Resource Bounded Symmetry of Information

Troy Lee
CWI and University of Amsterdam

Andrei Romashchenko
Institute for Information Transmission Problems

Kolmogorov Complexity

- Developed as a way to measure randomness in individual strings

- $C_T(x|y) = \min_p\{|p| : T(p,y) = x\}$

- Invariance Theorem: We define $C(x|y) = C_U(x|y)$ for a universal machine U. This choice affects our definition by at most an additive constant factor.

Symmetry of Information

- $C(x,y) = C(x) + C(y|x)$ for any x, y. Proven independently by Kolmogorov and Levin.

- One direction is easy: $C(x,y) \leq C(x) + C(y|x)$. The other has clever proof.

- Symmetry of information is a useful tool in the Kolmogorov complexity toolbox. Proofs using symmetry of information are usually difficult to directly replace by counting arguments.

Resource Bounded Symmetry of Information

- $C^t(x\,|\,y) = \min_p\{|p| : U(p,y) = x$ in $t(|x|+|y|)$ steps.$\}$

- Standard proof works for exponential time, or polynomial space. Things become interesting for polynomial time bounds.

- Before P and NP, Kolmogorov suggested time bounded symmetry of information as a good way to show exhaustive search cannot be eliminated.

- We call polynomial time symmetry of information the statement: for any polynomial time bound q there exists polynomial q':

$$C^q(x,y) \geq C^{q'}(x) + C^{q'}(y\,|\,x)$$

What is Known

- If P=NP then polynomial time symmetry of information holds (Longpré-Watanabe, 95)

- If polynomial time symmetry of information holds, then poly time computable functions can be inverted on large fraction of range (Longpré-Mocas, 93).
  - $C^q(f(x)\,|\,x) = O(1)$
  - If $C^q(x\,|\,f(x)) = O(\log n)$ then we can invert f on $f(x)$.

- Can a weaker form of symmetry of information hold?
  Can symmetry of information hold for other complexity measures?

Nondeterministic Printing Complexity

We define $\mathrm{CN}^t(x \mid y)$ as the length of a shortest program p such that

- $U_n(p, y)$ has at least one accepting path
- $U_n(p, y)$ outputs x on every accepting path
- $U_n(p, y)$ runs in $O(t(|x|))$ steps.

Similarly we define $\mathrm{CAM}^t(x \mid y)$ based on the complexity class AM.

Language Compression Theorems

Language Compression Theorem (Buhrman-L-van Melkebeek, 04):
For any $A \in \mathrm{NP}$, there is a polynomial q s.t. for all $x \in A^{=n}$

- $\mathrm{CN}^q(x) \leq \log \|A^{=n}\| + \tilde{O}(\sqrt{\log \|A^{=n}\|})$

- $\mathrm{CAM}^q(x) \leq \log \|A^{=n}\| + O(\log^3 n)$

A Negative Result

There is an oracle A where

$$(2-\varepsilon)\mathrm{CN}^{q,A}(x,y) \leq \mathrm{CN}^{q',A}(x) + \mathrm{CN}^{q',A}(y\,|\,x)$$

- Notice this is tight as $\mathrm{CN}^{q}(x,y) \geq \max\{CN^{q}(x), CN^{q}(y)\}$
- Proof uses language compression theorem

The Hard Direction, Prerequisites

To show $C(x,y) \geq C(x) + C(y\,|\,x)$ we will use three facts:

