Lecture 2: Geometric Graph Embeddings: Isometric and Nearly Isometric Embeddings of Geometric Graphs.

Radu Balan

Department of Mathematics, AMSC, CSCAMM and NWC
University of Maryland, College Park, MD

February 11, 2020

Embeddings with Full Data

Problem Statement and Ambiguities

Main Problem

Isometric Embedding: Given the set of all squared-distances $\{d_{i,j}^2;\ 1 \leq i,j \leq n\}$ *find a dimension* d *and a set of* n *points* $\{y_1, \cdots, y_n\} \subset \mathbb{R}^d$ *so that* $\|y_i - y_j\|^2 = d_{i,j}^2$, $1 \leq i,j \leq n$.

Main Problem

Nearly Isometric Embedding: Given the set of all squared-distances $\{d_{i,j}^2;\ 1 \leq i,j \leq n\}$ *find a dimension* d *and a set of* n *points* $\{y_1, \cdots, y_n\} \subset \mathbb{R}^d$ *so that* $\|y_i - y_j\|^2 \approx d_{i,j}^2$, $1 \leq i,j \leq n$.

Note the set of points is unique up to rigid transformations: translations, rotations and reflections: $\mathbb{R}^d \times O(d)$. This means two sets of n points in $\mathbb{R}^d$ have the same pairwise distances if and only if one set is obtained from the other set by a combination of rigid transformations.

Isometric Embeddings with Full Data

Converting pairwise distances into the Gram matrix

Let $S = (S_{i,j})_{1 \leq i,j \leq n}$ denote the $n \times n$ symmetric matrix of squared pairwise distances:

$$S_{i,j} = d_{i,j}^2 \quad , S_{i,i} = 0$$

Denote by 1 the n-vector of 1's (the Matlab $ones(n,1)$). Let $\nu = (\|y_i\|^2)_{1 \leq i \leq n}$ denote the unknown n-vector of squared-norms. Finally, let $G = (\langle y_i, y_j \rangle)_{1 \leq i,j \leq n}$ denote the Gram matrix of scalar products between y_i and y_j.

We can remove the translation ambiguity by fixing the center:

$$\sum_{i=1}^{n} y_i = 0$$

Isometric Embeddings with Full Data

Converting pairwise distances into the Gram matrix

Expand the square:

$d_{i,j}^2 = \|y_i - y_j\|^2 = \|y_i\|^2 + \|y_j\|^2 - 2\langle y_i, y_j\rangle \Rightarrow 2\langle y_i, y_j\rangle = \|y_i\|^2 + \|y_j\|^2 - d_{i,j}^2$

Rewrite the system as:

$$2G = \nu \cdot 1^T + 1 \cdot \nu^T - S \quad (*)$$

The center condition reads: $G \cdot 1 = 0$, which implies:

$$0 = 2n\nu^T \cdot 1 - 1^T \cdot S \cdot 1$$

Let $\rho := \nu^T \cdot 1 = \sum_{i=1}^n \|y_i\|^2$. We obtain:

$$\rho = \frac{1}{2n} 1^T \cdot S \cdot 1 = \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^n d_{i,j}^2$$

$$\nu = \frac{1}{n}(S \cdot 1 - \rho 1) = \frac{1}{n}(S - \rho I) \cdot 1$$

that you substitute back into (*).

Isometric Embeddings with Full Data

Converting pairwise squared-distances into the Gram matrix: Algorithm

Algorithm (Alg 1)

Input: Symmetric matrix of squared pairwise distances $S = (d_{i,j}^2)_{1 \le i,j \le n}$.

