Matrix Perturbation: Theory and Applications

Lekshmi Ramesh

Indian Institute of Science
Bangalore

07 October 2017

Outline

- Matrix perturbation
  - Introduction
  - Applications (PCA, Clustering in networks)
- Clustering: more details
- Perturbation theory
  - Weyl's inequality
  - The Davis-Kahan theorem, Proof
- Guarantees for spectral clustering

Matrix perturbation

- Matrix perturbation theory tries to characterize the effect of an unknown perturbation on certain properties of a matrix
- For $A, E \in \mathbb{R}^{n \times n}$:
  - how are the eigenvalues of A and $A + E$ related?
  - how are the eigenvectors of A and $A + E$ related?
  - other questions..
- Perturbation theory is useful for analysing algorithms that are based on eigenvalue/eigenvector computations. We will see two examples
  - Principal Components Analysis
  - Spectral clustering in networks

Matrix perturbation: an example

- Let $A = \begin{bmatrix} 1-\epsilon & 0 \\ 0 & 1+\epsilon \end{bmatrix}$

  eigenvalues: $\{1-\epsilon, 1+\epsilon\}$ eigenvectors: $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

- Let $\hat{A} = A + \begin{bmatrix} \epsilon & \epsilon \\ \epsilon & -\epsilon \end{bmatrix} = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix}$

  eigenvalues: $\{1-\epsilon, 1+\epsilon\}$ eigenvectors: $\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$

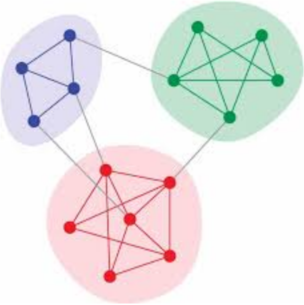
- Perturbation caused a large rotation of eigenvectors
- Can guarantee eigenvectors of A and $\hat{A}$ are "close" under restrictions on eigenvalues of A

Applications

- Principal Components Analysis
  - given n data points $X_1, \ldots, X_n$ in $\mathbb{R}^d$, find a lower dimensional subspace that best fits the data
  - optimal subspace determined by leading eigenvectors of covariance matrix of data
  - how far are the corresponding eigenvectors of the population covariance matrix and sample covariance matrix?
- Spectral clustering in networks
  - given a graph $G = (V, E)$, partition its vertices into clusters
  - under a certain generative model, the second leading eigenvector of the *expected* adjacency matrix gives cluster labels
  - how far are the second leading eigenvectors of the adjacency matrix and the expected adjacency matrix?

Clustering in networks: introduction

- The network is represented as a graph $G = (V, E)$
- We want to partition the vertex set V into clusters such that
  - there are many edges within a cluster
  - there are few edges across clusters



Clustering in networks: the stochastic block model

- Stochastic Block Model (SBM): A generative model for graphs with clusters

- Two-cluster case
  For $n \in \mathbb{N}$ and $p, q \in (0, 1)$, let $\mathcal{G}(n, p, q)$ be the class of random graphs where
  - each vertex v is assigned a label $\sigma_v \in \{+1, -1\}$ (independently and uniformly at random)
  - each possible edge (u, v) is included with probability p if $\sigma_u = \sigma_v$ and with probability q if $\sigma_u \neq \sigma_v$

Clustering in networks

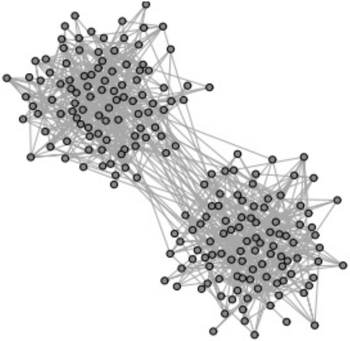


Figure 1: A random graph $G \sim \mathcal{G}(200, \frac{1}{20}, \frac{1}{200})$

Clustering in networks: spectral algorithm

- Let $G \sim \mathcal{G}(n, p, q)$ and A be the adjacency matrix of G. The expected adjacency matrix $D := \mathbb{E}A$ has a block structure (after reordering rows and columns)
- For example, with $n = 4$:

$$D = \begin{bmatrix} p & p & q & q \\ p & p & q & q \\ q & q & p & p \\ q & q & p & p \end{bmatrix}$$

- Eigenvalues of D: $\lambda_1^{\mathrm{D}} = 2(p+q), \quad \lambda_2^{\mathrm{D}} = 2(p-q)$

  Corresponding eigenvectors: $v_1^{\mathrm{D}} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, v_2^{\mathrm{D}} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$

