On the Orthographic Dimension of Definable Sets

Stavros Cosmadakis
University of Patras

Gabriel Kuper
Bell Labs

Leonid Libkin*
University of Toronto and Bell Labs

Abstract

A formula $\varphi(x_1, \ldots, x_n)$ conforms to a partition P of $\{x_1, \ldots, x_n\}$ if it is equivalent to a Boolean combination of formulae that do not have free variables from more than one block of P. We show that if φ conforms to two partitions P_1 and P_2, it also conforms to their greatest lower bound in the partition lattice. As a corollary, we obtain that the concept of orthographic dimension of a constraint-definable set, introduced in the field of constraint databases, is well-defined.

1 Introduction

The primary motivation for the concept of orthographic dimension is from the field of constraint databases [5, 6], but it may also be of independent interest. The idea of constraint databases is that database relations are first-order definable sets over some structure. This way one can store a finite representation of an infinite set in a database, and query it as if the entire set were stored. Query evaluation over constraint databases then reduces to constraint processing. For large constraint datasets, this could become very expensive if the dimension (number of variables) is high.

However, in many applications data of high dimensionality can be described by a combination of several components of lower dimension. For example, a cadastral database may contain spatial information (who owns what land) and time (who owns it when). Or, in a 3-dimensional GIS, it may be possible to losslessly describe the data as a set of 2-dimensional layers. In other words, some variables may be independent of each other. One can often benefit from such independence: for example, query evaluation can be considerably improved [4, 3]. It was also shown in [1], that spatial aggregates can be safely introduced into query languages under such independence conditions.

A related notion is that of the "orthographic dimension" – the size of the largest block of dependent variables. This concept is important since many algorithms on constraint databases have running time of the form $O(n^{O(d)})$ where n is the size of the constraint representation and d the orthographic dimension; thus, establishing a lower orthographic dimension decreases complexity. This concept was introduced in [4], but it was not clear whether it was well-defined: conceivably, there could be various ways of breaking free variables into independent blocks of variables, and for two such ways the sizes of the largest blocks may not coincide.

In this note, we show that this is not the case – for any constraint-definable set, the orthographic

*Contact author. Address: before November 1, 2000 – Bell Laboratories, Room 2C-407, 600 Mountain Avenue, Murray Hill, NJ 07974, USA. After November 1, 2000 – Department of Computer Science, University of Toronto 6 King's College Road, Toronto, Ontario, Canada M5S 3H5. Email: libkin@cs.toronto.edu.

dimension is well-defined. In fact, we prove a more general result: if there are two ways to partition variables into independent blocks, then there is another way to do so that refines both partitions.

This result, and its algorithmic consequences described in [7], may be related to work on variable elimination and early projection in CLP (see, e.g., [2]), and perhaps to other constraint processing algorithms. It also seems close in spirit to various interpolation theorems in logic, but despite these similarities, the notion of orthographic dimension does not appear to have been studied outside of the field of constraint databases.

2 Definitions and Main Theorem

Suppose we are given a first-order language L, an L-structure $\mathcal{M}$, and an L-formula $\varphi(x_1, \ldots, x_n)$ with free variables $x_1, \ldots, x_n$. Let P be a partition on $\{x_1, \ldots, x_n\}$. We say that φ *$\mathcal{M}$-conforms* to P (written as $\varphi \sim_{\mathcal{M}} P$) if there exists a finite collection of formulae γ_i, $i \in I$, such that:

1. φ is equivalent to a Boolean combination of formulae γ_i, and
2. no γ_i has free variables from two different blocks of the partition P.

For example, if $L = (+, *, 0, 1, <)$, $\mathcal{M}$ is the real field, $\varphi(x_1, x_2) = ((x_1 > 0) \wedge (x_2 > 0) \wedge (x_1 \geq x_2)) \vee ((x_1 > 0) \wedge (x_2 > 0) \wedge (x_2 \geq x_1))$, then $\varphi \sim_{\mathcal{M}} \{\{x_1\}, \{x_2\}\}$, since for $\gamma_1(x_1) = (x_1 > 0)$ and $\gamma_2(x_2) = (x_2 > 0)$ one has $\varphi(x_1, x_2) \leftrightarrow \gamma_1(x_1) \wedge \gamma_2(x_2)$.

