CS395T Numerical Optimization for Graphics and AI — Linear Algebra

Qixing Huang
The University of Texas at Austin
huangqx@cs.utexas.edu

1 Basic

Linear algebra is widely used in AI and Graphics research. If you have not seriously learned this before, please take a class. At UT, you can take CS 383C NUMERICAL ANLY: LINEAR ALGEBRA. If you have learned it before but forgot the technical details (most likely), I recommend you to go through the following wikipedia page for review:

- Introduction to linear algebra [1].

1.1 Notations

We will use the following convention throughout all the lectures:

- Capital letters denote matrices, e.g., $A, B, C, \cdots$.
- Lowercase bold face letters denote vectors, e.g., $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \cdots$.
- Lowercase letters denote scalars, e.g., $s, t, \cdots$.
- $\boldsymbol{e}_i = (0, \cdots, 1, \cdots, 0)^T$ is the reserved for the canonical basis of $\mathbb{R}^n$.

2 Spectral of Normalized Adjacency Matrix

Eigenvalues and eigenvectors of a square matrix are very fundamental concepts in matrix theory. We are particularly interested in symmetric matrices. Specifically, given a symmetric matrix $A \in \mathbb{R}^{n \times n}$, it has n eigenvalues $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ and n eigenvectors $\boldsymbol{u}_1(A), \cdots, \boldsymbol{u}_n(A)$. The eigenvalues and eigenvectors are related by the following equality

$$A\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \quad 1 \leq i \leq n.$$

Equivalently, we can write out the eigen-decomposition of A as

$$A = U \Lambda U^T, \tag{1}$$

where

$$U = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n), \quad \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n).$$

In this lecture, we use some basic facts of spectral graph theory to study properties of eigenvalues and eigenvectors of square matrices. Spectral techniques are widely used in Graphics and AI, we will have three lectures on this topic later this semester.

[1] https://en.wikipedia.org/wiki/Linear_algebra

Consider a connected graph of n vertices $\mathcal{G} = (\{1, \cdots, n\}, \mathcal{E})$. With d_i we denote the vertex degree of i-th vertex. Consider so-called normalized adjacency matrix $\overline{A} \in \mathbb{R}^{n\times n}$, whose elements are given by

$$\overline{A}_{ij} = \begin{cases} \frac{1}{\sqrt{d_i d_j}} & (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

We will prove the following facts:

Fact 2.1. *1 is an eigenvalue of $\overline{A}$, and the corresponding eigenvector is* $\boldsymbol{u}_1 = \left(\sqrt{\frac{d_1}{\sum_i d_i}}, \cdots, \sqrt{\frac{d_n}{\sum_i d_i}}\right)$.

Proof. The proof reviews matrix-vector multiplication. In fact, let $\mathcal{N}(i) \subset \{1, \cdots, n\}$ collects indices of the neighboring vertices of vertex i, then

$$\begin{aligned} \boldsymbol{e}_i^T A \boldsymbol{u}_1 &= \sum_{j\in\mathcal{N}(i)} \frac{1}{\sqrt{d_i d_j}} \cdot \sqrt{\frac{d_j}{\sum_k d_k}} \\ &= \frac{1}{\sqrt{d_i}} \sum_{j\in\mathcal{N}(i)} \sqrt{\frac{1}{\sum_k d_k}} \\ &= \sqrt{\frac{d_i}{\sum_k d_k}}. \end{aligned}$$

□

Fact 2.2. *The eigenvalues of $\overline{A}$ are between -1 and 1.*

The proof of the following fact will use a different definition of eigenvalues for symmetric matrices:

$$\lambda_1(A) = \max_{\boldsymbol{x}\in\mathbb{R}^n} \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}, \tag{2}$$

$$\lambda_n(A) = \min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}. \tag{3}$$

The proof is easy — using (1), we have

$$\begin{aligned} \max_{\boldsymbol{x}\in\mathbb{R}^n} \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} &= \max_{\boldsymbol{x}\in\mathbb{R}^n} \frac{(U^T\boldsymbol{x})^T \Lambda (U^T\boldsymbol{x})}{(U^T\boldsymbol{x})^T (U^T\boldsymbol{x})} \\ &= \max_{\boldsymbol{y}\in\mathbb{R}^n} \frac{\boldsymbol{y}^T \Lambda \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{y}} \\ &= \max_{\boldsymbol{y}\in\mathbb{R}^n} \frac{\sum_{i=1}^n \lambda_i y_i^2}{\sum_{i=1}^n y_i^2} \\ &= \lambda_1. \end{aligned} \tag{4}$$

We can generalize (2) and (3) to other eigenvalues. For example,

$$\lambda_i(A) = \max_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{u}_1^T\boldsymbol{x}=0, \cdots, \boldsymbol{u}_{i-1}^T\boldsymbol{x}=0} \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}},$$

$$\lambda_i(A) = \min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{u}_n^T\boldsymbol{x}=0, \cdots, \boldsymbol{u}_{i+1}^T\boldsymbol{x}=0} \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}. \tag{5}$$

Alternatively, we have

$$\lambda_i(A) = \min_{U, dim(U)=n-i} \max_{\boldsymbol{x} \in U} \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}},$$
$$\lambda_i(A) = \max_{U, dim(U)=i} \min_{\boldsymbol{x} \in U} \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}. \tag{6}$$

The proofs are more details can be found at Rayleigh quotient (`https://en.wikipedia.org/wiki/Rayleigh_quotient`) and min-max theorem (`https://en.wikipedia.org/wiki/Min-max_theorem`). Now we give the proof of Fact 2.2.

