Doubly stochastic matrices & solving sparse linear systems

Bora Uçar

CNRS & ENS Lyon, France

The 11th Scheduling For Large Scale Systems Workshop

From joint work with:

Michele Benzi
Emory Univ., US

Fanny Dufossé
Inria, France

Kamer Kaya
Sabancı Univ., Turkey

Alex Pothen
Purdue Univ., US

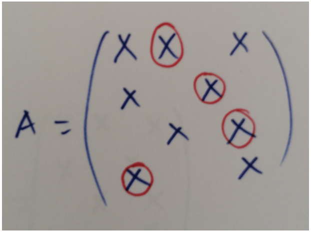
Context

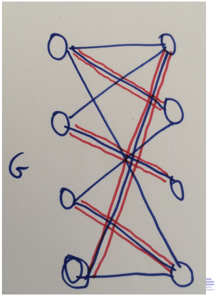
Sparse matrix

- A matrix with many zeros (which are not stored).
- Permutation matrix: An $n \times n$ matrix with exactly one 1 in each row and in each column (other entries are 0)

Bipartite graphs

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





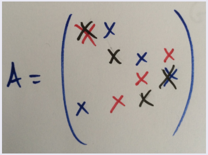
Context

$\mathbf{A} \Longleftrightarrow G_{\mathbf{A}}$: edge between r_i, c_j if $a_{ij} \neq 0$. ($n \times n$ in this talk).

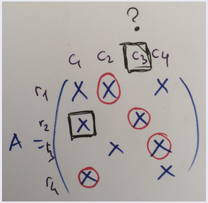
Total support

A has *total support* if every edge in $G_{\mathbf{A}}$ is in a perfect matching.

A has *total support* if every nonzero in **A** is in a permutation matrix.



Our sample matrix does not have total support.



Context

An $n \times n$ matrix $\mathbf{A}$ is doubly stochastic if $a_{ij} \geq 0$, and row sums and column sums are 1.

A doubly stochastic matrix has total support.

Birkhoff's Theorem: $\mathbf{A}$ is a doubly stochastic matrix

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:

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

- Also called Birkhoff-von Neumann (BvN) decomposition.
- Not unique, neither k, nor $\mathbf{P}_i$s in general.

Problem

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.

We show that the problem is NP-hard.

We also propose a heuristic and investigate some its properties theoretically and experimentally.

Motivation

Consider solving $\alpha \mathbf{P} x = b$ for x where $\mathbf{P}$ is a permutation matrix.

$$\alpha \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} \text{ yields } \begin{array}{l} x_4 = b_1/\alpha \\ x_3 = b_2/\alpha \\ x_1 = b_3/\alpha \\ x_2 = b_4/\alpha \end{array}$$

We just scale the input and write at unique (permuted) positions in the output. Should be very efficient.

Next consider solving $(\alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2) x = b$ for x.

Motivation

Consider solving $(\alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2)x = b$ for x.

Matrix splitting and stationary iterations

for an invertible $\mathbf{A} = \mathbf{M} - \mathbf{N}$ with invertible $\mathbf{M}$

$$x^{(i+1)} = \mathbf{H}x^{(i)} + c, \quad \text{where} \quad \mathbf{H} = \mathbf{M}^{-1}\mathbf{N} \quad \text{and} \quad c = \mathbf{M}^{-1}b$$

where $k = 0, 1, \ldots$ and $x^{(0)}$ is arbitrary.

- Computation: At every step, multiply with $\mathbf{N}$ and solve with $\mathbf{M}$.
- Converges to the solution of $\mathbf{A}x = b$ for any $x^{(0)}$ if and only if $\rho(\mathbf{H}) < 1$ [largest magnitude of an eigenvalue is less than 1].

Motivation

Theorem

Let $\mathbf{A} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2$ *and* $\alpha_1 \geq \alpha_2$. *Then,* $\mathbf{A}$ *is invertible if*

(i) $\alpha_1 \neq \alpha_2$,

(ii) $\alpha_1 = \alpha_2$ *and all connected components of* $G_{\mathbf{A}}$ *have an odd number of rows (and columns). If any such block is of even order,* $\mathbf{A}$ *is singular.*

Define the splitting $\mathbf{A} = \alpha_1 \mathbf{P}_1 - (-\alpha_2 \mathbf{P}_2)$.