1. *Compute:*
$$\rho = \frac{1}{2n} 1^T \cdot S \cdot 1 = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{i,j}^2$$
2. *Set:*
$$\nu = \frac{1}{n}(S \cdot 1 - \rho 1) = \frac{1}{n}(S - \rho I) \cdot 1$$
3. *Compute:*
$$G = \frac{1}{2}\nu \cdot 1^T + \frac{1}{2} 1 \cdot \nu^T - \frac{1}{2} S = \frac{1}{2n}(S - \rho I) 1 \cdot 1^T + \frac{1}{2n} 1 \cdot 1^T (S - \rho I) - \frac{1}{2} S.$$

Output: Symmetric Gram matrix G

Isometric/Nearly Isometric Embeddings with Full Data

Factorization of the G matrix

In the absence of noise (i.e. if $S_{i,j}$ are indeed the Euclidean distances), the Gram matrix G should have rank d, the minimum dimension of the isometric embedding.
If S is noisy, then G has approximate rank d.
To find d and Y, the matrix of coordinates, perform the eigendecomposition:

$$G = Q\Lambda Q^T$$

where Λ is the diagonal matrix of eigenvalues, ordered monotonically decreasing. Choose d as the number of significant positive eigenvalues (i.e. truncate to zero the negative eigenvalues, as well as the smallest positive eigenvalues). Note G has always at least one zero eigenvalue: $rank(G) \leq n-1$.

Isometric Embeddings with Full Data

Factorization of the G matrix

Then we obtain an approximate factorization of G (exact in the absence of noise):

$$G \approx Q_1 \Lambda_1 Q_1^T$$

where Q_1 is the $n \times d$ submatrix of Q containing the first d columns.
Set $Y = \Lambda_1^{1/2} Q_1^T$, so that $G \approx Y^T Y$.
The $d \times n$ matrix Y contains the embedding vectors $y_1, \cdots, y_n$ as columns:

$$Y = [y_1 | y_2 | \cdots | y_n] .$$

Question: What optimization problem is solved by the eigendecomposition? We shall discuss it after Algorithm 2.

Isometric Embeddings with Full Data

Gram matrix factorization: Algorithm

Algorithm (Alg 2)

Input: Symmetric $n \times n$ *Gram matrix* G.

1. *Compute the eigendecomposition of* G, $G = Q\Lambda Q^T$ *with diagonal of* Λ *sorted in a descending order;*
2. *Determine the number* d *of significant positive eigevalues;*
3. *Partition*
$$Q = [Q_1 \quad Q_2] \quad , \text{and} \quad \Lambda = \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix}$$
where Q_1 *contains the first* d *columns of* Q, *and* Λ_1 *is the* $d \times d$ *diagonal matrix of significant positive eigenvalues of* G.
4. *Compute:*
$$Y = \Lambda_1^{1/2} Q_1^T$$

Output: Dimension d *and* $d \times n$ *matrix* Y *of vectors* $Y = [y_1 | \cdots | y_n]$

Optimality of Eigendecompositions

Assume $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix, $A = A^T$.
Fix $1 \leq d \leq n$. Consider the following problem: Find d vectors $\hat{f}_1, \cdots, \hat{f}_d \in \mathbb{R}^n$ that minimize

$$J = \underset{\{f_1, \cdots, f_d\} \subset \mathbb{R}^n}{\text{minimize}} \quad \|A - \sum_{k=1}^{d} f_k f_k^T\|_F \tag{1.1}$$

where the Frobenius norm is defined by $\|X\|_F = \left(\sum_{1 \leq i,j \leq n} |X_{i,j}|^2\right)^{1/2}$.
Claim 1: Without loss of generality (W.L.O.G.) we can assume $\{\hat{f}_1, \cdots, \hat{f}_d\}$ is orthogonal, i.e., $\langle \hat{f}_i, \hat{f}_j \rangle = 0$ for $i \neq j$.
Why?

$$I = \underset{\{g_1, \cdots, g_d\} \text{ orhogonal set}}{\text{minimize}} \quad \|A - \sum_{k=1}^{d} g_k g_k^T\|_F \tag{1.2}$$

i) Obviously: $J \leq I$ because less constraints in (1.1).

