

Tensor Decomposition and Approximation Schemes for Constraint Satisfaction Problems

W. Fernandez de la Vega ^{*} Ravi Kannan [†] Marek Karpinski [‡]
Santosh Vempala [§]

Abstract

The only general class of MAX- r CSP problems for which Polynomial Time Approximation Schemes (PTAS) are known are the dense problems. In this paper, we give PTAS's for a much larger class of weighted MAX- r CSP problems which includes as two important special cases — the dense problems and, for $r = 2$, all metric instances (where the weights satisfy the triangle inequality) and quasimetric instances; for general r , our class includes a generalization of metrics. (We note that a PTAS is known specifically for the metric case of the MAX-CUT and Partitioning problems.) Our algorithms are based on low-rank approximations. Besides the algorithmic result, our method has two novel features — first, a way of approximating a tensor by the sum of a small number of “rank 1” tensors, akin to the traditional Singular Value Decomposition which may be of independent interest and, second, a natural and simple way of scaling the weights.

1 Introduction

The singular value decomposition is a useful tool in the design of efficient algorithms for a variety of problems (e.g., [FKV98, Mc01]). In this paper, motivated by boolean constraint satisfaction problems (CSP's) with r variables per constraint, we propose an extension of low-rank approximation to *tensors*, i.e., r -dimensional real arrays. We give an efficient algorithm for finding such an approximation and apply it to weighted MAX- r CSP, i.e., the problem of finding a boolean assignment that maximizes the total weight of satisfied constraints. As a consequence, for any MAX- r CSP that satisfies a certain density condition, we obtain a polynomial-time approximation scheme. In the past, there has been much progress on special cases, in particular there are polynomial-time approximation schemes for *dense* unweighted problems [AKK95, F96, FK96, GGR96, FK00, AFKK02], and several cases of 2-CSP with metric weights including maxcut and partitioning [FK98, I99, FK03, FKK04]. We will show that our density condition captures all known special cases for which PTAS's exist as well as the metric MAX-2CSP (for which no PTAS was known before) and some generalizations.

^{*}LRI, Université Paris-Sud, Orsay, Paris. Email: lalo@lri.lri.fr

[†]Dept. of Computer Science, Yale University, New Haven, CT 06520. Email: kannan@cs.yale.edu

[‡]Dept. of Computer Science, University of Bonn. Email: marek@cs.uni-bonn.de

[§]Dept. of Mathematics, MIT, Cambridge, MA 02139. Email: vempala@math.mit.edu

A MAX- r CSP problem can be formulated as a problem of maximizing a homogeneous degree r polynomial in the variables $x_1, x_2, \dots, x_n, (1 - x_1), (1 - x_2), \dots, (1 - x_n)$. Let $\mathbf{S} = \{y = (x_1, x_2, \dots, x_n, (1 - x_1), (1 - x_2), \dots, (1 - x_n)) : x_i \in \{0, 1\}\}$ be the solution set. Then the problem is

$$\text{Max}_{y \in \mathbf{S}} \sum_{i_1, i_2, \dots, i_r} A_{i_1, i_2, \dots, i_r} y_{i_1} y_{i_2} \dots y_{i_r}.$$

where A is a given symmetric r -dimensional array (i.e., $A_{i_1, i_2, \dots, i_r} = A_{i_{\sigma(1)}, i_{\sigma(2)}, \dots, i_{\sigma(r)}}$ for any permutation σ) with nonnegative real entries. The entries of the r -dimensional array A can be viewed as the weights of an r -uniform hypergraph on $n = |V|$ vertices.

Our main tool to solve this problem is a generalization of low-rank approximation. A rank-1 tensor is the *outer product* of r vectors $x^{(1)}, x^{(2)}, \dots, x^{(r-1)}, x^{(r)}$, is the tensor or r -dimensional array whose (i_1, i_2, \dots, i_r) 'th entry is $x_{i_1}^{(1)} x_{i_2}^{(2)} \dots x_{i_r}^{(r)}$; it is denoted $x^{(1)} \otimes x^{(2)} \otimes \dots \otimes x^{(r)}$. A rank one matrix is the outer product of two vectors and a low rank matrix is the sum of a small number of these. We will show the following:

1. For any r -dimensional array A , there exists a good approximation by the sum of a small number of rank-1 tensors.
2. We can algorithmically find such an approximation.

