Lecture 6: SoS and Gaussian mixtures:

Let $x_1, \ldots, x_n \in \mathbb{R}^d$ be samples from

$$q = \frac{1}{k} \sum_{j=1}^{k} N(\mu_i, \text{Id})$$

Define $\Delta \equiv \min_{i \neq j} \|\mu_i - \mu_j\|_2$

$$N \equiv \frac{n}{k} \quad (\# \text{pts in each component})$$

Let $t$ (SoS degree) be power of 2. Suppose

$$\Delta \gg t^t k^{t}.$$  

\underline{SoS program}

Variables: $a_1, \ldots, a_n$ (1-dimensional)

$\mu$ (d-dimensional)

Constraints:

1. $a_i^2 = a_i : \quad (\text{Boolean indicators})$

2. $\sum_{i=1}^{n} a_i = N$ (selects out enough points for one component)

3. $\frac{1}{N} \sum_{i=1}^{n} a_i x_i = \mu$ (\(\mu\) is empirical mean of points selected)

4. $\frac{1}{N} \sum_{i=1}^{n} a_i (u, x_i - \mu)^t \leq 2t^{t/2} \|u\|^t$ (selected points have Gaussian moment bounds)

Let $S_j c(n)$ denote samples from $N(\mu_i, \text{Id})$

For convenience, we will pretend $|S_j| = N$ exactly $\forall j$.

* Seems like infinite constraints... see below for how to quantify over all $u \in S^{d-1}$.
For now, assume $d=1$.

So constraint 4) becomes $\frac{1}{n} \sum_{i} a_i (x_i - m)^t \leq 2^{t/2}$.

**Warmup lemma:** Let $S = S_j, \mu' = \mu_j$ for any $j \in [k]$.

There is a deg-$O(t)$ proof that:

$$(\sum_{i \in S} a_i)^t (\mu - \mu')^t \leq 2^t (\sum_{i \in S} a_i)^{t-1} N \cdot t^{t/2}.$$

**Proof:**

$$(\sum_{i \in S} a_i)^t (\mu - \mu')^t = \left( \sum_{i \in S} a_i (\mu - \mu') \right)^t$$

$$= \left( \sum_{i \in S} a_i \left[ (x_i - \mu) - (\mu' - x_i) \right] \right)^t$$

by degree 1

Hölder's inequality (SOS-able)

$$(\sum_{i \in S} b_i c_i)^t = \left( \sum_{i \in S} b_i^{\frac{t}{L_p}} c_i^{\frac{t}{L_q}} \right)^t \leq \left( \sum_{i \in S} b_i \right)^{t-1} \left( \sum_{i \in S} c_i \right)^t$$

for $p = \frac{t}{L_p}, q = t$

so $\frac{1}{t} + \frac{1}{q} = 1$

$$\leq \left( \sum_{i \in S} a_i \right)^{t-1} \left( \sum_{i \in S} a_i \left[ (x_i - \mu) - (\mu' - x_i) \right] \right)^t$$

$$(a-b)^t \leq 2^t (a-b)^{t-1}$$

$$(a \leq b \leq c \leq d)$$

Note: $\sum_{i \in S} a_i (x_i - m)^t \leq \sum_{i \in S} a_i (m - x_i)^t \leq N \cdot 2^t t^{t/2}$

moment bound, i.e. constraint 4
\[ \sum_{i \in S^0} a_i (m' - x_i)^t \leq \sum_{i \in S^0} (m' - x_i)^t \leq N \cdot 2^{t/2} \]

Bodeanuy, i.e. constraint \( i \) assuming \( t \)th empirical moment of actual Samples from Component Concentrate

\[ \left( \sum_{i \in S^0} a_i \right)^t (m - m')^t \leq 2^t \left( \sum_{i \in S^0} a_i \right)^{t-1} \cdot N \cdot t^{t/2}. \]

So,

Note, if we could "divide on both sides" and take \( t \)th roots, we would get

\[ |m - m'| \leq \left( \frac{1}{N} \sum_{i \in S^0} a_i \right)^{-1/4} \cdot \sqrt{t} \quad (8) \]

i.e. if overlap between our points (chosen by \( a_i \)) and true points in component \( S' \) is large, then our \( m \) is close to the mean of \( S' \).

Claim: If \( a_i \)'s were real indicators of a set \( S \) satisfying \( (8) \) for every center \( m' = m_j \), then
Component $S_j$ with largest overlap with $S$ satisfies $|S \cap S_j| = (1 - \delta)N$ for $\delta \leq k^{t/2} \cdot o(1/\delta)^t$.

