

Further notes on Birkhoff-von Neumann decomposition of doubly stochastic matrices

Fanny Dufossé, Kamer Kaya, Ioannis Panagiotas, Bora Uçar

▶ **To cite this version:**

Fanny Dufossé, Kamer Kaya, Ioannis Panagiotas, Bora Uçar. Further notes on Birkhoff-von Neumann decomposition of doubly stochastic matrices. Linear Algebra and its Applications, 2018, 554, pp.68–78. 10.1016/j.laa.2018.05.017 . hal-01586245v2

HAL Id: hal-01586245
https://inria.hal.science/hal-01586245v2

Submitted on 17 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Further notes on Birkhoff–von Neumann decomposition of doubly stochastic matrices

Fanny Dufossé[a], Kamer Kaya[b], Ioannis Panagiotas[c], Bora Uçar[d,*]

[a] *Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, 38000 Grenoble, France*
[b] *Faculty of Engineering and Natural Sciences, Sabanci University, İstanbul, Turkey*
[c] *ENS Lyon, France*
[d] *Univ Lyon, CNRS, ENS de Lyon, Inria, Université Claude-Bernard Lyon 1, LIP UMR 5668, F-69007 LYON, FRANCE*

Abstract

The well-known Birkhoff-von Neumann (BvN) decomposition expresses a doubly stochastic matrix as a convex combination of a number of permutation matrices. For a given doubly stochastic matrix, there are many BvN decompositions, and finding the one with the minimum number of permutation matrices is NP-hard. There are heuristics to obtain BvN decompositions for a given doubly stochastic matrix. A family of heuristics is based on the original proof of Birkhoff and proceeds step by step by subtracting a scalar multiple of a permutation matrix at each step from the current matrix, starting from the given matrix. At every step, the subtracted matrix contains nonzeros at the positions of some nonzero entries of the current matrix and annihilates at least one entry, while keeping the current matrix nonnegative. Our first result, which supports a claim of Brualdi [Canad. Math. Bull. 25 (1982), pp. 191–199], shows that this family of heuristics can miss optimal decompositions. We also investigate the performance of two heuristics from this family theoretically.

Keywords: doubly stochastic matrix, Birkhoff-von Neumann decomposition
2010 MSC: 15B51, 05C50, 05C70

1. Introduction

A square matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n \times n}$ is doubly stochastic if $a_{ij} \geq 0$ for all i, j and $\mathbf{A}e = \mathbf{A}^T e = e$, where e is the column vector of all ones. In other words, the entries of a doubly stochastic matrix are nonnegative, and the sum of entries in any row or column is equal to one. By Birkhoff's Theorem, there exist

*Corresponding author. Tel: +33 472728932, 46 allée d'Italie, 69007, Lyon, France
Email addresses: fanny.dufosse@inria.fr (Fanny Dufossé), kaya@sabanciuniv.edu (Kamer Kaya), ioannis.panagiotas@ens-lyon.fr (Ioannis Panagiotas), bora.ucar@ens-lyon.fr (Bora Uçar)

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

$$\mathbf{A} = \alpha_1 \mathbf{P}_1 + \alpha_2 \mathbf{P}_2 + \cdots + \alpha_k \mathbf{P}_k \; . \tag{1}$$

This representation is also called Birkhoff-von Neumann (BvN) decomposition. For a given matrix there are many different BvN decompositions. *Marcus–Ree Theorem* [1] states that for a dense matrix, there are decompositions with $k \leq n^2 - 2n + 2$ permutation matrices. For a sparse, fully indecomposable matrix with τ nonzeros, the same result holds [2, 3] with $k \leq \tau - 2n + 2$. Brualdi [3] gives lower bounds on the number of permutation matrices in any BvN decomposition of a given matrix. A recent work [4] shows that the problem of finding a BvN decomposition with the smallest number k of permutation matrices is strongly NP-complete.

