An overview of a problem —A tale of matrices and graphs

Bora Uçar
CNRS and LIP, ENS de Lyon, France

Oct 2023

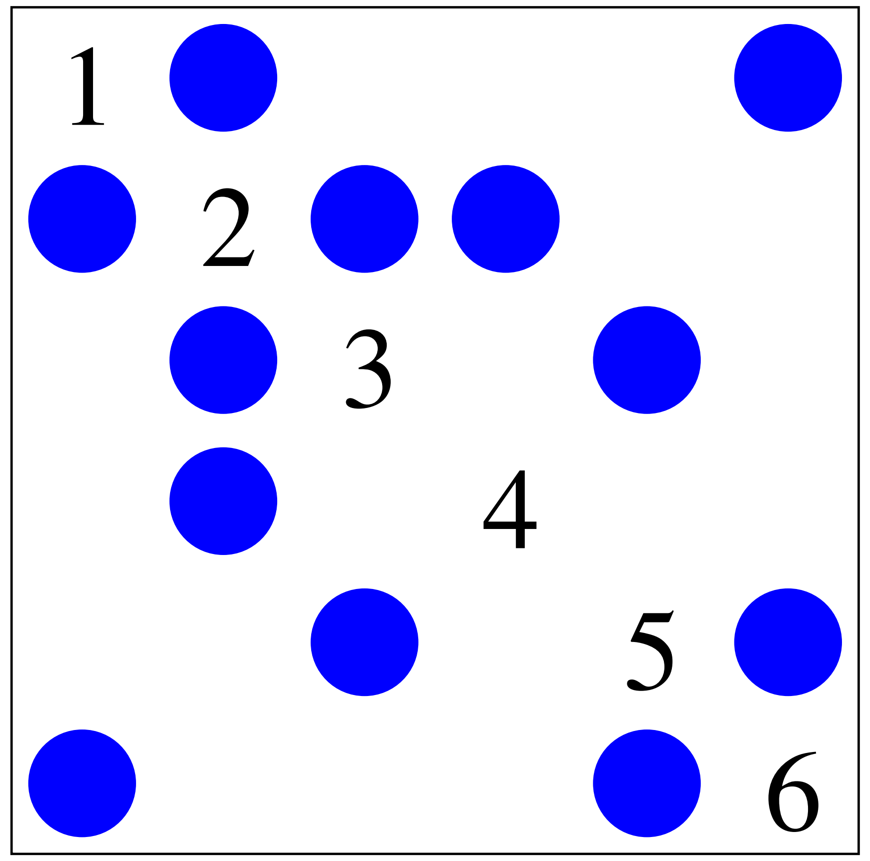
Research

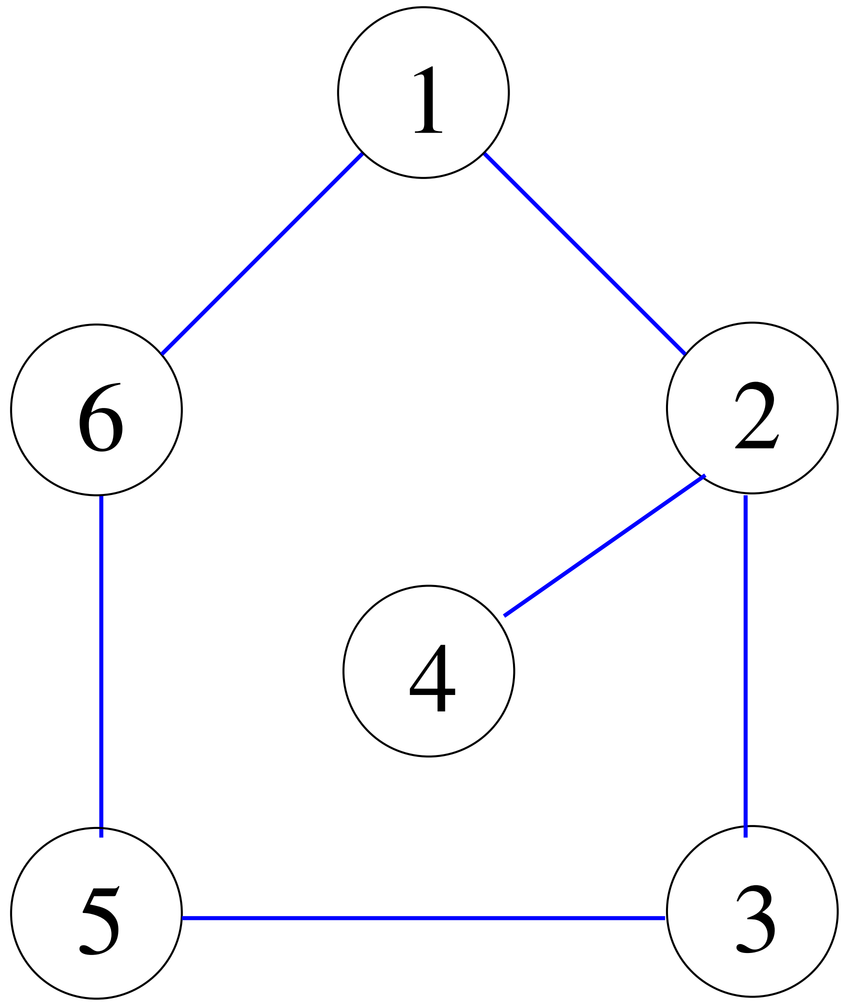
Combinatorial Scientific Computing

Development, application and analysis of combinatorial algorithms to enable scientific and engineering computations

- Parallel, high performance computing with matrices, graph algs.
- Matrices, sparse matrices, graphs…

$$\mathbf{A} = \begin{bmatrix} 4 & -1 & 0 & 0 & 0 & -1 \\ -1 & 3 & -1 & -1 & 0 & 0 \\ 0 & -1 & 3 & 0 & -1 & 0 \\ 0 & -1 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 & 2 & -1 \\ -1 & 0 & 0 & 0 & -1 & 2 \end{bmatrix}$$