Optimality of Eigendecompositions

Equivalence betwen I and J

ii) For the converse inequality $I \leq J$, we proceed as follows.
Let $\{\hat{f}_1, \cdots, \hat{f}_d\}$ be an optimizer of (1.1). Consider the eigenfacorization of matrix $R = \sum_{k=1}^{d} \hat{f}_k \hat{f}_k^T$. Say $R = ED_1 R^T$ where R is the $n \times d$ matrix formed by the first d eigenvectors of R and D_1 is the $d \times d$ matrix of top d eigenvalues of R. Note that R has rank at most d (its range is the span of d vectors), hence at most d eigenvalues are nonzero; the other $n-d$ eigenvalues are 0. Let $\{e_1. \cdots, e_d\}$ be the normalized eigenvectors of R that are columns in E, so that $E = [e_1 | \cdots | e_d]$. Let $\lambda_1, \cdots, \lambda_d$ be the top eigenvalues of R that are also on the diagonal of D_1. Then, for $g_1 = \sqrt{\lambda_1} e_1, \ldots, g_d = \sqrt{\lambda_d} e_d$, we have $R = g_1 g_1^T + g_2 g_2^T + \cdots g_d g_d^T$. On the other hand $\langle g_i, g_j \rangle = \sqrt{\lambda_1 \lambda_j} \langle e_i, e_j \rangle = 0$, where the last equality comes from the fact that we the eigenvectors $\{e_1, \cdots, e_d\}$ were chosen orthonormal. It follows $\{g_1, \cdots, g_d\}$ is a feasible set for problem (1.2). Hence $I \leq \|A - R\|_F = J$.

Optimality of Eigendecompositions
Reduction to one vector

Assume $(\hat{f}_1, \cdots, \hat{f}_d)$ is an orthogonal set minimizer in (1.2). Then $\hat{f}_d$ is the minimizer of

$$H = \underset{f \in \mathbb{R}^n}{\text{minimize}} \quad \|A - \sum_{k=1}^{d-1} \hat{f}_k \hat{f}_k^T - ff^T\|_F \tag{1.3}$$

Why?: Similarly, $J \leq H$ (because less constraints). And $H \leq I$ (because less constraints).
Consequence: we can solve the sequential optimization problems, i.e., peel-off one rank one at a time:

$$\underset{f \in \mathbb{R}^n}{\text{minimize}} \quad \|A_k - ff^T\|_F \tag{1.4}$$

where $A_0 = A$ and $A_k = A_{k-1} - \hat{f}\hat{f}^T$.

Optimality of Eigendecompositions

Solution for one vector optimization

We are left to solve the minimization of $\|A - xx^T\|_F$ for a symmetric matrix $A = A^T \in \mathbb{R}^{n \times n}$ and $x \in \mathbb{R}^n$.

Expand the Frobenius norm:

$$\|A - xx^T\|_F^2 = trace((A - xx^T)(A - xx^T)) = trace(A^2) - 2trace(Axx^T) + \\ + trace(xx^Txx^T) == \|A\|_F^2 - 2\langle Ax, x\rangle + \|x\|^4$$

(check!)

Let $x = t \cdot e$ where $t > 0$ is a scalar and $e \in \mathbb{R}^n$ is a unit vector $\|e\| = 1$, i.e., $t = \|x\|$ and $e = \frac{x}{\|x\|}$. Then

$$\|A - xx^T\|_F^2 = \|A\|_F^2 - 2t^2\langle Ae, e\rangle + t^4$$

Minimization over t produces a bi-quadratic problem whose solution is

$$\hat{t} = \sqrt{max(0, \langle Ae, e\rangle)}$$

Optimality of Eigendecompositions

Solution for one vector optimization - 2

Substitute back $\hat{f}$ into $\|A - xx^T\|_F^2$:

$$\|A - xx^T\|_F^2 = \begin{cases} \|A\|_F^2 & \text{if} \quad \langle Ax, x\rangle < 0 \\ \|A\|_F^2 - (\langle Ax, x\rangle)^2 & \text{if} \quad \langle Ax, x\rangle \geq 0 \end{cases}$$