In the case of matrices, traditional Linear Algebra algorithms find good approximations. Indeed, we can find the *best* approximations under both the Frobenius and L_2 norms using the Singular Value Decomposition (SVD). Unfortunately, there is no such theory (or algorithm) for r -dimensional arrays when $r \geq 2$. Here, we will develop sampling-based algorithms for finding low-rank approximations as defined above. These claims are formalized in the next lemma and theorem (see Section 2 for the definition of the generalized norms).

Lemma 1. *For any A , $\epsilon > 0$, there exist $k \leq 1/\epsilon^2$ rank-1 tensors, B_1, B_2, \dots, B_k such that*

$$\|A - (B_1 + B_2 + \dots + B_k)\|_2 \leq \epsilon \|A\|_F.$$

Theorem 2. *For any A , $\epsilon > 0$, we can find k rank 1 tensors B_1, B_2, \dots, B_k , where $k \leq 4/\epsilon^2$, in time $(n/\epsilon)^{O(1/\epsilon^4)}$ such that with high probability at least $3/4$ we have*

$$\|A - (B_1 + B_2 + \dots + B_k)\|_2 \leq \epsilon \|A\|_F.$$

The proofs and the algorithm for low-rank tensor approximation are given in Section 2.

Next, we give a density condition so that if a MAX- r CSP viewed as a weighted r -uniform hypergraph satisfies this condition, then there is a PTAS for the problem. This condition is a new unified framework for a large class of weighted MAX- r CSP's.

Define the node weights D_1, \dots, D_n of A and their average as

$$D_i = \sum_{i_2, i_3, \dots, i_r \in V} A_{i, i_2, \dots, i_r} \quad \bar{D} = \frac{1}{n} \sum_{i=1}^n D_i.$$

Definition 1. *The core-strength of a weighted r -uniform hypergraph given by an r -dimensional tensor $A : V \times V \times \dots \times V \rightarrow \mathbf{R}$ is defined as*

$$\left(\sum_{i=1}^n D_i \right)^{r-2} \sum_{i_1, i_2, \dots, i_r \in V} \frac{A_{i_1, \dots, i_r}^2}{\prod_{j=1}^r (D_{i_j} + \bar{D})}$$

We say that a class of weighted hypergraphs (MAX- r CSP's) is *core-dense* if the core-strength is $O(1)$ (i.e., independent of A, n).

To motivate the definition, first suppose the class consists of unweighted hypergraphs. Then if a hypergraph in the class has E as the edge set with m edges, then the condition says that

$$m^{r-2} \sum_{(i_1, \dots, i_r) \in E} \frac{1}{\prod_{j=1}^r (D_{i_j} + \bar{D})} = O(1). \quad (1)$$

Note that here the D_i 's are the degrees of the hypergraph vertices in the usual sense of the number of edges incident to the vertex. It is easy to see this condition is satisfied for dense hypergraphs, i.e., for r -uniform hypergraphs with $\Omega(n^r)$ edges, because in this case, $\bar{D} \in \Omega(n^{r-1})$.

The condition can be specialized to the case $r = 2$, where it says that

$$\sum_{i,j} \frac{A_{ij}^2}{(D_i + \bar{D})(D_j + \bar{D})} = O(1). \quad (2)$$

We will show that all metrics satisfy this condition. Also, so do *quasimetrics*. These are weights that satisfy the triangle inequality up to a constant factor (e.g., powers of a metric) and arise in clustering applications [FKKR03, SS73, FK00]. So, as a special case of our main result, we get PTAS's for metrics and quasimetrics for $r = 2$ case. (While PTAS's were known for general r for the dense case, they were not known previously for the metric case.) Our main algorithmic result is the following.

Theorem 3. *There is a PTAS for any class of core-dense weighted MAX- r CSP problems.*

The algorithm and proof are given in Section 3. We will also show (in Section 3.1) that a generalization of the notion of metric for higher r also satisfies our core-dense condition.

Theorem 4. *Suppose for a class of MAX- r CSP problems, the tensors A satisfy the following local density condition:*

$$\forall i_1, \dots, i_r \in V, \quad A_{i_1, \dots, i_r} \leq \frac{c}{r n^{r-1}} \sum_{j=1}^r D_{i_j}$$

where c is a constant. Then there is a PTAS for the MAX- r CSP defined by A .

