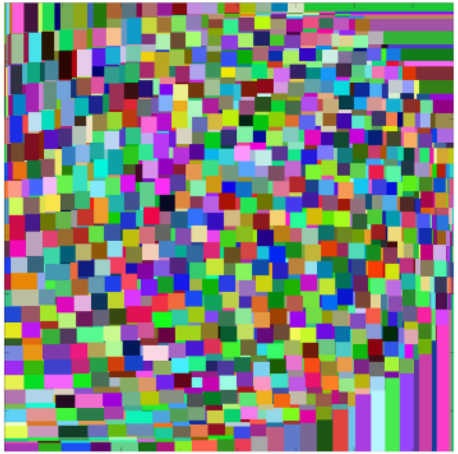
Nonparametric Relational Models with Superrectangulation

Masahiro Nakano   Ryo Nishikimi   Yasuhiro Fujiwara
Akisato Kimura   Takeshi Yamada   Naonori Ueda

NTT Communication Science Laboratories, NTT Corporation

Key takeaways

- **Super Bayes - New strategy for Bayesian machine learning**
  - Inspiration - Bayesian analogy of *lottery ticket hypothesis* for neural networks
  - Idea - Universal objects in extremal set theory (combinatorics)
  - Application - Relational data analysis with empirically stable solutions

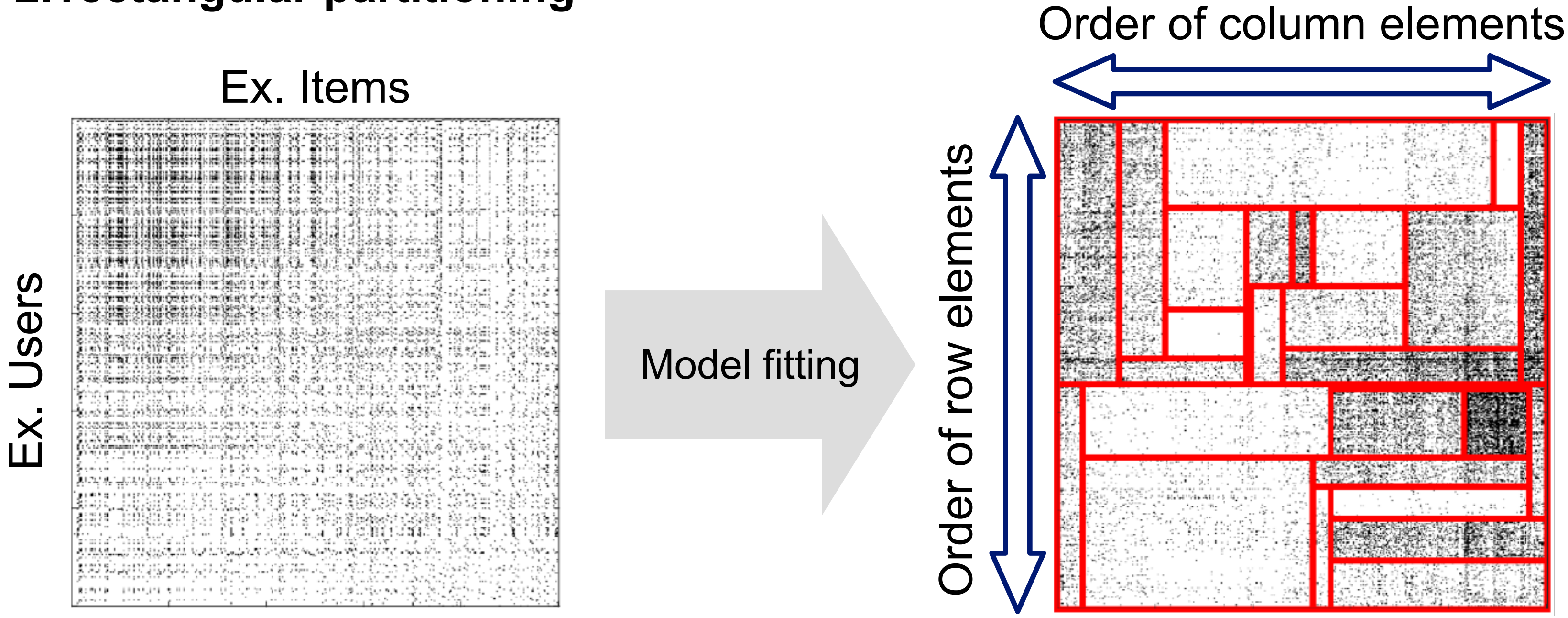


- **Superrectangulation - New notion for universal objects**
  - Sketch - Universal object that *contains* any rectangular partition with $n \in \mathbb{N}$ or fewer blocks as its part

