CONSTANT TIME APPROXIMATION SCHEMES: PARADIGMS & RECENT DEVELOPMENTS

MAREK KARPINSKI,
UNIV. OF BONN.

A TALE ON EFFICIENT APPROXIMATIONS AND SAMPLING, IN FACT, ON IMPROVING BOTH OF THEM BETWEEN 1994–2006.

WORK DONE AROUND 1994:

F. DE LA VEGA (FOR FRIENDS LALO) [F96]

COMBINATORIAL METHOD FOR DENSE MAX-CUT.

ARORA, KARGER, K.

[AKK95]

SMOOTH HIGH DEG. POLY. INTEGER PROGRAMS FOR DENSE MAX-CSP.

- PROPERTY TESTING
- STREAMING ALGS.

⋮

General Framework:

- Problem with Massive Data. Too Large to be Stored in RAM.
- Natural Approach:

  Draw a Small Sample from RAM and Process Sample by Special Approx. Algorithm.

MOTIVATED ALSO BY VARIOUS CLASSICAL OPT. PROBLEMS, LIKE PARTITIONING PROBLEMS.

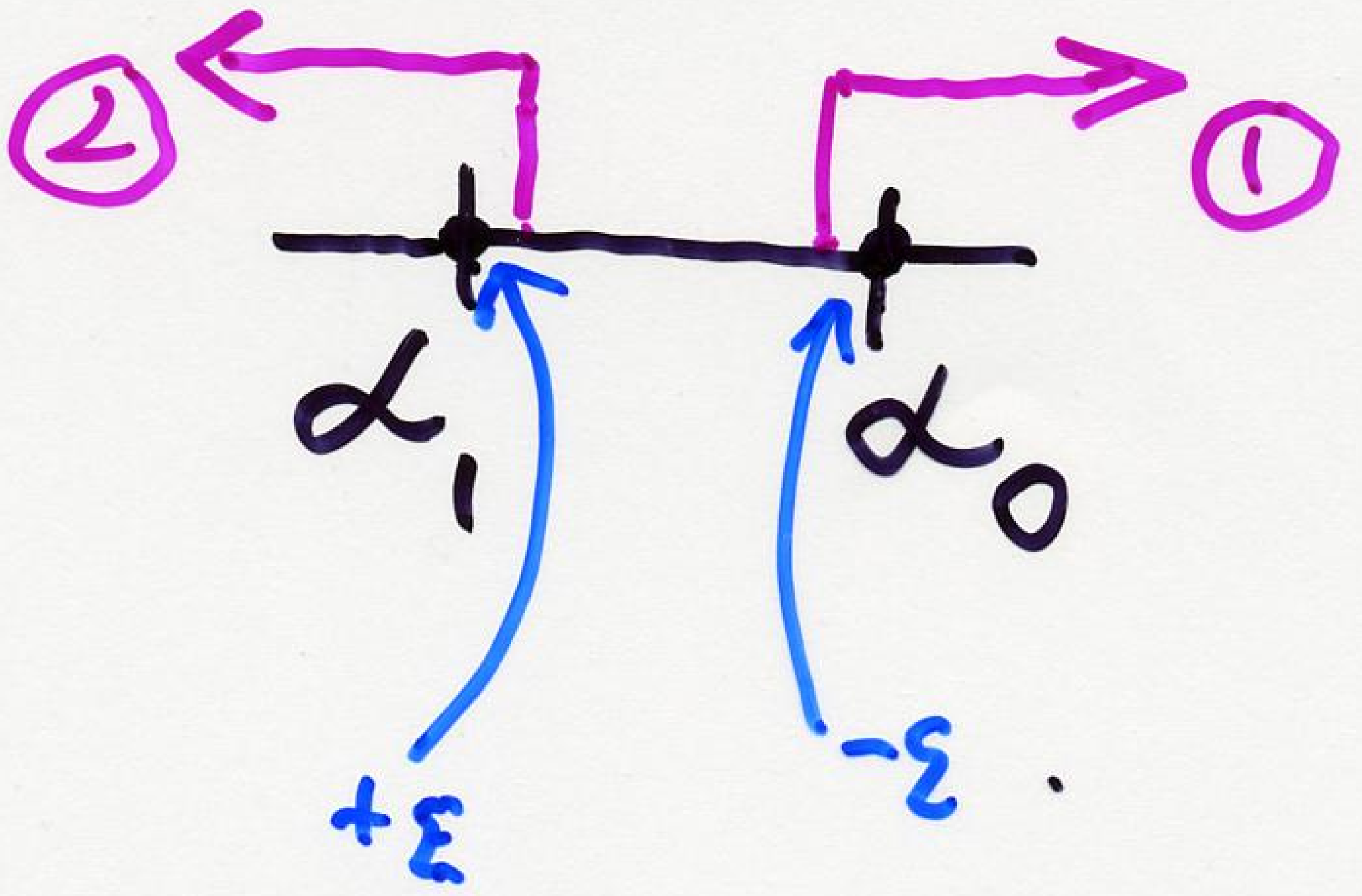
TWO IMPORTANT PARADIGMS DISCOVERED IN THE LAST DECADE:

- CONNECTIONS TO PROBABILISTIC PROOF VERIFICATION THEORY (APPROX. L.B);
- NEW PROBABILISTIC METHODS FOR DESIGNING A.A

(3')

[EXPLICIT PCP-METHOD:]

⇒ HGL-METHOD:



⇒ A.R. = $\frac{\alpha_0}{\alpha_1} - \varepsilon$

$(\forall \varepsilon > 0)$

(1) OR (2) NP-HARD TO DECIDE.

FOR **MAX-CUT**:

$$\frac{\alpha_0}{\alpha_1} = \frac{17}{16}.$$

UNDER THE UNIQUE GAME CONJECTURE:

$$\frac{\alpha_0}{\alpha_1} = 1.1383$$

APPROX. ISSUES OF MAX-CSP.

MAX/MIN r-CSP

EQUIVALENT TO THE FOLLOWING:

GIVEN A SET OF r-ARY BOOLEAN FNCT.'s $f_1, f_2, \ldots, f_m$. CONSTRUCT AN ASSIGNMENT $x \in \{0,1\}^n$ SO AS TO MAXIMIZE/MINIMIZE THE NUMBER OF SATISFIED FNCT.'s f_i.

Two **Notions** of **Approx.'s**:

- **Absolute**:

$$|\mathrm{OPT} - y| \leq r$$

- **Relative**:

$$\mathrm{Max}\left\{ \frac{y}{\mathrm{OPT}}, \frac{\mathrm{OPT}}{y} \right\} \leq r$$

[y = Costs of Solution S.]

AF-CLASS

III

PTAS

AN EPTAS IS CONSTANT TIME AS (CTAS) IF ITS RUNNING TIME IS

$$\underbrace{f(1/\varepsilon)}.$$

ρ-ABSOLUTE PTAS (ρ-APTAS)

CONDITION:

$|\text{OPT} - y| \leq \varepsilon\rho$,

FOR ALL $\varepsilon > 0$.

IF OPT IS OF ORDER ρ ($\text{OPT} = \Theta(\rho)$) ($\rightarrow$ DENSE INSTANCES), EXIST. OF ρ-APTAS $\Rightarrow$ EXIST. OF PTAS)

- DENSE MAX-CSP HAS PTASs [AKK95].

- SUBDENSE MAX-CSP HAS PTASs [FK05]

$\sim\left(\frac{n^k}{\log n}\right)$ - CLAUSES.

- MILDLY-SPARSE MAX-CSP HAS QPTAS

[FK06]

$\tilde{\Theta}\left(\frac{n^k}{\log^{O(1)} n}\right)$-CLAUSES

RUNNING TIME IMPROVEMENTS FOR MAX-CSP:

'95 ----→ 2003.
[AFKK03]

