Self-Correlation and Maximum Independence
in Finite Relations

Dilian Gurov

KTH Royal Institute of Technology, Stockholm, Sweden

`dilian@csc.kth.se`

Minko Markov

"St. Kliment Ohridski" University of Sofia, Sofia, Bulgaria

`minkom@fmi.uni-sofia.bg`

We consider relations with no order on their attributes as in Database Theory. An independent partition of the set of attributes S of a finite relation R is any partition $\mathfrak{X}$ of S such that the join of the projections of R over the elements of $\mathfrak{X}$ yields R. Identifying independent partitions has many applications and corresponds conceptually to revealing orthogonality between sets of dimensions in multidimensional point spaces. A subset of S is termed self-correlated if there is a value of each of its attributes such that no tuple of R contains all those values. This paper uncovers a connection between independence and self-correlation, showing that the maximum independent partition is the least fixed point of a certain inflationary transformer α that operates on the finite lattice of partitions of S. α is defined via the minimal self-correlated subsets of S. We use some additional properties of α to show the said fixed point is still the limit of the standard approximation sequence, just as in Kleene's well-known fixed point theorem for continuous functions.

1 Introduction

The problem of discovering independence between sets of points in a multidimensional space is a fundamental problem in science. It arises naturally in many areas of Computer Science. For instance, with respect to relational data, discovering such independence allows exponential gains in storage space and processing of information [11], [1], and can facilitate the problem of machine learning [13]. With respect to problem clusterisation of multidimensional relational data, finding independence helps finding the desired clusters [5], [8]. Decomposing data into smaller units that are independent except at their interfaces has been known to be essential for understanding large legacy systems [17]. Independence has also been the subject of recent works in logic, giving rise to so-called logics of dependence and independence [4].

The concrete motivation for the present work derives from the area of *software product line engineering*, a discipline that aims at planning for and developing a *family* of products through managed reuse in order to decrease time to market and improve software quality [12]. A software family can be modelled as a relation whose attributes are the software's functionalities. The various implementations of each functionality in the form of software artefacts are the attributes' *values*. The individual products of a family are thus modelled as the tuples of that relation over the attributes. In previous works [6, 15] we considered a restricted class of software families called *simple families* (later on we changed the term "families" to the more abstract term "relations"), where discovery of independence and a compositional model checking technique are utilised to derive a *divide-and-conquer verification strategy*. Simple relations constitute the least class that contains the single-attribute, single-value relations and is closed under join of relations with disjoint attribute sets and unions of relations over the same set of attribute names but with disjoint value sets. In the present work we generalise these previous results to discovering independence in arbitrary relations. We investigate decompositions of a relation R with disjoint attributes such that R equals the join of the component relations. Every decomposition is represented by a partition of the set of attributes of R. Such partitions are termed *independent partitions*.

The problem of computing a maximum decomposition of this kind has previously been studied in [10], where it is referred to as *prime factorisation*, and an efficient algorithmic solution is proposed. In this paper we investigate an alternative approach that works purely on the level of the attributes of R and is based on the concept of *correlation* between attributes. We have discovered a nontrivial connection between independence and correlation and the major goal of this paper is to demonstrate that connection.

A first observation is that the decomposition problem cannot be solved purely based on analysis of pairs of attributes. In the aforementioned work [6] we compute dependence (or independence) in simple relations by computing correlation between pairs of attributes. That approach does not generalise for arbitrary relations as we show in this paper. Our solution is to introduce *self-correlation* of sets (of arbitrary cardinality) of attributes. In other words, the current notion of correlation is a hypergraph whose hyperedges are the self-correlated sets, rather than an ordinary graph as were the case with the simple relations. Since self-correlated sets are upward closed under set inclusion (Proposition 2), the minimal self-correlated sets, or the *mincors* (Definition 4), are the foundation of our analysis. A second observation is that mincors do not cross independent partitions (Lemma 5), hence one can safely merge overlapping mincors to compute the maximum independent partition. In the case of simple relations that merger indeed yields the maximum independent partition [6] but in arbitrary relations merging the mincors *does not* necessarily output an independent partition, as the example on page 67 shows. We overcome this hindrance with the help of a final important insight. Let $\mathfrak{X}$ be the partition of the set of attributes that results from merging overlapping mincors. The relation can be factored on $\mathfrak{X}$, producing a *quotient relation*. In other words, the elements of $\mathfrak{X}$ are considered atomic now; the subsets of $\mathfrak{X}$ may or may not be self-correlated in their turn, and the said quotient relation is defined via those new mincors. We show that the procedure of identifying mincors and merging overlapping ones can be repeated on this quotient relation and this can be iterated until stabilisation, yielding the desired maximum independent partition.

The above insights suggest that relational decomposition can be presented in terms of a transformer over the finite lattice of quotient relations, or conceptually even simpler, over *the lattice of the partitions* ordered by refinement, inducing the former lattice. The transformer α on partitions introduced here essentially corresponds to identifying the mincors of the quotient relation induced by a partition, merging the overlapping ones, and extracting from the result the corresponding partition (Definition 5). We prove that the independent partitions correspond exactly to the fixed points of α (Theorem 1).

If α is monotone, one can utilise two well-known fixed point theorems on complete lattices (having in mind that monotone functions over finite lattices are continuous). First, by Tarski's fixed point theorem for complete lattices [16], the set of fixed points forms a lattice itself with respect to the same ordering, hence there is a unique *least fixed point* (LFP), which in our case would be precisely the maximum independent partitioning that we are after. And second, one can utilise Kleene's fixed point theorem [7], to the effect that the LFP can be computed *iteratively*, starting from the bottom of the lattice, *i.e.* the partition into singletons, and applying α until stabilisation, *i.e.*, until the fixed point is reached. It turns out, however, that α in general is *not monotone* as demonstrated by the example on page 70 and therefore the above reasoning is not applicable.

On the other hand, we show that α is *inflationary* (Proposition 4). The existence of a LFP is established by showing that there exists a fixed point and the set of all fixed points is closed under intersection (Lemma 6). Furthermore, the downward closure of LFP, *i.e.*, the set of all partitions refining it, is closed under α (Lemma 8). Since the lattice is finite, these results give rise to a modified version of Kleene's fixed point theorem—formulated in terms of inflationary transformers rather than monotone ones (Theorem 2)—justifying the same iterative fixed point computation procedure (Corollary 3). The proposed characterisation reduces relational decomposition to the problem of identifying the mincors of a relation.

Organisation   The paper is organised as follows. Section 2 recalls some known notions and results about sets and families, partitions, lattices, fixed points, relations, attributes, and relation schemes, quotient relations, and defines independent partitions of the attributes set. Section 3 develops the theory of self-correlated sets in quotient relations and how they relate w.r.t. partition abstraction. Section 4 presents many useful lemmas that concern independence. Section 5 defines the transformer α and contains our main result, Theorem 2. Section 6 discusses what we currently know about the area of decomposition of relations, also called factorisation of relations, and compares the approach and the results of this paper with similar works. The final Section 7 draws some conclusions and outlines directions for future work.

2   Background

In this section we recall some standard set-theoretical notions and notation needed for our theoretical developments.

2.1   Sets, covers, and partitions

In this work we consider only finite sets. The powerset of a set A is denoted by $\mathsf{POW}(A)$ and $\mathsf{P}^+(A)$ denotes $\mathsf{POW}(A) \setminus \{\emptyset\}$. *Ground sets* are nonempty sets over which we construct the families that are our subject of research.

Let A be a ground set. *A family over A* is any nonempty subset of $\mathsf{P}^+(A)$. A family F is *Sperner family* if $\forall X, Y \in \mathrm{F} : X \nsubseteq Y$. F is *connected* if $\forall X, Z \in \mathrm{F}\colon X \cap Z \neq \emptyset$ or F has elements $Y_1, Y_2, \ldots, Y_k$ for some $k \geq 1$, such that $X \cap Y_1 \neq \emptyset$, $Y_i \cap Y_{i+1} \neq \emptyset$ for $1 \leq i \leq k-1$, and $Y_k \cap Z \neq \emptyset$. *A connected component of a family* is any maximal connected subfamily in it. We use $\mathsf{CC}(\mathrm{F})$ to denote the family $\{\cup B \,|\, B$ is a connected component of F$\}$. *A superfamily over A* is any nonempty subset of $\mathsf{P}^+(\mathsf{P}^+(A))$.