1. The set $\{x : C(x) \leq m\}$ is of size less than 2^{m+1}.

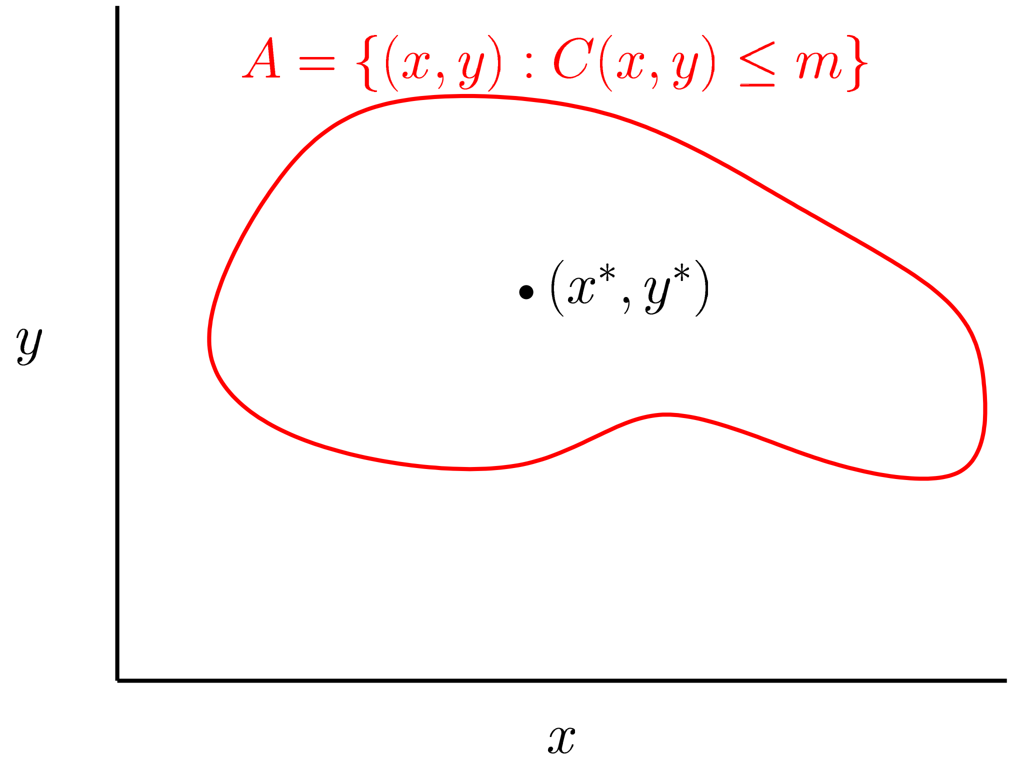
2. The set $\{x : C(x) \leq m\}$ is recursively enumerable.

3. Language Compression Theorem: For any recursively enumerable set A, and all $x \in A^{=n}$, $C(x) \leq \log \|A^{=n}\| + O(\log n)$.

Proof of Symmetry of Information, Resource Unbounded Case

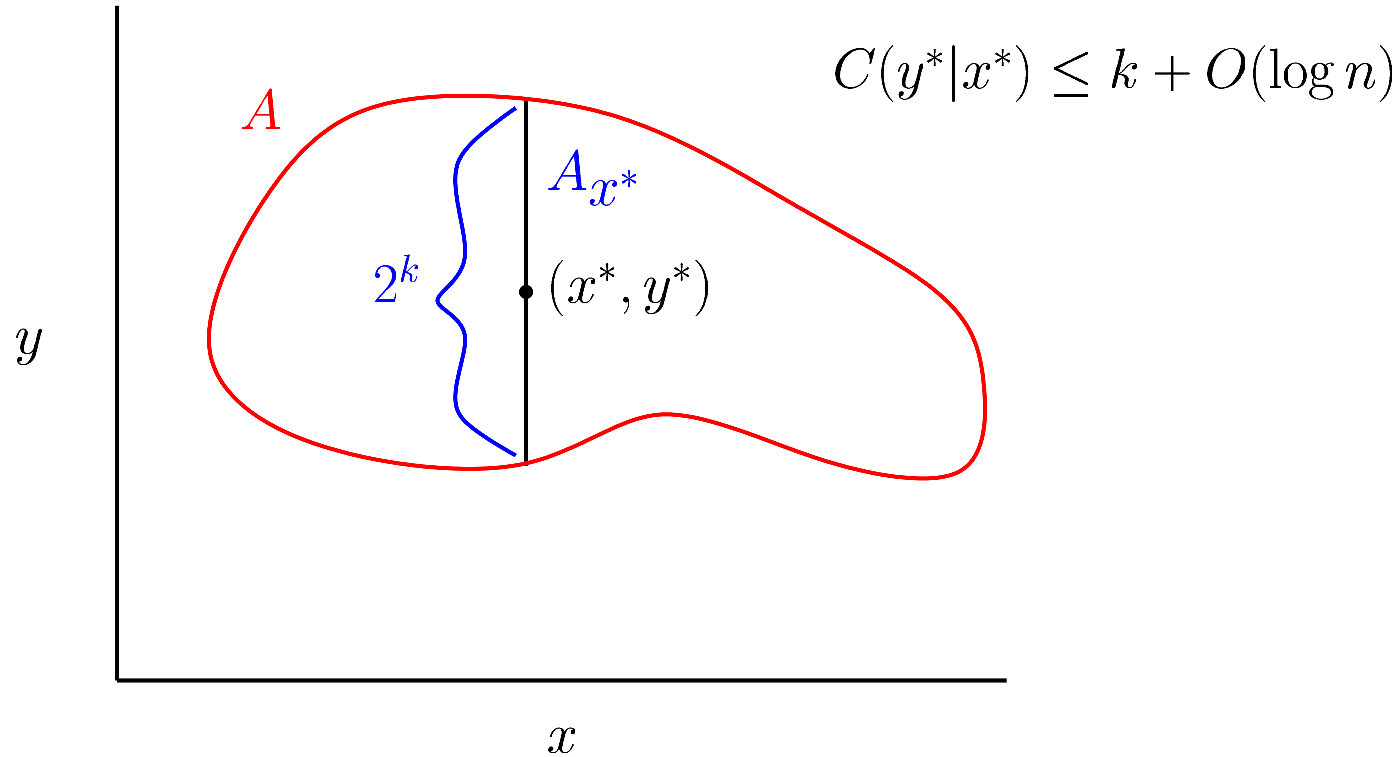
To show: $C(x,y) \geq C(x) + C(y|x)$.

