Final Report: Lower bounds on the size of semidefinite programming relaxations

Weihao Kong

June 14, 2017

1 Problem Formulation and Main result

An instance $\xi$ of Max-CSP problem is defined as

$$\text{maximize } \sum P_i(x)$$
$$x \in \{0,1\}^n$$

where $P_i : \{0,1\}^n \to 0,1$ are all predicates. We use $\xi(x)$ to denote $\sum P_i(x)$ and $\max(\xi)$ to denote the optimal value of the instance. A Max-CSP problem is defined as a set of the Max-CSP instances.

Definition 1. The degree $d$ SOS upperbound for function $f$, $\text{sos}_d(f)$, is defined to be smallest $c$ such that $c - f$ has a degree $d$ SOS proof.

Definition 2. The subspace $U$ SOS upperbound for function $f$, $\text{sos}_U(f)$, is defined to be smallest $c$ such that $c - f = \sum f_i^2$ where $f_i \in U$.

$sos_d(f)$ is the upperbound of $f$ given by a degree $d$ sos algorithm. $sos_U(f)$ is the upperbound of $f$ given by a subspace $U$ sos algorithm. Now for a Max-CSP problem, we need the following definition to capture how good approximation does a subspace U sos algorithm give.

Definition 3. We say that the subspace $U$ achieves $(c,s)$-approximation of problem $\Pi$ if for any $\xi \in \Pi$, $\max(\xi) \leq s \Rightarrow sos_U(\xi) \leq c$.

The authors claim that any SDP formulation with instance oblivious constraints actually is equivalent to computing $sos_U$ for a certain $U$ where the running time is $\text{dim}(U)$. Hence we can focus on showing that $U$ must have large dimension in order for $sos_U(\xi)$ to be close to $\max(\xi)$. Indeed, the following theorem states that if polynomial sos need high degree to achieve good approximation, no $U$ with much smaller dimension can achieve the same approximation.

Theorem 1 (Main Theorem). Let $\Pi$ be Max-CSP problem and let $\Pi_m$ be the set of instances of $\Pi$ on $n$ variables. Suppose that for some $m, d \in N$, the subspace of degree-$d$ functions $f : \{0,1\}^m \to R$ fails to achieve a $(c,s)$-approximation for $\Pi_m$. For all $n \geq 2m$, every subspace $U$ of functions $f : \{0,1\}^n \to R$ with $\text{dim}(U) = nd^{18}$ fails to achieve a $(c,s)$-approximation for $\Pi_n$.

Before going further to prove the main theorem, let’s see what would happen if $U$ achieves $(c,s)$-approximation for problem $\Pi$ and has dimension $d$. Given any instance $\xi \in \Pi$, the function $c - \xi$ has a subspace $U$ sos proof: $c - \xi = \sum f_i^2$ where $f_i \in U$. Let $\{g_i\}, i = 1, \ldots d$ be a set of orthogonal basis of subspace $U$. Define a matrix $A$ such that $f_i = \sum_j g_j A_{ji}$.

Define matrix $B \in R^{2n \times d}$ such that $B(x,i) = g_i(x)$. $c - \xi(x)$ can be written as $\text{tr}(BA'BA') = \text{tr}(AA'BB')$ which means there exists two $d \times d$ PSD matrix $P = AA', Q = BB'$ such that $c - \xi(x) = \text{tr}(PQ)$. Notice that $P$ is a function of $\xi$ and $Q$ is a function of $x$, so we also use $P(\xi)$ and $Q(x)$ to denote the two PSD matrices. Let’s define matrix $M_{\Pi}^c(\xi,x) = c - \xi(x)$, by the definition of $M_{\Pi}^c$ and previous observation, $M_{\Pi}^c(\xi,x) = \text{tr}(P(\xi)Q(x))$ where $P(\xi), Q(x)$ are $d \times d$ PSD matrices. Now we introduce a useful definition called PSD rank of a matrix.

Definition 4. Let $M \in R^{p \times q}$ be a matrix with non-negative entries. We say that $M$ admits a rank-$r$ psd factorization if there exist positive semidefinite matrices $\{P_i : i \in [p]\}, \{Q_j : j \in [q]\} \subset S^+_r$ such that $M_{i,j} = \text{tr}(P_i Q_j)$ for all $i \in [p], j \in [q]$. We define $r_{psd}(M)$ to be the smallest $r$ such that $M$ admits a rank-$r$ psd factorization. We refer to this value as the PSD rank of $M$. 

Since we have constructed a rank \( d \) psd factorization of matrix \( M^c_{\Pi} \). We conclude that \( rk_{psd}(M^c_{\Pi}) \leq d \) assuming \( U \) achieves \((c, s)\)-approximation of \( \Pi \). In order to show the hardness result, we will dedicate the rest of the report for proving the psd rank of matrix \( M^c_{\Pi} \) is large.