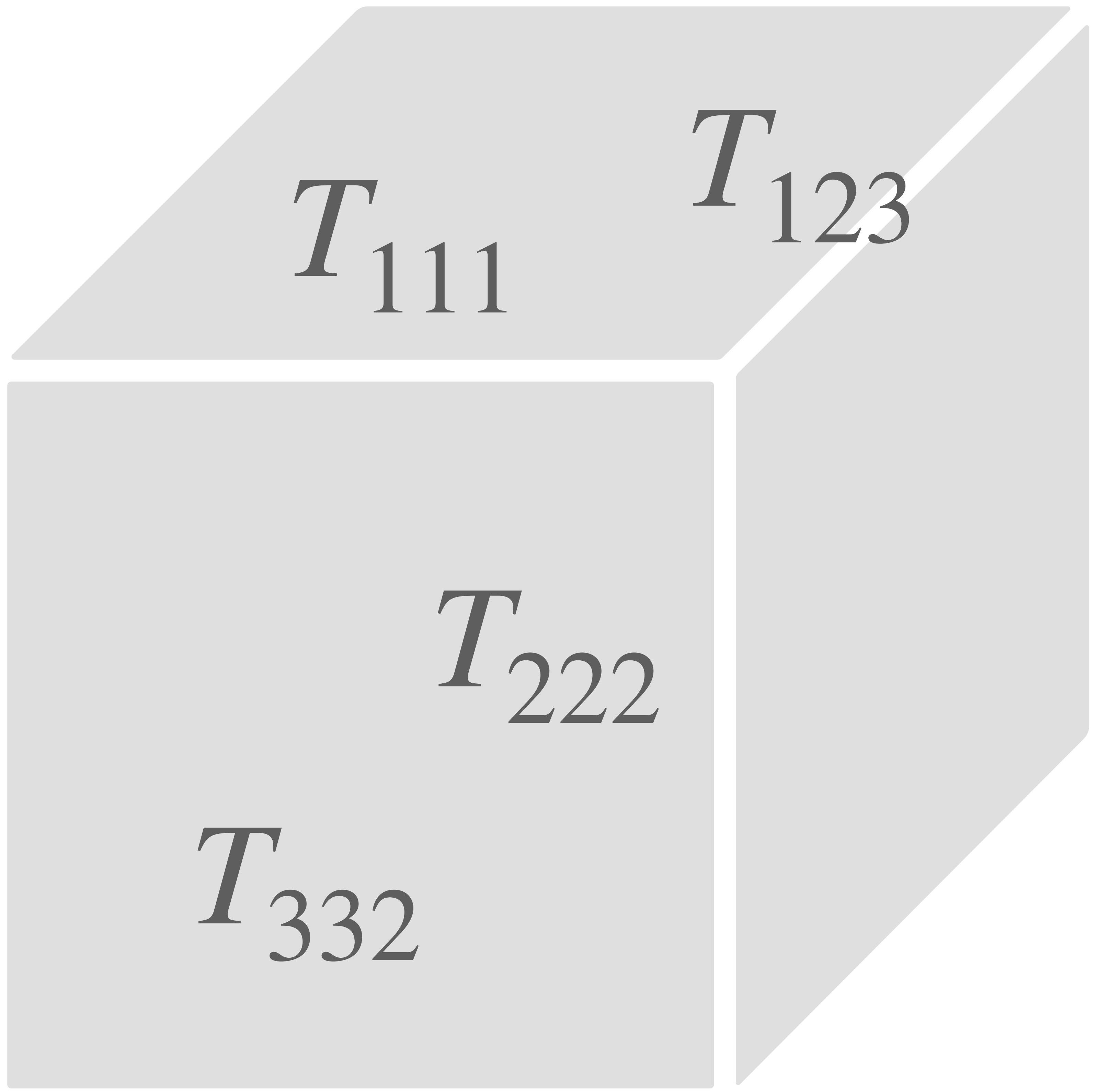


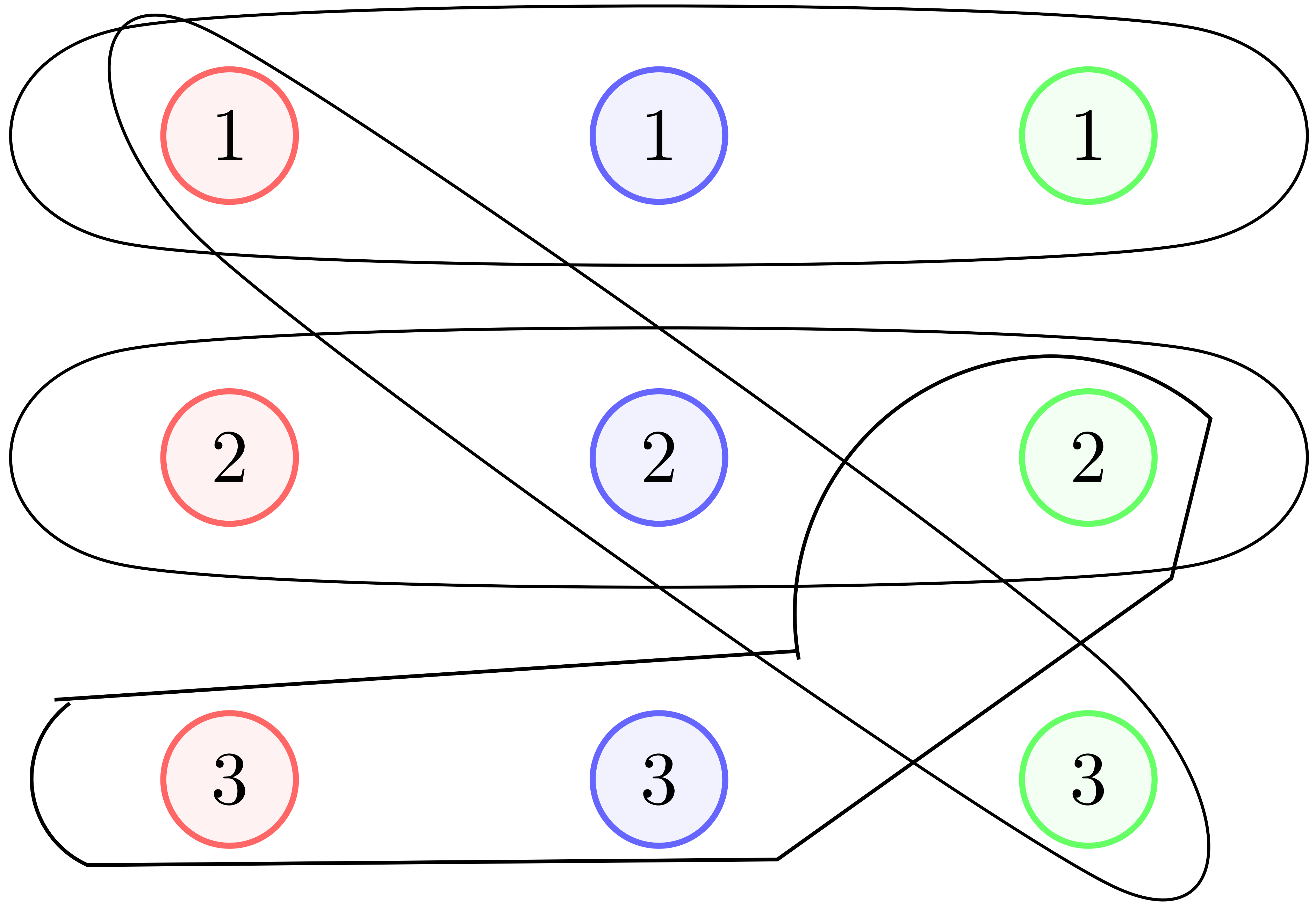


Research

Combinatorial Scientific Computing

…and tensors, sparse tensors,
hypergraphs





A sample problem:
Birkhoff — von Neumann decomposition

The problem

Birkhoff—von Neumann decomposition

Definition:

An $n \times n$ matrix $\mathbf{A}$ is doubly stochastic

$a_{ij} \geq 0$, row sums and column sums are 1.