Suppose A is a set. *A cover of A* is any family F over A such that $\cup \mathrm{F} = A$. The set of all covers of A is denoted by $K(A)$. If $\mathfrak{X} \in K(A)$ and $Y \cap Z = \emptyset$ for all distinct $Y, Z \in \mathfrak{X}$, we say $\mathfrak{X}$ is *a partition of A*. If $|\mathfrak{X}| = 1$ the partition is *trivial* and if $|\mathfrak{X}| = |A|$ the partition is *partition into singletons*. Note that $\mathsf{CC}(\mathrm{F})$ defined above is a partition of the ground set. We denote by $\mathfrak{Y} \Subset \mathfrak{X}$ the fact that for some $B \subseteq A$, $\mathfrak{Y}$ is a family over B such that every element of $\mathfrak{Y}$ is a subset of precisely one element of $\mathfrak{X}$ and every element of $\mathfrak{X}$ is a superset of at most one element of $\mathfrak{Y}$. For example, if $A = \{a, b, c, d, e, f, g, h, k\}$ then $\{\{b\}, \{c\}, \{d, g\}\} \Subset \{\{a, b\}, \{c\}, \{d, e, f, g\}, \{h, k\}\}$.

The set of all partitions of A is denoted by $\Pi(A)$. For any $P_1, P_2 \in \Pi(A)$, P_1 *refines* P_2, which we denote by $P_1 \sqsubseteq P_2$, if

$$\forall X \in P_1 \ \exists Y \in P_2 : X \subseteq Y$$

Conversely, we say that P_2 *abstracts* P_1. If $P_1 \sqsubseteq P_2$ and $P_1 \neq P_2$ we write $P_1 \sqsubset P_2$.

2.2   Partial orders, lattices, and chains

We denote generic partial orders by "$\preccurlyeq$". If $(A, \preccurlyeq)$ is a poset, *a least element* of A is any $x \in A$ such that $\forall y \in A : x \preccurlyeq y$ and *a greatest element* of A is any $x \in A$ such that $\forall y \in A : y \preccurlyeq x$. A least element may not exist but if it exists it is unique; the same holds for a greatest element. The least element is called *bottom* and is denoted by $\bot$. The greatest element is called *top* and is denoted by $\top$. *A chain* in a poset $(A, \preccurlyeq)$ is any $B \subseteq A$ such that $\forall x, y \in B : x \preccurlyeq y \lor y \preccurlyeq x$.

A lattice is a poset $(A, \preccurlyeq)$, shortly A when $\preccurlyeq$ is understood, such that for any $x, y \in A$ there exists a (unique) greatest lower bound in A called *meet* and denoted by $x \sqcap y$ and a (unique) least upper bound

in A called *join* and denoted by $x \sqcup y$. Collectively, $\sqcap$ and $\sqcup$ are *the lattice operations of A*. They are commutative and associative [2, pp. 8]. We generalise the lattice operations on subsets of A in the obvious way. *A complete lattice* is a lattice such that every $B \subseteq A$ has a meet $\sqcap B$ and a join $\sqcup B$. In particular, A has a meet $\sqcap A = \bot$ and a join $\sqcup A = \top$. Every finite lattice is complete [3, pp. 46], therefore from now on by lattice we mean complete lattice. For any $x \in A$, the sets $\{y \in A \,|\, y \preccurlyeq x\}$ and $\{y \in A \,|\, x \preccurlyeq y\}$ are called *down-x* and *up-x* and are denoted by $\uparrow x$ and $\downarrow x$, respectively [3, pp. 20].

It is well-known that $(\Pi(A), \sqsubseteq)$ is a lattice. Furthermore, $\bot$ is the partition into singletons, $\top$ is the trivial partition, and for any $P_1, P_2 \in \Pi(A)$, $P_1 \sqcap P_2 = \{X \cap Y \,|\, X \in P_1, Y \in P_2\} \setminus \{\emptyset\}$ and $P_1 \sqcup P_2 = \mathbb{CC}(P_1 \cup P_2)$ (see [2, pp. 15]). We extend the "$\sqcap$" notation to subsets of partitions: for any $\mathfrak{X}, \mathfrak{Y} \in \Pi(A)$, for any nonempty $\mathfrak{X}' \subseteq \mathfrak{X}$ and any nonempty $\mathfrak{Y}' \subseteq \mathfrak{Y}$ such that $\mathfrak{X}' \cap \mathfrak{Y}' \neq \emptyset$, $\mathfrak{X}' \sqcap \mathfrak{Y}'$ denotes the set $\{B \cap C \,|\, B \in \mathfrak{X}', C \in \mathfrak{Y}'\} \setminus \{\emptyset\}$.

2.3 Functions and fixed points

Suppose A is a set and $f : A \to A$ is a function. For every $x \in A$: $f^0(x) \stackrel{\text{def}}{=} x$ and for every $n \in \mathbb{N}^+$, $f^n(x) \stackrel{\text{def}}{=} f \circ f^{n-1}(x)$. For every $n \in \mathbb{N}$, $f^n(x)$ is *the n-th iterate of f*. *A fixed point* of f is every $x \in A$ such that $f(x) = x$. Let $(A, \preccurlyeq)$ be a poset. A function $f : A \to A$ is *monotone* if $\forall x, y \in A : x \preccurlyeq y \to f(x) \preccurlyeq f(y)$ and f is *inflationary* if $\forall x \in A : x \preccurlyeq f(x)$ [14, pp. 263].

A well-known fixed point theorem is Tarski's fixed point theorem for continuous functions over complete lattices [16], stating that the set of fixed points is non-empty and forms a lattice itself with respect to the same ordering, and hence the function has a unique *least fixed point* (LFP). Another well-known theorem due to Kleene states the existence of an LFP for continuous functions on chain-complete partial orders [7], and that the LFP can be computed *iteratively*, starting from the bottom of the lattice and applying the function until stabilisation.

2.4 Schemes, relations, and quotient relations

The following definitions are close to the ones in [9]. *A scheme* is a nonempty set $S = \{A_1, \ldots, A_n\}$ whose elements, called *the attributes*, are nonempty sets. For every attribute, its elements are said to be its *values*. *A relation over S* is a nonempty set of total functions $\{t_1, t_2, \ldots, t_p\}$, which we call *the tuples*, such that for $1 \leq j \leq p$, $t_j : S \to \cup S$, with the restriction that $t_j(A_i) \in A_i$, for $1 \leq i \leq n$. We assume that every value of every attribute occurs in at least one tuple.

The relations we have in mind are as in Relational Database Theory, *i.e.* with unordered tuples, rather than as in Set Theory, *i.e.* with ordered tuples.

We further postulate that the said attributes are mutually disjoint sets. That allows a simplification of the definition of relation: a relation over S is nonempty set of tuples, each tuple being an n-element set with precisely one element from every attribute. To save space, we often write the tuples without commas between their elements. For example, let $n = 3$, $A_1 = \{a_1, a_2\}$, $A_2 = \{b_1, b_2\}$, and $A_3 = \{c_1, c_2, c_3\}$. One of the relations over the scheme $\{A_1, A_2, A_3\}$ is written as $\{\{a_1 b_1 c_1\}, \{a_1 b_2 c_2\}, \{a_2 b_2 c_3\}\}$.

Let $S_1, S_2, \ldots, S_k$ be schemes such that for $1 \leq i < j \leq k$, $\forall A \in S_i \; \forall B \in S_j : A \cap B = \emptyset$. Let R_i be a relation over S_i, for $1 \leq i \leq k$. *The join of R_1, ..., R_k* is the relation

$$R_1 \bowtie R_2 \bowtie \cdots \bowtie R_k = \{\cup\{x_1, x_2, \ldots, x_k\} \,|\, x_1 \in R_1, x_2 \in R_2, \ldots, x_k \in R_k\}$$

The complete relation over $S = \{A_1, \ldots, A_n\}$ is $\bowtie_{i=1}^n \{\{x\} \,|\, x \in A_i\}$. Clearly, its cardinality is $\prod_{i=1}^n |A_i|$.

Let $S = \{A_1, \ldots, A_n\}$ be a scheme. *A subscheme of* S is any nonempty subset of S. The notation $f|_Z$ stands for the restriction of f to Z, for any function $f : X \to Y$ and any $Z \subseteq X$. Let $R = \{t_1, t_2, \ldots, t_p\}$ be a relation over S and let T be a subscheme of S. *The projection of R on* T *is* $R \upharpoonright T = \{t_j|_T : 1 \leq j \leq p\}$.

