Algebraic Optimization of Database Queries with Preferences*

Radim Nedbal

4th year of PGS, email: radned@seznam.cz
Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, CTU
advisor: Ing. Július Štuller, CSc., Institute of Computer Science, ASCR

Abstract. The paper resumes a logical framework for formulating preferences and proposes their embedding into relational algebra through a single *preference operator* parameterized by a set of user preferences of sixteen various kinds, inclusive of *ceteris paribus* preferences, and returning only the most preferred subsets of its argument relation. Most importantly, conflicting set of preferences is permitted and preferences between sets of elements can be expressed.

Formal foundation for algebraic optimization, applying heuristics like *push preference*, also is provided: abstract properties of the preference operator and a variety of algebraic laws describing its interaction with other relational algebra operators are presented.

Abstrakt. Příspěvek shrnuje logické přístupy k vyjadřování preferencí a navrhuje jejich začlenění do relační algebry pomocí jediného *preferenčního operátoru* parametrizovaného množinou až šestnácti různých druhů preferencí, včetně preferencí *ceteris paribus*, a vracejícího nejpreferovanější podmnožiny relace, která je v jeho argumentu. Podstatné je, že koncept zahrnuje preference, které mohou být navzájem v konfliktu a umožňuje reprezentovat i preference mezi množinami.

Navrženy jsou také základní principy algebraické optimalizace jako je např. propagování preferenčního operátoru výrazem relační algebry směrem ke vstupním relacím. Podobné heuristické metody vycházejí z algebraických vztahů operací relační algebry – v tomto případě preferenčního operátoru, které jsou také prezentovány.

1 Introduction

If users have requirements that are to be satisfied completely, their database queries are characterized by *hard constraints*, delivering exactly the required objects if they exist and otherwise empty result. This is how traditional database query languages treat all the requirements on the data. However, requirements can be understood also in the sense of wishes: in case they are not satisfied, database users are usually prepared to accept worse alternatives and their database query is characterized by *soft constraints*. Requirements of the latter type are called preferences.

Building on a logical framework for formulating preferences and their embedding into relational algebra (RA) through a single *preference operator*, introduced in [10] to combat

*This has been supported by the project 1ET100300419 of the Program Information Society (of the Thematic Program II of the National Research Program of the Czech Republic) "Intelligent Models, Algorithms, Methods and Tools for the Semantic Web Realization" and by the Institutional Research Plan AV0Z10300504 "Computer Science for the Information Society: Models, Algorithms, Applications".

the *empty result* and the *flooding effects*, this paper presents an approach to algebraic optimization of relational queries with various kinds of preferences. The preference operator selects from its argument relation the *best-matching alternatives* with regard to user preferences, but *nothing worse*.[1] Preferences are specified using a propositional logic notation and their semantics is related to that of a disjunctive logic program. The language for expressing preferences i) is declarative, ii) includes various kinds of preferences, iii) is rich enough to express preferences between sets of elements, iv) and has an intuitive, well defined semantics allowing for conflicting preferences.

In Sect. 2, the above mentioned framework for formulating preferences and in Sect. 3 an approach to their embedding into RA are revisited. Presenting a variety of algebraic laws that describe interaction with other RA operators to provide a formal foundation for algebraic optimization, Sect. 4 provides the main contribution of this paper. A brief overview of related work in Sect. 5 and conclusions in Sect. 6 end this paper. All the nontrivial proofs are given.

To improve the readability, $\succeq(x, y) \wedge \neg \succeq(y, x)$ and $\succeq(x, y) \wedge \succeq(y, x)$ is substituted by $\succ(x, y)$ and $=(x, y)$, respectively.

2 User Preferences

A user preference is expressed by a preference statement, e.g. "a is preferred to b", or symbolically by an appropriate preference formula. Preference formulas comprise a simple declarative language for expressing preferences. To capture its declarative aspects, model-theoretic semantics is defined: considering a set of states of affaires S and a set $W = 2^S$ of all its subsets – worlds, if $\mathcal{M} = \langle W, \succeq \rangle$ is an order $\succeq$ on W such that $w \succeq w'$ holds for some words w, w' from W, then $\mathcal{M}$ is termed a *preference model* of $w > w'$ – a preference of the world w over the world w', which we express symbolically as $\mathcal{M} \models w > w'$.

The basic differentiation between preferences is based on notions of optimism and pessimism. Defining a-world as a world in which a occurs, if we are optimistic about a and pessimistic about b for example, we expect some a-world to precede at least one b-world in each preference model of a preference statement "a is preferred to b". This kind of preference is called *opportunistic*. By contrast, if we are pessimistic about a and optimistic about b, we expect every a-world to precede each b-world in each preference model of a preference statement "a is preferred to b". This kind of preference is called *careful.* Alternatively, we might be optimistic or pessimistic about both a and b. Then we expect some a world to precede each b-world or each a-world to precede some b-world in each preference model of a preference statement "a is preferred to b". This kind of preference is called *locally optimistic* or *locally pessimistic*, respectively. Locally optimistic, locally pessimistic, opportunistic and careful preferences are symbolically expressed by preference formulas of the form: $a\ {}^{M}{>}^{M}\ b$, $a\ {}^{m}{>}^{m}\ b$, $a\ {}^{M}{>}^{m}\ b$, and $a\ {}^{m}{>}^{M}\ b$, respectively.

