Latent Distance Estimation For Random Geometric Graphs

Ernesto Araya Valdivia† and Yohann De Castro‡

†Laboratoire de Mathématiques d'Orsay, Université Paris-Sud

‡École Centrale de Lyon

Random Geometric Graph Model

Popular latent space model for graphs that has been used to model wireless, social and biological networks.

Angular version:

- **Latent points**: $\{X_i\}_{i=1}^n$ are i.i.d uniformly distributed in the unit sphere $\mathbb{S}^{d-1}$. The population Gram matrix is defined by $(\mathcal{G})_{ij} := \langle X_i, X_j \rangle$.
- **Connection function**: $f : [-1,1] \to [0,1]$ such that

$$\Theta_{ij}(n) = \rho_n f(\langle X_i, X_j \rangle)$$

where ρ_n is the sparsity parameter.

- **Random geometric graph**: The adjacency matrix $A(n)$ is given by

$$\mathbb{P}(A_{ij}(n) = 1 | \{X_1, \cdots, X_n\}) = \Theta_{ij}$$

Problem statement

- We observe a simple graph $G = (V, E)$.
- We assume it was generated using the angular RGG model (see above), but we do not know which function f was used.
- **Objective**: Recover the Gram matrix $\mathcal{G}$, with small error, only using the adjacency matrix of G.
- **Error criteria**: We define the error $\mathbf{E}(A, B) = \frac{1}{n^2} \|A - B\|_F^2$ for A, B matrices of the same size.

Strategy

- Use the theory of graph limits [1]: $A(n)$ converges towards the compact integral operator $T_f : L^2(\mathbb{S}^{d-1}) \to L^2(\mathbb{S}^{d-1})$

$$T_f g(x) = \int_{\mathbb{S}^{d-1}} f(\langle x, y \rangle) g(y) d\mu(y)$$

with μ = the surface measure on $\mathbb{S}^{d-1}$. In particular, the spectrum of $\Theta(n)$ converges to the spectrum of T_f in a modified ℓ_2 metric [3].

- Use harmonic analysis on $\mathbb{S}^{d-1}$ to derive properties of T_f:
  - Eigenfunctions do not depend on f: they are the spherical harmonics $\{\phi_k\}_{k=1}^{\infty}$.
  - Non zero eigenvalues have fixed multiplicities that depend only on the sphere dimension.
  - Reconstruction formula: we have $\forall x, y \in \mathbb{S}^{d-1}$, $P_{E_j}(x, y) \propto G_j(\langle x, y \rangle)$, where is a Gegenbauer polynomial of degree k.

Main assumptions

- **Eigengap condition**: $\exists \Delta^* > 0$ such that $\min_{\lambda \in \lambda(T_f), \lambda \notin \Lambda_1^*} |\lambda - \lambda^*| > \Delta^*$
- **Regularity**: f has regularity s if the eigenvalues of T_f satisfy $\sum_k |\lambda_k|^2 k^{2s} < \infty$.

Main result

Theorem 1 (Informal statement) *Let G be a graph generated with the RRG model with connection function f (unknown), which satisfies the eigengap condition. Then*

- *$\exists \Lambda_1 = (v_1, \cdots, v_d)$ subset of the eigenvectors of $A(n)$, such that $\mathbf{E}(\mathcal{G}, \hat{\mathcal{G}}) \xrightarrow[n \to \infty]{} 0$, where $\hat{\mathcal{G}} := \frac{1}{c} \sum_{i=1}^d v_i v_i^T$ and c is a known constant.*
- *Furthermore, if has regularity s, we have $\mathbf{E}(\mathcal{G}, \hat{\mathcal{G}}) = O(\Delta^{*-1} n^{-s/2s+d-1})$.*

