Available online at www.sciencedirect.com

ScienceDirect

Fuzzy Sets and Systems 360 (2019) 97–116

FUZZY sets and systems

www.elsevier.com/locate/fss

# Factorization of matrices with grades via essential entries

Radim Belohlavek, Marketa Trneckova *

*Dept. Computer Science, Palacký University, Olomouc, Czech Republic*

Received 1 November 2017; received in revised form 24 May 2018; accepted 9 July 2018
Available online 14 July 2018

**Abstract**

We present new results regarding the problem of factorization of matrices with grades, or, equivalently, decomposition of fuzzy relations. In particular, we examine geometry of factorizations and the role of fuzzy concept lattices in factorizations of matrices with grades. The results make it possible to reduce input data and enable a more focused search for factors in the search space, and are intended to guide the design of greedy and other approximation algorithms for the decomposition problem, which itself is NP-hard. To demonstrate usefulness of these results, we propose a new factorization algorithm based on these results. Our experiments demonstrate improvements in the quality of factorizations due to the new approach. We conclude by presenting further research topics implied by our findings.


*Keywords:* Fuzzy logic; Matrix decomposition; Decomposition of fuzzy relations; Fuzzy Galois connection; Fuzzy concept lattice

## 1. Problem description

The problem we consider may be formulated in terms of matrices or, equivalently, in terms of relations. We proceed for matrices, which framework is commonly used for this problem. Let $L$ denote a partially ordered set of grades bounded by 0 and 1. We primarily interpret the grades as truth degrees and denote by $L^{n\times m}$ the set of all matrices $I$ with $n$ rows and $m$ columns. That is, $I_{ij}$ is a truth degree in $L$ and we interpret it as the degree to which the object represented by $i$ has the attribute represented by $j$. We assume that $L$ is equipped with a binary operation $\otimes$ with respect to which it forms a residuated lattice [11]; see also below. This operation is involved in the matrix composition $\circ$, which is defined by

$$(A \circ B)_{ij} = \bigvee_{l=1}^{k} A_{il} \otimes B_{lj}. \tag{1}$$

The formulation of our problem involves the measure $s : L^{n\times m} \times L^{n\times m} \to [0, 1]$ of similarity, or approximate equality, of matrices. For this purpose, two direct generalizations of the similarity measure routinely utilized for Boolean matrices were considered in [5]. The first one is defined by $s(I, J) = \frac{|\langle i,j\rangle; I_{ij}=J_{ij}|}{n\cdot m}$. Since this measure only takes into

* Corresponding author.
*E-mail addresses:* radim.belohlavek@acm.org (R. Belohlavek), marketa.trneckova@gmail.com (M. Trneckova).

https://doi.org/10.1016/j.fss.2018.07.004
0165-0114/

account the entries in which $I$ and $J$ are equal and hence disregards entries in which $I$ and $J$ are possibly highly similar rather than equal, we use the second one, namely:

$$s(I, J) = \frac{\sum_{i,j=1}^{n,m} I_{ij} \leftrightarrow J_{ij}}{n \cdot m}, \tag{2}$$

where $\leftrightarrow$ is the biresiduum, which is defined in terms of the residuum $a \to b = \bigvee\{c \mid a \otimes c \leq b\}$ via $a \leftrightarrow b = \min(a \to b, b \to a)$. The biresiduum naturally measures closeness of truth degrees. For instance, $a \leftrightarrow b = 1 - |a - b|$ for the Łukasiewicz conjunction $\otimes$. Observe that both functions satisfy $s(I, J) = 1$ iff $I = J$, and that both generalize the Boolean matrix similarity in that for $L = \{0, 1\}$, $s(I, J) = 1 - d(I, J)$ with $d$ begin the Hamming distance.

The problem, which we call the *approximate factorization problem*, shortly *AFP(L)* to explicitly mention $L$, is defined as follows:

Given a matrix $I \in L^{n\times m}$ and a prescribed precision (or similarity threshold) $\varepsilon \in [0, 1]$, find matrices $A \in L^{n\times k}$ and $B \in L^{k\times m}$ with the least $k$ number of factors possible such that $s(I, A \circ B) \geq \varepsilon$.

**Remark 1** *(Interpretation).* (a) Clearly, if $\varepsilon = 1$, the problem is to find an exact decomposition $I = A \circ B$ which is optimal in that the number $k$ of factors is the smallest possible.

(b) The problem is naturally interpreted as a problem of *factor analysis*; see e.g. [8]: Compute from the input object-attribute relationship represented by $I$ a small number $k$ of factors that explain the input data. Explanation here means the following. The degrees $A_{il}$ and $B_{lj}$ in the object-factor matrix $A$ and the factor-attribute matrix $B$ are interpreted as the degree to which factor $l$ applies to object $i$, and the degree to which attribute $j$ is one of the particular manifestations of factor $l$. Then, $I = A \circ B$ means that object $i$ has attribute $j$ iff there exists factor $l$ such that $l$ applies to $i$ and $j$ is a particular manifestation of $l$. We return to the significance of this factor analysis interpretation below in our experimental evaluation.

**Remark 2** *(Connection to decomposition of fuzzy relations).* (a) Since matrices $M$ with entries in $L$ may equivalently be represented by fuzzy relations $R_M$, and since matrix composition defined by (1) corresponds to the well-known $\circ$-composition of fuzzy relations, the AFP problem may equivalently be formulated as a problem of decomposition of fuzzy relations.

(b) Notice that the problem is, nevertheless, very different from the problem of solving fuzzy relational equations, because there, two fuzzy relations, $R_I$ and $R_A$, or $R_I$ and $R_B$, are assumed to be known in the equation $R_I = R_A \circ R_B$. In AFP, only $R_I$ is known. This has fundamental algorithmic consequences: While solving fuzzy relational relations may be done in polynomial time, AFP is NP-hard as we shall prove.

**Remark 3** *(Directly related work).* Our problem is an obvious generalization of the problem of Boolean matrix factorization, which is its particular case for $L = \{0, 1\}$. For an overview of recent works in Boolean matrix factorization we refer, e.g., to [6].

Matrices with grades in partially ordered sets $L$ (matrices over $L$), and fuzzy relations with truth degrees in $L$, are examined in many papers; for those with $L$ being residuated structures of truth degrees we refer e.g. to [1,11]. As far as factorization of matrices over $L$ is concerned, we refer to [3,8], in which both the fundamental properties of formal concepts of $I$ as optimal factors as well as the first decomposition algorithm are presented. [5] provides factor analyses of various sports datasets by means of this algorithm and examines additional theoretical problems inspired by the analyses. A related problem, in a sense dual to the present one, in that one attempts to extract a given (small) number of factors with the highest coverage is the subject of our recent paper [4]. Methods of analysis of ordinal data also appear in the psychological literature but the tools employed are basically variations of classical factor analysis. That is, grades are represented by and treated like numbers which leads to loss of interpretability, similarly as in the case of Boolean data, see e.g. [14].

## 2. Theoretical analysis

### 2.1. Hardness of the approximate factorization problem

We start by the following observation which is crucially important for algorithmic considerations. In view of the fact that the problem of exact decomposition is NP-hard ([13], see also [7,12]), this theorem is actually not surprising.

**Theorem 1.** *The approximate factorization problem is an NP-hard optimization problem for any $L$.*

**Proof.** The proof proceeds by adaptation of the proofs of NP-hardness of the exact decomposition problem in the Boolean case (see e.g. [7]). We need to take into account that instead of the two-element Boolean algebra, we work with an arbitrary complete residuated lattice $L$, and that instead of exact factorization, our argument has to cover approximate factorization. We prove the claim by showing that the restriction of the problem to instances with $\varepsilon = 1$ is NP-hard. Due to our assumptions, $s(I, A \circ B) \geq \varepsilon$ is equivalent to $A \circ B = I$ in this case. According to the definition of NP-hardness, it suffices to verify that the corresponding decision problem, $\Pi$, is NP-complete. $\Pi$ consists in deciding whether for a given $I \in L^{n\times m}$ and $k$ there exists $A \in L^{n\times k}$ and $B \in L^{k\times m}$ with $A \circ B = I$.

The Boolean version of $\Pi$ is NP-complete because it is a reformulation (see e.g. [7]) of the set basis problem whose NP-completeness is due to [13]. To finish the proof it thus suffices to check that the restriction of $\Pi$ to Boolean input matrices $I$ is NP-complete. But the latter fact follows since for a Boolean $I$, there exist $A \in L^{n\times k}$ and $B \in L^{k\times m}$ with $A \circ B = I$ iff there exist Boolean matrices $A \in \{0,1\}^{n\times k}$ and $B \in \{0,1\}^{k\times m}$ with $A \circ B = I$. Namely, if $A \circ B = I$ for $A \in L^{n\times k}$ and $B \in L^{k\times m}$ then $A' \circ B' = I$ for the Boolean $A'$ and $B'$ defined by $A'_{il} = 1$ if $A_{il} = 1$, $A'_{il} = 0$ if $A_{il} < 1$, and the same for $B'$, which is easily seen from the isotony of $\otimes$.

In view of the preceding theorem, we need to resort to algorithms providing approximate solutions to these problems. Before presenting our algorithm, we proceed with new theoretical results regarding geometry of decomposition which are directly utilized in the design of GREESS$_L$, a new decomposition algorithm presented in Section 3.

### 2.2. Essential parts of matrices with grades

#### 2.2.1. Formal concepts and their optimality as factors

Our new algorithm utilizes formal concepts of the fuzzy concept lattice $\mathcal{B}(I)$ associated to the input matrix $I$; see e.g. [1,2]. Using formal concepts as factor of matrices with grades is not new; see [3,8]. The novelty in our approach consists in obtaining new results regarding geometry of decompositions and utilizing these results for an efficient selection of formal concepts as factors. In addition, we prove that formal concepts are optimal factors not only for exact decompositions but also for particular approximate decompositions which we call from-below approximations (see Theorem 2 and the definition preceding the theorem).

We start by recalling some basic notions; see e.g. [3,8]. We call $J \in L^{n\times m}$ a rectangular matrix, shortly a *rectangle*, if for some column $C \in L^{n\times 1}$ and row $D \in L^{1\times m}$ we have $J = C \circ D$. Contrary to the Boolean case, the $C$ and $D$ for which $J = C \circ D$ are not unique, which makes the situation more difficult as we shall see. We say that a matrix $J$ is included in matrix $I$, in symbols $J \leq I$, if $J_{ij} \leq I_{ij}$ for every entry $\langle i, j\rangle$. $J$ is a *rectangle in* $I$ iff $J$ is a rectangle that is included in $I$. A rectangle $J$ *covers* entry $\langle i, j\rangle$ in $I$ if $J_{ij} = I_{ij}$. The next lemma extends the observation in [3]:

**Lemma 1.** *For any matrix $I \in L^{n\times m}$, the following conditions are equivalent:*

(a) *$I = A \circ B$ for some $A \in L^{n\times k}$ and $B \in L^{k\times m}$.*
(b) *$I = J_1 \vee \cdots \vee J_k$ for some rectangles $J_1, \ldots, J_k \in L^{n\times m}$, where $(J_1 \vee \cdots \vee J_k)_{ij} = \max_{l=1}^{k}(J_l)_{ij}$.*

*If $L$ is linearly ordered, then (a) and (b) are equivalent to the following condition:*

(c) *There exist rectangles $J_1, \ldots, J_k \in L^{n\times m}$ in $I$ such that every $\langle i, j\rangle$ in $I$ is covered by some $J_l$.*

**Proof.** (a) ⇒ (b): Denote for $l = 1, \ldots, k$ by $J_l$ the rectangle which results as the product $J_l = A_{\_l} \circ B_{l\_}$ of the $l$th column $A_{\_l}$ of $A$ and the $l$th row $B_{l\_}$ of $B$, i.e. $(J_l)_{ij} = A_{il} \otimes B_{lj}$. Then

$$I_{ij} = (A \circ B)_{ij} = \max_{l=1}^{k} A_{il} \otimes B_{lj} = \max_{l=1}^{k} (J_l)_{ij}.$$

(b) ⇒ (a): Let $J_l = C_l \circ D_l$ for $l = 1, \ldots, k$. Consider the matrices $A$ and $B$ such that for each $l$, the $l$th column of $A$ equals $C_l$ and the $l$th row of $B$ equals $D_l$. A direct computation as above concludes the proof, namely:

$$I_{ij} = \max_{l=1}^{k} (J_l)_{ij} = \max_{l=1}^{k} (C_l)_i \otimes (D_l)_j = \max_{l=1}^{k} A_{il} \otimes B_{lj} = (A \circ B)_{ij}.$$

The equivalence (b) ⇔ (c) when $L$ is linearly ordered, is obvious.

The above lemma thus makes it possible to regard a decomposition of $I$ as a coverage of entries in $I$ by rectangles contained in $I$.