There are heuristics to compute a BvN decomposition for a given matrix $\mathbf{A}$. In particular, the following family of heuristics is based on the constructive proof of Birkhoff. Let $\mathbf{A}^{(0)} = \mathbf{A}$. At every step $j \geq 1$, find a permutation matrix $\mathbf{P}_j$ having its ones at the positions of the nonzero elements of $\mathbf{A}^{(j-1)}$, use the minimum nonzero element of $\mathbf{A}^{(j-1)}$ at the positions identified by $\mathbf{P}_j$ as α_j, set $\mathbf{A}^{(j)} = \mathbf{A}^{(j-1)} - \alpha_j \mathbf{P}_j$, and repeat the computations in the next step $j + 1$ until $\mathbf{A}^{(j)}$ becomes void. Any heuristic of this type is called *generalized Birkhoff heuristic*. The original Birkhoff heuristic chooses a permutation matrix $\mathbf{P}_j$ at step j which contains a one at the position of the minimum nonzero element of $\mathbf{A}^{(j-1)}$. Dufossé and Uçar [4] propose choosing a permutation matrix $\mathbf{P}_j$ where the minimum nonzero element of $\mathbf{A}^{(j-1)}$ identified by $\mathbf{P}_j$ is maximum. We refer to this last heuristic as the *Greedy* heuristic.

Doubly stochastic matrices and their associated BvN decompositions have been used in several operations research problems and applications. Classical examples are concerned with allocating communication resources, where an input traffic is routed to an output traffic in stages [5]. Each routing stage is a (sub-)permutation matrix and is used for handling a disjoint set of communications. The number of stages correspond to the number of (sub-)permutation matrices. A recent variation of this problem appears in routing in data centers [6]. BvN decompositions are also used to build preconditioners for solving sparse linear systems [7]. Here, the number k of the permutation matrices is related to the cost of applying the preconditioner.

Our contribution in this note is threefold. First, we answer an open question asked by Brualdi [3, pp. 197–198]. Brualdi predicts that there are BvN decompositions which cannot be written as a convex combination of other decompositions and which cannot be obtained by a generalized Birkhoff heuristic. We give an example affirming the prediction of Brualdi. Second, we investigate the worst case performance guarantee of the original Birkhoff heuristic and show that it can be very bad. To do so, we describe a family of matrices where the smallest number of permutation matrices required is 3 while the Birkhoff heuristic finds decompositions with n permutation matrices, for a matrix of size $n \times n$. Kulkarni et al. [6, Theorem 7] also present examples to show that the worst case

performance of Birkhoff heuristic can be very bad. Our construction is more explicit and seems simpler. Our third contribution is to show an approximation guarantee of the *Greedy* heuristic.

2. A result on the polytope of BvN decompositions

Let $\mathcal{S}(\mathbf{A})$ be the polytope of all BvN decompositions for a given, doubly stochastic matrix $\mathbf{A}$. The extreme points of $\mathcal{S}(\mathbf{A})$ are the ones that cannot be represented as a convex combination of the other decompositions. Brualdi [3, pp. 197–198] observes that any generalized Birkhoff heuristic obtains an extreme point of $\mathcal{S}(\mathbf{A})$, and predicts that there are extreme points of $\mathcal{S}(\mathbf{A})$ which cannot be obtained by a generalized Birkhoff heuristic. In this section, we substantiate this claim by showing an example.

Any BvN decomposition of a given matrix $\mathbf{A}$ with the smallest number of permutation matrices is an extreme point of $\mathcal{S}(\mathbf{A})$; otherwise the other BvN decompositions expressing the said point would have smaller number of permutation matrices.

Lemma 1. *There are doubly stochastic matrices whose polytopes of BvN decompositions contain extreme points that cannot be obtained by a generalized Birkhoff heuristic.*

We will prove the lemma by giving an example. We use computational tools based on a mixed integer linear programming (MILP) formulation of the problem of finding a BvN decomposition with the smallest number of permutation matrices. We first describe the MILP formulation.