Definition 1 (quotient relation) *Let R be a relation over some scheme* S. *For any* $\mathfrak{X} = \{X_1, X_2, \ldots, X_n\} \in \Pi(S)$, $R/\mathfrak{X} \subseteq \bowtie_{i=1}^n (R \upharpoonright X_i)$ *is the following relation:*

$$\forall \{y_1 y_2 \ldots y_n\} \in \bowtie_{i=1}^n (R \upharpoonright X_i) :$$
$$\{y_1 y_2 \ldots y_n\} \in R/\mathfrak{X} \text{ iff } \exists t \in R \, \forall i_{1 \leq i \leq n} (t \upharpoonright X_i = y_i)$$

We term $R/\mathfrak{X}$ *the* quotient relation of R relative to $\mathfrak{X}$. *When* $\mathfrak{X}$ *is understood we say simply* the quotient relation of R.

We emphasise the quotient relation is not over S but over a partition of S.

Here is an example of a quotient relation. Let $S = \{A, B, C, D\}$, let each attribute have precisely two values, say $A = \{a_1, a_2\}$ and so on, let $\mathfrak{X}_1 = \{\{A, B\}, \{C, D\}\}$, let $\mathfrak{X}_2 = \{\{A\}, \{B\}, \{C\}, \{D\}\}$, and let

$$R' = \{\{a_1 b_1 c_1 d_1\}, \{a_1 b_1 c_2 d_2\}, \{a_1 b_2 c_1 d_2\}, \{a_2 b_2 c_1 d_1\}, \{a_2 b_2 c_2 d_2\}\} \tag{1}$$

be a relation over S. Then

$$R'/\mathfrak{X}_1 = \{\{\{a_1, b_1\}\{c_1, d_1\}\}, \{\{a_1, b_1\}\{c_2, d_2\}\}, \{\{a_1, b_2\}\{c_1, d_2\}\},$$
$$\{\{a_2, b_2\}\{c_1, d_1\}\}, \{\{a_2, b_2\}\{c_2, d_2\}\}\} \tag{2}$$
$$R'/\mathfrak{X}_2 = \{\{\{a_1\}\{b_1\}\{c_1\}\{d_1\}\}, \{\{a_1\}\{b_1\}\{c_2\}\{d_2\}\}, \{\{a_1\}\{b_2\}\{c_1\}\{d_2\}\},$$
$$\{\{a_2\}\{b_2\}\{c_1\}\{d_1\}\}, \{\{a_2\}\{b_2\}\{c_2\}\{d_2\}\}\} \tag{3}$$

A quotient relation is but a grouping together of the tuples of the original relation into subtuples according to the partition. It trivially follows that $|R/\mathfrak{X}| = |R|$ for any relation R over any attribute set S and any $\mathfrak{X} \in \Pi(S)$.

2.5   Independent partitions

For a given relation *R* over some scheme *S*, we are after decompositions of *R* such that *R* equals the join of the obtained components. Each decomposition of this kind corresponds to a certain partition of *S*.

Definition 2 (independent partition) *Let R be a relation over some scheme S. For any* $\mathfrak{X} \in \Pi(S)$, $\mathfrak{X}$ *is an independent partition of S with respect to R if* $R = \bowtie_{Y \in \mathfrak{X}} R \upharpoonright Y$. *The set of all independent partitions of* S *with respect to R is denoted by* $\amalg_R(S)$, *or shortly* $\amalg(S)$ *if R is understood. If a partition is not independent, it is* dependent.

Note that $\amalg(S)$ is nonempty since it necessarily contains the trivial partition.

Proposition 1 *For every independent partition* $\mathfrak{X}$, $R/\mathfrak{X}$ *is the complete relation over* $\mathfrak{X}$.

Informally speaking, the object of the present study is the independent partition with the maximum number of equivalence classes, provided it is unique.

3   Correlation in Relations

In this section we define correlation in relations and quotient relations. From now on assume an arbitrary but fixed scheme S and relation *R* over it.

3.1 Correlated subsets of ground sets

In this subsection, the ground sets are schemes.

Definition 3 (correlated subsets of schemes) *Let* $S = \{A_1, A_2, \ldots, A_n\}$ *and let* T *be some nonempty sub-scheme* $\{A_{i_1}, A_{i_2}, \ldots, A_{i_m}\}$ *where* $1 \le i_1 < i_2 < \cdots < i_m \le n$. T *is* self-correlated *with respect to R, or shortly* correlated *with respect to R, iff*

$$\exists x_1 \in A_{i_1} \; \exists x_2 \in A_{i_2} \; \cdots \; \exists x_m \in A_{i_m} : \{x_1 x_2 \cdots x_m\} \notin R \restriction T \tag{4}$$

We denote that fact by $corr_R(T)$ *or* $corr(T)$ *if R is understood. The opposite concept is* uncorrelated. *The family* $\{T \subseteq A \,|\, corr_R(T)\}$, *in case it is nonempty, is called* the correlation family of R.

Note that no minimal correlated subset is a singleton. The following result re-states correlation of a subscheme in terms of the projection of the relation on it.

Lemma 1 *Let* $T \subseteq S$. *Then* $corr(T)$ *iff* $R \restriction T \subsetneq \bowtie_{X \in T} R \restriction \{X\}$.

Proof: First assume $corr(T)$. By Definition 3, there is an element in every attribute from T such that the tuple of those elements does not occur in $R \restriction T$. On the other hand, the tuples of $\bowtie_{X \in T} R \restriction \{X\}$ are all possible combinations of the elements of the attributes in T. Therefore, $R \restriction T \subsetneq \bowtie_{X \in T} R \restriction \{X\}$.

In the other direction, assume $\neg corr(T)$. The negation of expression (4) in Definition 3 is but another way to write $R \restriction T = \bowtie_{X \in T} R \restriction \{X\}$. □

As the next result establishes, with respect to the poset $(S, \subseteq)$, every correlated subset is upward closed, while every uncorrelated subset is downward closed.

Proposition 2 *If* $corr(T)$ *for some* $T \subseteq S$ *then* $\forall Z_{T \subseteq Z \subseteq S} : corr(Z)$. *If* $\neg corr(T)$ *for some* $T \subseteq S$ *then* $\forall Z_{Z \subseteq T} : \neg corr(Z)$.

It is obvious that the correlation family, if it exists, is a cover of the scheme. Furthermore, it does not exist iff the relation is complete. The interesting part of a correlation family is the sub-family comprising the minimal correlated sets. However, that sub-family does not necessarily cover the scheme. We want to define a family that both covers the scheme—because we are ultimately interested in a partition of the scheme—and is a Sperner family, since the implied members of the family are of no interest.

Definition 4 (mincor family) A mincor of *R is every minimal, self-correlated with respect to R, sub-scheme* $T \subseteq S$. *Further,* $\mathsf{mincors}(R) \overset{def}{=} \{T \subseteq S \,|\, T \text{ is a mincor}\}$ *and* $\mathsf{singletons}(R) \overset{def}{=} \{\{A\} \,|\, A \in S \wedge \neg \exists X \in \mathsf{mincors}(R) : A \in X\}$. The mincor family *of R, denoted by* $\mathsf{MF}(R)$, *is* $\mathsf{MF}(R) = \mathsf{mincors}(R) \cup \mathsf{singletons}(R)$.

For example, consider R' defined in (1) on the facing page. Clearly, $corr_{R'}(\{A, B\})$ and $corr_{R'}(\{C, D\})$ because of the lacks of both a_2 and b_1 in any tuple and the lack of both c_2 and d_1 in any tuple, respectively. The other four two-element subsets of S are uncorrelated. Then $\mathsf{singletons}(R') = \emptyset$ and therefore $\mathsf{MF}(R') = \{\{A, B\}, \{C, D\}\}$.

Proposition 3 *With respect to* S *and R,* $\mathsf{MF}(R)$ *exists and is unique.*

If *R* is complete then $\mathsf{MF}(R)$ consists of singletons. Clearly, $\mathsf{MF}(R) \in K(S)$, and thus $\mathsf{CC}(\mathsf{MF}(R)) \in \Pi(S)$.

3.2 Correlation in quotient relations

The following result establishes an important connection between self-correlation in a partition of the scheme and self-correlation in the scheme itself. More specifically, Lemma 2 is used to prove Lemma 3, and the latter is used in the proof of .