**Example 1.** Consider the matrix

$$I = \begin{pmatrix} 0.5 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.5 \\ 0.0 & 0.5 & 0.0 \end{pmatrix},$$

which we shall use as in our illustrative examples in the subsequent parts, and the Łukasiewicz operations on $L = \{0, 0.5, 1\}$. The matrix may be decomposed as follows (as we shall see in Example 4, the decomposition may be obtained using our new algorithm):

$$I = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \\ 0.5 & 0.5 \end{pmatrix} \circ \begin{pmatrix} 0.5 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.5 \end{pmatrix}.$$

We thus have $I = A \circ B$ as in Lemma 1 (a) with $k = 2$. According to Lemma 1 (b) and the above proof, $I$ may be expressed as a max-superposition of rectangles $J_1$ and $J_2$ which are obtained from the columns of $A$ and rows of $B$ as described in the proof. In particular,

$$I = \begin{pmatrix} 0.5 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.0 \\ 0.0 & 0.5 & 0.0 \end{pmatrix} \vee \begin{pmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.5 \\ 0.0 & 0.0 & 0.0 \end{pmatrix}.$$

We now extend the basic property of formal concepts in $I$ as optimal factors for exact decompositions of $I$ to the case when also approximate decompositions are desired. Recall that formal concepts are studied within formal concept analysis (for Boolean case see [10], for fuzzy setting see e.g. [1]) and that they are in fact fixpoints of Galois connections associated to relations (Boolean setting) and fuzzy relations (fuzzy setting). Since we use the language of vectors and matrices rather than sets and relations, we accommodate the notions of formal concept analysis to the language of vectors and matrices, but switch freely to sets and relations when convenient. A *formal concept* in a matrix $I \in L^{n\times m}$ (representing a fuzzy relation) is a pair $\langle C, D\rangle$ consisting of a vector $C \in L^{1\times n}$ (representing a fuzzy set of $n$ objects) and a vector $D \in L^{1\times m}$ (representing a fuzzy set of $m$ attributes) which satisfy $C^{\uparrow_I} = D$ and $D^{\downarrow_I} = C$; here, the operators are defined by

$$(C^{\uparrow_I})_j = \textstyle\bigwedge_{i=1}^{n} (C_i \to I_{ij}) \text{ and } (D^{\downarrow_I})_i = \bigwedge_{j=1}^{m} (D_j \to I_{ij}). \tag{3}$$

Note that when the inducing matrix $I$ is obvious, the subscripts are omitted in $^{\uparrow_I}$ and $^{\downarrow_I}$ and hence one uses $^{\uparrow}$ and $^{\downarrow}$ only. The collection of all fixpoints of $^{\uparrow}$ and $^{\downarrow}$, i.e.

$$\mathcal{B}(I) = \{\langle C, D\rangle \mid C^{\uparrow} = D, D^{\downarrow} = C\},$$

is called a (fuzzy) *concept lattice* of $I$. It is indeed a complete lattice, in which the partial order $\le$ is defined by $\langle C_1, D_1\rangle \le \langle C_2, D_2\rangle$ iff $C_1 \le C_2$, or, dually $D_2 \le D_1$.

Table 1
All formal concepts of $\mathcal{B}(I)$.

| $c_i$ | Extent | Intent |
|---|---|---|
| $c_0$ | $\{^{1}/x, {}^{1}/y, {}^{1}/z\}$ | $\{^{0}/u, {}^{0.5}/v, {}^{0}/w\}$ |
| $c_1$ | $\{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}$ | $\{^{0}/u, {}^{0.5}/v, {}^{0.5}/w\}$ |
| $c_2$ | $\{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}$ | $\{^{0.5}/u, {}^{1}/v, {}^{0}/w\}$ |
| $c_3$ | $\{^{0.5}/x, {}^{0.5}/y, {}^{0.5}/z\}$ | $\{^{0.5}/u, {}^{1}/v, {}^{0.5}/w\}$ |
| $c_4$ | $\{^{0.5}/x, {}^{0}/y, {}^{0}/z\}$ | $\{^{1}/u, {}^{1}/v, {}^{0.5}/w\}$ |
| $c_5$ | $\{^{0}/x, {}^{0.5}/y, {}^{0}/z\}$ | $\{^{0.5}/u, {}^{1}/v, {}^{1}/w\}$ |
| $c_6$ | $\{^{0}/x, {}^{0}/y, {}^{0}/z\}$ | $\{^{1}/u, {}^{1}/v, {}^{1}/w\}$ |

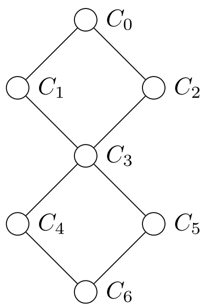


Fig. 1. Diagram of $\mathcal{B}(I)$.

**Example 2.** Consider again the matrix (cf. Example 1)

$$I = \begin{pmatrix} 0.5 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.5 \\ 0.0 & 0.5 & 0.0 \end{pmatrix}.$$

Let $X = \{x, y, z\}$ and $Y = \{u, v, w\}$ denote the sets of objects (matrix rows) and attributes (matrix columns) and consider the Łukasiewicz operations on $L = \{0, 0.5, 1\}$. The set of all formal concepts of the corresponding fuzzy concept lattice $\mathcal{B}(I)$ is listed in Table 1. Fig. 1 displays the Hasse diagram of $\mathcal{B}(I)$.

It turns out that the rectangles $C^{\mathrm{T}} \circ D$ ($C^{\mathrm{T}}$ denotes the transpose of $C$) corresponding to formal concepts $\langle C, D\rangle$ of $I$ play an important role in decompositions of $I$. Namely, denote for an indexed collection $\mathcal{F} = \{\langle C_1, D_1\rangle, \ldots, \langle C_k, D_k\rangle\} \subseteq \mathcal{B}(I)$ by $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ the $n \times k$ and $k \times m$ matrices defined as follows:

$$(A_{\mathcal{F}})_{il} = (C_l)_i \text{ and } (B_{\mathcal{F}})_{lj} = (D_l)_j. \tag{4}$$

That is, the $l$th column $A_{\mathcal{F}}$ is $C_l^{\mathrm{T}}$ and the $l$th row of $B_{\mathcal{F}}$ is $D_l$. Optimality of formal concepts as factors for certain approximate decompositions of $I$ into $A$ and $B$, which satisfy $A \circ B \le I$ and which we call *from-below approximations* of $I$, is shown by the following theorem (notice that these include exact decompositions).

**Theorem 2.** *Let $A \in L^{n\times k}$ and $B \in L^{k\times m}$ provide a from-below approximation of a given $I \in L^{n\times m}$, i.e. $A \circ B \le I$. Then there is a set $\mathcal{F} \subseteq \mathcal{B}(I)$ of at most $k$ formal concepts, i.e. $|\mathcal{F}| \le k$, such that the $n \times |\mathcal{F}|$ and $|\mathcal{F}| \times m$ matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ provide at least as good from-below approximation of $I$ as $A$ and $B$, i.e.*

$$s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \ge s(I, A \circ B).$$

**Proof.** Since $A \circ B \le I$, Lemma 1 implies that every rectangle $J_l = A_{\_l} \circ B_{l\_}$ is contained in $I$ (here again, $A_{\_l}$ and $B_{l\_}$ denote the $l$th column of $A$ and the $l$th row of $B$). Consider the pairs $\langle (A^{\mathrm{T}}_{\_l})^{\uparrow\downarrow}, (A^{\mathrm{T}}_{\_l})^{\uparrow}\rangle$. Every $\langle (A^{\mathrm{T}}_{\_l})^{\uparrow\downarrow}, (A^{\mathrm{T}}_{\_l})^{\uparrow}\rangle$ is a formal concept in $\mathcal{B}(I)$ (a well-known fact in FCA).

Moreover $A^{\mathrm{T}}_{\_l} \le (A^{\mathrm{T}}_{\_l})^{\uparrow\downarrow}$, because ${}^{\uparrow\downarrow}$ is a closure operator. Since, $A_{\_l} \circ B_{l\_}$ is contained in $I$, a straightforward computation using adjointness of $\otimes$ and $\to$ implies $B_{l\_} \le (A^{\mathrm{T}}_{\_l})^{\uparrow}$. Now consider the set

$$\mathcal{F} = \{\langle (A^{\mathrm{T}}_{\_1})^{\uparrow\downarrow}, (A^{\mathrm{T}}_{\_1})^{\uparrow}\rangle, \ldots, \langle (A^{\mathrm{T}}_{\_k})^{\uparrow\downarrow}, (A^{\mathrm{T}}_{\_k})^{\uparrow}\rangle\} \subseteq \mathcal{B}(I)$$

and the matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$. Clearly $\mathcal{F}$ contains at most $k$ elements (it may happen $|\mathcal{F}| < k$). It is easy to check that the rectangle corresponding to $\langle (A^{\mathrm{T}}_{\_l})^{\uparrow\downarrow}, (A^{\mathrm{T}}_{\_l})^{\uparrow}\rangle$, i.e. the cross-product $(A_{\mathcal{F}})_{\_l} \circ (B_{\mathcal{F}})_{l\_}$, is contained in $I$ and, due to the above observation, contains $J_l = A_{\_l} \circ B_{l\_}$. Hence,

$$A \circ B \le \max_{l=1}^{k} J_l \le \max_{l=1}^{k} (A_{\mathcal{F}})_{\_l} \circ (B_{\mathcal{F}})_{l\_} = A_{\mathcal{F}} \circ B_{\mathcal{F}} \le I.$$

Since $a \le b \le c$ implies $a \leftrightarrow c \le b \leftrightarrow c$, we readily obtain $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \ge s(I, A \circ B)$, finishing the proof.

Our interest in the from-below approximations of $I$ is twofold. First, these approximations are constructed when computing an exact decomposition of $I$ by successively computing the factors. Second, even if we are interested in obtaining an approximate (rather than an exact) factorization of $I$ only, the algorithms providing from-below approximations perform very well (see our experimental evaluation). Part of the reason is the fact that the from-below approximations are amenable to theoretical analysis in terms of closure and order-theoretic structures, which we demonstrate by the results below.

### 2.3. Essential entries in matrices with grades

We now examine in detail the coverage problem by rectangles, to which the decomposition problem may be transformed (previous section). A closer examination of the concept lattice $\mathcal{B}(I)$ associated to the input matrix $I$ shows a possibility to identify entries in $I$ that are crucial in search of factors, and therefore to differentiate matrix entries that are essential for decompositions from the other entries. We thus identify what we call the essential part of $I$: a minimal set of entries whose coverage by factors guarantees exact decomposition of $I$ by these factors. We show later that the number of such entries is significantly smaller than the number of all entries. Most importantly, the essential part may be seen as the part to focus on when computing decompositions. This view is utilized in the design of a decomposition algorithm in the next section. Note that the idea of differentiating the role of entries is inspired by [6], but the situation is considerably more involved in the setting of grades compared to the Boolean case.

**Definition 1.** We call a matrix $J \leq I$ an *essential part* of $I$ if $J$ is minimal w.r.t. $\leq$ such that for every $\mathcal{F} \subseteq \mathcal{B}(I)$: $J \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$ implies $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

That is, if a certain collection $\mathcal{F}$ of formal concepts of $I$ cover an essential part $J$ of $I$, these concepts are guaranteed to cover all entries in $I$ and thus provide an exact decomposition of $I$.

In these considerations, certain intervals in $\mathcal{B}(I)$ are of fundamental importance. Let for $C \in L^{1\times n}$ and $D \in L^{1\times m}$ denote

$$\gamma(C) = \langle C^{\uparrow\downarrow}, C^{\uparrow}\rangle \text{ and } \mu(D) = \langle D^{\downarrow}, D^{\downarrow\uparrow}\rangle,$$

and denote furthermore by $\mathcal{I}_{C,D}$ the interval

$$\mathcal{I}_{C,D} = [\gamma(C), \mu(D)]$$

in the lattice $\mathcal{B}(I)$, i.e. the set

$$[\gamma(C), \mu(D)] = \{\langle E, F\rangle \in \mathcal{B}(I) \mid \gamma(C) \leq \langle E, F\rangle \leq \mu(D)\}.$$

Note that we need not have $\gamma(C) \leq \mu(D)$ in which case the interval is the empty set. The following lemma shows that all the rectangles corresponding to the formal concepts in $\mathcal{I}_{C,D}$ cover the rectangle $C^{\mathrm{T}} \circ D$ in $I$.

**Lemma 2.** *If $\langle E, F\rangle \in \mathcal{I}_{C,D}$ then $C^{\mathrm{T}} \circ D \leq E^{\mathrm{T}} \circ F$.*

**Proof.** Since $\langle E, F\rangle \in \mathcal{I}_{C,D}$, we have $C^{\uparrow\downarrow} \leq E$ and $D^{\downarrow\uparrow} \leq F$. As $\langle E, F\rangle$ is a formal concept of $I$, we have $E = E^{\uparrow\downarrow}$ and $F = F^{\downarrow\uparrow}$. Since $^{\uparrow\downarrow}$ and $^{\downarrow\uparrow}$ are closure operators, we obtain $C \leq C^{\uparrow\downarrow} \leq C^{\uparrow\downarrow\uparrow\downarrow} = E^{\uparrow\downarrow} \leq E$ and similarly $D \leq F$. The claim now easily follows.