Application: relational data analysis

- Problem of finding suitable
  1. **row and column reorderings** and
  2. **rectangular partitioning**

*Typical challenge for **Bayesian nonparametrics***
***-** It is difficult to determine model complexity (number of blocks) is manually in advance.*



Input relational matrix

Rectangular partitioning

Related work: Bayesian nonparametric array models

- Extraordinarily valuable survey papers: [Orbanz & Roy, 2013, Fan et al., 2021]

- **One-stage** model construction inspired by Chinese restaurant process (CRP)
  - Infinite relational model (product of CRPs) [Kemp et al., 2006]
  - Permuton-induced Chinese restaurant process [Nakano et al., 2021]
  - Hierarchical infinite relational model [Saad & Mansinghka, 2021]

Note: This strategy must take care of projectivity and exchangeability at the same time, which is very difficult in general.

- **Two-stage** model construction inspired by Aldous-Hoover-Kallenberg rep. using
  - Mondrian process [Roy & Teh, 2009]
  - Ostomachion process [Fan et al., 2016]
  - Binary space partitioning-tree process/forest [Fan et al., 2018]
  - Random tessellation process/forest [Ge et al., 2019]

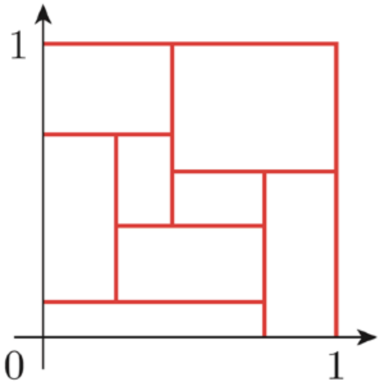
Note: Since this strategy only requires taking care of projectivity and exchangeability separately, the construction is more likely to be adopted in many cases.

Standard strategy of Bayesian nonparametric (1/2)

- **Two-stage model** construction based on Aldous-Hoover-Kallenberg representation
  1. Draw random rectangulation of [0,1]x[0,1] from some stochastic process
  2. Draw virtual locations on [0,1] for each *i*-th row and *j*-th column of input matrix:

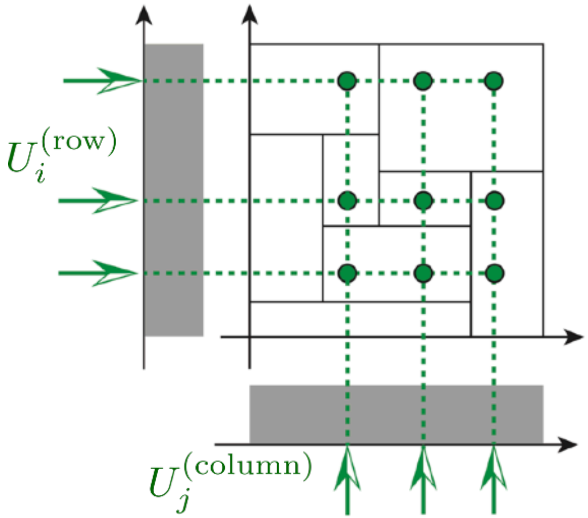
$$U_i^{(\mathrm{row})} \sim \mathrm{Uniform}([0,1]), \quad U_j^{(\mathrm{column})} \sim \mathrm{Uniform}([0,1])$$