Fix $x^*, y^* \in \{0,1\}^n$, and say $C(x^*, y^*) = m$.



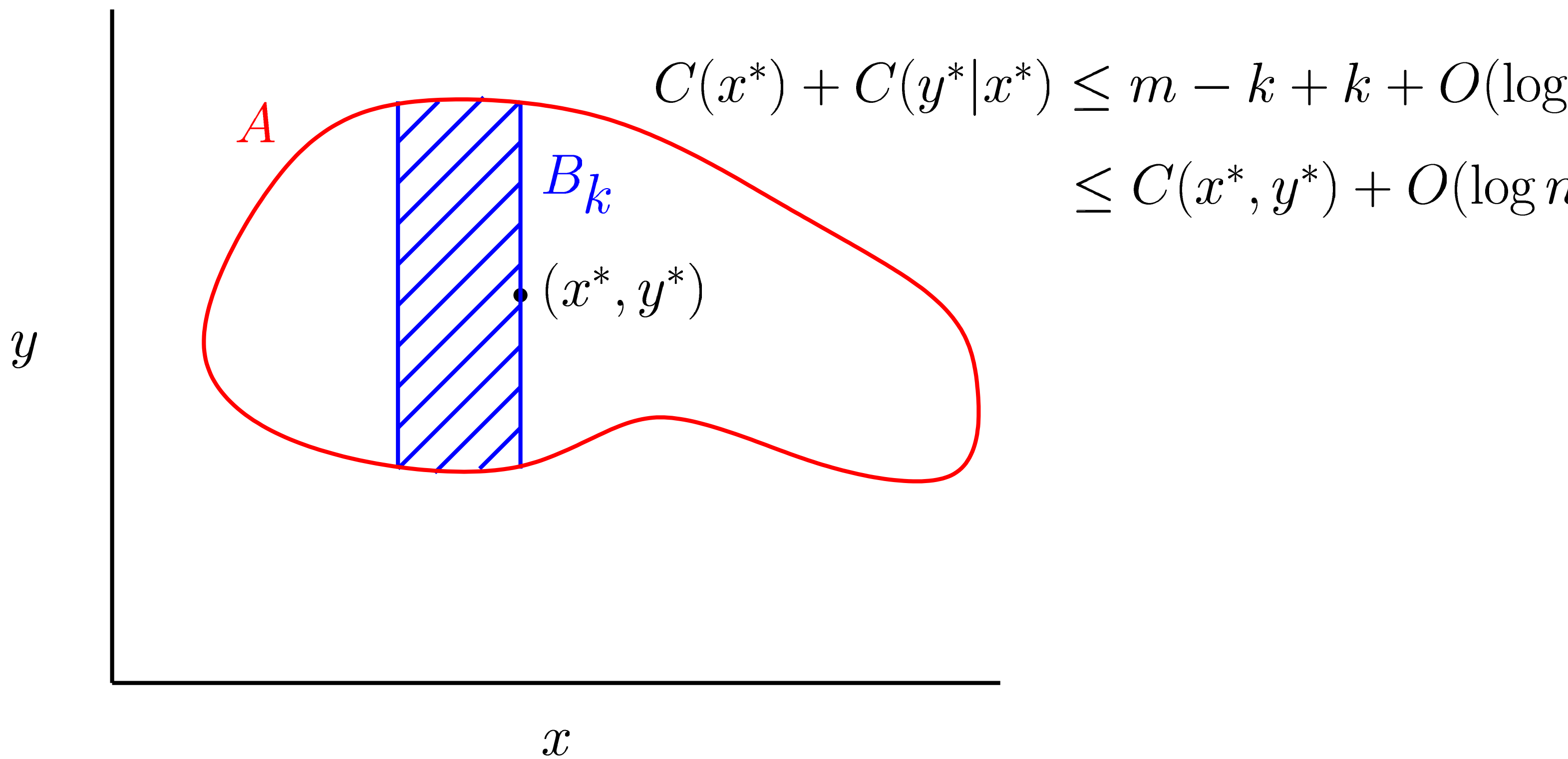
Proof of Symmetry of Information, Resource Unbounded Case
To show: $C(x,y) \geq C(x) + C(y|x)$.