**Proof:** Note $1 - \delta \geq \frac{1}{k}$ by averaging, and $|S \cap S_j| \geq \frac{\delta}{k} \cdot N$ for some $j \neq j'$. So:

$$|\mu_j - \mu| \leq (1 - \delta)^{-1/4} \sqrt{t} \leq k^{t/4} \sqrt{t} \ll \Delta/2,$$

So $|\mu_j - \mu| > \frac{\delta}{2}$. Thus, by (8) applied to comp. $j$,

$$\frac{\Delta}{2} < |\mu_j - \mu| \leq \left(\frac{\delta}{k}\right)^{-1/4} \sqrt{t},$$

so $S^{t/4} \geq \frac{k^{t/4} \sqrt{t}}{\Delta} \ll 1$,

and thus $S \leq k^{t/2} \cdot o(1/\delta)^t$ as claimed. $\square$

i.e. $a_i$'s must have $1 - o(1)$ overlap with some component!

---

**SOS's:**

- SOS version of claim 1? closely related
- Rounding SOS solution? related
- $d > 1$?
Issue with Claim 1 is it breaks symmetry across clusters. Makes it unclear how to round.

Claim 2 (symmetric version of Claim 1 — still not SAS):
If $a_i$'s are indicator of $S$, then
\[
\sum_{j=1}^{k} \left( \frac{|S_j \cap S|}{N} \right)^2 \geq 1 - k^2 + \frac{4}{k} \cdot O(1/k^t)
\]

Note: This implies Claim 1. Define $c_j = \frac{|S_j \cap S|}{N}$
so $\sum_j c_j = 1$. Thus
\[
1 - k^2 + \frac{4}{k} \cdot O(1/k^t) \leq \sum_j c_j \leq (\max_j c_j) \cdot \sum_j c_j = \max_j c_j,
\]
i.e. exists $j$ s.t. $\frac{1}{N} |S_j \cap S| \geq 1 - k^2 + \frac{4}{k} \cdot O(1/k^t)$,
which recovers Claim 1 w/ extra (but unimportant) $k$ factor.

Proof: Define $c_j = \frac{|S_j \cap S|}{N}$. Then
\[
1 = \left( \sum_j c_j \right)^2 = \sum_j c_j^2 + 2 \sum_{i<j} c_i c_j
\]
we'll show these are small
\[
\frac{1}{\Delta} \leq C_i C_j \leq \frac{C_i}{\Delta} \frac{|\mu_i - \mu|}{\Delta} + \frac{C_j}{\Delta} \frac{|\mu_j - \mu|}{\Delta}
\]

Recall by warmup lemma, in particular (6),
\[
|\mu_i - \mu| \leq C_i^{-1/4} \sqrt{t}, \text{ so}
\]
\[
C_i \leq \frac{\sqrt{t}}{\Delta},
\]

thus \[
\sum_{i \neq j} C_i C_j \leq k^2 (\frac{\sqrt{t}}{\Delta})^t
\]
as desired. \( \square \)

Next: “SoS-ize” Claim 2

Claim 3 (SoS version of Claim 2):
For any deg-\( t \) pseudodistribution over \( \{a_i\}, \mu, \)
\[
\mathbb{E} \left[ \sum_{j=1}^{k} \left( \frac{1}{N} \sum_{i \in S_j} a_i \right)^2 \right] \geq 1 - k^{2 + \frac{t}{2}} \cdot O(1/\Delta)^t
\]
**Proof** Define $c_j = \frac{1}{N} \sum_{i \in S_j} a_i$ (now a deg-1 polynomial).

Recall the only thing we have proved about $\Theta$ is that for all $j \in (k)$,

$$c_j \cdot (m - \mu_j) \leq O(t)^{1/2} \cdot c_j^{t-1}. \quad (1)$$

Next, note that

$$\Delta^+ \leq (m_i - \mu_j)^+$$

$$= \left[ (m_i - \mu) - (\mu_j - \mu) \right]^+$$

$$\leq 2^+ \left[ (m_i - \mu)^+ + (\mu_j - \mu)^+ \right],$$

so

$$\frac{(m_i - \mu)^+ + (\mu_j - \mu)^+}{(\Delta/2)^+} \leq 1 \quad (2)$$

Combining (1) and (2) yields

$$c_i^+ c_j^+ \leq (2)^+ \cdot \left[ c_j^+ c_i^+ (m_i - \mu)^+ + c_i^+ c_j^+ (\mu_j - \mu)^+ \right] \leq (2|\Delta|^+ \cdot \left[ c_j^+ c_i^+ (m_i - \mu)^+ + c_i^+ c_j^+ (\mu_j - \mu)^+ \right]$$
\[
(1) \leq (z^{t+1}/\Delta)^t \left( c_i c_i^{t-1} + c_j c_j^{t-1} \right) \\
\leq 2 (z^{t+1}/\Delta)^t c_i c_j^{t-1} c_j^{t-1}
\]

