# Sampling Techniques For Approximating CSP Problems.

Marek Karpinski,

University of Bonn.

MAREK KARPINSKI
UNIV. OF BONN

マレク カルピンスキー
ボン大学

## General Framework:

- Problem With Massive Data.

  Too Large To Be Stored In RAM,

  Must Read From External Memory.

- Natural Approach:

  Draw A Small Sample Storable in RAM.

  A Special Approximation Algorithm

  Processes Sample And Yield A Good Estimate

  Of An Answer To The Whole Problem.

OPT. PROBLEM $P$

(MAX-CUT, MAX-3SAT, ...)


CONSTANT SIZE (VERY TINY) SAMPLE, DOES $OPT_{I^S}$ TELLS YOU "SOMETHING" ABOUT $OPT_I$??

BASIC QUESTION
OF PROPERTY TESTING
OR EQUIVALENTLY
ABSOLUTE APPROX.

HISTORICALLY,
THE FIRST RESULTS
WERE DISCOVERED
FOR DENSE GRAPHS
(AT FIRST LACKING
THE HARD CORE
ESTIMATES)

- FIRST-CONNECTION TO
  THE DENSE INSTANCES
  OF OPTIMIZATION PROBLEMS
- SUBDENSE AND NONDENSE INSTANCES
- SAMPLE SIZES
- METRIC & QUASIMETRIC INSTANCES

PART I:

FIRST-TRY:

DENSE MAX-CSP INSTANCES AND

SMOOTH INTEGER PROGRAMS

PART II:

MIN-CSP INSTANCES

PART III:

APPLICATIONS:

METRIC AND

QUASI-METRIC PROBLEMS

# 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 $f_i$.

# PRELIMINARIES:

## Coping With Computational Hardness

- New Approx. Methods (U.B.'s.)
- New PCP-Techn. (L.B.'s.)

DENSE <u>INSTANCES</u>

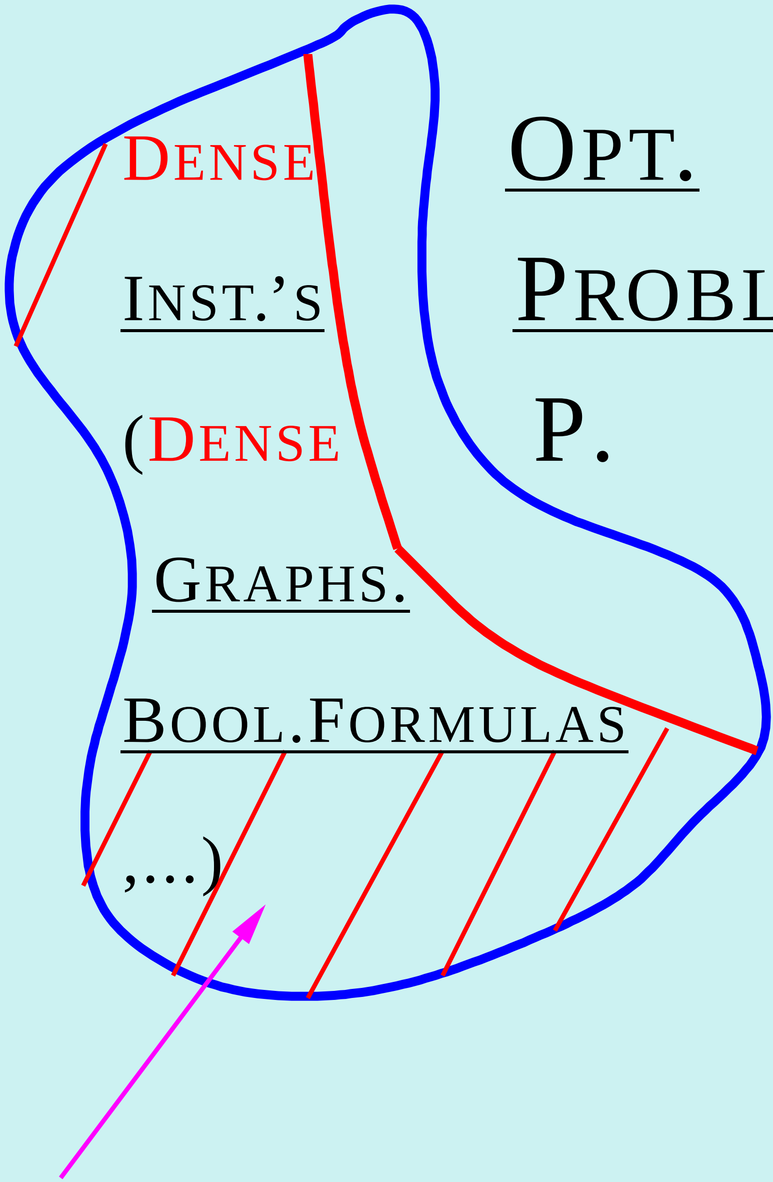
<u>ARE</u> IN MANY

CASES -VERY-

“NATURAL” <u>INSTANCES</u>.

$\Rightarrow$ ( DENSITY IN <u>MANY</u> CASES-

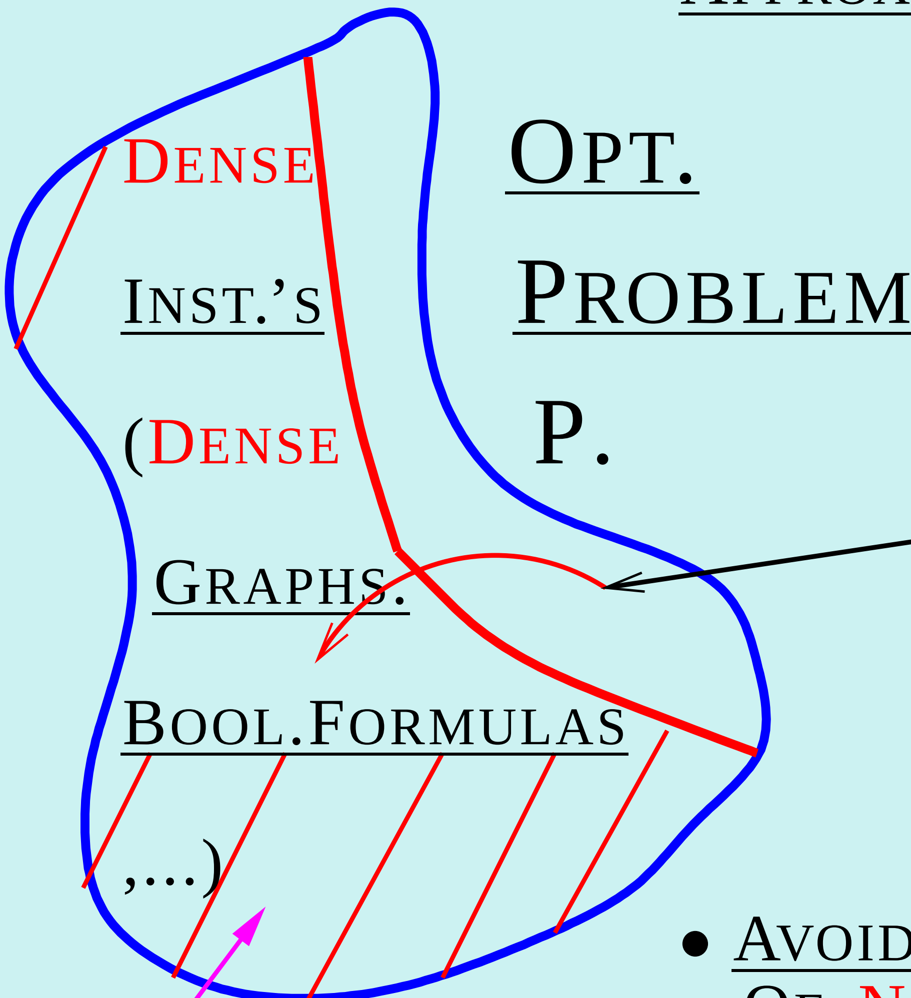
A SENSIBLE PARAMETER )




OPT.
PROBLEM
P.

'ALMOST ALL' GRAPHS, k-SAT FORMULAS,...

APPROX.HARDNESS?
(DENSE PCP??)




CLASSIC.NP-H.

- AVOIDING PATHOLOGIES OF NON-DENSE HARDNESS.

'ALMOST ALL' GRAPHS, k-SAT FORMULAS, ...

DENSITY

IN OPT. :

- $\Omega(n^2)$ EDGES
- $\Omega(n^k)$ CLAUSES

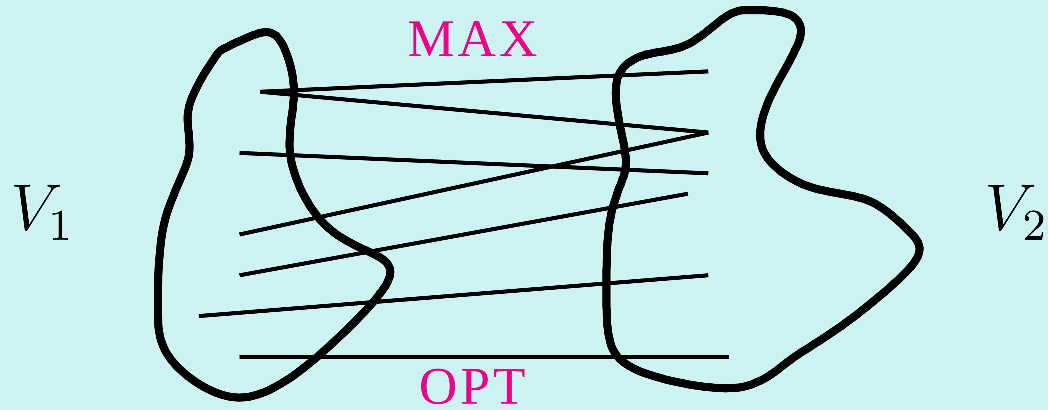
IN k-SAT FORM.'S

$\vdots$

## ⇒EXAMPLE:

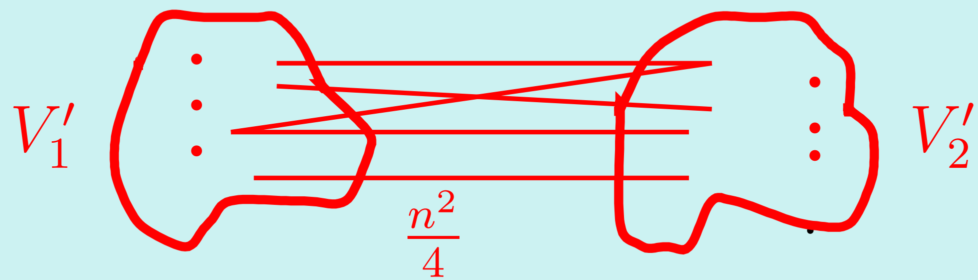
### MAX-CUT.