Consider the line $A_{x^*} = \{y : C(x^*, y) \leq m\}$, and say that $2^k \leq \|A_{x^*}\| < 2^{k+1}$.



Proof of Symmetry of Information, Resource Unbounded Case
To show: $C(x,y) \geq C(x) + C(y\,|\,x)$.

Consider $B_k = \{x : \exists^{\geq 2^k} y$ such that $C(x,y) \leq m\}$.
As $\|A\| \leq 2^{m+1}$, we have $\|B_k\| \leq 2^{m-k+1}$.



Adapting Proof to Resource Bounded Case

How do our three facts translate?

1. We still have $\{x : C^t(x) \leq m\}$ is of size less than 2^{m+1}.

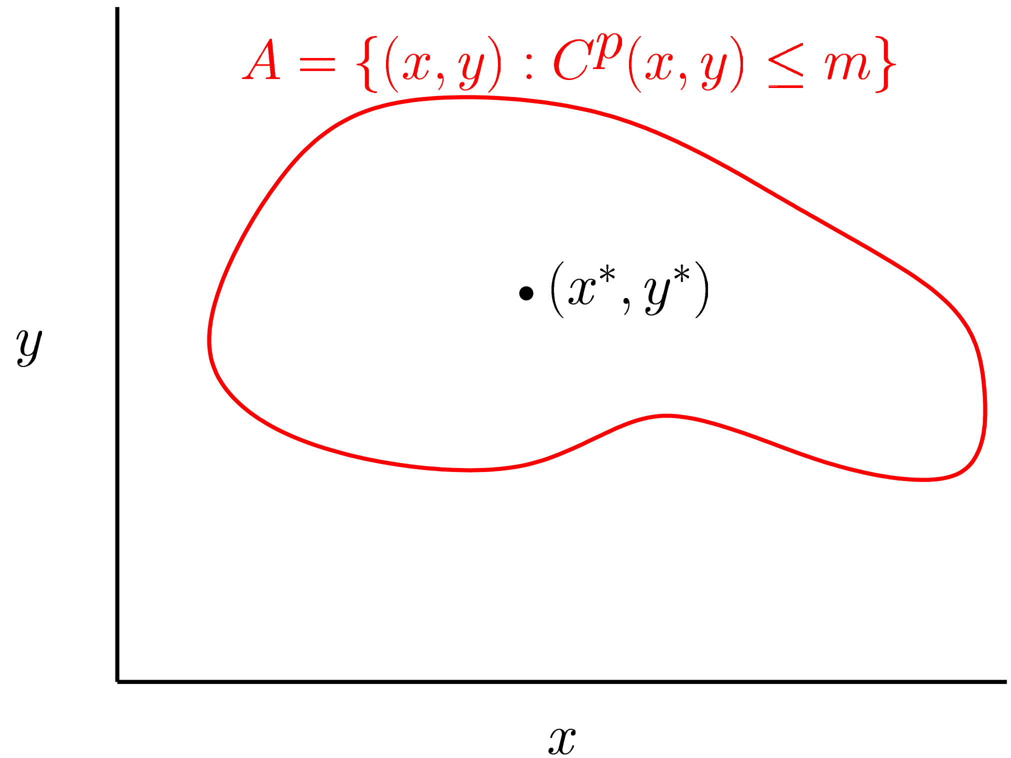
2. If $t(n)$ is polynomial, then the set $\{x : C^t(x) \leq m\}$ is in NP (but probably not in P).