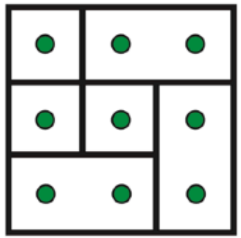
1. Draw rectangulation



from stochastic process

2. Draw virtual locations



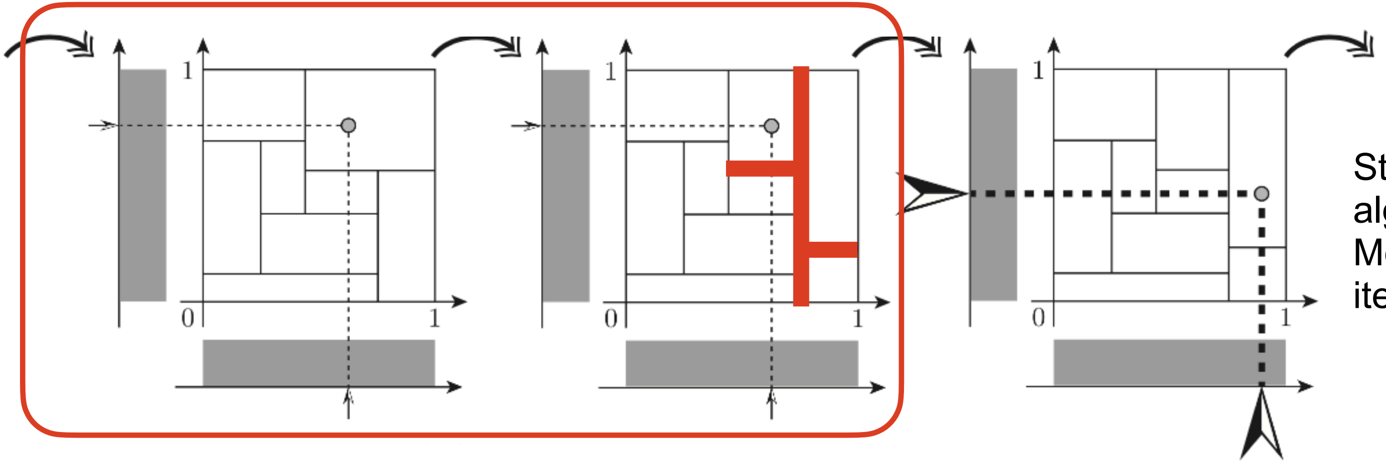


Obtain partition of input matrix

Standard strategy of Bayesian nonparametric (2/2)

- **Bayesian inference** alternating between two update rules:
  - Update **random rectangulation** of [0,1]x[0,1]
  - Update **virtual locations** on [0,1] for each *i*-th row and *j*-th column of input matrix

Rectangulation update



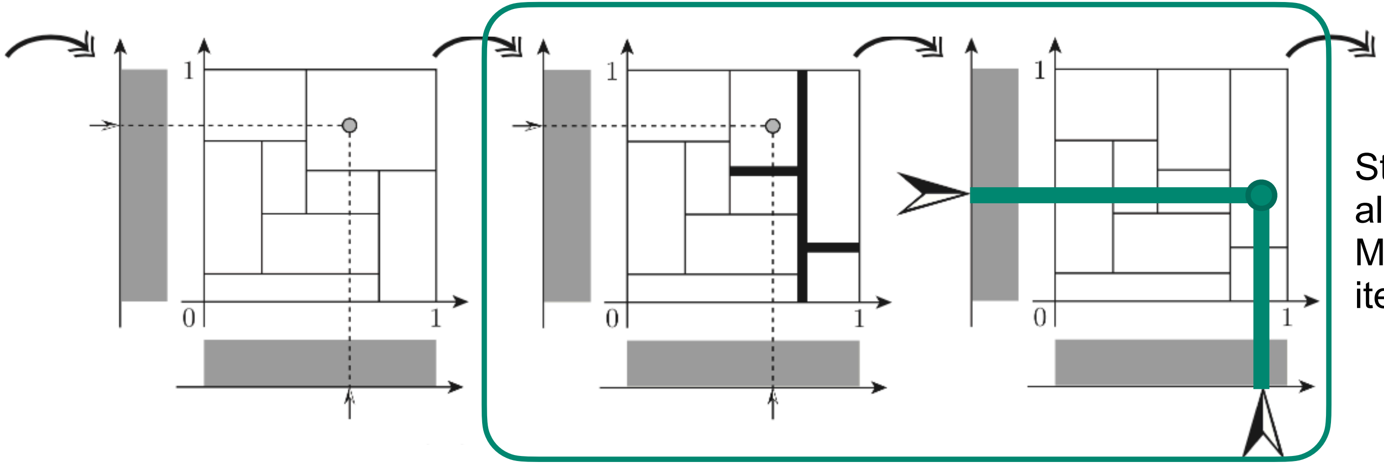
Standard Bayesian inference algorithm (e.g., Markov chain Monte Carlo) that alternately iterates between two updates

Typically, updates of rectangular partitions are often caught in bad local optima.