The condition in the theorem says that no entry of A is “wild” in that it is at most a constant times the average entry in the r “planes” passing through the entry. The reason for calling such tensors “metric tensors” will become clear when we show in Section 3.1 that for $r = 2$, metrics do indeed satisfy this condition. Theorem 4 has the following corollary for “quasi-metrics”, where the triangle inequality is only satisfied within constant factors - $A_{ik} \leq c(A_{ij} + A_{jk})$.

Corollary 5. *There exists a PTAS for metric and quasimetric instances of MAX-CSP.*

2 Fast tensor approximation via sampling

Corresponding to A , there is a r -linear form which for a set of r vectors $x^{(1)}, x^{(2)}, \dots, x^{(r-1)}, x^{(r)}$, is defined as

$$A(x^{(1)}, x^{(2)}, \dots, x^{(r)}) = \sum_{i_1, i_2, \dots, i_r} A_{i_1, i_2, \dots, i_{r-1}, i_r} x_{i_1}^{(1)} x_{i_2}^{(2)}, \dots, x_{i_r}^{(r)}.$$

We will use the following two norms of r -dimensional arrays corresponding to the Frobenius norm and L_2 norm for matrices.

$$\|A\|_F = \left(\sum A_{i_1, i_2, \dots, i_r}^2 \right)^{\frac{1}{2}} \quad \|A\|_2 = \max_{x^{(1)}, x^{(2)}, \dots, x^{(r)}} \frac{A(x^{(1)}, x^{(2)}, \dots, x^{(r-1)}, x^{(r)})}{|x^{(1)}| |x^{(2)}| \dots}.$$

We begin with a proof of Lemma 1 about the existence of a low-rank tensor decomposition.

Proof. If $\|A\|_2 \leq \epsilon \|A\|_F$, then we are done. If not, there are $x^{(1)}, x^{(2)}, \dots, x^{(r)}$, all of length 1 such that $A(x^{(1)}, x^{(2)}, \dots, x^{(r)}) \geq \epsilon \|A\|_F$. Now consider the r -dimensional array

$$B = A - (A(x^{(1)}, x^{(2)}, \dots, x^{(r)})) x^{(1)} \otimes x^{(2)} \otimes \dots \otimes x^{(r)}.$$

It is easy to see that $\|B\|_F^2 = \|A\|_F^2 - (A(x, y, z, \dots))^2$. We may repeat on B and clearly this process will only go on for at most $1/\epsilon^2$ steps. \square

From the proof of Lemma 1, it suffices to find $x^{(1)}, x^{(2)}, \dots, x^{(r)}$ all of length 1, maximizing $A(x^{(1)}, x^{(2)}, \dots, x^{(r)})$ to within *additive error* $\epsilon \|A\|_F / 2$. We will give an algorithm to solve this problem. We need a bit more notation. For any $r - 1$ vectors $x^{(1)}, x^{(2)}, \dots, x^{(r-1)}$, we define $A(x^{(1)}, x^{(2)}, \dots, x^{(r-1)}, \cdot)$ as the vector whose i 'th component is

$$\sum_{i_1, i_2, \dots, i_{r-1}} A_{i_1, i_2, \dots, i_{r-1}, i} x_{i_1}^{(1)} x_{i_2}^{(2)}, \dots, x_{i_{r-1}}^{(r-1)}.$$

Here is the idea behind the algorithm. Suppose $z^{(1)}, z^{(2)}, \dots, z^{(r)}$ are the (unknown) unit vectors that maximize $A(x^{(1)}, x^{(2)}, \dots)$. Since

$$A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, z^{(r)}) = z^{(r)} \cdot A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot),$$

we have

$$z^{(r)} = \frac{A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot)}{|A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot)|}.$$

Thus, if we had $z^{(1)}, z^{(2)}, \dots, z^{(r-1)}$, then we could find $z^{(r)}$. In fact, we can estimate the components of $z^{(r)}$ if we had sufficiently many random terms in the sum $A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot)$. It turns out that we need only $s = O(1/\epsilon^2)$ terms for a good estimate. Now we do not need to know the $z^{(1)}, z^{(2)}, \dots, z^{(r-1)}$ completely; only $s(r-1)$ of their coordinates in total are needed for the estimate. We enumerate all possibilities for the values of these coordinates (in steps of a certain size) and one of the sets of coordinates we enumerate will correspond to the optimal $z^{(1)}, z^{(2)}, \dots, z^{(r-1)}$, whence we get the an estimate of $z^{(r)}$. For each candidate $z^{(r)}$, we can reduce the problem to maximizing an $(r - 1)$ -dimensional tensor and we solve this recursively.

