Aside: Perturbation Bounds

So far:

$\boxed{\text{generative model}} \xrightarrow{\text{conditional independence}} \boxed{\text{low rank tensor}}$

can be estimated from samples

Phylogenetic trees
HMMs
GMMs
community detection

$\xrightarrow{\text{Jennrich's Algorithm}} \boxed{\text{parameter learning}}$

Main Question: what about sampling noise?

when is Jennrich's algorithm stable?

def: The condition number is

$$\kappa(M) = \frac{\sigma_{max}(M)}{\sigma_{min}(M)}$$

Lemma: Suppose $Mx = b$ and we are given $\tilde{b}$.
Let
$$\tilde{x} = M^{-1}\tilde{b}$$

Then
$$\frac{\|\tilde{x} - x\|}{\|x\|} \leq \kappa(M) \frac{\|\tilde{b} - b\|}{\|b\|}$$

i.e. the condition number controls the relative error

Proof: We have

$$\tilde{x} - x = M^{-1}(b + \tilde{b} - b) - x$$

$$= M^{-1}(\tilde{b} - b)$$

$$\Rightarrow \|\tilde{x} - x\| \leq \frac{\|\tilde{b} - b\|}{\sigma_{min}(M)}$$

$$\Rightarrow \frac{\|\tilde{x} - x\|}{\|x\|} \leq \frac{\|\tilde{b} - b\|}{\|x\| \sigma_{min}(M)} \qquad (1)$$

Moreover $\|b\| \leq \sigma_{max}(M) \|x\| \Rightarrow$

$$\frac{1}{\|x\|} \leq \frac{\sigma_{max}(M)}{\|b\|} \qquad (2)$$

(1) + (2) completes proof $\boxtimes$

Ultimately we want perturbation bounds for eigenvalues / eigenvectors:

If $M = UDU^{-1}$ and $\tilde{M}$ is close to M, when can we accurately estimate U?

First we'll bound change in eigenvalues

Thm [Gershgorin Disk Thm] The eigenvalues of M are contained in

$$\bigcup_i D(M_{ii}, R_i)$$

disk in $\mathbb{C}$ around M_{ii}, of radius R_i

where $R_i = \sum_{j \neq i} |M_{ij}|$

Proof: Let (x, λ) be an eigenvector-eigenvalue pair for M. Let

$$i = \operatorname{argmax}_i |x_i| \quad \text{(break ties arbitrarily)}$$

Then from the eigenvector eqn

$$\sum_j M_{ij} x_j = \lambda x_i \Rightarrow$$

$$\sum_{j \neq i} M_{ij} x_j = \lambda x_i - M_{ii} x_i \Rightarrow$$

$$|\lambda - M_{ii}| = \left| \frac{\sum_{j \neq i} M_{ij} x_j}{x_i} \right|$$

$$\leq \sum_{j \neq i} \left| \frac{M_{ij} x_j}{x_i} \right| \leq \sum_{j \neq i} M_{ij} = R_i \quad \blacksquare$$

Now suppose $M = UDU^{-1}$ and $\tilde{M} = M + E$

Let $\delta = \min_{i \neq j} |D_{ii} - D_{jj}|$

Claim: If E is small enough, compared to δ and $1/\kappa(U)$, then $\tilde{M}$ has distinct eigenvalues

Proof: we have

$$U^{-1}\tilde{M}U = D + \underbrace{U^{-1}EU}_{\substack{\text{can be controlled by } \kappa(U) \\ \text{and norm of } E}}$$

So if E is small enough, the Gershgorin disks will be disjoint

Need a continuity argument to ensure each disk contains one and only one eigenvalue. $\boxtimes$

What about the eigenvectors?

Now that we know $\tilde{M}$ is diagonalizable, can write

$$\tilde{M} = \tilde{U}\tilde{D}\tilde{U}^{-1}$$

Moreover there is a natural pairing

btwn eigenvectors of $M / \tilde{M}$.

$$u_i \longleftrightarrow \tilde{u}_i$$

Suppose $\tilde{u}_i = \sum_j c_j u_j$. We'd like to show that $|c_j|$ is small for any $j \neq i$

we have $\tilde{M} \hat{u}_i = \hat{\lambda}_i \hat{u}_i \Rightarrow$

$$\sum_j c_j \lambda_j u_j + E \tilde{u}_i = \tilde{\lambda}_i \tilde{u}_i \Rightarrow$$

$$\sum_j c_j (\lambda_j - \tilde{\lambda}_i) u_j = -E \hat{u}_i$$

Now let $w_j^T = j^{th}$ row of U^{-1}. Then

$$\underbrace{w_j^T \left(\sum_k c_k (\lambda_k - \tilde{\lambda}_i) u_k \right)}_{c_j (\lambda_j - \tilde{\lambda}_i)} = -w_j^T E \hat{u}_i$$

Thus u_i and $\hat{u}_i$ get close as E gets smaller

Going back to Jennrich's algorithm, as we take more samples

$$\tilde{T} \to T$$

This in turn implies

$$\tilde{T}_a \to T_a \text{ and } \tilde{T}_b \to T_b$$

when T_a and T_b are invertible it can be shown that

$$\tilde{T}_a^{-1} \to T_a^{-1} \text{ and } \tilde{T}_b^{-1} \to T_b^{-1}$$

And so finally

$$\tilde{T}_a \tilde{T}_b^{-1} \to T_a T_b^{-1}, \text{etc}$$

and we get estimators for the factors that converge at an inverse polynomial rate.

Note: The expression is complicated, and involves condition numbers and separation btwn eigenvalues, etc