The iterations are convergent with the rate α_2/α_1 for $\alpha_1 > \alpha_2$.

Next generalize to more than two permutation matrices.

Motivation: Let's generalize to solve $Ax = b$

Let $\mathbf{A} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2 + \cdots + \alpha_k \mathbf{P}_k$ be a BvN.

Assume $\alpha_1 \geq \cdots \geq \alpha_k$. Pick an integer r between 1 and $k-1$ and split $\mathbf{A}$ as $\mathbf{A} = \mathbf{M} - \mathbf{N}$ where

$$\mathbf{M} = \alpha_1 \mathbf{P}_1 + \cdots + \alpha_r \mathbf{P}_r, \quad \mathbf{N} = -\alpha_{r+1} \mathbf{P}_{r+1} - \cdots - \alpha_k \mathbf{P}_k.$$

($\mathbf{M}$ and $-\mathbf{N}$ are doubly substochastic matrices.)

Computation: At every step $\mathbf{M}^{-1}\mathbf{N}x^{(i)}$

- multiply with $\mathbf{N}$ ($k - r$ parallel steps).
- apply $\mathbf{M}^{-1}$ (or solves with the doubly stochastic matrix $\frac{1}{1-\sum_{i=r+1}^{k} \alpha_i}\mathbf{M}$); a recursive solver.

Motivation: Let's generalize

Splitting $\mathbf{A} = \mathbf{M} - \mathbf{N}$ where

$$\mathbf{M} = \alpha_1 \mathbf{P}_1 + \cdots + \alpha_r \mathbf{P}_r, \quad \mathbf{N} = -\alpha_{r+1} \mathbf{P}_{r+1} - \cdots - \alpha_k \mathbf{P}_k.$$

Theorem

A sufficient condition for $\mathbf{M} = \sum_{i=1}^{r} \alpha_i \mathbf{P}_i$ to be invertible: α_1 is greater than the sum of the remaining ones.

Theorem

Suppose that α_1 is greater than the sum of all the other α_i. Then $\rho(\mathbf{M}^{-1}\mathbf{N}) < 1$ and the stationary iterative method converges for all x^0 to the unique solution of $\mathbf{A}x = b$.

This is a sufficient condition; ...and it is rather restrictive in practice.☹

Motivation: Let's generalize

Open question: Identify other, less restrictive conditions on the α_i (with $1 \leq i \leq r$) that will ensure convergence (by the help of permutation matrices).

The natural idea (put the mass in **M**) did not always work: we have examples with $k = 3$, $r = 2$ and $\alpha_1 + \alpha_2 > \alpha_3$.

Another option is to renounce convergence and use **M** as a preconditioner for a Krylov subspace method like GMRES.

Motivation: Let's generalize to any A

M as a preconditioner for a Krylov subspace method like GMRES.

... need generalize to matrices with negative and positive entries.

Scaling fact

Any nonnegative matrix **A** with total support can be scaled with two (unique) positive diagonal matrices **R** and **C** such that **RAC** is doubly stochastic [Sinkhorn & Knopp,'67 and Knight, Ruiz and U.,'14].

Let **A** be $n \times n$ with total support and positive and negative entries.

B = abs(**A**) is nonnegative and **RBC** is doubly stochastic.

We can write $\mathbf{RBC} = \sum \alpha_i \mathbf{P}_i$.