In particular, consider $C = \{^{a}/i\}$, by which we denote the "singleton" vector with zero components except $C_i = a$, and $D = \{^{b}/j\}$ with analogous meaning. Then every concept $\langle E, F\rangle$ in $\mathcal{I}_{C,D} = \mathcal{I}_{\{^{a}/i\},\{^{b}/j\}}$ covers the entry $\langle i, j\rangle$ in $C^{\mathrm{T}} \circ D$. This means that if $a \otimes b = I_{ij}$, then every concept in $\mathcal{I}_{\{^{a}/i\},\{^{b}/j\}}$ covers the entry $\langle i, j\rangle$ in $I$. Conversely, however, the entry $\langle i, j\rangle$ in $I$ is covered also by other concepts than those in $\mathcal{I}_{\{^{a}/i\},\{^{b}/j\}}$. The following lemma is crucial in understanding this issue.

**Lemma 3.** *Let $\langle E, F\rangle \in \mathcal{B}(X, Y, I)$ and $a, b \in L$. Then $a \otimes b \leq E_i \otimes F_j$ if and only if for some $c, d$ with $a \otimes b \leq c \otimes d$ we have $\langle E, F\rangle \in \mathcal{I}_{\{^{c}/i\},\{^{d}/j\}}$.*

**Proof.** If $a \otimes b \leq E_i \otimes F_j$, one may put $c = E_i$ and $d = F_j$. Namely, we then have to check $\langle E, F\rangle \in \mathcal{I}_{\{^{E_i}/i\},\{^{F_j}/j\}}$ which is equivalent to $\gamma(\{^{E_i}/i\}) \leq E$ and $\mu(\{^{F_j}/j\}) \leq F$. The first inequality is equivalent to $\{^{E_i}/i\}^{\uparrow\downarrow} \leq E$ which is true. Namely, from the obvious fact $\{^{E_i}/i\} \leq E$ we obtain by isotony and idempotency of ${}^{\uparrow\downarrow}$ that $\{^{E_i}/i\}^{\uparrow\downarrow} \leq E^{\uparrow\downarrow} = E$. The second inequality is obtained symmetrically.

Conversely, assume that for some $c, d$ with $a \otimes b \leq c \otimes d$ we have $\langle E, F\rangle \in \mathcal{I}_{\{^c/i\},\{^d/j\}}$. Lemma 2 then implies $\{^c/i\}^{\mathrm{T}} \circ \{^d/j\} \leq E^{\mathrm{T}} \circ F$, which entails $c \otimes d \leq E_i \otimes F_j$. Since $a \otimes b \leq c \otimes d$, the proof is finished.

Now, for a given matrix $I \in L^{n\times m}$ and an entry position $\langle i, j\rangle$, let

$$\mathbf{I}_{ij} = \{\mathcal{I}_{\{^a/i\},\{^b/j\}} \mid a, b \in L, a \otimes b = I_{ij}\}$$

be the system of intervals in $\mathcal{B}(I)$ associated to $\langle i, j\rangle$ and let

$$\mathcal{I}_{ij} = \bigcup \mathbf{I}_{ij},$$

i.e. $\mathcal{I}_{ij}$ is the union of this system of intervals.

Note that the situation is much easier in the Boolean case. Namely, if $I_{ij} > 0$, then $\mathcal{I}_{ij}$ consists of a single interval in the Boolean case because the only $a$ and $b$ for which $a \otimes b = 1$ are $a = b = 1$. In the setting of grades, there may be several pairs of $a$ and $b$ for which $I_{ij} = a \otimes b$, hence several intervals of which $\mathcal{I}_{ij}$ consists. Next, we obtain an important theorem which shows that $\mathcal{I}_{ij}$ is just the set of all formal concepts of $I$ that cover $\langle i, j\rangle$.

**Theorem 3.** *The rectangle corresponding to $\langle E, F\rangle \in \mathcal{B}(X, Y, I)$ covers $\langle i, j\rangle$ in $I$ iff $\langle E, F\rangle \in \mathcal{I}_{ij}$.*

**Proof.** If $E^{\mathrm{T}} \circ F$ covers $\langle i, j\rangle$, i.e. $I_{ij} = E_i \otimes F_j$, then since $I_{ij} = I_{ij} \otimes 1$, we obtain $\langle E, F\rangle \in \mathcal{I}_{ij}$ by Lemma 3.

Conversely, let $\langle E, F\rangle \in \mathcal{I}_{ij}$, i.e. $\langle E, F\rangle \in \mathcal{I}_{\{^a/i\},\{^b/j\}}$ for some $a, b$ with $a \otimes b = I_{ij}$. Lemma 3 then implies $I_{ij} = a \otimes b \leq E_i \otimes F_j$. Since the definition of a formal concept of $I$ along with adjointness yield that we always have $E_i \otimes F_j \leq I_{ij}$, we readily obtain $E_i \otimes F_j = I_{ij}$, finishing the proof.

Denote now by $\mathcal{E}(I) \in L^{n\times m}$ the matrix over $L$ defined by

$$(\mathcal{E}(I))_{ij} = \begin{cases} I_{ij} & \text{if } \mathcal{I}_{ij} \text{ is } \neq \emptyset \text{ and minimal w.r.t. } \subseteq, \\ 0 & \text{otherwise.} \end{cases}$$

Note that minimality of $\mathcal{I}_{ij}$ w.r.t. the set inclusion $\subseteq$ in the above definition of $(\mathcal{E}(I))_{ij}$ means that there is no $\mathcal{I}_{i'j'}$ properly contained in $\mathcal{I}_{ij}$, i.e. that if $\mathcal{I}_{i'j'} \subseteq \mathcal{I}_{ij}$ then $\mathcal{I}_{i'j'} = \mathcal{I}_{ij}$ for any $i'$ and $j'$.

**Example 3.** Consider our illustrative matrix from Example 1,

$$I = \begin{pmatrix} 0.5 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.5 \\ 0.0 & 0.5 & 0.0 \end{pmatrix},$$

and its concept lattice $\mathcal{B}(I)$ from Example 2. Consider the entry $I_{\langle y,v\rangle} = 0.5$ of $I$. Since $I_{\langle y,v\rangle}$ may be decomposed in two ways, namely $I_{\langle y,v\rangle} = 0.5 \otimes 1$ and $I_{\langle y,v\rangle} = 1 \otimes 0.5$, there exist two intervals in $\mathcal{B}(I)$ corresponding to this entry, namely

$$\mathcal{I}_{\{^{0.5}/y\},\{^1/v\}} = [c_5, c_2] = [\langle \{^{0.5}/y\}^{\uparrow\downarrow}, \{^{0.5}/y\}^{\uparrow}\rangle, \langle \{^1/v\}^{\downarrow}, \{^1/v\}^{\downarrow\uparrow}\rangle]$$

and

$$\mathcal{I}_{\{^1/y\},\{^{0.5}/v\}} = [c_1, c_0] = [\langle \{^1/y\}^{\uparrow\downarrow}, \{^1/y\}^{\uparrow}\rangle, \langle \{^{0.5}/v\}^{\downarrow}, \{^{0.5}/v\}^{\downarrow\uparrow}\rangle].$$

The corresponding system of intervals, $\mathbf{I}_{yv}$, is therefore given by

$$\mathbf{I}_{yv} = \{\mathcal{I}_{\{^{0.5}/y\},\{^1/v\}}, \mathcal{I}_{\{^1/y\},\{^{0.5}/v\}}\},$$

and we have

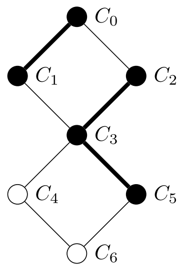


Fig. 2. $\mathcal{B}(I)$ with the two intervals comprising $\mathbf{I}_{yv}$.

$$\mathcal{I}_{yv} = \mathcal{I}_{\{0.5/y\},\{1/v\}} \cup \mathcal{I}_{\{1/y\},\{0.5/v\}};$$

see Fig. 2.

Now, $\mathcal{I}_{yv}$ is nonempty but is not minimal w.r.t. set inclusion. Namely, as one may check,

$$\mathcal{I}_{xv} = [c_2, c_2] = \{c_2\} \subset \mathcal{I}_{yv}.$$

We thus obtain $\mathcal{E}(I)_{yv} = 0$. Proceeding entry by entry, we finally obtain

$$\mathcal{E}(I) = \begin{pmatrix} 0.5 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.5 \\ 0.0 & 0.5 & 0.0 \end{pmatrix}.$$

The following two theorems provide the main result in this section and are utilized in the new algorithm.

**Theorem 4.** *$\mathcal{E}(I)$ is an essential part of $I$.*

**Proof.** First, $\mathcal{E}(I) \leq I$ follows from the definition of $\mathcal{E}(I)$. Second, consider any $\mathcal{F} \subseteq \mathcal{B}(I)$ for which $\mathcal{E}(I) \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$. We need to show $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.

On one hand, $I \geq A_{\mathcal{F}} \circ B_{\mathcal{F}}$ is a consequence of the fact that every $\langle C, D\rangle \in \mathcal{F}$ is a formal concept of $I$. Namely, adjointness easily yields $C_i \otimes D_j \leq I_{ij}$ from which the required inequality directly follows.

It remains to prove $I \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$. Consider any $\langle i, j\rangle$ and the corresponding set $\mathcal{I}_{ij}$. Take any $\mathcal{I}_{i'j'} \subseteq \mathcal{I}_{ij}$ that is non-empty and minimal w.r.t. $\subseteq$. The definition of $\mathcal{E}(I)$ implies $\mathcal{E}(I)_{i'j'} = I_{i'j'}$. Since $\mathcal{E}(I) \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$, the definition of $\circ$ and of $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ imply the existence of $\langle C, D\rangle \in \mathcal{F}$ for which $\mathcal{E}(I)_{i'j'} \leq C_{i'} \otimes D_{j'}$. Since $\langle C, D\rangle$ is a formal concept of $I$, we also have $C_{i'} \otimes D_{j'} \leq I_{i'j'} \leq \mathcal{E}(I)_{i'j'}$, hence the rectangle corresponding to $\langle C, D\rangle$ covers $\langle i', j'\rangle$. Thanks to Theorem 3 we get $\langle C, D\rangle \in \mathcal{I}_{i'j'}$ and since $\mathcal{I}_{i'j'} \subseteq \mathcal{I}_{ij}$, also $\langle C, D\rangle \in \mathcal{I}_{ij}$. Applying Theorem 3 again now yields that the rectangle corresponding to $\langle C, D\rangle$ covers $\langle i, j\rangle$, i.e. $I_{ij} = C_i \otimes D_j \leq (A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij}$ and since we always have $(A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij} \leq I_{ij}$, we obtain $(A_{\mathcal{F}} \circ B_{\mathcal{F}})_{ij} = I_{ij}$. Since $\langle i, j\rangle$ is arbitrary, $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ follows.

The next theorem shows how a factorization of $\mathcal{E}(I)$ may be used to obtain a factorization of $I$.

**Theorem 5.** *Let $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$ be a set of factor concepts of $\mathcal{E}(I)$, i.e. $\mathcal{E}(I) = A_{\mathcal{G}} \circ B_{\mathcal{G}}$. Then every set $\mathcal{F} \subseteq \mathcal{B}(I)$ containing for each $\langle C, D\rangle \in \mathcal{G}$ at least one concept from $\mathcal{I}_{C,D}$ is a set of factor concepts of $I$, i.e. $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$.*

**Proof.** Let for $\langle C, D\rangle \in \mathcal{G}$ denote by $\langle E, F\rangle_{\langle C,D\rangle}$ a concept in $\mathcal{F} \cap \mathcal{I}_{C,D}$ (it exists by assumption). Due to Lemma 2, $\langle E, F\rangle_{\langle C,D\rangle}$ covers the rectangle corresponding to $\langle C, D\rangle$. Since this is true for every $\langle C, D\rangle \in \mathcal{G}$, it is easy to see that $A_{\mathcal{G}} \circ B_{\mathcal{G}} \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$. The assumption $\mathcal{E}(I) = A_{\mathcal{G}} \circ B_{\mathcal{G}}$ now yields $\mathcal{E}(I) \leq A_{\mathcal{F}} \circ B_{\mathcal{F}}$. As $\mathcal{E}(I)$ is an essential part of $I$, we get $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$, finishing the proof.