If $\mathcal{M}$ admits quantifier-elimination (that is, for every $\varphi(\vec{x})$, there exists a quantifier-free $\psi(\vec{x})$ such that $\mathcal{M} \models \forall \vec{x}\ \varphi(\vec{x}) \leftrightarrow \psi(\vec{x})$), the definition of $\varphi \sim_{\mathcal{M}} P$ can be restated as the existence of a family $\gamma_i^j(\vec{x}_{B_j})$, $i = 1, \ldots, k$, $j = 1, \ldots, m$ of quantifier-free formulae, where $P = \{B_1, \ldots, B_m\}$, and $\vec{x}_{B_j}$ is the subtuple of $\vec{x}$ consisting of x_ls with $l \in B_j$, such that

$$\mathcal{M} \models \varphi(\vec{x}) \leftrightarrow \bigvee_{i=1}^{k} (\gamma_i^1(\vec{x}_{B_1}) \wedge \cdots \wedge \gamma_i^m(\vec{x}_{B_m})) \tag{1}$$

The set of partitions of a set X forms a lattice $\mathrm{Part}(X)$ where $P \sqsubseteq P'$ iff P is a refinement of P'. The meet operation $\sqcap$ in $\mathrm{Part}(X)$ is defined in the following way: the blocks of $P_1 \sqcap P_2$ are the nonempty sets of the form $B_1 \cap B_2$, where B_i is a block of P_i, $i = 1, 2$.

Given $\mathcal{M}$ with a universe U, a formula $\varphi(x_1, \ldots, x_n)$, the set *definable* by φ is the set $\varphi(\mathcal{M}) = \{\vec{a} \in U^n \mid \mathcal{M} \models \varphi(\vec{a})\}$. The *orthographic dimension* of definable sets was defined as follows [4]. Let $\mathcal{P}_{\mathcal{M}}(\varphi) = \{P \in \mathrm{Part}(\{x_1, \ldots, x_n\}) \mid \varphi \sim_{\mathcal{M}} P\}$. Then the orthographic dimension of $\varphi(\mathcal{M})$ is the maximal size of a block of a minimal element $P \in \mathcal{P}_{\mathcal{M}}(\varphi)$ (that is, any $P' \sqsubset P$ does not belong to $\mathcal{P}_{\mathcal{M}}(\varphi)$). In other words, one takes a partition that φ conforms to, and that cannot be further refined while preserving this property, and defines the orthographic dimension to be the maximum size of a block of such a partition.

The problem with this definition is that it is not clear whether one can have two minimal elements $P_1, P_2 \in \mathcal{P}_{\mathcal{M}}(\varphi)$ with different maximum block sizes. We prove that this is impossible. In fact, we prove a more general result, implying that $\mathcal{P}_{\mathcal{M}}(\varphi)$ always has a unique minimal element.

Theorem 1 *Let $\mathcal{M}$ be an L-structure, and $\varphi(x_1, \ldots, x_n)$ be an L-formula. Assume that $\varphi \sim_{\mathcal{M}} P_1$ and $\varphi \sim_{\mathcal{M}} P_2$ for $P_1, P_2 \in \mathrm{Part}(\{x_1, \ldots, x_n\})$. Then $\varphi \sim_{\mathcal{M}} P_1 \sqcap P_2$.*

Corollary 1 *For any $\mathcal{M}$ and φ as above, the set $\mathcal{P}_{\mathcal{M}}(\varphi) = \{P \mid \varphi \sim_{\mathcal{M}} P\}$ is nonempty and has a unique least element.*

Thus, the concept of the orthographic dimension is well-defined for any constraint-definable set.

3 Proof of the theorem

The structure of the proof is as follows. We start by proving an easy case of two partitions of the form $A, B \cup C$ and $A \cup B, C$ (Lemma 2), which is preceded by Lemma 1 establishing a certain kind of unambiguous representations of formulae. After that we use Lemma 2 to prove a general case of two two-block partitions (Lemma 4), which again relies on certain unambiguous representations (Lemma 3). With this preparatory work, we conclude the proof by induction: the base case is provided by Lemma 4, and the induction step follows from Lemma 2.

