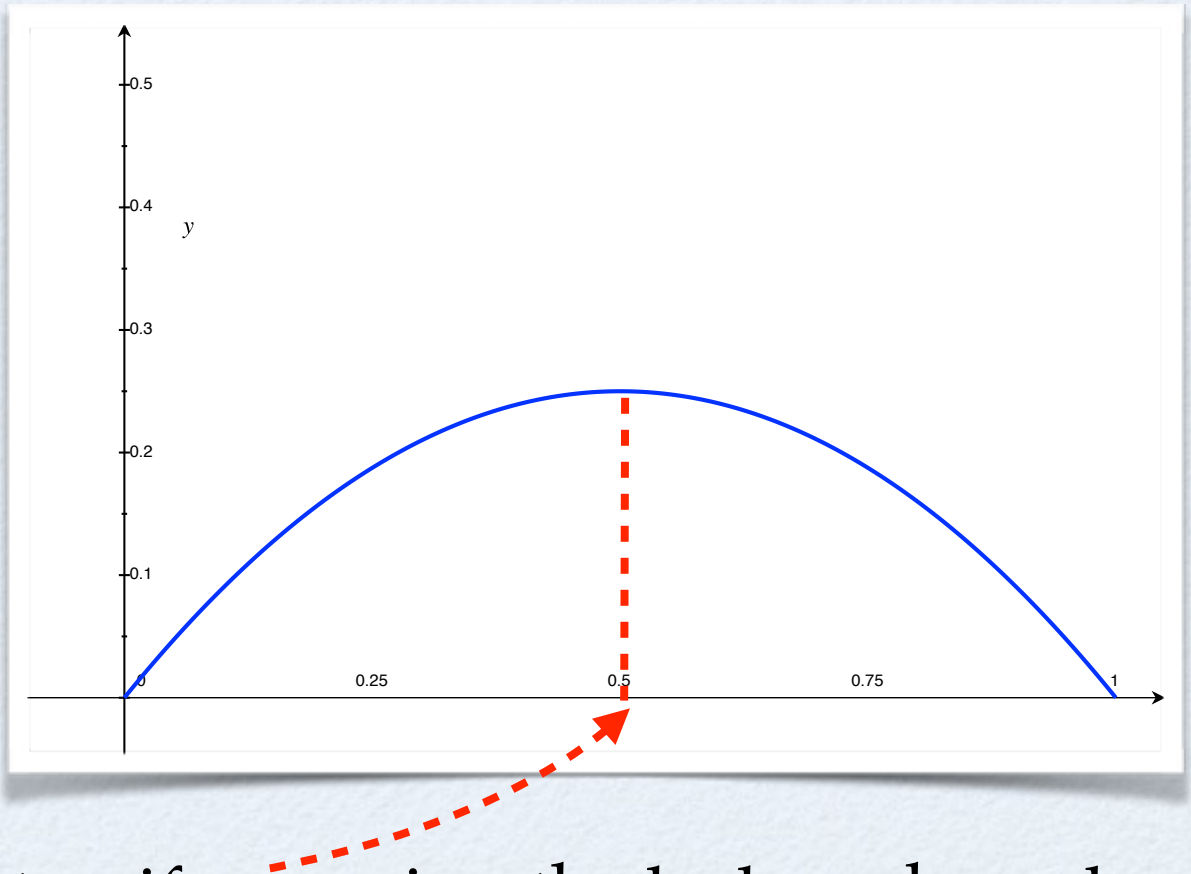
$$\begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix}$$



$$[\pi, C]_{\mathbb{E}^1 L} = x - x^2$$

$$\begin{pmatrix} x \\ 1-x \end{pmatrix}$$

What is the maximal leakage, and where does it occur?