Permutation matrix: An $n \times n$ matrix with exactly one 1 in each row and in each column (other entries are 0)

For a doubly stochastic matrix $\mathbf{A}$

there exist $\alpha_1, \alpha_2, \ldots, \alpha_k \in (0, 1]$ with $\sum_{i=1}^{k} \alpha_i = 1$ and permutation matrices $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_k$ such that:

$$\mathbf{A} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2 + \cdots + \alpha_k \mathbf{P}_k \,.$$

BvN decomposition

Applications

- **Switch design**: connections are established and the traffic from inputs to output is routed (one permutation = one set of connections).
- Similar problem in **data center networks**.
- Classical applications in assignment problems and economics.
- A (mathematical) tool in linear algebra (von Neumann trace ineq., Hoffmann — Wielandt theorem).

Background

Sparse Matrices

- A sparse matrix is a matrix with many zeros, which are not stored, and operations with them are avoided (PageRank, network analysis, simulations of all sorts, graph neural networks …)

When we started putting these models together they became very large as compared to linear programming capability at that time... It struck me that our matrices were mostly full of zeros, and if you have a set of simultaneous equations that are mostly zeros, if you pick your pivots right, you could just solve it by hand. Then I thought, well, maybe we could get the computer to do the same thing. This led "Sparse Matrices". As far as I know, I coined the word Sparse Matrix.

Harry Markowitz: Nobel Prize in 1990 (work ’52)

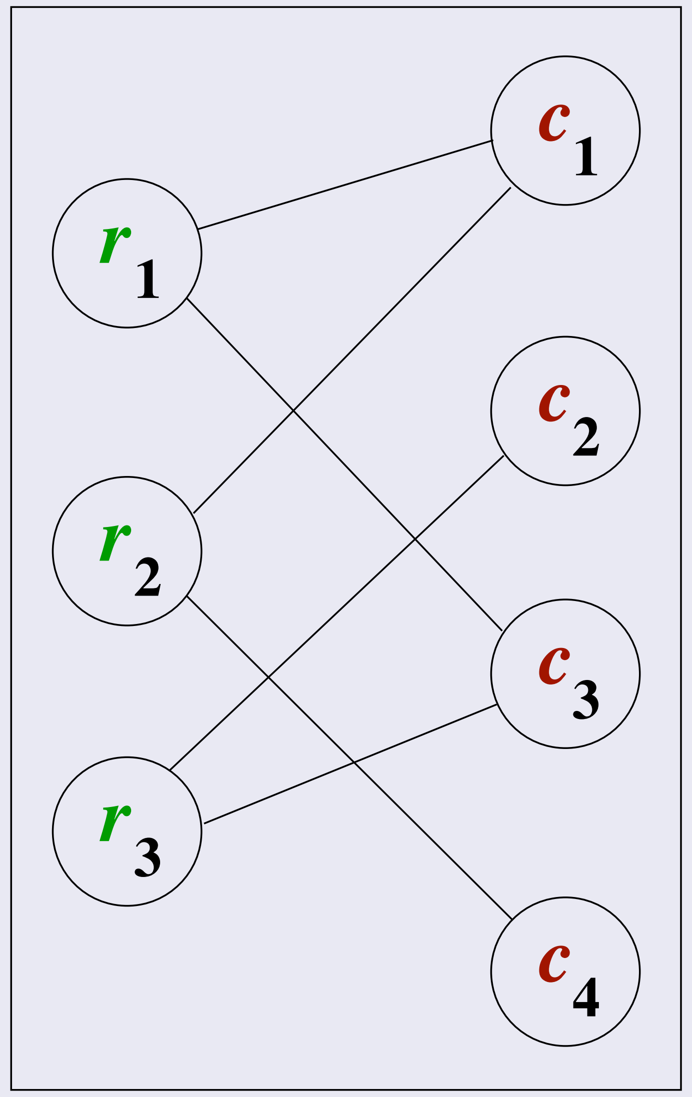
Sparse matrices and graphs (bipartite here)

The rows/columns and nonzeros of a given sparse matrix correspond (with the natural labelling) to the vertices and edges, respectively, of a graph.

Rectangular matrices

$$
A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
\begin{pmatrix} \times & & \times & \\ \times & & & \times \\ & \times & \times & \end{pmatrix}
\end{array}
$$

Bipartite graph



The set of rows corresponds to R, the set of columns corresponds to C such that for each $a_{ij} \neq 0$, (r_i, c_j) is an edge.

Matchings in (bipartite) graphs

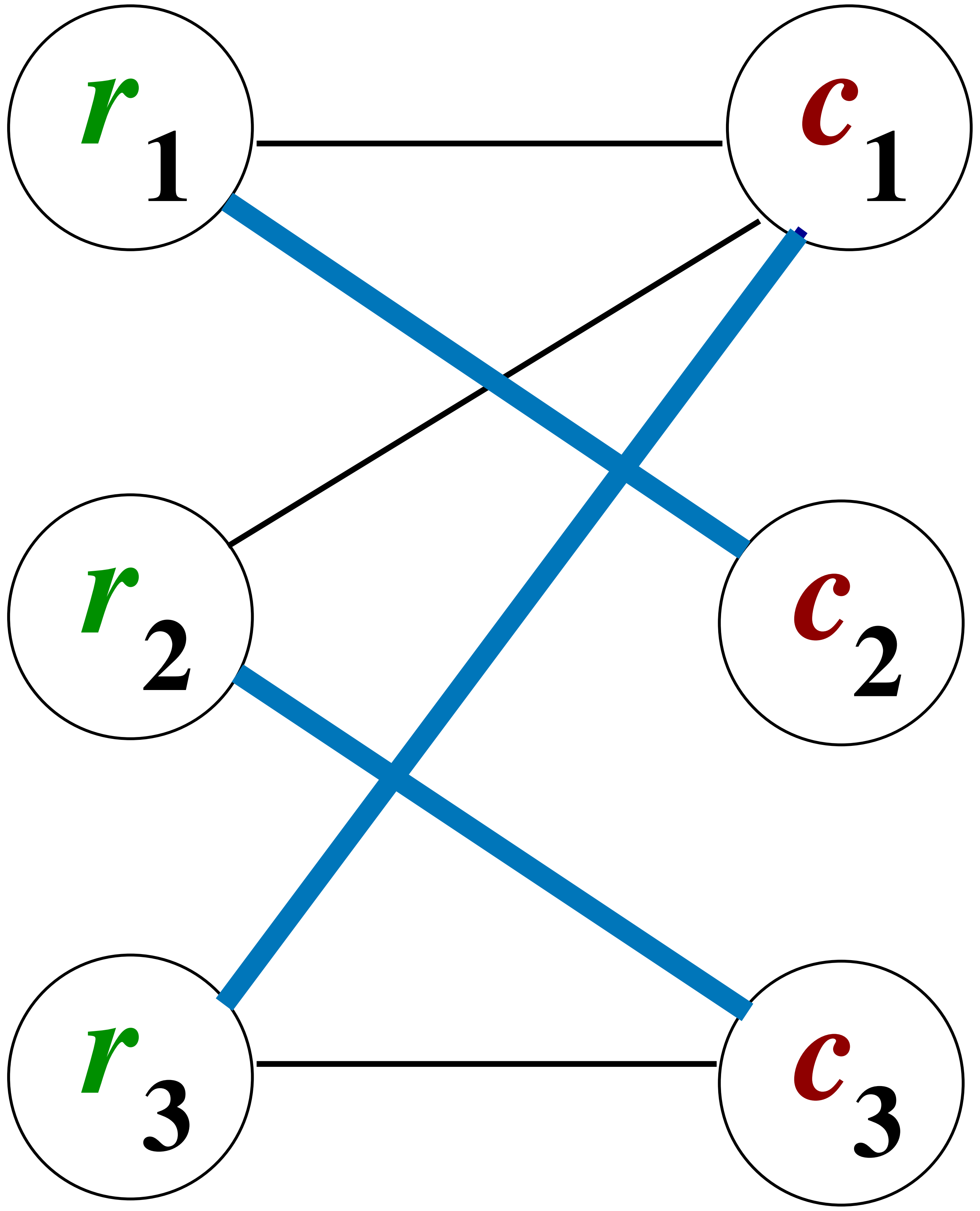
- A **matching** in a graph is a set of edges no two of which share a common vertex. We will be mostly dealing with matchings in bipartite graphs.