Lemma 2 *For any* $\mathfrak{X} \in \Pi(S)$ *and* $\mathfrak{X}' \subseteq \mathfrak{X}$:

$$corr_{R/\mathfrak{X}}(\mathfrak{X}') \leftrightarrow corr_R(\cup \mathfrak{X}')$$

Proof: Assume $corr_{R/\mathfrak{X}}(\mathfrak{X}')$. Let $\mathfrak{X}' = \{Y_1, Y_2, \ldots, Y_m\}$. So, $(R/\mathfrak{X}) \upharpoonright \mathfrak{X}'$ does not contain some m-tuple $\{U_1, U_2, \ldots, U_m\}$ such that $U_i \in R \upharpoonright Y_i$ for $1 \leq i \leq m$. Then $R \upharpoonright \cup \mathfrak{X}'$ does not contain $\cup\{U_1, U_2, \ldots, U_m\}$.

In the other direction, assume $corr_R(\cup \mathfrak{X}')$ where $\cup \mathfrak{X}'$ is a subset S′ of S. Let $S' = \{A_1, A_2, \ldots, A_n\}$. That is, $R \upharpoonright S'$ does not contain some n-tuple $\{W_1, W_2, \ldots, W_n\}$ such that $W_i \in A_i$ for $1 \leq i \leq n$. Let $\mathfrak{X}' = \{Y_1, Y_2, \ldots, Y_m\}$. Then $(R/\mathfrak{X}) \upharpoonright \mathfrak{X}'$ does not contain the m-tuple $\{U_1, U_2, \ldots, U_m\}$ where $U_i \in R \upharpoonright Y_i$ for $1 \leq i \leq m$. $\square$

As an example that illustrates Lemma 2, consider R' and $\mathfrak{X}_1$ on page 64. Clearly, $\mathfrak{X}_1 = \{\{A, B\}, \{C, D\}\}$ is self-correlated with respect to $\tilde{R}/\mathfrak{X}_1$ as $\tilde{R}/\mathfrak{X}_1$ does not contain, among others, the tuple $\{\{a_1, b_1\}\{c_1, d_2\}\}$. That implies $\cup \mathfrak{X}_1 = \{A, B, C, D\}$ is self-correlated with respect to $\tilde{R}$: since $\{\{a_1, b_1\}\{c_1, d_2\}\}$ is not an element of $\tilde{R}/\mathfrak{X}_1$, it must be the case that $\{a_1 b_1 c_1 d_2\}$ is not element of $\tilde{R}$ (and indeed it is not). In the other direction, the fact that $\{a_1 b_1 c_1 d_2\} \notin \tilde{R}$ implies $\{\{a_1, b_1\}\{c_1, d_2\}\} \notin \tilde{R}/\mathfrak{X}_1$.

The next result establishes that for every mincor Y of a quotient relation there is a way to pick elements from every element of Y such that the collection of those elements is a mincor of the original relation R.

Lemma 3 $\forall \mathfrak{X} \in \Pi(S) \; \forall \mathfrak{Y} \in \mathsf{mincors}(R/\mathfrak{X}) \; \exists Z \Subset \mathfrak{Y} : |Z| = |\mathfrak{Y}| \wedge \cup Z \in \mathsf{mincors}(R)$.

Proof: Assume $\mathfrak{Y} \in \mathsf{mincors}(R/\mathfrak{X})$. Clearly, there is some $Z \Subset \mathfrak{Y}$ such that $\cup Z$ is correlated with respect to R because $\Subset$ is reflexive and $\cup \mathfrak{Y}$ is correlated with respect to R by Lemma 2. Now consider any $Z' \Subset \mathfrak{Y}$ such that $|Z'| < |\mathfrak{Y}|$. There exists some $\mathfrak{Y}' \subset \mathfrak{Y}$ such that $Z \Subset \mathfrak{Y}'$. But $\mathfrak{Y}'$ is uncorrelated with respect to $R/\mathfrak{X}$ because $\mathfrak{Y}$ is a mincor of $R/\mathfrak{X}$ and so every proper subset of $\mathfrak{Y}$ is uncorrelated with respect to $R/\mathfrak{X}$. Note that $\mathfrak{Y}'$ being uncorrelated with respect to $R/\mathfrak{X}$ implies $\cup Z'$ is uncorrelated with respect to R by Lemma 2. It follows that for any $Z \Subset \mathfrak{Y}$ such that $corr_R(\cup Z)$—and we established such a Z exists—it is the case that $|Z| = |\mathfrak{Y}|$.

So, there exists a $Z \Subset \mathfrak{Y}$ such that $|Z| = |\mathfrak{Y}|$ and $\cup Z$ is correlated with respect to R. Furthermore, there does not exist $Z \Subset \mathfrak{Y}$ such that $|Z| < |\mathfrak{Y}|$ and $\cup Z$ is correlated with respect to R. Consider any $\tilde{Z} \Subset \mathfrak{Y}$ such that $\cup \tilde{Z}$ is correlated with respect to R. As $|\tilde{Z}| = |\mathfrak{Y}|$, every element of $\mathfrak{Y}$ is a superset of precisely one element of $\tilde{Z}$.

First assume all elements of $\tilde{Z}$ are singletons. In this case no proper subset of $\cup \tilde{Z}$ is correlated with respect to R. Suppose the contrary, namely that some $W \subset \cup \tilde{Z}$ is correlated with respect to R and deduce there is some $Z'' \Subset \mathfrak{Y}$ such that $W = \cup Z''$, thus $|Z''| < |\mathfrak{Y}|$, such that $\cup Z''$ is correlated with respect to R. Since no proper subset of $\cup \tilde{Z}$ is correlated with respect to R, $\cup \tilde{Z}$ is a mincor with respect to R and we are done with the proof.

Now assume not all elements of $\tilde{Z}$ are singletons. It trivially follows there exists a minimal set $\hat{Z} \Subset \tilde{Z}$ such that $|\hat{Z}| = |\tilde{Z}|$ (thus $|\hat{Z}| = |\mathfrak{Y}|$) such that $\cup \tilde{Z}$ is correlated with respect to R. $\square$

4   Results on Independent Partitions

This section provides important auxiliary results concerning independent partitions. In subsection 4.1 we investigate the connection between independence and self-correlation. In subsection 4.2 we prove the meet of independent partitions is an independent partition.

4.1 Independence and the mincor family

The following lemma establishes that partition independence is preserved under removal of attributes.

Lemma 4 $\forall \mathfrak{Y} \in \mathrm{III}(S) \ \forall \mathfrak{X} \Subset \mathfrak{Y} : \mathfrak{X} \in \mathrm{III}_{R \restriction \cup \mathfrak{X}}(\cup \mathfrak{X})$.

Proof: Let $Q = R \restriction \cup \mathfrak{X}$. We prove that $Q = \underset{Z \in \mathfrak{X}}{\bowtie} (Q \restriction Z)$. In one direction, $Q \subseteq \underset{Z \in \mathfrak{X}}{\bowtie} (Q \restriction Z)$ follows immediately from the definitions of relation join and projection. In the other direction, consider any tuple t in $\underset{Z \in \mathfrak{X}}{\bowtie} (Q \restriction Z)$. Let v be any tuple in $\underset{Z \in \mathfrak{Y}}{\bowtie} (R \restriction Z)$ such that $t = v\big|_{\cup \mathfrak{X}}$. But $v \in R$ because $\mathfrak{Y}$ is independent and thus $R = \underset{Z \in \mathfrak{Y}}{\bowtie} (R \restriction Z)$. As $v \in R$, it follows that $v\big|_{\cup \mathfrak{X}} \in Q$. But $v\big|_{\cup \mathfrak{X}}$ is t, therefore $t \in Q$, and so $\underset{Z \in \mathfrak{X}}{\bowtie} (Q \restriction Z) \subseteq Q$. $\qquad\square$

The next lemma is pivotal. It shows that the mincors respect independent partitions, in the sense that no mincor can intersect more than one element of an independent partition.

Lemma 5 $\forall \mathfrak{Y} \in \mathrm{III}(S) \ \forall W \in \mathsf{mincors}(R) \ \exists Y \in \mathfrak{Y} : W \subseteq Y$.