Clustering in networks: spectral algorithm

- Eigenvector corresponding to second largest eigenvalue of D gives vertex labelings
- But we only have access to A. We will see that v_2^{D} and v_2^{A} are close under some conditions on the spectrum of D

Weyl's inequality

- Weyl's inequality gives a characterization of the maximum deviation caused in eigenvalues by an additive perturbation

  Let A and B be $n \times n$ symmetric matrices. Then,

  $$\max_i |\lambda_i^A - \lambda_i^B| \leq \|A - B\|.$$

Davis-Kahan theorem

Some notation:

- A and B are symmetric $n \times n$ matrices and $E = B - A$
- $\lambda_1^A \geq \ldots \geq \lambda_n^A$ are the eigenvalues of A with corresponding eigenvectors $v_1^A, \ldots, v_n^A$
- $\lambda_1^B \geq \ldots \geq \lambda_n^B$ be the eigenvalues of B with corresponding eigenvectors $v_1^B, \ldots, v_n^B$
- θ_i is the angle between the lines through v_i^A and v_i^B

The Davis-Kahan theorem states that

$$\sin \theta_i \leq \frac{2\|E\|}{\min_{j \neq i} |\lambda_i^A - \lambda_j^A|}.$$

Davis-Kahan theorem

Proof

- Consider "shifted" versions $A - \lambda_i^A I$ and $B - \lambda_i^A I$ (does not affect the eigenvectors)
- After shifting, $\lambda_i^A = 0$. Also,

$$\|E\| = \|B - A\| \geq |\lambda_i^B|$$

- Assume all eigenvectors are of unit length

Davis-Kahan theorem

- Expand v_i^B in the eigenbasis for A

$$v_i^B = \sum_j c_j v_j^A,$$

where $c_j = \langle v_j^A, v_j^B \rangle$

- $\delta := \min_{j \neq i} |\lambda_j^A|$
- Then,

$$\begin{aligned} \|Av_i^B\|_2^2 &= \sum_j c_j^2 (\lambda_j^A)^2 \\ &\geq \sum_{j \neq i} c_j^2 \delta^2 \\ &= \delta^2 (1 - c_i^2) \\ &= \delta^2 \sin^2 \theta_i \end{aligned} \tag{1}$$

- Also,

$$\begin{aligned} \|Av_i^B\|_2 &= \|(B-E)v_i^B\|_2 \\ &\leq \|Bv_i^B\|_2 + \|Ev_i^B\|_2 \\ &= \lambda_i^B + \|Ev_i^B\|_2 \\ &\leq 2\|E\| \end{aligned} \tag{2}$$

- Using (1) and (2)

$$\sin\theta_i \leq \frac{2\|E\|}{\delta}$$

- We can also show that

$$\sin\theta_i \leq \frac{2\|E\|}{\delta}$$

implies that there exists $\alpha \in \{-1, +1\}$ such that

$$\|v_i^A - \alpha v_i^B\|_2 \leq \frac{2\sqrt{2}\|E\|}{\delta}.$$

That is, v_i^A and v_i^B are close upto sign.

Spectral clustering

- Recall: $G \sim \mathcal{G}(n, p, q)$ with adjacency matrix A and $D = \mathbb{E}A$
- We want $\operatorname{sgn}(v_2^A) \approx \operatorname{sgn}(v_2^D)$
- Using the Davis-Kahan theorem,

$$\|v_i^A - \alpha v_i^D\|_2 \leq \frac{2\sqrt{2}\|A - D\|}{\delta}$$

- Computing δ:

$$\delta = \min\left(\frac{p-q}{2}, q\right)n =: \mu n$$

- Using concentration results, we can show that

$$\|A - D\| \leq c\sqrt{n}$$

  with probability at least $1 - 4e^{-n}$

- And so,

$$\|v_i^A - \alpha v_i^D\|_2 \leq \frac{c}{\mu\sqrt{n}}$$

  with probability at least $1 - 4e^{-n}$

- Can show:
  $\#\{\text{indices where signs disagree}\} \leq \|v_i^A - \alpha v_i^D\|_2^2 \leq \frac{c}{\mu^2}$

- We thus have the following result:

  Let $G \sim \mathcal{G}(n, p, q)$ with $p > q$ and $\mu = \min(q, p - q)$. Then, with probability at least $1 - 4e^{-n}$, the spectral clustering algorithm identifies the communities of G upto $\frac{c}{\mu^2}$ misclassified vertices.

References

McSherry, Frank. "Spectral Partitioning of Random Graphs". In: *FOCS*. 2001, pp. 529–537.

Roughgarden, Tim. *Lecture notes CS264*. Stanford University. 2014.

Thank you