We now fix a structure $\mathcal{M}$ and a formula $\varphi(x_1, \ldots, x_n)$. For a subset B of $\{x_1, \ldots, x_n\}$, we let $\vec{x}_B$ be the subtuple of $(x_1, \ldots, x_n)$ that consists of variables in B. For blocks of partitions, we sometimes write AB instead of $A \cup B$ and subsequently $\vec{x}_{AB}$, or $\vec{x}_A\vec{x}_B$. When $\mathcal{M}$ is clear from the context, we write $\varphi \sim P$.

Suppose we have a partition $P = \{B_1, \ldots, B_m\}$ and let $\varphi \sim P$. By putting φ in DNF we assume that it is of the form

$$\bigvee_{i=1}^{N} (\alpha_i^1(\vec{x}_{B_1}) \wedge \cdots \wedge \alpha_i^m(\vec{x}_{B_m})) \tag{2}$$

We call the above representation *unambiguous in B_l* if $\mathcal{M} \models \neg\exists\vec{y}(\alpha_i^l(\vec{y}) \wedge \alpha_j^l(\vec{y}))$ for $i \neq j$. That is, formulae α_j^l are mutually exclusive: no $\vec{y}_{B_l}$ can satisfy more than one α_j^l.

Lemma 1 *Let φ and P be as above, with $\varphi \sim P$. Fix $1 \leq l \leq m$. Then φ admits a representation unambiguous in l. Furthermore, if in the representation (2) every α_i^l conforms to some fixed partition P_l on B_l, then there is a representation unambiguous in l in which all formulae that depend on $\vec{x}_{B_l}$ also conform to P_l.*

Proof. Make each disjunct in (2) into 2^{N-1} ones by replacing each α_i^l with all the possible conjunctions $\alpha_i^l \wedge \bigwedge_{j \neq i} (\alpha_j^l)^{\sigma(j)}$, where σ is a map from an $N-1$ element set to $\{+, -\}$ and $(\alpha)^+ = \alpha$, $(\alpha)^- = \neg\alpha$. Then use distributivity to ensure that no two identical formulae in the variables $\vec{x}_{B_l}$ are present in the DNF. $\square$

Now, suppose a partition P is given, and let C be a union of some of its blocks. Then P naturally induces a partition P_C on C. Suppose we have a formula $\psi(\vec{x}_C)$. If it conforms to P_C, we shall say that it conforms to P, in order to simplify the notation.

Lemma 2 *Let $P = \{A, B, C\}$ be a partition. Consider two partitions P_A on A and P_C on C. Let P' be the partition whose blocks are those of P_A and $B \cup C$, and let P'' be the partition whose blocks are those of P_C and $A \cup B$. Assume that $\varphi \sim P'$ and $\varphi \sim P''$. Then $\varphi \sim P' \sqcap P''$.*

Proof. Note that the blocks of $P' \sqcap P''$ are those of P_A, P_C and B. By the preceding lemma, we write φ in two equivalent ways, as

$$\bigvee_{i=1}^{N} (\alpha_i(\vec{x}_A) \wedge \beta_i(\vec{x}_B \vec{x}_C))$$

and as

$$\bigvee_{j=1}^{M} (\gamma_j(\vec{x}_A \vec{x}_B) \wedge \delta_j(\vec{x}_C)) ,$$

with $\exists \vec{x}_A . \alpha_{i_1}(\vec{x}_A) \wedge \alpha_{i_2}(\vec{x}_A)$ being false for $i_1 \neq i_2$ and with α_is conforming to P_A and δ_js conforming to P_C. We now construct for each $1 \leq i \leq N$ and $1 \leq j \leq M$ a formula $\psi_{ij}(\vec{x}) \equiv \alpha_i(\vec{x}_A) \wedge \delta_j(\vec{x}_C) \wedge \chi_{ij}(\vec{x}_B)$ where $\chi_{ij}(\vec{x}_B)$ is $\exists \vec{z} (\alpha_i(\vec{z}) \wedge \gamma_j(\vec{z}\vec{x}_B))$. We claim that φ is equivalent to $\bigvee_{ij} \psi_{ij}$. Clearly this will suffice as subformulae of ψ_{ij}s conform to partitions P_A and P_C.