Tensor decomposition

Set $\eta = \epsilon^2/100r\sqrt{n}$ and $s = 10^5r/\epsilon^2$.

1. Pick s random $(r-1)$ -tuples $(i_1, i_2, \dots, i_{r-1})$ with probabilities proportional to the sum of squared entries on the line defined by it:

$$p(i_1, i_2, \dots, i_{r-1}) = \frac{\sum_i A_{i_1, i_2, \dots, i_{r-1}, i}^2}{\|A\|_F^2}.$$

Let I be the set of s $r-1$ tuples picked.

2. For each $i_1, i_2, \dots, i_{r-1} \in I$, enumerate all possible values of $z_{i_1}^{(1)}, z_{i_2}^{(2)}, \dots, z_{i_{r-1}}^{(r-1)}$ whose coordinates are in the set

$$J = \{-1, -1 + \eta, -1 + 2\eta, \dots, 0, \dots, 1 - \eta, 1\}^{s(r-1)}.$$

- (a) For each set of $\hat{z}^{(t)}$, for each $i \in V_r$, compute

$$y_i = \sum_{(i_1, i_2, \dots, i_{r-1}) \in I} A(i_1, i_2, \dots, i_{r-1}, i) \hat{z}_{i_1}^{(1)} \hat{z}_{i_2}^{(2)} \dots \hat{z}_{i_{r-1}}^{(r-1)}.$$

and normalize the resulting vector y to be a unit vector (a candidate for $z^{(r)}$).

- (b) Consider the $(r-1)$ -dimensional array $A(y)$ defined by

$$(A(y))_{i_1, i_2, \dots, i_{r-1}} = \sum_i A_{i_1, i_2, i_3, \dots, i_{r-1}, i} y_i$$

and apply the algorithm recursively to find the optimum

$$A(y)(x^{(1)}, x^{(2)}, \dots, x^{(r-1)}) \quad |x^{(1)}| = \dots = |x^{(r-1)}| = 1$$

to within additive error $\epsilon \|A(y)\|_F/2$. (Note that $\|A(y)\|_F \leq \|A\|_F$ by Cauchy-Schwartz).

3. Output the set of vectors that given the maximum among all these candidates.

We will now analyze the algorithm and consequently prove Theorem 2. We begin by showing the discretization does not cause any significant loss.

Lemma 6. *Let $z^{(1)}, z^{(2)}, \dots, z^{(r-1)}$ be the optimal unit vectors. Suppose $w^{(1)}, w^{(2)}, \dots, w^{(r-1)}$ are obtained from the $z^{(t)}$'s by rounding each coordinate down to the nearest integer multiple of η . Then,*

$$\left| A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot) - A(w^{(1)}, w^{(2)}, \dots, w^{(r-1)}, \cdot) \right| \leq \frac{\epsilon^2}{100} \|A\|_F.$$

Proof. We may write

$$\begin{aligned} & \left| A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot) - A(w^{(1)}, w^{(2)}, \dots, w^{(r-1)}, \cdot) \right| \\ & \leq \left| A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot) - A(w^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot) \right| + \\ & \left| A(w^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot) - A(w^{(1)}, w^{(2)}, z^{(3)}, \dots, z^{(r-1)}, \cdot) \right| \dots \end{aligned}$$

A typical term above is

$$\begin{aligned} & \left| A(w^{(1)}, w^{(2)}, \dots, w^{(t)}, z^{(t+1)}, \dots, z^{(r-1)}, \cdot) - A(w^{(1)}, w^{(2)}, \dots, w^{(t)}, w^{(t+1)}, z^{(t+2)}, \dots, z^{(r-1)}, \cdot) \right| \\ & \leq \left| B(z^{(t+1)} - w^{(t+1)}) \right| \leq \|B\|_2 |z^{(t+1)} - w^{(t+1)}| \leq \|B\|_F \eta \sqrt{n} \leq \|A\|_F \eta \sqrt{n}. \end{aligned}$$

Here, B is the matrix defined as

$$B_{ij} = \sum_{j_1, j_2, \dots, j_t, j_{t+2}, \dots, j_{r-1}} A_{j_1, j_2, \dots, j_t, i, j_{t+2}, \dots, j_{r-1}, j} w_{j_1}^{(1)} \dots w_{j_t}^{(t)} z_{j_{t+2}}^{(t+2)} \dots z_{j_{r-1}}^{(r-1)}$$

The claim follows. \square

Next, we analyze the error incurred by sampling.

