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] Arora, Karger K. [AKK95]

Combinatorial method for dense Max-Cut. Smooth high deg. poly. integer programs for dense Max-CSP.

Property testing
Streamline 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 \textbf{opt.} classical 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.
Explicit PCP-Method:

HGL-Method:

⇒ A.R. = \frac{\lambda_0}{\lambda_1} - 3

\forall \lambda_0 \geq 3

1 \text{ OR } 2 \text{ NP-HARD TO DECIDE.}
For MAX-CUT:

\[
\frac{\lambda_0}{\lambda_1} = \frac{17}{16}
\]

Under the Unique Game Conjecture:

\[
\frac{\lambda_0}{\lambda_1} = 1.1383
\]
APPROX.

ISSUES

OF MAX-CSP.
Max/Min CSP

Equivalent to the following:

Given a set of $r$-ary Boolean functions $f_1, f_2, \ldots, f_m$. Construct an assignment $x \in \{0,1\}^n$ so as to maximize/minimize the number of satisfied function's $f_i$. 
Two notions of approx.'s:

- **Absolute:**
  \[ |\text{OPT} - y| \leq \tau \]

- **Relative:**
  \[ \max \left\{ \frac{y}{\text{OPT}}, \frac{\text{OPT}}{y} \right\} \leq \tau \]

\[ y = \text{costs of solution } s. \]
AF-CLASS

III

PTAS
An \textbf{EPTAS} is \textbf{constant} time as \textbf{(CTAS)} if its running time \textbf{is} 
\[ f\left(\frac{1}{\varepsilon}\right). \]
$\delta$-Absolute

PTAS ($\delta$-APTAS)

Condition:

$|\text{OPT} - y| \leq \epsilon \delta$, for all $\epsilon > 0$.

If $\text{OPT}$ is of order $\delta$ ($\text{OPT} = \Theta(\delta)$) (dense instances),

exist. of $\delta$-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]
\( \sim \left( \frac{n^k}{\log^{O(\sqrt{n})} n} \right) \)-Clauses
RUNNING TIME IMPROVEMENTS
FOR MAX-CSP:

'95 ----> 2003.
[AFKK03]

• Absolute CTASs

(0^\sim(\frac{1}{3}4) \text{ SAMPLE},
2^\sim(\frac{1}{3}2) \text{ TIME})

(\text{ THAT AREA BECAME TO BE KNOWN AS THE PROPERTY TESTING })
\[
\downarrow
\]

- Dense instances of MAX-CSP have CTASs \(O^{\Theta(\frac{1}{\sqrt[4]{3}})}\) sample, \(O^{\Theta(\frac{1}{\sqrt[2]{3}})}\) time.
Opt. Problem \( P \) (Max-Cut, Max-3Sat, ...)

\[ S \]

"Hard" Core Sample.

Constant Size (Very Tiny) Sample, Does Opt Is I Tells You "Something" About Opt 2.2

I
OLD METHOD:
Given any \textsc{Max-}$r$\textsc{CSP} problem $P$, "simulate" it by an SPIP $P'$ of degree $r$.

There exists an $r$-APTAS for $P'$, $\frac{\text{OPT}_{P'}}{\text{OPT}_P} - \epsilon \leq \varepsilon_n$ for all $\varepsilon > 0$. 
RUNNING

TIME:

$O(\frac{1}{\varepsilon^2})$

$\nu$
GETTING DOWN TO CONSTANT TIME.
METHOD:

PROBABILITYSTIC

$\tau$-DIMENSIONAL

SUB-ARRAYS

AND THEIR

CUT-NORMS

\downarrow

CONSTANT TIME

APPROX. OF MAX-$\tau$CSP
NOTATION:

Given finite sets \( V_1, V_2, \ldots, V_r \), an \( r \)-dimensional array \( A \) on \( V_1, \ldots, V_r \) is a function

\[
A : V_1 \times V_2 \times \cdots \times V_r \rightarrow \mathbb{R}
\]

