3.1   SA Review

Recall: Each feasible solution α to the degree-d SA LP corresponds to a linear functional $E_\alpha : [x_1, ..., x_n]_d \to \mathbb{R}$. These functionals are the *pseudo-distributions*. The set of all such functions is $\mathcal{E}_d(\mathcal{H})$ where $\mathcal{H} = \{Ax - b \geq 0, 1 \geq x \geq 0, 1 \geq 0\}$ are our original inequalities (which are linear).

Lemma 3.1 $\mathcal{H} = \{Ax - b \geq 0, 1 \geq x \geq 0, 1 \geq 0\}$ *and* $\mathcal{F}$ *as before. Then*

$$\min\{E(P) : E \in \mathcal{E}_d(\mathcal{H})\}$$

equals $\max\{c_0 : \text{there is a degree-}d \text{ SA derivation of } P \geq c_0 \text{ from } \mathcal{H}, \mathcal{F}\}$.

3.1.1   How hard is it to find SA proofs?

If the LP is infeasible then we can find a SA refutation of degree-d in time $n^{o(d)}$ since we can make a system of n^d linear equations in the coefficients and solve.

If the degree d SA LP has a feasible solution solving the $n^{o(d)}$ linear constraints takes time $poly(d)$.

There was also a comment about size automatizability, but it is not clear what the notes are trying to say there.

3.1.2   Proof Systems/ LP tightening related to SA

1. Dynamic SA (LS_d which stands for Lovasz-Schrijver Lift and Project SA): normal SA is a one shot system. The dynamic variant takes the original LP, lifts to degree-d, then project down to your original dimension. If you did not get the solution that you want yet, add some of the higher degree inequalities to your original set and repeat.

   Formally a proof is a sequence of $P_1 \geq 0, P_2 \geq 0, ...$ where $P_i \geq 0$ is an axiom, a given inequality, or derived by an inference. We say:
   $$p_1 \geq 0 \cdots p_m \geq 0 \vdash_D^H P \geq 0$$
   which means that $P_1 \geq 0 \cdots P_m \geq 0$ *derive* $P \geq 0$.

   Given $P_1 \geq 0, ..., P_m \geq 0$ as initial inequalities and $x_i^2 - x_i \geq 0$ as axioms an **inference** is a set $\{Q_i\}$ of non-negative juntas with
   $$Q_0 + P_1 Q_1 + \cdots + P_m Q_m = P.$$

   This is a derivation of $P \geq 0$. If we draw the derivation DAG, we get a start since this is a one-shot proof system. If instead we apply inference rules repeatedly, we would be able to prove linear inequalities that SA of lower degree can prove.

2. Nullsatz (polynomial calculus, PC): over any field, not just $\mathbb{R}$, were we only allow equalities. Suppose we have polynomial equalities $P_1 = 0, ..., P_m = 0$ (which correspond to the pair of inequalities $P_i \geq 0$ and $P_i \leq 0$) as well as $x_i^2 - x_i = 0$. Let

$$\sum P_i Q_i + \sum (x_i^2 - x_i) R_i = P$$

where Q_i, R_i are arbitrary polynomials. When we want to derive $P = 0$.

When applying Nullstellensatz for unsatisfiable CNF we can consider the k-CNF $F = C_1 \wedge \cdots \wedge C_m$ over $x_1, ..., x_n$ and convert each clause C_i to an equality as follows (an example will suffice):

$$x_1 \vee \bar{x_2} \vee x_3 \implies (1 - x_1)x_2(1 - x_3) = 0.$$

Observe that the only way the polynomial would *not* be satisfied is if the clause is false.

Theorem 3.2 *If $f = C_1 \wedge \cdots \wedge C_m$ has a degree d Nullstellensatz refutation, then it has a degree d SA refutation.*

Theorem 3.3 *There are unsatisfiable k-CNFs that have small degree SA refutations but no small degree Nullstellensatz refutation.*

From the above, it seems as though SA is strictly stronger than Nullstellensatz, but I am not sure if this is the case. Some proof systems can be incomparable. A degree-d Nullstellensatz proof can be found in time $n^{o(d)}$ since we have a linear system of $n^{o(d)}$ equations and variables.