Consider an $(r-1)$ -tuple $(i_1, i_2, \dots, i_{r-1}) \in I$ and define the random variables variables X_i for i by

$$X_i = \frac{A_{i_1, i_2, \dots, i_{r-1}, i} w_{i_1}^{(1)} w_{i_2}^{(2)} \dots w_{i_{r-1}}^{(r-1)}}{p(i_1, i_2, \dots, i_{r-1})}.$$

It follows that

$$E(X_i) = A(w^{(1)}, w^{(2)} \dots w^{(r-1)}, \cdot)_i.$$

We estimate the variance:

$$\begin{aligned} \sum_i \text{Var}(X_i) & \leq \sum_i \sum_{i_1, i_2, \dots} \frac{A_{i_1, i_2, \dots, i_{r-1}, i}^2 (w_{i_1}^{(1)} \dots w_{i_{r-1}}^{(r-1)})^2}{p(i_1, i_2, \dots)} \\ & \leq \sum_{i_1, i_2, \dots} \frac{(z_{i_1}^{(1)} \dots z_{i_{r-1}}^{(r-1)})^2}{p(i_1, i_2, \dots)} \sum_i A_{i_1, i_2, \dots, i_{r-1}, i}^2 \\ & \leq \|A\|_F^2. \end{aligned}$$

Consider the y_i computed by the algorithm when all $\hat{z}_{i_t}^{(t)}$ are set to $w_{i_t}^{(t)}$. This will clearly happen sometime during the enumeration. This y_i is just the sum of s i.i.d. copies of X_i , one for each element of I . Thus, we have that

$$E(y) = sA(w^{(1)}, w^{(2)} \dots w^{(r-1)}, \cdot) \quad \text{Var}(y) = E(|y - E(y)|^2) \leq s\|A\|_F^2.$$

We will sketch the rest of the argument. Define

$$\zeta = A(z^{(1)}, z^{(2)}, \dots, z^{(r-1)}, \cdot) \quad \text{and} \quad \Delta = y - s\zeta.$$

From the above, it follows that with probability at least $1 - (1/10r)$, we have

$$|\Delta| \leq 10r\sqrt{s}\|A\|_F.$$

Using this,

$$\begin{aligned} \left| \frac{y}{|y|} - \frac{\zeta}{|\zeta|} \right| &= \frac{|(y|\zeta| - \zeta|y|)|}{|y||\zeta|} \\ &= \frac{1}{|y||\zeta|} |(\Delta + s\zeta)|\zeta| - \zeta(|y| - s|\zeta| + s|\zeta|)| \\ &\leq \frac{2|\Delta|}{(s|y|)} \leq \frac{\epsilon}{50}, \end{aligned}$$

assuming $|y| \geq \epsilon\|A\|_F/100$. If this assumption does not hold, we know that the $|\zeta| \leq \epsilon\|A\|_F/20$ and in this case, the all-zero tensor is a good approximation to the optimum. From this, it can be shown that

$$\|A\left(\frac{y}{|y|}\right) - A\left(\frac{\zeta}{|\zeta|}\right)\|_F \leq \frac{\epsilon}{10}\|A\|_F.$$

Thus, for any $r - 1$ unit length vectors $a^{(1)}, a^{(2)}, \dots, a^{(r-1)}$, we have

$$\left| A(a^{(1)}, a^{(2)}, \dots, a^{(r-1)}, \frac{y}{|y|}) - A(a^{(1)}, a^{(2)}, \dots, a^{(r-1)}, \frac{\zeta}{|\zeta|}) \right| \leq \frac{\epsilon}{10}\|A\|_F.$$

This implies that the optimal set of vectors for $A(y/|y|)$ are nearly optimal for $A(\zeta/|\zeta|)$. Since $z^{(r)} = \zeta/|\zeta|$, the optimal vectors for the latter problem are $z^{(1)}, \dots, z^{(r-1)}$.