Also, we distinguish between strict and non-strict preferences. For example, if w precedes w' strictly in a preference model, then we strictly prefer w to w'.

In addition, we distinguish between preferences with and without *ceteris paribus* pro-

[1] A similar concept was proposed independently by Kießling et al. [6, 7] and Chomicki et al. [2] and, in a more restricted form, by Börzsönyi et al. [1] (for more detail refer to Sect. 5).

viso – a notion introduced by von Wright [11] and generalized by Doyle and Wellman [3] by means of contextual equivalence relation – an equivalence relation on W.[2] For example, a preference model of a preference statement "a is *carefully* preferred to b *ceteris paribus*" is such an order on W that a-worlds precede b-worlds in the same contextual equivalence class. Specifically, the preference statement "I prefer playing tenis to playing golf ceteris paribus" might express by means of an contextual equivalence that I prefer playing tenis to playing golf only if the context of weather is the same, i.e., it is not true that I prefer playing tenis in strong winds to playing golf during a sunny day.

Next, we revisit the basic definitions introducing syntax and model-theoretic semantics of the language for expressing user preferences:

Definition 1 (Language)**.** Given a finite set of propositional variables $p, q, \ldots$, the set L_0 of *propositional formulas* and the set L of *preference formulas* is defined as the smallest set satisfying the following:

$L_0 \ni \varphi, \psi\colon\ p \mid (\varphi \wedge \psi) \mid \neg\varphi$
$L \ni \Phi, \Psi\colon\ \varphi\ {}^{x}{>}^{y}\ \psi \mid \varphi\ {}^{x}{\geq}^{y}\ \psi \mid \neg\Phi \mid (\Phi \wedge \Psi) \quad \text{for } x, y \in \{m, M\}$

If we identify propositional variables with tuples over a relation schema R, then the elements of L are termed *preference formulas over* R. A relation instance $I(R)$, i.e., a set of tuples over R, creates a *world* w, an element of a set W.

The preference model is defined so that any set of (possibly conflicting) preferences is consistent: the partial pre-order, i.e., a binary relation which is reflexive and transitive, in the definition of the preference model, enables to express some kind of conflict by incomparability:

Definition 2 (Preference model)**.** A preference model $\mathcal{M} = \langle W, \succeq \rangle$ over a relation schema R is a couple in which W is a set of worlds, relation instances of R, and $\succeq$ is a *partial pre-order* over W, the *preference relation* over R.

A set of user preferences of various kinds can by represented symbolically by a *preference specification*, which corresponds to an appropriate complex preference formula in the above defined language.

Definition 3 (Preference specification)**.** Let R be a relation schema and $\mathcal{P}_\rhd$ a set of preference formulas over R of the form $\{\varphi_i \rhd \psi_i : i = 1, \ldots, n\}$. A preference specification $\mathcal{P}$ over R is a tuple $\langle \mathcal{P}_\rhd | \rhd \in \{\ {}^{x}{>}^{y},\ {}^{x}{\geq}^{y}\ | x, y \in \{m, M\}\} \rangle$, and $\mathcal{M}$ is its model, i.e., a *preference specification model*, iff it models all elements $\mathcal{P}_\rhd$ of the tuple:

$$\mathcal{M} \models \mathcal{P}_\rhd \iff \forall (\varphi_i \rhd \psi_i) \in \mathcal{P}_\rhd : \mathcal{M} \models \varphi_i \rhd \psi_i\ .$$

[2] As it has been shown [5] that any preference with contextual equivalence specification can be expressed by a set of preferences without contextual specification, we can restrict ourselves only to preferences without ceteris paribus proviso.

3 Preference Operator

To embed preferences into RQL, the *preference operator* $\omega_{\mathcal{P}}$ returning only the best sets of tuples in the sense of user preferences $\mathcal{P}$ is defined:

Definition 4 (Preference operator)**.** If R is a relation schema, $\mathcal{P}$ a preference specification over R, and $\mathscr{M}$ the set of its models; then the preference operator $\omega_{\mathcal{P}}$ is defined for all instances $I(R)$ of R as follows:

$$\omega_{\mathcal{P}}(I(R)) = \{w \in W \mid w \subseteq I(R) \wedge \exists \mathcal{M}_k = \langle W, \succeq_k \rangle \in \mathscr{M} \text{ s.t. } \forall w' \in W : \\ w' \subseteq I(R) \wedge \succeq_k (w', w) \Rightarrow \ \succeq_k (w, w')\} \ .$$

Remark 1 (Preference operator notation)*.* To be precise, we should write $\omega_{\mathcal{P}}(2^{I(R)})$ instead of $\omega_{\mathcal{P}}(I(R))$. Thus it makes sense to write $\omega_{\mathcal{P}}(\{a, b, c\})$, where the argument of preference operator is a set of elements a, b, and c.

