Lecture 2: Tensors, Jennrich’s algorithm, applications

**Tensor Contraction:**

Given \(a_1, \ldots, a_k \in \mathbb{R}^d\), \(T \in (\mathbb{R}^d)^{\otimes k}\),

\[
T(a_1, \ldots, a_k) = \sum_{i_1, \ldots, i_k} T_{i_1 \ldots i_k} (a_1)_{i_1} \cdots (a_k)_{i_k}
\]

can be interpreted as evaluating polynomial w/ coefficients given by entries of \(T\) at the point \((a_1, \ldots, a_k)\).

**Partial Contraction:**

\[
T(a_1, \ldots, a_m, ; \ldots ;) \in (\mathbb{R}^d)^{\otimes k-m}
\]

\[
T(a_1, \ldots, a_m, ; \ldots ;)_{i_1 \ldots i_{k-m}} = T(a_1, \ldots, a_m, e_{i_1}, \ldots, e_{i_{k-m}})
\]

"higher-order" generalization of matrix-vector mult:

if \(T\) is matrix, i.e. order-2 tensor,

\[
T(:,z) = T_z
\]

Ex. If \(T = u \otimes v \otimes w\), then

\[
T(:,z) = (u \otimes v) \cdot \langle w, z \rangle
\]

So if \(T = \sum_i u_i \otimes v_i \otimes w_i\), then

\[
T(:,z) = \sum_i (u_i \otimes v_i) \cdot \langle w_i, z \rangle
\]
Jennrich's algorithm:

Suppose we are given \( T = \sum_{i=1}^{k} u_i \otimes v_i \otimes w_i \).

Sample \( z, z' \) randomly from \( S^{d-1} \).

Form

\[
M_z \coloneqq T(:, :, z) = \sum_i (u_i \otimes v_i) \cdot \langle w_i, z \rangle \\
M_{z'} \coloneqq T(:, :, z') = \sum_i (u_i \otimes v_i) \cdot \langle w_i, z' \rangle
\]

Note: if \( U, V, W \in \mathbb{R}^{d \times k} \) are matrices whose columns are \( \{ u_i \}, \{ v_i \}, \{ w_i \} \) respectively, and

\[
D_z \coloneqq \text{diag} (\langle w_1, z \rangle, \ldots, \langle w_k, z \rangle) \\
D_{z'} \coloneqq \text{diag} (\langle w_1, z' \rangle, \ldots, \langle w_k, z' \rangle),
\]

then

\[
M_z = U D_z V^T \\
M_{z'} = U D_{z'} V^T
\]

Now we use the same "simultaneous diagonalization" trick as in matrix pencil method!
\[ M_z M_z^+ = U D_z V^T (U D_z', V^T)^+ \]
\[ = U D_z V^T (V^T)^+ D_z^{-1} U^+ \]
\[ = U D_z D_z^{-1} U^+ \]

So the eigenvectors of \( M_z M_z^+ \) are the columns of \( U \) i.e. we get \( \{u_i\} \)

Similarly,
\[ M_z^+ M_z, = (U D_z V^T)^+ U D_z V^T \]
\[ = (V^T)^+ D_z^{-1} U^+ U D_z V^T \]
\[ = (V^T)^+ D_z^{-1} D_z V^T \]

So eigenvectors of \( (M_z^+ M_z,)^T \) are the columns of \( V \), i.e we get \( \{v_i\} \)

Q: How to "pair up" \( \{u_i\} \) and \( \{v_i\} \)? (we only know them up to permutation.)
A: Eigenvalues of \( M_2 M_2^+ \) and \((M_2^+ M_2)\) are in 1-1 correspondence:

- Those of \( M_2 M_2^+ \) are \( \{ \frac{\langle w_i, z \rangle}{\langle w_i, z \rangle} \} \) \( i = 1, \ldots, k \)

- Those of \( (M_2^+ M_2)\) are \( \{ \frac{\langle w_i, z' \rangle}{\langle w_i, z \rangle} \} \) \( i = 1, \ldots, k \)

N.B.: Here, we use randomness of \( z, z' \) to ensure that \( \langle w_i, z \rangle < \langle w_i, z' \rangle \neq 0 \), and together with non-collinearity of \( w_i \)'s, that there are no "accidental" reciprocals so eigenvalues really are in 1-1 correspondence as claimed.