The running time of algorithm is dominated by the number of candidates we enumerate, and is

$$\text{poly}(n) \left(\frac{1}{\eta}\right)^{s^2 r} = \left(\frac{n}{\epsilon}\right)^{O(1/\epsilon^4)}.$$

2.1 Optimizing in the constant rank case

We have a vector x of and $A(x, 1 - x)$ is a form of degree r in these variables (as noted in the introduction, any MAX- r CSP problem can be posed as a question of maximizing this form over $(x, 1 - x) \in \{0, 1\}^{2n}$. From the above, it now suffices to deal with the case when A is of constant rank (is the sum of $O(1)$ rank 1 tensors.) The method here is similar to that in [AFKK02] and we give a brief sketch.

Suppose $v^{(1)}, v^{(2)}, \dots, v^{(l)}$ are the $O(1)$ vectors which together span the space of all the vectors occurring in the constant rank expressions for A . Then the value of the polynomial at x is completely determined by the l dimensional vector $\{v^{(i)} \cdot x\}$. One can prove bounds on $\{v^{(i)} \cdot x\}$, so there is box $B \subseteq \mathbf{R}^l$ where the vectors lie. We cut B up into small boxes so that inside each small box, $A(x, 1 - x)$ does not vary much. The number of small boxes will be exponential only in l which is $O(1)$.

We then determine which small boxes are “feasible”. i.e., which can be realized by 0-1 x_i ’s. This is in general an Integer Program, but it can be shown that if we relax this to a Linear Program, any basic feasible solution to the LP has only $O(1)$ fractional variables, so rounding causes only a small error. So, it can be shown that it suffices to check LP feasibility for each small box. Now we take the feasible boxes and in each of them, compute approximately the value of $A(x, 1 - x)$ and take the best of these.

3 Approximation schemes for core-dense MAX- r CSP's

In this section, we give a PTAS for core-dense weighted MAX- r CSP's proving Theorem 3.

We wish to solve the problem

$$\max_{x \in \{0,1\}^n} A(x, x, \dots, x).$$

The algorithm first scales the entries of A to get an r -dimensional tensor B , as follows :

$$B_{i_1, \dots, i_r} = \frac{A_{i_1, \dots, i_r}}{\prod_{j=1}^r \alpha_{i_j}}$$

where $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbf{R}^n$ is defined by $\alpha_j = \sqrt{\bar{D} + D_j}$.

Note that for any x , using the substitution, $x_j = y_j/\alpha_j$, we get

$$A(x, \dots, x) = B(y, \dots, y).$$

Then, applying the sampling algorithm from Section 2 to get a tensor \hat{B} of rank at most k satisfying

$$\|B - \hat{B}\|_2 \leq \frac{\epsilon}{2} \|B\|_F.$$

We then solve the following problem approximately to within additive error $\epsilon \|B\|_F/2$.

$$\max_{y: y_j \in \{0, \alpha_j\}} \hat{B}(y, y, \dots, y).$$

The error of the approximation is bounded by

$$\begin{aligned} \max_{y: y_j \in \{0, \alpha_j\}} |(B - \hat{B})(y, \dots, y)| &\leq \max_{y: |y| \leq |\alpha|} |(B - \hat{B})(y, \dots, y)| \\ &\leq |\alpha|^r \|B - \hat{B}\|_2 \\ &\leq \epsilon |\alpha|^r \|B\|_F. \\ \epsilon \left(\sum_{i=1}^n (\bar{D} + D_i) \right)^{r/2} \left(\sum_{i_1, \dots, i_r} \frac{A_{i_1, \dots, i_r}^2}{\prod_{j=1}^r D_{i_j}} \right)^{1/2} &\leq \epsilon 2^{r/2} c \left(\sum_{i=1}^n D_i \right) \end{aligned}$$

where c is the bound on the core-strength, noting that $\sum_i (\bar{D} + D_i) = 2 \sum_i D_i$. This proves Theorem 3.

Remark We could have carried this out with any “scaling” vector α . The current choice turns out to be useful for the two important special cases here. Note that we are able to add the \bar{D} almost “for free” since we have $\sum_i D_i + \bar{D} \leq 2 \sum D_i$. If we did not add the \bar{D} term, then nodes with very low degrees would cause us trouble.

3.1 Metric tensors

Lemma 7. *Let A be an r -dimensional tensor satisfying the following local density condition:*

$$\forall i_1, \dots, i_r \in V, \quad A_{i_1, \dots, i_r} \leq \frac{c}{rn^{r-1}} \sum_{j=1}^r D_{i_j}$$

where c is a constant. Then A is a core-dense hypergraph with core-strength c .