3.1 Basic Properties.

The following propositions are essential for investigation of algebraic properties describing interaction of the preference operator with other RA operations:

Proposition 1. *Given a relation schema R and a preference specification $\mathcal{P}$ over R, for all instances $I(R)$ of R the following properties hold:*

$$\begin{aligned} \omega_{\mathcal{P}}(I(R)) &\subseteq 2^{I(R)} \ , \\ \omega_{\mathcal{P}}\left(\omega_{\mathcal{P}}(I(R))\right) &= \omega_{\mathcal{P}}(I(R)) \ , \\ \omega_{\mathcal{P}_{\text{empty}}}(I(R)) &= 2^{I(R)} \ , \end{aligned}$$

where $\mathcal{P}_{\text{empty}}$ is the empty preference specification, i.e., containing no preference.

Preference operator is not *monotone* or *antimonotone* with respect to its relation argument. However, partial antimonotonicity holds:

Proposition 2 (Partial antimonotonicity)**.** *Given a relation schema R and a preference specification $\mathcal{P}$ over R, for all instances $I(R), I'(R)$ of R the following property holds:*

$$I(R) \subseteq I'(R) \Rightarrow 2^{I(R)} \cap \omega_{\mathcal{P}}(I'(R)) \subseteq \omega_{\mathcal{P}}(I(R)) \ .$$

Proof. Assume $w \in 2^{I(R)} \cap \omega_{\mathcal{P}}(I'(R))$. It follows that $w \subseteq I(R)$ and from the definition (Def. 4) of preference operator $w \subseteq I'(R) \wedge \exists \mathcal{M}_k \in \mathscr{M}$ s.t. $\forall w' \in W : w' \subseteq I'(R) \wedge \succeq_k (w', w) \Rightarrow \succeq_k (w, w')$. As $I(R) \subseteq I'(R)$, we can conclude that $\exists \mathcal{M}_k \in \mathscr{M}$ s.t. $\forall w' \in W : w' \subseteq I(R) \wedge \succeq_k (w', w) \Rightarrow \succeq_k (w, w')$, which together with $w \subseteq I(R)$ implies $w \in \omega_{\mathcal{P}}(I(R))$. $\square$

The following theorem enables to reduce cardinality of an argument relation of the preference operator without changing the return value:

Theorem 1 (Reduction). *Given a relation schema R, a preference specification $\mathcal{P}$ over R, for all instances $I(R), I'(R)$ of R the following property holds:*

$$I(R) \subseteq I'(R) \wedge \omega_{\mathcal{P}}(I'(R)) \subseteq 2^{I(R)} \Rightarrow \omega_{\mathcal{P}}(I(R)) = \omega_{\mathcal{P}}(I'(R)) \ .$$

Proof. $\subseteq$: Assume $w \in \omega_{\mathcal{P}}(I(R))$. Then, it follows from the definition of the preference operator $w \subseteq I(R) \wedge \exists \mathcal{M}_k \in \mathscr{M}$ s.t. $\forall w' \in W : w' \subseteq I(R) \wedge \succeq_k (w', w) \Rightarrow \succeq_k (w, w')$. The assumption $\omega_{\mathcal{P}}(I'(R)) \subseteq 2^{I(R)}$ implies $\forall w' \in 2^{I'(R)} - 2^{I(R)} : \neg \succeq_k (w', w)$, and we can conclude $\exists \mathcal{M}_k \in \mathscr{M}$ s.t. $\forall w' \in W : w' \subseteq I'(R) \wedge \succeq_k (w', w) \Rightarrow \succeq_k (w, w')$, which together with the assumption $I(R) \subseteq I'(R)$ implies $w \in \omega_{\mathcal{P}}(I'(R))$.

$\supseteq$: Immediately follows from Prop. 2. $\square$

The following theorem ensures that the empty query result effect is successfully eliminated:

Theorem 2 (Non-emptiness). *Given a relation schema R, a preference specification $\mathcal{P}$ over R, then for every finite, nonempty instance $I(R)$ of R, $\omega_{\mathcal{P}}(I(R))$ is nonempty.*

3.2 Multidimensional Composition.

In multidimensional composition, we have a number of preference specifications defined over several relation schemas, and we define preference specification over the Cartesian product of those relations: the most common ways are Pareto and lexicographic composition.