Proof: Assume the contrary. Then there is a mincor W that has nonempty intersection with more than one set from $\mathfrak{Y}$. Suppose W has nonempty intersection with precisely t sets from $\mathfrak{Y}$ for some t such that $2 \le t \le q$. Let $Y_1, Y_2, \ldots, Y_t$ be precisely those sets from $\mathfrak{Y}$ that have nonempty intersection with W. Let $W_i = W \cap Y_i$, for $1 \le i \le t$. Clearly, $\bigcup_{i=1}^{t} W_i = W$. By Lemma 4:

$$R \restriction W = \underset{1 \le i \le t}{\bowtie} R \restriction W_i$$

Every W_i is a proper subset of W. But W is a minimal correlated set. That implies $\neg \mathrm{corr}(W_i)$, for $1 \le i \le t$. Apply Lemma 1 to conclude that $R \restriction W_i = \underset{x \in W_i}{\bowtie} R \restriction \{x\}$. Then,

$$R \restriction W = \underset{1 \le i \le t}{\bowtie} \ \underset{x \in W_i}{\bowtie} R \restriction \{x\}$$

Obviously, $\underset{1 \le i \le t}{\bowtie} \ \underset{x \in W_i}{\bowtie} R \restriction \{x\} = \underset{x \in W}{\bowtie} R \restriction \{x\}$. Then, $R \restriction W = \underset{x \in W}{\bowtie} R \restriction \{x\}$. By Lemma 1 that implies $\neg \mathrm{corr}(W)$. $\qquad\square$

Furthermore, merging mincors also yields sets that respect independent partitions.

Corollary 1 $\forall \mathfrak{Y} \in \mathrm{III}(S) : \mathsf{CC}(\mathsf{MF}(R)) \sqsubseteq \mathfrak{Y}$.

Proof: Assume the contrary. Then for some R on S and $\mathfrak{Y} \in \mathrm{III}(S)$:

$$\exists X \in \mathsf{CC}(\mathsf{MF}(R)) \ \forall Y \in \mathfrak{Y} \ \exists A \in X : A \notin Y$$

First note that X is not a singleton, otherwise X would be contained in some set from $\mathfrak{Y}$. So, $|X| \ge 2$ and according to Definition 4, X is the union of one or more mincors, each of size ≥ 2, and X is connected. But by assumption X is not a subset of any set from $\mathfrak{Y}$ and so there has to be some mincor $W \in X$ that has nonempty intersection with at least two sets from $\mathfrak{Y}$. However, that contradicts Lemma 5. $\qquad\square$

Note that $\mathsf{CC}(\mathsf{MF}(R))$ is not necessarily an independent partition. For example, consider R' defined in (1) on page 64. As explained on page 65, $\mathsf{MF}(R') = \{\{A,B\},\{C,D\}\}$ and thus $\mathsf{CC}(\mathsf{MF}(R')) = \{\{A,B\},\{C,D\}\}$, too. But $\{\{A,B\},\{C,D\}\}$ is not an independent partition with respect to R'. In fact, there is no independent partition of S except for the trivial partition as $|R'|$ is a prime number.

Now consider another relation R'' on the same scheme:

$$R'' = \{\{a_1b_1c_1d_1\}, \{a_1b_1c_1d_2\}, \{a_1b_1c_2d_2\}, \{a_1b_2c_1d_1\}, \{a_1b_2c_1d_2\}, \{a_1b_2c_2d_2\},$$
$$\{a_2b_2c_1d_1\}, \{a_2b_2c_1d_2\}, \{a_2b_2c_2d_2\}\}$$

But $\mathsf{MF}(R'') = \{\{A,B\},\{C,D\}\} = \mathsf{CC}(\mathsf{MF}(R''))$ just as in the case of R'. Now $\{\{A,B\},\{C,D\}\}$ is an independent partition with respect to R'' because $R'' = R'' \restriction \{A,B\} \bowtie R'' \restriction \{C,D\}$.

So, in the case of R'', the connected components of the mincor family constitute an independent partition, while that is not true for R', although the mincor families of both relations are the same. We conclude that computing the mincor family does not suffice to obtain an independent partition. Therefore, we use a more involved approach in which the computation of the mincor family is but the first step towards the computation of the maximum independent partition.

4.2 The meet of independent partitions

The following lemma allows us to define the maximum independent partition as the meet of all independent partitions.

Lemma 6 $\forall \mathfrak{X}, \mathfrak{Y} \in I\Pi(S) : \mathfrak{X} \sqcap \mathfrak{Y} \in I\Pi(S)$.

Proof: (sketch) Let $\mathfrak{X}, \mathfrak{Y} \in I\Pi(S)$. We assume $\mathfrak{X} \sqcup \mathfrak{Y}$ is connected. There is no true loss of generality in that because the proof below can be done componentwise if $\mathfrak{X} \sqcup \mathfrak{Y}$ is not connected. Relative to an arbitrary element of $\mathfrak{X}$, say X_1, we define the family $\mathfrak{Z} = \{Z_0, Z_1, \ldots, Z_k\}$ over S as follows. $\mathfrak{Z}$ is a partition of S and its elements are constructed in an ascending order of the index according to the following rule:

$$Z_i = \begin{cases} X_1, & \text{if } i = 0 \\ \bigcup\{A \setminus Z_{i-1} \,|\, A \in \mathfrak{Y} \wedge A \cap Z_{i-1} \neq \emptyset\}, & \text{if } i \text{ is odd} \\ \bigcup\{A \setminus Z_{i-1} \,|\, A \in \mathfrak{X} \wedge A \cap Z_{i-1} \neq \emptyset\}, & \text{if } i \text{ is even and } i > 0 \end{cases}$$

Let us define $B_i = \{\bigcup_{j=0}^i Z_j\} \sqcap \mathfrak{X} \sqcap \mathfrak{Y}$ for $0 \leq i \leq k$. Clearly, $B_0 = \{X_1\} \sqcap \mathfrak{Y}$, $B_i = B_{i-1} \cup (\{Z_i\} \sqcap \mathfrak{X} \sqcap \mathfrak{Y})$ for $1 \leq i \leq k$ and $B_k = \mathfrak{X} \sqcap \mathfrak{Y}$. Furthermore, $\cup B_k = S$ and thus $R \restriction \cup B_k = R$. We prove by induction on i that for all i such that $0 \leq i \leq k$:

$$R \restriction \cup B_i = \bowtie_{C \in B_i} R \restriction C \tag{5}$$

and hence the result follows.

Basis. Let $i = 0$. Let the elements of $\mathfrak{Y}$ that have nonempty intersection with X_1 be called $Y_1, \ldots, Y_j$. Obviously, there is at least one of them. The claim is that $R \restriction X_1 = \bowtie_{i=1}^j R \restriction (X_1 \cap Y_i)$. That follows immediately from Lemma 4.

Inductive Step. Assume the claim holds for some B_{i-1} such that $0 \leq i-1 < k$ and consider B_i. As already mentioned, $B_i = B_{i-1} \cup (\{Z_i\} \sqcap \mathfrak{X} \sqcap \mathfrak{Y})$.

Without loss of generality, assume i is odd. Very informally speaking, Z_i is the union of some elements of $\mathfrak{Y}$ that overlap with some elements (from $\mathfrak{X}$) in B_{i-1}, minus the overlap. Therefore, we can write $B_i = B_{i-1} \cup (\{Z_i\} \sqcap \mathfrak{X})$ because under the current assumption, it is $\mathfrak{X}$ rather than $\mathfrak{Y}$ that dictates the grouping together of the elements of Z_i in B_i. More specifically, since $i \neq k$, there are elements from $\mathfrak{X}$ whose elements do not appear in the current B_i; those elements of $\mathfrak{X}$ dictate the aforementioned grouping.

So, B_i is the union of two disjoint sets whose elements are from $\mathfrak{X} \sqcap \mathfrak{Y}$, namely B_{i-1} and $\{Z_i\} \sqcap \mathfrak{X}$. By the inductive hypothesis, $R \upharpoonright \cup B_{i-1} = \underset{C \in B_{i-1}}{\bowtie} R \upharpoonright C$.