Let $\mathbf{A}$ be a given $n \times n$ doubly stochastic matrix, and Ω_n be the set of all $n \times n$ permutation matrices. There are $n!$ matrices in Ω_n; for brevity, let us refer to these permutations by $\mathbf{P}_1, \ldots, \mathbf{P}_{n!}$. We associate an incidence matrix $\mathbf{M}$ of size $n^2 \times n!$ with Ω_n. We fix an ordering of the entries of $\mathbf{A}$ so that each row of $\mathbf{M}$ corresponds to a unique entry in $\mathbf{A}$. Each column of $\mathbf{M}$ corresponds to a unique permutation matrix in Ω_n. We set $m_{ij} = 1$ if the ith entry of $\mathbf{A}$ appears in the permutation matrix $\mathbf{P}_j$, and set $m_{ij} = 0$ otherwise. Let $\vec{a}$ be the n^2-vector containing the values of the entries of $\mathbf{A}$ in the fixed order. Let $\vec{x}$ be a vector of $n!$ elements, x_j corresponding to the permutation matrix $\mathbf{P}_j$. With these definitions, MILP formulation for finding a BvN decomposition of $\mathbf{A}$ with

the smallest number of permutation matrices can be written as:

$$\text{minimize} \sum_{j=1}^{n!} s_j \tag{2}$$

$$\text{subject to } \mathbf{M}\vec{x} = \vec{a}\,, \tag{3}$$

$$1 \geq x_j \geq 0, \quad \text{for } j = 1, \ldots, n!\,, \tag{4}$$

$$\sum x_j = 1, \quad \text{for } j = 1, \ldots, n!\,, \tag{5}$$

$$s_j \geq x_j, \quad \text{for } j = 1, \ldots, n!\,, \tag{6}$$

$$s_j \in \{0, 1\}, \quad \text{for } j = 1, \ldots, n!\,. \tag{7}$$

In this MILP, s_j is a binary variable which is 1 only if $x_j > 0$, otherwise 0. The equality (3), the inequalities (4), and the equality (5) guarantee that we have a BvN decomposition of $\mathbf{A}$. This MILP can be used only for small problems. In this MILP, we can exclude any permutation matrix $\mathbf{P}_j$ from a decomposition by setting $s_j = 0$.

Proof of Lemma 1. Let the following 10 letters correspond to the numbers underneath

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, hence can be considered as doubly stochastic

$$\mathbf{A} = \begin{pmatrix} a+b & d+i & c+h & e+j & f+g \\ e+g & a+c & b+i & d+f & h+j \\ f+j & e+h & d+g & b+c & a+i \\ d+h & b+f & a+j & g+i & c+e \\ c+i & g+j & e+f & a+h & b+d \end{pmatrix}. \tag{8}$$

Observe that the entries containing the term 2^ℓ form a permutation matrix, for $\ell = 0, \ldots, 9$. Therefore, this matrix has a BvN decomposition with 10 permutation matrices. We created the MILP above and found 10 as the smallest number of permutation matrices by calling the CPLEX solver [8] via the NEOS Server [9, 10, 11]. Hence the described decomposition is an extreme point of $\mathbf{S}(\mathbf{A})$. None of the said permutation matrices annihilate any entry of the matrix. Therefore, at the first step no entry of $\mathbf{A}$ gets reduced to zero, regardless of the order of permutation matrices. Thus, this decomposition cannot be obtained by a generalized Birkhoff heuristic.

One can create a family of matrices with arbitrary sizes by embedding the same matrix into a larger one of the form

$$\mathbf{B} = \begin{pmatrix} 1023 \cdot \mathbf{I} & O \\ O & \mathbf{A} \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} 3 & 264 & 132 & 528 & 96 \\ 80 & 5 & 258 & 40 & 640 \\ 544 & 144 & 72 & 6 & 257 \\ 136 & 34 & 513 & 320 & 20 \\ 260 & 576 & 48 & 129 & 10 \end{pmatrix}$$

(a) The sample matrix

129	511	257	63	33	15	7	3	2	2	1
3	4	2	5	**5**	4	**2**	**3**	**4**	1	**1**
5	**5**	3	1	4	1	**4**	2	**1**	**2**	**3**
2	1	**5**	3	**1**	**2**	3	4	**3**	4	**4**
1	3	4	**4**	2	5	**1**	5	**5**	**3**	**2**
4	2	1	2	3	**3**	5	**1**	**2**	5	**5**

