Signed Determinantal Point Processes

Victor-Emmanuel Brunel

ENSAE Paris-Saclay

Selection models

- Items labeled $1, 2, \ldots, N$.
- Selection model: Random subset of items.
- E.g., ferromagnetic / antiferromagnetic Ising models (statistical physics, e.g., particles with positive spin)
- Objective: Develop a simple model ($\ll 2^N$ parameters), that is tractable, computationally simple and accurate for given applications.
- ML application: Recommender systems.

DPP's

- Model for **random binary vectors** $X = (X_1, \ldots, X_N) \in \{0,1\}^N$
- Equivalently, **random subset** $Y \subseteq [N]$ s.t.

$$\mathbb{P}[J \subseteq Y] = \det(K_J), \forall J \subseteq [N]$$

for some matrix $K \in \mathbb{R}^{N \times N}$.

$K = \begin{pmatrix} & \overset{J}{\leftrightarrow} & \\ & [K_J] & \\ & & \end{pmatrix} \in \mathbb{R}^{N \times N}$

- Ex: $\mathbb{P}[1 \in Y] = K_{1,1}$, $\mathbb{P}[1,2 \in Y] = K_{1,1}K_{2,2} - K_{1,2}K_{2,1}$.
- If $I_N - K$ is invertible, DPP(K) is also an **L-ensemble**:

$$\mathbb{P}[Y = J] = \frac{\det(L_J)}{\det(I_N + L)}, \quad \forall J \subseteq [N]$$

where $L = K(I_N - K)^{-1}$ ($\Leftrightarrow K = L(I_N + L)^{-1}$).

- If K is **symmetric**: DPPs can model **repulsive** interactions: $(X_1, X_2, \ldots, X_N)$ are **negatively associated** ($\gg$ negative correlation), i.e.,

$$cov\big(f(X_i, i \in S), g(X_j, j \in T)\big) \le 0,$$

for all disjoint $S, T \subseteq [N]$ and coordinatewise nondecreasing functions f, g.

E.g., $cov(X_i, X_j) = -K_{i,j}^2 \le 0$.

- In general, $cov(X_i, X_j) = -K_{i,j}K_{j,i}$.
- **Signed DPP**: $K_{i,j} = \varepsilon_{i,j}K_{j,i}$, $\varepsilon_{i,j} \in \{-1,1\}$: $cov(X_i, X_j) = -\varepsilon_{i,j}K_{i,j}^2$

  $\Longrightarrow$ Allow for both **positive** and **negative** dependence.
- **Admissible kernels:** When $I_N - K$ is invertible, DPP(K) is well defined iff $L = K(I_N - K)^{-1}$ is a **P_0-matrix** (i.e., all its principal minors are nonnegative).
- Examples of admissible kernels:
  - Any symmetric K with $0 \le K \le I_N$
  - Any $K = D + \lambda A$ for some $\lambda \in \left(0, \frac{1}{2}\right)$, diagonal matrix D with $D_{i,i} \in [\lambda, 1-\lambda]$ and $A \in [-1,1]^N$.

Identification and learning

Given $Y \sim$ DPP(K), identify and learn K:
The Principal Minor Assignment Problem

- DPP(K) is completely determined by the principal minors of K.
- Given a class $\mathcal{T} \subseteq \mathbb{R}^{N \times N}$, PMA asks:

  I. *What is the collection of all matrices $H \in \mathcal{T}$ that have the same list of principal minors as K ?*

  II. *Given an available list of prescribed principal minors, how to find a matrix $H \in \mathcal{T}$ whose principal minors are given by that list, using as few queries from that list as possible ?*

  I $\Leftrightarrow$ **Identification of K**

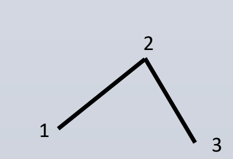
  II $\Leftrightarrow$ **Learning K efficiently**

- Here, $\mathcal{T}$ is the class of **signed kernels**, i.e., $K_{i,j} = \varepsilon_{i,j}K_{j,i}$, $\varepsilon_{i,j} \in \{-1,1\}$.

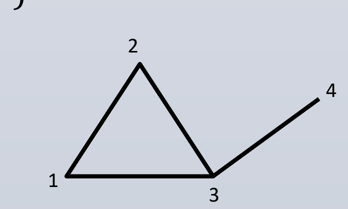
More precisely, given a family $(a_J)_{J \subseteq [N], J \neq \emptyset} \subseteq \mathbb{R}$, characterize $\mathcal{S} = \{H \in \mathcal{T}: det(H_J) = a_J, \forall J \subseteq [N], J \neq \emptyset\}$ and find $H \in \mathcal{S}$ efficiently.

- Main idea:
  - $J = \{i\}$: $H_{i,i} = a_{\{i\}}$.
  - $J = \{i,j\}$: $\varepsilon_{i,j}H_{i,j}^2 = a_{\{i\}}a_{\{j\}} - a_{\{i,j\}}$ $\Longrightarrow$ These determine the **adjacency graph** of any solution H and the values of $\varepsilon_{i,j}$'s.
- **Adjacency graph** of a solution $H \in \mathcal{S}$: $G_H = ([N], E_H)$, with $E_H = \{\{i,j\}: i \neq j, H_{i,j} \neq 0\} = \{\{i,j\}: i \neq j, a_{\{i\}}a_{\{j\}} - a_{\{i,j\}} \neq 0\}$.