Standard strategy of Bayesian nonparametric (2/2)

- **Bayesian inference** alternating between two update rules:
  - Update **random rectangulation** of [0,1]x[0,1]
  - Update **virtual locations** on [0,1] for each *i*-th row and *j*-th column of input matrix

Virtual location updates



Standard Bayesian inference algorithm (e.g., Markov chain Monte Carlo) that alternately iterates between two updates

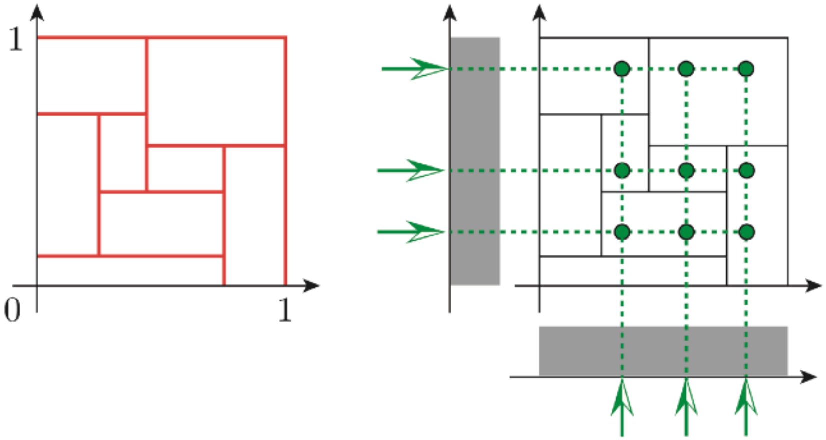
Typically, updates of rectangular partitions are often caught in bad local optima.

Nonparametric Bayes vs. Super Bayes

Our strategy: introducing sufficiently **redundant and deterministic rectangulation** instead of random rectangulation to avoid the search problem of rectangulation.

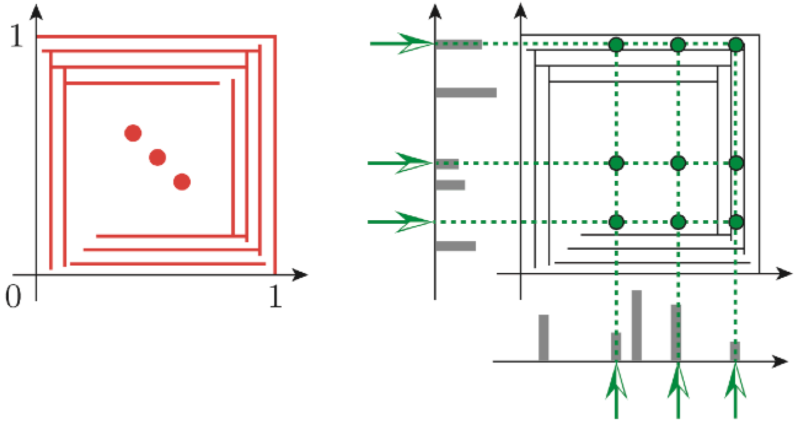
Nonparametric Bayes

- Random and parsimonious rectangulation
- Uniform random coordinates



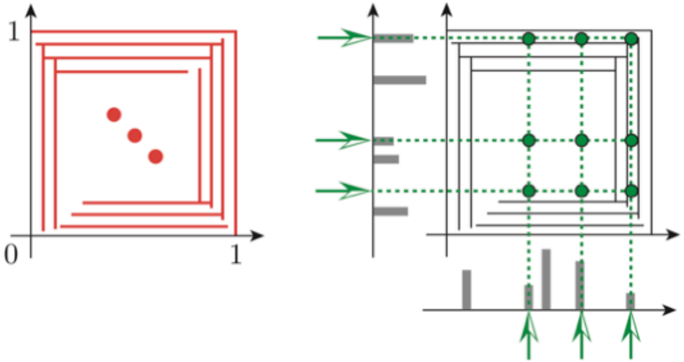
Super Bayes (1st contribution)

- **Deterministic** and **redundant** rectangulation
- Sparse atomic random coordinates



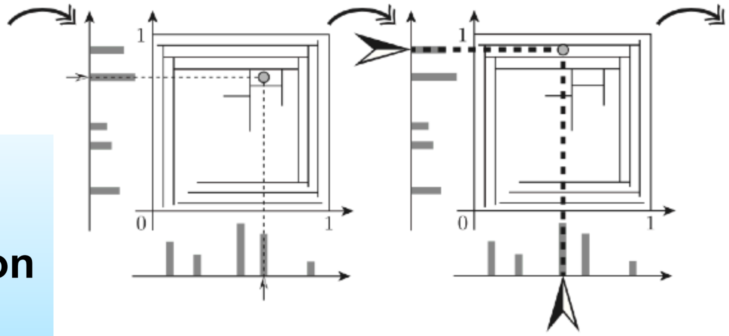
Super Bayesian relational model

- **Two-stage model construction**
  - Set **superrectangulation** as **deterministic rectangulation** of [0,1]x[0,1]
  - Draw **virtual locations** on [0,1] for row and column from Dirichlet process (DP) mix.