Finally, consider the optimization problem

$$\begin{array}{cc} \text{maximize} & \langle Ae, e\rangle \\ e \in \mathbb{R}^n, \|e\| = 1 & \end{array}$$

Use Lagrange multiplier technique to solve it:

$$L(e, \lambda) = \langle Ae, e\rangle - \lambda(\langle e, e\rangle - 1) \Rightarrow \nabla L = 0$$

Obtain:

$$Ae - \lambda e = 0 \quad , \quad \langle e, e\rangle - 1 = 0$$

Hence (λ, e) is an eigenpair. Solution: $\hat{e}$ is the *principal unit-norm eigenvector* of matrix A.

Optimality of Eigendecompositions
Summary

Theorem

Let $A = A^T \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Fix an integer $1 \le d \le n$. Let $\{(\lambda_k, e_k); 1 \le k \le d\}$ be the top d eigenpairs, i.e. $Ae_k = \lambda_k e_k$, $\|e_k\| = 1$ and $\{\lambda_1, \cdots, \lambda_d\}$ the largest d eigenvalues. An optimizer of the problem:

$$J = \underset{\{f_1,\cdots,f_d\}\subset\mathbb{R}^n}{\text{minimize}} \quad \|A - \sum_{k=1}^d f_k f_k^T\|_F \tag{1.5}$$

is given by $\hat{f}_k = \sqrt{\max(0,\lambda_k)}e_k$, $1 \le k \le d$. Equivalently, the optimizer of the problem

$$J = \underset{\substack{R=R^T\in\mathbb{R}^{n\times n}\\ rank(R)=d\\ R\ge 0}}{\text{minimize}} \quad \|A - R\|_F \tag{1.6}$$

is given by $R = \sum_{k=1}^d \max(0,\lambda_k)e_k e_k^T$.

Review of the Eigenproblems Theory
Definitions

Recall: An *eigenpair* (λ, v) of a square matrix $A \in \mathbb{C}^{n\times n}$ is pair composed of a non-zero vector v (called *eigenvector*) and a scalar λ (called *eigenvalue*) that satisfy $Av = \lambda v$. In general, we normalize v so that $\|v\| = 1$.

Any $n \times n$ matrix admits exactly n (maybe complex and repeated) eigenvalues. They all are roots of the *characteristic polynomial*, $P_A(z) = det(zI - A)$. If A admits n linearly independent eigenvectors $\{v_1, \cdots, v_n\}$ then A *diagonalizes*, that is, with $V = [v_1|v_2|\cdots|v_n]$ and $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$, $A = V\Lambda V^{-1}$.

It is a remarkable fact that all symmetric matrices ALWAYS diagonalize. In fact more can be said about these matrices.

First, a bit of terminology:

A real matrix $A \in \mathbb{R}^{n\times n}$ is said *symmetric*, or *self-adjoint*, if $A = A^T$.

A complex matrix $A \in \mathbb{C}^{n\times n}$ is said *hermitian*, or *self-adjoint*, if $A = \bar{A}^T$ (i.e., complex-conjugate and transpose). In general, we denote $A^* = \bar{A}^T$.

Review of the Eigenproblems Theory
Matrix Factorization

Theorem (Factorization of self-adjoint matrices)

Assume $A = A^*$ (either real or complex matrix).