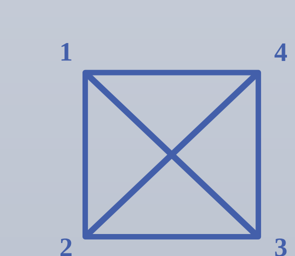
$$H = \begin{pmatrix} * & * & 0 \\ * & * & * \\ 0 & * & * \end{pmatrix}$$



$$H = \begin{pmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & * \\ 0 & 0 & * & * \end{pmatrix}$$



- Remark: $\{i,j\} \notin E_K \Leftrightarrow X_i, X_j$ are independent ($X_i = 1_{i \in Y}$, where $Y \sim$ DPP(K).)
- To recover the signs of $H_{i,j}$'s, use higher order principal minors, associated with small **positive cycles** in G (i.e., $\prod_{\{i,j\} \in C} \varepsilon_{i,j} = 1$)
- Key idea: $\det H_J = \sum_{\sigma \in \mathfrak{S}_J} (-1)^\sigma \prod_{i \in J} H_{i,\sigma(i)}$ $\Longrightarrow$ Decompose each $\sigma \in \mathfrak{S}_J$ as a product of cycles
- Example: If C is an induced cycle (i.e., with no chords) in G_K, with vertex set J, then $\det K_J = F\big(K_{i,i}, K_{i,j}^2 : i,j \in J\big) \pm \big(1 + \prod_{\{i,j\} \in C} \varepsilon_{i,j}\big) \prod_{\{i,j\} \in C} K_{i,j}$
- Ideal situation: There is a basis of induces cycles that are all positive.
- Main issue: Induced cycles may be negative and non induced positive cycles may have more than one positive **traveling**:



$1 \to 2 \to 3 \to 4 \to 1$

$1 \to 3 \to 2 \to 4 \to 1$

$1 \to 2 \to 4 \to 3 \to 1$

- **Positive travelings** can not be dissociated from the principal minors (e.g., $H_{12}H_{23}H_{34}H_{41} + H_{13}H_{32}H_{24}H_{41} + H_{12}H_{24}H_{43}H_{32}$)
- Assumptions:
  - H is dense, i.e., the graph G_H is complete (this allows to only consider cycles of size 3 and 4)
  - The magnitudes $|H_{i,j}|$ are in *general position* (this allows to separate positive travelings of a given cycle)

$\Longrightarrow$ **Theorem:** The signs can be (not uniquely) recovered using the a_J's, for $\#J \le 4$.

Algorithm

Input: Family $(a_J)_{J \subseteq [N], J \neq \emptyset} \subseteq \mathbb{R}$ of prescribed principal minors.

Output: Matrix $H \in \mathcal{T}$ with $\det H_J = a_J$, $\forall J \subseteq [N], J \neq \emptyset$.

Step 1: Set $\mathrm{H}_{i,i} = a_{\{i\}}$ for all $i \in [N]$.

Step 2: Set $|\mathrm{H}_{i,j}| = |a_{\{i\}}a_{\{j\}} - a_{\{i,j\}}|$ for all $i,j \in [N], i \neq j$.

Step 3: Set $\varepsilon_{i,j} = sign\big(a_{\{i\}}a_{\{j\}} - a_{\{i,j\}}\big)$ for all $i,j \in [N], i \neq j$ s.t. $\mathrm{A}_{i,j} \neq 0$.

Step 4: Find the set $\mathcal{J}^+$ of all triples (i,j,k) such that $\varepsilon_{i,j}\varepsilon_{j,k}\varepsilon_{i,k} = 1$ and find the sign of $H_{i,j}H_{j,k}H_{i,k}$ from $a_{\{i,j,k\}}$.

Step 5: For all $S \subseteq [N]$ if size 4, use a_S in order to find $\prod_{\{i,j\} \in C: i<j} K_{i,j}$ for all the (at most three) positive cycles C that have vertex set S.

Step 6: By Gaussian elimination on $\{+,-\}$, find a sign assignment of all the $H_{i,j}$'s that agree with signs of the products found in Steps 4 and 5.

Conclusions

Theorem: Under the previous assumptions, the set of solutions $\mathcal{S}$ is completely determined by the a_J's, for $\#J \le 4$, and there is a polynomial time algorithm that outputs one solution.

In general, the signs of the $H_{i,j}$'s would be (not uniquely) determined by the a_J's, where J is the vertex set of a positive cycle in some *simple* family of spanning cycles.

Open questions:

1. In general, how to find such a family of spanning cycles efficiently?
2. What properties (analogous to negative association) are satisfied by signed DPP's?
3. For signed DPP's, the eigenstructure of the kernel no longer plays a significant role (e.g., for sampling). How to sample a signed DPP efficiently?

References

Determinantal Point Processes for Machine Learning. Alex Kulesza and Ben Taskar. Now Publishers Inc., Hanover, MA, USA. (2012)

Learning determinantal point processes with moments and cycles. John Urschel, Victor-Emmanuel Brunel, Ankur Moitra, and Philippe Rigollet. ICML (2017).

Negative dependence and the geometry of polynomials. Julius Borcea, Petter Branden, and Thomas Liggett. Journal of the American Mathematical Society (2009).

Convex sets of nonsingular and P-matrices. Charles R. Johnson and Michael J. Tsatsomeros. Linear and Multilinear Algebra (1995).