One must now show $\mathcal{M} \models \forall \vec{x} (\varphi(\vec{x}) \leftrightarrow \bigvee_{ij} \psi_{ij}(\vec{x}))$. If $\varphi(\vec{a})$, we have that for some i, j, $\alpha_i(\vec{a}_A) \wedge \beta_i(\vec{a}_B \vec{a}_C)$ is true and $\gamma_j(\vec{a}_A \vec{a}_B) \wedge \delta_j(\vec{a}_C)$, and thus $\psi_{ij}(\vec{a})$ is true. Conversely, if $\psi_{ij}(\vec{a})$ is true, we have that for some $\vec{u}$ of the same length as $\vec{x}_A$, it is the case that $\alpha_i(\vec{u})$ is true and $\gamma_j(\vec{u}\vec{a}_B)$ is true; furthermore, $\alpha_i(\vec{a}_A)$ and $\delta_j(\vec{a}_C)$ are true. We thus obtain that $\varphi(\vec{u}\vec{a}_B\vec{a}_C)$ is true. Hence, for some index $1 \leq l \leq N$, we have $\alpha_l(\vec{u})$ and $\beta_l(\vec{a}_B \vec{a}_C)$. Since we know that $\alpha_i(\vec{u})$ is true, by unambiguity of the first representation in A-variables, we have $l = i$; hence, $\beta_i(\vec{a}_B \vec{a}_C)$ is true and thus $\varphi(\vec{a})$ is true. This proves the lemma. □

Next, we need a variant of Lemma 1 that involves two different formulae.

Lemma 3 *Let $\Phi(\vec{x}, \vec{y})$ be $\bigvee_{i=1}^{n} (\zeta_i(\vec{x}) \wedge \psi_i(\vec{y}))$ and let $\Psi(\vec{x}, \vec{z})$ be $\bigvee_{i=1}^{m} (\chi_i(\vec{x}) \wedge \xi_i(\vec{z}))$ where $\vec{x}$ and $\vec{y}$ (and likewise $\vec{x}$ and $\vec{z}$) have no variables in common. Then there exist formulae $\alpha_i(\vec{x})$, $\beta_i(\vec{y})$, $\gamma_i(\vec{z})$, $i = 1, \ldots, k$, such that*

$$\mathcal{M} \models \Phi(\vec{x}, \vec{y}) \leftrightarrow \bigvee_{i=1}^{k} (\alpha_i(\vec{x}) \wedge \beta_i(\vec{y})) ,$$

$$\mathcal{M} \models \Psi(\vec{x}, \vec{y}) \leftrightarrow \bigvee_{i=1}^{k} (\alpha_i(\vec{x}) \wedge \gamma_i(\vec{y})) ,$$

and furthermore $\mathcal{M} \models \neg \exists \vec{x} (\alpha_i(\vec{x}) \wedge \alpha_j(\vec{x}))$ for any $i \neq j$.

Proof. Assume without loss of generality that n, $m > 0$. Let Σ be the set of all 2^{n+m} mappings from $\{1, \ldots, n, n+1, \ldots, n+m\}$ to $\{+, -\}$. Let $\delta^+ = \delta$ and $\delta^- = \neg\delta$ for any formula δ. For each $\sigma \in \Sigma$, let

$$\alpha_\sigma(\vec{x}) \equiv \bigwedge_{i=1}^{n} \zeta_i^{\sigma(i)}(\vec{x}) \wedge \bigwedge_{j=1}^{m} \chi_j^{\sigma(n+j)}(\vec{x}) .$$

Then Φ is equivalent to the following formula $\Phi'(\vec{x}, \vec{y})$:

$$\bigvee_{i=1}^{n} \left(\bigvee_{\sigma \in \Sigma, \sigma(i) = +} (\alpha_\sigma(\vec{x}) \wedge \psi(\vec{y})) \right) \vee \left(\bigvee_{\sigma \in \Sigma, \sigma(i) = -} (\alpha_\sigma(\vec{x}) \wedge F(\vec{y})) \right) ,$$

where $F(\vec{y})$ is any unsatisfiable formula (e.g., $\vec{y} \neq \vec{y}$). Thus, Φ' is equivalent to the formula of the form

$$\bigvee_{\sigma \in \Sigma} (\alpha_\sigma(\vec{x}) \wedge \beta_\sigma(\vec{y})) ,$$

where each β_σ is either a disjunction of some ψ_is, or an unsatisfiable formula. Clearly, any α_σ and $\alpha_{\sigma'}$ for $\sigma \neq \sigma'$ are inconsistent. The proof that Ψ admits a similar representation with the same family $\{\alpha_\sigma\}$ is identical. □

We next consider the key case of two two-block partitions, e.g., $\{AB, CD\}$ and $\{AC, BD\}$.