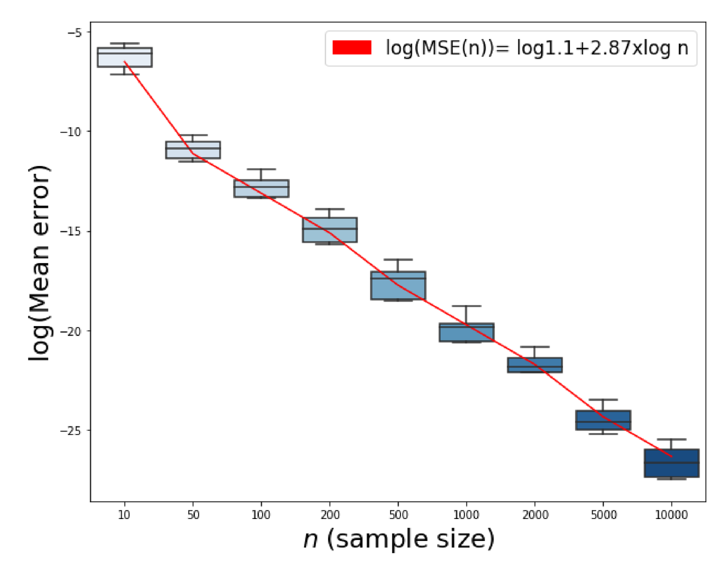
Gram matrix recovery

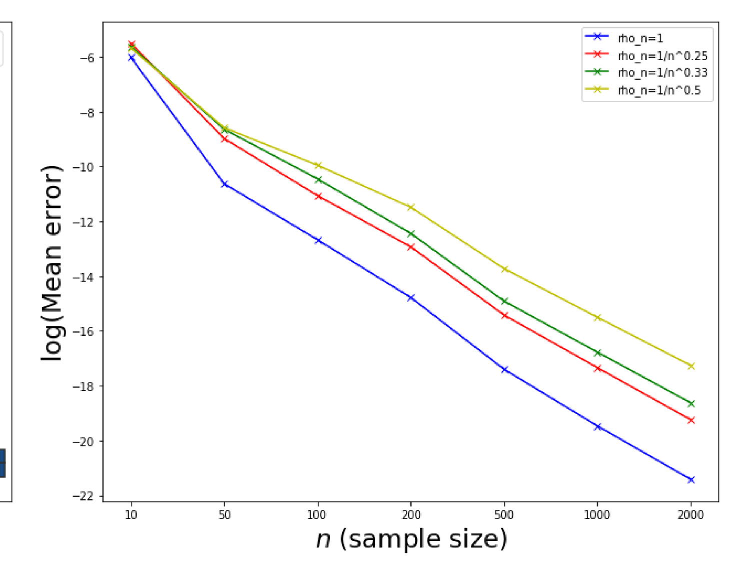
Harmonic Eigen-Cluster (HEiC) algorithm:

- **Input**: (A, d) adjacency matrix and dimension of the sphere.
- $\lambda^{\text{sort}}(A) \leftarrow$ Compute $\lambda(A) = \{\lambda_0^{\text{sort}}, \cdots, \lambda_n^{\text{sort}}\}$ and order decreasingly.
- $\text{Gap}_1 \leftarrow \max \left\{ \max_{1 \le i \le n-d-1} \min \{\text{ljump}(i), \text{rjump}(i+d)\}, \text{ljump}(n-d+1) \right\}$, with

$$\text{ljump}(i) := |\hat{\lambda}_i^{\text{sort}} - \hat{\lambda}_{i-1}^{\text{sort}}| \quad \text{and} \quad \text{rjump}(i) := \text{ljump}(i+1)$$

- $i^* \leftarrow \begin{cases} n - d \text{ if } \max_{1 \le i \le n-d-1} \min \{\text{ljump}(i), \text{rjump}(i+d)\} < \text{ljump}(n-d+1) \\ \arg\max_{1 \le i \le n-d-1} \min \{\text{ljump}(i), \text{rjump}(i+d)\}, \text{ljump}(n-d+1) \text{ otherwise} \end{cases}$
- **Return**:(Gap_1, i^*)