- ABSOLUTE CTASs

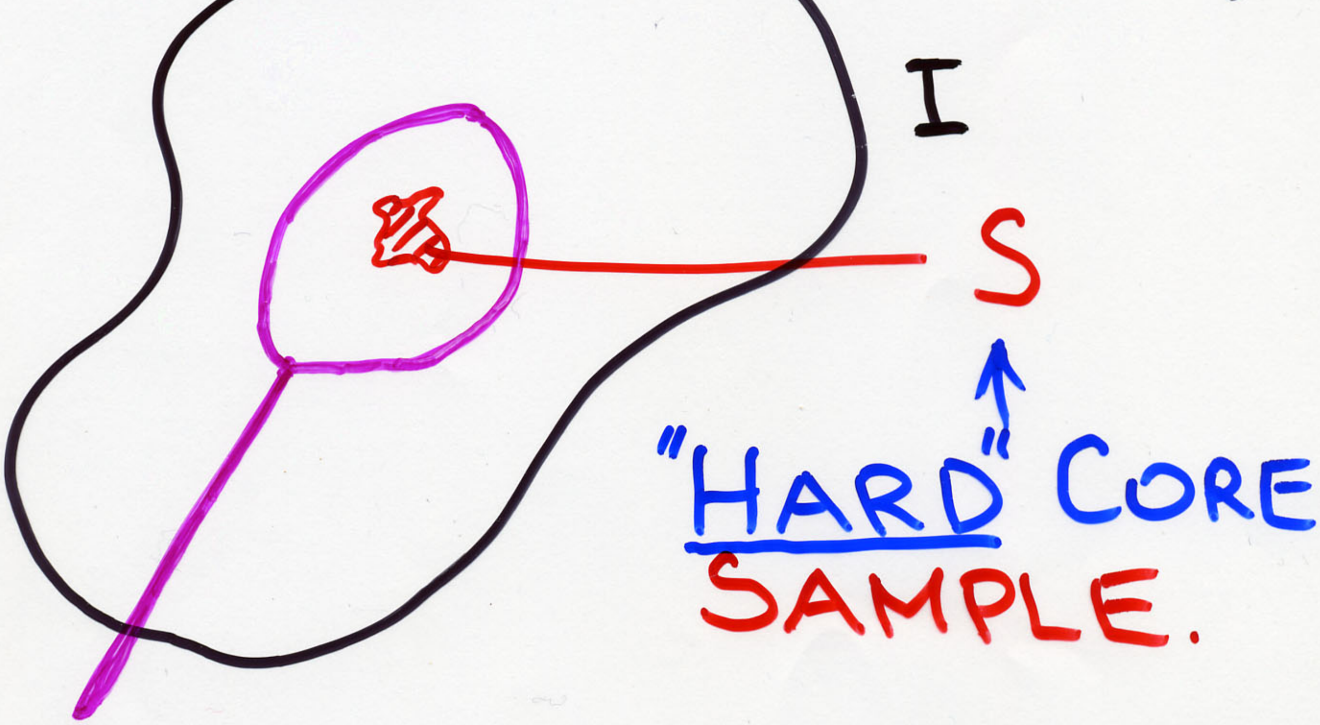
($\tilde{O}(1/\varepsilon^4)$ SAMPLE,
$2^{\tilde{O}(1/\varepsilon^2)}$ TIME)

(THAT AREA BECAME TO BE KNOWN AS THE PROPERTY TESTING)

$(\Downarrow)$

- DENSE INST. OF MAX-CSP HAVE CTASs ($\tilde{O}(1/\varepsilon^4)$ SAMPLE, $2^{\tilde{O}(1/\varepsilon^2)}$ TIME).

OPT. PROBLEM P
(MAX-CUT, MAX-3SAT, ...)



CONSTANT SIZE (VERY TINY) SAMPLE, DOES OPT_{I_S} TELLS YOU "SOMETHING" ABOUT OPT_I ??

<u>OLD METHOD</u> :

GIVEN ANY MAX-rCSP PROBLEM P.

$\Downarrow$

"SIMULATE" IT BY AN SPIP P' OF DEGREE r.

TH. THERE EXISTS AN n^r-APTAS FOR P', $|OPT_{P'} - Y| \le \varepsilon n^r$ FOR ALL $\varepsilon > 0$.

RUNING

TIME:

$n^{O(1/\varepsilon^2)}$.

GETTING DOWN TO CONSTANT TIME.

METHOD:

PROBABILISTIC r-DIMENSIONAL SUB-ARRAYS AND THEIR CUT-NORMS.

$\Downarrow$

CONSTANT TIME APPROX. OF MAX-rCSP

NOTATION:

GIVEN FINITE SETS $V_1, V_2, \dots, V_r$,

AN r-DIMENSIONAL ARRAY A ON $V_1, \dots, V_r$ IS A FUNCTION

$$A : V_1 \times V_2 \times \dots V_r \to \mathbb{R}$$

($A(i_1, i_2, \dots, i_r)$ IS AN ENTRY OF A).