Lemma 4 *Let $\Phi_1(\vec{x},\vec{y},\vec{u},\vec{v})$ be $\bigvee_{i=1}^n(\alpha_i(\vec{x},\vec{y}) \wedge \beta_i(\vec{u},\vec{v}))$, and let $\Phi_2(\vec{x},\vec{y},\vec{u},\vec{v})$ be $\bigvee_{j=1}^m(\gamma_j(\vec{x},\vec{u}) \wedge \delta_j(\vec{y},\vec{v}))$, where $\vec{x},\vec{y},\vec{u},\vec{v}$ are pairwise disjoint nonempty tuples of variables. Assume that $\mathcal{M} \models \Phi_1 \leftrightarrow \Phi_2$. Then there exists a collection of formulae $\psi_k(\vec{x}), \chi_k(\vec{y}), \xi_k(\vec{u}), \rho_k(\vec{v})$ such that*

$$\mathcal{M} \models \Phi_1 \leftrightarrow \bigvee_k (\psi_k(\vec{x}) \wedge \chi_k(\vec{y}) \wedge \xi_k(\vec{u}) \wedge \rho_k(\vec{v}))\ .$$

Proof. First, we assume, in view of Lemma 1, that the representation for Φ_1 is unambiguous in $\vec{x}$, $\vec{y}$ (that is, $\mathcal{M} \models \neg\exists\vec{x}\vec{y}(\alpha_i(\vec{x},\vec{y}) \wedge \alpha_{i'}(\vec{x},\vec{y}))$ for $i \neq i'$). Consider $\Phi_1'(\vec{x},\vec{y},\vec{v}) \equiv \exists\vec{u}(\Phi_1(\vec{x},\vec{y},\vec{u},\vec{v}))$. Let $\gamma_j'(\vec{x})$ be $\exists\vec{u}\gamma_j(\vec{x},\vec{u})$ and let $\beta_i'(\vec{v})$ be $\exists\vec{u}\beta_i(\vec{u},\vec{v})$. Then Φ_1' is equivalent to both

$$\bigvee_{i=1}^{m}(\alpha_i(\vec{x},\vec{y}) \wedge \beta_i'(\vec{v}))$$

and

$$\bigvee_{j=1}^{m}(\gamma_j'(\vec{x}) \wedge \delta_j(\vec{y},\vec{v}))\ .$$

Applying Lemma 2, we find a collection of formulae $\tilde{\psi}_l(\vec{x})$, $\tilde{\chi}_l(\vec{y})$, $\tilde{\rho}_l(\vec{v})$ such that Φ_1' is equivalent to $\bigvee_l(\tilde{\psi}_l(\vec{x}) \wedge \tilde{\chi}_l(\vec{y}) \wedge \tilde{\rho}_l(\vec{v}))$.

Applying the same argument to $\Phi_1''(\vec{x},\vec{y},\vec{u}) \equiv \exists\vec{v}(\Phi_1(\vec{x},\vec{y},\vec{u},\vec{v}))$, we find a collection of formulae $\hat{\psi}_s(\vec{x})$, $\hat{\chi}_s(\vec{y})$, $\hat{\xi}_s(\vec{u})$ such that Φ_1'' is equivalent to $\bigvee_s(\hat{\psi}_l(\vec{x}) \wedge \hat{\chi}_s(\vec{y}) \wedge \hat{\xi}_s(\vec{u}))$. Now using Lemma 3, we find a collection of formulae $\psi_k(\vec{x})$, $\chi_k'(\vec{y})$, $\chi_k'(\vec{y})$, $\xi_k(\vec{u})$, $\rho_k(\vec{v})$ such that Φ_1' is equivalent to

$$\bigvee_k \psi_k(\vec{x}) \wedge \chi_k'(\vec{y}) \wedge \xi_k(\vec{v})\ ,$$

Φ_1'' is equivalent to

$$\bigvee_k \psi_k(\vec{x}) \wedge \chi_k''(\vec{y}) \wedge \rho_k(\vec{u})\ ,$$

and $\mathcal{M} \models \neg\exists\vec{x}(\psi_i(\vec{x}) \wedge \psi_j(\vec{x}))$ for $i \neq j$ (this is because the formulae β_i and γ_j produced in the proof of Lemma 3 conform to the same partitions as the formulae in the $\vec{y}$ and $\vec{z}$ variables in the original DNF formulae).