- In matrix terms, a matching in the bipartite graph of a matrix corresponds to a set of nonzero entries no two of which are in the same row or column.

- A vertex is said to be **matched** if there is an edge in the matching incident on the vertex, and to be **unmatched** otherwise. In a **perfect matching**, all vertices are matched.

- The cardinality of a matching is the number of edges in it. A **maximum cardinality matching** or a maximum matching is a matching of maximum cardinality. Solvable in polynomial time.

Matchings in (bipartite) graphs



$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \left(\begin{array}{ccc} \times & \bullet & \\ \times & & \bullet \\ \bullet & & \times \end{array} \right) \end{array}$$

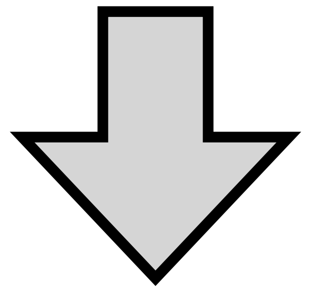
Matchings in (bipartite) graphs

- A **path** is a finite sequence of distinct vertices $v_i, v_{i+1}, \ldots, v_{i+k}$ such that $\{v_j, v_{j+1}\}$ is an edge
- An **alternating path** with respect to a matching M is a path whose edges alternate between edges in the matching M and edges not in the matching.
- If the first and last vertices on an M-alternating path P are unmatched, then P is an **M-augmenting path**. An M-augmenting path P necessarily has an odd number of edges, since if there are k matching edges, there are $k + 1$ non-matching edges in P.
- The number of edges in the matching M can be increased by computing the matching M' = M \ P + P \ M. This is symmetric difference of M and P.

Matchings in (bipartite) graphs

M-augmenting path

P:



M’ = M \ P + P \ M

Matchings in graphs

Theorem (Berge '57). *Let G be a graph and* M *be a matching in G. Then* M *is of maximum cardinality iff there is no* M*-augmenting path in G.*

Proof (hints):

(i) [*If* **M** *has max card, no augmenting path*] Let **M** be a maximum matching, and assume that there is an **M**-augmenting path P.

(ii) [*if no augmenting path, then* **M** *maximum*]. Assume that there is a **M** such that there is no **M**-augmenting path, but that **M** does not have maximum cardinality. Then there exists a matching **N** with larger cardinality than **M**. How does the symmetric difference **M** \ **N** + **N** \ **M** look?

Matchings in (bipartite) graphs

Theorem (Hall, 1935). *A bipartite graph $G = (A, B, E)$ has a matching* M *that matches all vertices of A if and only if* $|\mathrm{adj}(S)| \geq |S|$, *for all* $S \subseteq A$.

(Also known as **Hall's mariage theorem**.)

Proof (hints):

(i) If **M** matches all vertices of *A, then …*

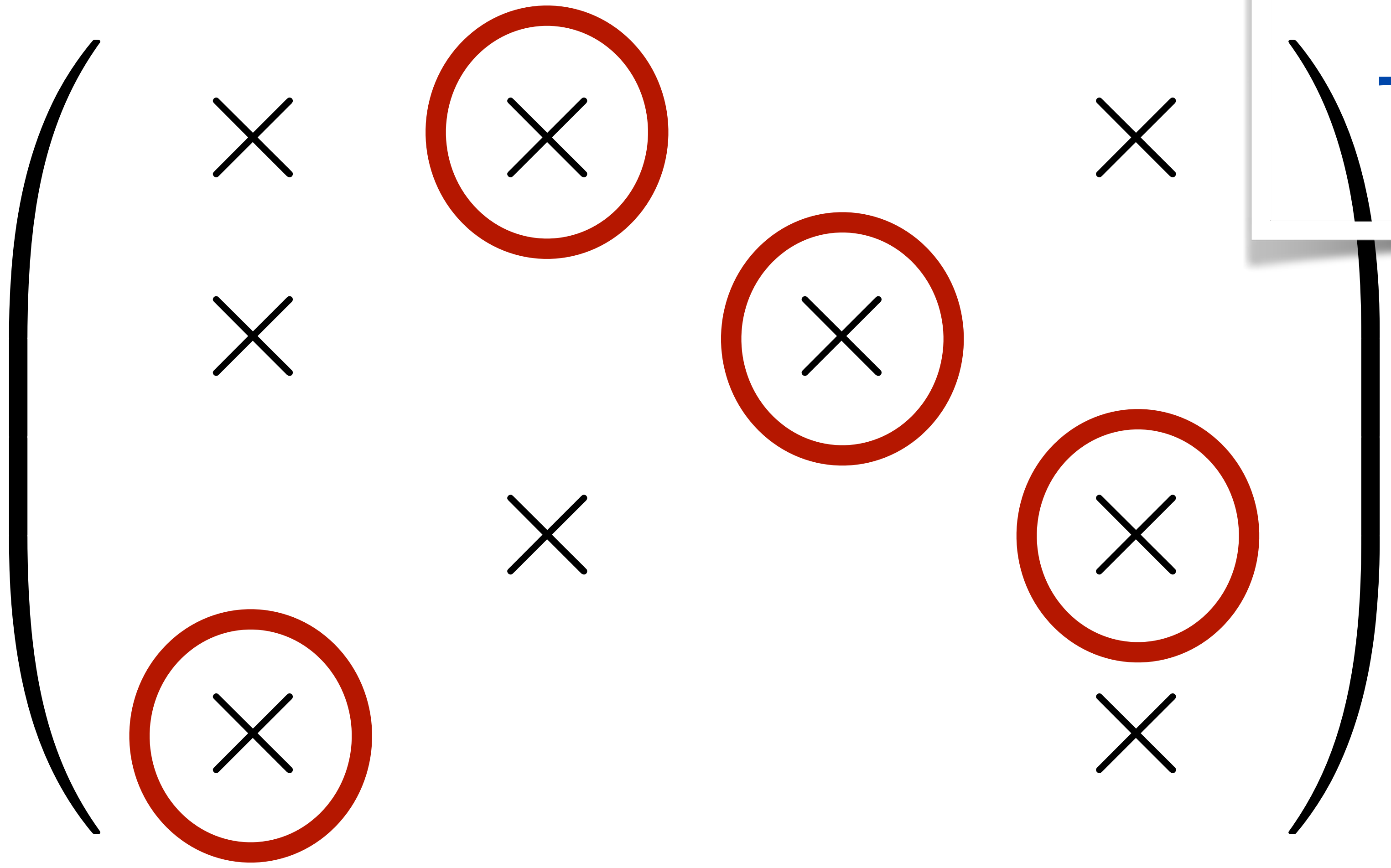
(ii) [If $|adj(S)| \geq |S|$, for all $S \subseteq A$, then perfect matching on *A*] Show: if **M** is not perfect, then there is a set S where $|adj(S)| < |S|$.