\( A(i_1, i_2, \ldots, i_r) \) is an entry of \( A \).
**Frobenius Norm of a**: 

\[ \|A\|_F = \text{square root of the sum of squares of all entries}. \]
Let $S_1 \leq V_1$, $S_2 \leq V_2$, ..., $S_r \leq V_r$, define the quantity:

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

A cut-norm of $A$:

$$\|A\|_C = \max_{S_1 \leq V_1, ..., S_r \leq V_r} |A(S_1, ..., S_r)|.$$
CUT-NORM RELATES FOR $r=2$ TO GROTHENDIECK NORM, AND SD-
PROGRAMMING,

$$\|\text{All}_G\| = \max \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} x_i y_j : x_i, y_j \in \{0, 1\} \right\}$$

$$\|\text{All}_C\| \leq \|\text{All}_G\| \leq 4 \|\text{All}_C\|$$
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 $G = (V, E)$,

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

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

$E' = \{e_1', \ldots, e_m', e'_1, \ldots, e'_m\}$. 
ORIENT $G = (V, E)$ ARBITRARILY, E.G.

$V_i \rightarrow V_j$ IFF $i > j$.

LET $e_k = (V_i \rightarrow V_j)$.

SET: $\begin{bmatrix}
a_{2k-1}, i = 1 \\
a_{2k}, j = 1 \\
a_{2k-1}, j = a_{2k}, i = -1
\end{bmatrix}$.
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 entries in $\mathbf{A}$. Thus, $\|\mathbf{A}_{|c}| = \text{max-cut}(G)$. 
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 $\|\text{All For } r\text{-Dim. Arrays } A\|$.
APPROXIMATION
OF $\|A\|_{\infty}$.

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

ASSUMPTIONS (*):
ON $A$:

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

$\|A\|_{\infty} \leq 2^r n^{r/2}$.
Then, a \textbf{random} \underline{induced subarray} $H$ of a $S\text{atisfies}$

\[ \|H\|_c \leq C(r) \cdot q \]

\underline{W.H.P.}

The other direction is easy:

If $\|A\|_c$ is \textbf{high},

then so is $\|H\|_c$. 
More explicitly:

For

\[ q \geq 10^6 \tau^{12} \frac{1}{\delta^5 \varepsilon^4} \log \left( \frac{4}{\varepsilon^2} \right), \]

\[ \delta, \varepsilon > 0, \]

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

with \( \text{Prob.} \geq 1 - \delta. \)
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: $O\left(\frac{1}{\varepsilon^2}\right)$. 2
A SITUATION:

\[ \beta \leq ||A|| \leq \alpha \]

\[ \beta' \leq ||H|| \leq \alpha' \]

\[ \frac{n}{q} \]

\[ \text{ILP P GIVES NORMAL. FACTOR} \]

\[ \text{LP} \]
Let $F$ be an instance of $\textsc{Max-}$-$r$-$\text{CSP}$ 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-instance induced by $Q$.

Main result:

$$\left|\frac{1}{n} \sum_{q=1}^{n} \frac{\text{OPT}_{F^q} - \text{OPT}_F}{\text{OPT}_F}\right| \leq \varepsilon \ln^n$$

for $|Q| = q = O(\log(1/\varepsilon) / \varepsilon^4)$.
Very recent improvement of a part to hard core size $\sim (\frac{1}{3})^2$ 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"
  \[\text{[FKKR03]}\]

No general sublinear PTASs known.
\[ \sum_{i=1}^{K} \sum_{\{u, v\} \in C_i} d(u, v) \]

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

\text{INTRA-CLUSTER DISTANCES}

\text{MIN-SUM CLUSTERING}
VERY RECENT
GENERAL RESULT:
MAX-$k$CSP ON INST.
WITH $\sim \left(\frac{n^k}{\Delta}\right)$-CLAUSES
HAS A $2^{O(\Delta)}$-TAS.

[FKO6].
Ongoing work on Subdense version of Szemeredi's lemma (with Density $\log n$) 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" PROBLEM INTRACTABILITY BARRIERS FOR GETTING DOWN TO, SAY, $\Theta(\frac{1}{\varepsilon^2})$ 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?