G::




$\bigoplus$

$K_n$




MAX-CUT (G) = OPT

MAX-CUT $(G \oplus K_n) =$ OPT $+ \frac{n^2}{4}$

$G \oplus K_n$ HAS

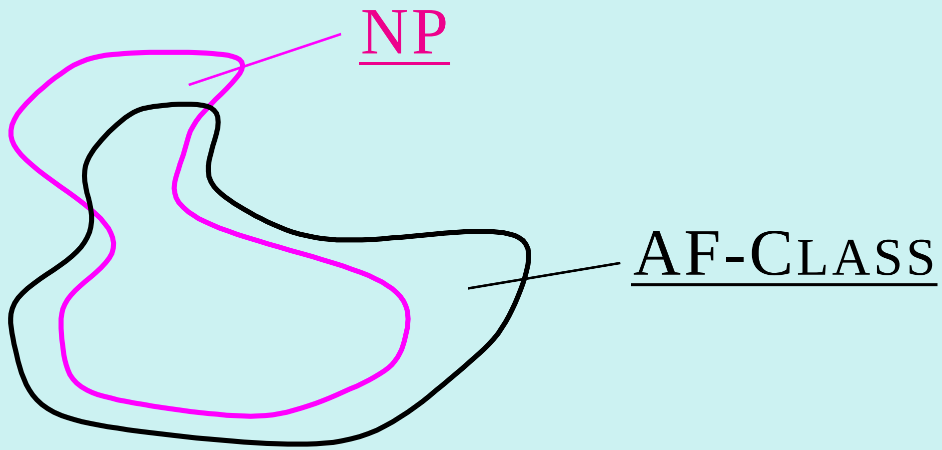
$\Omega(n^2)$ EDGES

$\Rightarrow$ BEYOND DENSITY.

$\Rightarrow$ "METRIC / GEOMETRIC SITUATIONS"

$\Rightarrow$ [ APPROXIMATION FEASIBILITY ]

VS. NP.




APPROX. ANALOG TO

THE NP-CLASS: MAX-SNP

# TWO NOTIONS OF APPROX.'S:

- **ABSOLUTE:**

$$|\underline{OPT} - Y| \leq \underline{r}$$

- **RELATIVE:**

$$\mathrm{MAX}\left\{ \frac{Y}{\underline{OPT}}, \frac{\underline{OPT}}{Y} \right\} \leq \underline{r}$$

[Y = COSTS OF SOLUTION S.]

## ⇛ PTAS:

- $\mathcal{A}$ IS CALLED A PTAS FOR $\pi$ IF FOR EVERY FIXED $\varepsilon > 0$, $\mathcal{A}$ IS A POLY TIME ALGORITHM IN THE SIZE OF I (AN INSTANCE OF $\pi$) WITH APPROX. RATIO $1+\varepsilon$ (MEANING $\mathcal{A}$ OUTPUTS A SOLUTION S TO I S.T. $\max\left\{\frac{S}{OPT(I)}, \frac{OPT(I)}{S}\right\} \leq 1+\varepsilon$)

AF-CLASS

$\Downarrow$

EPTAS

- $\mathcal{A}$ IS CALLED AN EFFICIENT PTAS (EPTAS) IF ITS RUNNING TIME IS

$$\underbrace{f(1/\varepsilon)}_{} \cdot n^{O(1)}.$$

[ OUGHT TO BE "SMALL" GROWTH FUNCTION. ]

An EPTAS is Constant Time AS (CTAS) if its running time is

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

## $\rho$-ABSOLUTE PTAS ($\rho$-APTAS)

CONDITION:

$$|OPT - y| \leq \varepsilon \rho,$$

FOR ALL $\varepsilon > 0$.

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

- DENSE MAX-CSP

HAS PTASs

[AKK95].

REF. [2]

- SUBDENSE MAX-CSP

HAS PTASs

[FK05]

REF. [4]

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

- APPROXIMATION HARDNESS

  ( WITH RESPECT TO APPROX. RATIO $r$ ):

  ACHIEVING A.R. $r$

  IN (RANDOMIZED) POLY TIME

  $\Rightarrow$ $\underbrace{\mathrm{NP} = \mathrm{P}}$ ( NP = RP, NP = coRP, . . . )

- APPROXIMATION HARDNESS

  ( WITH RESPECT TO APPROX. RATIO $r$ ):

  ACHIEVING A.R. $r$

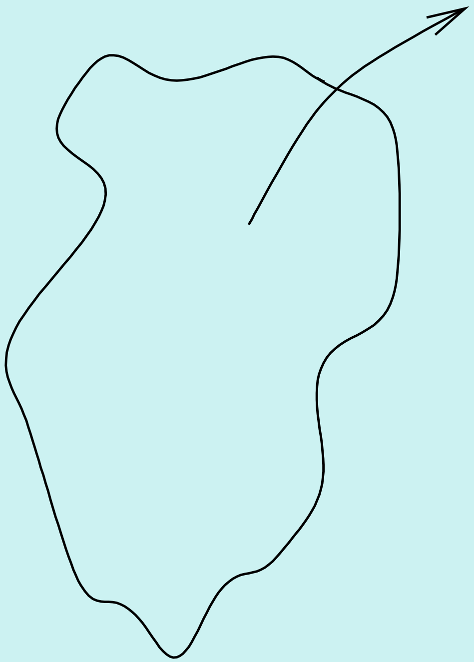
  IN (RANDOMIZED) POLY TIME

  $\Rightarrow \underbrace{\text{NP} = \text{P}}$ ( NP = RP, NP = coRP, . . . )

  ( PCP-TH. $\Rightarrow \forall$ X MAX-SNP-HARD

  $\exists\, \varepsilon > 0$ [X IS APPROX. HARD WITHIN $1+\varepsilon$].)

# "WORLDS" OF MAX-OPT. APPROXIMATION:

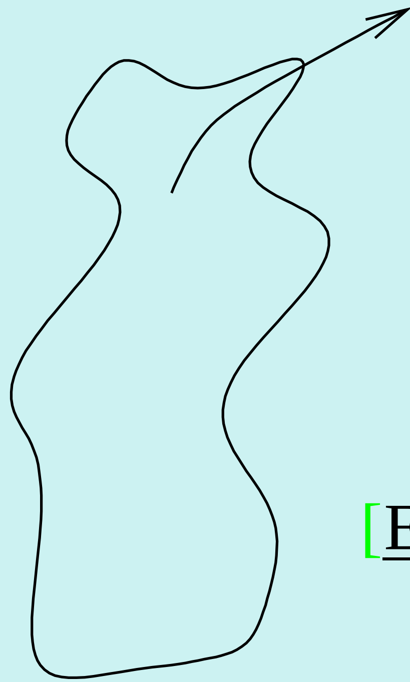


PTAS

(SPIPs

PROGR.,

..., '95)



NO PTAS

(MAX-SNP-

HARD FOR

$b = 3$,

'91, '92)

[EXPL. CONST.]

[ • DENSE ]

MIN. DEG.

$= \Theta(n)$ ($\Theta(n^2)$-EDGES)

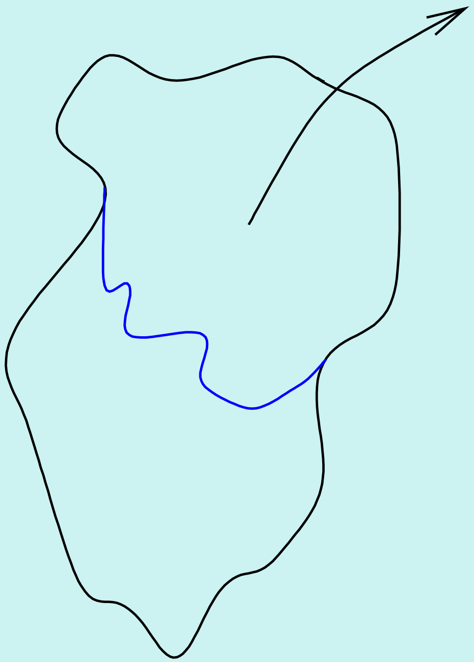
[ • SPARSE ]

MAX. DEG

$\leq b$ (CONST.)

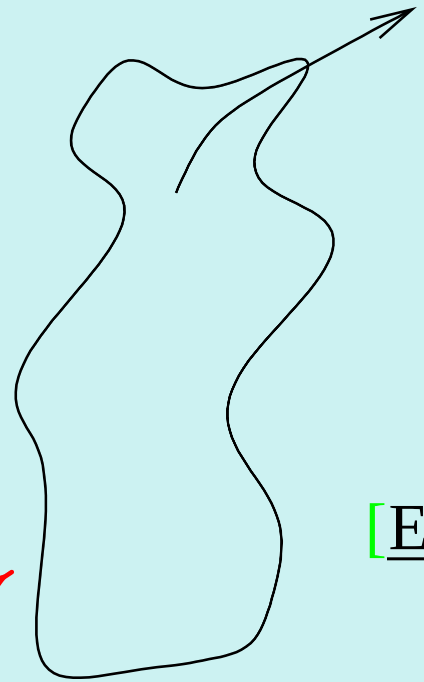
'95 $\Rightarrow$ 2000/'01 MIN

# "WORLDS" OF ~~MAX~~-OPT. APPROXIMATION:



PTAS

(SPIPs PROGR., ..., '95)



NO PTAS

(MAX-SNP-HARD FOR $b = 3$, '91, '92)

[EXPL. CONST.]

[ • DENSE ]

MIN. DEG.

$= \Theta(n)$ ( ~~$\Theta(n^2)$-EDGES~~ )

$\boxed{\Theta(n^2)\text{-EDGES}}$

[ • SPARSE ]

MAX. DEG

$\leq b$ (CONST.)

- Dense Instances:

  $\alpha$-Dense Graphs

  ( Min Degree Of $G = (V, E)$ is $\geq \alpha|V|$ )

- A Graph Is Dense

  If Its Min Degree Is $\Theta(n)$

( A Graph Is Dense In Average

If It Has $\Theta(n^2)$ Edges )

## ⇒ $\boxed{\text{Optimization Classes:}}$

$\boxed{\text{0-Class:}}$

- Longest Path
- TSP

$\boxed{\text{1st Class:}}$ (Local Constr.)

- Max-CUT
- Bisection

$\boxed{\text{2nd Class:}}$

- Set Cover
- Steiner Tree

$\boxed{\text{3rd Class:}}$

- Bandwidth

REF. [7]

# A MAP OF "DENSE" (ABSOLUTE) APPROXIMATIONS:

| PROBLEM | APPROX. RATIO | APPROX. HARDNESS | REF. |
|---|---|---|---|
| DENSE MAX-SNP | PTAS | — | [AKK95] |
| DENSE MAX-CUT | PTAS | — | [AKK95],[FV96] |
| DENSE MAX-DCUT | PTAS | — | [AKK95] |
| DENSE MAX-HYPERCUT(d) | PTAS | — | [AKK95] |
| DENSE DENSE-K-SUBGRAPH | PTAS | — | [AKK95] |
| EVERYWHERE DENSE SEPERATOR | PTAS | — | [AKK95] |
| EVERYWHERE DENSE BISECTION | PTAS | — | [AKK95] |
| EVERYWHERE DENSE MIN-K-CUT | PTAS | — | [AKK95] |

| Problem | Approx. Ratio | Approx. Hardness | Ref. |
|---|---|---|---|
| DENSE MIN-LINEAR-ARRANGEMENT | PTAS | — | [AFK96] |
| DENSE d-DIMENSIONAL-ARRANGEMENT | PTAS | — | [AFK96] |
| DENSE MIN-CUT-LINEAR-ARRANGEMENT | PTAS | — | [AFK96] |
| DENSE SET COVER | $\bigcap_c c \cdot \ln n$ | OPEN | [KZ97b] |
| DENSE STEINER TREE | PTAS | — | [KZ97b] |
| DENSE VERTEX COVER | $\frac{2}{2-\sqrt{1-\varepsilon}}$ | MAX-SNP-hard | [KZ97b] |
| EVERYWHERE DENSE VERTEX COVER | $\frac{2}{1+\varepsilon}$ | MAX-SNP-hard | [KZ97b] |
| EVERYWHERE DENSE BANDWIDTH | 3 | OPEN | [KWZ97] |
| EVERYWHERE DENSE DBANDWIDTH | 2 | OPEN | [KWZ97] |

( $\Rightarrow$ $\boxed{\text{4TH CLASS:}}$

(METRIC PROBLEMS)

- METRIC MAX-CUT )

---

$\Rightarrow$ $\boxed{\text{4}^{\star}\text{TH CLASS:}}$

(GEOMETRIC AND METRIC GRAPH PROBLEMS)

- PARTITIONING PROBLEMS

■

( $\Rightarrow$'0'-CLASS

DENSE 'METRIC' TSP,

LONGEST PATH. )

QUALITATIVE RESULTS CONNECTED

TO BOTH

DENSE AND VERY

SPARSE SITUATIONS (!)