(b) A BvN decomposition

Figure 1: The matrix $\mathbf{A}$ from Lemma 1, and a BvN decomposition with 11 permutation matrices which can be obtained by a generalized Birkhoff heuristic. Each column in (b) corresponds to a permutation, where the first line gives the associated coefficient, and the following lines give the column indices matched to the rows 1 to 5 of $\mathbf{A}$.

where $\mathbf{I}$ is the identity matrix with the desired size. All BvN decompositions of $\mathbf{B}$ can be obtained by extending the permutation matrices in $\mathbf{A}$'s BvN decompositions with $\mathbf{I}$. That is, for a permutation matrix $\mathbf{P}$ in a BvN decomposition of $\mathbf{A}$, $\begin{pmatrix} \mathbf{I} & O \\ O & \mathbf{P} \end{pmatrix}$ is a permutation matrix in the corresponding BvN decomposition of $\mathbf{B}$. Furthermore, all permutation matrices in an arbitrary BvN decomposition of $\mathbf{B}$ must have $\mathbf{I}$ as the principle sub-matrix, and the rest should correspond to a permutation matrix in $\mathbf{A}$, defining a BvN decomposition for $\mathbf{A}$. Hence, the extreme point $\mathcal{S}(\mathbf{B})$ corresponding to the extreme point of $\mathcal{S}(\mathbf{A})$ with 10 permutation matrices cannot be found by a generalized Birkhoff heuristic. □

Let us investigate the matrix $\mathbf{A}$ and its BvN decomposition given in the proof of Lemma 1. Let $\mathbf{P}_a, \mathbf{P}_b, \ldots, \mathbf{P}_j$ be the 10 permutation matrices of the decomposition, corresponding to $a, b, \ldots, j$. We solve 10 MILPs in which we set $s_t = 0$ for one $t \in \{a, b, \ldots, j\}$. This way, we try to find a BvN decomposition of $\mathbf{A}$ without $\mathbf{P}_a$, without $\mathbf{P}_b$ and so on, always with the smallest number of permutation matrices. The smallest number of permutation matrices in these 10 MILPs were 11. This certifies that the only BvN decomposition with 10 permutation matrices necessarily contains $\mathbf{P}_a, \mathbf{P}_b, \ldots, \mathbf{P}_j$. It is easy to see that there is a unique solution to the equality $\mathbf{M}\vec{x} = \vec{a}$ of the MILP with $x_t = 0$ for $t \notin \{a, b, \ldots, j\}$, as the submatrix $\mathbf{M}$ containing only the corresponding 10 columns has full column rank.

Any generalized Birkhoff heuristic obtains at least 11 permutation matrices for the matrix $\mathbf{A}$ of the proof of Lemma 1. One such decomposition is shown in a tabular format in Fig. 1. In Fig. 1a, we write the matrix of the lemma explicitly for convenience. Then in Fig. 1b, we give a BvN decomposition. The column headers (the first line) in the table contains the coefficients of the permutation matrices. The nonzero column indices of the permutation matrices are stored by rows. For example, the first permutation has the coefficient 129 and columns 3, 5, 2, 1, and 4 are matched to the rows 1–5. The bold indices signify entries whose values are equivalent to the coefficients of the corresponding permutation matrices, at the time where the permutation matrices are found. Therefore, the

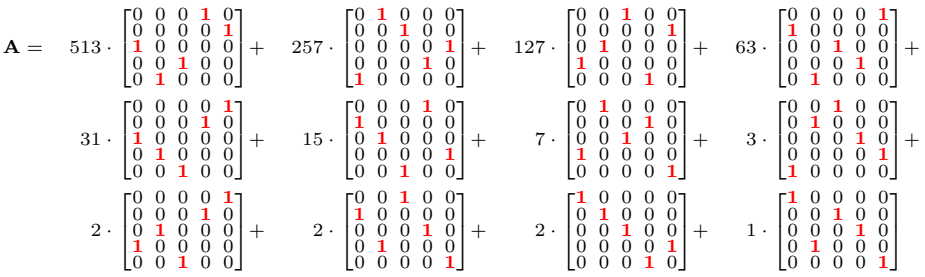


Figure 2: The output of the *Greedy* heuristic for the matrix given in the proof of Lemma 1.

corresponding entries become zero after the corresponding step. For example, in the first permutation, $a_{5,4} = 129$.