Proof. We need to bound the core-strength of A . To this end,

$$\begin{aligned}
\sum_{i_1, i_2, \dots, i_r \in V} \frac{A_{i_1, \dots, i_r}^2}{\prod_{j=1}^r (D_{i_j} + \bar{D})} &\leq \frac{c}{rn^{r-1}} \sum_{i_1, i_2, \dots, i_r \in V} \frac{A_{i_1, \dots, i_r} \sum_{j=1}^r D_{i_j}}{\prod_{j=1}^r (D_{i_j} + \bar{D})} \\
&\leq \frac{c}{rn^{r-1}} \sum_{i_1, i_2, \dots, i_r \in V} A_{i_1, \dots, i_r} \sum_{j=1}^r \frac{1}{\prod_{k \in \{1, \dots, r\} \setminus \{j\}} (D_{i_k} + \bar{D})} \\
&\leq \frac{c}{rn^{r-1}} \left(\sum_{i_1, i_2, \dots, i_r \in E} A_{i_1, \dots, i_r} \right) \frac{r}{\bar{D}^{r-1}} \\
&= \frac{c}{(\sum_{i=1}^n D_i)^{r-2}}.
\end{aligned}$$

Thus, the core-strength is at most

$$\left(\sum_{i=1}^n D_i \right)^{r-2} \sum_{i_1, i_2, \dots, i_r \in E} \frac{A_{i_1, \dots, i_r}^2}{\prod_{j=1}^r (D_{i_j} + \bar{D})} \leq c.$$

□

Theorem 4 follows directly from Lemma 7 and Theorem 3. We next prove Corollary 5 for metrics.

Proof. (of Corollary 5) For $r = 2$, the condition of Theorem 4 says that for any $i, j \in V$,

$$A_{i,j} \leq \frac{c}{2n} (D_i + D_j).$$

We will verify that this holds for a metric MAX-2CSP with $c = 2$. When the entries of A form a metric, for any i, j, k , we have

$$A_{i,j} \leq A_{i,k} + A_{k,j}$$

and so

$$\begin{aligned}
A_{i,j} &\leq \frac{1}{n} \left(\sum_{k=1}^n A_{i,k} + \sum_{k=1}^n A_{j,k} \right) \\
&= \frac{1}{n} (D_i + D_j).
\end{aligned}$$

□

We are going to extend now the notion of a metric. A nonnegative real function d defined on $M \times M$ is called *quasimetric* (cf. [MS79], [S03]; [MP00]) if

$$d(x, y) = 0 \text{ exactly when } x = y, d(x, y) = d(y, x), \text{ and } d(x, z) \leq C(d(x, y) + d(y, z)) \quad (3)$$

the last for some positive real number C , and all $x, y, z \in M$. Thus if (3) holds with $C = 1$, then d is a metric on M . The proof of Corollary 5 easily extends to quasimetrics. An interesting property of a quasimetric $d(x, y)$ is that $d(x, y)^a$ is also a quasimetric for every positive real number a (cf. [MS79]). Thus this notion encompasses a large number of interesting distance functions which are not metrics, like the squares of Euclidean distances used in clustering applications.

3.2 Core-dense graphs

We now confine attention to the case of graphs. As we saw already, dense graphs are core-dense graphs, but the converse is not in general true. One simple example is a graph consisting of a dense graph on $\Omega(n^{3/4})$ vertices, up to $O(n)$ edges in the subgraph defined by the rest of the vertices and up to $O(n^{5/4})$ edges from high-degree vertices of the dense subgraph to the rest. We show below that as in this example, in fact there are always “large” dense subgraphs in a core-dense graph.

Theorem 8. *A core-dense graph with m edges contains a dense induced subgraph with $\theta(\sqrt{m})$ vertices.*

Proof. Since G is core-dense, we have

$$\sum_{i,j \in E} \frac{1}{(d_i + \bar{d})(d_j + \bar{d})} \leq c$$

for some c .

We assume that $m < n^2/16c$; otherwise, G itself is a dense graph.

We partition the vertices of the graph into 3 subsets R, S, T according to their degrees:

$$\begin{aligned} R &= \{i \in V : d_i \geq 8\sqrt{m}\} \\ S &= \{i \in V : \frac{\sqrt{m}}{64c} \leq d_i < 8\sqrt{m}\} \\ T &= \{i \in V : d_i < \frac{\sqrt{m}}{64c}\} \end{aligned}$$

We will prove that $|S| \geq \sqrt{m}/8$. Suppose not for a contradiction.

