# Sum of Squares Hierarchy



## Non-negativity of a polynomial

Input: An n-variate polynomial $f$.

Goal: Either certify that $f(x) \geq 0 \ \forall x \in \{0,1\}^n$

or find $x \in \{0,1\}^n$ s.t. $f(x) < 0$.

**Max-Cut** $\quad -\sum_{(i,j)\in E} (x_i - x_j)^2 + C \geq 0 \quad \longleftrightarrow$ cut given by $x_1 \ldots x_n$ is of size $\leq C$.

Similarly, can express Max-3SAT, Max-3LIN, ....

as non-negativity of poly problem.

**This lecture** A hierarchy of SDP algorithms, attempting to certify $f(x) \geq 0 \ \forall x \in \{0,1\}^n$.

- The $(2d)$th algorithm $d = 1, 2, \ldots, n$ runs in time $n^{O(d)}$
- For $d = n$, algo will definitely either certify $f(x) \geq 0 \ \forall x \in \{0,1\}^n$ or find $x \in \{0,1\}^n$ s.t. $f(x) < 0$.
- For $d < n$, algo will either certify $f(x) \geq 0 \ \forall x \in \{0,1\}^n$ or fail but will give useful information for approx.

# Sum of Square Proof

If $f(x)=\sum g_i(x)^2 \quad \forall x\in\{0,1\}^n$ then

$f(x)\geq 0 \quad \forall x\in\{0,1\}^n$

If all the $g_i$'s are of degree $\leq d/2$ then we say that they form a degree $\leq d$ SoS proof.

<u>Lemma</u> If $f(x)\geq 0 \ \forall x\in\{0,1\}^n \Rightarrow$ degree $\leq 2n$ SoS proof.

<u>Pf</u> $f(x_1 \dots x_n) = x_1^2 \cdots x_n^2 f(1 \dots 1) + \dots + (1-x_1)^2 \cdots (1-x_n)^2 f(0 \dots 0)$.

<u>Lemma</u> Degree-d SoS proofs can be found in $n^{O(d)}$ time via SDP. The variables are the coefficients of the $g_i$'s.

$g_i(x_1 \dots x_n) \Leftrightarrow$ coefficients vector $\overbrace{( \dots )}^{n^{O(d)}} = g_i$

$= ( \dots g_i \dots )(m_x)$

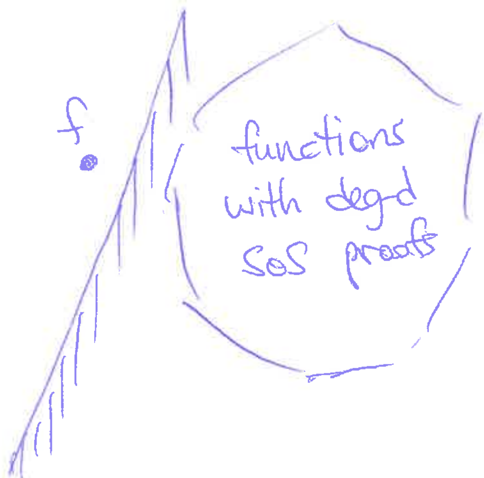
monomials vector $(1 \ x_1 \ x_2 \dots x_1x_2 \dots) = m_x$

$\sum_i g_i g_i^T$ = PSD matrix

$\sum_i g_i(x)^2 = \sum_i ( \dots m_x \dots )(g_i)( \ g_i \ )(m_x)$

# Pseudo-distributions (Dual of SoS proofs)

If SDP algorithm failed to find degree-$d$ SoS pf, then




there must exist a separating hyperplane, i.e., $\mu : \{0,1\}^n \to \mathbb{R}$ s.t.

* $$\sum_{x \in \{0,1\}^n} f(x)\,\mu(x) < 0$$

* $\forall g$ with deg-$d$ SoS proof,

$$\sum g(x)\mu(x) \geq 0$$

---

If $\mu(x) \geq 0 \ \forall x \in \{0,1\}^n$ then $\dfrac{\mu(x)}{\sum_{x \in \{0,1\}^n} \mu(x)}$ is a probability distribution.

$\Rightarrow$ By drawing $x \sim \mu$ get $\mathbb{E} f(x) < 0$

In general, $\mu$ may have negative values ...

<u>Def</u> $\mu$ is a <u>degree-$d$ pseudo-distribution</u> if

$\forall g$ of $\deg \le d/2$

$$\sum_{x \in \{0,1\}^n} g(x)^2 \mu(x) \geq 0$$

Define $\tilde{\mathbb{E}}_{\mu} f(x) = \sum_{x \in \{0,1\}^n} f(x)\mu(x)$. <u>pseudo-expectation</u>

---

While we can't draw $x = (x_1 \cdots x_n)$ from $\mu$, $\mu$ does give reasonable answers to questions like: what's the correlation between $x_i$ and $x_j$? — i.e., $\tilde{\mathbb{E}}_{\mu} x_i x_j$ (similarly for other correlations)

---

- <u>Linearity</u> $\tilde{\mathbb{E}}_{\mu} f+g = \tilde{\mathbb{E}}_{\mu} f + \tilde{\mathbb{E}}_{\mu} g$
- <u>Variance</u> $\tilde{\mathbb{E}}_{\mu} (f - \tilde{\mathbb{E}}_{\mu} f)^2 \geq 0$

Many inequalities only involve low degree polynomials, e.g., Cauchy-Schwarz, Hölder, hypercontractivity, and hence extend to pseudo-expectations.