# NP-Hardness And Density

$\alpha$-Dense Graphs
( Min Degree Of $G = (V, E)$ Is $\geq \alpha|V|$ )

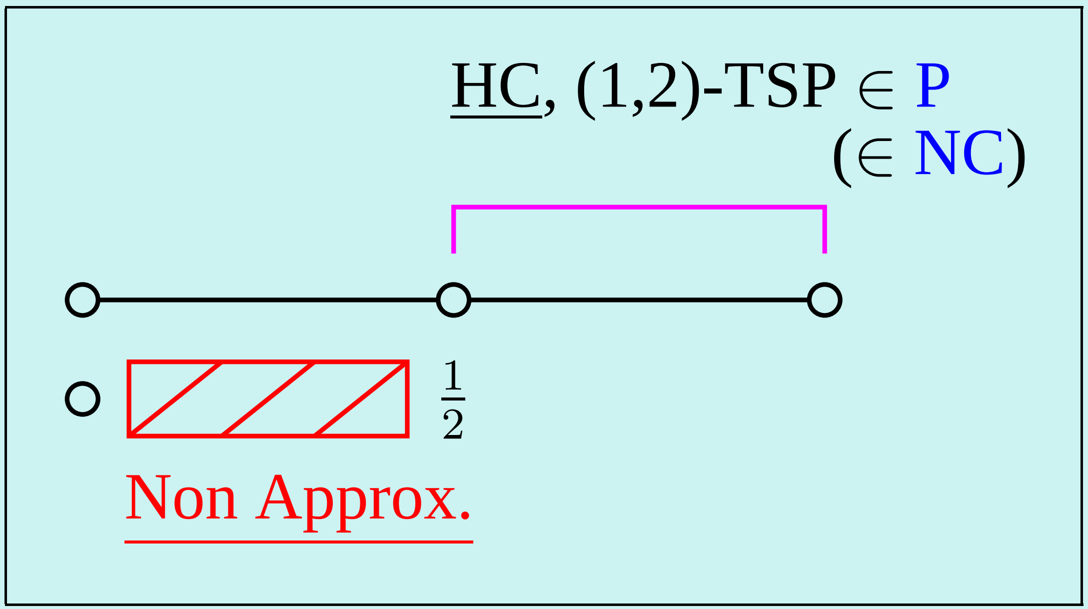
- HC ( Hamiltonian Cycle Problem )
  Is In P For $\alpha \geq 1/2$,
- HC Is NP-Hard For $\alpha < 1/2$.

- LONGEST-PATH IS APPROX. HARD

  FOR $\alpha < 1/2$ ([FVK98]).

- (1,2)-TSP IS APPROX. HARD

  FOR $\alpha < 1/2$ ([FVK98]).



$\Rightarrow$ ANOTHER 'METRIC' PROBLEM

DENSE (1,2)-TSP

$\uparrow$

SUBGRAPH SPANNED BY THE EDGES OF

LENGTH 1 IS DENSE

THEOREM ([FVK98],[FS98]).

FOR EVERY $0 < d < 1/2$,

(1,2)-TSP$\restriction$INST. OF MIN. DEG. $\geq dn$

HAS A PTAS $\Rightarrow$ (1,2)-TSP HAS A PTAS.

- Approx. Hardness Of (1,2)-TSP:

  $r^* \cong 1 + 1.34 \cdot 10^{-3} \quad (743/742)$ [EK00]

  $\Downarrow$

  1/4-Dense-TSP:

  $r^* = 1 + 1.07 \cdot 10^{-3}$ [FVK98]

  ( Best Upper Bound:

  $1.1429 \quad (8/7)$, [BK06]. )

## 3-Occ-E3-Lin2

Lower B.:

1.0163 [BK99]

$$r_d^* = 1 + \varepsilon^* \cdot \frac{1}{1+d}$$

FOR

$$r^* = 1 + \varepsilon^* .$$

# Classes Of Dense Opt. Problems

⇒ 1st Class (Local-Constraint)

- Max-Cut
- Bisection
- Max-SNP

⇒ 2nd Class (Covering Problems)

- Set Cover
- Steiner Tree
- Vertex Cover

⇒ 3rd Class (Bandwidth Problems)

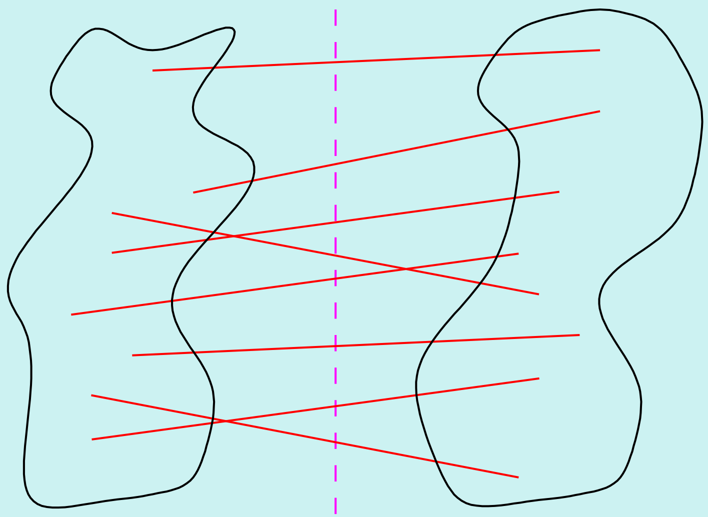
- Bandwidth
- DBandwidth

$\Rightarrow$ $\boxed{\underline{\text{1ST CLASS}}}$

( [AKK95],

[FV96]. )

- MAX-CUT




---

- BISECTION

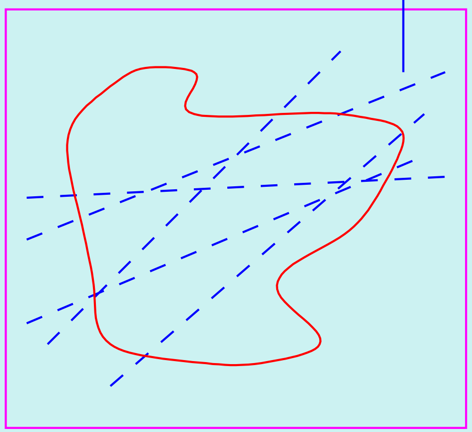
MIN IN "50/50"-CUT

(EQUI-CUT) $|V_1| = |V_2|$

("STATUS" WIDE OPEN!)

# [“DENSE”-IDEA::]

SAMPLE CUTS



EXHAUSTIVE SAMPLING

$O(\log n)$-SAMPLE

# PLACEMENT METHOD BY EXHAUSTIVE SAMPLING:

- TAKE A SAMPLE OF $O(\log n)$ VERTICES.
- EXHAUSTIVELY TRY ALL POSSIBLE PLACEMENTS (IN L, R), $2^{O(\log n)}$-MANY DECIDE WHERE EACH VERTEX OF THE SAMPLE BELONGS IN OPT CUT.

Density $\Rightarrow$ With High Probability ($1 - n^{-\beta}$, $\beta$ Const.) Some Of The Neighbors Of $V$ Where Sampled.

| Easy | Decisions |
| --- | --- |
| If | $\lvert\Gamma_{\mathrm{L}}(V)\rvert \ll \lvert\Gamma_{\mathrm{R}}(V)\rvert$ |
| Or | $\lvert\Gamma_{\mathrm{L}}(V)\rvert \gg \lvert\Gamma_{\mathrm{R}}(V)\rvert$ |

"Difficulty"

If

$$\boxed{\lvert\Gamma_{\mathrm{L}}(V)\rvert \approx \lvert\Gamma_{\mathrm{R}}(V)\rvert}$$

$( \Rightarrow$ SPIPs $)$

Smooth-

-Programming

$\oplus$ Global

Method

1

# SPIP-

# PROGRAMS.

1

# SPIP-PROGRAMS.

SMOOTH POLYNOMIAL INTEGER PROGRAMS.

# SPIPs.

LET $p \in \mathbb{Z}[x_1, x_2, \ldots, x_n]$ BE A DEGREE $k$ POLYNOMIAL WITH INTEGER COEFF.'S.

LET

$$\boxed{p = \sum_{i=1}^{m} \underbrace{a_i M_i}}$$

COEFF. MONOMIAL.

P IS CALLED c-SMOOTH, FOR A CONSTANT $c > 0$, IFF

$$|a_i| \leq c \cdot n^{k - \mathrm{DEG}(M_i)}$$

FOR ALL $1 \leq i \leq n$.

DEF. A POLYNOMIAL INTEGER PROGRAM OF DEGREE $K$ IS A SPIP (OF DEGREE $K$) IF BOTH "OBJECTIVE" AND "CONSTRAINT" POLYNOMIALS ARE $C$-SMOOTH FOR SOME CONSTANT $C>0$.