BvN decomposition

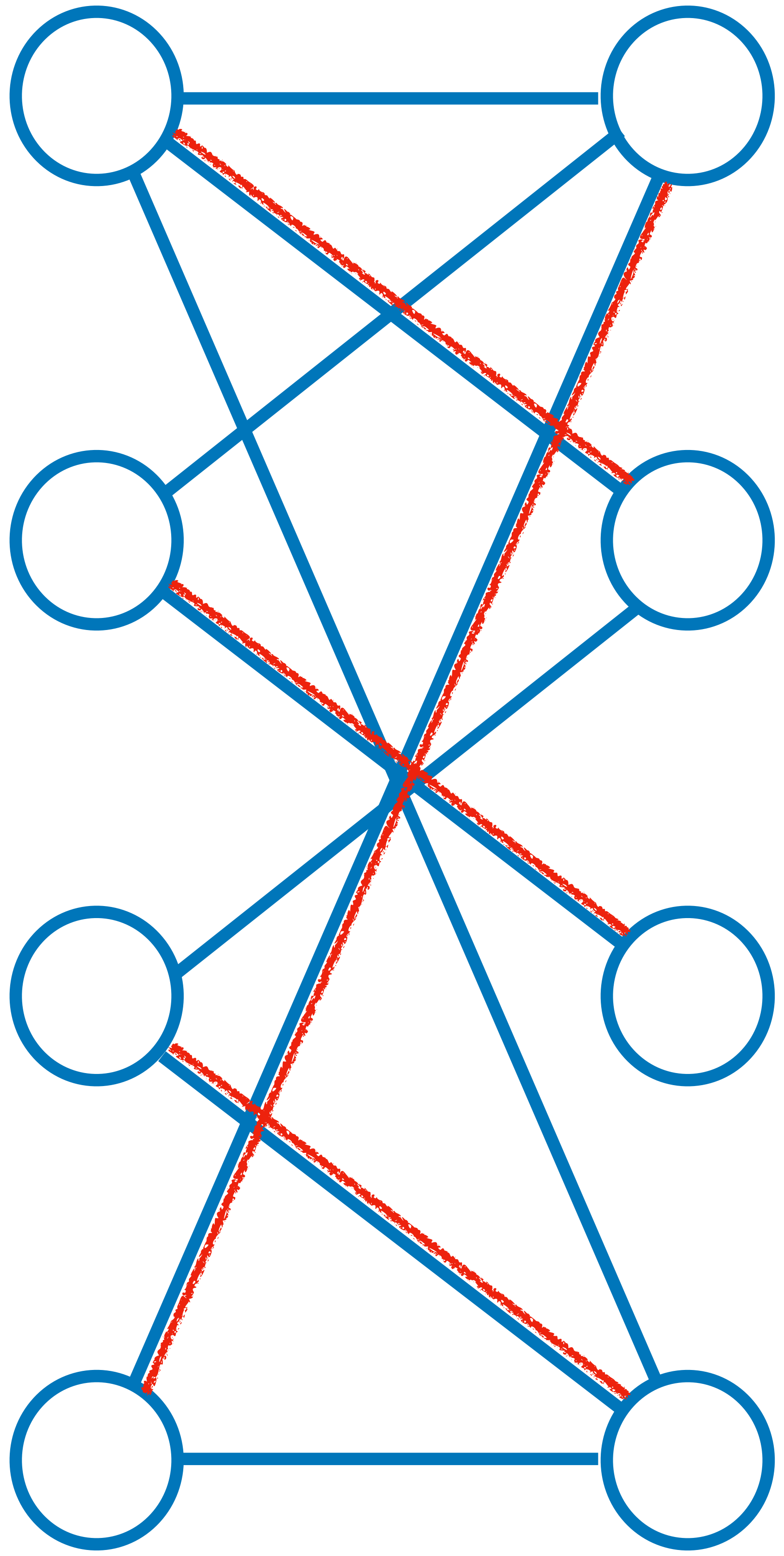
Birkhoff—von Neumann decomposition

$$\mathbf{A} = \sum \alpha_i \mathbf{P}_i$$

Permutation matrix: An $n \times n$ matrix with exactly one 1 in each row and in each column (other entries are 0)



$$\mathbf{P}_i \subseteq \mathbf{A}$$



Perfect matching in $(\mathcal{R} \cup \mathcal{C}, E)$ with $|\mathcal{R}| = |\mathcal{C}| = n$: a set of n edges no two share a common vertex.