Definition 5 (Pareto and lexicographic composition). Given two relation schemas R_1 and R_2, preference specifications $\mathcal{P}_1$ over R_1 and $\mathcal{P}_2$ over R_2, and their sets of models $\mathscr{M}_1$ and $\mathscr{M}_2$, the *Pareto composition* $P(\mathcal{P}_1, \mathcal{P}_2)$ and the *lexicographic composition* $L(\mathcal{P}_1, \mathcal{P}_2)$ of $\mathcal{P}_1$ and $\mathcal{P}_2$ is a preference specification $\mathcal{P}_0$ over the Cartesian product $R_1 \times R_2$, whose set of models $\mathscr{M}_0$ is defined as:

$$\forall \mathcal{M}_m = \langle W_1 \times W_2, \succeq_m \rangle \in \mathscr{M}_0, \exists \mathcal{M}_k = \langle W_1, \succeq_k \rangle \in \mathscr{M}_1, \exists \mathcal{M}_l = \langle W_2, \succeq_l \rangle \in \mathscr{M}_2 \text{ s.t.}$$

$$\forall w_1, w_1' \in W_1, \forall w_2, w_2' \in W_2 : \succeq_m (w_1 \times w_2, w_1' \times w_2') \quad \equiv \quad \succeq_k (w_1, w_1') \wedge \succeq_l (w_2, w_2')$$

and

$$\forall w_1, w_1' \in W_1, \forall w_2, w_2' \in W_2 :$$
$$\succeq_m (w_1 \times w_2, w_1' \times w_2') \quad \equiv \quad \succ_k (w_1, w_1') \vee (=_k (w_1, w_1') \wedge \succeq_l (w_2, w_2')) \ ,$$

respectively.

4 Algebraic Optimization

As the preference operator extends RA, the optimization of queries with preferences can be realized as an extension of a classical relational query optimization. Most importantly, we can inherit all well known laws from RA, which, together with algebraic laws governing the commutativity and distributivity of the preference operator with respect to RA operations, constitute a formal foundation for rewriting queries with preferences using the standard strategies (*push selection*, *push projection*) aiming at reducing the sizes of intermediate relations.

4.1 Commuting with Selection

The following theorem identifies a sufficient condition under which the preference operator commutes with RA selection:

Theorem 3 (Commuting with selection)**.** *Given a relation schema R, a preference specification $\mathcal{P}$ over R, the set of its preference models $\mathscr{M}$, and a selection condition φ over R, if the formula*

$$\forall \mathcal{M}_k = \langle W, \succeq_k \rangle \in \mathscr{M}, \forall w, w' \in W : \quad \succ_k (w', w) \wedge w = \sigma_\varphi(w) \Rightarrow w' = \sigma_\varphi(w')$$

is valid, then for any relation instance $I(R)$ of R:

$$\omega_{\mathcal{P}}(\sigma_\varphi(I(R))) = \sigma_\varphi(\omega_{\mathcal{P}}(I(R))) \stackrel{def}{=} \{w \in \omega_{\mathcal{P}}(I(R)) | \sigma_\varphi(w) = w\} \ .$$

Proof. Observe that:

$$w \in \omega_{\mathcal{P}}(\sigma_\varphi(I(R))) \equiv w \subseteq I(R) \wedge \sigma_\varphi(w) = w \wedge \\ \neg(\forall \mathcal{M}_k \in \mathscr{M}, \exists w' : (w' \subseteq I(R) \wedge \sigma_\varphi(w') = w' \wedge \succ_k (w', w)) \ .$$

$$w \in \sigma_\varphi(\omega_{\mathcal{P}}(I(R))) \equiv w \subseteq I(R) \wedge \sigma_\varphi(w) = w \wedge \\ \neg(\forall \mathcal{M}_k \in \mathscr{M}, \exists w' : (w' \subseteq I(R) \wedge \succ_k (w', w)) \ ,$$

Obviously, the second formula implies the first. To see that the opposite implication also holds, we assume $w \notin \sigma_\varphi(\omega_{\mathcal{P}}(I(R)))$ and prove that than also $w \notin \omega_{\mathcal{P}}(\sigma_\varphi(I(R)))$. There are three cases when $w \notin \sigma_\varphi(\omega_{\mathcal{P}}(I(R)))$. If $w \nsubseteq I(R)$ or $\sigma_\varphi(w) \neq w$, it is immediately clear that $w \notin \omega_{\mathcal{P}}(\sigma_\varphi(I(R)))$. In the third case, $\forall \mathcal{M}_k \in \mathscr{M}, \exists w' : (w' \subseteq I(R) \wedge \succ_k (w', w)$. However, due to assumption of the theorem, $\forall \mathcal{M}_k \in \mathscr{M}, \exists w' : (w' \subseteq I(R) \wedge \sigma_\varphi(w') = w' \wedge \succ_k (w', w)$, which completes the proof. □

4.2 Commuting with Projection

The following theorem identifies sufficient conditions under which the preference operator commutes with RA projection. To prepare the ground for the theorem, some definitions have to be introduced:

Definition 6 (Restriction of a preference relation)**.** Given a relation schema R, a set of attributes X of R, and a preference relation $\succeq$ over R, the restriction $\theta_X(\succeq)$ of $\succeq$ to X is a preference relation $\succeq_X$ over $\pi_X(R)$ defined using the following formula:

$$\succeq_X (w_X, w'_X) \quad \equiv \quad \forall w, w' \in W : \pi_X(w) = w_X \wedge \ \pi_X(w') = w'_X \Rightarrow \ \succeq (w, w') \ .$$

Definition 7 (Restriction of the preference model)**.** Given a relation schema R, a set of relation attributes X of R, and a preference model $\mathcal{M} = \langle W, \succeq \rangle$ over R, the restriction $\theta_X(\mathcal{M})$ of $\mathcal{M}$ to X is a preference model $\mathcal{M}_X = \langle W_X, \succeq_X \rangle$ over $\pi_X(R)$ where $W_X = \{\pi_X(w) \mid w \in W\}$.

Definition 8 (Restriction of the preference operator)**.** Given a relation schema R, a set of attributes X of R, a preference specification $\mathcal{P}$ over R, and the set $\mathscr{M}_X$ of its models restricted to X, the restriction $\theta_X(\omega_{\mathcal{P}})$ of the preference operator $\omega_{\mathcal{P}}$ to X is the preference operator $\omega^X_{\mathcal{P}}$ defined as follows:

$$\omega^X_{\mathcal{P}}(\pi_X(I(R))) = \{w_X \in W_X \mid w_X \subseteq \pi_X(I(R)) \wedge \exists \mathcal{M}_X \in \mathscr{M}_X \text{ s.t. } \\ \forall w'_X \in W_X : w'_X \subseteq \pi_X(I(R)) \wedge \succeq_X (w'_X, w_X) \Rightarrow \ \succeq_X (w_X, w'_X)\} \ .$$

Theorem 4 (Commuting with projection)**.** *Given a relation schema R, a set of attributes X of R, a preference specification $\mathcal{P}$ over R, and the set of its preference models $\mathscr{M}$, if the following formulae*

$$\forall \mathcal{M}_k \in \mathscr{M}, \forall w_1, w_2, w_3 \in W : \\ \pi_X(w_1) = \pi_X(w_2) \wedge \pi_X(w_1) \neq \pi_X(w_3) \wedge \succeq_k (w_1, w_3) \Rightarrow \ \succeq_k (w_2, w_3) \ ,$$

$$\forall \mathcal{M}_k \in \mathscr{M}, \forall w_1, w_3, w_4 \in W : \\ \pi_X(w_3) = \pi_X(w_4) \wedge \pi_X(w_1) \neq \pi_X(w_3) \wedge \succeq_k (w_1, w_3) \Rightarrow \ \succeq_k (w_1, w_4)$$

are valid, then for any relation instance $I(R)$ of R:

$$\omega^X_{\mathcal{P}}(\pi_X(I(R))) = \pi_X(\omega_{\mathcal{P}}(I(R))) \overset{def}{=} \{\pi_X(w) \mid w \in \omega_{\mathcal{P}}(I(R))\} \ .$$

Proof. We prove: $\pi_X(w) \notin \omega^X_{\mathcal{P}}(\pi_X(I(R))) \iff \pi_X(w) \notin \pi_X(\omega_{\mathcal{P}}(I(R)))$.

- $\Rightarrow$: Assume $\pi_X(w_3) \notin \omega^X_{\mathcal{P}}(\pi_X(I(R)))$. The case $\pi_X(w_3) \nsubseteq \pi_X(I(R))$ is trivial. Otherwise, it must be the case that $\forall \mathcal{M}_X \in \mathscr{M}_X, \exists w_X$ s.t. $w_X \subseteq \pi_X(I(R))$ and $\succ_X (w_X, \pi_X(w_3))$, which implies $\forall \mathcal{M}_k \in \mathscr{M}, \forall w_1, w_4 \in W : \pi_X(w_1) = w_X \wedge \pi_X(w_4) = \pi_X(w_3) \Rightarrow \succ_k (w_1, w_4)$ and thus $\pi_X(w_3) \notin \pi_X(\omega_{\mathcal{P}}(I(R)))$.
- $\Leftarrow$: Assume $\pi_X(w_3) \notin \pi_X(\omega_{\mathcal{P}}(I(R)))$. Then $\forall \mathcal{M}_k \in \mathscr{M}$ and $\forall w_4 \subseteq I(R)$ s.t. $\pi_X(w_4) = \pi_X(w_3)$, there is $w_1 \subseteq I(R)$ s.t. $\succ_k (w_1, w_4)$ and $\pi_X(w_1) \neq \pi_X(w_4)$. From the assumption of the theorem, it follows that $\forall w_2, w_4 \subseteq I(R) : \pi_X(w_2) = \pi_X(w_1) \wedge \pi_X(w_4) = \pi_X(w_3) \Rightarrow \succ_k (w_2, w_4)$, which implies $\theta_X(\succ_k)(\pi_X(w_1), \pi_X(w_3))$ and thus $\pi_X(w_3) \notin \omega^X_{\mathcal{P}}(\pi_X(I(R)))$. □