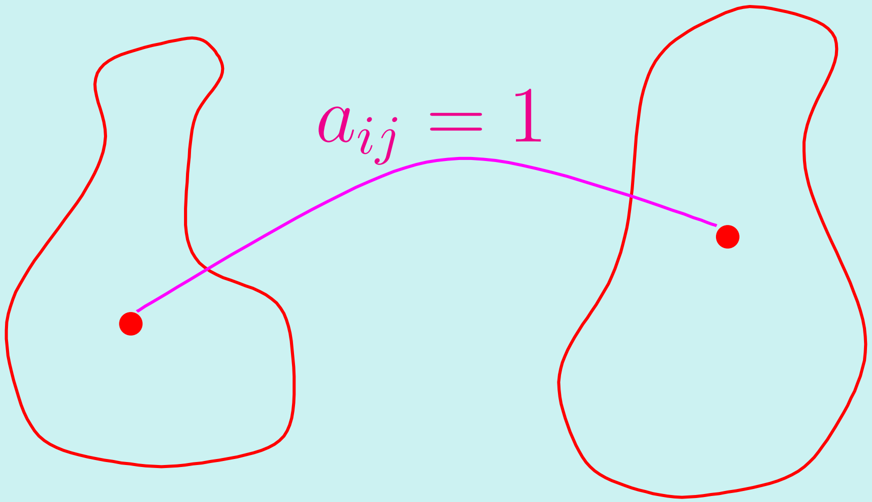
$$\begin{cases} \underset{x}{\text{MAX}} / \underset{x}{\text{MIN}}\ p(x), \\ p \in \mathbb{Z}[x_1, \ldots, x_n], \\ c\text{-}\underline{\text{SMOOTH}}, \\ \underline{\text{DEG}}(p) \leq k, \\ x_i \in \{0,1\}; \\ q_j(x) \geq 0 \\ \underline{\text{FOR}}\ q_j \in \mathbb{Z}[x_1, \ldots, x_n] \\ c\text{-}\underline{\text{SMOOTH}}, \text{ AND} \\ \text{OF } \underline{\text{DEG}}. \leq k. \end{cases}$$

EXAMPLE:

# [ “SMOOTH” PROGRAMS ]

( MAX-CUT:

$L$




$R$

$\boxed{x_i = 0}$ $\boxed{x_j = 1}$

$(V = L \cup R)$ )

$\Rightarrow$ WRITE A

QUADRATIC INT. PROGRAM:

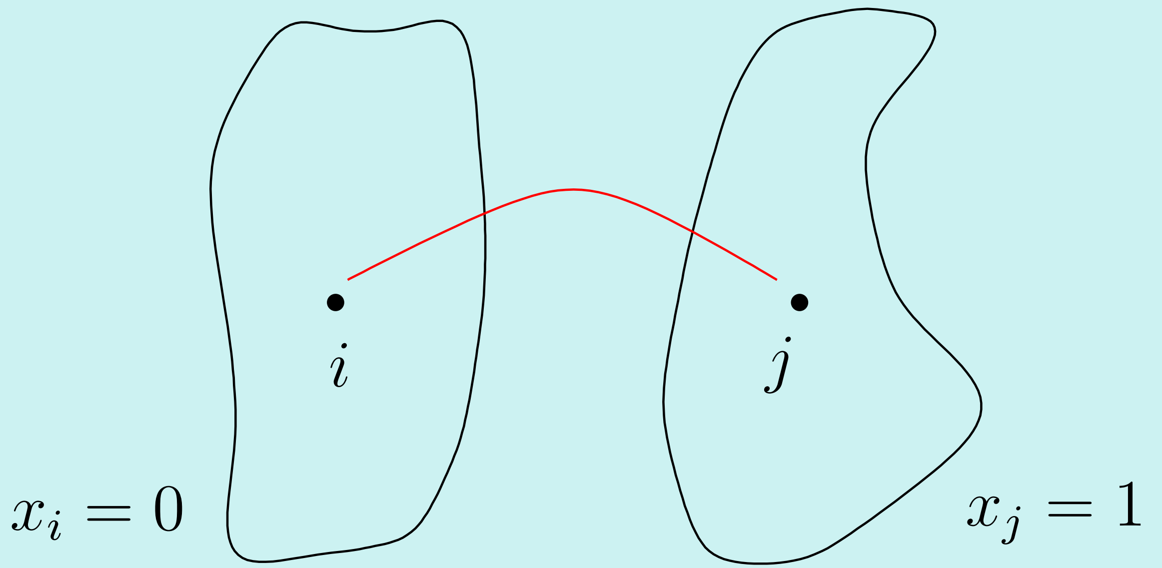
P::

$$\boxed{\max_{x_i} \left\{ \sum_{i<j} (a_{ij} \cdot (x_i(1-x_j) + x_j(1-x_i))) \right\}}$$

- $\boxed{x_i \in \{0, 1\}}$

( $A = [a_{ij}]$ IS THE ADJ. MATRIX OF $G$ )

$$\boxed{\underline{\text{MAX-CUT}}(G) = \underline{OPT}_{\mathrm{P}}}$$




$$s_{ij} \quad = \quad x_i(1 - x_j) + x_j(1 - x_i)$$

$$\boxed{s_{ij} \quad = \quad 1 \Longleftrightarrow x_i \neq x_j \ , \quad (x_i, x_j \in \{0, 1\})}$$

$$(s_{ij} \quad = \quad x_i + x_j - 2x_i x_j)$$

$$
\begin{aligned}
Q &= \sum_{i<j} a_{ij} s_{ij} \\
&= \sum_{i<j} -2a_{ij} x_i x_j + \sum_i b_i x_i
\end{aligned}
$$

---

- <u>Coefficients</u> Of <u>Deg. 2</u>
  Monomials $\in \{0, -2\}$
- <u>Coefficients</u> Of <u>Deg. 1</u>
  Monomials $= O(n)$

DEF.

A POLYNOMIAL $P \in \mathbb{Q}[x_1, \ldots, x_n]$ OF

DEG. $d$ IS $c$-SMOOTH ($c$-CONST)

IF

$$\boxed{\begin{array}{l} P = \sum_{i=1}^{m} a_i M_i \\ \text{AND} \\ |a_i| \leq c \cdot n^{d - \mathrm{DEG}(M_i)} \end{array}}$$

FOR ALL $1 \leq i \leq m$.

$$\Rightarrow Q = \sum_{i<j} a_{ij} s_{ij}$$

$\boxed{\text{FOR \underline{MAX-CUT} IS \underline{2-SMOOTH}.}}$

( • $|-2| \leq 2n^0$ FOR DEG $= 2$ ;

• $|cn| \leq 2n^{2-1}$ FOR DEG $= 1$ $(c \leq 1)$ ).

DEF.

A POLYNOMIAL INTEGER PROGRAM (PIP) OF DEG. $d$ IS $c$-SMOOTH IF ITS OBJECTIVE FNCT. $P$ IS A $c$-SMOOTH POLYNOMIAL OF DEG. $d$ .

ABS. APPROX. OF SPIPS NEEDED FOR BOTH MAX- & MIN-CSP.

SPIP-Theorem

(Arora, Karger, Karpinski '95)

Let $P$ Be An SPIP Of Deg. $d$.

Let OPT Be The Opt-Value Of $P$.

Then For Every $\varepsilon > 0$ There Is A Poly Time Approx. Alg. Producing An Assignment $s \in \{0,1\}^n$,

S.T.

$$\boxed{P(s) \geq \text{OPT} - \varepsilon n^d}$$

- Running Time: $n^{O(1/\varepsilon^2)}$ (As Of '95)

$\boxed{\underline{\mathrm{II}}.}$

- $\underline{\text{RUNING}}$ $\underline{\text{TIME}}$: $\tilde{O}(1/\varepsilon^4)$, [AFKK'03], $\underline{\text{REF}}$ [1].

## TI'.

IMPROVEMENT OF ABSOLUTE APPROX. RATIO TO $\boxed{\varepsilon n^d / \log n}$ :

$$|P(s) - OPT| \leq \underbrace{\varepsilon n^d / \log n}$$

(DE LA VEGA, KARPINSKI, 2006)

→ cf. REF. [4].

WORKS FOR

BOTH MAX- & MIN-OBJ.

$\Downarrow$

YIELDS PTASs FOF “LARGE”-VALUES

OF OBJ. FNCTS. FOR ’MAX/MIN’.

$\Downarrow$

DENSE MAX-SNP

DOES HAVE A PTAS.

SO, ’CONSTRUCTIVELY’

$\Downarrow$

$$\Downarrow$$

DENSE

MAX-CUT $\in$ PTAS

---

■ (1ST CLASS).

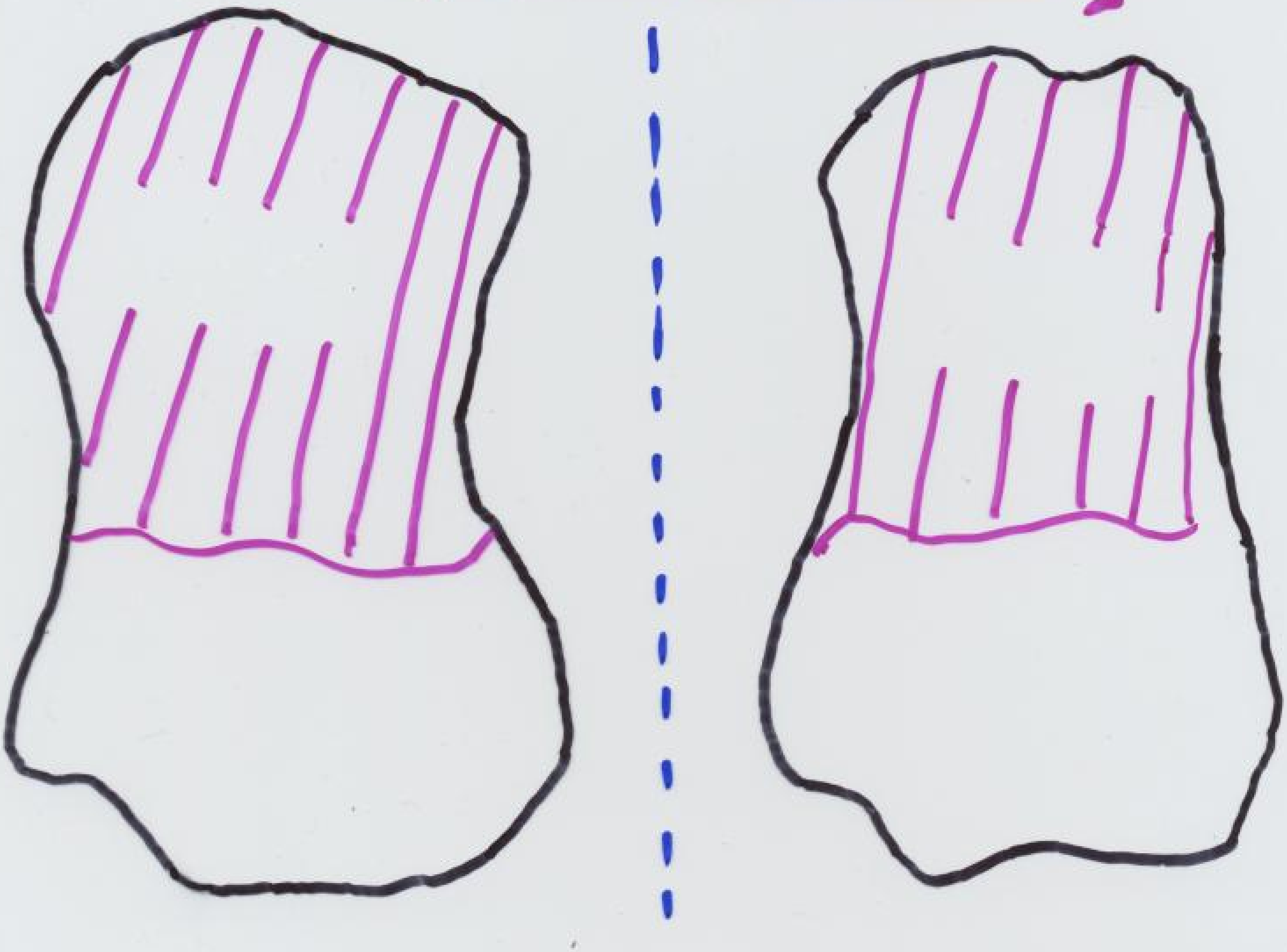
# PART II.