The output of the *Greedy* heuristic for the matrix $\mathbf{A}$ is given in Fig. 2 for reference. It contains 12 permutation matrices.

3. Analysis of two known heuristics for computing BvN decompositions

Empirical results in the literature demonstrated that the *Greedy* heuristic [4] can obtain a much smaller number of permutation matrices compared to the Birkhoff heuristic. Here, we compare these two heuristics theoretically. First we show that the original Birkhoff heuristic does not have any constant ratio approximation guarantee. Furthermore, for an $n \times n$ matrix, its worst-case approximation ratio is $\Omega(n)$.

We begin with a small example shown in Fig. 3. We decompose a 6×6 matrix $\mathbf{A}^{(0)}$ which has an optimal BvN decomposition with three permutation matrices; the main diagonal, the one containing the entries equal to 4, and the one containing the remaining entries. For simplicity, we used integer values in our example. However, since the row and column sums of $\mathbf{A}^{(0)}$ is equal to 6, it can be converted to a doubly stochastic matrix by dividing all the entries to 6. Instead of the optimal decomposition, in the figure, we obtain the permutation matrices as the original Birkhoff heuristic does. Each red-colored entry set is a permutation and contains the minimum possible value, 1.

In the following, we show how to generalize the idea for having matrices of arbitrarily large size, with three permutation matrices in an optimal decomposition, while the original Birkhoff heuristic obtains n permutation matrices.

Lemma 2. *The worst-case approximation ratio of the Birkhoff heuristic is* $\Omega(n)$.

Proof. For any given integer $n \geq 3$, we show that there is a matrix of size $n \times n$ whose optimal BvN decomposition has 3 permutations, whereas the Birkhoff heuristic obtains a BvN decomposition with exactly n permutation matrices.

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

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

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

Figure 3: A sample matrix to show that the original Birkhoff heuristic can obtain BvN decomposition with n permutation matrices while the optimum one has 3.

The example in Fig. 3 is a special case for $n = 6$ for the following construction process.

Let $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ be the function $f(x) = (x \bmod n) + 1$. Given a matrix $\mathbf{M}$, let $\mathbf{M}' = \mathcal{F}(\mathbf{M})$ be another matrix containing the same set of entries where the function is $f(\cdot)$ is used on the coordinate indices to redistribute the entries of $\mathbf{M}$ on $\mathbf{M}'$. That is $m_{i,j} = m'_{f(i),f(j)}$. Since $f(\cdot)$ is one-to-one and onto, if $\mathbf{M}$ is a permutation matrix then $\mathcal{F}(\mathbf{M})$ is also a permutation matrix. We will start with a permutation matrix, and run it through $\mathcal{F}$ for $n - 1$ times to obtain n permutation matrices, which are all different. By adding these permutation matrices, we will obtain a matrix $\mathbf{A}$ whose optimal BvN decomposition has three permutation matrices, while the n permutation matrices used to create $\mathbf{A}$ correspond to a decomposition that can be obtained by the Birkhoff heuristic.

Let $\mathbf{P}_1$ be the permutation matrix whose ones, which are partitioned into three sets, are at the positions

$$\overbrace{(1,1)}^{\text{1st set}}, \overbrace{(n,2)}^{\text{2nd set}}, \overbrace{(2,3),(3,4),\ldots,(n-1,n)}^{\text{3rd set}} . \tag{9}$$

Let us use $\mathcal{F}(\cdot)$ to generate a matrix sequence $\mathbf{P}_i = \mathcal{F}(\mathbf{P}_{i-1})$ for $2 \leq i \leq n$. For

example, $\mathbf{P}_2$'s nonzeros are at the positions

$$\overbrace{(2,2)}^{\text{1st set}}, \overbrace{(1,3)}^{\text{2nd set}}, \overbrace{(3,4),(4,5),\ldots,(n,1)}^{\text{3rd set}}.$$