**Example 4.** Consider again the matrix from Example 1, its concept lattice $\mathcal{B}(I)$ from Example 2, and its essential part $\mathcal{E}(I)$ from Example 3. Now, according to Theorem 5, to obtain a factorization of $I$, it is sufficient to obtain a factorization $\mathcal{G}$ of $\mathcal{E}(I)$ and proceed accordingly, i.e. select for each $\langle C, D\rangle \in \mathcal{G}$ a formal concept in the interval $\mathcal{I}_{C,D}$ of $\mathcal{B}(I)$.

A particular factorization of $\mathcal{E}(I)$, selected in our new algorithm GREESS$_L$ by the function COMPUTEINTERVALS (see the next section) is:

$$\mathcal{G} = \{\langle\{^1/x, {}^{0.5}/y, {}^{0.5}/z\}, \{^{0.5}/u, {}^{0.5}/v\}\rangle, \langle\{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}, \{^{0.5}/w\}\rangle\}.$$

From the set $\mathcal{G}$, GREESS$_L$ computes concepts $c_2$ and $c_1$ as factors. For the corresponding set $\mathcal{F} = \{c_1, c_2\}$, we thus obtain a decomposition $I = A_\mathcal{F} \circ B_\mathcal{F}$, i.e.

$$I = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \\ 0.5 & 0.5 \end{pmatrix} \circ \begin{pmatrix} 0.5 & 1.0 & 0.0 \\ 0.0 & 0.5 & 0.5 \end{pmatrix}.$$

Let $\mathrm{rank}_L(J)$ denote the rank of matrix $J \in L^{n\times m}$, i.e. the least $k$ for which there exist matrices $A \in L^{n\times k}$ and $B^{k\times m}$ with $J = A \circ B$. The rank of the simpler matrix $\mathcal{E}(I)$ provides a proper upper bound on the rank of the given $I$:

**Theorem 6.** *For any $I \in L^{n\times m}$: $\mathrm{rank}_L(I) \le \mathrm{rank}_L(\mathcal{E}(I))$. Moreover, it may happen that $\mathrm{rank}_L(I) < \mathrm{rank}_L(\mathcal{E}(I))$.*

**Proof.** Let $k = \mathrm{rank}_L(\mathcal{E}(I))$. Due to optimality of concepts as formal factors [3], we may assume the existence of $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$ containing $k$ formal concepts that factorize $\mathcal{E}(I)$, i.e. $|\mathcal{G}| = k$ and $\mathcal{E}(I) = A_\mathcal{G} \circ B_\mathcal{G}$. Let $\mathcal{F} \subseteq \mathcal{B}(I)$ be any set of formal concepts which contains for every concept $\langle C, D\rangle \in \mathcal{G}$ exactly one concept in the interval $\mathcal{I}_{C,D}$ of $\mathcal{B}(I)$.

Such $\mathcal{F}$ indeed exists because due to $C^\mathrm{T} \circ D \le \mathcal{E}(I)$ and $\mathcal{E}(I) \le I$, we have $C^\mathrm{T} \circ D \le I$. Using basic properties of the operators $^\uparrow$ and $^\downarrow$, it is now easy to verify that $\mathcal{I}_{C,D}$ is nonempty. Since each such concept $\langle E, F\rangle \in \mathcal{I}_{C,D}$ satisfies $C^\mathrm{T} \circ D \le E^\mathrm{T} \circ F$, we obtain $A_\mathcal{G} \circ B_\mathcal{G} \le A_\mathcal{F} \circ B_\mathcal{F}$, and since $\mathcal{E}(I) = A_\mathcal{G} \circ B_\mathcal{G}$, Theorem 4 yields $A_\mathcal{F} \circ B_\mathcal{F} = I$. The proof is finished because, obviously, $\mathrm{rank}_L(I) \le |\mathcal{F}| \le |\mathcal{G}| \le \mathrm{rank}_L(\mathcal{E}(I))$.

**Remark 4.** The estimation is not tight. Namely, as one easily verifies,

$$I = \begin{pmatrix} 1 & 1 & 0.5 \\ 1 & 1 & 0 \\ 0.5 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0.5 \\ 0 & 0.5 \end{pmatrix} \circ \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0.5 & 0.5 \end{pmatrix},$$

i.e. $\mathrm{rank}_L(I) = 2$. For the essential part of $I$ we get

$$\mathcal{E}(I) = \begin{pmatrix} 0 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0 \end{pmatrix}.$$

One may easily check that $\mathrm{rank}_L(\mathcal{E}(I)) = 3 > \mathrm{rank}_L(I)$.

The following theorem shows that as far as exact decompositions of $I$ are concerned, it is in fact sufficient to select formal concepts of the set

$$\mathcal{B}_\mathcal{E}(I) = \bigcup \{\mathcal{I}_{ij} \mid (\mathcal{E}(I))_{ij} \ne 0\},$$

rather than formal concepts in the whole $\mathcal{B}(I)$ as suggested by Theorem 2 (note that $\mathcal{B}_\mathcal{E}(I)$ is in general a subset of $\mathcal{B}(I)$).

**Theorem 7.** *The rank $\mathrm{rank}_L(I)$ may be achieved by using formal concepts in $\mathcal{B}_\mathcal{E}(I)$ as factors.*

**Proof.** First, the optimality of formal concepts as factors [3] implies that $I = A_\mathcal{F} \circ B_\mathcal{F}$ for some $\mathcal{F} \subseteq \mathcal{B}(I)$ that consists of $\mathrm{rank}_L(I)$ elements. Therefore, it suffices to verify that $\mathcal{F} \subseteq \mathcal{B}_\mathcal{E}(I)$. We prove this claim by contradiction. Suppose $\langle C, D\rangle$ is in $\mathcal{F}$ but not in $\mathcal{B}_\mathcal{E}(I)$ for some $\langle C, D\rangle$. As $\mathcal{F}$ covers $I$, and hence also $\mathcal{E}(I)$, for every entry $\langle p, q\rangle$ with $(\mathcal{E}(I))_{pq} \ne 0$ there exists a formal concept $\langle E, F\rangle \in \mathcal{F}$ covering $\langle p, q\rangle$. Now, due to the definition of $\mathcal{I}_{pq}$, $\langle E, F\rangle \in \mathcal{I}_{pq}$, and hence $\langle E, F\rangle \ne \langle C, D\rangle$. If we exclude $\langle C, D\rangle$ from $\mathcal{F}$, we therefore obtain a set $\mathcal{F}'$ that still covers $\mathcal{E}(I)$, i.e. for $\mathcal{E}(I) \le A_\mathcal{F} \circ B_\mathcal{F}$. Theorem 4 then implies $A_\mathcal{F} \circ B_\mathcal{F} = I$. However, $|\mathcal{F}'| = |\mathcal{F}| - 1 < \mathrm{rank}_L(I)$, a contradiction to the fact that $\mathcal{F}'$ factorizes $I$.

Note also that the previous theorem may not be extended to from-below factorizations, i.e. the reduction in the search space (smaller $\mathcal{B}_{\mathcal{E}}(I)$ instead of the larger $\mathcal{B}(I)$ suffices) is only applicable when exact factorizations are concerned. This observation is easily proved by a counterexample.

## 3. A new factorization algorithm

The GREESS$_L$ algorithm (see Algorithms 1 and 2), which we now present, is based on the results presented in the previous section and some other properties mentioned below in this section. Due to algorithmic considerations, we assume that the set $L$ of grades is finite throughout this section. GREESS$_L$ computes, for a given matrix $I \in L^{n\times m}$ and a prescribed precision $\varepsilon > 0$ a set $\mathcal{F}$ of formal concepts of $I$, i.e. $\mathcal{F} \subseteq \mathcal{B}(I)$, such that the corresponding matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ provide an approximate decomposition of $I$ with the prescribed precision, i.e. $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \varepsilon$.

The algorithm is inspired by the GREESS algorithm [6], which is an algorithm for factorizing Boolean matrices. Our algorithm shares with GREESS the basic idea the essential part of the input matrix $I$ is utilized. However, the move from Boolean, i.e. yes/no, data to data with grades makes the problem considerably more involved, which is apparent when comparing the theoretical results in this paper to those from [6] (cf. also the relationship between the mathematical structures that lie behind both frameworks, i.e. ordinary closure structures vs. closure structures over complete residuated lattices; see e.g. [1]).

Before presenting a detailed description of the pseudocode of GREESS$_L$ (Algorithms 1 and 2), let us present the concept of our algorithm. To compute the set $\mathcal{F}$ of factors for which $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \varepsilon$, the algorithm exploits the essential part $\mathcal{E}(I)$ of $I$ and its properties as follows. In accordance with Theorem 5, which justifies correctness of the algorithm, GREESS$_L$ performs the following steps:

1. It computes $\mathcal{E}(I)$, i.e. the essential part of $I$.
2. It computes a set $\mathcal{G}$ of formal concepts in $\mathcal{B}(\mathcal{E}(I))$ that factorizes $\mathcal{E}(I)$, i.e. $\mathcal{E}(I) = A_{\mathcal{G}} \circ B_{\mathcal{G}}$.
3. It computes a set $\mathcal{F} \subseteq \mathcal{B}(I)$ of formal concepts by iteratively selecting concepts $\langle E, F\rangle$ from the intervals $\mathcal{I}_{C,D}$ of the concept lattice $\mathcal{B}(I)$, for all $\langle C, D\rangle \in \mathcal{G}$, in such a way that at most one concept is chosen from each $\mathcal{I}_{C,D}$, until $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \varepsilon$.

Let us now consider the pseudocodes. In these pseudocodes, $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$ denotes the similarity measure defined in (2). In addition, $\emptyset$ denotes the empty set (in $\mathcal{F} \leftarrow \emptyset$) or the vector full of zeroes (in $F \leftarrow \emptyset$). $F \vee \{^a/_j\}$ denotes the vector whose components coincide with those of $F$ except for the $j$th component, which is updated to $F_j \vee a$. $C \otimes D$ denotes the product $C^{\mathrm{T}} \circ D$ of $C^{\mathrm{T}}$ (the transpose of $C$) and $D$, i.e. the rectangular matrix in which for every row $i$ and column $j$ we have $(C \otimes D)_{ij} = C_i \otimes D_j$. Recall also that $^{\uparrow_K}$ and $^{\downarrow_K}$ denote the operators induced by the matrix $K$; see (3). Moreover, $U$ represents the collection of entries $\langle i, j\rangle$ which are not covered by the factors that have been computed so far (this pertains to both GREESS$_L$ and COMPUTEINTERVALS). $cov(U, F, J)$ represents the number of entries $\langle i, j\rangle \in U$ covered in $I$ by the rectangle $F^{\downarrow_J} \otimes F^{\downarrow_J \uparrow_J}$, while $cov_I(U, D, \mathcal{E})$ denotes the number of entries $\langle i, j\rangle \in U$ covered in $I$ by the rectangle $(D^{\downarrow_{\mathcal{E}}})^{\uparrow_I \downarrow_I} \otimes (D^{\downarrow_{\mathcal{E}} \uparrow_{\mathcal{E}}})^{\downarrow_I \uparrow_I}$. The variables $s$ and $s_{\langle C,D\rangle}$ store information about the largest values of $cov$ found in the particular loop. The notation $\{^a/_j\} \in C^{\uparrow_I} \setminus F$ means $F_j < a \leq C_j^{\uparrow_I}$.

Let us start with the function COMPUTEINTERVALS. The goal of this function is to compute a set $\mathcal{G}$ of formal concepts in $\mathcal{B}(\mathcal{E}(I))$, which represents a set of intervals in $\mathcal{B}(I)$, namely the intervals $\mathcal{I}_{C,D}$ for $\langle C, D\rangle \in \mathcal{G}$. For this purpose, the function COMPUTEINTERVALS first computes the essential part $\mathcal{E}(I)$ which is done by definition. Consequently, COMPUTEINTERVALS computes a set $\mathcal{G}$ of factors of the matrix $\mathcal{E}(I)$ in view of Theorem 5 (the computation itself is described in the next paragraph). Thus, each $\langle C, D\rangle \in \mathcal{G}$ represents the interval $\mathcal{I}_{C,D}$ in $\mathcal{B}(I)$. According to Theorem 5, a set $\mathcal{F}$ of factors of the matrix $I$, i.e. a decomposition of $I$, may then be obtained by searching the intervals $\mathcal{I}_{C,D}$. In our algorithm, we actually use an improvement of Theorem 5, whose proof is easy and thus omitted: The set $\mathcal{G}$ of concepts in $\mathcal{B}(\mathcal{E}(I))$ need not provide an exact factorization of $\mathcal{E}(I)$. Rather, it is sufficient that the rectangles obtained as $C^{\uparrow_I \downarrow_I} \otimes D^{\downarrow_I \uparrow_I}$, which correspond to $\langle C, D\rangle \in \mathcal{G}$, cover all entries in $I$ (l. 11). The next two paragraphs describe the computation of $\mathcal{G}$ and then, given $\mathcal{G}$, the computation of the required set $\mathcal{F}$ of factors.