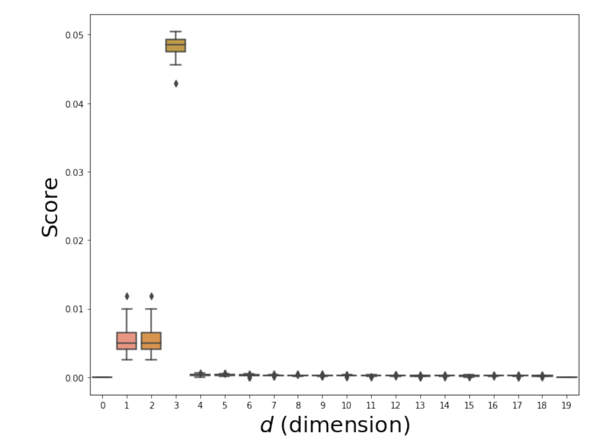


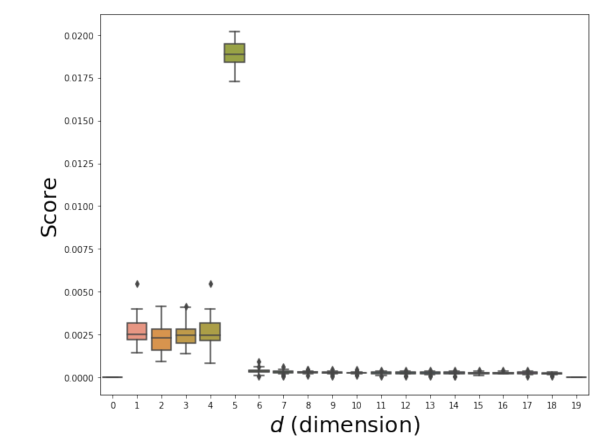


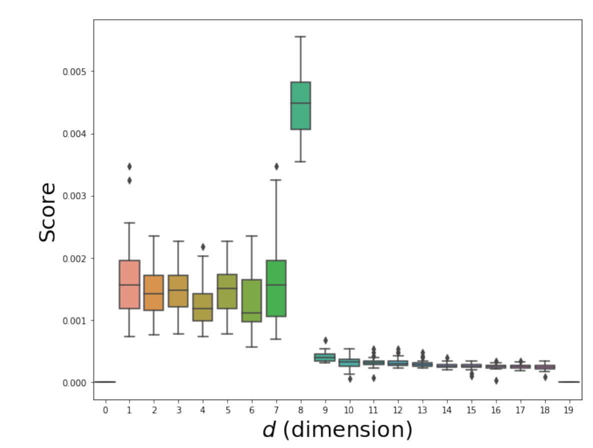
Dimension recovery

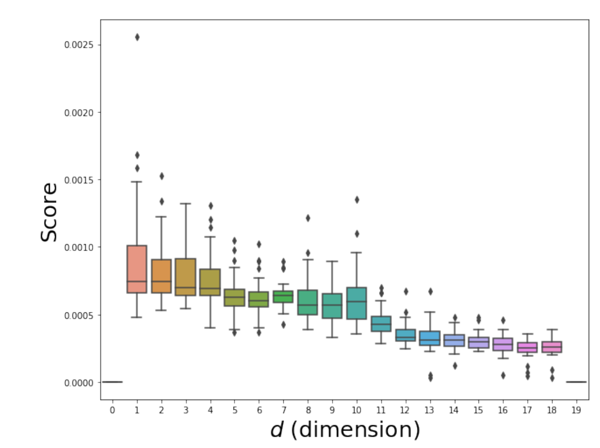
Algorithm to recover the dimension if belongs to a candidate set $\mathcal{D}$:

- **Input**: $(A, \mathcal{D})$ the adjacency matrix and $\mathcal{D} = \{d_1, \cdots, d_{\max}\}$
- $\text{Score}(d) \leftarrow \text{Gap}_1(d)$ obtained running $\text{HEiC}(A, d)$ as subroutine.
- **Return**: $d^* \leftarrow \arg\max_{d \in \mathcal{D}} \text{score}(d)$.









Extensions

- The algorithm can be extended, almost without modification to compact lie groups and symmetric spaces [2].
- Consider the closed unit ball in $\mathbb{R}^d$ as underlying space and probability matrix $\Theta_{ij} = f(\langle x, y \rangle)$. The harmonic analysis on the ball allows for an extension of the previous methods, with slight modifications.

References

[1] C Borgs et al. "Convergent sequences of dense graphs II. Multiway cuts and statistical physics". In: *Annals of Mathematics* 176.1 (2012), pp. 151–219.

[2] Y. De Castro, C. Lacour, and T.M Pham Ngoc. "Adaptive estimation of nonparametric geometric graphs". In: (2017). URL: `https://arxiv.org/pdf/1708.02107.pdf`.

[3] V Koltchinskii and E Giné. "Random matrix approximation of spectra of integral operators". In: *Bernoulli* (2000), pp. 113–167.