Proof of Fact 2.2. First of all,

$$\begin{aligned}
\boldsymbol{x}^T \overline{A} \boldsymbol{x} &= \sum_{(i,j) \in \mathcal{E}} \frac{x_i x_j}{\sqrt{d_i d_j}} \\
&= \frac{1}{2} \sum_{(i,j) \in \mathcal{E}} \Big((\frac{x_i}{\sqrt{d_i}} + \frac{x_j}{\sqrt{d_j}})^2 - \frac{x_i^2}{d_i} - \frac{x_j^2}{d_j} \Big) \\
&\geq -\frac{1}{2} \sum_{(i,j) \in \mathcal{E}} \Big(\frac{x_i^2}{d_i} + \frac{x_j^2}{d_j} \Big) \\
&= -\frac{1}{2} (\sum_{i=1}^{n} \sum_{j \in \mathcal{N}(i)} \frac{x_i^2}{d_i} + \sum_{j=1}^{n} \sum_{i \in \mathcal{N}(j)} \frac{x_j^2}{d_j}) \\
&= -\frac{1}{2} (2 \sum_{i=1}^{n} x_i^2) \\
&= -\sum_{i=1}^{n} x_i^2.
\end{aligned} \tag{7}$$

In other words,

$$\lambda_n(\overline{A}) \geq -1.$$

In the other direction,

$$\begin{aligned}
\boldsymbol{x}^T \overline{A} \boldsymbol{x} &= \sum_{(i,j) \in \mathcal{E}} \frac{x_i x_j}{\sqrt{d_i d_j}} \\
&= \frac{1}{2} \sum_{(i,j) \in \mathcal{E}} \Big(-(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}})^2 + \frac{x_i^2}{d_i} + \frac{x_j^2}{d_j} \Big) \\
&\leq \frac{1}{2} \sum_{(i,j) \in \mathcal{E}} \Big(\frac{x_i^2}{d_i} + \frac{x_j^2}{d_j} \Big) \\
&= \sum_{i=1}^{n} x_i^2,
\end{aligned} \tag{8}$$

which means

$$\lambda_1(\overline{A}) \leq 1.$$

□

3 Registration

We proceed to consider singular values and singular value decompositions of matrices. Given a matrix $A \in \mathbb{R}^{m \times n}$(assuming $m < n$ for simplicity), its singular value decompositions admit the form

$$A = U \Sigma V^T, \tag{9}$$

where $U = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m) \in O(m)$ and $V = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n) \in O(n)$ are unitary matrices that collect singular vectors $\boldsymbol{u}_i, \boldsymbol{v}_j$; $\Sigma = (\mathrm{diag}(\sigma_1, \cdots, \sigma_m), 0)$ is a generalized diagonal matrix that collects its singular values $\sigma_i \geq 0, 1 \leq i \leq m$.

Now we apply singular value decomposition to a concrete problem. We are interested in estimating the best rigid translation between two point clouds with known correspondences. Formally speaking, we are given two point clouds in $\mathbb{R}^k$: $P = (\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n)$ and $Q = (\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n)$. $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ are linked. You can think that these point clouds come from the feature points of two 3D shapes, and the correspondences come from matching feature descriptors.

We estimate the best rigid transformation $R \in SO(m), \boldsymbol{t}$ by solving the following minimization problem:

$$\min_{R,\boldsymbol{t}} \sum_{i=1}^{n} \|R\boldsymbol{p}_i + \boldsymbol{t} - \boldsymbol{q}_i\|^2 \tag{10}$$

In other words, the rigid transform aligns the corresponding points in L^2 norm. Rotations are formally called signed unitary matrices, i.e., $det(R) = 1, R \in O(m)$. Note that $det(R) = -1, R \in O(m)$ encode matrices that encode reflection symmetries (which also preserve pair-wise distances). Please refer to `https://en.wikipedia.org/wiki/Rotation_group_SO(3)` and `https://en.wikipedia.org/wiki/Orthogonal_group` for further reading.