3. For any set $A \in \mathrm{NP}$ there is a polynomial $p(n)$ such that for all $x \in A^{=n}$

$$\mathrm{CAM}^p(x) \leq \log \|A^{=n}\| + O(\log^3 n)$$

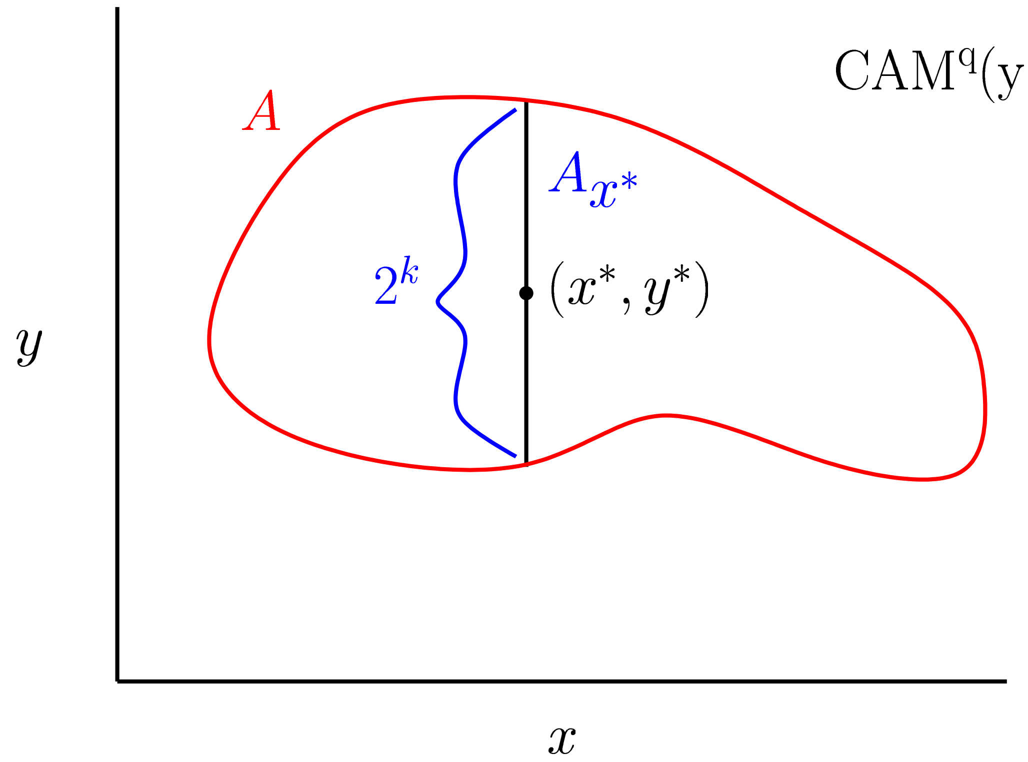
Symmetry of Information, Resource Bounded Case

Fix $x^*, y^* \in \{0,1\}^n$, let $p(n)$ be a polynomial, and say $C^p(x^*, y^*) = m$.