Proof that BvN exists (Birkhoff'46)

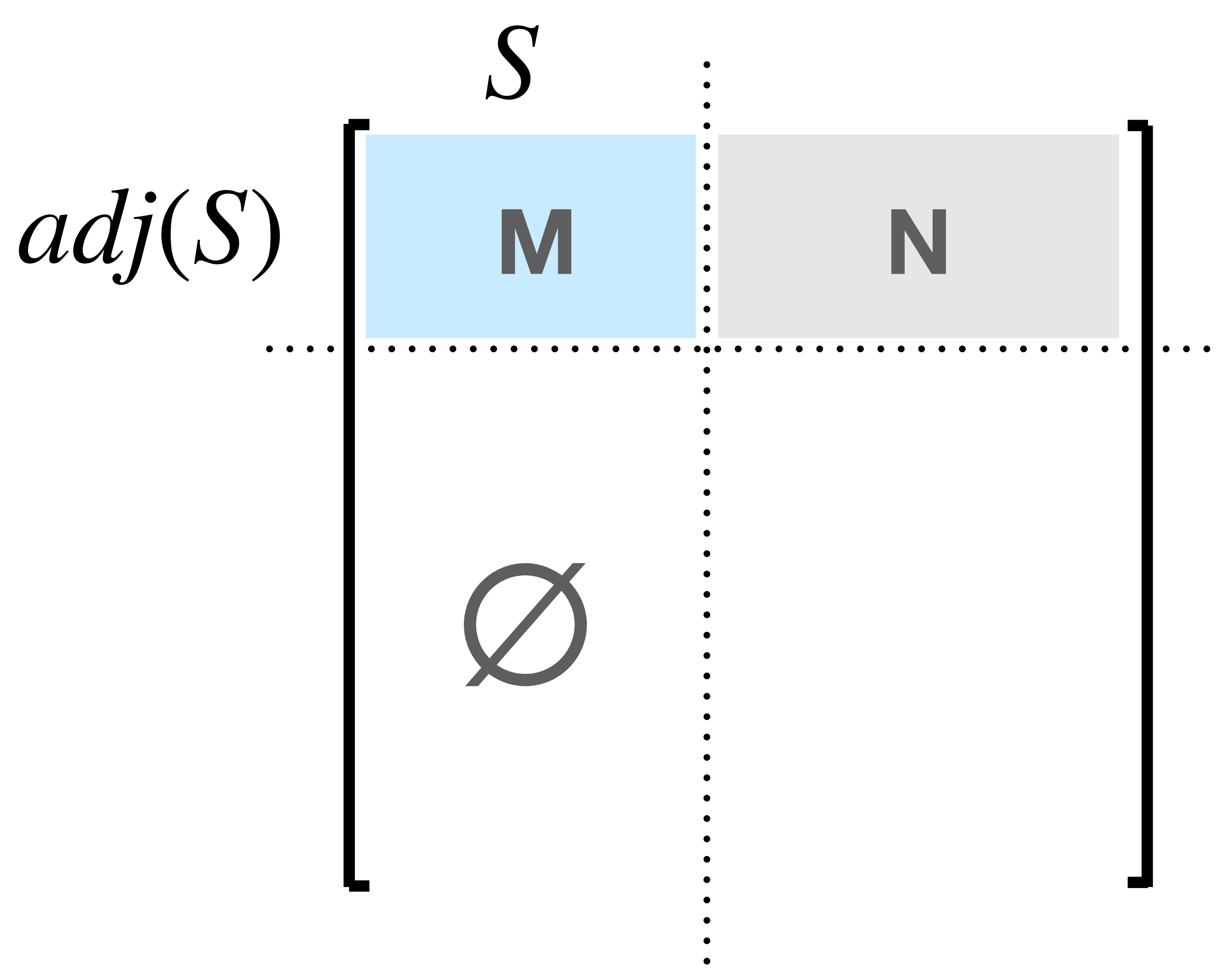
Hall's mariage theorem (Hall'35):

A bipartite graph $G = (A, B, E)$ has a matching that matches all vertices of A if and only

$|adj(S)| \geq |S|$, for all $S \subseteq A$

…holds for doubly stochastic matrices:

There is always a perfect matching in the bipartite graph of a doubly stochastic matrix.



$\sum m_{ij} = |S|$ as each column adds up to 1

$\sum m_{ij} + \sum n_{ij} = |adj(S)|$ as each row adds up to 1

If $|S| > |adj(S)|$ then **N** must contain negative entries, a contradiction.

Birkhoff—von Neumann decomposition (recall)

Definition:

An $n \times n$ matrix $\mathbf{A}$ is doubly stochastic

$a_{ij} \geq 0$, row sums and column sums are 1.

Permutation matrix: An $n \times n$ matrix with exactly one 1 in each row and in each column (other entries are 0)

For a doubly stochastic matrix $\mathbf{A}$

there exist $\alpha_1, \alpha_2, \ldots, \alpha_k \in (0,1]$ with $\sum_{i=1}^{k} \alpha_i = 1$ and permutation matrices $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_k$ such that:

$$\mathbf{A} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2 + \cdots + \alpha_k \mathbf{P}_k \,.$$

We can find perfect matchings in a bipartite graph in $\mathcal{O}(\sqrt{V}E)$

Proof that BvN exists

We got one permutation from Hall's theorem; its coefficient? Then, others by induction

1: $\mathbf{A}^{(0)} = \mathbf{A}$
2: **for** $j = 1, \ldots$ **do**
3: find a permutation matrix $\mathbf{P}_j \subseteq \mathbf{A}^{(j-1)}$
4: the minimum element of $\mathbf{A}^{(j-1)}$ at the nonzero positions of $\mathbf{P}_j$ is α_j
5: $\mathbf{A}^{(j)} \leftarrow \mathbf{A}^{(j-1)} - \alpha_j \mathbf{P}_j$

At step 5, we subtract the same value from each row and column sum, hence

$$\frac{1}{1-\alpha}\mathbf{A}^{(j)}$$

is doubly stochastic, and has at least one less nonzero. Continue until we have a single permutation matrix (n entries only).

BvN is not unique

Neither the permutations nor their number

$$\frac{1}{6}\begin{bmatrix} 1 & 4 & 0 & 1 \\ 2 & 1 & 3 & 0 \\ 2 & 1 & 1 & 2 \\ 1 & 0 & 2 & 3 \end{bmatrix}$$

$$= \frac{1}{6}\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} + \frac{1}{6}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \frac{2}{6}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \frac{2}{6}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{6}\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \frac{1}{6}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \frac{1}{6}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \frac{1}{6}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} + \frac{2}{6}\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

BvN decomposition with min. terms

INPUT: A doubly stochastic matrix $\mathbf{A}$.

OUTPUT: A Birkhoff-von Neumann decomposition of $\mathbf{A}$ as $\mathbf{A} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2 + \cdots + \alpha_k \mathbf{P}_k$.

MEASURE: The number k of permutation matrices in the decomposition.

- This problem is NP-hard; not fixed parameter tractable (in k).
- Design and analyze heuristics.

Known results

An upper bound on minimum k

Marcus-Ree Theorem ('59): for a dense matrix there are decompositions where $k \leq n^2 - 2n + 2$

- can be seen using **Carathéodory's** theorem (1911): if a point $\mathbf{x}$ of $\mathbb{R}^d$ lies in the convex hull of a set ***P***, then $\mathbf{x}$ can be written as the convex combination of at most d+1 points in ***P***.