Now we describe how to solve (10). First, we notice that when R is fixed, the objective function is quadratic in $\boldsymbol{t}$, and the optimal solution admits the following form

$$\boldsymbol{t}^{\star} = R(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{p}_i) - (\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{q}_i). \tag{11}$$

In fact, for points $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n$,

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{a}_i := \underset{\boldsymbol{x}}{\mathrm{argmin}} \sum_{i=1}^{n} \|\boldsymbol{x} - \boldsymbol{a}_i\|^2.$$

Denote

$$\overline{\boldsymbol{p}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{p}_i, \quad \overline{\boldsymbol{q}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{q}_i,$$

and let

$$P = (\boldsymbol{p}_1 - \overline{\boldsymbol{p}}, \cdots, \boldsymbol{p}_n - \overline{\boldsymbol{p}}), \quad Q = (\boldsymbol{q}_1 - \overline{\boldsymbol{q}}, \cdots, \boldsymbol{q}_n - \overline{\boldsymbol{q}}).$$

Substituting (11) into (10), it is easy to see that the optimal $R^{\star}$ is given by

$$R^{\star} = \underset{R}{\mathrm{argmin}} \|RP - Q\|_{\mathcal{F}}^2. \tag{12}$$

To proceed, we will use the so-called matrix inner product, which is tied to both the Frobenius norm and trace of matrices `https://en.wikipedia.org/wiki/Trace_%28linear_algebra%29`. More precisely, the inner product between two matrices $A, B \in \mathbb{R}^{m \times n}$ is given by

$$\langle A, B \rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij} = \mathrm{Trace}(AB^T) = \mathrm{Trace}(A^T B).$$

The matrix inner product can be used to encode the Frobenius norm:

$$\|X\|_{\mathcal{F}}^2 = \langle X, X \rangle.$$

It also possesses many standard properties such as

$$\langle X, Y + Z \rangle = \langle X, Y \rangle + \langle X, Z \rangle.$$

Please refer to`https://en.wikipedia.org/wiki/Frobenius_inner_product` for more details.

With this setup, we can expand

$$\begin{aligned}
\|RP - Q\|_{\mathcal{F}}^2 &= \langle RP - Q, RP - Q\rangle \\
&= \|RP\|_{\mathcal{F}}^2 + \|Q\|_{\mathcal{F}}^2 - 2\langle RP, Q\rangle \\
&= \|RP\|_{\mathcal{F}}^2 + \|Q\|_{\mathcal{F}}^2 - 2\mathrm{Trace}(RPQ^T). \qquad (13)
\end{aligned}$$

Denote $S = PQ^T$, the optimal solution is given by

$$R^\star = \max_{R}\ \mathrm{Trace}(RS). \qquad (14)$$

In the following, we will derive an explicit formula for $R^\star$. First of all, when S is a diagonal matrix with non-negative diagonal entries, i.e., $S = \mathrm{diag}(s_1, \cdots, s_m)$, we have

$$\mathrm{Trace}(RS) = \sum_{i=1}^{m} R_{ii} s_i.$$

Since the largest possible value for each R_{ii} is 1. So the optimal solution for (14) is $R^\star = I_m$.

Now how to turn S into a diagonal matrix? Well, one possibility is to use singular value decomposition. Let $S = U\Sigma V^T$ be the SVD of S. We have

$$\mathrm{Trace}(RS) = \mathrm{Trace}(RU\Sigma V^T) = \mathrm{Trace}(V^T RU\Sigma),$$

where we have used the most importance property about matrix trace, i.e., $\mathrm{Trace}(AB) = \mathrm{Trace}(BA)$. Applying what we have just derived, the optimal solution shall satisfy $V^T R^\star U = I_m$, or in other words,

$$R^\star = V \cdot U^T.$$

Are we done?.... Unfortunately not, because we have forgot the constraint that $\det(R^\star) = 1$. When $\det(S) \geq 0$, we have $\det(U) \cdot \det(V) = 1$[2], which means $\det(R^\star) = 1$. However, when $\det(S) < 0$, $\det(U) \cdot \det(V) = -1$, and we have to do more work. In fact, we have to solve the following optimization problem: Given $s_1 \geq \cdots \geq s_m \geq 0$,

$$\max_{R \in O(k), \det(R) = -1} \sum_{i=1}^{k} R_{ii} s_i. \qquad (15)$$

This becomes non-trivial problem. We will have to use a fact that was proven in [1]: Consider the diagonal entries of a matrix $R \in O(k), \det(R) = -1$ as the vector in $\mathbb{R}^k$. The convex-hull of these vectors are given by the convex hull of the points $(\hat{A}\pm 1, ..., \hat{A}\pm 1)$ with an odd number of -1s. In this convex hull, a linear objective function is attached at the vertices. So the optimal solution is given by $R_{ii} = 1, 1 \leq i \leq k-1, R_{kk} = -1$.

References

[1] A. Horn., "Doubly stochastic matrices and the diagonal of a rotation matrix." *Amer. J. Math.*, pp. 620–630, 1954. 3

[2] When $\det(S) = 0$, you can flip at least one column of U without changing the decomposition.