## Framework for approximation algorithms

- Decide on $d$.
- Use SDP to find degree-$d$ pseudo distribution $\mu$.
- Use $\mu$ to find $x \in \{0,1\}^n$ with $f(x) \lesssim 0$

---

**Lemma** For every degree-2 pseudo-dist $\mu$ over $\{0,1\}^n$, $\rho$ defined next is a probability distribution over $\mathbb{R}^n$ with

* $\mathbb{E}_{\rho} X_i = \tilde{\mathbb{E}}_{\mu} X_i$
* $\mathbb{E}_{\rho} X_i X_j = \tilde{\mathbb{E}}_{\mu} X_i X_j$ $\quad \forall i,j.$

To sample $r_1 \dots r_n \in \mathbb{R}$ from $\rho$:

* Pick gaussian $g \in \mathbb{R}^n$

  Let the covariance matrix of $\mu$ be

$$i \rightarrow \begin{pmatrix} \cdots \tilde{\mathbb{E}}_{\mu} X_i X_j - \tilde{\mathbb{E}}_{\mu} X_i \tilde{\mathbb{E}}_{\mu} X_j \end{pmatrix}$$

$\leftarrow$ This is a PSD matrix so can be written as $VV^T$.

* Let $r_i = \tilde{\mathbb{E}}_{\mu} X_i + V_i \cdot g.$

# Proof of Lemma

$$\mathbb{E}_{p} x_i = \mathbb{E}_{g}\ \tilde{\mathbb{E}}_{\mu} x_i + v_i \cdot g = \tilde{\mathbb{E}}_{\mu} x_i + \overbrace{\mathbb{E}_{g} v_i g}^{\approx 0} = \tilde{\mathbb{E}}_{\mu} x_i$$

$$\mathbb{E}_{p} x_i x_j = \mathbb{E}_{g} \left(\tilde{\mathbb{E}}_{\mu} x_i + v_i g\right)\left(\tilde{\mathbb{E}}_{\mu} x_j + v_j g\right)$$

$$= \tilde{\mathbb{E}}_{\mu} x_i\, \tilde{\mathbb{E}}_{\mu} x_j + \underbrace{\mathbb{E}_{g} v_i g}_{\approx 0}\ \tilde{\mathbb{E}}_{\mu} x_j + \underbrace{\mathbb{E}_{g} v_j g}_{\approx 0} \cdot \tilde{\mathbb{E}}_{\mu} x_i + \mathbb{E}_{g} v_i g \cdot v_j g$$

$$= \tilde{\mathbb{E}}_{\mu} x_i\, \tilde{\mathbb{E}}_{\mu} x_j + \underbrace{\mathbb{E}_{g} \sum_{l,k} v_{i,l} g_l\, v_{j,k} g_k}_{\boxed{\mathbb{E}\, g_l g_k = \begin{cases} 1 & l = k \\ 0 & l \neq k \end{cases}}}$$

$$= \tilde{\mathbb{E}}_{\mu} x_i \cdot \tilde{\mathbb{E}}_{\mu} x_j + \underbrace{\sum_{l} v_{i,l} \cdot v_{j,l}}_{\boxed{v_i v_j^T = \tilde{\mathbb{E}}_{\mu} x_i x_j - \tilde{\mathbb{E}}_{\mu} x_i \cdot \tilde{\mathbb{E}}_{\mu} x_j}}$$

$$= \tilde{\mathbb{E}}_{\mu} x_i x_j$$

$\square$

# Max-Cut approximation algorithm

* Set $d=2$
* Find $\mu$ using SDP. Construct $\rho$ over $\mathbb{R}^n$ with matching pairwise correlations.
* Sample $r_1 - - r_n$ from $\rho$.
* Output $\text{Sign}(r_1 - \frac{1}{2}), - -, \text{Sign}(r_n - \frac{1}{2})$.

To analyze need to solve the following question in probability:

If $X, Y$ normal random variables with $\overbrace{\text{co-variance } \rho}^{\mathbb{E}XY = \rho}$,

then $\mathbb{E}\,\text{Sign}X \cdot \text{Sign}Y = ?$

( Take $X = 1-2r_i$ to translate $0,1$ $r_i$'s to $1,-1$ $X, Y$ )
$Y = 1-2r_j$
$(i,j) \in E$