Symmetry of Information, Resource Bounded Case

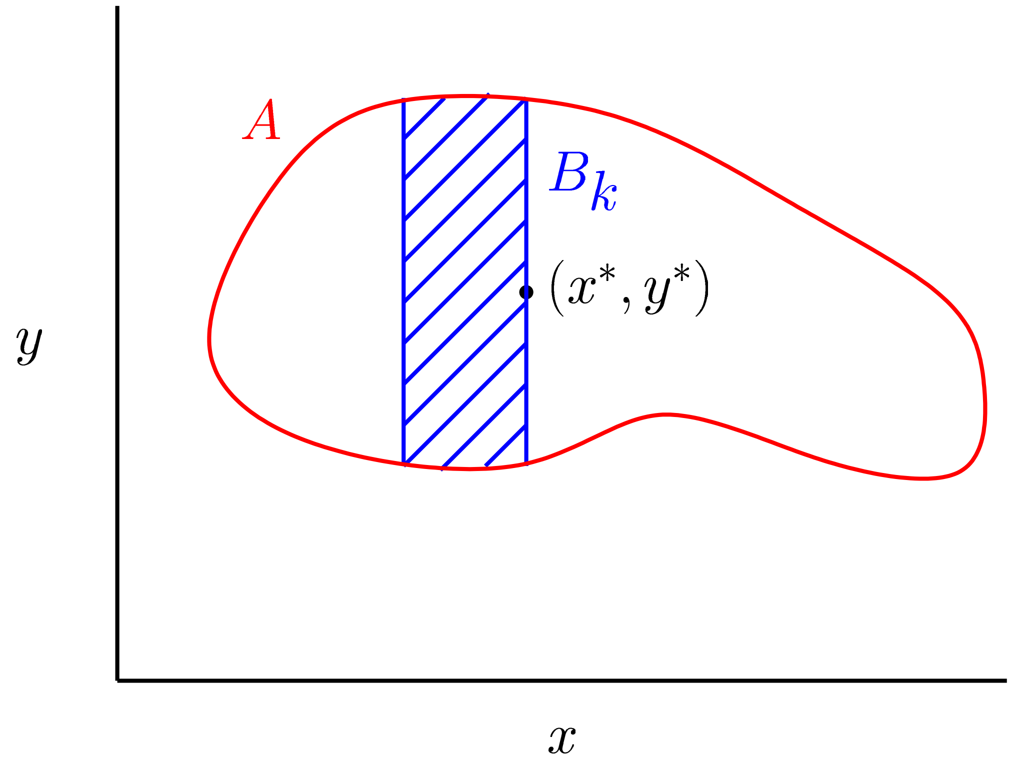
Consider the line $A_{x^*} = \{y : C^p(x^*, y) \leq m\}$, and say that $2^k \leq \|A_{x^*}\| < 2^{k+1}$.



$\mathrm{CAM}^q(y^*|x^*) \leq k + O(\log^3 n)$

Proof of Symmetry of Information, Resource Bounded Case

Consider $B_k = \{x : \exists^{\geq 2^k} y$ such that $C^p(x,y) \leq m\}$.
Again $\|B_k\| \leq 2^{m-k+1}$, but how can we decide if $x \in B_k$?



Lower Bound Counting and AM

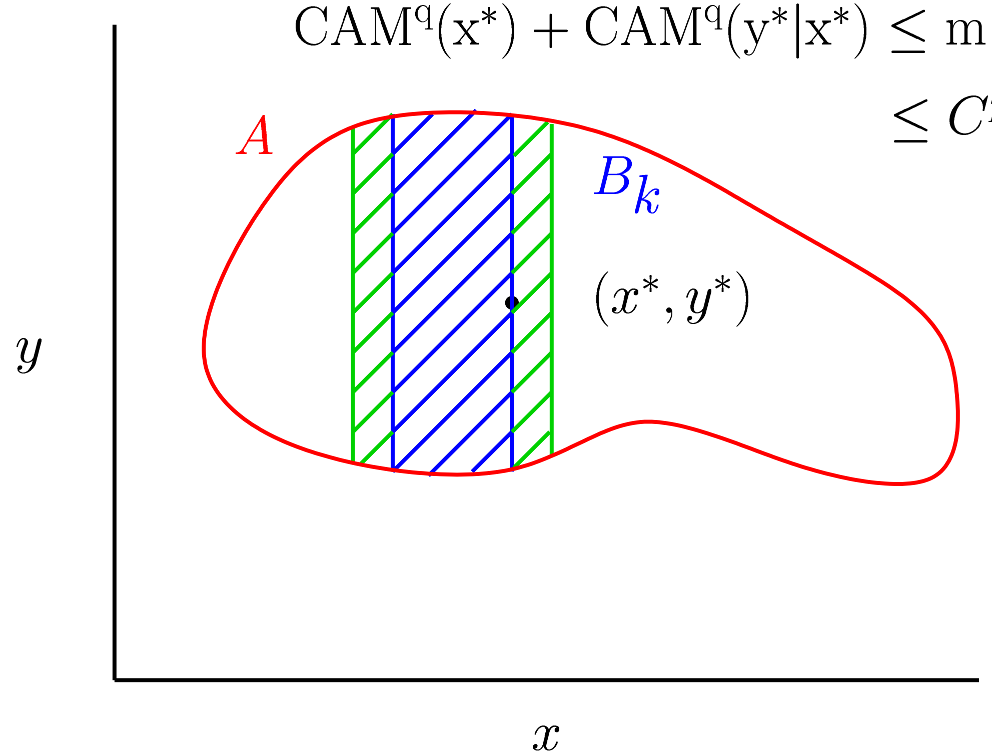
- Sipser's Coding Lemma: If A_x is a set in NP then there is a NP predicate M such that
  - if $\|A_x\| \geq 2^k$ then $\Pr_r[M(x,r) = 1] \geq 2/3$
  - if $\|A_x\| \leq 2^{k-1}$ then $\Pr_r[M(x,r) = 1] < 1/3$
- Extend Language Compression Theorem to work for these AM "gap" sets: Let $B \subseteq \{0,1\}^*$, suppose there is an NP predicate M such that:
  - for all $x \in B^{=n}$, $\Pr_r[M(x,r) = 1] \geq 2/3$
  - $\|\{x : \Pr_r[M(x,r) = 1] > 1/3\}\| \leq 2^\ell$