## MIN-OPTIMIZATION

## (MIN-CSP).

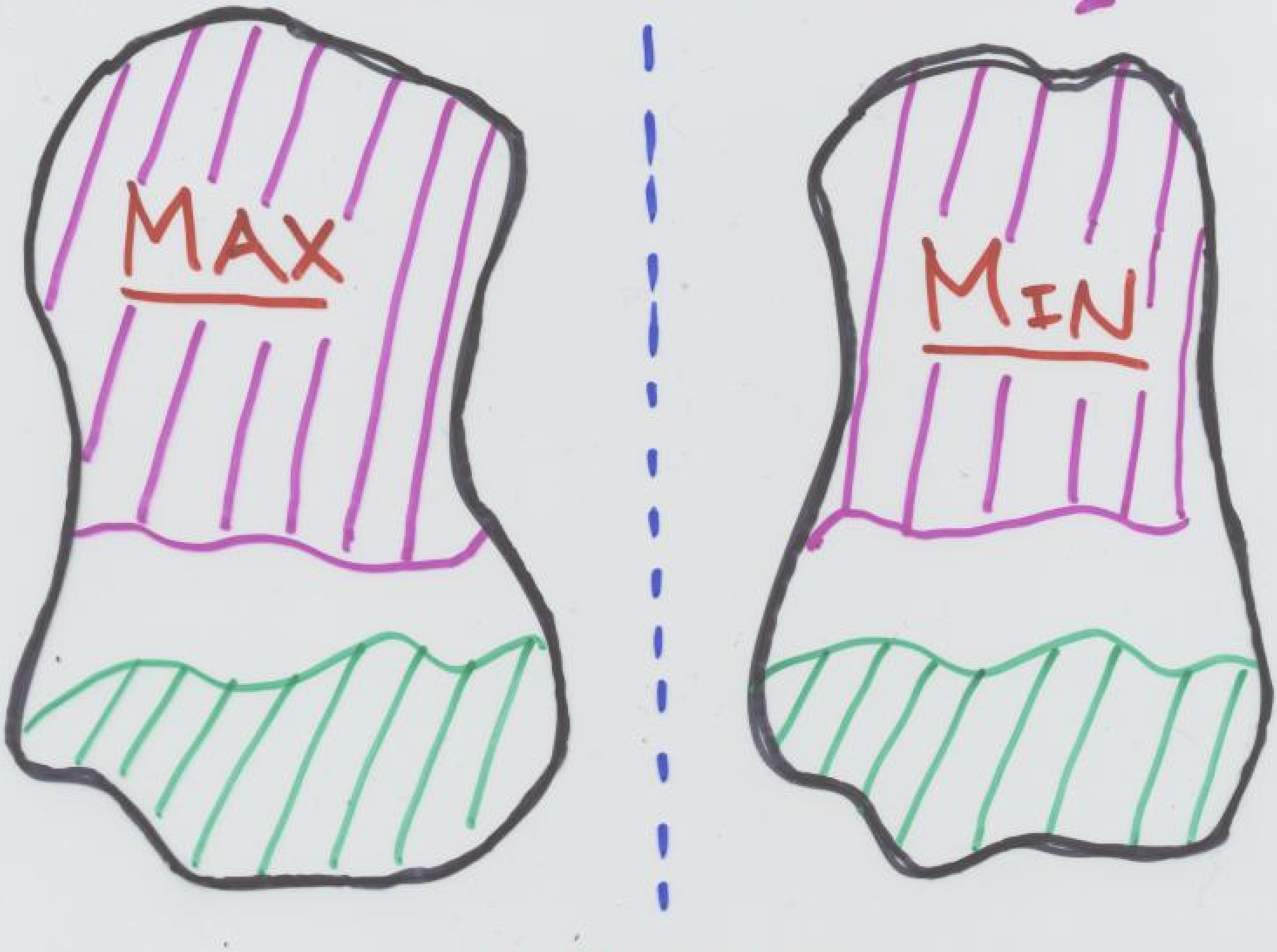
# CSP:

## DENSE INSTANCES



# CSP:

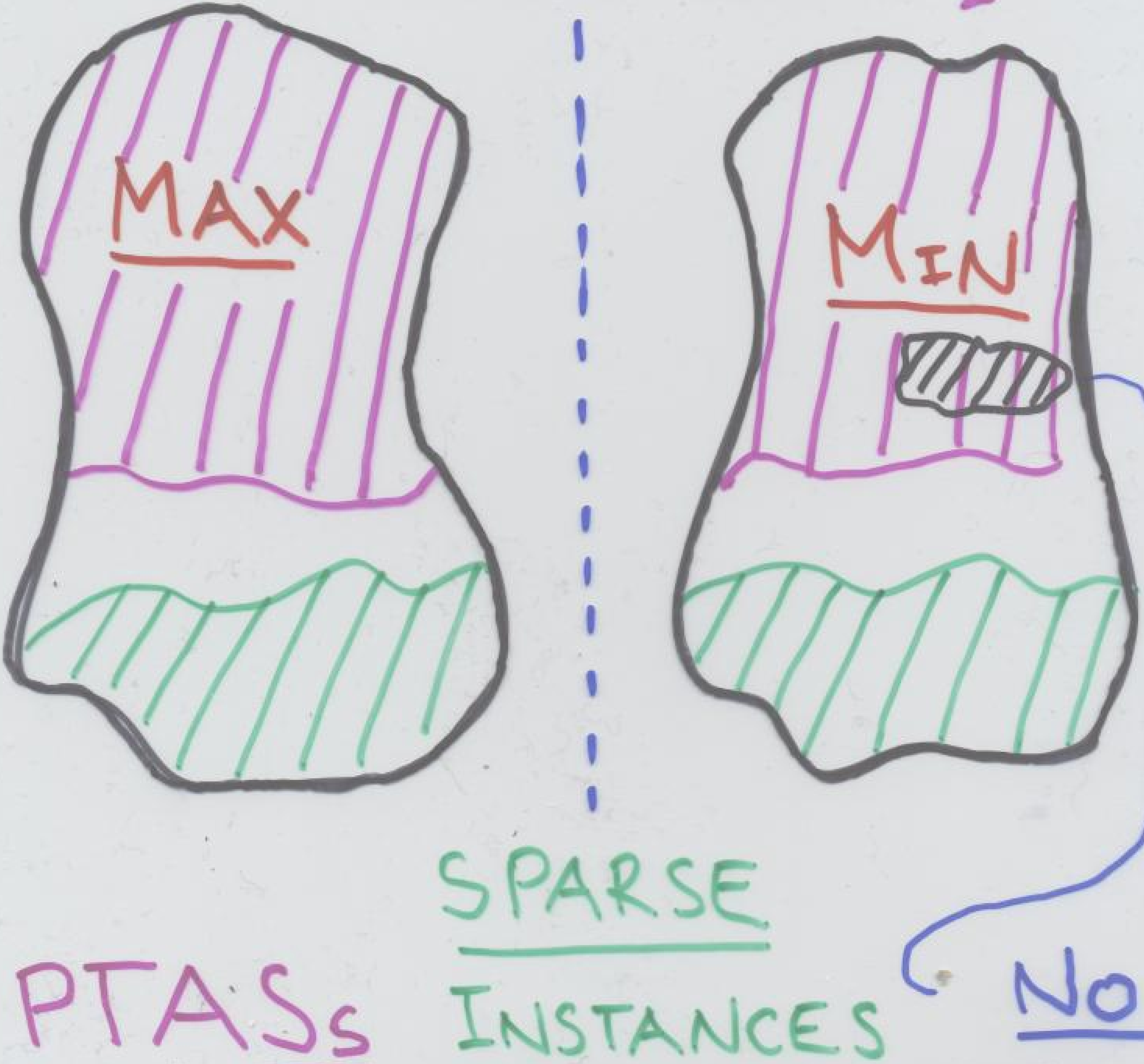
DENSE INSTANCES




SPARSE INSTANCES

# CSP:

[DENSE INSTANCES]




SPARSE INSTANCES

PTASs

[AKK '95]

SMOOTH POLY. INTEGER PROGRAMS.

No PTASs

(→ E.G. - MIN-2DNF)

# SPARSITY

(vs. DENSITY)

TAKE 2DNF-CONJ.

*) $\boxed{\text{MIN-CSP}(x \wedge y, \neg x \wedge \neg y)}$

$\uparrow$

MIN-2CNF-DELETION

KNOWN TO BE MAX-SNP-HARD ([KPRT'96]).

FOR $f$ AN INSTANCE OF *) WITH $n$ VAR'S $x_1, \ldots, x_n$, CONSTRUCT A DENSE INST. $f'$ BY ADDING ALL CLAUSES $\boxed{x_i \wedge y_j}$ FOR $1 \le i, j \le n$. WE HAVE CLEARLY $OPT(f') = OPT(f)$.

NEW VAR'S

$$f' = f \bigwedge_{i,j} (x_i \wedge y_j)$$

IN OPT($f'$)

SET TO 0.

$f'$ IS "EVERYWHERE"

DENSE $\Rightarrow$ DENSE

MIN-2DNF IS

MAX-SNP-HARD.

TWO

OUTSTANDING PROBLEMS:

MIN-$k$SAT

AND

MIN-$k$LIN

$\equiv k$NCP

NEAREST CODEWORD PROBLEM

HERE

"DENSE"

MEANS

"EVERYWHERE DENSE"

(EVERY VAR. $x_i$ OR ITS NEGATION OCCURS $\Theta(n^{k-1})$ TIMES).

$\boxed{\text{MIN-}k\text{SAT}}$

IS

MAX-SNP-HARD ✓

AND

APPROX. WITHIN

$\boxed{2(1-1/2^k)}$

FOR EVERY $k \geq 2$

( [BTV 96], [KKM 94] ).

'06:

APPROX.
UPPER : $O(n/\log n)$
APPROX.
LOWER : $n^{\Omega(1/\log\log n)}$

3NCP =

MIN-CSP $(x \oplus y \oplus z = 0,$
$x \oplus y \oplus z = 1)$

SET OF CONSTRAINT APPLICATIONS IS A SYSTEM OF LIN. EQUATIONS WITH EXACTLY 3 VAR'S PER EQUATION.

→ REF. [7].

HOW

ABOUT

DENSE

MIN-kSAT ?

(MAX-SNP-HARD OR

EXISTENCE OF PTAS?)

( IT IS NP-HARD IN

EXACT SETTING)

- "OPTIMALITY"

OF PTAS $(\forall \varepsilon > 0)$

HOW ABOUT DENSE MIN-kSAT ?

ALIKE MIN-kDNF ??

(MAX-SNP-HARD OR EXISTENCE OF PTAS?)