To obtain a collection $\mathcal{G}$ of formal concepts $\langle C, D\rangle$ in $\mathcal{B}(\mathcal{E}(I))$ with the above properties, the formal concepts are computed from $\mathcal{E}(I)$ in a greedy manner adopted from the algorithm FIND-FACTORS [8] as follows: One first computes the essential part $\mathcal{E}$ of $I$ and assigns to $U$ the set of all entries $\langle i, j\rangle$ for which $\mathcal{E}_{ij} > 0$, and which thus need

**Algorithm 1:** GreEss$_L$.

**Input**: matrix $I$ with entries in $L$, threshold $\varepsilon$
**Output**: set $\mathcal{F}$ of factors for which $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$
$\mathcal{G} \leftarrow$ ComputeIntervals($I$)
$U \leftarrow \{\langle i, j\rangle | I_{ij} > 0\}$; $\mathcal{F} \leftarrow \emptyset$
**while** $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) < \varepsilon$ **do**
  $s \leftarrow 0$
  **foreach** $\langle C, D\rangle \in \mathcal{G}$ **do**
    $J \leftarrow D^{\downarrow_I} \otimes C^{\uparrow_I}$; $F \leftarrow \emptyset$; $s_{\langle C,D\rangle} \leftarrow 0$
    **while** *exists* $\{{}^{a}/_{j}\} \in C^{\uparrow_I} \setminus F$ *s.t.* $cov(U, F \vee \{{}^{a}/_{j}\}, J) > s_{\langle C,D\rangle}$ **do**
      select $\{{}^{a}/_{j}\}$ maximizing $cov(U, F \vee \{{}^{a}/_{j}\}, J)$
      $F \leftarrow (F \vee \{{}^{a}/_{j}\})^{\downarrow_J \uparrow_J}$; $E \leftarrow (F \vee \{{}^{a}/_{j}\})^{\downarrow_J}$
      $s_{\langle C,D\rangle} \leftarrow cov(U, F, J)$
    **end**
    **if** $s_{\langle C,D\rangle} > s$ **then**
      $\langle E', F'\rangle \leftarrow \langle E, F\rangle$
      $\langle C', D'\rangle \leftarrow \langle C, D\rangle$
      $s \leftarrow s_{\langle C,D\rangle}$
    **end**
  **end**
  **add** $\langle E', F'\rangle$ **to** $\mathcal{F}$
  **remove** $\langle C', D'\rangle$ **from** $\mathcal{G}$
  **remove** from $U$ all $\langle i, j\rangle$ covered by $E' \otimes F'$ in $I$
**end**
**return** $\mathcal{F}$

**Algorithm 2:** ComputeIntervals.

**Input**: matrix $I$ with entries in $L$
**Output**: set $\mathcal{G} \subseteq \mathcal{B}(\mathcal{E}(I))$
1 $\mathcal{E} \leftarrow \mathcal{E}(I)$
2 $U \leftarrow \{\langle i, j\rangle | \mathcal{E}_{ij} > 0\}$
3 **while** $U$ *is non-empty* **do**
4   $D \leftarrow \emptyset$; $s \leftarrow 0$
5   **while** *exists* $\{{}^{a}/_{j}\} \in D$ *s.t.* $cov_I(U, D \vee \{{}^{a}/_{j}\}, \mathcal{E}) > s$ **do**
6     select $\{{}^{a}/_{j}\}$ maximizing $cov_I(U, D \vee \{{}^{a}/_{j}\}, \mathcal{E})$
7     $D \leftarrow (D \vee \{{}^{a}/_{j}\})^{\downarrow_{\mathcal{E}} \uparrow_{\mathcal{E}}}$; $C \leftarrow (D \vee \{{}^{a}/_{j}\})^{\downarrow_{\mathcal{E}}}$
8     $s \leftarrow cov_I(U, D, \mathcal{E})$
9   **end**
10   **add** $\langle C, D\rangle$ **to** $\mathcal{G}$
11   **remove** from $U$ entries $\langle i, j\rangle$ covered by $C^{\uparrow_I \downarrow_I} \otimes D^{\downarrow_I \uparrow_I}$ in $I$
12 **end**
13 **return** $\mathcal{G}$

to be covered by factors in $\mathcal{G}$ (l. 1 and 2). After each factor $\langle C, D\rangle$ is computed as described below, the entries covered by $\langle C, D\rangle$ are removed from $U$. To compute a new factor $\langle C, D\rangle$, a candidate fuzzy set $D$, from which the pair $\langle C, D\rangle$ is obtained, is initialized to the empty fuzzy set $\emptyset$. Then, one sequentially increases in the membership function of $D$ the most promising truth degree $a$ of the most promising attribute $j$ (l. 5–9), until such increase is impossible. This step represents a purely greedy strategy. In l. 7, one obtains from $D$ the corresponding formal concept in $\mathcal{B}(\mathcal{E}(I))$ by means of the closure operators induced by the matrix $\mathcal{E}(I)$. This formal concept is then added to $\mathcal{G}$ in l. 10. The entries covered by the corresponding rectangle $C^{\uparrow_I \downarrow_I} \otimes D^{\downarrow_I \uparrow_I}$ in $I$, namely entries $\langle i, j\rangle$ for which $(C^{\uparrow_I \downarrow_I} \otimes D^{\downarrow_I \uparrow_I})_{ij} > I_{ij}$, are then deleted from $U$, because we know that every formal concept in the interval $\mathcal{I}_{C,D}$ covers all these entries (see the previous paragraph). The greedy selection of formal concepts $\langle C, D\rangle$ is performed until $U$ is empty.

After $\mathcal{G}$ is computed by ComputeIntervals, GreEss$_L$ starts a search for formal concepts as factors of the input matrix $I$. The search, is performed in a greedy manner different from that utilized in ComputeIntervals, and proceeds in the intervals $\mathcal{I}_{C,D}$ for $\langle C, D\rangle \in \mathcal{G}$; see l. 3–21. For each $\mathcal{I}_{C,D}$, one picks the best formal concept in $\mathcal{I}_{C,D}$, i.e. the concept with the best coverage, in l. 6–11. The particular selection of this formal concept proceeds by initializing a candidate fuzzy set $F$ to the empty fuzzy set $\emptyset$ and then extending this $F$ by most promising attributes $j$ and truth degrees $a$. The variable $J$, which is initialized by $J \leftarrow D^{\downarrow_I} \otimes C^{\uparrow_I}$, represents a matrix that acts as a restriction to speed up the computation. This restriction guarantees that we do not leave the interval $\mathcal{I}_{C,D}$ during this search. The best formal concept, $\langle E', F'\rangle$, found this way in the intervals $\mathcal{I}_{C,D}$ is then inserted to the collection $\mathcal{F}$ of computed factors in l. 18. The respective interval $\mathcal{I}_{C',D'}$, from which the concept $\langle E', F'\rangle$ has been selected, is then deleted from the set $\mathcal{G}$ in l. 19. Therefore, the interval $\mathcal{I}_{C',D'}$ is no longer searched during the remaining iterations. The value $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$ is then updated accordingly. This selection of formal concepts $\langle E', F'\rangle$ is performed until the prescribed precision is reached, i.e. until $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \varepsilon$.

**Theorem 8.** GreEss$_L$ *is correct, i.e. it computes from the input matrix $I$ a collection $\mathcal{F}$ of formal concepts for which $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \varepsilon$.*

**Proof.** In above description of the algorithm, we have seen that GreEss$_L$ performs steps 1., 2., and 3. described at the beginning of this section. Now, observe that if $\varepsilon = 1$, condition $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \varepsilon$ becomes $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. Due to Theorem 5, conditions 1., 2. and 3. imply that in this case (i.e. $\varepsilon = 1$), $\mathcal{F}$ provides an exact decomposition of $I$, i.e. $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. Clearly, for $\varepsilon < 1$, the algorithm is guaranteed to deliver the required solution, i.e. a set $\mathcal{F}$ for which $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \varepsilon$: We have just shown that the algorithm is capable of attaining the stronger condition $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) = 1$, and thus the weaker condition shall mostly be attained even before processing all the intervals $\mathcal{I}_{C,D}$ of $\mathcal{B}(I)$ for $\langle C, D\rangle \in \mathcal{G}$.

**Example 5.** We now illustrate the algorithm GreEss$_L$ using the matrix $I$ from Example 1. The algorithm first calls ComputeIntervals, which computes the matrix $\mathcal{E}(I)$, see Example 3, and initializes $U$ by assigning to it non-zero entries in $\mathcal{E}(I)$. As described above, the factors of $\mathcal{G}$ are computed by selecting consecutively attribute $j$ and degree $a$ whose addition (i.e. attribute $j$ is added with degree $a$) to the intent $D$ of the so-far computed formal concept $\langle C, D\rangle$ in $\mathcal{B}(\mathcal{E}(I))$ maximizes the coverage of the still uncovered part of matrix $\mathcal{E}(I)$, until such extension is possible. This search starts with the closure $\{\}^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}}$ of empty set as intent, i.e. $D = \{\}^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}}$.

In our case, we thus start with $D = \{\}^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}} = \{\}$. Adding of attributes $j$ and degrees $a$ proceeds as follows. One first selects $j = x$ and $a = 0.5$. For this choice, we have $((D \vee \{^{0.5}/_x\})^{\downarrow_\mathcal{E}})^{\uparrow_\mathcal{I}\downarrow_\mathcal{I}} = \{^1/x, {}^{0.5}/y, {}^{0.5}/z\}$ and $((D \vee \{^{0.5}/_x\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}})^{\downarrow_\mathcal{I}\uparrow_\mathcal{I}} = \{^{0.5}/u, {}^1/v, {}^0/w\}$. The thus obtained concept $\langle C, D\rangle$ covers 3 so far uncovered entries in $\mathcal{E}(I)$. Not other addition of $j$ and $a$ results in a better coverage. For instance, selecting $j = x$ and $a = 1$ we obtain $((D \vee \{^1/_x\})^{\downarrow_\mathcal{E}})^{\uparrow_\mathcal{I}\downarrow_\mathcal{I}} = \{^{0.5}/x, {}^0/y, {}^0/z\}$ and $((D \vee \{^1/_x\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}})^{\downarrow_\mathcal{I}\uparrow_\mathcal{I}} = \{^1/u, {}^1/v, {}^{0.5}/w\}$, which concept covers only a single uncovered entry in $\mathcal{E}(I)$. We therefore select $j = x$ and $a = 0.5$, and put $C = (\{^{0.5}/_x\})^{\downarrow_\mathcal{E}} = \{^1/x, {}^{0.5}/y, {}^{0.5}/z\}$ and $D = (\{^{0.5}/_x\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}} = \{^{0.5}/u, {}^{0.5}/v, {}^0/w\}$, finishing the selection in l. 6–7 of ComputeIntervals. Next we try to extend the current $D$ by adding other $j$ and $a$ which improves coverage, as specified in l. 5. In our case, no further extension is possible, finishing thus the loop in l. 5–9. The selected concept $\langle C, D\rangle$ is hence added to $\mathcal{G}$, as specified in l. 10, and the 3 entries covered by $\langle C, D\rangle$ are removed from $U$.