4.3 Distributing over Cartesian Product

For preference operator to distribute over the Cartesian product of two relations, the preference specification, which is the parametr of the preference operator, needs to be decomposed into the preference specifications that will distribute into the argument relations:

Theorem 5 (Distributing over Cartesian product)**.** *Given two relation schemas R_1 and R_2, and preference specifications $\mathcal{P}_1$ over R_1 and $\mathcal{P}_2$ over R_2, for any two relation instances $I(R_1)$ and $I(R_2)$ of R_1 and R_2, the following property holds:*

$$\omega_{\mathcal{P}_0}(I(R_1) \times I(R_2)) = \omega_{\mathcal{P}_1}(I(R_1)) \times \omega_{\mathcal{P}_2}(I(R_2)) \overset{def}{=} \\ \{w_1 \times w_2 \mid w_1 \in \omega_{\mathcal{P}_1}(I(R_1)) \wedge w_2 \in \omega_{\mathcal{P}_2}(I(R_2))\} \ ,$$

where $\mathcal{P}_0 = P(\mathcal{P}_1, \mathcal{P}_2)$ is a Pareto composition of $\mathcal{P}_1$ and $\mathcal{P}_2$.

Proof. We prove:

$$w_1 \times w_2 \notin \omega_{\mathcal{P}_0}(I(R_1) \times I(R_2)) \Longleftrightarrow w_1 \times w_2 \notin \omega_{\mathcal{P}_1}(I(R_1)) \times \omega_{\mathcal{P}_2}(I(R_2)) \ .$$

- $\Rightarrow$: Assume $w_1 \times w_2 \notin \omega_{\mathcal{P}_0}(I(R_1) \times I(R_2))$. Then $\forall \mathcal{M}_m \in \mathscr{M}_0$, models of $\mathcal{P}_0$, there are $w_1' \subseteq I(R_1), w_2' \subseteq I(R_2)$ s.t. $\succ_m (w_1' \times w_2', w_1 \times w_2)$. Consequently, $\forall \mathcal{M}_k \in \mathscr{M}_1, \forall \mathcal{M}_l \in \mathscr{M}_2$, models of $\mathcal{P}_1$ and $\mathcal{P}_2$, there are $w_1' \subseteq I(R_1), w_2' \subseteq I(R_2)$ s.t. $\succ_k (w_1', w_1)$ or $\succ_l (w_2', w_2)$, which implies $w_1 \notin \omega_{\mathcal{P}_1}(I(R_1))$ or $w_2 \notin \omega_{\mathcal{P}_2}(I(R_2))$ and thus $w_1 \times w_2 \notin \omega_{\mathcal{P}_1}(I(R_1)) \times \omega_{\mathcal{P}_2}(I(R_2))$.
- $\Leftarrow$: Assume $w_1 \times w_2 \notin \omega_{\mathcal{P}_1}(I(R_1)) \times \omega_{\mathcal{P}_2}(I(R_2))$. Then $w_1 \notin \omega_{\mathcal{P}_1}(I(R_1))$ or $w_2 \notin \omega_{\mathcal{P}_2}(I(R_2))$. Assume the first. Then $\forall \mathcal{M}_k \in \mathscr{M}_1$, models of $\mathcal{P}_1$, there must be $w_1' \subseteq I(R_1)$ s.t. $\succ_k (w_1', w_1)$. Consequently, $\forall \mathcal{M}_m \in \mathscr{M}_0$, models of $\mathcal{P}_0$, $\exists w_1' \subseteq I(R_1) : \succ_m (w_1' \times w_2, w_1 \times w_2)$, which implies $w_1 \times w_2 \notin \omega_{\mathcal{P}_0}(I(R_1) \times I(R_2))$. The second case is symmetric. □

For lexicographic composition, we obtain the same property as for Pareto composition:

Theorem 6 (Distributing over Cartesian product)**.** *Given two relation schemas R_1 and R_2, and preference specifications $\mathcal{P}_1$ over R_1 and $\mathcal{P}_2$ over R_2, for any two relation instances $I(R_1)$ and $I(R_2)$ of R_1 and R_2, the following property holds:*

$$\omega_{\mathcal{P}_0}(I(R_1) \times I(R_2)) = \omega_{\mathcal{P}_1}(I(R_1)) \times \omega_{\mathcal{P}_2}(I(R_2)) \stackrel{def}{=} \{w_1 \times w_2 \mid w_1 \in \omega_{\mathcal{P}_1}(I(R_1)) \wedge w_2 \in \omega_{\mathcal{P}_2}(I(R_2))\} \ ,$$

where $\mathcal{P}_0 = L(\mathcal{P}_1, \mathcal{P}_2)$ is a lexicographic composition of $\mathcal{P}_1$ and $\mathcal{P}_2$.