- **Bayesian inference**
  - Keep fixed superrectangulation
  - Update **virtual locations** on [0,1]

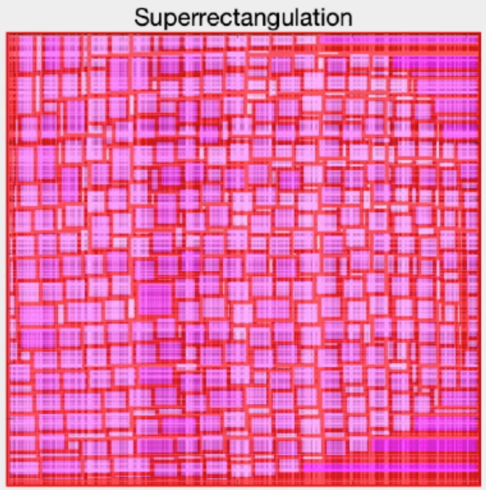
Virtual location updates



The superrectangulation, once set, can remain fixed, and a rectangulation that fits the data well can be attributed to the problem **of which sub-rectangulation of the superrectangulation to extract**.

Super Bayesian relational model

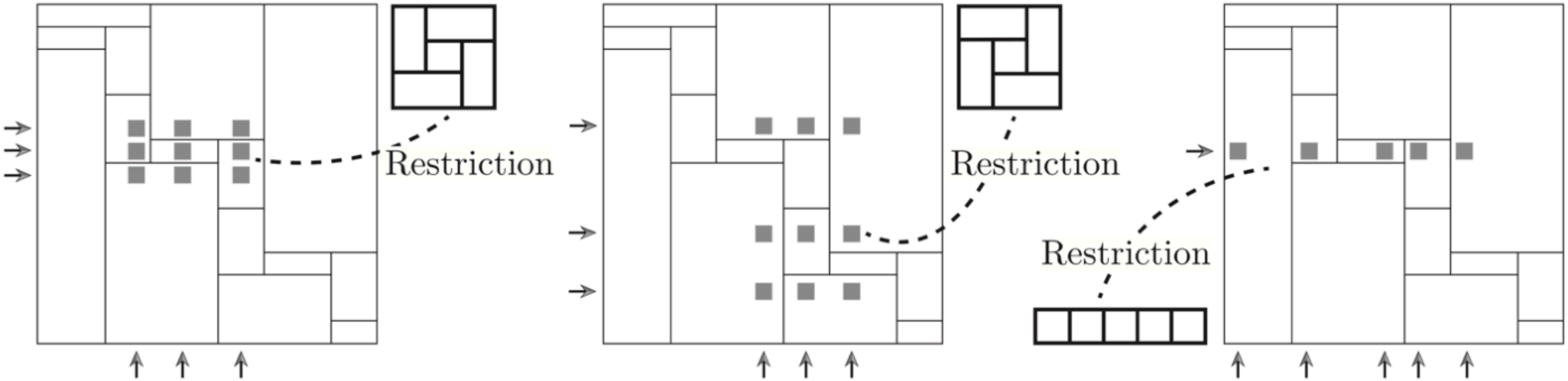
Obtain the target rectangular partition by extracting only some of the partition structure from the superrectangulation and discarding all other partitions.



After this, we will discuss in detail how the requirements for a superrectangulation and the challenges to its construction method.

2nd contribution - Notion of superrectangulation

Definition (sketch) - Universal object that *contains* any rectangular partition with $n \in \mathbb{N}$ or fewer blocks as its part (i.e., restriction to its subregion).



- This idea may be reminiscent of
  - The lottery ticket hypothesis in deep neural networks [Frankle & Carbin, 2019] and
  - Universal objects in extremal combinatorics [Engen & Vatter, 2021].

Two inspirations from existing wisdom

- **Superpermutation** (Comprehensive reference: [Engen & Vatter, 2021])

Theorem (Theorem 3.1 in [Miller, 2009]):
There is a word over the alphabet $\{1,2,\ldots,n+1\}$ of length $(n^2+n)/2$ containing subsequences order-isomorphic to every permutation of length n.