  Then for all $x \in B^{=n}$, $\mathrm{CAM}^q(x) \leq \ell + O(\log^3 n)$.

Proof of Symmetry of Information, Resource Bounded Case

$B_k = \{x : \exists^{\geq 2^k} y$ such that $C^p(x,y) \leq m\}$. If $x \notin B_{k-1}$ then success probability of M is less than $1/3$. Thus number of elements accepted with success probability greater than $1/3$ is less than 2^{m-k+2}.

$$\begin{aligned} \mathrm{CAM^q(x^*) + CAM^q(y^*|x^*)} &\leq \mathrm{m - k + k + O(\log^3 n)} \\ &\leq C^p(x^*, y^*) + O(\log^3 n) \end{aligned}$$



What We've Shown

- For any polynomial p and $x, y \in \{0,1\}^n$ there is a polynomial q such that

$$C^p(x,y) \geq \mathrm{CAM}^q(x) + \mathrm{CAM}^q(y|x) - O(\log^3 n)$$

- Recall: It was known that P=NP implies polynomial time symmetry of information

- Thus we obtain polynomial time symmetry of information holds under the (seemingly weaker) assumption: for all $x, y \in \{0,1\}^n$,

$$C^p(x\,|\,y) \leq \mathrm{CAM}^q(x\,|\,y) + O(\log n) \qquad (*)$$

- It turns out that $(*)$ implies P=NP

$(*) \Rightarrow \mathrm{P} = \mathrm{NP}$

- Recall $(*) = C^p(x\,|\,y) \leq \mathrm{CAM}^q(x\,|\,y) + O(\log n)$

- $C^p(x\,|\,y) \leq \mathrm{CN}^q(x\,|\,y) + O(\log n) \Rightarrow \mathrm{NP} = \mathrm{RP}$
  (Buhrman-Fortnow-Laplante, 02)

  – let ϕ be formula with exactly one satisfying assignment a. Then $\mathrm{CN}^q(a\,|\,\phi) = O(1)$.

- There is a string x^* with $C^q(x^*) = |x^*|$ and $C^{q',\Sigma_2^p}(x^*) = O(\log n)$.

- By the collapse, $\mathrm{CAM}^{q'}(x^*) = O(\log n)$ and so $C^p(x^*) = O(\log n)$ by $(*)$.

- $C^q(x^*) = |x^*|$ and $C^{q'}(x^*) = O(\log n)$ lets us derandomize RP.

Summary and Open Problems

- $C^q(x,y) \geq \mathrm{CAM}^{q'}(x) + \mathrm{CAM}^{q'}(y\,|\,x) - O(\log^3 n)$

- Can you improve this to $C^q(x,y) \geq \mathrm{CN}^{q'}(x) + \mathrm{CN}^{q'}(y\,|\,x) - O(\log n)$?
  - Doing this would imply $\mathrm{FP}^{\mathrm{NP}_{||}} = \mathrm{FP}^{\mathrm{NP}[\log n]} \Rightarrow \mathrm{P} = \mathrm{NP}$

- Does $C^q(x\,|\,y) \leq \mathrm{CN}^{q'}(x\,|\,y) + O(\log n)$ imply $P = NP$?