We then add the $\mathbf{P}_i$s to build the matrix

$$\mathbf{A} = \mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_n \ .$$

We have the following observations about the nonzero elements of $\mathbf{A}$:

1. $a_{i,i} = 1$ for all $i = 1, \ldots, n$, and only $\mathbf{P}_i$ has a one at the position (i, i). These elements are from the first set of positions of the permutation matrices, as identified in (9). When put together, these n entries form a permutation matrix $\mathbf{P}^{(1)}$.
2. $a_{i,j} = 1$ for all $i = 1, \ldots, n$ and $j = ((i+1) \bmod n)+1$, and only $\mathbf{P}_h$, where $h = (i \bmod n) + 1$, has a one at the position (i, j). These elements are from the second set of positions of the permutation matrices, as identified in (9). When put together, these n entries form a permutation matrix $\mathbf{P}^{(2)}$.
3. $a_{i,j} = n - 2$ for all $i = 1, \ldots, n$ and $j = (i \bmod n) + 1$, where all $\mathbf{P}_\ell$ for $\ell \in \{1, \ldots, n\} \setminus \{i, j\}$ have a one at the position $a_{i,j}$. These elements are from the third set of positions of the permutation matrices, as identified in (9). When put together, these n entries form a permutation matrix $\mathbf{P}^{(3)}$ multiplied by the scalar $(n - 2)$.

In other words, we can write

$$\mathbf{A} = \mathbf{P}^{(1)} + \mathbf{P}^{(2)} + (n-2) \cdot \mathbf{P}^{(3)} \ ,$$

and see that $\mathbf{A}$ has a BvN decomposition with three permutation matrices. We note that each row and column of $\mathbf{A}$ contains three nonzeros; and hence three is the smallest number of permutation matrices in a BvN decomposition of $\mathbf{A}$.

Since the minimum element in $\mathbf{A}$ is 1, and each $\mathbf{P}_i$ contains one such element, the Birkhoff heuristic can obtain a decomposition using $\mathbf{P}_i$ for $i = 1, \ldots, n$. Therefore, the Birkhoff heuristic's approximation is no better than $\frac{n}{3}$, which can be made arbitrarily large. □

We note that *Greedy* will optimally decompose the matrix $\mathbf{A}$ used in the proof above.

We now analyze the performance of the *Greedy* heuristic. We start with a few experiments which indicate that there is a connection between the performance of *Greedy* and the values in the matrix.

We create a set of $n \times n$ matrices. To do that, we first fix a set of z permutation matrices $\{\mathbf{C}_1, \ldots, \mathbf{C}_z\}$ of size $n \times n$. These permutation matrices with varying values of α will be used to generate the matrices. The matrices are parametrized by the subscript i and each $\mathbf{A}_i$ is created as follows: $\mathbf{A}_i = \alpha_1 \cdot \mathbf{C}_1 + \alpha_2 \cdot \mathbf{C}_2 + \cdots + \alpha_z \cdot \mathbf{C}_z$ where each α_j for $j = 1, \ldots, z$ is a

Table 1: Experiments showing the dependence of the performance of *Greedy* on the values of the matrix elements. n is the matrix size; $i \in \{10, 20, 30, 40, 50\}$ is the parameter for creating matrices using $\alpha_j \in [1, 2^i]$; z is the number of permutation matrices used in creating $\mathbf{A}_i$. Five experiments for a given pair of n and i. *Greedy* obtains k_i permutation matrices. The average and the maximum number of permutation matrices obtained by *Greedy* for five random instances are given. k_i/z is a lower bound to the performance of *Greedy*, as $z \geq Opt$.