Motivation: Let's generalize to any A

$\mathbf{B} = \text{abs}(\mathbf{A})$ and $\mathbf{RBC} = \sum_i^k \alpha_i \mathbf{P}_i$.

$$\mathbf{RAC} = \sum_i^k \alpha_i \mathbf{Q}_i \,.$$

where $\mathbf{Q}_i = [q_{jk}^{(i)}]_{n \times n}$ is obtained from $\mathbf{P}_i = [p_{jk}^{(i)}]_{n \times n}$ as follows:

$$q_{jk}^{(i)} = \text{sgn}(a_{jk}) p_{jk}^{(i)} \,.$$

Generalizing Birkhoff–von Neumann decomposition

Any (real) matrix $\mathbf{A}$ with total support can be written as a convex combination of a set of signed, scaled permutation matrices.

We can then use the same construct to define $\mathbf{M}$ (for splitting or for defining the preconditioner).

Linear algebraic problem and combinatorial problem

Reduce the complexity of the solver by reducing the cost of applying $\mathbf{M}$.

Find a BvN decomposition with the smallest number of perm. matrices.

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 is NP-hard.

Known results for the num. of permutation matrices: Upper bound

Marcus–Ree Theorem ['59]: $k \leq n^2 - 2n + 2$ for dense matrices;

[Brualdi& Gibson,'77] and [Brualdi,'82]: for a sparse matrix with τ nonzeros
$k \leq \tau - 2n + \ell + 1$

(containing ℓ submatrices with total support; take $\ell = 1$)

Known results for the num. of permutation matrices: lower bound

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 (Brualdi,'82)

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 $\mathbf{A}$.

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

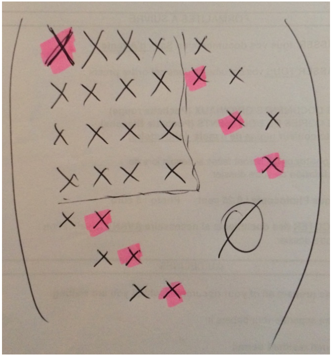
Known results for the num. of permutation matrices: lower bound

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

[Brualdi,'82] shows that for any integer t

$$1 \leq t \leq \lceil n/2 \rceil \lceil (n+1)/2 \rceil$$

there exists an $n \times n$ doubly stochastic matrix $\mathbf{A}$ such that $\gamma(\mathbf{A}) = t$.



Heuristics: Generalized Birkhoff heuristic

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 M in the graph of $\mathbf{A}^{(j-1)}$ containing μ.

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

Experiments (heuristics)

τ: the number of nonzeros in a matrix.

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

				Birkhoff		greedy	
matrix	n	τ	$d_{\max}$	$\sum_{i=1}^{k} \alpha_i$	k	$\sum_{i=1}^{k} \alpha_i$	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
s2rmq4m1	5489	263351	54	0.00	2000	1.00	208

[at most 2000 permutation matrices, or accumulated a sum of at least 0.9999]

Experiments (linear systems)

Number of GMRES iterations

matrix	ilu(0)	**M** with different r			
		1	2	16	32
bp_200	2	38	31	22	23
gemat11	168	1916	1606	620	297
gemat12	254	2574	1275	570	386
lns3937	348	1702	801	48	36
mahindas	37	225	158	43	32
orani678	23	172	140	71	58

num. perm (k)

matrix	k
bp_200	5
gemat11	48
gemat12	34
lns3937	57
mahindas	154
orani678	1053

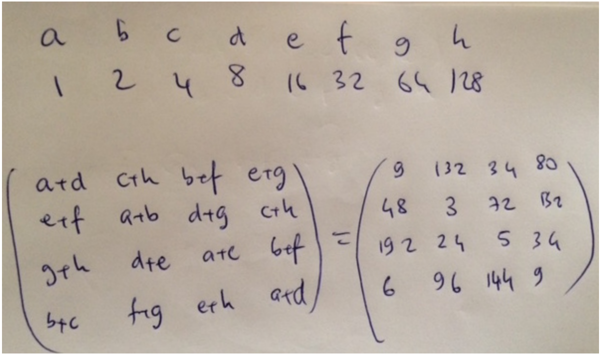
Number of nonzeros in **M**

matrix	ilu(0)	1	2	16	32
bp_200	125	0.32	0.52	0.75	0.75
gemat11	31425	0.15	0.21	0.69	0.88
gemat12	31184	0.15	0.20	0.64	0.79
lns3937	24002	0.15	0.25	0.59	0.65
mahindas	4744	0.12	0.16	0.29	0.36
orani678	47823	0.04	0.04	0.05	0.06

Concluding remarks (recently shown)

Closed (or not so) problem: are there optimal points of the polytope $f^{-1}(\mathbf{S})$ other than those obtained by a generalized Birkhoff heuristic?