Using the density condition, the number of edges in the subgraph induced by T is at most

$$c \left(\frac{\sqrt{m}}{64c} + \frac{m}{n} \right)^2 < \frac{m}{16}.$$

Similarly, the number of edges between S and T is at most

$$c(8\sqrt{m} + \frac{m}{n}) \left(\frac{\sqrt{m}}{64c} + \frac{m}{n} \right) < \frac{m}{8}.$$

Next, the number of vertices in R is at most $2m/8\sqrt{m} = \sqrt{m}/4$. Thus the total number of edges in the graph induced by R is at most $m/32$. Also, the number of edges between R and S is at most

$$|S| \frac{\sqrt{m}}{4} < \frac{m}{32}.$$

Adding up these bounds, the total number of edges in G not in the subgraph induced by S is at most $m/2$. Therefore, the number of vertices in S is at least

$$\frac{m}{8\sqrt{m}} = \frac{\sqrt{m}}{8}$$

which contradicts our assumption.

Thus G contains an induced subgraph with $\sqrt{m}/8$ vertices and minimum degree $\sqrt{m}/64c$. \square

References

- [AFKK02] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski, *Random Sampling and MAX-CSP Problems*, Proc. 34th ACM STOC (2002), pp. 232-239.
- [AKK95] S. Arora, D. Karger and M. Karpinski, *Polynomial Time Approximation Schemes for Dense Instances of NP-Hard Problems*, Proc. 27th STOC (1995), pp. 284-293; J. Computer and System Sciences 58 (1999), pp. 193-210.
- [F96] W. Fernandez de la Vega, *MAX-CUT has a Randomized Approximation Scheme in Dense Graphs*, Random Structures and Algorithms 8 (1996), pp. 187-198.
- [FK00] W. Fernandez de la Vega and M. Karpinski, *Polynomial time approximation of dense weighted instances of MAX-CUT*, Random Structures and Algorithms 16 (2000), pp. 314-332.
- [FK98] W. Fernandez de la Vega and C. Kenyon, *A randomized approximation scheme for metric MAX-CUT*, Proc. 39th IEEE FOCS (1998), pp. 468-471, final version in J. Computer and System Sciences 63 (2001), pp. 531-541.
- [FKKR03] W. Fernandez de la Vega, M. Karpinski, C. Kenyon and Y. Rabani, *Approximation schemes for clustering problems*, Proc. 35th ACM STOC (2003), pp. 50-58.
- [FKK04] W. Fernandez de la Vega, M. Karpinski and C. Kenyon, *Approximation Schemes for Metric Bisection and Partitioning*, Proc. 15th ACM-SIAM SODA (2004), pp. 499-508.
- [FK96] A. M. Frieze and R. Kannan, *The Regularity Lemma and Approximation Schemes for Dense Problems*, Proc. 37th IEEE FOCS (1996), pp. 12-20.
- [FK97] A. M. Frieze and R. Kannan, *Quick Approximation to Matrices and Applications*, Combinatorica 19 (2) (1999), pp. 175-120.
- [FKV98] A. M. Frieze, R. Kannan and S. Vempala, *Fast Monte-Carlo Algorithms for Finding Low-Rank Approximations*, Proc. 39th IEEE FOCS (1998), pp. 370-378.
- [GGR96] O. Goldreich, S. Goldwasser and D. Ron, *Property Testing and its Connection to Learning and Approximation*, Proc. 37th IEEE FOCS (1996), pp. 339-348; J. ACM 45 (1998), pp. 653-750.
- [I99] P. Indyk, *A Sublinear Time Approximation Scheme for Clustering in Metric Spaces*, Proc. 40th IEEE FOCS (1999), pp. 154-159.
- [MS79] R. Macias and C. Segovia, *Lipschitz functions on spaces of homogenous type*, Advances in Mathematics **33** (1979), pp. 257-270.
- [Mc01] F. McSherry, *Spectral Partitioning of Random Graphs*, FOCS 2001, pp.529-537
- [MP00] R. R. Mettu and C. G. Plaxton, *The Online Median Problem*, Proc. 41st IEEE FOCS (2000), pp. 339-348.
- [S03] S. Semmes, *A brief introduction to Gromov's notion of hyperbolic groups*, Mathematics, arXiv: math CA/021341 (2003), pp. 1-10.

[SS73] P. H. Sneath and R. R. Sokal, *Numerical Taxonomy*, Freeman, London, 1973.