Consider $\{Z_i\} \sqcap \mathfrak{X}$ and call its elements, $T_1, \ldots, T_m$. Without loss of generality, consider T_1. Our immediate goal is to prove that $R \upharpoonright ((\cup B_{i-1}) \cup T_1) = \underset{C \in B_{i-1} \cup \{T_1\}}{\bowtie} R \upharpoonright C$. Note that T_1 is a subset of some $Y' \in \mathfrak{Y}$ such that Y' has nonempty intersection with $\cup B_{i-1}$, T_1 itself being disjoint with B_{i-1}. Furthermore, T_1 is the intersection of Y' with some $X' \in \mathfrak{X}$. X' is disjoint with $\cup B_{i-1}$, otherwise the elements of T_1 would be part of $\cup B_{i-1}$. Furthermore, every element of B_{i-1} is a subset of some element of $\mathfrak{X}$ that is not X'. Let the elements of $\mathfrak{X}$ that have subsets-elements of B_{i-1} be $X_1, \ldots, X_p$. Note that $X_1 \cup \cdots \cup X_p = \cup B_{i-1}$. By Lemma 4, it is the case that

$$R \upharpoonright (X_1 \cup \cdots \cup X_p \cup T_1) = R \upharpoonright X_1 \bowtie \cdots \bowtie R \upharpoonright X_p \bowtie R \upharpoonright T_1 \tag{6}$$

since T_1 is a subset of X' and X' is none of $X_1, \ldots, X_p$. However, $X_1 \cup \cdots \cup X_p \cup T_1 = (\cup B_{i-1}) \cup T_1$ by an earlier observation and $R \upharpoonright X_1 \bowtie \cdots \bowtie R \upharpoonright X_p = \underset{C \in B_{i-1}}{\bowtie} R \upharpoonright C$. Substitute that in equation 6 to obtain

$$R \upharpoonright (\cup B_{i-1} \cup T_1) = \left(\underset{C \in B_{i-1}}{\bowtie} R \upharpoonright C \right) \bowtie R \upharpoonright T_1 = \underset{C \in B_{i-1} \cup \{T_1\}}{\bowtie} R \upharpoonright C \tag{7}$$

which is what we wanted to prove with respect to T_1.

We can use (7) as the basis of a nested induction. More specifically, we prove that

$$R \upharpoonright ((\cup B_{i-1}) \cup T_1 \cup \cdots \cup T_k) = \left(\underset{C \in B_{i-1}}{\bowtie} R \upharpoonright C \right) \bowtie R \upharpoonright T_1 \bowtie \cdots \bowtie R \upharpoonright T_k$$

implies

$$R \upharpoonright ((\cup B_{i-1}) \cup T_1 \cup \cdots \cup T_{k+1}) = \left(\underset{C \in B_{i-1}}{\bowtie} R \upharpoonright C \right) \bowtie R \upharpoonright T_1 \bowtie \cdots \bowtie R \upharpoonright T_{k+1}$$

for any $k \in \{1, 2, \ldots, m-1\}$. The nested induction can be proved in a straightforward manner, having in mind the proof of (7). That implies the desired:

$$R \upharpoonright ((\cup B_{i-1}) \cup T_1 \cup \cdots \cup T_m) = \left(\underset{C \in B_{i-1}}{\bowtie} R \upharpoonright C \right) \bowtie R \upharpoonright T_1 \bowtie \cdots \bowtie R \upharpoonright T_m$$

And that concludes the proof because $\cup B_i = \cup B_{i-1} \cup T_1 \cup \cdots \cup T_m$. □

The proof of Lemma 6 relies on the fact that all sets we consider are finite.

As a corollary of Lemma 6, the maximum independent partition, which is the object of our study, is well-defined: $\sqcap \Pi \Pi(S)$ exists, it is unique, and is an element of $\Pi \Pi(S)$. For notational convenience we introduce another term for that object. We say that $\sqcap \Pi \Pi_R(S)$ is the *focus* of R and denote it by $\mathsf{foc}(R)$. A trivial observation is that $\Pi \Pi_R(S)$ coincides with $\uparrow \mathsf{foc}(R)$.

5 A Fixed Point Characterisation of the Maximum Independent Partition

In this section we identify the object of our study as the least fixed point of α, where α is a transformer on the lattice of all partitions of S. Furthermore, we present an iterative fixed point approximation procedure for computing the maximum independent partition.

5.1   Function α

First we introduce a helper function. Let A be a ground set. The function ξ maps superfamilies over A to families over A as follows. For any superfamily $\mathfrak{F}$:

$$\xi(\mathfrak{F}) \overset{\text{def}}{=} \{\cup Z | Z \in \mathfrak{F}\}$$

Syntactically speaking, ξ removes the innermost pairs of parentheses. For instance, suppose $A = \{a,b,c,d\}$ and $\mathfrak{F} = \{\{\{a\},\{b,c\}\},\{\{d\}\}\}$. Then $\xi(\mathfrak{F}) = \{\{a,b,c\},\{d\}\}$.

   We now define the central function of the present study. It takes a partition of S, identifies the mincors of the corresponding quotient relation, merges the overlapping mincors, and uses ξ to map the result back to a partition of S.

Definition 5 (function α) $\alpha_R : \Pi(S) \to \Pi(S)$*, shortly α when R is understood, is defined as follows for any $\mathfrak{X} \in \Pi(S)$:*

$$\alpha_R(\mathfrak{X}) \overset{\text{def}}{=} \xi(\text{CC}(\text{MF}(R/\mathfrak{X})))$$

   Notably, α is *not monotone* in general as demonstrated by the following example. Let $\widetilde{S} = \{A,B,C,D,E\}$ and let each attribute have precisely two values, say $A = \{a_1,a_2\}$ and so on. Let Q be the relation obtained from the complete relation over $\widetilde{S}$ after deleting all tuples containing $a_1b_1c_1$, all tuples containing d_2e_2, and the tuples $\{a_2b_1c_1d_2e_1\},\{a_2b_2c_1d_2e_1\}$. In other words,

$$Q = \{\{a_1b_1c_2d_1e_1\},\{a_1b_1c_2d_1e_2\},\{a_1b_1c_2d_2e_1\},\{a_1b_2c_1d_1e_1\},\{a_1b_2c_1d_1e_2\},\{a_1b_2c_1d_2e_1\}$$
$$\{a_1b_2c_2d_1e_1\},\{a_1b_2c_2d_1e_2\},\{a_1b_2c_2d_2e_1\},\{a_2b_1c_1d_1e_1\},\{a_2b_1c_1d_1e_2\},\{a_2b_1c_2d_1e_1\}$$
$$\{a_2b_1c_2d_1e_2\},\{a_2b_1c_2d_2e_1\},\{a_2b_2c_1d_1e_1\},\{a_2b_2c_1d_1e_2\},\{a_2b_2c_2d_1e_1\},\{a_2b_2c_2d_1e_2\},\{a_2b_2c_2d_2e_1\}\}$$

   Let us see which sets of attributes are self-correlated with respect to Q. The only two-element subset of $\widetilde{S}$ that is self-correlated is $\{D,E\}$. Further, $\{A,B,C\}$ is self-correlated. It follows $\text{MF}(Q) = \{\{A,B,C\},\{D,E\}\}$. Consider the following two partitions of $\widetilde{S}$: $\mathfrak{X}_1 = \{\{A\},\{B\},\{C\},\{D\},\{E\}\}$ and $\mathfrak{X}_2 = \{\{A\},\{B,D\},\{C,E\}\}$. Obviously, $\mathfrak{X}_1 \sqsubseteq \mathfrak{X}_2$. It is clear that $\alpha(\mathfrak{X}_1) = \{\{A,B,C\},\{D,E\}\}$. Consider $\alpha(\mathfrak{X}_2)$. The set $\{\{B,D\},\{C,E\}\}$ is self-correlated because of the lack of $\{b_1,d_2\}$ and $\{c_1,e_2\}$ in any tuple, which in its turn is due to the fact that d_2 and e_2 do not occur in any tuple of R. The sets $\{\{A\},\{B,D\}\}$ and $\{\{A\},\{C,E\}\}$ are uncorrelated. It follows that $\alpha(\mathfrak{X}_2) = \{\{A\},\{B,C,D,E\}\}$, and thus $\alpha(\mathfrak{X}_1) \not\sqsubseteq \alpha(\mathfrak{X}_2)$.

However, we have the following property of α that shall later be exploited.

Proposition 4 α *is an inflationary function on* $(\Pi(S),\sqsubseteq)$.

5.2   Independence and function α

The following central result establishes that the independent partitions are precisely the fixed points of α.

Theorem 1 $\forall \mathfrak{X} \in \Pi(S) : \mathfrak{X} \in I\Pi(S) \leftrightarrow \alpha(\mathfrak{X}) = \mathfrak{X}$.

Proof: In one direction, assume $\mathfrak{X} \in I\Pi(S)$. $R/\mathfrak{X}$ is complete by Proposition 1. By definition, that is $R/\mathfrak{X} = \times_{Y \in \mathfrak{X}} Y$. By the definition of $\bowtie$, $(R/\mathfrak{X}) \restriction \mathfrak{X} = \bowtie_{Y \in \mathfrak{X}} (R/\mathfrak{X}) \restriction \{Y\}$. It follows that $\neg\text{corr}(\mathfrak{X})$ by Lemma 1. So, $\text{mincors}(R/\mathfrak{X}) = \emptyset$ and $\text{MF}(R/\mathfrak{X}) = \text{singletons}(R/\mathfrak{X})$ by Definition 4. Then $\text{CC}(\text{MF}(R/\mathfrak{X})) =$

$\{\{A\}\,|\,A \in \mathfrak{X}\}$. Therefore, $\xi(\mathsf{CC}(\mathsf{MF}(R/\mathfrak{X}))) = \{A\,|\,A \in \mathfrak{X}\} = \mathfrak{X}$. But $\xi(\mathsf{CC}(\mathsf{MF}(R/\mathfrak{X})))$ is $\alpha(\mathfrak{X})$ by defini-
tion. Therefore, $\alpha(\mathfrak{X}) = \mathfrak{X}$.