FROBENIUS NORM OF A:

$\|A\|_F$ = SQUARE ROOT OF THE SUM OF SQUARES OF ALL ENTRIES.

<u>LET</u> $S_1 \subseteq V_1$, $S_2 \subseteq V_2$, ..., $S_r \subseteq V_r$, <u>DEFINE</u> THE <u>QANTITY</u>:

$$A(S_1, S_2, \dots, S_r) = \sum_{(i_1, \dots, i_r) \in S_1 \times S_2 \times \dots \times S_r} A(i_1, i_2, \dots, i_r).$$

A <u>CUT-NORM</u> OF $\underline{A}$:

$$\|A\|_C = \max_{S_1 \subseteq V_1, \dots, S_r \subseteq V} |A(S_1, \dots, S_2)|$$

CUT-NORM RELATES FOR $r=2$ - TO GROTHENDIECK NORM, AND SD-PROGRAMMING,

$$\|A\|_G = \underset{\substack{x_i,\, y_j \in \\ \{-1,\, 1\}}}{\mathrm{MAX}} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} x_i y_j \right\}.$$

$$\boxed{\|A\|_C \leqslant \|A\|_G \leqslant 4\|A\|_C}$$

EXAMPLE:

$r=2$, A IS A MATRIX WITH A SET OF ROWS INDEXED BY E' AND A SET OF COLUMNS INDEXED BY V, FOR A GIVEN GRAPH $G=(V,E)$,

$V=\{v_1, v_2, \dots, v_n\}$,

$E=\{e_1, e_2, \dots, e_m\}$, AND

$E'=\{e_1, \dots, e_m, e'_1, \dots, e'_m\}$.

$$A = E'\left\{\overbrace{\begin{bmatrix} & \vdots & \\ \cdots & a_{ki} & \cdots \\ & \vdots & \end{bmatrix}}^{V}\right.$$