Since $U$ is not yet empty, the loop in l. 3–12 continues by selecting next concept $\langle C, D\rangle$. As with the previous one, the algorithm starts with $D = \{\}^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}} = \{\}$ and seeks for the best $j$ and $a$. For the first possible choice, $j = x$ and $a = 0.5$, one obtains $((D \vee \{^{0.5}/_x\})^{\downarrow_\mathcal{E}})^{\uparrow_\mathcal{I}\downarrow_\mathcal{I}} = \{^1/x, {}^{0.5}/y, {}^{0.5}/z\}$ and $((D \vee \{^{0.5}/_x\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}})^{\downarrow_\mathcal{I}\uparrow_\mathcal{I}} = \{^{0.5}/u, {}^1/v, {}^0/w\}$. Clearly, since this concept was obtained already, it does not cover any uncovered entries. Proceeding with $j = x$ and $a = 1$, we get $((D \vee \{^1/_x\})^{\downarrow_\mathcal{E}})^{\uparrow_\mathcal{I}\downarrow_\mathcal{I}} = \{^{0.5}/x, {}^0/y, {}^0/z\}$ and $((D \vee \{^1/_x\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}})^{\downarrow_\mathcal{I}\uparrow_\mathcal{I}} = \{^1/u, {}^1/v, {}^{0.5}/w\}$. The corresponding concept does not cover any still uncovered entries either, one thus proceeds with $j = y$ and $a = 0.5$. Neither the resulting concept, which consists of $((\{^{0.5}/_y\})^{\downarrow_\mathcal{E}})^{\uparrow_\mathcal{I}\downarrow_\mathcal{I}} = \{^1/x, {}^1/y, {}^1/z\}$ and $((\{^{0.5}/_y\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}})^{\downarrow_\mathcal{I}\uparrow_\mathcal{I}} = \{^0/u, {}^{0.5}/v, {}^1/w\}$, nor the concept obtained for $j = y$ and $a = 1$, which consists of $((\{^1/_y\})^{\downarrow_\mathcal{E}})^{\uparrow_\mathcal{I}\downarrow_\mathcal{I}} = \{^1/x, {}^{0.5}/y, {}^{0.5}/z\}$ and $((\{^1/_y\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}})^{\downarrow_\mathcal{I}\uparrow_\mathcal{I}} = \{^{0.5}/u, {}^1/v, {}^0/w\}$, covers still uncovered entries. However, for $j = z$ and $a = 0.5$, we obtain $((\{^{0.5}/_z\})^{\downarrow_\mathcal{E}})^{\uparrow_\mathcal{I}\downarrow_\mathcal{I}} = \{^{0.5}/x, {}^1/y, {}^{0.5}/z\}$ and $((\{^{0.5}/_z\})^{\downarrow_\mathcal{E}\uparrow_\mathcal{E}})^{\downarrow_\mathcal{I}\uparrow_\mathcal{I}} = \{^0/u, {}^{0.5}/v, {}^{0.5}/w\}$, and the corresponding concept covers the last uncovered entry in $U$. For $j = z$ and $a = 1$, we do not get a better coverage, hence $j = z$ and $a = 0.5$ is selected and one puts $C = (\{^{0.5}/_z\})^{\downarrow\mathcal{E}} = \{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}$ and $D = (\{^{0.5}/_x\})^{\downarrow\mathcal{E}\uparrow\mathcal{E}} = \{^{0}/u, {}^{0}/v, {}^{0.5}/w\}$ in the loop in l. 6–7. This factor is then added to $\mathcal{G}$ in l. 10 and the last entry is removed from $U$. The set $\mathcal{G}$ returned by ComputeIntervals then contains two formal concepts of $\mathcal{B}(\mathcal{E}(I))$, namely $\langle\{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}, \{^{0.5}/u, {}^{0.5}/v\}\rangle$ and $\langle\{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}, \{^{0.5}/w\}\rangle$.

GreEss$_L$ then continues by initializing $U$ and $\mathcal{F}$ in l. 2. The loop in l. 3–21 proceeds by selecting the best, i.e. with largest coverage, concept from the intervals $\mathcal{I}_{\langle C,D\rangle}$ for $\langle C, D\rangle \in \mathcal{G}$, then deleting the interval from which the best concept was chosen from $\mathcal{G}$ (l. 19), selecting the next best concept from the remaining intervals, and so on until the condition in l. 3 is not met (in our case, until $U$ is empty i.e. all entries in $I$ are covered).

In our case, the selection thus proceeds as follows. The algorithm inspects both intervals $\mathcal{I}_{\langle C,D\rangle}$ corresponding to the two concepts $\langle C, D\rangle$ in $\mathcal{G}$ (l. 5). Let us first consider the interval $\mathcal{I}_{\langle C,D\rangle}$ corresponding to $\langle C, D\rangle = \langle\{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}, \{^{0.5}/u, {}^{0.5}/v\}\rangle$. The set $F$ is initially set to $F = \{\}$ (l. 6) and extended by adding to $F$ attribute $j$ in degree $a$ such that $\{^{a}/_j\} \in C^{\uparrow_I}$ and such that the resulting $F \vee \{^{a}/_j\}$ maximizes coverage of the still uncovered entries in $U$ (l. 8). Note that we have $C^{\uparrow_I} = \{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}^{\uparrow_I} = \{^{0.5}/u, {}^{1}/v\}$. By adding $\{^{0.5}/u\}$ to $F$, we obtain a concept corresponding to $(\{^{0.5}/_u\})^{\downarrow\mathcal{I}} = \{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}$ and $(\{^{0.5}/_u\})^{\downarrow\mathcal{I}\uparrow\mathcal{I}} = \{^{0.5}/u, {}^{1}/v, {}^{0}/w\}$, which covers 4 entries in $U$. By adding $\{^{1}/v\}$ do $F$, we obtain $(\{^{1}/_v\})^{\downarrow\mathcal{I}} = \{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}$ and $(\{^{1}/_v\})^{\downarrow\mathcal{I}\uparrow\mathcal{I}} = \{^{0.5}/u, {}^{1}/v, {}^{0}/w\}$, and the corresponding concept also covers 4 entries, i.e. not more than the previous one. We thus add $\{^{0.5}/u\}$ to $F$, i.e. we have $E = \{^{0.5}/u\}^{\downarrow\mathcal{I}} = \{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}$ and $F = E^{\uparrow_I} = \{^{0.5}/u, {}^{1}/v, {}^{0}/w\}$. Now, the present set $F$ cannot be extended by adding another by $\{^{a}/_j\} \in C^{\uparrow_I}$ (there is no attribute to add) and so the concept $\langle E, F\rangle$ is selected from the present interval $\mathcal{I}_{\langle C,D\rangle}$. Proceeding the same way for the second interval $\mathcal{I}_{\langle C,D\rangle}$, i.e. the one corresponding to $\langle C, D\rangle = \langle\{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}, \{^{0.5}/w\}\rangle$, we find out that the coverage of the best concept selected from $\mathcal{I}_{\langle C,D\rangle}$ is not better. The above concept $\langle E, F\rangle$, selected from the first interval, is therefore and added to $\mathcal{F}$ in l. 18. Then concept $\langle C, D\rangle$ is then removed from $\mathcal{G}$ (l. 19) and the entries covered by $\langle E, F\rangle$ are removed from $U$ (l. 20).

Since the condition in l. 3 is still true, we proceed the same way for the remaining concept in $\mathcal{G}$, namely $\langle\{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}, \{^{0.5}/w\}\rangle$. $F$ is initially set to $F = \{\}$. We have $C^{\uparrow_I} = \{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}^{\uparrow_I} = \{^{0.5}/v, {}^{0.5}/w\}$. By adding $\{^{0.5}/v\}$ do $F$, we obtain $(\{^{0.5}/_v\})^{\downarrow\mathcal{I}} = \{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}$ and $(\{^{0.5}/_v\})^{\downarrow\mathcal{I}\uparrow\mathcal{I}} = \{^{0}/u, {}^{0.5}/v, {}^{0.5}/w\}$. The corresponding concept covers 2 entries in $U$. By adding $\{^{0.5}/w\}$ do the set $F$, we obtain $(\{^{0.5}/_w\})^{\downarrow\mathcal{I}} = \{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}$ and $(\{^{0.5}/w\})^{\downarrow\mathcal{I}\uparrow\mathcal{I}} = \{^{0}/u, {}^{0.5}/v, {}^{0.5}/w\}$, and the corresponding concept also covers 2 entries. We thus add $\{^{0.5}/v\}$ to $F$, i.e. we have $E = \{^{0.5}/v\}^{\downarrow\mathcal{I}} = \{^{1}/x, {}^{1}/y, {}^{1}/z\}$ and $F = E^{\uparrow_I} = \{^{0}/u, {}^{0.5}/v, {}^{0}/w\}$. The set $F$ may still be extended by $\{^{0.5}/_w\} \in C^{\uparrow_I}$. Doing so, we obtain $(\{^{0.5}/_v, {}^{0.5}/_w\})^{\downarrow\mathcal{I}} = \{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}$ and $(\{^{0.5}/_v, {}^{0.5}/w\})^{\downarrow\mathcal{I}\uparrow\mathcal{I}} = \{^{0}/u, {}^{0.5}/v, {}^{0.5}/w\}$. Since this also covers 2 new entries, the algorithm chooses $E = \{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}$ and $F = \{^{0.5}/u, {}^{1}/v, {}^{0}/w\}$ from present interval and adds it to $\mathcal{F}$. Then the present concept $\langle C, D\rangle$ is removed from $\mathcal{G}$ and the entries covered by $\langle E, F\rangle$ are removed from $U$.

Now, $U$ is empty and the algorithm obtained the set $\mathcal{F} = \{\langle\{^{1}/x, {}^{0.5}/y, {}^{0.5}/z\}, \{^{0.5}/u, {}^{1}/v, {}^{0}/w\}\rangle, \langle\{^{0.5}/x, {}^{1}/y, {}^{0.5}/z\}, \{^{0}/u, {}^{0.5}/v, {}^{0.5}/w\}\rangle\}$ of factors of $I$.

**Remark 5.** We now derive an upper estimation of the worst-case time complexity of GreEss$_L$. For this purpose, let us consider the number $|L|$ of truth degrees in $L$ a constant, and assume that $\max(n, m) \leq \|I\|$, i.e. assume that the number $n$ of objects (rows) and the number $m$ of attributes (columns) are both smaller than or equal to the number $\|I\|$ of nonzero entries in $I$. Observe first that for $J \in L^{p\times q}$, computing the result of ${}^{\uparrow_J}$ and ${}^{\downarrow_J}$ takes time $O(pq)$. Computing $\mathcal{E}(I)$ may be done in time $O(n^2m^2)$ by computing first for each of the at most $nm$ non zero entries $\langle i, j\rangle$ the sets $\mathbf{I}_{ij}$ and comparing these sets to see which are minimal w.r.t. inclusion. Consider now ComputeIntervals. The loop in l. 3–12 repeats at most $\|\mathcal{E}(I)\| = O(\|I\|)$ times, since in the worst case, only one entry in $\mathcal{E}(I)$ gets covered in each execution of the loop. Most expensive in the loop is the while loop in l. 5–9. In this while loop, the most expensive operation is the computation of $(D \vee \{^{a}/j\})^{\downarrow\mathcal{E}\uparrow\mathcal{E}}$, which takes time $O(nm)$. The while loop in l. 5–9 repeats at most $O(m\|L\|) = O(m)$ times since at most $m|L|$ pairs $\{^{a}/j\}$ may be added to $D$. In each cycle of the while loop in l. 5–9, at most $O(m|L|) = O(m)$ attributes may be added to extend the constructed $D$, because again, at most $m|L|$ pairs $\{^{a}/j\}$ come into play. To sum up, the most expensive operation, $(D \vee \{^{a}/j\})^{\downarrow\mathcal{E}\uparrow\mathcal{E}}$, is executed at most $O(m) \cdot O(m) = O(m^2)$ times, which takes time $O(nm) \cdot O(m^2) = O(nm^3)$. Since the loop in l. 3–12 executes at most $O(\|I\|)$ times, this loop takes time $O(\|I\|nm^3)$. As we assume $\max(n, m) \leq \|I\|$, the total time of ComputeIntervals is $O(n^2m^2) + O(\|I\|nm^3) = O(\|I\|nm^3)$.

Table 2
Five most popular cat breeds.

| | Activity | Playfulness | Need for attention | Affection | Need to vocalize | Docility | Intelligence | Independence | Healthiness and hardiness | Grooming needs | Good with children | Good with other pets |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Persian | 2 | 7 | 9 | 9 | 3 | 11 | 5 | 4 | 4 | 11 | 9 | 9 |
| Maine Coon | 7 | 9 | 7 | 9 | 3 | 9 | 8 | 8 | 10 | 8 | 9 | 9 |
| Exotic Shorthair | 2 | 7 | 9 | 10 | 3 | 11 | 4 | 4 | 4 | 4 | 9 | 9 |
| Abyssinian | 11 | 10 | 9 | 8 | 3 | 2 | 8 | 7 | 5 | 4 | 5 | 6 |
| Siamese | 9 | 10 | 10 | 9 | 10 | 3 | 10 | 2 | 5 | 3 | 5 | 6 |

When ComputeIntervals is finished, GreEss$_L$ proceeds with at most $O(\|I\|)$ executions of the loop in l. 3–21 (indeed, only one entry in $I$ gets covered during each execution in the worst case). Within this loop, the loop in l. 5–17 is executed at most $|\mathcal{G}| = O(\|I\|)$ times. In the loop in l. 5–17, the construction of $\langle E, F\rangle$ by successive extension proceeds analogously as in ComputeIntervals; an analogous analysis shows that the construction takes time $O(nm^3)$. The loop in l. 3–21 thus takes $O(\|I\|^2 nm^3)$ steps. Putting this together with the time of ComputeIntervals, we conclude that GreEss$_L$ runs in time $O(\|I\|nm^3) + O(\|I\|^2 nm^3) = O(\|I\|^2 nm^3)$ in the worst case. Note however that our analysis is not tight. Obtaining a better upper estimate remains an interesting problem.

## 4. Experiments

In our experimental evaluation, we performed two kinds of experiments. The first consists in performing factor analysis of real data. Since it has been repeatedly demonstrated that factor analysis of matrices with grades using formal concepts yields well-interpretable, informative factors (see e.g. [5] for numerous examples), our aim is to show that our new algorithm retains this convenient property. We performed analyses of several datasets and, for space reasons, present in Section 4.1 a detailed illustrative example. The second kind of experiments, presented in Section 4.2, focuses on performance of our algorithm in terms of the number of factors needed to explain input data.