(IT IS NP-HARD IN EXACT SETTING)

- "OPTIMALITY" OF PTAS ($\forall \varepsilon > 0$)

$\boxed{\text{T1.}}$

$\Rightarrow$ SUGGESTS

A PARTITION OF INSTANCES OF MIN-$k$ SAT INTO INSTANCES $f$ WITH

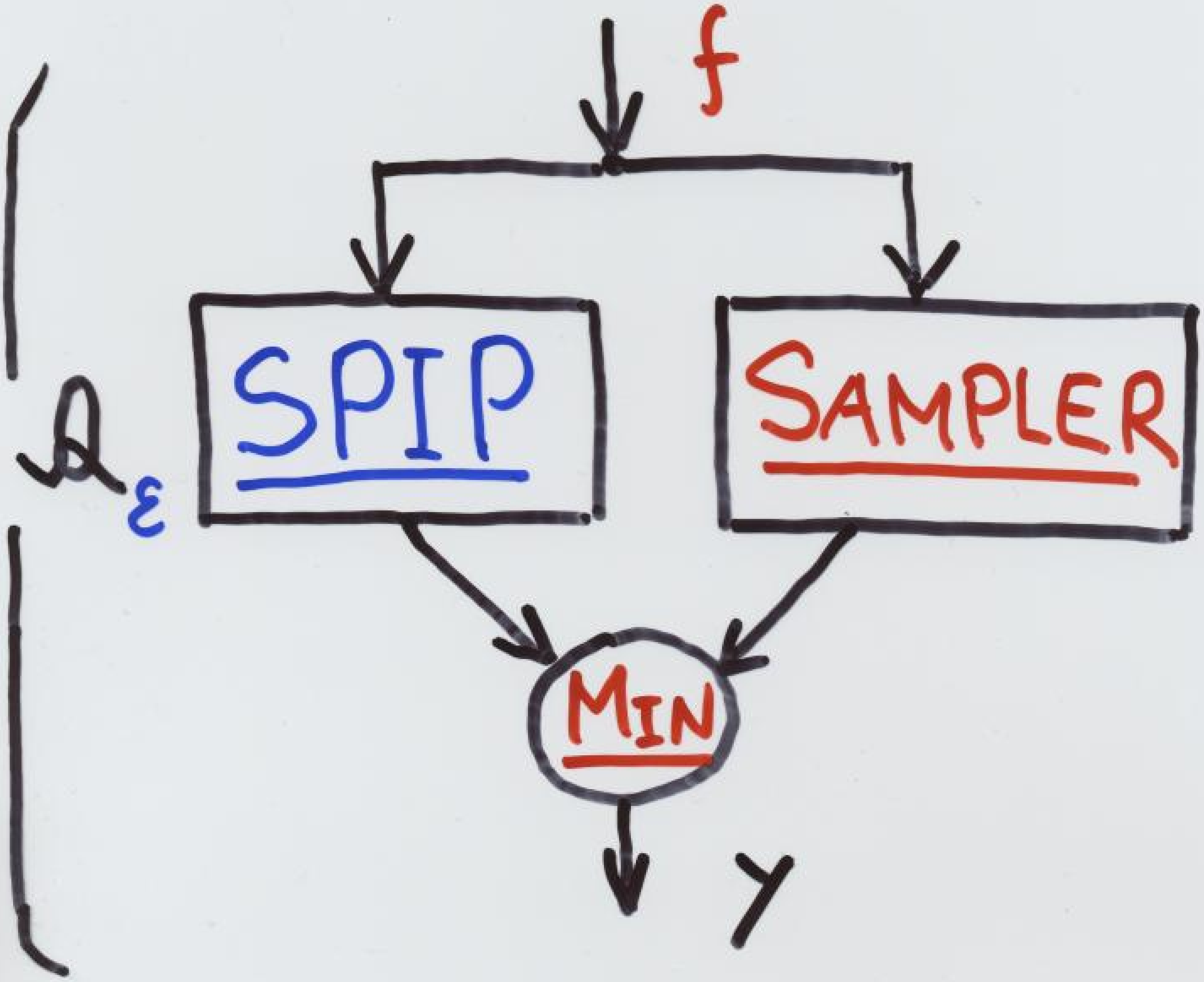
- $\text{OPT}(f) \geq \alpha n^k$,

AND

- $\text{OPT}(f) < \alpha n^k$.

# ALGORITHMS

$\mathcal{A}_\varepsilon$ OF THE FORM:



## METHOD:

SAMPLER DESIGN FOR 'RETRACT DENSITY' OF $(k-1)$-UNIFORM HYPERGRAPHS.

OBJ. FNCT. $\leq \alpha n^k$

**THEOREM.**

DENSE MIN-kSAT AND kNCP PROBLEMS HAVE PTASs (FOR EVERY $k \geq 2$)

∎

# A PTAS for Dense Nearest Codeword Problem.

# A PTAS BASED ON SAMPLERS FOR 3NCP.

I.: A $\delta$-DENSE INSTANCE OF MIN-E3-LIN-2, AND $\varepsilon > 0$.

1. APPLY SPIP-ALG. OF [AKK95]
   OUTPUT $= \bar{x}_1^* \in \{0,1\}^n$.
   (→'START' NOW A NEW ALG. BASED ON SAMPLERS)

2. PICK 2 RANDOM SUBSETS $S_1, S_2$ OF THE SET $X = \{x_1, \ldots, x_n\}$ OF VAR's OF SIZE $\Theta\left(\frac{1}{\delta\varepsilon^2}\log n\right)$ $(S_1 \cap S_2 = \emptyset)$.

3. FOR ALL POSSIBLE ASSIGN's $a \in \{0,1\}^{|S_1 \cup S_2|}$ DO: $\boxed{\text{SAMPLER}}$

# A PTAS BASED ON SAMPLERS FOR 3NCP.

I.: A $\delta$-DENSE INSTANCE OF MIN-E3-LIN-2, AND $\varepsilon > 0$.

OBJ. 1. APPLY SPIP-ALG. OF [AKK95]

FNCT.: OUTPUT $= \bar{x}_1^* \in \{0,1\}^n$.

$\geq \alpha n^3$ (→ 'START' NOW A NEW ALG. BASED ON SAMPLERS)

2. PICK 2 RANDOM SUBSETS OF THE SET $X = \{x_1, \ldots, x_n\}$

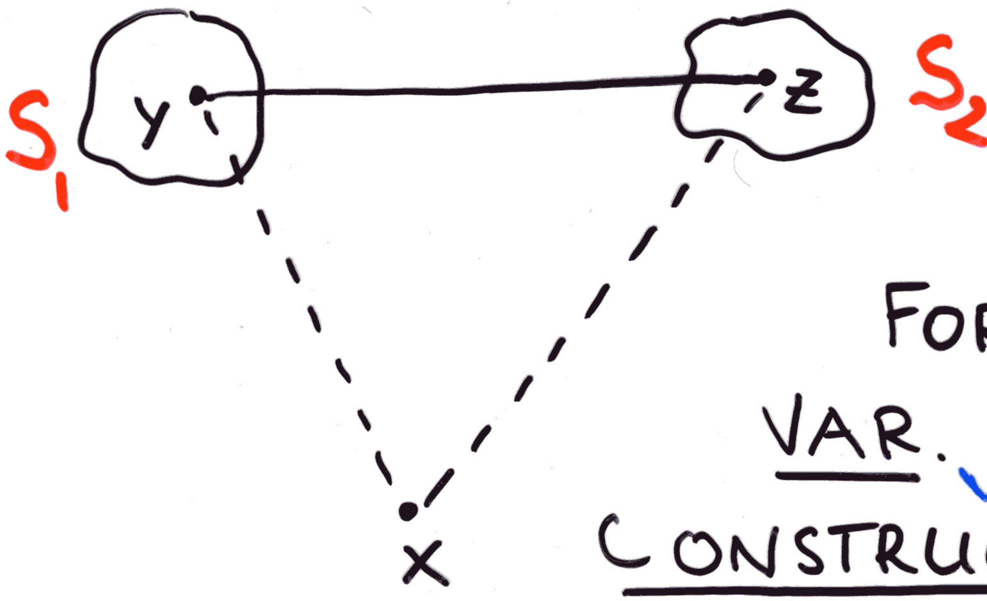
$\leq \alpha n^3$ $\underbrace{S_1}$, $\underbrace{S_2}$ OF VAR's OF SIZE $\Theta\left(\frac{1}{\delta\varepsilon^2}\log n\right)$

$(S_1 \cap S_2 = \phi)$.

3. FOR ALL POSSIBLE ASSIGN's $a \in \{0,1\}^{|S_1 \cup S_2|}$ DO: $\boxed{\text{SAMPLER}}$

# SAMPLER:

(FOR $a \in \{0,1\}^{|S_1 \cup S_2|}$)



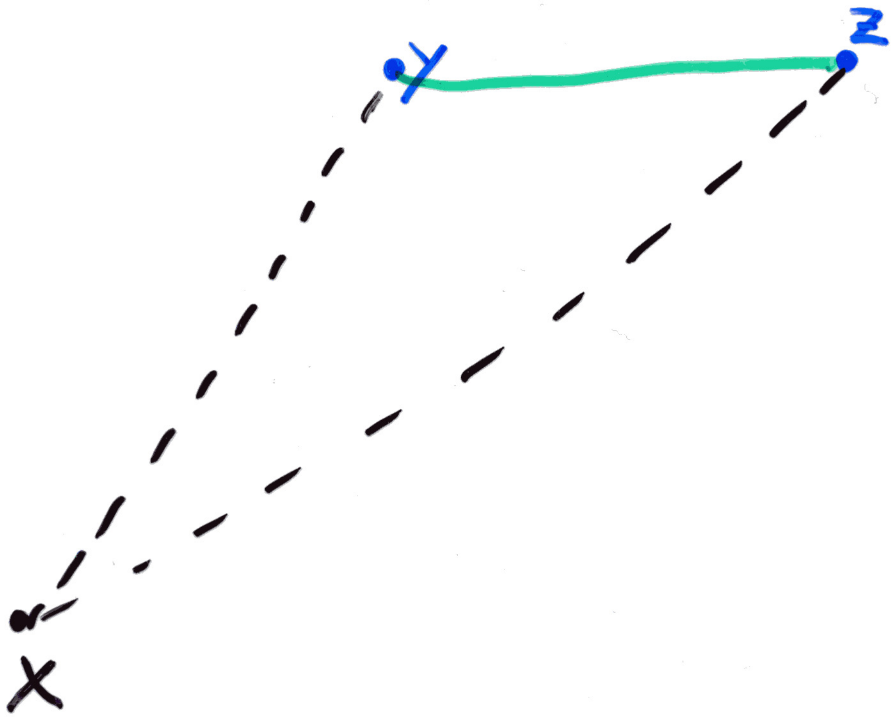

FOR EACH VAR. $x \notin S_1 \cup S_2$ CONSTRUCT A GRAPH $G_x = (S_1 \cup S_2, E_x)$,

$E_x = \{\{y,z\} \mid x \oplus y \oplus z = b$ IS AN EQ., $y \in S_1$ AND $z \in S_2$, OR $y \in S_2$ AND $z \in S_1\}$

$m_0^a$ = # OF EDGES IN $E_x$ SAT. BY $x=0$,

$m_1^a$ = # — " — $x=1$.

FOR ALL $a \in \{0,1\}^{|S_1 \cup S_2|}$ DO:




(DEC. WHETHER

$x := 0$

OR

$x := 1$)

- TEST
  - IF $m_0^a \geq \underbrace{\frac{2}{3}(m_0^a + m_1^a)}$

    THEN SET $\underbrace{x := 1}$;
  - IF $m_1^a \geq \underbrace{\frac{2}{3}(m_0^a + m_1^a)}$

    THEN SET $\underbrace{x := 0}$;

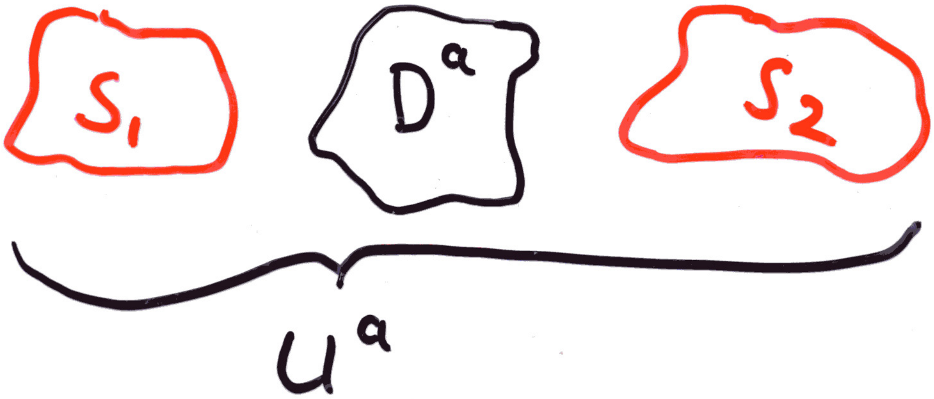
  OTHERWISE SET

  $x :=$ UNDEFINED.