In the other direction, assume $\alpha(\mathfrak{X}) = \mathfrak{X}$. That is, $\xi(\mathsf{CC}(\mathsf{MF}(R/\mathfrak{X}))) = \mathfrak{X}$, which in its turn implies
$\mathsf{CC}(\mathsf{MF}(R/\mathfrak{X})) = \{\{A\}\,|\,A \in \mathfrak{X}\}$ because $\mathsf{CC}(\mathsf{MF}(R/\mathfrak{X}))$ is a superfamily such that every element from S is in
precisely one element of precisely one element of it. The remainder of the proof mirrors the above one.
□

Having in mind the observation on page 69 that $\mathrm{I}\Pi_R(\mathrm{S})$ coincides with $\uparrow\!\mathsf{foc}(R)$, we derive the follow-
ing corollary of Theorem 1.

Corollary 2 $\uparrow\!\mathsf{foc}(R)$ *is closed with respect to* α.

The following lemma says that the mincors of a quotient relation respect the focus of the relation in the
sense that for every mincor of $R/\mathfrak{X}$, the union of its elements is a subset of some element of the focus.

Lemma 7 $\forall \mathfrak{X} \in \downarrow\!\mathsf{foc}(R)\ \forall T \in \mathsf{mincors}(R/\mathfrak{X})\ \exists \mathrm{Y} \in \mathsf{foc}(R): \cup T \subseteq \mathrm{Y}$.

Proof: Assume the contrary. That is, for some partition $\mathfrak{X}$ that refines the focus there is a mincor T
of $R/\mathfrak{X}$ such that $\cup T$ has nonempty intersection with at least two subsets, call them Y_1 and Y_2, of the
focus. Use Lemma 3 to conclude there is some $\mathrm{Z} \Subset T$ such that $|\mathrm{Z}| = |T|$ and $\cup \mathrm{Z} \in \mathsf{mincors}(R)$. Since
$|\mathrm{Z}| = |T|$, it must be the case that $\cup \mathrm{Z}$ has nonempty intersection with both Y_1 and Y_2. But the focus is
an independent partition. We derived that a mincor of R, namely $\cup \mathrm{Z}$, intersects two distinct elements of
an independent partition. That contradicts Lemma 5 directly. □

We already established (see Proposition 4) that α is an inflationary function. The next lemma, however,
establishes a certain restriction: the application of α on a dependent partition can yield another dependent
partition or at most the focus, and never an independent partition "above" the focus.

Lemma 8 $\downarrow\!\mathsf{foc}(R)$ *is closed with respect to* α.

Proof: We prove that $\forall \mathfrak{X} \in \downarrow\!\mathsf{foc}(R): \alpha(\mathfrak{X}) \sqsubseteq \mathsf{foc}(R)$. Recall that $\alpha(\mathfrak{X})$ is a partition of S and it abstracts
$\mathfrak{X}$. Assume the claim is false. Then there is a partition $\mathfrak{X}$ such that $\mathfrak{X} \sqsubseteq \downarrow\!\mathsf{foc}(R)$ but $\alpha(\mathfrak{X}) \not\sqsubseteq \downarrow\!\mathsf{foc}(R)$. Then
there is some $\mathrm{P} \in \alpha(\mathfrak{X})$ such that P has nonempty intersection with at least two elements, call them Y_1
and Y_2, of $\mathsf{foc}(R)$. However, P is $\xi(C)$ for some C that is a connected component—relative to the ground
set $\mathfrak{X}$—of the mincor family of $R/\mathfrak{X}$. Consider C. It is the union of one or more mincors of $R/\mathfrak{X}$, those
mincors being subsets of $\mathfrak{X}$.

Since $\mathfrak{X} \sqsubseteq \mathsf{foc}(R)$, no element of $\mathfrak{X}$ can intersect both Y_1 and Y_2. It follows that at least one mincor
$\mathrm{M} \in C$ is such that $\cup \mathrm{M}$ intersects both Y_1 and Y_2. But that contradicts Lemma 7. □

The next and final central result allows us to compute the focus of R by an iterative application of α,
starting with the partition into singletons.

Theorem 2 *For some m such that* $1 \leq m \leq |\mathrm{S}|$, $\alpha^m(\bot) = \mathsf{foc}(R)$.

Proof: Consider the sequence:

$$C = \bot,\ \alpha(\bot),\ \alpha^2(\bot),\ \dots$$

It is a chain in the lattice $(\Pi(\mathrm{S}), \sqsubseteq)$, as $\alpha(\mathfrak{X})$ abstracts $\mathfrak{X}$ for all $\mathfrak{X}$ (see Proposition 4), therefore all those
elements are comparable with respect to $\sqsubseteq$. C has only a finite number of distinct elements as the said
lattice is finite.

First note that every element of C is in $\downarrow\!\mathsf{foc}(R)$. Indeed, assuming the opposite immediately contra-
dicts Lemma 8.

Then note that for every $\mathfrak{X} \in {\downarrow}\mathsf{foc}(R) \setminus \{\mathsf{foc}(R)\}$, it is the case that $\alpha(\mathfrak{X}) \neq \mathfrak{X}$. Assuming the opposite implies $\mathfrak{X}$ is a fixed point of α, contradicting Corollary 2. Proposition 4 implies a stronger fact: for every $\mathfrak{X} \in {\downarrow}\mathsf{foc}(R) \setminus \{\mathsf{foc}(R)\}$, it is the case that $\mathfrak{X} \sqsubset \alpha(\mathfrak{X})$. But ${\downarrow}\mathsf{foc}(R)$ is a finite lattice. It follows immediately that for some value m not greater than $|S|$, $\alpha^m(\bot)$ equals the top of ${\downarrow}\mathsf{foc}(R)$, *viz.* $\mathsf{foc}(R)$. $\qquad\square$

We thus obtain Kleene's iterative least fixed point approximation procedure [7], however for inflationary functions instead of monotone ones.

Corollary 3 *The following algorithm:*

> $\mathfrak{X} \leftarrow \bot$
> **while** $\mathfrak{X} \neq \alpha(\mathfrak{X})$
> $\qquad \mathfrak{X} \leftarrow \alpha(\mathfrak{X})$
> **return** $\mathfrak{X}$

computes the least fixed point of α, i.e.*, the maximum independent partition of* S *with respect to R.* $\quad\square$

Here is a small example illustrating the work of that algorithm. Consider S and R' defined in (1) on page 64. $\bot$ is $\{\{A\},\{B\},\{C\},\{D\}\}$. Let us compute $\alpha(\bot)$, that is, $\xi\big(\mathsf{CC}\big(\mathsf{MF}\big(R'/\bot\big)\big)\big)$. $R'/\bot$ is the same as $R'/\mathfrak{X}_2$ on page 64, namely:

$$R'/\bot = \{\{\{a_1\}\{b_1\}\{c_1\}\{d_1\}\}, \{\{a_1\}\{b_1\}\{c_2\}\{d_2\}\}, \{\{a_1\}\{b_2\}\{c_1\}\{d_2\}\},$$
$$\{\{a_2\}\{b_2\}\{c_1\}\{d_1\}\}, \{\{a_2\}\{b_2\}\{c_2\}\{d_2\}\}\}$$

Let us compute $\mathsf{CC}\big(\mathsf{MF}\big(R'/\bot\big)\big)$. Having in mind that $\mathsf{MF}(R') = \{\{A,B\},\{C,D\}\}$ as explained on page 65, conclude that $\mathsf{CC}\big(\mathsf{MF}\big(R'/\bot\big)\big) = \{\{\{A,B\}\},\{\{C,D\}\}\}$. Therefore, $\xi\big(\mathsf{CC}\big(\mathsf{MF}\big(R'/\bot\big)\big)\big) = \{\{A,B\},\{C,D\}\}$. That differs from $\bot$ and the **while** loop is executed again. $R'/\alpha(\bot)$ is the same as $R'/\mathfrak{X}_1$ on page 64, namely:

$$R'/\alpha(\bot) = \{\{\{a_1 b_1\}\{c_1 d_1\}\}, \{\{a_1 b_1\}\{c_2 d_2\}\}, \{\{a_1 b_2\}\{c_1 d_2\}\},$$
$$\{\{a_2 b_2\}\{c_1 d_1\}\}, \{\{a_2 b_2\}\{c_2 d_2\}\}\}$$