Proof. We prove:

$$w_1 \times w_2 \notin \omega_{\mathcal{P}_0}(I(R_1) \times I(R_2)) \Longleftrightarrow w_1 \times w_2 \notin \omega_{\mathcal{P}_1}(I(R_1)) \times \omega_{\mathcal{P}_2}(I(R_2)) \ .$$

- $\Rightarrow$: Assume $w_1 \times w_2 \notin \omega_{\mathcal{P}_0}(I(R_1) \times I(R_2))$. Then $\forall \mathcal{M}_m \in \mathscr{M}_0$, models of $\mathcal{P}_0$, there are $w_1' \subseteq I(R_1), w_2' \subseteq I(R_2)$ s.t. $\succ_m (w_1' \times w_2', w_1 \times w_2)$. Consequently, $\forall \mathcal{M}_k \in \mathscr{M}_1, \forall \mathcal{M}_l \in \mathscr{M}_2$, models of $\mathcal{P}_1$ and $\mathcal{P}_2$, there are $w_1' \subseteq I(R_1), w_2' \subseteq I(R_2)$ s.t. $\succ_k (w_1', w_1)$ or $=_k (w_1', w_1) \wedge \succ_l (w_2', w_2)$, which implies $w_1 \notin \omega_{\mathcal{P}_1}(I(R_1))$ or $w_2 \notin \omega_{\mathcal{P}_2}(I(R_2))$ and thus $w_1 \times w_2 \notin \omega_{\mathcal{P}_1}(I(R_1)) \times \omega_{\mathcal{P}_2}(I(R_2))$.
- $\Leftarrow$: Assume $w_1 \times w_2 \notin \omega_{\mathcal{P}_1}(I(R_1)) \times \omega_{\mathcal{P}_2}(I(R_2))$. Then $w_1 \notin \omega_{\mathcal{P}_1}(I(R_1))$ or $w_2 \notin \omega_{\mathcal{P}_2}(I(R_2))$. Assume the first. Then $\forall \mathcal{M}_k \in \mathscr{M}_1$, models of $\mathcal{P}_1$, there must be $w_1' \subseteq I(R_1)$ s.t. $\succ_k (w_1', w_1)$. Consequently, $\forall \mathcal{M}_m \in \mathscr{M}_0$, models of $\mathcal{P}_0$, there must be w_1' s.t. $\succ_m (w_1' \times w_2, w_1 \times w_2)$, which implies $w_1 \times w_2 \notin \omega_{\mathcal{P}_0}(I(R_1) \times I(R_2))$. The second case is symmetric. □

Both Theorem 5 and Theorem 6 make it possible to derive the transformation rule that pushes preference operator with a one-dimensional preference specification down the appropriate argument of the Cartesian product:

Corollary 1. *Given two relation schemas R_1 and R_2, a preference specifications $\mathcal{P}_1$ over R_1, and an empty preference specification $\mathcal{P}_2$ over R_2, for any two relation instances $I(R_1)$ and $I(R_2)$ of R_1 and R_2, the following property holds:*

$$\omega_{\mathcal{P}_0}(I(R_1) \times I(R_2)) = \omega_{\mathcal{P}_1}(I(R_1)) \times 2^{I(R_2)} \stackrel{def}{=} \{w_1 \times w_2 \mid w_1 \in \omega_{\mathcal{P}_1}(I(R_1)) \wedge w_2 \subseteq I(R_2)\} \ ,$$

where $\mathcal{P}_0 = P(\mathcal{P}_1, \mathcal{P}_2)$ is a Pareto of lexicographic composition of $\mathcal{P}_1$ and $\mathcal{P}_2$.

Proof. Follows from previous theorems and from the equality $\omega_{\mathcal{P}_{\text{empty}}}(I(R)) = 2^{I(R)}$. □

4.4 Distributing over Union

The following theorem shows how the preference operator distributes over the union of two relations:

Theorem 7 (Distributing over union)**.** *Given two compatible relation schemas*[3] *R and S, and a preference specification $\mathcal{P}$ over R (and S), if the following formula*

$$\omega_{\mathcal{P}}(I(R) \cup I(S)) \subseteq 2^{I(R)} \cup 2^{I(S)}$$

is valid for relation instances $I(R)$ and $I(S)$ of R and S, then the following property holds:

$$\omega_{\mathcal{P}}(I(R) \cup I(S)) = \omega_{\mathcal{P}}(\omega_{\mathcal{P}}(I(R)) \cup \omega_{\mathcal{P}}(I(S))) \ .$$

Proof. Obviously, $\omega_{\mathcal{P}}(I(R)) \cup \omega_{\mathcal{P}}(I(S)) \subseteq 2^{I(R) \cup I(S)}$. If we show that $\omega_{\mathcal{P}}(I(R) \cup I(S)) \subseteq \omega_{\mathcal{P}}(I(R)) \cup \omega_{\mathcal{P}}(I(S))$, the theorem immediately follows from Theorem 1.

