Tensor Decomposition and Approx Schemes for CSP's

W. Fernandez de la Vega

Ravi Kannan

Marek Karpinski

Santosh Vempala

Max-r CSP : Well-known class of problems. Example : Maximize the number (total weight) of clauses satisfied in a $r-$CNF Boolean formula.

Throughout: n is the number of variables. $n \to \infty$.

"arity" $r \in O(1)$.

Here : A Max$-r$CSP can be formulated as maximizing a homogeneous degree r polynomial in the variables $x_1, x_2, \dots x_n, (1 - x_1), (1 - x_2), \dots (1 - x_n)$ over

$S = \{y = (x_1, x_2, \dots x_n, (1-x_1), \dots (1-x_n) : x_i \in \{0, 1$

Coefficient array - A- has 0,1 entries in the unweighted case.

Case $r = 2$:

$$\text{Max}_{y \in S} \sum A_{ij}\, y_i y_j \quad (= y^T A y).$$

What was known (1)

$O(1)$ factor approx alg.s. (SDP based - GW ....)

What class of Max$-r$CSP admit Poly time approx schemes ?

Dense unweighted Class : $\Omega(n^r)$ clauses. A has $\Omega(n^r)$ 1's.

Arora, Karger, Karpinski: PTAS for general r. (STOC 1995).

Goldreich, Goldwasser, Ron $O(1)$ time alg. for $r = 2$ - Property Testing Combinatorial arguments (FOCS 96)

Frieze, Kannan PTAS, then $O(1)$ time alg. for general r... Linear Algebra based arguments (FOCS 96, 99)

What was known (2)

Alon, de la Vega, Kannan, Karpisnki : Arbitrary $r-$CNF formula F in $x_1, x_2, \ldots x_n$. Q uniform random subset of $x_1, x_2, \ldots x_n$ of cardinality q. F^Q the induced formula on the variables in Q. Then, whp

$$\left|\text{Max}(F) - \frac{n^r}{q^r}\text{Max}(F^Q)\right| \leq \epsilon n^r,$$

provided $q \in \Omega^*(1/\epsilon^4)$, where Max$(F)$ is the maximum number of simultaneosly satisfiable clauses of F. (STOC 02).

$|Q|$ = Sample Complexity = poly $(1/\epsilon)$.

Rudelson, Vershynin Improvements using functional analysis techniques. (05)

Engebretson and Anderson (02) : General r, poly Sample complexity $= 1/\epsilon^7$ using combinatorial methods.

Question Can we get PTAS's for a more general class than unweighted dense problems ?

Main Theorem here : There is a PTAS for any core-dense family of MAX$-r$CSP 's.

Includes the unweighted dense case, the "metric" case, powers of metrics and generalizations of metrics to higher r.

First, we focus on the $r = 2$ case. Here

$$D_i = \text{ “degree of node } i = \sum_j A_{ij}$$

$$\bar{D} = \text{average degree } = \frac{1}{n}\sum_i D_i.$$

$$\mathit{Scaling} \quad B_{ij} = \frac{A_{ij}}{\sqrt{(D_i + \bar{D})(D_j + \bar{D})}}.$$

$$\mathit{Core-Dense} \ : \ ||B||_F^2 = \sum_{ij} B_{ij}^2 \in O(1).$$

$$\mathit{Scaling} \quad B_{ij} = \frac{A_{ij}}{\sqrt{(D_i + \bar{D})(D_j + \bar{D})}}.$$

$$\mathit{Core - Dense} \ : \ ||B||_F^2 = \sum_{ij} B_{ij}^2 \in O(1).$$

Dense Case : $\bar{D} = \Omega(n)$, so denominator $= \Omega(n^2)$ for each i, j.

Triangle Inequality : $A_{ij} \leq (D_i + D_j)/n$.

Core-Dense roughly means that such an inequality is satisfied on the average.

Algorithm for the case $r = 2$ + Bisection etc.

(1) Scale A to get B.

(2) Find a low-rank approximation $\widehat{B}$ of B using Singular Value Decomposition. $\hat{A}$ the "de-scaled" version of $\widehat{B}$ is then a low-rank approx to A.

(3) Solve $\text{Max}_{y \in S} y^T \hat{A} y$ (in place of the original problem - $\text{Max} y^T A y$.).

$\hat{A} \approx A$ implies this is an approx solution.

$\text{Max} y^T \hat{A} y$ is a rank($\hat{A}$)-variable problem, rather than a n variable problem. So it can be solved ("brute-force") in time exponential in rank($\hat{A}$).

