- Jenrich's algorithm is not very stable
  needs eigenvalue gaps in \((M_0 M_1)^{-1}\) to be large.
- In practice, power method is more popular.
- Tensor \(T = \sum_{i=1}^{r} \lambda_i a_i \otimes a_i \otimes a_i\)
  \(a_i\)'s are orthonormal. \(\text{wlog } \lambda_i \geq 0\)

\[
\begin{align*}
  u^{(0)} &\sim N(0, \frac{1}{d} I) \\
  \text{for } i = 1 \to +\infty \\
  u^{(i)} &= T(u^{(i-1)}, u^{(i-1)}, I) \\
  &= \sum_{j=1}^{d} \lambda_j \langle u^{(i-1)}, a_j \rangle^2 a_j
\end{align*}
\]

(\text{optional: } u^{(i)} = \frac{u^{(i)}}{\|u^{(i)}\|})

Theorem: With high probability, power method converges to \(a_i\) with largest \(|\lambda_i \langle u^{(0)}, a_i \rangle|\) in \(\log \log \frac{d}{\varepsilon} + \log d\) steps.

\textbf{Proof:} Let \(u^{(i)} = \sum_{j=1}^{r} c_j^{(i)} a_j\)

\[
\begin{align*}
  c_j^{(i)} &= \lambda_j (c_j^{(i-1)})^2 \\
  c_j^{(1)} &= \lambda_j (c_j^{(0)})^2 \\
  c_j^{(2)} &= \lambda_j \left( \lambda_j (c_j^{(0)})^2 \right)^2 = \lambda_j^3 (c_j^{(0)})^4 \\
  c_j^{(3)} &= \lambda_j \left( \lambda_j^3 (c_j^{(0)})^4 \right)^2 = \lambda_j^7 (c_j^{(0)})^8
\end{align*}
\]
\[ C_j^{(i)} = \lambda_j^{2^{-i}} (C_j^{(0)})^2 \]
\[ = (\lambda_j C_j^{(0)})^{2^{-i-1}} C_j^{(0)} \]

with high probability

\[ \max \left| \lambda_j C_j^{(0)} \right| \geq (1 + \frac{1}{d^2}) \lambda_{j'} C_j^{(0)} \quad (j \neq \arg \max) \]

\[ C_j^{(0)} \geq \frac{1}{d^2} \]

choose \( t \) such that \( 2^t - 1 \geq 3d^2 \log \frac{d}{\varepsilon} \)

\[ \frac{C_j^{(t)}}{C_j^{(t+1)}} > \frac{\varepsilon}{d} \implies \frac{\| u_t \|}{\| u_{t+1} \|} \] is \( \varepsilon \)-close to \( a_j \)

- Problem: \( a_j \)'s need to be orthogonal.

- Solution: Whitening.

\[ T = \sum_{i=1}^{r} \lambda_i u_i \otimes u_i \otimes u_i \]

\[ M = \sum_{i=1}^{r} \lambda_i u_i u_i^\top \]

does not have to be the same, but need \( \lambda_i > 0 \)

\[ M = U D U^\top \]

\[ W = U D^{-\frac{1}{2}} \]

property: \( W^T M W = I \)

\[ a_i = \lambda_i^{\frac{1}{2}} W u_i \]

then: \( \| a_i \| = 1, \quad \langle a_i, a_j \rangle = 0 \)

\[ T(w, w, w) = \sum_{i=1}^{r} \lambda_i (W u_i) \otimes (W u_i) \otimes (W u_i) \]
\[
\sum_{i=1}^{r} \lambda_i^{-\frac{1}{2}} (\lambda_i^{-\frac{1}{2}} W u_i) \Theta^3
\]

- How to find tensor structure?

- Method of moments.
  - recall: model \( D(\theta) \), want to estimate \( \theta \).
  - MoM: compute the moments of \( x \sim D(\theta) \)
    \[
    E[x], E[xx^T], E[x \otimes x \otimes x]
    \]
    then solve equations.
  - Example: \( x \sim N(\mu, \sigma^2) \)
    \[
    E[x] = \mu \quad E[x^2] = \mu^2 + \sigma^2
    \]
    - identifiability: the equations have unique solution.

- Pearson’s Crabs
  - measures a certain ratio for crabs, distribution

\[
\text{Conjecture: mixture of two Gaussians.}
\]

Can compute \( E[x], E[x^2], \ldots, E[x^6] \) and solve for parameters.

- Simple example: Pure topic model
- Each document has only 1 topic

\[ Q_{i,j} = \Pr[\text{lst word} = i, \text{2nd word} = j] \]

\[ = \sum_{l=1}^{r} \Pr[\text{topic} = l] \cdot A_{i,l} \cdot A_{j,l} \]

\[ Q = \sum_{l=1}^{r} \Pr[\text{topic} = l] \mathbf{a}_l \mathbf{a}_l^T \]

- \( T_{i,j,k} = \Pr[\text{lst word} = i, \text{2nd word} = j, \text{3rd word} = k] \)

\[ = \sum_{l=1}^{r} \Pr[\text{topic} = l] A_{i,l} A_{j,l} A_{k,l} \]

\[ T = \sum_{l=1}^{r} \Pr[\text{topic} = l] \mathbf{a}_l \otimes \mathbf{a}_l \otimes \mathbf{a}_l \]