Now let $\chi_k(\vec{y}) \equiv \chi_k'(\vec{y}) \wedge \chi_k''(\vec{y})$ and define $\Psi(\vec{x},\vec{y},\vec{u},\vec{v})$ to be $\bigvee_k(\psi_k(\vec{x}) \wedge \chi_k(\vec{y}) \wedge \xi_k(\vec{u}) \wedge \rho_k(\vec{v}))$. We now claim that $\mathcal{M} \models \Phi_1(\vec{a},\vec{b},\vec{c},\vec{d}) \leftrightarrow \Psi(\vec{a},\vec{b},\vec{c},\vec{d})$ for any $\vec{a},\vec{b},\vec{c},\vec{d}$ of appropriate arity over U.

First, if $\Phi_1(\vec{a},\vec{b},\vec{c},\vec{d})$ holds, then for some k_1, k_2, both $(\psi_{k_1}(\vec{a}) \wedge \chi_{k_1}(\vec{b}) \wedge \xi_{k_1}(\vec{c}))$ and $(\psi_{k_2}(\vec{a}) \wedge \chi_{k_2}(\vec{b}) \wedge \rho_{k_2}(\vec{d}))$ hold, which implies $k_1 = k_2$ and thus that $\Psi(\vec{a},\vec{b},\vec{c},\vec{d})$ holds.

Next, let m be the length of $\vec{x}$. Define L^* to be an extension of L with m new constant symbols, and let $\mathcal{M}^{\vec{a}}$ be the L^*-expansion of $\mathcal{M}$, where the extra constants are interpreted as $\vec{a}$. For every

L-formula $\chi(\vec{x}, \vec{s})$, where $\vec{x}$ is of length m, let $\chi^*(\vec{s})$ be an L^* formula in which each free occurrence of a variable x_i in $\vec{x}$ is replaced by the corresponding constant symbol in L^*. Then $\mathcal{M} \models \chi(\vec{a}, \vec{e})$ iff $\mathcal{M}^{\vec{a}} \models \chi^*(\vec{e})$.

Now assume that $\Psi(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ holds. Then $\exists u(\Phi_1(\vec{a}, \vec{b}, \vec{u}, \vec{d}))$ holds, and thus $\Phi_1(\vec{a}, \vec{b}, \vec{c}_0, \vec{d})$ holds in $\mathcal{M}$ for some $\vec{c}_0$. Similarly, $\Phi_1(\vec{a}, \vec{b}, \vec{c}, \vec{d}_0)$ holds in $\mathcal{M}$ for some $\vec{d}_0$.

The formula $\Phi_1^*(\vec{y}, \vec{u}, \vec{v})$ is equivalent, over $\mathcal{M}^{\vec{a}}$, to

$$\bigvee_i (\alpha_i^*(\vec{y}) \wedge \beta_i(\vec{u}, \vec{v}))$$

and

$$\bigvee_j (\gamma_j^*(\vec{u}) \wedge \delta_j(\vec{y}, \vec{v}))\ .$$

Applying Lemma 2 to $\mathcal{M}^{\vec{a}}$, we find that Φ_1^* is equivalent to a formula of the form

$$\bigvee_r (\mu_r^*(\vec{y}) \wedge \eta_r^*(\vec{u}) \wedge \nu_r^*(\vec{v}))\ ,$$