<u>ORIENT</u> $G=(V,E)$

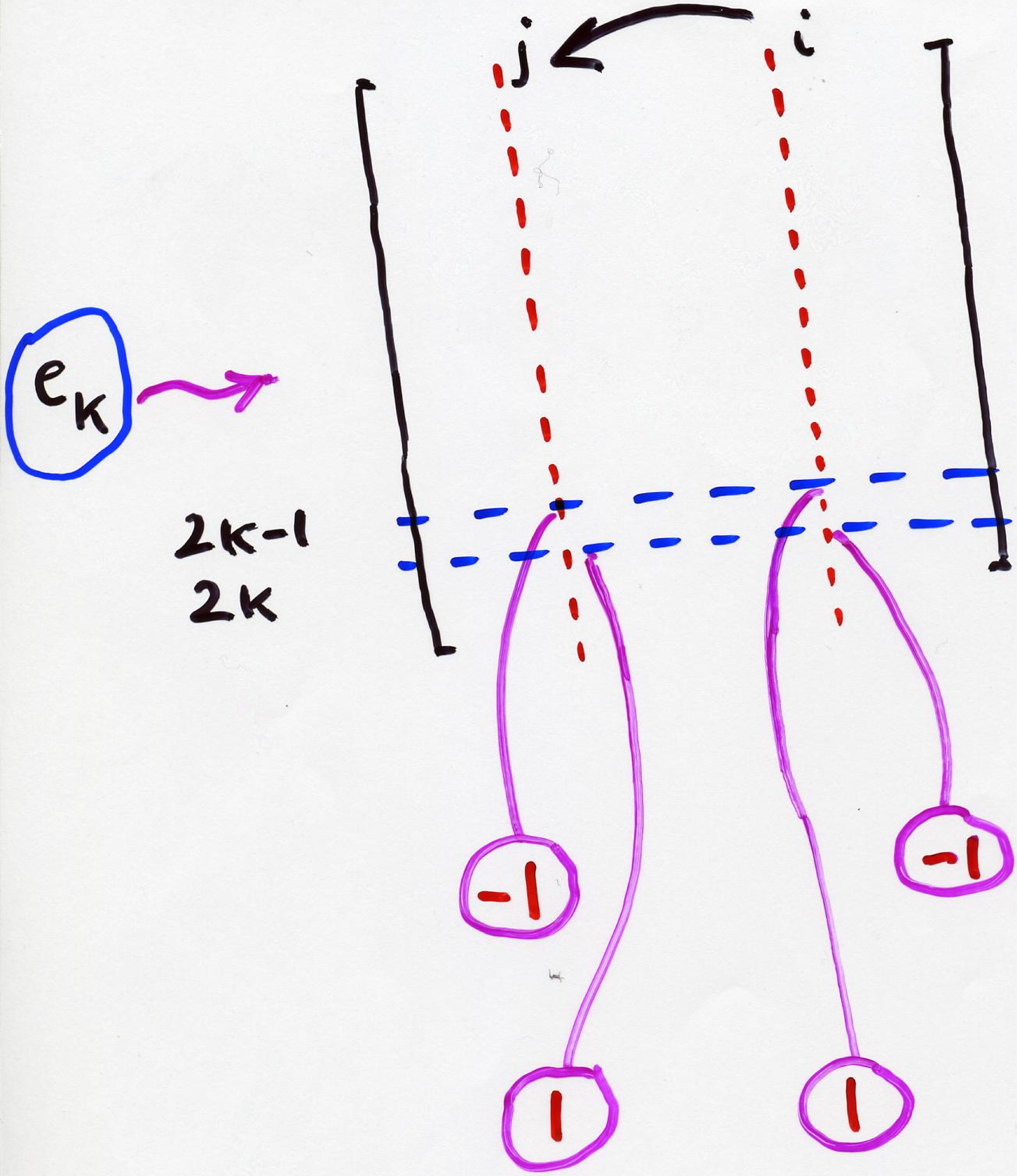
<u>ARBITRARILY</u>, <u>E.G.</u>

$\underset{v_i}{\bullet}\longrightarrow\underset{v_j}{\bullet}$ <u>IFF</u> $i>j$.

<u>LET</u> $e_k = \left(\underset{v_i}{\bullet}\longrightarrow\underset{v_j}{\bullet}\right)$,

<u>SET</u>: $\left[a_{2k-1,i} = a_{2k,j} = 1\right]$;

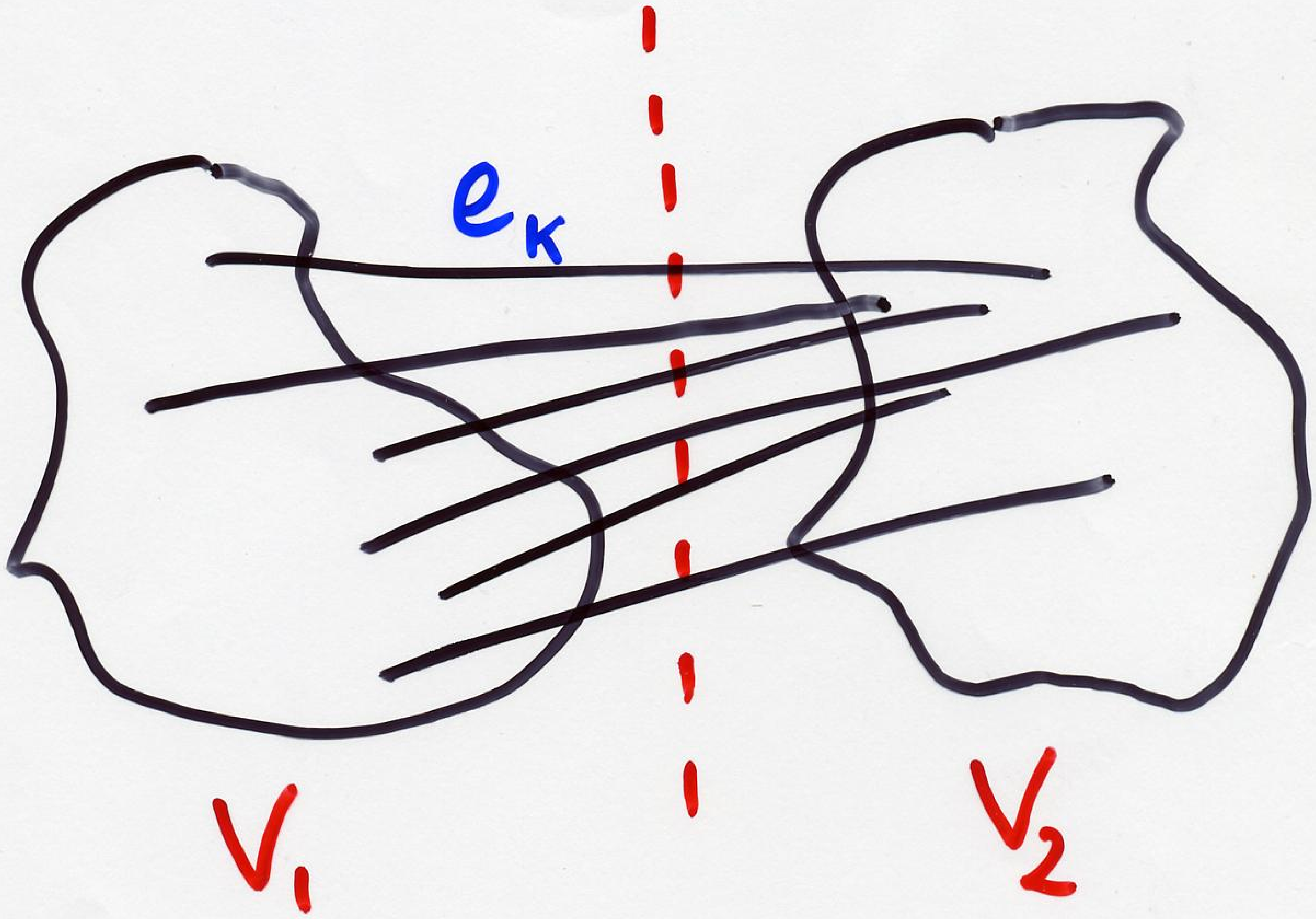
<u>AND</u> $\left[a_{2k-1,j} = a_{2k,i} = -1\right]$.