Thus, can recover \( \{ u_i \otimes v_i \} \) \( i = 1, \ldots, k \).

Finally, need to recover \( w_1, \ldots, w_k \).

Idea: set up a linear system
Note, \( T_{abc} = \sum_i (u_i)_a (v_i)_b (w_i)_c \) (\#)

This gives \( d^3 \) equations in \( kd \) unknowns.

Need to show the equations have a unique soln.

Let \( (a,b) \in \mathbb{R}^k \) have entries \( \{(u_i)_a (v_i)_b\} \) \( i = 1, \ldots, k \).

Then (\#) \( \iff \ T_{abc} = \langle (a,b) , W_c \rangle \),

where \( W_c \) denotes \( c \)-th row of \( W \).

To show uniqueness of soln., the following suffices:

**Lem:** \( \{ (a,b) \} \) \( a,b \in \{d\} \) spans \( \mathbb{R}^k \)

**PF:** Let \( \Lambda \in \mathbb{R}^{d^2 \times k} \) have columns \( \{ (a,b) \} \).

Because \( d^2 > k \), suffices to show \( \Lambda \) has full column rank.

Suppose to contrary that \( \exists \ c \in \mathbb{R}^k \text{ s.t.} \)

\[ \sum_i c_i \Lambda_i = 0 \] (\#)
Note $\Lambda_i = \underbrace{\text{vec}(u_i \otimes v_i)}_{\text{i.e. flatten matrix}}$

i.e. flatten matrix $u_i \otimes v_i$ into vector

So (66) implies

$$\sum_i c_i u_i v_i^T = 0$$

Suppose WLOG $c_i \neq 0$.

Let $x$ be s.t. $\langle u_1, x \rangle \neq 0$, yet $\langle u_i, x \rangle = 0 \forall i > 1$.

(exists b/c $u_1, \ldots, u_k$ linearly independent)

Then

$$x^T \sum_i c_i u_i v_i^T = 0$$

$$\Rightarrow c_i \langle u_i, x \rangle v_i^T = 0$$

contradiction. $\square$

Q1: What about higher-order tensors?

e.g. $T = \sum_{i=1}^k u_i \otimes v_i \otimes w_i \otimes x_i$

A "reshape" $T$ into 3rd-order tensor, e.g. take

$$T' = \sum_{i=1}^k \underbrace{\text{vec}(u_i \otimes v_i) \otimes w_i \otimes x_i}_{\mathbb{R}^d}$$
To apply Jennrich's to $T'$, need

1). $\{ u_i \otimes v_i \}$ are linearly indep

2). $\{ w_i \}$ linearly indep

3). $x_i, x_j$ non-collinear for $i \neq j$

Note: by lemma above, 1) holds if $\{ u_i \}$ linearly indep and $\{ v_i \}$ linearly indep

Q2: What if

$$T = \sum u_i \otimes v_i \otimes w_i + \text{[noise]}$$

A: Instead of $U, V$ being full column rank, need that $\sigma_{\min}(U), \sigma_{\min}(V)$ not too small

See Pset 1 Q#2
### APPLICATIONS:

1. **MIXTURES OF EXPONENTIALS:**

Recall from lecture 1:

get access to

\[ G: w \mapsto \frac{1}{k} \sum_{j=1}^{k} e^{2\pi i \langle w, \nu_j \rangle} \]

for any \( \|w\| \leq 1 \) where \( w, \nu_1, \ldots, \nu_k \in \mathbb{R} \)

Goal: recover \( \nu_1, \ldots, \nu_k \) for \( m \) sufficiently large

**Alg:**

1) Sample \( w_1, \ldots, w_m \sim \mathcal{B}(0.49) \)

2) Define \( X_1 = 0.02 \cdot v, \quad X_2 = 0.02 \cdot v' \)

3) Define \( T \in \mathbb{R}^{m \times m \times 2} \)

by \( T_{abc} = G[\nu_a + \nu_b + \nu_c] \)

\( \|v\| \leq 1 \) by design
Then \( T_{abc} = \frac{1}{k} \sum_{j=1}^{k} \left( x_j \cdot w_j \right) \cdot \left( x_j \cdot w_j \right) \cdot \left( x_j \cdot w_j \right) \)

where \( u_i \in \mathbb{R}^d \) and \( w_i \in \mathbb{R}^2 \)

So \( T = \frac{1}{k} \sum_{j=1}^{k} u_j \otimes u_j \otimes w_j \), and we can apply Jemrich's provided \( \{u_i\} \) linearly independent and \( w_i \)'s non-collinear.