So we have proved in deg-C(t) SoS that
\[
c_i^{t+1} c_j^{t+1} \leq O(\sqrt{t}/\Delta)^t c_i^{t-1} c_j^{t-1}
\]

To avoid working with odd powers, square both sides to get
\[
c_i^{2t} c_j^{2t} \leq O(\sqrt{t}/\Delta)^t c_i^{2t-2} c_j^{2t-2}
\]

Thus,
\[
\mathbb{E} \left[ c_i^{2t} c_j^{2t} \right] \leq O(\sqrt{t}/\Delta)^t \mathbb{E} \left[ c_i^{2t-2} c_j^{2t-2} \right] \quad (**) 
\]

Q: How do we simulate taking \( t \)-th roots in SoS? 
A: "pseudo-expectation Cauchy-Schwarz/Hölder's inequalities" 

Fact ("Cauchy-Schwarz"): 
\[
\mathbb{E} \left[ p(x) q(x) \right] \leq \mathbb{E} \left[ p(x) \right]^{1/2} \cdot \mathbb{E} \left[ q(x) \right]^{1/2}
\]
for any \( p, q \) of degree \( \leq t/2 \) and \( \mathbb{E} \) any deg-\( t \) pseudo-expectation.
Fact ("Hölder's"): $$\tilde{E}(p(x)^{t-2}) \leq \tilde{E}[p(x)^t]^{\frac{t-2}{t}}$$

For any degree sum of squares polynomial $p$ and any degree pseudo-expectation.

Proof: Part 2.

Applying pseudo-exp. Hölder's to (10), we get
$$\tilde{E}[c_i^2, c_j^2] \leq O(t/\Delta^2) + \tilde{E}[c_i^{2t}, c_j^{2t}]^{\frac{t-1}{t}}$$

Now we can divide freely $$\Rightarrow \tilde{E}[c_i^2, c_j^2] \leq O(t/\Delta^2)^{\frac{t}{2}}$$

Applying pseudo-exp Cauchy-Schwarz, we have
$$\tilde{E}(c_i c_j) = \tilde{E}(c_i c_j \cdot 1) \leq \tilde{E}[c_i^2, c_j^2]^{\frac{t}{2}} + \tilde{E}[c_i c_j | c_i^2, c_j^2]^{\frac{t}{2}}$$
and repeating this $\log_2 n$ times, get

$$\hat{\mathbb{E}}(c; c_j) \leq \hat{\mathbb{E}}(c_{i}; c_j)^{1/2}.$$

Thus, $\hat{\mathbb{E}}(c; c_j) \leq O(n/\Delta^2)^{1/2}$ for $i \neq j$,

and thus

$$\hat{\mathbb{E}}\left[ \sum_j c_j^2 \right] = \hat{\mathbb{E}}\left[ \left( \sum_j c_j \right)^2 - \sum_{i \neq j} c_i c_j \right]$$

$$\leq \left| - k + t^2 O(1/\delta) \right|$$

as desired. \(\square\)

**Rounding:**

Can't just output $\hat{\mathbb{E}}[n]$ b/c $\{a_i\}$'s don't preserve any particular component...

How do we know $\{a_i\}$'s are indicating a fixed component, or a dist over components?

**Trick:** entropy maximization
Want pseudo-dist over \( \{a_i^k\} \)'s to resemble uniform distribution over true indicators \( \{a(i)\} \)'s, where
\[
a(i) \sim \mathbb I [ x_i \text{ from } N(\mu_i, \Sigma)]
\]

This distribution has high "entropy" as quantified by
\[
\left\| \mathbb E \left[ a(i) a(i)' \right] \right\|_F^2 \quad \text{(ENT)}
\]

Note:
\[
\text{(ENT)} = \sum_{i,j=1}^{k} \frac{1}{k^2} \langle a(i) a(i)' , a(j) a(j)' \rangle^2
\]
\[
= \sum_{i,j=1}^{k} \frac{1}{k^2} \left( \langle a(i) , a(j)' \rangle \right)^2
\]
\[
= 0 \quad \text{if } j \neq j'
\]
\[
= \text{tr} \prod_{j \neq j'} a_j a_j' \quad \text{if components disjoint,}
\]
\[
= \frac{1}{k^2} \sum_{j=1}^{k} \| a(i) \|_2^4 = \frac{N^2}{k}
\]

We pick the pseudo-distribution solving
\[
\max \left\| \mathbb E \left[ (a_1, \ldots, a_n) (a_1, \ldots, a_n)' \right] \right\|_F
\]
subject to \( \mathbb E \) satisfying constraints of the program.
Lemma: The $\hat{E}$ satisfies

$$\left\| \hat{E} [aa^T] - E \left[ a^{(j)}(a^{(j)})^T \right] \right\|_F^2 \leq \left\| E \left[ a^{(j)}(a^{(j)})^T \right] \right\|_F^2 \left( k^2 + \frac{1}{2} \cdot O(1/\Delta^2) \right)$$