$n = 30$ and $z = 20$

	average		worst case	
i	k_i	k_i/z	k_i	k_i/z
10	59	2.99	63	3.15
20	105	5.29	110	5.50
30	149	7.46	158	7.90
40	184	9.23	191	9.55
50	212	10.62	227	11.35

$n = 200$ and $z = 100$

	average		worst case	
i	k_i	k_i/z	k_i	k_i/z
10	268	2.69	280	2.80
20	487	4.88	499	4.99
30	716	7.16	726	7.26
40	932	9.33	947	9.47
50	1124	11.25	1162	11.62

randomly chosen integer in the range $[1, 2^i]$, and we also set a randomly chosen α_j equivalent to 2^i to guarantee the existence of at least one large value even in the unlikely case that all other values are not large enough. As can be seen, each $\mathbf{A}_i$ has the same structure and differs from the rest only in the values of α_j's that are chosen. As a consequence, they all can be decomposed by the same set of permutation matrices.

We present our results in two sets of experiments shown in Table 1, for two different n. In both cases we create five random $\mathbf{A}_i$ for $i \in \{10, 20, 30, 40, 50\}$, that is we have five matrices with the parameter i, and there are five different i. We have $n = 30$ and $z = 20$ in the first set, and $n = 200$ and $z = 100$ in the second set. Let k_i be the number of permutation matrices *Greedy* obtains for a given $\mathbf{A}_i$. The table gives the average and the maximum k_i of five different $\mathbf{A}_i$, for given n and i. By construction, each $\mathbf{A}_i$ has a BvN with z permutation matrices. Since z is no smaller than the optimal value, the ratio $\frac{k_i}{z}$ gives a lower bound on the performance of *Greedy*. As seen from the experiments, as i increases, the performance of *Greedy* gets increasingly worse. This shows that a constant ratio worst case approximation of *Greedy* is unlikely. While the performance depend on z (for example, for small z, *Greedy* is likely to obtain near optimal decompositions), it seems that the size of the matrix does not largely affect the relative performance of *Greedy*.

Now we attempt to explain the above results theoretically.

Lemma 3. *Let $\alpha_1^\star \mathbf{P}_1^\star + \cdots + \alpha_k^\star \mathbf{P}_{k^\star}^\star$ be a BvN decomposition of a given doubly stochastic matrix $\mathbf{A}$ with the smallest number $k^\star$ of permutation matrices. Then, for any BvN decomposition of $\mathbf{A}$ with $\ell \geq k^\star$ permutation matrices, we have $\ell \leq k^\star \cdot \frac{\max_i \alpha_i^\star}{\min_i \alpha_i}$. If the coefficients are integers (e.g., when $\mathbf{A}$ is a matrix with constant row and column sums of integral values), we have $\ell \leq k^\star \cdot \max_i \alpha_i^\star$.*

Proof. Consider a BvN decomposition $\alpha_1 \mathbf{P}_1 + \cdots + \alpha_\ell \mathbf{P}_\ell$ with $\ell \geq k^\star$. Assume without loss of generality that $\alpha_1^\star \geq \cdots \geq \alpha_{k^\star}^\star$ and $\alpha_1 \geq \cdots \geq \alpha_\ell$.

We know that the coefficients of these two decompositions sum up to the same value. That is

$$\sum_{i=1}^{\ell} \alpha_i = \sum_{i=1}^{k^\star} \alpha_i^\star \ .$$

Since α_ℓ is the smallest of α, and $\alpha_1^\star$ is the largest of $\alpha^\star$, we have

$$\ell \cdot \alpha_\ell \leq k^\star \cdot \alpha_1^\star \ ,$$

and hence

$$\frac{\ell}{k^\star} \leq \frac{\alpha_1^\star}{\alpha_\ell} \ .$$

By assuming integer values, we see that $\alpha_\ell \geq 1$ and thus

$$\ell \leq k^\star \cdot \max_i \alpha_i^\star \ .$$

□

This lemma evaluates the approximation guarantee of a given BvN decomposition. It does not seem very useful, because of the fact that even if we have $\min_i \alpha_i$, we do do not have $\max_i \alpha_i^\star$. Luckily, we can say more in the case of *Greedy*.