# 2 Main Lemma

We will prove a stronger result by bounding the psd rank of a submatrix of \( M \) from below. Given a function \( f : \{0, 1\}^m \to R_+ \), define a \( (\binom{n}{m}) \times 2^n \) matrix \( M^f \) where \( M^f(S, x) = f(x, S) \). Let \( deg_{sos}(f) \) be the smallest \( d \) such that \( f \) has a degree-\( d \) SOS proof.

**Lemma 1** (Main Lemma). For every \( m \geq 1 \) and \( f : \{0, 1\}^m \to R_+ \), there exists a constant \( C > 0 \) such that for \( n \geq 2m \), \( rk_{psd}(M^f) > n^{deg_{sos}(f)/8} \).

Now we are ready to prove the main theorem.

**Proof of the Main Theorem.** Suppose for some \( n \geq 2m \), there is subspace \( U \) with \( \dim(U) \leq n^{d/8} \) achieves \((c, s)\)-approximation of \( \Pi_n \). Then by the previous argument, the matrix \( M^c_{\Pi_n} \) has psd rank less than or equal to \( n^{d/8} \). Since degree \( d \) SOS fails to achieve a \((c, s)\) approximation of \( \Pi_m \), there must be a \( \xi \) such that \( \max(\xi) \leq s \) and \( deg_{sos}(c - \xi(x)) > d \). By Lemma 1, for \( n \geq 2m \) \( rk_{psd}(M^c_{\Pi_m}) \geq n^{d/8} \). Since \( M^c_{\Pi_m} \) is a submatrix of \( M^c_{\Pi_n} \), we conclude that \( rk_{psd}(M^c_{\Pi_n}) \geq n^{d/8} \) and there is a contradiction. Actually for the submatrix property to hold, we need some assumptions on the Max-CSP problem \( \Pi \). Without formally state the assumption, we just verify this property for Max Cut and Max 3-SAT here. A max cut problem on a graph with \( n \) vertices is valid even if there are only \( m \) nodes which are incident to some edges. A Max 3-SAT on \( n \) variable is valid even if there are only \( m \) variables involved in the formula.

Now we give a plan to prove the main lemma. First there must be a degree \( d = deg_{sos}(f) - 1 \) pseudo distribution \( D \) such that \( E(D(x)f(x)) < -1 \). Then we define the following linear functional on matrices \( M^f : \binom{[n]}{m} \times \{0, 1\}^n \to R \):

\[
L_D(M^f) = E_{|S|=m} E_x D(x, S) M^f(x, S, x)
\]

By the definition, suppose \( L_D(M^f) < -1 \). It is known that we can find a set of matrices \( \{P(S), Q(x)\} \) such that \( M^f(S, x) = tr(P(S)Q(x)) \) and \( \|P(S)\| \cdot \|Q(x)\| \leq rk_{psd}(M^f)^2 \leq n^{d/4} \). Define the quantum relative entropy of \( X \) with respect to \( Y \) to be the quantity \( S(X|Y) = tr(X \cdot (log X - log Y)) \). Then the relative entropy between \( Q = \frac{1}{tr(Q)} E_x e_x e_x^T \otimes Q(x) \) and uniform distribution \( U = \frac{1}{tr(U)} \) is small (roughly \( log rk_{psd}(M^f) \)). Given that, we have the following proposition showing that it can be approximated by a low degree polynomial.

**Proposition 1** (Low degree polynomial approximation). Let \( F \) be a symmetric matrix. Then, for every \( \epsilon > 0 \), there exists a degree-\( k \) univariate polynomial \( p \) with \( k \leq (1 + S(Q||U)) \cdot \|F\|/\epsilon \) such that the \( Q = \frac{1}{tr(F)} p(F)^2 \) satisfies

\[
tr(FQ) = tr(FQ) + \epsilon.
\]

Using the low degree polynomial approximation, we can now show that \( L_D(M^f) > -1 \). Let \( F(x) = E_{|S|=m} D(x, S) P(S) \) and \( F = \sum_x e_x e_x^T \otimes F(x) \)

\[
L_D(M^f) = E_{|S|=m} E_x D(x, S) M^f(S, x)
= E_{|S|=m} E_x D(x, S) tr(P(S)Q(x)) = tr(FQ)
= tr(FQ) - \epsilon = E_x E_x P(S) p(F(x))^2 - \epsilon
\]

The degree of \( p(F(x))^2 \) can be much larger than \( d \), but notice that for a fixed set \( S \), the degree of \( p(F(x)) \) in terms of the variables in \( S \) is typically smaller than \( d \). The probability that the degree in terms of the variables in \( S \) is larger than \( d \) is on the order of \( O(\frac{1}{(n-m)^D}) \). Since \( D \) is a degree-\( d \) pseudo distribution, \( E_x P(S) p(F(x))^2 \) must be non-negative unless the \( \frac{1}{(n-m)^D} \) probability event happens. In that case, the pseudo expectation can be \( -\|F_S\| \) which is larger than \( -rk_{psd}(M^f) \). Hence when \( rk_{psd}(M^f)^2 = \frac{1}{(n-m)^D} \), we have find \( L_D(M^f) \) is both smaller than \(-1 \) and larger than \(-1 \).