- PLACEMENT OF UNDEFINED VAR'S.

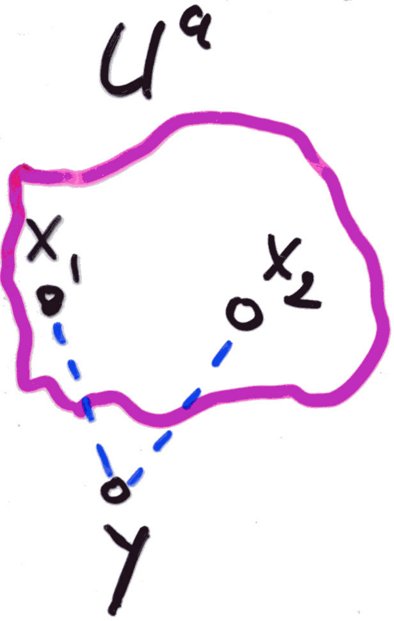
$$\boxed{U^a = S_1 \cup S_2 \cup D^a}$$

VAR'S WHICH ARE ALREADY DEFINED.




SUPPOSE

$y \in X \setminus U^a$ ($y$ IS UNDEF.)




CONSTRUCT

A SET OF EQ.'s $\mathcal{S}_y$ WITH $y$ AS A VARIABLE

IF $\mathcal{S}_y = \emptyset$ WE SET $y$ TO 0 (OR 1 ARBITR.)

$n_0^a$ = # OF EQ.'s SAT. IN $\mathcal{S}_y$ FOR $y=0$;

$n_1^a$ = # ——— " ——— $y=1$

- IF $n_0^a \geq n_1^a$ THEN SET $y := 1$ ELSE SET $y := 0$.

DENOTE THE RESULTING ASSIGNMENT BY $\bar{x}_a^* \, (\in \{0,1\}^n)$.

4. $\boxed{\text{OUTPUT } \bar{x}^* \text{ S.T. } \underbrace{N^*}_{} = \text{MIN}\big(\bar{x}_1^*, \underbrace{\text{MIN}_a(\bar{x}_a^*)}_{\bar{x}_2^*}\big)}$

$\underbrace{N^*}_{}$ = # OF EQ.'S SAT. BY $\bar{x}^*$.

→ CLAIM: $\boxed{N^* \leq (1+\varepsilon) \cdot OPT \quad (\forall \varepsilon > 0)}$

# PART III.

## CONSTANT TIME ALGORITHMS & APPLICATIONS, GETTING BEYOND ABSOLUTE APPROX. BOUNDS.

GETTING THE RUNING TIME OF SPIP-THEOREM DOWN TO

$$\underbrace{\tilde{O}\left(1/\varepsilon^4\right)}.$$

SEE FOR THE INTERMEDIATE RESULTS, AND THEIR APPLIC's REF[1]

SPECIAL LINEAR ALGEBRAIC TECHNIQUES NEEDED - COMBINED WITH SOME CONSTANT TIME SAMPLING METHOD FOR LPs (LINEAR PROGRAMS).

FIRST A LOOK AT THE METHOD OF SPIP-THEOREM: LINEARIZING QUADRATIC PROGRAM P FOR MAX-CUT

$$P:\ \underset{x_i}{\mathrm{MAX}}\left\{\sum_{i} x_i \sum_{(i,j)\in E} a_{ij}\underbrace{(1-x_j)}\right\},$$

$$x_i, x_j \in \{0,1\}.$$

LET $x^*$ BE OPT.
ASSIGNMENT OF P.

LINEARIZE P BY
ESTIMATING

$$\underbrace{z_i - \varepsilon n \leq \sum_{(i,j)\in E} (1 - x_j^*) \leq z_i + \varepsilon n}$$

BY TAKING A
RANDOM SAMPLES OF
SIZE $\Theta(\log n / \varepsilon^2)$ AND
SETTING $z_i = \frac{n}{|S|} \sum_{(i,j)\in E,\, j\in S} (1 - x_j^*)$

SOLVE LINEARIZED LP AND ROUND FRACTIONAL SOLUTION (BY RANDOMIZED ROUNDING).

FOR HIGHER DEGREES USE RECURSION BASED ON POLYN. DECOMPOSITION

$$P(x) = \underset{\substack{\uparrow \\ \text{CONST.}}}{T} + \sum x_i \underbrace{p_i(x_i, \ldots, x_n)}_{\substack{\text{SMOOTH,} \\ \text{DEG} = d-1.}}$$

# Getting Down to Constant Time.

## 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.

**LET** $S_1 \subseteq V_1$, $S_2 \subseteq V_2$, $\ldots$, $S_r \subseteq V_r$, **DEFINE** THE **QANTITY**:

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

A **CUT-NORM** OF $A$:

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

## EXAMPLE:

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

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

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

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

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

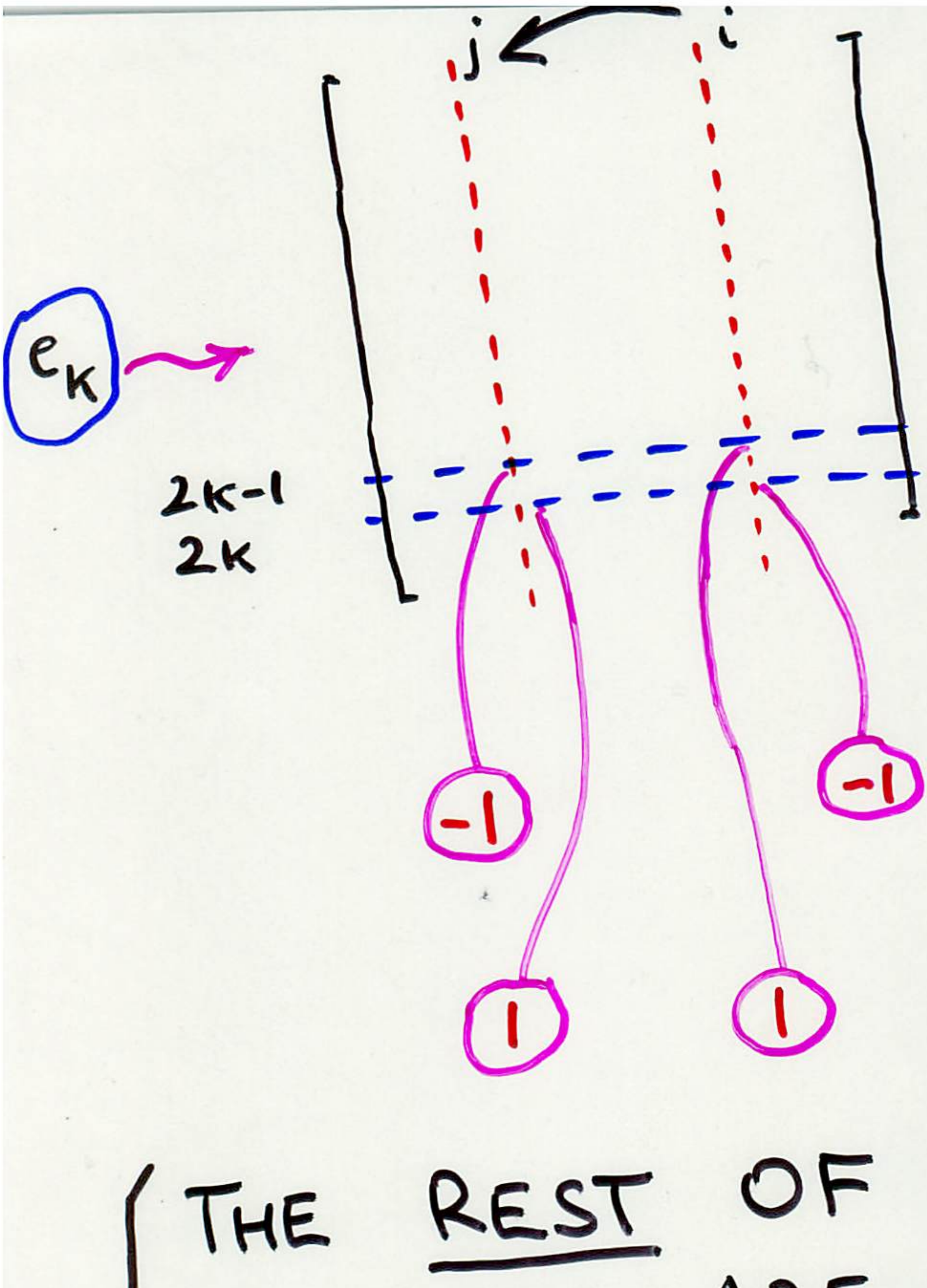
ORIENT $G = (V, E)$ ARBITRARILY, E.G.

$v_i \longrightarrow v_j$ IFF $i > j$.

LET $e_k = (v_i \longrightarrow v_j)$,

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

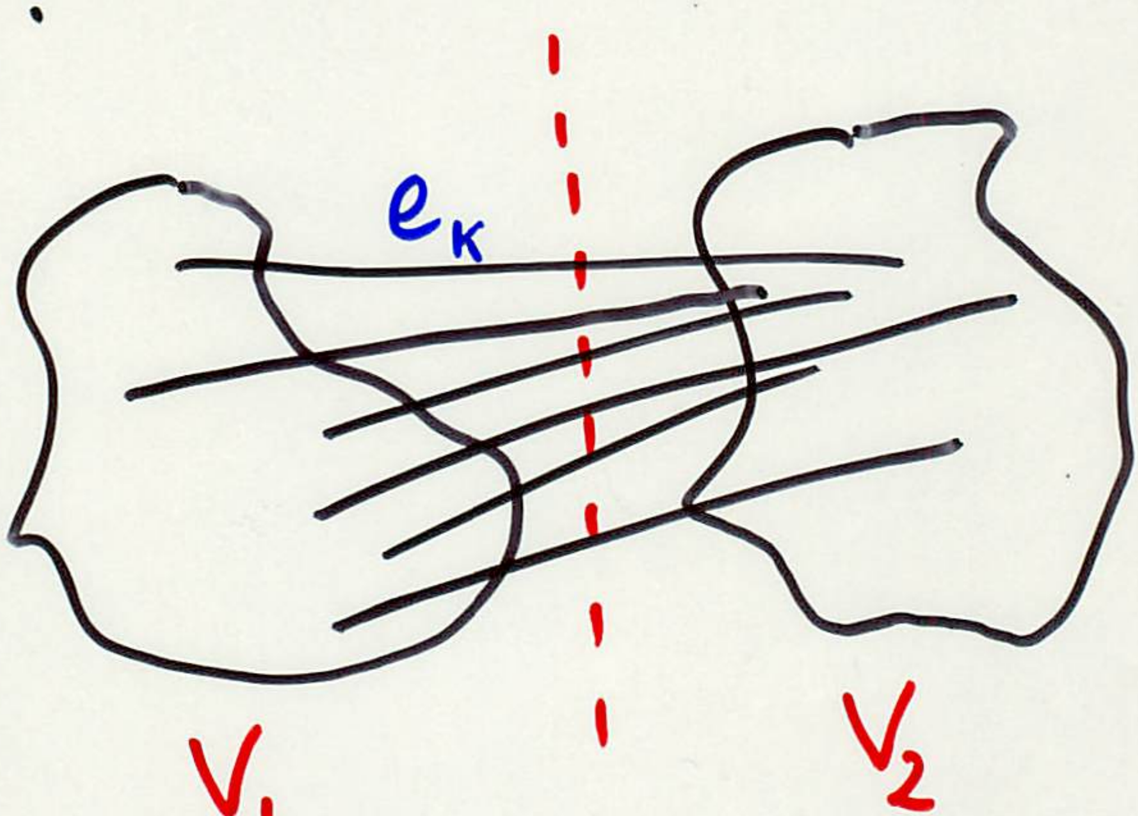
AND $\left[ a_{2k-1,j} = a_{2k,i} = -1 \right]$.




