Lecture 8

Interlude: tensors

Definition: an array of numbers with multiple indices, the order of a tensor is the number of its indices.

order	0	1	2	3 ⋯
	scalar	col/row	matrix	⋯⋯

order-l tensor X

$X_{i_1,\cdots,i_l}$

Tensor product:
- $(a\otimes a')_{i,i'} = a_i \cdot a'_{i'}$
- $a^{\otimes k} = \underbrace{a\otimes\cdots\otimes a}_{k \text{ times}}$
- $\langle a\otimes a', b\otimes b'\rangle = \langle a, b\rangle\cdot\langle a', b'\rangle$

Moment tensors

k-th moment tensor: $M_k = \mathbb{E}[\underline{y}^{\otimes t}]$

Connection to 1-dim moments:

$\mathbb{E}[\langle u, \underline{y}\rangle^k] = \mathbb{E}[\langle u^{\otimes k}, \underline{y}^{\otimes k}\rangle]$
$= \langle u^{\otimes k}, \mathbb{E}[\underline{y}^{\otimes k}]\rangle = \langle u^{\otimes k}, M_k\rangle$

Contractions:
- $a^T\cdot b = \langle a, b\rangle$
- $(a^T\otimes I)\cdot(b\otimes b') = \langle a, b\rangle\cdot b'$
- $(a^T\otimes I\otimes I)\cdot(b\otimes b'\otimes b'') = \langle a, b\rangle\cdot(b'\otimes b'')$
- For $i\in[r]$, $M_i\in\mathbb{R}^{p_i\times q_i}$, rank-1 tensors $V_1\otimes\cdots\otimes V_r\in\mathbb{R}^{p_1}\otimes\cdots\otimes\mathbb{R}^{p_r}$
  $(M_1\otimes\cdots\otimes M_r)\cdot(V_1\otimes\cdots\otimes V_r) = M_1V_1\otimes\cdots\otimes M_rV_r\in\mathbb{R}^{q_1}\otimes\cdots\otimes\mathbb{R}^{q_r}$

Random contraction algorithm

> **Theorem**
> Given an order-3 tensor $X\in(\mathbb{R}^d)^{\otimes 3}$ that admits a decomposition of the form
> $$X = a_1^{\otimes 3}+\cdots+a_r^{\otimes 3}$$
> for orthonormal vectors $a_1,\cdots,a_r\in\mathbb{R}^d$, we can efficiently compute the set $\{a_1,\cdots,a_r\}$.

or more general

> **Theorem**
> Let $X\in\mathbb{R}^{d_1}\otimes\mathbb{R}^{d_2}\otimes\mathbb{R}^{d_3}$ be an order-3 tensor that admits a decomposition of the form
> $$X = a_1\otimes b_1\otimes c_1+\cdots+a_r\otimes b_r\otimes c_r$$
> where the vectors $a_1\cdots a_r\in\mathbb{R}^{d_1}$ and $b_1\cdots b_r\in\mathbb{R}^{d_2}$ are orthonormal, $c_1\cdots c_r\in\mathbb{R}^{d_3}$ are pairwise distinct. We can efficiently compute the set of vectors $\{a_1\cdots a_r\}$, $\{b_1\cdots b_r\}$, $\{c_1\cdots c_r\}$ up to sign.

Proof:
- Choose $g\sim N(0, I_{d_3})$
- Compute the contraction $M = (I_{d_1}\otimes I_{d_2}\otimes \underline{g}^T)\cdot X$
- Compute the singular value decomposition of M viewed as d_1-by-d_2 matrix
- output the left and right singular vectors in this decomposition

So $M = \sum^r\langle \underline{g}, c_i\rangle\cdot a_i\otimes b_i$

$\underset{i=1}{}\ \underbrace{\quad}_{\text{singular}}\ \underbrace{\quad}_{\text{orthonormal}}$

singular value $\{|\langle g, c_i\rangle| : i \in [r]\}$. As c_i are pairwise distinct, therefore the singular value decomposition of M is unique up to sign.

Independent component analysis

parameters: $n, d, r \in \mathbb{N}$

unknown: $A^\circ \in \mathbb{R}^{d\times r}$, $\operatorname{rank}(A^\circ) = r$, product distribution $\{\underline{x}\}$ over $\mathbb{R}^r$

observe: $\underline{y}_1, \cdots, \underline{y}_n \overset{iid}{\sim} \{A^\circ \underline{x}\}$

goal: recover A° up to scaling & permuting columns.

example: $\underline{x} \sim \operatorname{Unif}(\{\pm 1\}^r)$, A° is orthonormal basis

usually: $r \le d \ll n$

matrix view: $\underline{Y} = A^\circ \underline{X}$

cocktail party problem (blind source separation)