Proof: Because unit distribution over $\{[a_i^j], \mu_0\}$ is a feasible solution, $\|\hat{E}[aa^T]\|_F^2 \leq \frac{N}{k}$, so

$$\left( \frac{N}{k} \right) = \frac{2N^2}{k} - \frac{2}{k} \sum_{j=1}^{k} E \left[ \langle a, a^{(j)} \rangle^2 \right]$$

$$= \frac{2N^2}{k} - \frac{2}{k} \sum_{j=1}^{k} \frac{N}{N} \sum_{i \in S_j} \left[ \sum a_i \right]^2$$

$$= \frac{2N^2}{k} \left( 1 - \sum_j c_j^2 \right) \leq \frac{N^2}{k} \left( k^2 + \frac{1}{2} \cdot O(1/\Delta^2) \right). \quad \Box$$
Note,
\[ \mathbb{E} \left[ a_i (a_{i})^\top \right] = \left( \begin{array}{cccc} \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \\ \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{k} & \frac{1}{k} & \cdots & \frac{1}{k} \end{array} \right) \] (after row/col permutation),
so Lemma implies that we can read off clustering from \( \mathbb{E} (a a^\top) \)!

---

Final IOU: ... this was all for \( d=1 \)!

Warmup lemma and main Claim 3 easy to generalize, e.g.

Before:
\[
\left( \sum_{i \in S_j} a_i \right)^\top (m - m_j)^\top \leq 2^H \left( \sum_{i \in S_j} a_i \right)^{+1} N + t/2. 
\]

After
\[
\left( \sum_{i \in S_j} a_i \right)^\top \| m - m_j \|_2^\top \leq 2^H \left( \sum_{i \in S_j} a_i \right)^{+1} N + t/2. 
\]
But $\|\mu - \mu_j\|^2 = \langle \mu - \mu_j, \mu - \mu_j \rangle$, so can just “project” data along $\mu - \mu_j$ direction and reduce to 10 proof.

(need to be careful b/c $\mu - \mu_j$ is not a real vector because $\mu$ is an SOS variable)

Trickier: how to impose constraint

$$\frac{1}{N} \sum_{i=1}^{n} a_i \langle u, x_i - \mu \rangle^{+ 1/2} \leq 2^{+1/2} \|u\|_2^+$$

For all $u \in \mathbb{R}^d$?

Because we will apply this to $u = \mu - \mu_j$, need this to make sense even when $u$ is not a real vector...

Idea: Constrain Via

$$(\textbf{a}) \left\| \frac{1}{N} \sum_{i=1}^{n} a_i (x_i - \mu)^{+ 1/2} \left[(x_i - \mu)^{+ 1/2}\right]^T - \mathbb{E}_{g \sim N(0, I_d)} \left(g^{+ 1/2} (g^{+ 1/2})^T \right) \right\|_F^2 \leq 1$$

(satisfied by $a_i = a(i)$ and $\mu = \mu_j$, if $n$ large enough)
i.e. pick out subset s.t. empirical order-$t$ moments are close to those of $N(0, \text{Id})$.

**Fact.** For an SoS variable $u$,

$$
\mathbb{E}_{g \sim N(0, \text{Id})} \langle g, u \rangle^+ \leq +t/2 \cdot \|u\|_2^t
$$

has a deg-$t$ SoS proof in $u$.

**Pf:**

$$
\mathbb{E} \langle g, u \rangle^+ = \sum_{\text{deg-$t$ monomials } \alpha} u_\alpha \mathbb{E} [ g_\alpha ]
$$

where every variable appears even # times

$$
\leq +t/2 \sum_{\alpha} u_\alpha
$$

$$
= +t/2 \|u\|_2^t.
$$

i.e. $N(0, \text{Id})$ is "certifiably $t$-hypercontractive".

Note, if we take constraint (\ref{eq:constraint}) and hit it on both sides with $\left[(\mu - m_1)^\top \cdots (\mu - m_j)^\top \right]^{t/2}$, we get:
\[
\frac{1}{N} \sum_{i=1}^{\hat{N}} a_i \langle \mu - m_j, x_i - \mu \rangle^+ - \mathbb{E} \langle m - m_j, g \rangle^+ \leq \|m - m_j\|_2^+
\]

\[\downarrow \text{ (using Fact above)}\]

\[
\frac{1}{N} \sum_{i=1}^{\hat{N}} a_i \langle \mu - m_j, x_i - \mu \rangle^+ \leq (1 + t^{t/2}) \|m - m_j\|_2^{t/2}
\]

\[\leq O(t)^{t/2} \|m - m_j\|_2^{t/2},\]

which is sufficient to prove high-dim generalization of warmup lemma and its consequences.