THE REST OF ENTRIES ARE ALL SET TO 0.

ANY CUT OF

G :



INDUCES A SET OF EDGES e_k IN A CUT, AND THE SET OF 1-ENTRIES IN A. THUS,

$$\|A\|_c = \text{MAX-CUT}(G).$$

$$\boxed{\underline{\text{MAIN RESULT:}}}$$

MAX-rCSP CAN BE REDUCED TO THE PROBLEMS OF MAXIMIZING POLYNOMIALS OF DEGREE r OVER THE BOOLEAN CUBE (AS WE DID USING AN SPIP FOR MAX-CUT PROBLEM), AND COMPUT. OF $\|A\|_c$ FOR r-DIM. ARRAYS A.

APPROXIMATION OF $\|A\|_C$.

ON A RANDOM SUBSET OF SIZE $\Theta(\log(1/\varepsilon)/\varepsilon^4)$ $(= \underbrace{q}_{})$.

ASSUMPTIONS (*) ON A:

$\|A\|_C \leq \varepsilon n^r$, $\|A\|_\infty \leq \frac{1}{\varepsilon} B(r)$,

$\|A\|_F \leq 2^{2^r} n^{r/2}$.

THEN, A RANDOM. INDUCED SUBARRAY H OF A SATISFIES

$$\|H\|_c \leq \underbrace{C(r)\cdot q^r}$$

W.H.P.

THE OTHER DIRECTION IS EASY:

IF $\|A\|_c$ IS HIGH, THEN SO IS $\|H\|_c$.

$\boxed{\text{W.H.P.}:}$

MORE EXPLICITLY:

FOR

$$q \geq 10^6 \, r^{12} \frac{1}{\delta^5 \varepsilon^4} \operatorname{LOG}\left(\frac{4}{\varepsilon^2}\right),$$