Corollary 1. *Let $k^\star$ be the smallest number of permutation matrices in a BvN decomposition of a given doubly stochastic matrix $\mathbf{A}$. Let α_1 and α_ℓ be the first and last coefficients obtained by the* Greedy *heuristic for decomposing $\mathbf{A}$. Then, $\ell \leq k^\star \cdot \frac{\alpha_1}{\alpha_\ell}$.*

Proof. This is easy to see as *Greedy* obtains the coefficients in a non-increasing order [4, Lemma 3], and $\alpha_1 \geq \alpha_j^\star$ for all $1 \leq j \leq k^\star$ for any BvN decomposition containing $\alpha_j^\star$. □

Lemma 3 and Corollary 1 give a posteriori estimates of the performance of the *Greedy* heuristic, in that one looks at the decomposition and tells how good it is. This potentially can reveal a good performance. For example, when *Greedy* obtains a BvN decomposition with all coefficients equivalent, then we know that it is an optimal BvN. The same cannot be told for the Birkhoff heuristic though (consider the example proceeding Lemma 2). We also note that the ratio given in Corollary 1 should usually be much larger than the practical performance.

4. Conclusion

We have investigated heuristics for obtaining Birkhoff–von Neumann decomposition of doubly stochastic matrices and presented three results. First, there are matrices whose decompositions with the smallest number of permutation matrices cannot be obtained by any Birkhoff-like heuristic. Second, the worst-case approximation ratio of the original Birkhoff heuristic is $\Omega(n)$. Third, the performance of the *Greedy* heuristic depends on the values of matrix elements, and a bound using the first and the last coefficients found by *Greedy* is obtained.

The shown bound for the performance of *Greedy* is expected to be much larger than what one observes in practice, as the bound can even be larger than the upper bound on the number of permutation matrices. A tighter analysis should be possible to explain the practical performance of the *Greedy* heuristic (which was demonstrated earlier [4]).

References

[1] M. Marcus, R. Ree, Diagonals of doubly stochastic matrices, The Quarterly Journal of Mathematics 10 (1) (1959) 296–302.

[2] R. A. Brualdi, P. M. Gibson, Convex polyhedra of doubly stochastic matrices: I. Applications of the permanent function, Journal of Combinatorial Theory, Series A 22 (2) (1977) 194–230.

[3] R. A. Brualdi, Notes on the Birkhoff algorithm for doubly stochastic matrices, Canadian Mathematical Bulletin 25 (2) (1982) 191–199.

[4] F. Dufossé, B. Uçar, Notes on Birkhoff–von Neumann decomposition of doubly stochastic matrices, Linear Algebra and its Applications 497 (2016) 108–115.

[5] C.-S. Chang, W.-J. Chen, H.-Y. Huang, On service guarantees for input-buffered crossbar switches: A capacity decomposition approach by Birkhoff and von Neumann, in: Quality of Service, 1999. IWQoS '99. 1999 Seventh International Workshop on, 1999, pp. 79–86.

[6] J. Kulkarni, E. Lee, M. Singh, Minimum Birkhoff-von Neumann decomposition, Preliminary version `http://www.cs.cmu.edu/~euiwoonl/sparsebvn.pdf` of the paper which appeared in Proc. 19th International Conference Integer Programming and Combinatorial Optimization (IPCO 2017), Waterloo, ON, Canada, pp. 343–354 (June 2017).

[7] M. Benzi, B. Uçar, Preconditioning techniques based on the Birkhoff–von Neumann decomposition, Computational Methods in Applied Mathematics 17 (2017) 201–215.

[8] IBM ILOG CPLEX Optimizer, `http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/` (Last accessed: 2017).

[9] E. D. Dolan, The NEOS Server 4.0 administrative guide, Technical Memorandum ANL/MCS-TM-250, Mathematics and Computer Science Division, Argonne National Laboratory (2001).

[10] J. Czyzyk, M. P. Mesnier, J. J. Moré, The NEOS Server, IEEE Journal on Computational Science and Engineering, 5 (3) (1998) 68–75.

[11] W. Gropp, J. J. Moré, Optimization environments and the NEOS Server, in: M. D. Buhman, A. Iserles (Eds.), Approximation Theory and Optimization, Cambridge University Press, 1997, pp. 167–182.