- **Surjective map from permutations to rectangulations** [Reading, 2012]

Theorem (Proposition 4.2 in [Reading, 2012]):
There is a surjective map from permutations to generic rectangulations.

Two inspirations from existing wisdom

- **Superpermutation** (Comprehensive reference: [Engen & Vatter, 2021])
  - **Zigzag word** [Miller, 2009] **-** For $n = 4$, **" 1 3 5 4 2 1 3 5 4 2 "** contains

every permutations of length $n = 4$ as subsequences $\left\{ \begin{array}{l} 1234, 1243, 1324, 1342, 1423, 1432, \ldots \\ 2134, 2143, 2314, 2341, 2413, 2431, \ldots \\ 3124, 3142, 3214, 3241, 3412, 3421, \ldots \\ \ldots \end{array} \right.$

- **Surjective map from permutations to rectangulations** [Reading, 2012]

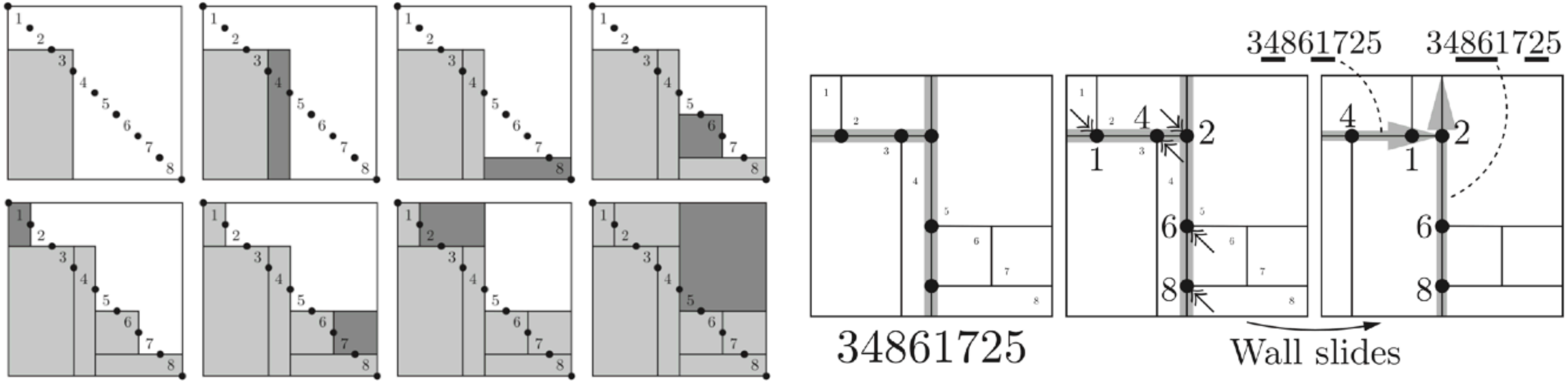
Theorem (Proposition 4.2 in [Reading, 2012]):
There is a surjective map from permutations to generic rectangulations.

Two inspirations from existing wisdom

- **Superpermutation** (Comprehensive reference: [Engen & Vatter, 2021])

> **Theorem** (Theorem 3.1 in [Miller, 2009]):
> There is a word over the alphabet $\{1,2,\ldots,n+1\}$ of length $(n^2+n)/2$ containing subsequences order-isomorphic to every permutation of length n.

- **Surjective map from permutations to rectangulations** [Reading, 2012]



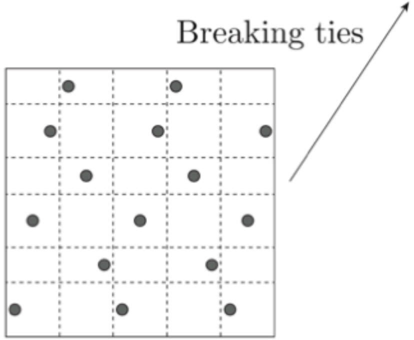
2nd contribution - superrectangulation candidates