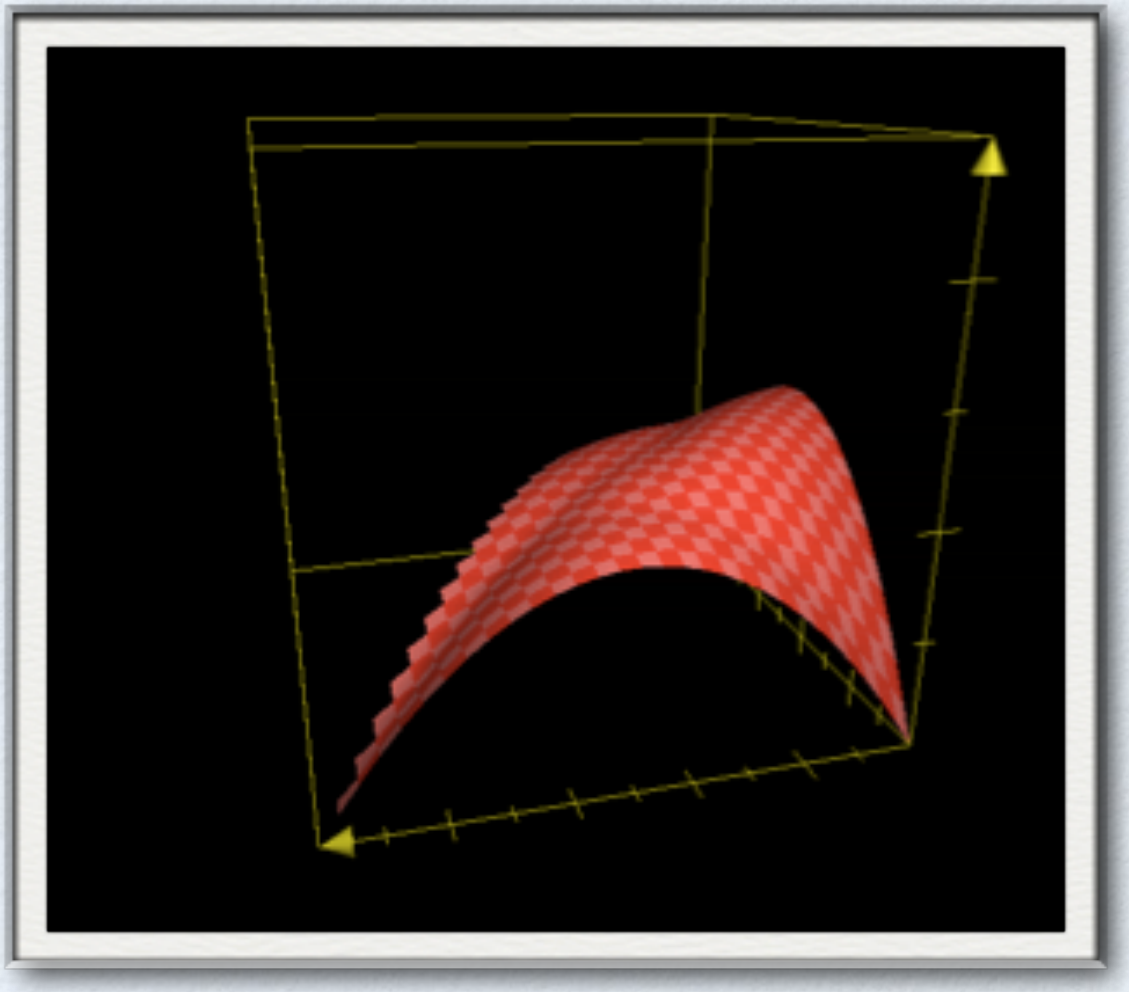


Optimal at uniform prior: the leakage here dominates all other leakages *whatever the initial gain function*.

Moreover, for all 2 x M channels, the optimal leak always occurs at the uniform prior.

For 3 x M matrices, the optimal does not always occur at the uniform prior.

Optimal occurs at (0.44,0.09,0.47)



Where next…

- Define an "Abstract Channel Additive Leakage" measure as an independent measure of the channel.

- Solve the optimisation problem of finding the prior that maximises the Abstract Channel Additive Leakage.

- Develop algebras and semantics for Channels by utilising the Giry/Kantorovich perspective;

- …