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, $\|A\|_C =$ MAX-CUT(G).

MAX-$r$CSP 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 $\Theta\left(\log(1/\varepsilon)/\varepsilon^4\right)$ $(= \underbrace{q}_{})$.

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

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

W.H.P.

THE OTHER DIRECTION IS EASY:

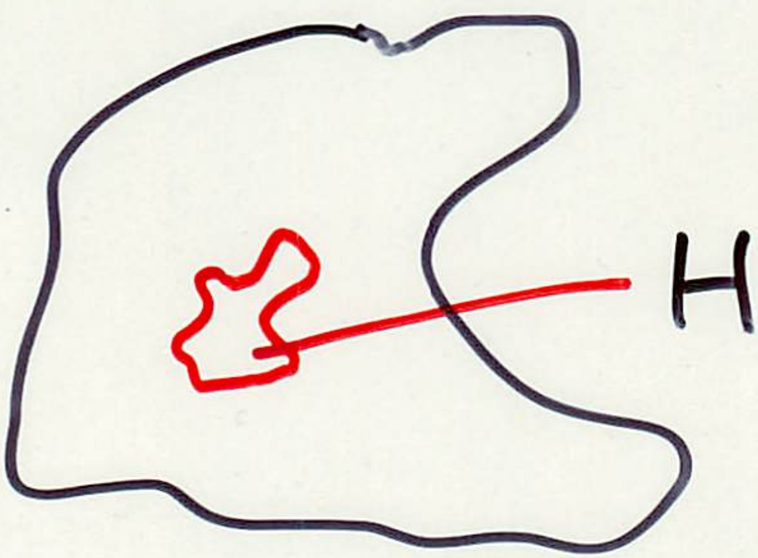
IF $\|A\|_c$ IS HIGH, THEN SO IS $\|H\|_c$.

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:

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

SO FAR, WE WERE ABLE TO DEAL WITH "DOWN-STAIRS" SITATION FROM THE WHOLE ARRAY TO THE SMALL RANDOM SUBARRAY H.



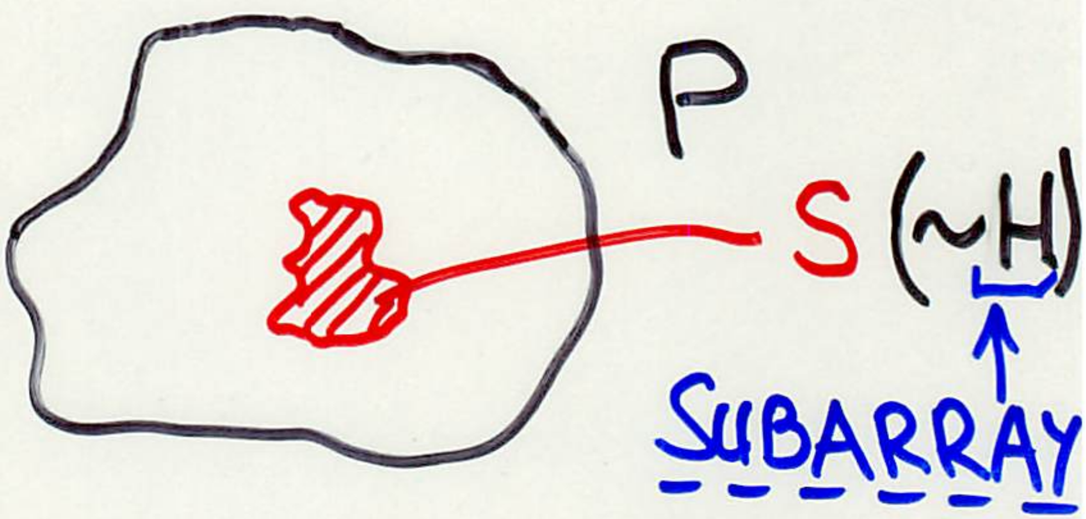
HOW <u>ABOUT</u> REVERSED ("UP-STAIRS") <u>DIRECTION</u>?

TO <u>RELATE</u> $\|H\|_c$ <u>WITH</u> THE $\|A\|_c$ (<u>OPT</u> OF AN <u>INSTANCE</u>).

## LP PART:

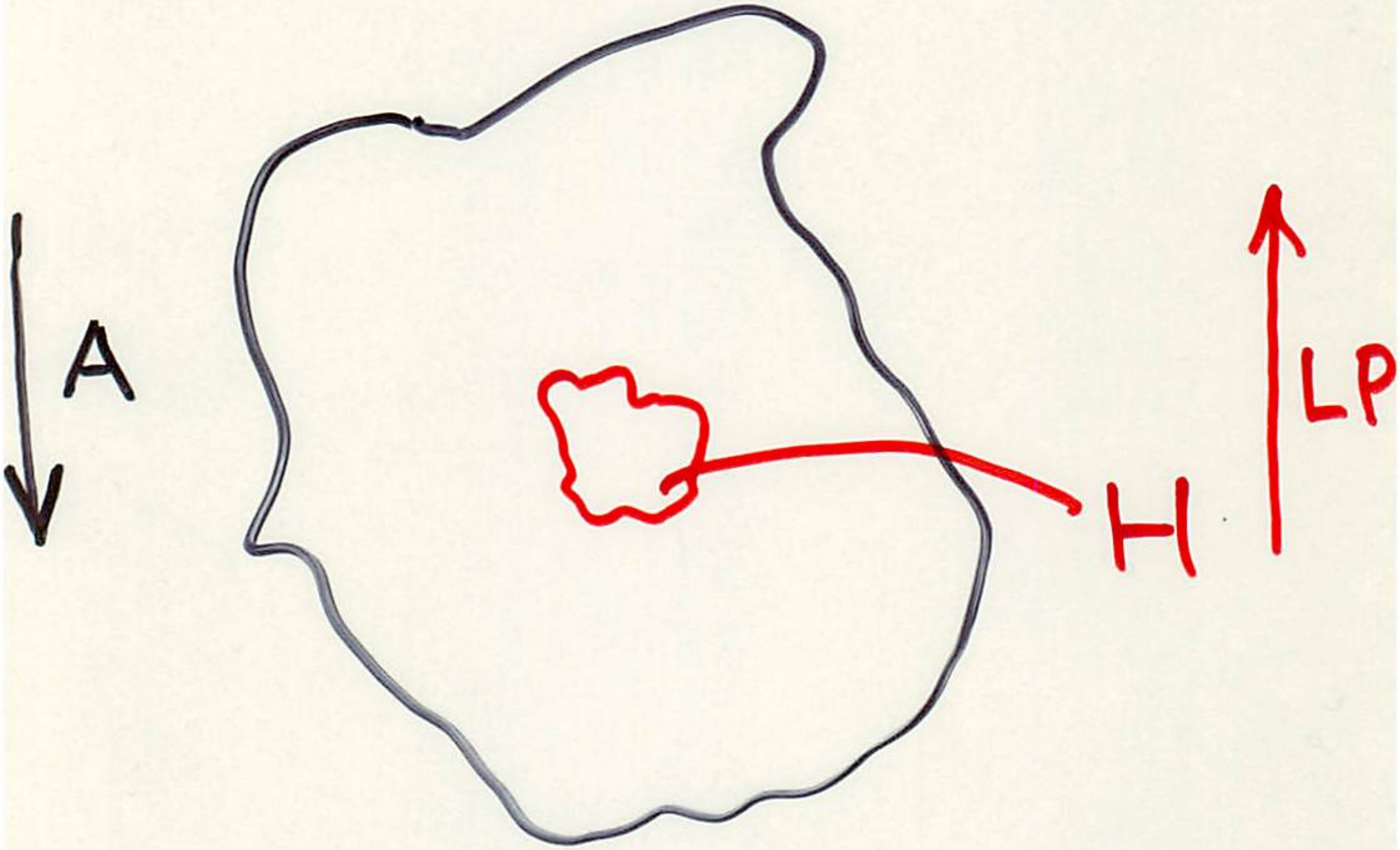
TAKE A LINEARIZED SPIP FOR AN INST. F WHICH IS AN ILP, CALL IT P.

$|S| = \Theta(1/\varepsilon^4)$




IF $\alpha \geqslant \mathrm{OPT}_P$, THEN $\alpha' \geqslant \mathrm{OPT}_{P^S}$.

# A SITUATION:




$$\beta \le \|A\|_c \le \varepsilon n^r$$

$$\Downarrow \qquad \Big\uparrow \text{LP}$$

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

[ ILP P GIVES
NORMAL. FACTOR $\frac{n^r}{q^r}$

# Main Result:

LET $F$ BE AN INST. OF MAX-$r$CSP WITH $n$ VAR.'S. FOR A RANDOM SAMPLE $Q$ OF THE SET OF VAR.'S $\{x_1, \ldots, x_n\}$ 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.

RECENT EXT.'S TO UNBOUNDED WEIGHT CLASSES, INCLUDING METRIC AND QUASIMETRIC CASES.

[FKKV05]

(REF. [3])

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

(BY "NON-HARD-CORE" METHOD)

[FKKR03]

---

NO GENERAL SUBLINEAR PTASs KNOWN.

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

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


MIN-SUM

CLUSTERING

# GETTING BEYOND THE ABSOLUTE BOUNDS.

GIVEN A $\Theta\left(\frac{n^r}{\Delta}\right)$-DENSE INST. OF MAX-rCSP, THERE EXISTS $2^{O(\Delta)}$-TAS WITH $\Theta(\Delta)$ SAMPLE SIZE. (FOR SUBDENSE CLASS, $\Delta = \log n$, WE HAVE PTAS).

PROOF METHOD BASED ON NEW ANALYSIS OF SPECIAL SPIPs FOR THAT PROBLEMS.

[FK05]

(cf. REF. [4])

■

## FURTHER RESEARCH:

IMPROVING THE HARD CORE COMPLEXITY OF THE ALGORITHMS.

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

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

■