### *4.1. Factor analysis of real data*

The data in Table 2 describes 5 most popular cat breeds and their 12 attributes[1] (the full set of 43 breeds is analyzed below in this section). Since the original attributes take their values in the set $\{1, \ldots, 11\}$, we transformed them to fuzzy attributes over the 11-element chain $L = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$. We used the Łukasiewicz operations on $L$.

To provide a graphical representation of the data and its factorization, we represent the grades in $L$ by shades of gray as follows:


The input data is thus represented by the $5 \times 12$ object-attribute matrix $I$ over $L$ which is depicted along with its decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$ into the object-factor and factor-attribute matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ in Fig. 3. This decomposition was obtained using GreEss$_L$ and, therefore, it utilizes formal concepts in $\mathcal{B}(I)$ as factors.

Each factor $F_l$ is represented by the $l$th column and the $l$th row in the matrices $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$, respectively. Each entry $(A_{\mathcal{F}})_{il}$ indicates the extent to which the factor $F_l$ applies to breed $i$, while $(B_{\mathcal{F}})_{lj}$ represents the extent to which attribute $j$ is a particular manifestation (i.e., is typical) of factor $F_l$.

[1] http://www.petfinder.com/.

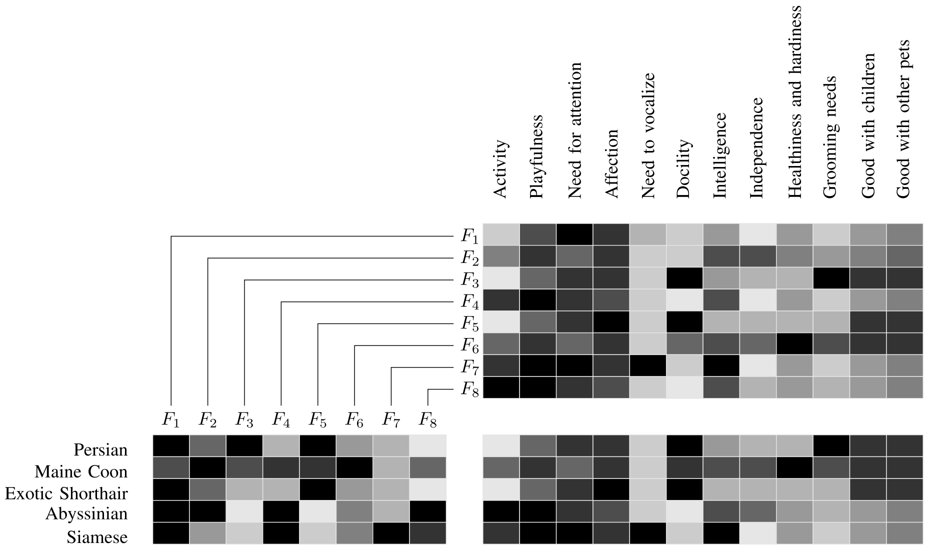


Fig. 3. Decomposition $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$. $I$, $A_{\mathcal{F}}$, and $B_{\mathcal{F}}$ are the bottom-right, bottom-left, and top matrix, respectively.

For instance, $F_3$ is manifested mainly by the attributes "docility", "grooming needs", "need for attention ", "affection", and "good with children and other pets." These are the attributes with high degrees in the third row of $B_{\mathcal{F}}$. In particular, the degrees to which these factors belong to (the intent of) $F_3$ are 1 for "docility" and "grooming needs," and 0.8 for "need for attention," "affection," and "good with children and other pets." The degrees for the other attributes are smaller, namely 0.6 for "playfulness," 0.4 for "intelligence," 0.3 for "independence" and "healthiness and hardiness," 0.2 for "need to vocalize," and 0.1 for "activity." Naturally, the factor may be termed *cats that need human*. $F_3$ applies in particular to Persian and Maine Coon, which is apparent from the third column of $A_{\mathcal{F}}$: the degrees to which $F_3$ applies to these two breeds are 1 and 0.7, while the degrees for Exotic Shorthair, Abyssinian, and Siamese are 0.3, 0.1, and 0.2, respectively.

In the same manner, one may interpret the other factors. In particular, $F_1$ is characterized mainly by the attributes "need for attention" (degree 0.9), "affection" (0.8), "playfulness" (0.7), and much less significantly by the other attributes, namely "good with other pets" (0.5), "healthiness and hardiness" (0.4), "good with children" (0.4), "need to vocalize" (0.3), "activity" (0.2), "docility" (0.2), "grooming needs" (0.2), and "independence" (0.1). The first column of $A_{\mathcal{F}}$ tells us that $F_1$ applies to Siamese (to degree 1), Persian, Exotic Shorthair, Abyssinian (all three to degree 0.9), and Maine Coon (0.7). As for $F_2$, this factor is characterized mainly by the attributes "playfulness" (to degree 0.8), "affection" (0.8), "intelligence" and 'independence" (both 0.7), and to smaller degrees by "need for attention," "good with other pets" (both 0.6), "activity," "healthiness and hardiness," "good with children," and "good with other pets" (all 0.5). The remaining attributes are characteristic to still smaller degrees for $F_2$. As for the breeds, $F_2$ applies to degree 1 to Maine Coon, to degree 0.9 to Abyssinian, which largely satisfy the characteristic attributes, to degree 0.6 to Persian and Exotic Shorthair, and to degree 0.4 to Siamese.

It is furthermore interesting to see that the first three factors (and thus the three most important ones), $F_1$, $F_2$, and $F_3$, explain, by and large, the whole data. Hence, the other factors, $F_4 \ldots, F_8$, may be neglected. Namely, denote by $A_{\mathcal{F}_3}$ and $B_{\mathcal{F}_3}$ the $5 \times 3$ and $3 \times 12$ matrices which are the parts of $A_{\mathcal{F}}$ and $B_{\mathcal{F}}$ corresponding to $F_1$, $F_2$, and $F_3$. The degree $s(I, A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3})$ of similarity of $I$ to $A_{\mathcal{F}_3} \circ B_{\mathcal{F}_3}$, which represents reconstructability of the original data $I$ from $F_1$, $F_2$, and $F_3$, equals 0.91. In accordance with the terminology used in the Boolean setting, we also say that 0.91 represents the coverage of the data by the first three factors.

The coverage of the data matrix $I$ by the set $\mathcal{F}_l = \{F_1, \ldots, F_l\}$ of the first $l$ factors, for $l = 1, \ldots k$, is shown in Fig. 4.

When analyzing the entire data ($43 \times 12$ input matrix $I$), GREESS$_L$ obtained 17 factors. These factors contain factors similar to those obtained from the smaller dataset described above. In fact, the factors from the five most popular cat breeds have the coverage of 0.94 of the extended dataset $I$ (the coverage in the extended data being assessed as in [5]). Moreover, among the factors obtained is a formal concept containing the attributes "playfulness",

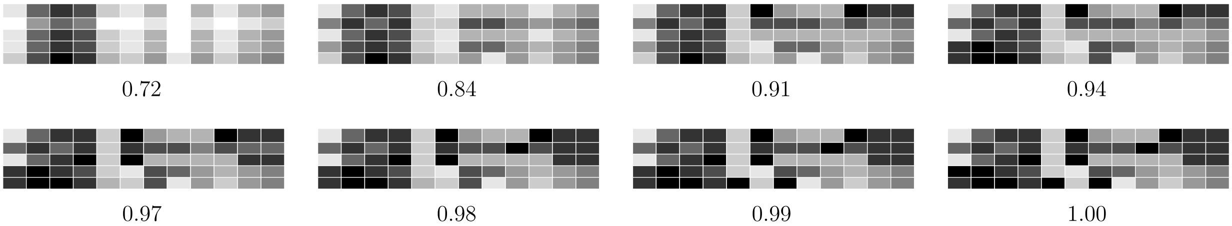


Fig. 4. Matrices $A_{\mathcal{F}_l} \circ B_{\mathcal{F}_l}$, representing the $\bigvee$-superposition of the set $\mathcal{F}_l$ of the first $l$ factors, $l = 1, \ldots, 8$, along with the coverage $s(I, A_{\mathcal{F}_l} \circ B_{\mathcal{F}_l})$ indicating the extent of the input data $I$ explained by the first $l$ factors.

Table 3
Real data.

| Dataset | Size | $\|L\|$ | $\|I\|$ | $\|\mathcal{E}(I)\|$ | $\|\mathcal{E}(I)\|/\|I\|$ |
|---|---|---|---|---|---|
| Dogs | 151 × 11 | 6 | 1963 | 362 | 0.184 |
| Decathlon | 28 × 10 | 5 | 266 | 59 | 0.221 |
| IPAQ | 4510 × 16 | 3 | 41624 | 1281 | 0.031 |
| Music | 900 × 26 | 7 | 20377 | 5952 | 0.292 |
| Rio | 87 × 31 | 4 | 402 | 332 | 0.820 |

"affection", and "intelligence" to degree 1 and "activity" and "need for attention" to a high degree. This factor may therefore be interpreted as *active and intelligent cats*. This factor applies to a high degree to the breeds Sphynx, Siamese, Javanese, and Balinese. Another factor obtained is a formal concept containing "good with children", "good with other pets", "affection", and "intelligence" to high degrees. This factor may be interpreted as *family cats* and applies e.g. to Sphynx, Siberian, Ragdoll, Manx, Cymric, Birman, American Curl, and American Shorthair.

Notice, at this point, the difference between providing a user, who supplies a data matrix $I$, with a set $\mathcal{F} \subseteq \mathcal{B}(I)$ of formal concepts of $I$, which provide an exact or approximate factorization of $I$, and providing the user with the set $\mathcal{B}(I)$ of all formal concepts of $I$. In the former case, the user is provided a small set of formal concepts fully or almost fully describing the input data in the sense of our factor model, while in the latter, the user is provided a set of concepts which also describe the input data in the sense of our factor model (namely, putting $\mathcal{F} = \mathcal{B}(I)$, one always has $I = A_{\mathcal{F}} \circ B_{\mathcal{F}}$; see [3]) but this set is usually very large and hence difficult to comprehend by the user.

### 4.2. *Quantitative evaluation of algorithms*

We now provide experimental evaluation of GREESS$_L$ on real and synthetic data. We observe the ability of extracted factors to explain input data. In particular, we measure this ability by the degree of similarity $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$, see (2), of the input matrix $I$ and the matrix $A_{\mathcal{F}} \circ B_{\mathcal{F}}$ obtained from the set $\mathcal{F}$ of extracted factors. The set $\mathcal{F}$ consists of the first $k$ factors extracted from $I$, where $k$ depends on the context. In view of Section 2.2.1, we also speak of $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$ as of the coverage of $I$ by factors in $\mathcal{F}$. In our evaluation, we compare GREESS$_L$ against the currently best algorithm to compute exact or almost exact decompositions of matrices with grades, namely the algorithm designed in [8], which was termed FIND-FACTORS in [8] and which we call GRECOND$_L$ (GREedy search for CONcepts on Demand) in this paper to give the algorithm a more particular name derived from the algorithm's strategy and to be compatible with our more recent terminology [6].

#### 4.2.1. *Real data*

The data we used are described in Table 3.

*Dog Breeds* is a data taken from http://www.petfinder.com/ which describes 151 popular breeds using 11 attributes regarding the breed characteristics such as "playfulness", "friendliness toward dogs", "friendliness toward strangers", "protection ability", "watchdog ability", "ease of training", and the like. For every breed and attribute, the data contains a grade out of a six-element set of grades, which indicates the extent to which the attribute applies to the particular breed. We therefore represented this data using a 151 × 11 matrix $I$ over a six-element set $L$, which we equipped with the Łukasiewicz operations. In a similar manner we represented by matrices $I$ over suitable sets $L$ the other datasets in Table 3. *Decathlon* is a $28 \times 10$ dataset from http://www.sports-reference.com/ describing the performance of 28 athletes in the 10 disciplines of decathlon in the 2004 Olympics. Note that a smaller portion of this data has been used in [8]. The data entries represent the actual performances of the athletes in the ten disciplines of decathlon and we represented these performances, as in [8], using a five-element set $L$. The *IPAQ* data comes from http://www.ipaq.ki.se/ and describes results of an international questionnaire examining physical activity of population. It involves 4510 respondents and 16 questions. The responses were made using a three-element set, hence we used $L = \{0, 0.5, 1\}$. This questionnaire is considered important from the health management point of view, particularly as a source for making government decisions regarding health policy, and includes questions regarding respondents' age, sex, sports activity, walking activity, health, body-mass-index (BMI), and the like. The *Music* data comes from an inquiry of how people perceive the speed of song [9]. In particular, the aim was to examine how the perception of speed of a given song depends on various features of the song. This data was obtained in [9] by questioning 30 participants. The participants were provided with 30 music samples—29 complex samples plus a simple tone of 528Hz. The participants then described their emotional experience by means of 26 attributes including "exciting", "restful", "happy", "pleasant", "intelligible", "ugly", "valuable", "interesting", "slow", "meaningful", "active", "violent", "strong","tense", "predictable", "closed", "known", "variable", or "like it". They used a six-element set to express the extent to which a given attribute applies to the given sample. They also evaluated a retrospective time duration and time passage. The core data in this study is thus represented by a $900 \times 26$ matrix with entries in a six-element set $L$. The *Rio* data is taken from https://www.rio2016.com/en/medal-count and is represented by a $87 \times 31$ matrix $I$ representing 87 countries that obtained a medal in one of 31 selected sports area (such as archery, athletics, badminton, basketball, or boxing) at the 2016 Olympics in Rio de Janeiro. The set $L$ contains four grades: $I_{ij} = 1$ means that country $i$ won at least one gold medal in the sports area $j$, $I_{ij} = \frac{2}{3}$ represents at least one silver medal, $\frac{1}{3}$ represents at least one bronze medal and 0 represents no medal in this sport. This dataset is very sparse in comparison with other presented datasets. A large portion of the input entries are essential: The ratio $\|\mathcal{E}(I)\|/\|I\|$ of the number of entries in $\mathcal{E}(I)$ to the corresponding number for $I$ is high.