NTT

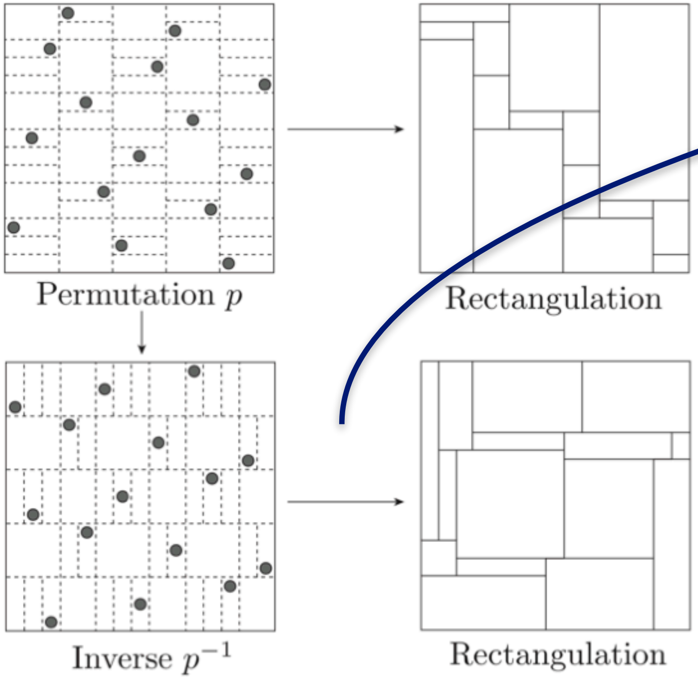
Our Strategy - Zigzag Rectangulation

1. Introduce zigzag word as a superpermutation
2. Convert zigzag words into permutations using “breaking ties” operations
3. Convert the corresponding permutations into rectangulations

Case of $n = 5$

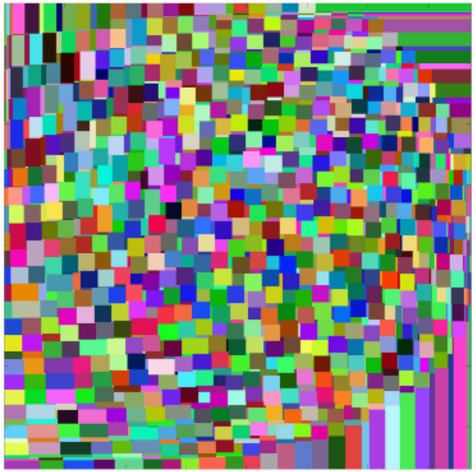


Geometric representation of **“zigzag word”** [Miller, 2009]



Transformation from permutation to rectangulation [Reading, 2012]

Case of $n = 55$

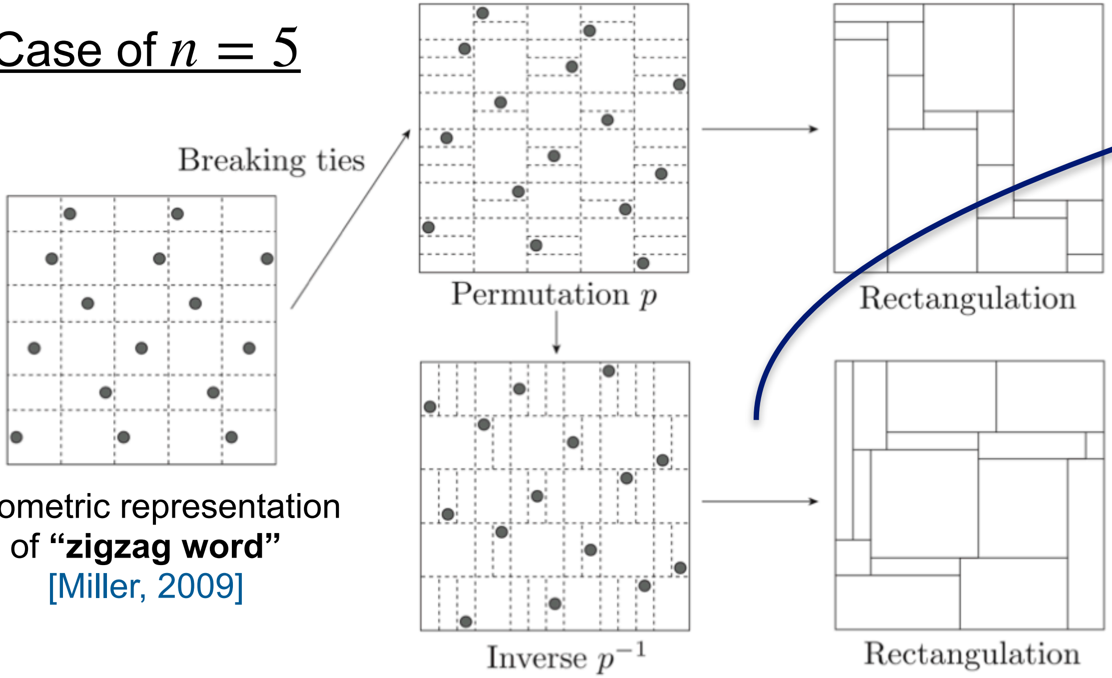


2nd contribution - superrectangulation candidates

Open question - Is zigzag rectangulation a superpermutation?