Let us compute $\mathsf{CC}\big(\mathsf{MF}\big(R'/\alpha(\bot)\big)\big)$. To that end, note that $\alpha(\bot) = \{\{A,B\},\{C,D\}\}$ is self-correlated with respect to $R'/\{\{A,B\},\{C,D\}\}$ because of the lack of, for instance, both $\{a_1,b_2\}$ and $\{c_1,d_1\}$ in any tuple of $R'/\alpha(\bot)$. It follows that $\mathsf{CC}\big(\mathsf{MF}\big(R'/\alpha(\bot)\big)\big) = \{\{\{A,B\},\{C,D\}\}\}$ and, therefore, $\alpha^2(\bot) = \xi\big(\mathsf{CC}\big(\mathsf{MF}\big(R'/\alpha(\bot)\big)\big)\big) = \{\{A,B,C,D\}\}$. That differs from $\alpha(\bot)$ and the **while** loop is executed once more. At the end of that execution, it turns out that $\alpha^3(\bot)$ equals $\alpha^2(\bot)$ and the algorithm terminates, returning as the result $\{\{A,B,C,D\}\}$, the trivial partition.

6   Related Work

An algorithm that factorizes a given relation into prime factors is proposed in [10, algorithm PRIME FACTORIZATION]. It runs in time $O(mn\lg n)$ where m is the number of tuples and n is the number of attributes. Since mn is the input size, that time complexity is very close to the optimum. The theoretical foundation of PRIME FACTORIZATION is a theorem (see [10, Proposition 10]) that says a given relation S has a factor F iff, with respect to any attribute A and any value v of its domain, F is a factor of both Q and R where Q and R are relations such that $Q \cup R = S$ and Q consists precisely of the tuples in which the value of A is v. In other words, the approach of [10] to the problem of computing the prime factors

is "horizontal splitting" of the given relation using the selection operation from relational algebra. The approach of this paper to that same problem is quite different. We utilise "vertical splitting", using the projection operation of relational algebra. The theoretical foundation of our approach is based on the concept of self-correlation of a subset of the attributes; that concept has no analogue in [10].

An excellent exposition of the benefits of the factorisation of relational data is [11]. The factorised representation both saves space, where the gain can potentially be as good as exponential, and time, speeding up the processing of information whose un-factorised representation is too big. [1] proposes a way of decomposing relational data that is incomplete and [13] proposes factorisation of relational data that facilitates machine learning.

Clusterisation of multidimensional data into non-intersecting classes called clusters is an important, hard and computationally demanding problem. [5] investigates clustering in high-dimensional data by detection of orthogonality in the latter. [8] proposes so called community discovering, which is a sort of clusterisation, in media social networks by utilising factorisation of a relational hypergraph.

The foundation of this paper is the work of Gurov *et al.* [6] that investigates relational factorisation of a restricted class of relations called there simple families. [6] introduces the concept of correlation between the attributes and proposes a fast and practical algorithm that computes the optimum factorisation of a simple family by using a subroutine for correlation. The fundamental approach of this paper is an extension of that, however now correlation is considerably more involved, being not a binary relation between attributes but a relation of arbitrary arity (this is the only place where "relation" means relation in the Set Theory sense, that is, a set of ordered tuples).

7 Conclusion

This paper illustrates the utility of fixed points to formally express maximum independence in relations by means of minimum correlated sets of attributes. By using minimum correlated sets, we define an inflationary transformer over a finite lattice and show the maximum independent partition is the least fixed point of this transformer. Then we prove the downward closure of that least fixed point is closed under the transformer. Hence, the least fixed point can be computed by applying the transformer iteratively from the bottom element of the lattice until stabilization. This iterative construction is the same as Kleene's construction, but does not rely on monotonicity of the transformer to guarantee that it computes the least fixed point.

A topic for future work is to introduce a quantitative measure for the degree of independence between sets of attributes and investigate approximate relational factorisation.

References

[1] Lyublena Antova, Christoph Koch & Dan Olteanu (2009): $10^{(10^6)}$ *worlds and beyond: efficient representation and processing of incomplete information*. VLDB J. 18(5), pp. 1021–1040, doi:10.1007/s00778-009-0149-y.

[2] Garrett Birkhoff (1967): *Lattice Theory*, 3rd edition. American Mathematical Society, Providence.

[3] Brian A. Davey & Hilary A. Priestley (2002): *Introduction to Lattices and Order*. Cambridge mathematical text books, Cambridge University Press, doi:10.1017/CBO9780511809088.

[4] Erich Grädel & Jouko A. Väänänen (2013): *Dependence and Independence*. Studia Logica 101(2), pp. 399–410, doi:10.1007/s11225-013-9479-2.

[5] Stephan Günnemann, Emmanuel Müller, Ines Färber & Thomas Seidl (2009): *Detection of Orthogonal Concepts in Subspaces of High Dimensional Data*. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, ACM, New York, NY, USA, pp. 1317–1326, doi:10.1145/1645953.1646120.

[6] Dilian Gurov, Bjarte M. Østvold & Ina Schaefer (2011): *A Hierarchical Variability Model for Software Product Lines*. In: *Leveraging Applications of Formal Methods, Verification, and Validation - International Workshops, SARS 2011 and MLSC 2011, Held Under the Auspices of ISoLA 2011 in Vienna, Austria, October 17-18, 2011. Revised Selected Papers*, pp. 181–199, doi:10.1007/978-3-642-34781-8_15.

[7] Jean-Louis Lassez, V. L. Nguyen & Liz Sonenberg (1982): *Fixed Point Theorems and Semantics: A Folk Tale*. Information Processing Letters 14(3), pp. 112–116, doi:10.1016/0020-0190(82)90065-5.

[8] Yu-Ru Lin, Jimeng Sun, Paul Castro, Ravi Konuru, Hari Sundaram & Aisling Kelliher (2009): *MetaFac: Community Discovery via Relational Hypergraph Factorization*. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, ACM, New York, NY, USA, pp. 527–536, doi:10.1145/1557019.1557080.

[9] David Maier (1983): *The Theory of Relational Databases*. Computer Science Press. Accessible online at `http://web.cecs.pdx.edu/~maier/TheoryBook/TRD.html`.

[10] Dan Olteanu, Christoph Koch & Lyublena Antova (2008): *World-set decompositions: Expressiveness and efficient algorithms*. Theoretical Computer Science 403(2-3), pp. 265–284, doi:10.1016/j.tcs.2008.05.004.

[11] Dan Olteanu & Jakub Závodný (2015): *Size Bounds for Factorised Representations of Query Results*. ACM Trans. Database Syst. 40(1), p. 2, doi:10.1145/2656335.

[12] Klaus Pohl, Günter Böckle & Frank van der Linden (2005): *Software Product Line Engineering - Foundations, Principles, and Techniques*. Springer, doi:10.1007/3-540-28901-1.

[13] Steffen Rendle (2013): *Scaling Factorization Machines to Relational Data*. PVLDB 6(5), pp. 337–348. Available at `http://www.vldb.org/pvldb/vol6/p337-rendle.pdf`.

[14] S. Roman (2008): *Lattices and Ordered Sets*. Springer. Available at `https://books.google.com/books?id=NZN8aum26LgC`.

[15] Ina Schaefer, Dilian Gurov & Siavash Soleimanifard (2010): *Compositional Algorithmic Verification of Software Product Lines*. In: *Formal Methods for Components and Objects - 9th International Symposium, FMCO 2010, Graz, Austria, November 29 - December 1, 2010. Revised Papers*, pp. 184–203, doi:10.1007/978-3-642-25271-6_10.

[16] Alfred Tarski (1955): *A Lattice-Theoretical Fixpoint Theorem and Its Applications*. Pacific journal of Mathematics 5(2), pp. 285–309, doi:10.2140/pjm.1955.5.285.

[17] Bruce W. Weide, Wayne D. Heym & Joseph E. Hollingsworth (1995): *Reverse Engineering of Legacy Code Exposed*. In: *Proceedings of the 17th International Conference on Software Engineering*, ICSE '95, ACM, New York, NY, USA, pp. 327–331, doi:10.1145/225014.225045.