For sparse matrices: $k \leq \mathrm{nnz} - 2n + 2$

Known results

A lower bound on minimum k

A set U of positions of the nonzeros of $\mathbf{A}$ is called strongly stable [Brualdi,'79]: if for each permutation matrix $\mathbf{P} \subseteq \mathbf{A}$, $p_{kl} = 1$ for at most one pair $(k, l) \in U$.

Lemma 1. *Let $\mathbf{A}$ be a doubly stochastic matrix. Then, in a BvN decomposition of $\mathbf{A}$, there are at least $\gamma(\mathbf{A})$ permutation matrices, where $\gamma(\mathbf{A})$ is the maximum cardinality of a strongly stable set of positions of $\mathbf{A}$.*

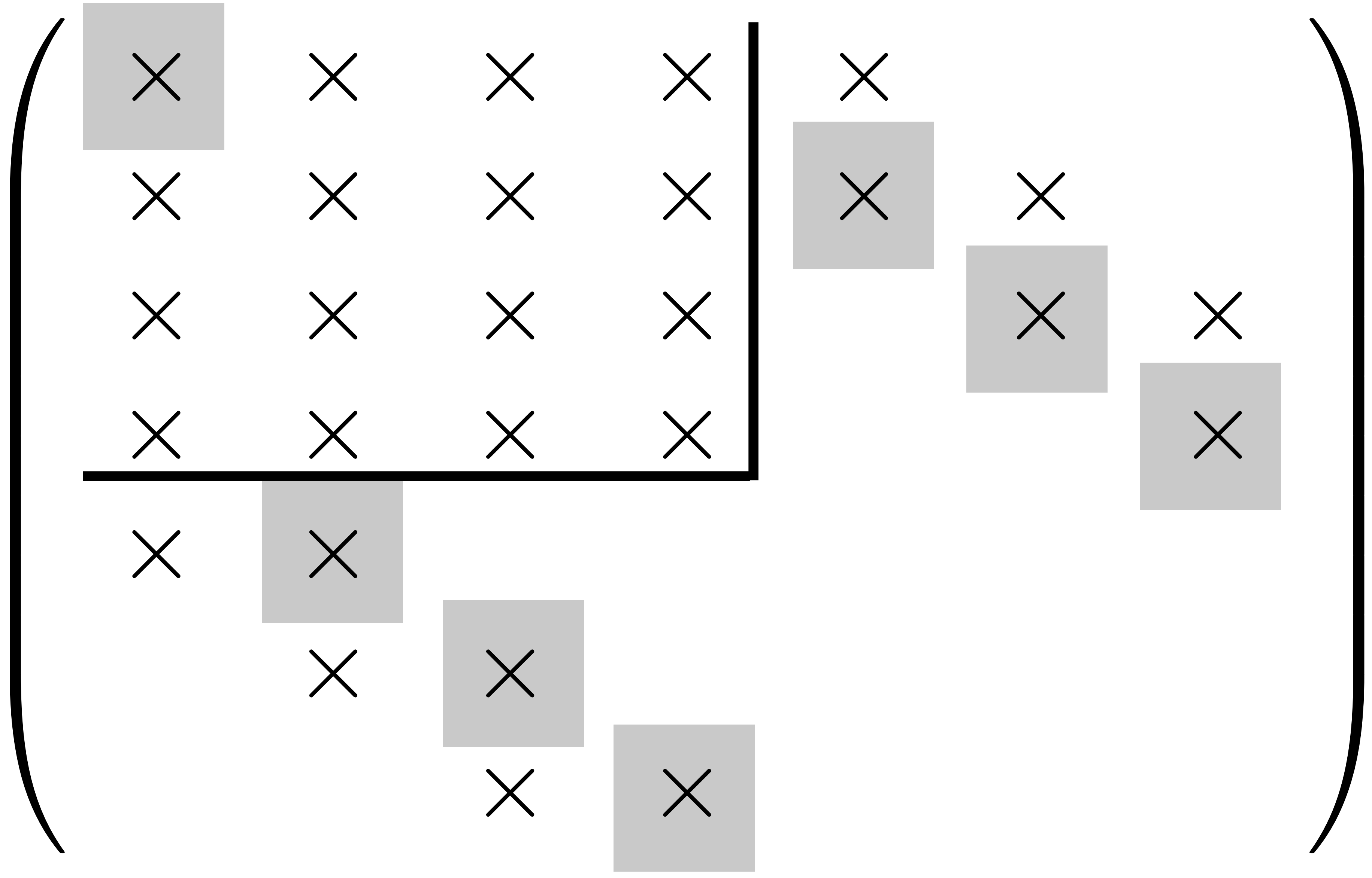
For example: $\gamma(\mathbf{A}) \geq$ the maximum number of nonzeros in a row or a column of $\mathbf{A}$

Known results

A lower bound on minimum k

$\gamma(\mathbf{A}) \geq$ the maximum number of nonzeros in a row or a column of $\mathbf{A}$

[Brualdi,'82] shows that for any integer t with $1 \leq t \leq \lceil n/2 \rceil \lceil (n+1)/2 \rceil$, there exists an $n \times n$ doubly stochastic matrix $\mathbf{A}$ such that $\gamma(\mathbf{A}) = t$.



Heuristics

Heuristics: Generalized Birkhoff heuristic

Finding $\mathbf{A} = \alpha_1\mathbf{P}_1 + \alpha_2\mathbf{P}_2 + \cdots + \alpha_k\mathbf{P}_k$.

1: $\mathbf{A}^{(0)} = \mathbf{A}$
2: **for** $j = 1, \ldots$ **do**
3: find a permutation matrix $\mathbf{P}_j \subseteq \mathbf{A}^{(j-1)}$
4: the minimum element of $\mathbf{A}^{(j-1)}$ at the nonzero positions of $\mathbf{P}_j$ is α_j
5: $\mathbf{A}^{(j)} \leftarrow \mathbf{A}^{(j-1)} - \alpha_j\mathbf{P}_j$

Birkhoff's heuristic: Remove the smallest element

- Let μ be the smallest nonzero of $\mathbf{A}^{(j-1)}$.
- A step 3, find a perfect matching in the graph of $\mathbf{A}^{(j-1)}$ containing μ.

Heuristics: Greedy

1: $\mathbf{A}^{(0)} = \mathbf{A}$
2: **for** $j = 1, \ldots$ **do**
3: find a permutation matrix $\mathbf{P}_j \subseteq \mathbf{A}^{(j-1)}$
4: the minimum element of $\mathbf{A}^{(j-1)}$ at the nonzero positions of $\mathbf{P}_j$ is α_j
5: $\mathbf{A}^{(j)} \leftarrow \mathbf{A}^{(j-1)} - \alpha_j \mathbf{P}_j$

Greedy heuristic: Get the maximum α_j at every step

► At step 3, among all perfect matchings in $\mathbf{A}^{(j-1)}$ find one whose minimum element is the maximum.