- For small n, it can be manually verified that most rectangular partitions are included.
- In application, the zigzag rectangulation can serve as a pseudo-superrectangulation substitute.

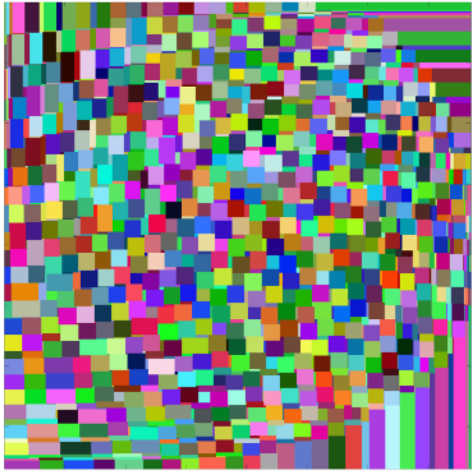
Case of $n = 5$



Geometric representation of **“zigzag word”** [Miller, 2009]

Transformation from permutation to rectangulation [Reading, 2012]

Case of $n = 55$



Experimental results

NTT

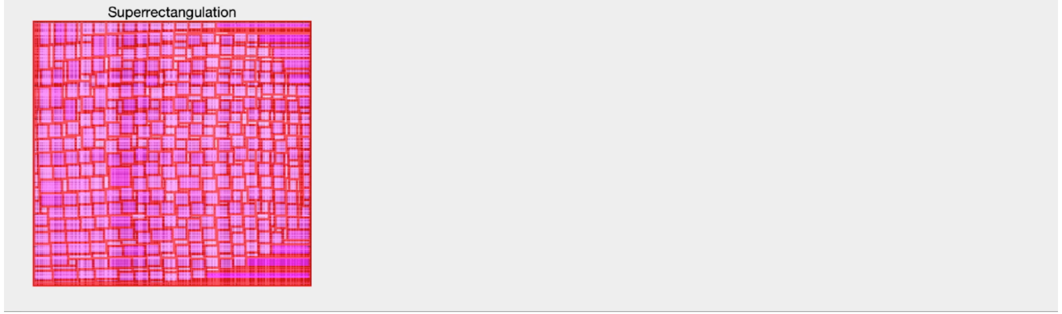
- **Task** - Matrix completion (Train:Test = 8:2), i.e., prediction for missing elements.
- **Dataset** - Four real-world network adjacency matrices (standard benchmark data).
    - Extract the most active 1000x1000 from the adjacency matrix of the sparse network and run 10 trials for each algorithm on 500x500 uniformly selected at random.
- **Comparison methods**:
    - Mondrian process (MP) [Roy & Teh, NeurIPS 2009]
    - Block-breaking process (BBP) [Nakano et al., NeurIPS 2020]
    - Permuton-induced Chinese restaurant process (PCRP) [Nakano et al., NeurIPS 2021]

	Nonparametric Bayes			Super Bayes	
	MP	BBP	PCRP	Zigzag	Random
Wiki	1.2838 ± 0.0094	1.2712 ± 0.0056	1.2583 ± 0.0041	$\mathbf{1.2565} \pm 0.0017$	1.2648 ± 0.0082
Facebook	1.1944 ± 0.0217	1.1818 ± 0.0197	1.1545 ± 0.0187	$\mathbf{1.1493} \pm 0.0095$	1.1682 ± 0.0234
Twitter	1.2316 ± 0.0209	1.2146 ± 0.0058	$\mathbf{1.2057} \pm 0.0092$	1.2077 ± 0.0071	1.2102 ± 0.0087
Epinions	1.4098 ± 0.0064	1.4006 ± 0.0044	1.3955 ± 0.0061	$\mathbf{1.3951} \pm 0.0054$	1.3979 ± 0.0056

Achieve prediction performance comparable to the latest results with smaller standard deviations

Conclusion and discussion

- **Super Bayes - New strategy for Bayesian machine learning**
  - Relational model with redundant and deterministic rectangulation



- **Superrectangulation - New universal objects**
  - Candidate: zigzag rectangulation induced from superpermutation.
  - Open issue: Is zigzag rectangulation a superrectangulation?

Zigzag rectangulation