3. PC (Polynomial Calculus): this is the dynamic variant of Nullstellensatz.

4. Cutting Planes: this is *not* a lift and project system (unlike SA and Nullsatz). The rules are

   (a) Add non-negative linear combinations of previously derived inequalities.
   (b) Perform division with rounding. That is, for

   $$\sum_{i=1}^{n} a_i x_i \geq a_0$$

   where $a_1, ..., a_n$ *is* divisible by k and a_0 is not, we can derive the inequality:

   $$\sum_{i=1}^{n} \frac{a_i}{k} x_i \geq \lceil a_0/k \rceil.$$

   It is unknown if this system is automatizable, but this is widely used in optimization applications for solving LP.

3.2   SOS

3.2.1   The Basics

We want to start with a quadratic integer program and relax it to an SDP program (this is going to be a generalization of degree-d SA, denoted SA_d).

Let as take only vectors over $\mathbb{R}$, where vectors $\mathbf{u}, \mathbf{v}$ are column vectors, $\mathbf{u}^T \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle$ is a scalar, $\mathbf{uv^T} = \mathbf{u} \otimes \mathbf{v}$ is a tensor product.

Note:
$$\mathbf{x}^T A \mathbf{x} = (\mathbf{x} \otimes \mathbf{x}) \cdot A$$
where $\cdot$ is the dot product of the columns of the matrices as vectors. Further
$$\mathsf{trace}(AB) = \mathsf{trace}(BA)$$
and if A, B are symmetric then
$$\mathsf{trace}(AB) = A \cdot B$$
where the matrix dot product $A \cdot B = a_{1,1}b_{1,1} + \cdots + a_{n,n}b_{n,n}$ which views A and B as vectors.

Here are some norms you should be familiar with: $\|\mathbf{v}\|_p = (\sum_{i=1}^n |v_i|^p)^{\frac{1}{p}}$ and $\|\mathbf{v}\|_\infty = \max_{i \in [n]} |v_i|$. If A, B are matrices with entries $A_{i,j}$ and $B_{i,j}$ then the Kronecker product of A, B, denoted $A \otimes B$ is a matrix with entries $(A \otimes B)_{ii,jj} = A_{i,j}B_{i,j}$. Further $A^{\otimes k} = A \otimes \cdots \otimes A$ k times.

If A, B are square matrices then $\mathsf{det}(AB) = \mathsf{det}(A) \cdot \mathsf{det}(B)$. Further with triangular matrices the determinant is just the product of the diagonal entries.

Theorem 3.4 *If A is an $n \times n$ matrix then the sum of the eigenvalues of A is equal to $\mathsf{trace}(A)$ and the product of the eigenvalues is equal to $\mathsf{det}(A)$.*

3.2.2 PSD Matrices

Theorem 3.5 *Let A be an $n \times n$ symmetric matrix. Then the following are all equal:*

1. *$A \succeq 0$ (A is PSD)*

2. *All eigenvalues of A are non-negative*

3. *The determinant of all upper left sub-matrices of A is non-negative (this requires A to be symmetric to work)*

4. *The quadratic polynomial $P_A(x) = \sum A_{i,j} x_i x_j$ is a SOS i.e.*

$$\exists L_1, ..., L_n : P_A = \sum_i (L_i)^2$$

   where each L_i is linear.

5. *$A = U^T U$ where U is upper triangular (Cholesky decomposition)*

6. *There are correlated random variables $x_1, ..., x_m$ such that for all i, j, $\mathbb{E}[x_i x_j] = A_{i,j}$ and x_i is has distribution approximately $\mathcal{N}(0, A_{i,j})$ the approximately is really quite suspicious...*