and we may assume, by Lemma 1, that this representation is unambiguous in $\vec{y}$. Since $\Phi_1^*(\vec{b}, \vec{c}_0, \vec{d})$ holds in $\mathcal{M}^{\vec{a}}$, we obtain $(\mu_{r_1}^*(\vec{b}) \wedge \eta_{r_1}^*(\vec{c}_0) \wedge \nu_{r_2}^*(\vec{d}))$ for some r_1. Since $\Phi_1^*(\vec{b}, \vec{c}, \vec{d}_0)$ holds in $\mathcal{M}^{\vec{a}}$, we obtain $(\mu_{r_2}^*(\vec{b}) \wedge \eta_{r_2}^*(\vec{c}) \wedge \nu_{r_2}^*(\vec{d}_0))$ for some r_2. Using unambiguity, we conclude $r_1 = r_2$, and thus $(\mu_{r_1}^*(\vec{b}) \wedge \eta_{r_1}^*(\vec{c}) \wedge \nu_{r_2}^*(\vec{d}))$ holds in $\mathcal{M}^{\vec{a}}$. Hence, $\Phi_1^*(\vec{b}, \vec{c}, \vec{d})$ holds in $\mathcal{M}^{\vec{a}}$, and thus $\Phi(\vec{a}, \vec{b}, \vec{c}, \vec{d})$ holds in $\mathcal{M}$, which concludes the proof. $\square$

Using Lemma 4, we prove the following result, which is the basis of the main induction argument. Let $D_1, \ldots, D_k$ enumerate the blocks of partition $P_1 \sqcap P_2$. Let $P_{12}(i)$ be the partition whose two blocks are D_i and $\cup_{j \neq i} D_j$.

Lemma 5 *For every i, $\varphi \sim P_{12}(i)$.*

Proof. Let D_i arise as $B \cap C$, where B is a block in P_1 and C is a block in P_2. Let $A = B - C$, $E = C - B$ and let F be the complement of $B \cup C$ (some of these sets may be empty). φ therefore conforms to $\{D_i A, EF\}$ and to $\{D_i E, AF\}$. If two or more sets among A, D_i, E, F are empty, the result is immediate. If one of these sets is empty, the result follows from Lemma 2. If all four are nonempty, it follows from Lemma 4. $\square$

We now complete the proof of the main theorem by a simple induction. Let $P(i)$ be the partition whose blocks are $D_1, \ldots, D_i$, and $D_{i+1} \cup \cdots \cup D_k$. We show that $\varphi \sim P(i)$ for all i; hence, $\varphi \sim P(k) = P_1 \sqcap P_2$. The base case, $\varphi \sim P(1)$, follows from Lemma 5. Suppose $\varphi \sim P(i)$; we have to show $\varphi \sim P(i+1)$. Let $A = D_1 \cup \cdots \cup D_i$, $B = D_{i+1}$ and $C = D_{i+2} \cup \cdots \cup D_k$. We then have that φ conforms to the partition with blocks A and $B \cup C$, as well as that with blocks B and $A \cup C$; furthermore, it conforms to the subpartition $D_1, \ldots, D_i$ on A. From Lemma 2 it then follows that φ conforms to the partition with blocks $D_1, \ldots, D_i$, $B = D_{i+1}$, C, that is, to $P(i+1)$. This completes the proof. $\square$

References

[1] J. Chomicki, D. Goldin and G. Kuper. Variable independence and aggregation closure. In *Principles of Database Systems (PODS'96)*, ACM Press, pages 40–48.

[2] A. Fordan, R. Yap. Early projection in CLP(R). In *Principles and Practice of Constraint Programming (CP'98)*, Springer LNCS vol. 1520, pages 177–191.

[3] S. Grumbach, P. Rigaux, L. Segoufin. The DEDALE system for complex spatial queries. In *SIGMOD'98*, ACM Press, pages 213–224.

[4] S. Grumbach, P. Rigaux, L. Segoufin. On the orthographic dimension of constraint databases. In *International Conference on Database Theory (ICDT'99)*, Springer LNCS vol. 1540, pages 199–216.

[5] P. Kanellakis, G. Kuper, and P. Revesz. Constraint query languages. *Journal of Computer and System Sciences*, 51 (1995), 26–52.

[6] G. Kuper, L. Libkin, and J. Paredaens (editors). *Constraint Databases*. Springer-Verlag, 2000.

[7] L. Libkin. Variable independence, quantifier elimination, and constraint representations. In *International Colloquium on Automata, Languages and Programming (ICALP'00)*, Springer LNCS vol. 1853, pages 260–271.