In Table 3, $|L|$ denotes the number of grades (truth degrees) in the set $L$. To demonstrate the reduction in the number of non-zero entries due to computation of the essential part, we also observe the number $\|I\|$ of non-zero entries in the input matrix $I$, the number $\|\mathcal{E}(I)\|$ of non-zero entries in the essential part of $I$, and the reduction ratio $\|\mathcal{E}(I)\|/\|I\|$. As one can see, the reduction in the number of non-zero entries in the essential part $\mathcal{E}(I)$ compared to that of $I$ is significant which is an important fact in view of the results in Section 2.3.

As usual in quantitative assessment of factorization algorithms, we observed the numbers of factors needed to achieve a prescribed precision $\varepsilon$. That is, we observed the number $|\mathcal{F}|$ of factors in the set $\mathcal{F}$ produced by $\text{GreEss}_L$ and $\text{GreConD}_L$ which are sufficient for $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}}) \geq \varepsilon$. As we can see from the results in Table 4, $\text{GreEss}_L$ mostly outperforms $\text{GreConD}_L$ for high-precision decompositions, sometimes significantly. Note that the numbers of factors necessary to obtain exact decompositions tend to be higher than the number of the original attributes, which is a feature shared by all the available algorithms for data with grades as well as for Boolean data. Nevertheless, as is well known, such decompositions are valuable for two reasons. First, the obtained factors are nontrivial and informative compared to the original attributes. Secondly, the number of factors needed to achieve very precise, though not exact, decompositions is considerably smaller than the number of input variables, as e.g. demonstrated in our example with cat breeds in the previous section.

### *4.3. Synthetic data*

The synthetic data we used consists of matrices organized in five collections. Each collection, Set 1–5, comprises 500 randomly generated matrices $I$ whose characteristics are shown in Table 5: The dimension $n \times m$, the size $|L|$ of the set of grades, the inner dimension $k$, and the probability distribution $p$ of the grades. Each matrix $I$ in each Set $i$ is obtained by multiplication on an $n \times k$ and $k \times m$ matrices $A$ and $B$. The matrices $A$ and $B$ contain entries in the particular set $L$ and are randomly generated, following a distribution $p$ of grades. For example, in Set 1 we used a three-element set $L = \{0, \frac{1}{2}, 1\}$ with uniform probabilities $p(a)$ of the degrees used in $A$ and $B$, i.e. $p(0) = p(\frac{1}{2}) = p(1) = \frac{1}{3}$. These probabilities generalize the densities of Boolean matrices commonly used in experiments on Boolean matrix decomposition: For example for $L = \{0, 1\}$, the distribution $[\frac{1}{5}\ \frac{4}{5}]$ corresponds to density 0.8.

Table 4
Quality of decompositions (real data).

| Dataset | $s$ | Number of factors needed | |
|---|---|---|---|
| | | GRECOND$_L$ | GREESS$_L$ |
| Dogs | 0.85 | 5 | 7 |
| | 0.95 | 9 | 11 |
| | 1 | 16 | 15 |
| Decathlon | 0.85 | 4 | 5 |
| | 0.95 | 8 | 8 |
| | 1 | 15 | 10 |
| IPAQ | 0.85 | 12 | 12 |
| | 0.95 | 18 | 15 |
| | 1 | 32 | 17 |
| Music | 0.85 | 13 | 14 |
| | 0.95 | 24 | 25 |
| | 1 | 36 | 29 |
| Rio | 0.85 | 16 | 6 |
| | 0.95 | 24 | 17 |
| | 1 | 35 | 32 |

Table 5
Synthetic data.

| Dataset | Size | $\|L\|$ | $k$ | Distribution on $L$ in $A$ and $B$ |
|---|---|---|---|---|
| Set 1 | 50×50 | 3 | 10 | $[\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}]$ |
| Set 2 | 50×50 | 5 | 10 | $[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$ |
| Set 3 | 100×50 | 5 | 25 | $[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$ |
| Set 4 | 100×100 | 5 | 20 | $[\frac{1}{8}\ \frac{1}{8}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$ |
| Set 5 | 500×100 | 6 | 25 | $[\frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}]$ |

Table 6
Characteristics of synthetic data.

| Dataset | Avg $\\|I\\|$ | Avg $\\|\mathcal{E}(I)\\|$ | avg $\\|\mathcal{E}(I)\\|/\\|I\\|$ |
|---|---|---|---|
| Set 1 | 2449 | 195 | 0.080 |
| Set 2 | 2503 | 355 | 0.141 |
| Set 3 | 4983 | 602 | 0.121 |
| Set 4 | 10000 | 2087 | 0.209 |
| Set 5 | 49997 | 14216 | 0.284 |

The average characteristics of the matrices in Set 1–5, which have the same meaning as for real data, are shown in Table 6. One may again observe that the reduction in number of nonzero entries, i.e. the reduction ratio $\|\mathcal{E}(I)\|/\|I\|$, is significant as in the case of real data.

The results comparing GREESS$_L$ and GRECOND$_L$ are displayed in Table 7. In the rows, we provide for selected numbers $k$ of factors their coverage of input data, i.e. the number $s(I, A_{\mathcal{F}} \circ B_{\mathcal{F}})$ achieved by the set $\mathcal{F}$ of the first $k$ factors computed by the algorithms. As with real data, we may observe that GREESS$_L$ tends to outperform GRECOND$_L$, particularly when observing higher precision decompositions. Moreover, we can also see that for the first few factors, GRECOND$_L$ tends to achieve higher coverage than GREESS$_L$, which is due to the fact that GRECOND$_L$ follows a purely greedy strategy. That is to say, the purely greedy strategy GRECOND$_L$ wins at the beginning, but as the factorization proceeds, the more sophisticated approach of GREESS$_L$, which concentrates on the essential entries, tends to perform better.

Table 7
Quality of decompositions (synthetic data).

| Dataset | $k$ | Coverage $s$ by the first $k$ factors | |
|---|---|---|---|
| | | GRECOND$_L$ | GREESS$_L$ |
| Set 1 | 1 | 0.576 | 0.525 |
| | 4 | 0.866 | 0.866 |
| | 11 | 0.992 | 1 |
| | 17 | 1 | – |
| Set 2 | 1 | 0.620 | 0.632 |
| | 2 | 0.782 | 0.820 |
| | 10 | 0.995 | 1 |
| | 13 | 1 | – |
| Set 3 | 1 | 0.684 | 0.728 |
| | 3 | 0.828 | 0.790 |
| | 19 | 0.966 | 0.979 |
| | 27 | 0.986 | 0.998 |
| | 39 | 0.998 | 1 |
| | 47 | 1 | – |
| Set 4 | 1 | 0.651 | 0.648 |
| | 4 | 0.827 | 0.854 |
| | 21 | 0.975 | 0.994 |
| | 27 | 0.998 | 1 |
| | 29 | 1/1 | – |
| Set 5 | 1 | 0.511 | 0.477 |
| | 5 | 0.821 | 0.798 |
| | 27 | 0.995 | 0.995 |
| | 36 | 0.999 | 1 |
| | 42 | 1 | – |

## 5. Conclusions

The contributions of our paper concern two aspects of the problem of factorization of matrices with grades. The first concerns theoretical insight regarding the problem. Such insight is largely needed because the existing factorization algorithms including those for Boolean matrices are mostly based on various greedy strategies and utilize only a very limited insight. Our results concern geometry of factorizations and make it possible to identify entries in the input matrix to which one may focus in computing factorizations. The second aspect concerns algorithms. We propose a new factorization algorithm that is largely based on the theoretical insight. By means of experimental evaluation, we demonstrated that the algorithm delivers informative and easily interpretable factors and that it outperforms the available algorithms as far as high-precision factorizations are concerned.

As far as future research is concerned, we believe the following topics are in need of further exploration. Most needed in our view is further advancement of theoretical results on matrices with grades, in particular those regarding factorizations. In a sense, our algorithm may be considered as an example demonstrating usefulness of such results. Naturally related to this topic is design of novel factorization algorithms for matrices with grades. The various algorithms developed for Boolean matrices may serve as an inspiration. Our experience shows that development of various case studies in factor analysis of real data play an important role, in particular it serves as an indispensable feedback in algorithms' design. Let us note that such studies are largely missing even in the Boolean case. As far as more detailed technical issues are concerned, of particular importance are problems in the setting with grades that are hidden in the Boolean case. An example is an appropriate choice of logical connectives. In the Boolean case, this problem is vacuous because the classical logical connectives have no alternative in the Boolean case. In the setting with grades, a number of logical connectives is available. Even though our results hold for any choice of connectives, the question of whether further analysis or other considerations may help in guiding a user in the selection of logical operations remains open. Last but not least, the choice of a proper matrix similarity function, for which we used (2), and its employment in the strategy to select factors is another problem to be looked at. Namely, the function is used to measure the extent to which the factors explain the data and should therefore naturally correspond to the intuitive assessment of this extent.

## Acknowledgement

Supported by grant No. 15-17899S of the Czech Science Foundation. Preliminary results in this paper were presented in Belohlavek, R., Krmelova M., Proc. ICDM 2013, pp. 961–966. R. Belohlavek acknowledges support by the ECOP (Education for Competitiveness Operational Programme) project no. CZ.1.07/2.3.00/20.0059, which is co-financed by the European Social Fund and the state budget of the Czech Republic. The present work was done during the sustainability period of this project.

## References

[1] R. Belohlavek, Fuzzy Relational Systems: Foundations and Principles, Kluwer, New York, 2002.
[2] R. Belohlavek, Concept lattices and order in fuzzy logic, Ann. Pure Appl. Log. 128 (1–3) (2004) 277–298.
[3] R. Belohlavek, Optimal decompositions of matrices with entries from residuated lattices, J. Log. Comput. 22 (2012) 1405–1425.
[4] R. Belohlavek, M. Krmelova, The discrete basis problem and Asso algorithm for fuzzy attributes, 2017, submitted for publication.
[5] R. Belohlavek, M. Krmelova, Factor analysis of ordinal data via decomposition of matrices with grades, Ann. Math. Artif. Intell. 72 (2014) 23–44.
[6] R. Belohlavek, M. Trnecka, From-below approximations in Boolean matrix factorization: geometry and new algorithm, J. Comput. Syst. Sci. 81 (2015) 1678–1697.
[7] R. Belohlavek, V. Vychodil, Discovery of optimal factors in binary data via a novel method of matrix decomposition, J. Comput. Syst. Sci. 76 (1) (2010) 3–20.
[8] R. Belohlavek, V. Vychodil, Factorization of matrices with grades, Fuzzy Sets Syst. 292 (1) (2016) 85–97 (preliminary version in LNAI 5548 (2009), 83–97).
[9] K. Flaska, P. Cakirpaloglu, Identification of the multidimensional model of subjective time experience, in: Int. Studies in Time Perspective, Imprensa da Universidade de Coimbra, 2013, pp. 259–273.
[10] B. Ganter, R. Wille, Formal Concept Analysis. Mathematical Foundations, Springer, Berlin, 1999.
[11] S. Gottwald, A Treatise on Many-Valued Logics, Research Studies Press, Baldock, Hertfordshire, England, 2001.
[12] P. Miettinen, T. Mielikäinen, A. Gionis, G. Das, H. Mannila, The discrete basis problem, IEEE Trans. Knowl. Data Eng. 20 (2008) 1348–1362.
[13] L. Stockmeyer, The Set Basis Problem is NP-Complete, Tech. Rep. RC5431, IBM, Yorktown Heights, NY, USA, 1975.
[14] N. Tatti, T. Mielikäinen, A. Gionis, H. Mannila, What is the dimension of your binary data?, in: Proc. IEEE ICDM, 2006, pp. 603–612.