Bottleneck perfect matching: efficient implementations exist [Duff & Koster,'01].

Some experiments (Comparing Birkhoff vs Greedy)

τ: the number of nonzeros in a matrix. $d_{\max}$: the maximum number of nonzeros in a row or a column.

matrix	n	τ	$d_{\max}$	Birkhoff $\sum_{i=1}^{k} \alpha_i$	Birkhoff k	*Greedy* $\sum_{i=1}^{k} \alpha_i$	*Greedy* k
aft01	8205	125567	21	0.16	2000	1.00	120
bcspwr10	5300	21842	14	0.38	2000	1.00	63
EX6	6545	295680	48	0.03	2000	1.00	226
flowmeter0	9669	67391	11	0.51	2000	1.00	58
fxm3_6	5026	94026	129	0.13	2000	1.00	383
g3rmt3m3	5357	207695	48	0.05	2000	1.00	223
mplate	5962	142190	36	0.03	2000	1.00	153
n3c6-b7	6435	51480	8	1.00	8	1.00	8
olm5000	5000	19996	6	0.75	283	1.00	14
s2rmq4m1	5489	263351	54	0.00	2000	1.00	208

The heuristics are run to obtain at most 2000 permutation matrices, or until they accumulated a sum of at least 0.9999 with the coefficients.

Some explanation

Birkhoff's performance

Lemma

The Birkhoff heuristic can obtain decompositions in which the number of permutation matrices is very large.

$n \geq 3$, Birkhoff obtains n, but optimal is 3.

$$\mathbf{A}^{(0)} = \begin{pmatrix} \mathbf{1} & 4 & 1 & 0 & 0 & 0 \\ 0 & 1 & \mathbf{4} & 1 & 0 & 0 \\ 0 & 0 & 1 & \mathbf{4} & 1 & 0 \\ 0 & 0 & 0 & 1 & \mathbf{4} & 1 \\ 1 & 0 & 0 & 0 & 1 & \mathbf{4} \\ 4 & \mathbf{1} & 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{A}^{(1)} = \begin{pmatrix} 0 & 4 & \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 3 & 1 & 0 & 0 \\ 0 & 0 & 1 & \mathbf{3} & 1 & 0 \\ 0 & 0 & 0 & 1 & \mathbf{3} & 1 \\ 1 & 0 & 0 & 0 & 1 & \mathbf{3} \\ \mathbf{4} & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{A}^{(2)} = \begin{pmatrix} 0 & \mathbf{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & \mathbf{2} & 1 \\ 1 & 0 & 0 & 0 & 1 & \mathbf{2} \\ \mathbf{3} & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{A}^{(3)} = \begin{pmatrix} 0 & \mathbf{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & \mathbf{1} \\ \mathbf{2} & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{A}^{(4)} = \begin{pmatrix} 0 & \mathbf{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \mathbf{1} \\ 1 & 0 & 0 & 0 & \mathbf{1} & 0 \\ \mathbf{1} & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{A}^{(5)} = \begin{pmatrix} 0 & \mathbf{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 0 \\ \mathbf{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} \end{pmatrix}$$

Closing

Open questions

1. Polytope of solutions

$$\mathbf{A} = \sum \alpha_i \mathbf{P}_i \qquad \mathbf{A} = \sum \beta_i \mathbf{Q}_i$$

Then

$$\mathbf{A} = c \cdot \sum \alpha_i \mathbf{P}_i + (1 - c) \cdot \sum \beta_i \mathbf{Q}_i$$

for $0 \leq c \leq 1$ and the decompositions form a polytope.

Open questions

1. Polytope of solutions

Let A be $n \times n$ doubly stochastic, and $\mathcal{S}(A)$ be the polytope of all BvN decompositions of A.

The extreme points of $\mathcal{S}(A)$ are the ones that cannot be represented as a convex combination of the other decompositions.

[Brualdi,'81]

A heuristic of the generalized Birkhoff family finds an extreme point of the convex polytope $\mathcal{S}(A)$.

Brualdi asks if there are other extreme points of $\mathcal{S}(A)$.

We show that there are.

Open questions

1. Polytope of solutions

a	b	c	d	e	f	g	h	i	j
1	2	4	8	16	32	64	128	256	512

Consider the following matrix whose row sums and column sums are 1023

$$\mathsf{A} = \begin{pmatrix} a+b & d+i & c+h & e+j & f+g \\ e+g & a+c & b+i & d+f & h+j \\ f+j & e+h & d+g & b+c & a+i \\ d+h & b+f & a+j & g+i & c+e \\ c+i & g+j & e+f & a+h & b+d \end{pmatrix}.$$

Generalize with $\mathsf{B} = \begin{pmatrix} 1023 \cdot \mathsf{I} & O \\ O & \mathsf{A} \end{pmatrix}.$

Open questions

1. Polytope of solutions

$$
A = \begin{pmatrix} a+b & d+i & c+h & e+j & f+g \\ e+g & a+c & b+i & d+f & h+j \\ f+j & e+h & d+g & b+c & a+i \\ d+h & b+f & a+j & g+i & c+e \\ c+i & g+j & e+f & a+h & b+d \end{pmatrix} .
$$

Have a decomposition with $a, b, \ldots, j$. No matter in which order, at the first step we do not annihilate an entry. Not Birkhoff.

Open questions

1. Polytope of solutions

- Our proof was computational: With (exponential time) integer linear program solvers, we have shown that there is no other solution with <=10 permutation matrices. The solution is thus extreme & optimal.

- **Looking for a more analytical/constructive proof than we did.**

Open questions

2. Better heuristics

- Better heuristics with/without approximation guarantees.

Greedy algorithms for computing the Birkhoff decomposition

Jérémy E. Cohen
CNRS, CREATIS,
INSA de Lyon
jeremy.cohen@cnrs.fr

and

Bora Uçar
CNRS, LIP,
ENS de Lyon
bora.ucar@ens-lyon.fr

CREATIS Lip cnrs ENS DE LYON Inria INVENTEURS DU MONDE NUMÉRIQUE

- An internship is available http://perso.ens-lyon.fr/bora.ucar/gabi/

Thank you

http://perso.ens-lyon.fr/bora.ucar/
http://perso.ens-lyon.fr/bora.ucar/gabi/

- R. A. Brualdi, **Notes on the Birkhoff algorithm for doubly stochastic matrices**, *Canadian Mathematical Bulletin* 25 (2) (1982) 191–199.
- F. Dufossé and B. Uçar, **Notes on Birkhoff-von Neumann decomposition of doubly stochastic matrices**, *Linear Algebra and its Applications*, vol. 497 (2016), 108–115.
- F. Dufossé, K. Kaya, I. Panagiotas, and B. Uçar, **Further notes on Birkhoff--von Neumann decomposition of doubly stochastic matrices**, *Linear Algebra and its Applications*, 554 (2018), pp. 68–78.