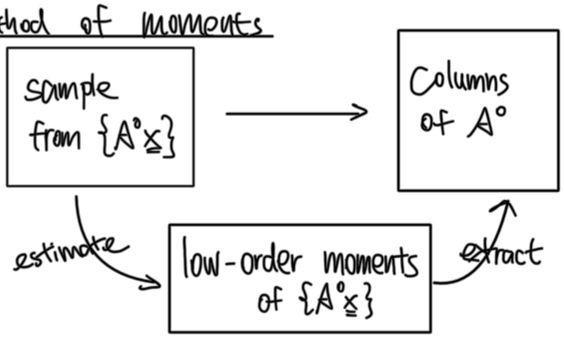
r conversations (signals)

d microphones

record n different superpositions

goal: disentangle signals.

Method of moments



moments: $u \in \mathbb{R}^d$

$$v \in (A^\circ)^T u \in \mathbb{R}^r \qquad (v_i = \langle A_i^\circ, u\rangle)$$

$$-\mathbb{E}[\langle u, A^\circ \underline{x}\rangle] = \langle v, \mathbb{E}[\underline{x}]\rangle, \quad \text{w.l.o.g } \mathbb{E}[\underline{x}] = 0$$

$$-\mathbb{E}[\langle u, A^\circ \underline{x}\rangle^2] = v^T (\underbrace{\mathbb{E}[\underline{x}\underline{x}^T]}_{\text{diagonal}}) v$$

$$\text{w.l.o.g. } \mathbb{E}[\underline{x}\underline{x}^T] = I_r, \quad (A^\circ)^T A^\circ = I_r, \ d = r.$$

$$-\mathbb{E}[\langle u, A^\circ \underline{x}\rangle^3] = \sum_{i,j,k=1}^{r} v_i v_j v_k \cdot \underbrace{\mathbb{E}[\underline{x}_i \underline{x}_j \underline{x}_k]}_{=0}$$

$$= \sum_{i=1}^{r} v_i^3 \cdot \mathbb{E}[\underline{x}_i^3]$$

$$-\mathbb{E}[\langle u, A^\circ \underline{x}\rangle^4] = \sum_{i,j,k,l} v_i v_j v_k v_l \cdot \mathbb{E}[\underline{x}_i \underline{x}_j \underline{x}_k \underline{x}_l]$$

$$= \sum v_i^4 \cdot (\mathbb{E}[\underline{x}_i^4] - 3) + 3 \cdot (\sum v_i^2)^2$$

Identifying columns of A^0

$F(u) = \mathbb{E}[\langle u, \underline{y}\rangle^4]$, $v = (A^0)^T u$

$F(u) = \sum v_i^4 \cdot (\mathbb{E}[\underline{x}_i^4] - 3) + 3 \cdot (\sum v_i^2)^2$

$\sum v_i^2 = \|u\|^2$ by assumption $(A^0)^T A^0 = I_r = I_d$

Turns Out: local first-order optima of $F(u)$ over unit sphere $\{u \in \mathbb{R}^d \mid \|u\| = 1\}$ are precisely $\{\pm A_1^0, \cdots, \pm A_r^0\}$.

$\rightarrow$ Can recover columns of A^0 up to sign using gradient descent on $F(u)$ with random initalization.

> **Theorem**
>
> Suppose $\mathbb{E}[\underline{x}_i^4] \neq 3$ for all $i \in [r]$, then, $\exists$ poly-time alg. to recover columns of A^0 up to sign approximately.

ICA:

$$M_3 = \mathbb{E}[(A^0\underline{x})^{\otimes 3}] = \sum_{ijk} A_i^0 \otimes A_j^0 \otimes A_k^0 \cdot \mathbb{E}[\underline{x}_i \underline{x}_j \underline{x}_k] = \sum_i (\mathbb{E}[\underline{x}_i^3]) \cdot (A_i^0)^{\otimes 3}$$

$$M_4 = \mathbb{E}[(A^0\underline{x})^{\otimes 4}] = \sum_i (\mathbb{E}[\underline{x}_i^4] - 3) \cdot (A_i^0)^{\otimes 4}$$

$$S = \begin{cases} + \sum_{ij} A_i^0 \otimes A_i^0 \otimes A_j^0 \otimes A_j^0 \\ + \sum_{ij} A_i^0 \otimes A_j^0 \otimes A_i^0 \otimes A_j^0 \\ + \sum_{ij} A_i^0 \otimes A_j^0 \otimes A_j^0 \otimes A_i^0 \end{cases}$$

$$S = \sum_i e_i^{\otimes 2} \otimes \sum_j e_j^{\otimes 2} + (\sum_{ij} e_i \otimes e_j)^{\otimes 2} + \sum_i e_i \otimes (\sum_j e_j^{\otimes 2}) \otimes e_i$$