YES:



Concluding remarks (still open)

Open problem 1: can we find less restrictive conditions for having a convergent solution?

Open problem 2: a better heuristic than the proposed greedy one? Approximation guarantee?

Open problem 3: special, interesting cases that we can solve?

Open problem 4: among the optimal ones, is there one obtained by a generalized Birkhoff.

Thanks!

Thanks for your attention.

More information

- R. A. Brualdi, Notes on the Birkhoff algorithm for doubly stochastic matrices, Canadian Mathematical Bulletin, 25 (1982), pp. 191–199.
- M. Benzi, A. Pothen, and B. Uçar, Preconditioning techniques based on the Birkhoff–von Neumann decomposition, Technical report RR-8914, Inria - Research Centre Grenoble – Rhône-Alpes, 2016.
- F. Dufossé and B. Uçar, Notes on Birkhoff–von Neumann decomposition of doubly stochastic matrices, Linear Algebra and its Applications 497 (2016) 108–115.

References I

- R. A. Brualdi, The diagonal hypergraph of a matrix (bipartite graph), Discrete Mathematics 27 (2) (1979) 127–147.
- R. A. Brualdi, *Notes on the Birkhoff algorithm for doubly stochastic matrices*, Canadian Mathematical Bulletin, 25 (1982), pp. 191–199.
- R. A. Brualdi and P. M. Gibson, *Convex polyhedra of doubly stochastic matrices: I. Applications of the permanent function*, Journal of Combinatorial Theory, Series A, 22 (1977), pp. 194–230.
- C. S. Chang , W. J. Chen, and H. Y. Huang, On service guarantees for input buffered crossbar switches: A capacity decomposition approach by Birkhoff and von Neumann, IEEE IWQos'99, 79–86, London, UK.
- I. S. Duff and J. Koster, On algorithms for permuting large entries to the diagonal of a sparse matrix, SIAM Journal on Matrix Analysis and Applications 22 (2001) 973–996.

References II

- F. Dufossé and B. Uçar, Notes on Birkhoff–von Neumann decomposition of doubly stochastic matrices, Linear Algebra and its Applications 497 (2016) 108–115.
- P. A. Knight, The Sinkhorn–Knopp algorithm: Convergence and applications, SIAM J. Matrix Anal. A. 30 (1) (2008) 261–275.
- P. A. Knight and D. Ruiz, A fast algorithm for matrix balancing, IMA Journal of Numerical Analysis 33 (3) (2013) 1029–1047.
- P. A. Knight, D. Ruiz, and B. Uçar, A symmetry preserving algorithm for matrix scaling, SIAM J. Matrix Anal. A. 35 (3) (2014) 931–955.
- M. Marcus and R. Ree, Diagonals of doubly stochastic matrices, The Quarterly Journal of Mathematics 10 (1) (1959) 296–302.
- A. Pothen and C.-J. Fan, Computing the block triangular form of a sparse matrix, ACM T. Math. Software 16 (4) (1990) 303–324.

References III

- D. Ruiz, A scaling algorithm to equilibrate both row and column norms in matrices, Tech. Rep. TR-2001-034, RAL (2001).
- R. Sinkhorn and P. Knopp, Concerning nonnegative matrices and doubly stochastic matrices, Pacific J. Math. 21 (1967) 343–348.
- D. de Werra, Variations on the Theorem of Birkhoff–von Neumann and extensions, Graphs and Combinatorics, 19 (2003), 263–278.
- D. de Werra, Partitioning the edge set of a bipartite graph into chain packings: Complexity of some variations, Linear Algebra and its Applications, 368 (2003) 315–327.