1. *All eigenvalues of A are real, i.e., the characteristic polynomial $p_A(z)$ has exactly n real zeros.*
2. *There exists an orthonormal basis $\{e_1, e_2, \cdots, e_n\}$ composed of eigenvectors associated to eigenvalues $\lambda_1, \cdots, \lambda_n\}$ so that, with $E = [e_1|e_2|\cdots|e_n]$ and $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$,*

$$A = E\Lambda E^*$$

   Furthermore, if A is a real matrix then all eigenvectors have real entries.
3. *For every $x, y \in \mathbb{C}^n$, $\langle Ax, y\rangle = \langle x, Ay\rangle$, and $\langle Ax, x\rangle \in \mathbb{R}$ is always a real number.*

Review of the Eigenproblems Theory

Matrix Factorization

The last property allows us to define a *non-negative matrix*, also called *positive semi-definite* (PSD) matrix A, that matrix so that: $A = A^*$ (i.e., it is self-adjoint), and for every $x \in \mathbb{C}^n$, $\langle Ax, x\rangle \geq 0$. We denote this by $A \geq 0$. If, in addition, the matrix satisfies, for every $x \in \mathbb{C}^n$, $x \neq 0$, $\langle Ax, x\rangle > 0$ then A is said *positive definite* (or just positive). We denote this by $A > 0$.

Given the factorization in this theorem, we conclude that:

Corollary

Assume $A = A^*$. Then,

1. *$A \geq 0$ if and only if all eigenvalues satisfy $\lambda \geq 0$.*
2. *$A > 0$ if and only if all eigenvalues satisfy $\lambda > 0$.*

As a side remark: If a matrix $A \in \mathbb{C}^{n\times n}$ satisfies, for every $x \in \mathbb{C}^n$, $\langle Ax, x\rangle \in \mathbb{R}$ then $A = A^*$.

Review of the Eigenproblems Theory
Optimization Problems solved by Eigenpairs

Assume $A = A^* \in \mathbb{R}^{n \times n}$ (the hermitian case is similar, but for ease of notation we assume all valiables are real).
Consider the following optimization problems:

$$\begin{array}{ll} maximize & \langle Ax, x\rangle \\ \|x\| = 1 & \end{array} \tag{1.7}$$

and

$$\begin{array}{ll} minimize & \langle Ax, x\rangle \\ \|x\| = 1 & \end{array} \tag{1.8}$$

Both problems can be solved using the Lagrange multiplier technique:

$$L(x, \lambda) = \langle Ax, x\rangle - \lambda(\langle x, x\rangle - 1) \Rightarrow \nabla L = 0$$

which produces eigenproblems for A: $Ax = \lambda x$. The first optimization problem has solution the largest eigenvalue of A, whereas the second problem has solution the smallest eigenvalue of A.

Review of the Eigenproblems Theory

Optimization Problems solved by Eigenpairs

To summarize:

Theorem

Let $A = A^* \in \mathbb{R}^{n\times n}$ be a self-adjoint matrix. Let $\{(\lambda_k, e_k); 1 \leq k \leq n\}$ be the eigenpairs with $\lambda_1 \geq \cdots \geq \lambda_n$ and $\|e_k\| = 1$. Then for any vector $x \in \mathbb{R}^n$, with $\|x\| = 1$,

$$\lambda_n = \langle Ae_n, e_n\rangle \leq \langle Ax, x\rangle \leq \langle Ae_1, e_1\rangle = \lambda_1.$$

If A is not symmetric, then it can be replaced by its symmetrization via

$$\langle Ax, x\rangle = \frac{1}{2}\langle Ax, x\rangle + \frac{1}{2}\langle x, A^*x\rangle = \langle \frac{1}{2}(A + A^*)x, x\rangle$$

Hence:

$$\lambda_{max}\left(\frac{1}{2}(A + A^*)\right) = \underset{\|x\|=1}{maximize} \quad \langle Ax, x\rangle \quad , \lambda_{min}\left(\frac{1}{2}(A + A^*)\right) = \underset{\|x\|=1}{minimize} \quad \langle Ax, x\rangle$$

References

- S. Boyd, L. Vandenberghe, **Convex Optimization**, available online at: http://stanford.edu/ boyd/cvxbook/
- [10]A. Javanmard, A. Montanari, Localization from Incomplete Noisy Distance Measurements, arXiv:1103.1417, Nov. 2012; also ISIT 2011.