$\delta, \varepsilon > 0$,

$$\|H\|_c \leq 2^{2^r + 9} \frac{\varepsilon}{\sqrt{\delta}} \, q^r$$

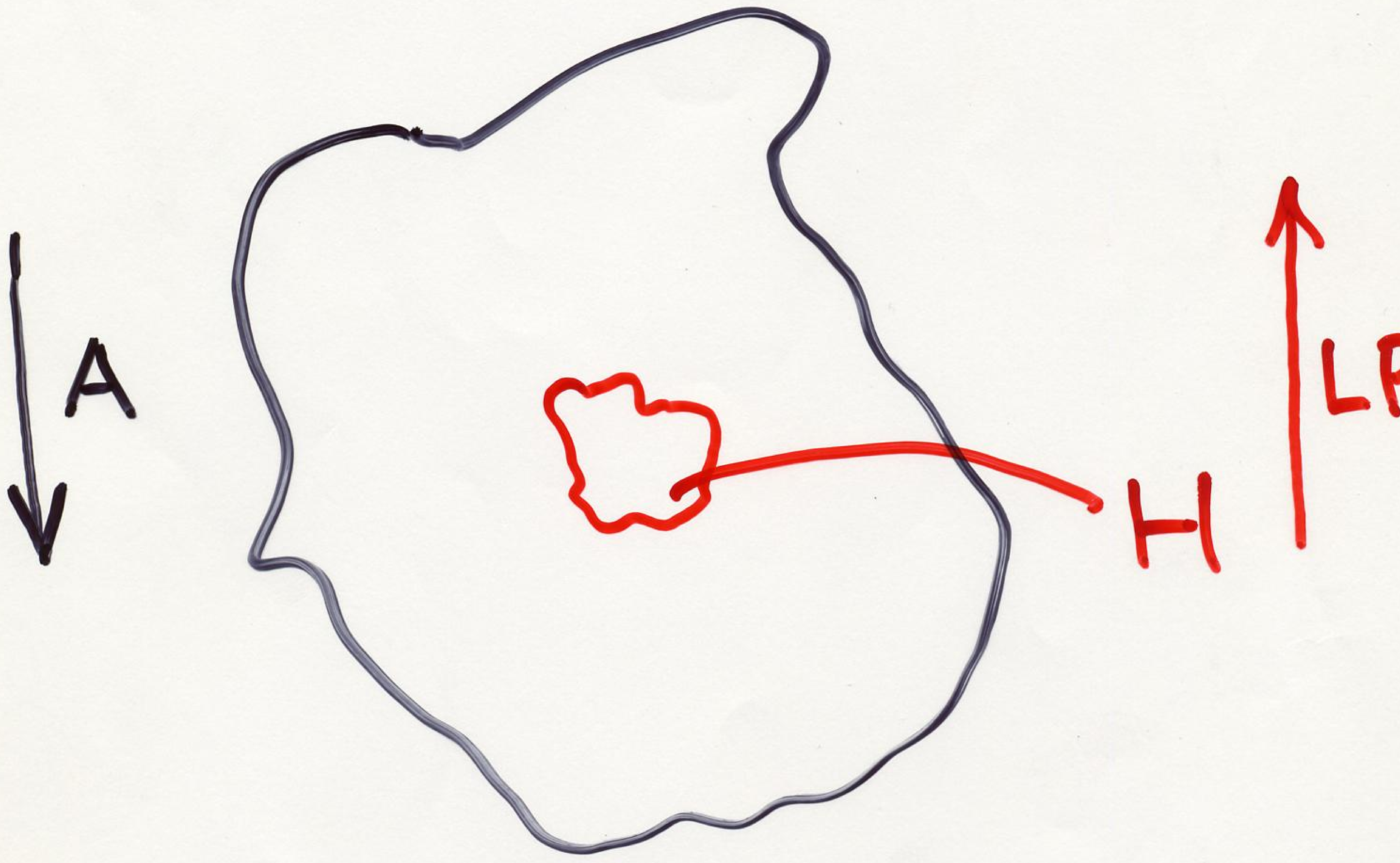
WITH PROB. $\geq 1 - \delta$.

THE COMPUTATION OF $\|H\|_c$ (WE NEED ONLY ABS. APPROX.) ON A SMALL SAMPLE CAN BE DONE BY KNOWN METHOD OF CUT-ARRAY DECOMPOSITION.

RESULTING TIME:

$$2^{O(1/\varepsilon^2)}.$$

A <u>SITUATION</u>:



$$\beta \le \|A\|_c \le \varepsilon n^{\gamma}$$

$$\Downarrow$$

$$\beta' \le \underbrace{\|H\| \le \alpha'}$$

LP

[<u>ILP</u> P <u>GIVES</u>
<u>NORMAL</u>. FACTOR $\frac{n^{\gamma}}{q^{\gamma}}$

LET F BE AN INST. OF MAX-rCSP WITH n VAR.'S. FOR A RANDOM SAMPLE Q OF THE SET OF VAR.'S $\{x_1, \dots, x_n\}$ LET F^Q BE A RANDOM SUB-INST. INDUCED BY Q.