(1), Bisection is new. (2), (3) use ideas from FK, other papers.

For the metric max-cut problem and metric bisection problem earlier PTAS in FOCS 98 by dela Vega, Kenyon and SODA 04 by above + Karpinski, Rabani

Norms and low-rank Approximations of Matrices

B an $m \times n$ matrix. Define

$$||B||_F^2 = \sum_{ij} B_{ij}^2 \quad ||B||_2 = \text{Max}_{x,y:|x|=1=|y|} x^T B y.$$

(1) For $\epsilon > 0$, there exits a rank $1/\epsilon^2$ matrix $\hat{B}$ such that

$$||B - \hat{B}||_2 \leq \epsilon ||B||_F.$$

(2) Such a $\hat{B}$ may be found in poly time by Linear Algebra.

For any cut (0-1 vector) x,

$$|x^T (B - \hat{B})(1 - x)| \leq n ||B - \hat{B}||_2,$$

so max-cut with weights $B \approx$ max-cut with weights $\hat{B}$.

For higher r, additional problem is (2).

Higher r : Tensors

An $(r-)$ tensor A on $[n]$ is an $r-$ dimensional array with generic entry $A_{i_1,i_2,\dots i_r}$.

Rank 1 tensor - $v^{(1)} \otimes v^{(2)} \otimes \dots v^{(r)}$ has generic entry $v^{(1)}_{i_1} v^{(2)}_{i_2} \dots v^{(r)}_{i_r}$.

Fundamental Question For a given tensor A, k, find the sum of k rank 1 tensors "closest" to A.

"An Eckart-Young" type of optimal rank-k approximation theorem for tensors continues to elude our investigations but can perhaps be eventually attained by using a different norm..." - Kolda 2001 survey.

Lim, Golub Such optimal approximations may not exist !!

Here - polynomial time algorithm for a "good" approximation - with proven upper bound on error. Enough for our algorithmic result (though cerainly not optimal.)

For ease of notation, consider only $r = 3$ here. [Proceedings paper deals with general r.] Notation (x, y, z are vectors of length 1.)

$$A(x,y,z) = \sum_{ijk} A_{ijk} x_i y_j z_k; \; ||A||_2 = \text{Max}_{x,y,z:} A(x,y,z).$$

$$\forall x, y, z, \quad ||A - A(x,y,z) x \otimes y \otimes z||_F^2 = ||A||_F^2 - (A(x,y,z)$$

Either (i) there exists a x, y, z with high $A(x,y,z)$ in which case, we can "peel off" $A(x,y,z) x \otimes y \otimes z$ from A and repeat.

Or (ii) $||A||_2$ is small and we stop.

(i) cannot occur too many times since $||A||_F^2$ cannot fall below 0. Remaining Problem : Find x, y, z with high $A(x,y,z)$

Finding Max$A(x,y,z)$

(1) If the maximizing x,y are known, then $z = A(x,y,.)/|A(x,y,.)|$ works.

(2) $A(x,y,.)_k = \sum_{ij} A_{ijk}x_iy_j$ can be estimated by drawing a random sample S of $O(1/\epsilon^2)$ of (i,j) and just summing over these (i,j).

non-uniform sampling : Prob$(i,j) \in S = \sum_k A_{ijk}^2/||A||$

(3) Need values of x_i, y_j only for $(i,j) \in S$. Enumerate "all possible" values of these - only exponential in $|S|$ possibilities.

(4) For each enumerated value, pretend it is the correct one. This yields an approx z. Now, for that z, find the best x,y - an arity $r-1$ problem.

(5) Take the best $A(x,y,z)$.

Connection to Low-Dim Embeddings

Theorem A metric on $[n]$. k positive integer. We can find in deterministic poly time a Euclidean $k-$ dimensional embedding σ of $[n]$ satisfying :

$$\forall S, T \subseteq [n],$$

$$\sum_{i \in S, j \in T} A_{ij} = \left(\sum_{i \in S} \sigma(i) \right) \left(\sum_{j \in T} \sigma(j) \right) + \Delta(S,T),$$

$$\text{where, } |\Delta(S,T)| \leq O\left(\frac{1}{\sqrt{k}}\right) \sum_{i,j \in [n]} A_{ij}.$$

The embedding has the property that distances are translated into dot products. The error is in an average sense. Third, $\Delta(S,T)$ may be positive or negative. But the dimension is constant for constant ϵ.

Open Problems

(i) Charecterise better the error of the tensor approximation.

(ii) Other Applications of the tensor approximation.

(iii) Connection to Embeddings ?