Note: This is essentially a generalization of matrix pencil method to higher dimensions!

---

2) SPHERICAL GAUSSIAN MIXTURES:

Unknown: \( \mu_1, \ldots, \mu_k \in \mathbb{R}^d \) linearly indep.
\( \lambda_1, \ldots, \lambda_k \in [0,1] \) s.t. \( \sum_j \lambda_j = 1 \)

Given: iid samples from mixture model \( q \), where
\[ q = \sum_{i=1}^{k} \lambda_i \cdot N(\mu_i, \text{Id}) \]
i.e. to sample from \( q \):

1) Sample \( i \in \{k\} \) w.p. \( \lambda_i \)

2) Sample \( g \sim N(0, I_d) \)

3) Output \( \mu_i + g \)

**Goal:** estimate \( \{\mu_i\}, \{\lambda_i\} \) up to small error

**Goal:** estimate \( \mu_1, \ldots, \mu_k \) up to error

\[
\text{Alg.: } E[X] = \sum_{i=1}^{k} \lambda_i E[\mu_i + g] = \sum_{i=1}^{k} \lambda_i \mu_i
\]

\[
E[X^{\otimes 3}] = \sum_{i=1}^{k} \lambda_i E[(\mu_i + g)^{\otimes 3}]
\]

\[
= \sum_{i=1}^{k} \lambda_i E[\mu_i^{\otimes 3} + g^{\otimes 3}] + \underbrace{E[g^{\otimes 3}]}_{= 0 \text{ by symmetry}} = 0 \text{ by symmetry}
\]

\[
= \sum_{i=1}^{k} \lambda_i \mu_i^{\otimes 3} + \left( \sum_{i=1}^{k} \lambda_i \mu_i \right)^{\otimes 3} I_d
\]

where \( z \otimes I_d \equiv \sum_{a=1}^{d} z a^a \otimes e_a + e_a \otimes z a^a + e_a \otimes e_a \otimes z a^a \)
So \( \mathbb{E} [ x^3 ] - \mathbb{E} (x) \otimes^3 \text{Id} = \sum_{i=1}^{k} \lambda_i \mu_i^3 \)

Can run Jennrich's to learn the parameters.

---

3. **INDEPENDENT COMPONENT ANALYSIS (BONUS)**

Unknown: \( A \in \mathbb{R}^{d \times d} \)

Given: iid samples of the form

\[ z = A x + g, \quad x \sim g \sim N(0, \text{Id}) \]

for a product distribution \( g \)

**interpretation:** \( g \) generates \( d \) independent signals, e.g. conversations at a dinner party, and we and we only observe noisy combinations of these signals, e.g. picked up by mics in the room. Want to unscramble by learning \( A \).

**Idea:** \( \mathbb{E}[z_1 z_2 z_3 z_4 z_5] \) unwieldy to write down (try it yourself), so use **cumulants** instead
(See slides for background on cumulants)

\[ K(z_i, z_j, z_k, z_l) = K(\sum_s A_{is} x_s + g_i, \sum_s A_{js} x_s + g_j, \sum_s A_{ks} x_s + g_k, \sum_s A_{ls} x_s + g_l) \]

(Gaussian cumulants vanish)

\[ = K(\sum_s A_{is} x_s, \ldots, \sum_s A_{ls} x_s) \]

(additivity)

\[ = \sum_s K(A_{is} x_s, \ldots, A_{ls} x_s) \]

(independent cumulants vanish)

\[ = \sum_s A_{is} A_{js} A_{ks} A_{ls} \cdot K(x_s, x_s, x_s, x_s) \]

So if \( T_{ijkl} = K(z_i, z_j, z_k, z_l) \), then

\[ T = \sum_s \lambda_s A_{is} \otimes^4 \]
Note $\lambda_5 = \mathbb{E} \left[ x_i^4 - 3 \right]_x$

If $q \sim N(0, \text{Id})$, then $\lambda_5 = 0$, but this case is actually impossible!

$z = Ax + g \sim N(0, \text{Id} + A^T A)$, so we would only be able to recover $A$ up to rotation of its rows...

But if $q$ “non-Gaussian” so $\lambda_5 \neq 0$, then Jennrich’s can be used to recover $A$ provided $A$ has full rank!