MAIN RESULT:

$$\left| \frac{n^r}{q^r} \mathrm{OPT}_{F^Q} - \mathrm{OPT}_F \right| \le \varepsilon n^r$$

FOR $|Q| = q = O(\log(1/\varepsilon)/\varepsilon^4)$

VERY RECENT IMPROVEMET OF $\boxed{\text{A-PART}}$ TO HARD CORE SIZE $\tilde{\Theta}\left(1/\varepsilon^2\right)$ BY RUDELSON AND VERSHYNIN USING SOME NEW TECHNIQUES OF BOURGAIN AND TZAFRIRI.

("DENSITY"

$\Downarrow$

METRIC SITUATIONS:

PARTITIONING OF FINITE METRIC SPACES (X, d).)

• THERE ARE PTASs FOR GENERAL QUASIMETRIC K-CLUSTERING PROBLEMS.

(BY "NON-HARD-CORE" METHOD)

[FKKR03]

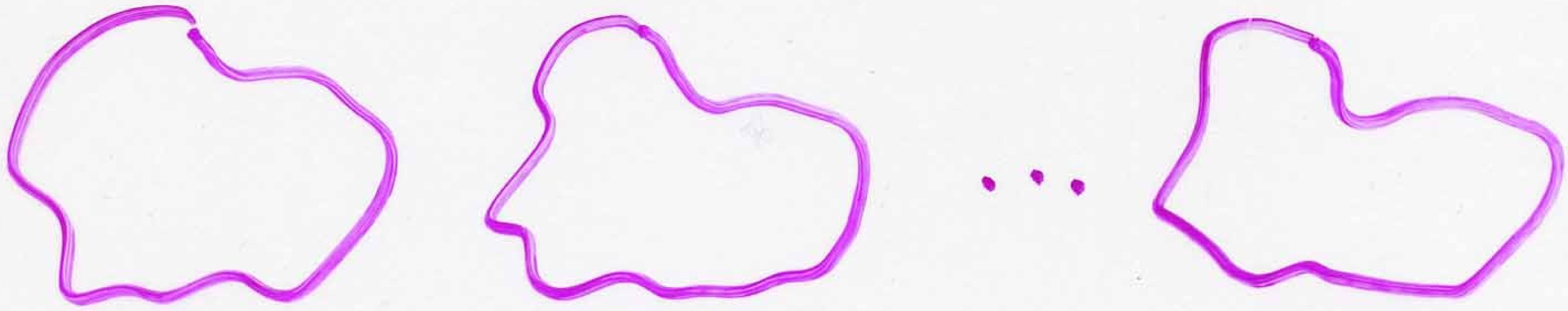
NO GENERAL SUBLINEAR PTASs KNOWN.

3)

$$\sum_{i=1}^{K} \underbrace{\sum_{\{u,v\}\in C_i} d(u,v)}_{\text{INTRA-CLUSTER DISTANCES}}$$

3)

$$\text{MIN} \sum_{i=1}^{K} \underbrace{\sum_{\{u,v\} \in C_i} d(u,v)}_{\text{INTRA-CLUSTER DISTANCES}}$$



C_1 C_2 ... C_K

MIN-SUM CLUSTERING

VERY RECENT GENERAL RESULT:

MAX-kCSP ON INST. WITH $\Theta\left(\frac{n^k}{\Delta}\right)$-CLAUSES HAS A $2^{O(\Delta)}$-TAS.

[FK06].

ONGOING WORK ON SUBDENSE VERSION OF SZEMEREDI'S LEMMA (WITH DENSITY $1/\log n$) AND TESTING HEREDITARY GRAPH PROPERTIES.

(WITH L. LOVÁSZ AND R. KANNAN)

FURTHER RESEARCH:

IMPROVING THE HARD CORE COMPLEXITY OF THE ALGORITHMS.

ARE THERE ANY "MYSTERIOUS" PROB. INTRACTABILITY BARRIERS FOR GETTING DOWN TO, SAY, $\underbrace{\Theta(1/\varepsilon^2)}$ HARD CORE BOUNDS?

SITUATION STILL FAR FROM RESOLVED:

RUNNING TIMES

$2^{O(\sqrt{\Delta})}$ - POSSIBLE?

PRACTICAL ISSUES

(10 BIT VS. 10^4 BIT COMPUTATIONS).

ANY SUBLINEAR HARD CORE PTASs (CTASs WITH METRIC PREPROCESSING) FOR k-CLUSTERING PROBLEMS?

∎