Indeed, if $w \in \omega_{\mathcal{P}}(I(R) \cup I(S))$, then it follows from the definition of the preference operator $w \subseteq I(R) \cup I(S) \wedge \exists \mathcal{M}_k \in \mathscr{M}$ s.t. $\forall w' \in W : w' \subseteq I(R) \cup I(S) \wedge \succeq_k (w', w) \Rightarrow \succeq_k (w, w')$. As we know that $w \subseteq I(R) \vee w \subseteq I(S)$ from the assumption of the theorem, we can conclude $w \in \omega_{\mathcal{P}}(I(R)) \cup \omega_{\mathcal{P}}(I(S))$. □

4.5 Distributing over Difference

Only in the trivial case, the preference operator can be distributed over difference:

Theorem 8 (Distributing over difference)**.** *Given two compatible relation schemas R and S, and a preference specification $\mathcal{P}$ over R (and S), for any two relation instances $I(R)$ and $I(S)$ of R and S, the following property holds:*

$$\omega_{\mathcal{P}}(I(R) - I(S)) = \omega_{\mathcal{P}}(I(R)) - \omega_{\mathcal{P}}(I(S))$$

iff the preference specification $\mathcal{P}$ is empty.

4.6 *Push Preference*

The question arises how to integrate the above algebraic laws into the classical, well-known hill-climbing algorithm. In particular, we want to add heuristic strategy of *push preference*, which is based on the assumption that early application of the preference operator reduces intermediate results. Indeed, the Theorem 1 provides a formal evidence that it is correct to pass exactly all the tuples that have been included in any world returned by the preference operator to the next operator in the operator tree. This leads to a better performance in subsequent operators.

[3] We call two relation schemas *compatible* if they have the same number of attributes and the corresponding attributes have identical domains.

5 Related Work

The study of preferences in the context of database queries has been originated by Lacroix and Lavency [8]. They, however, haven't addressed the issue of algebraic optimization.

Nevertheless, only at the turn of the millennium this area attracted broader interest again. Kießling [6] and Chomicki et al. [2] have pursued independently a similar, *qualitative* approach within which preferences between tuples are specified directly, using binary *preference relations*. They have defined an operator returning only the best preference matches. However, they, by contrast to the approach presented in this paper, don't consider preferences between *sets* of elements and are concerned only with one type of preference. Moreover, the relation to a preference logic unfortunately is unclear. On the other hand, both Chomicki et. al. [2] and Kießling [7, 4] have laid the foundation for preference query optimization that extends established query optimization techniques.

A special case of the same embedding represents *skyline operator* introduced by Börzsönyi et al. [1]. Some examples of possible rewritings for skyline queries are given but no general rewriting rules are formulated.

In [9], actual values of an arbitrary attribute were allowed to be partially ordered according to user preferences. Accordingly, RA operations, aggregation functions and arithmetic were redefined. However, some of their properties were lost, and the the query optimization issues were not discussed.

6 Conclusions

We build on the framework of embedding preferences into RQL through the preference operator that is parameterized by user preferences expressed in a declarative, logical language containing sixteen kinds of preferences and that returns the most preferred sets of tuples of its argument relation. Most importantly, the language is suitable for expressing preferences between sets of elements and its semantics allows for conflicting preferences.

The main contribution of the paper consists in presenting basic properties of the preference operator and a number of algebraic laws describing its interaction with other RA operators. Particularly, sufficient conditions for commuting the preference operator with RA selection or projection and for distributing over Cartesian product, set union, and set difference have been identified. Thus key rules for rewriting the preference queries using the standard algebraic optimization strategies like *push preference* or *push projection* have been established. Moreover, a new optimization strategy of *push preference* has been suggested.

Future work directions include identifying further algebraic properties and finding the best possible ordering of transformations for optimization of RA statements with the preference operator. Also, expressiveness and complexity issues have to be addressed in detail.

References

[1] S. Börzsönyi, D. Kossmann, and K. Stocker. The skyline operator. In 'Proceedings of the 17th International Conference on Data Engineering', 421–430, Washington, DC, USA, (2001). IEEE Computer Society.

[2] J. Chomicki. *Preference Formulas in Relational Queries.* ACM Trans. Database Syst. **28** (2003), 427–466.

[3] J. Doyle and M. P. Wellman. Representing preferences as ceteris paribus comparatives. In 'Decision-Theoretic Planning: Papers from the 1994 Spring AAAI Symposium', 69–75. AAAI Press, Menlo Park, California, (1994).

[4] B. Hafenrichter and W. Kießling. Optimization of relational preference queries. In 'CRPIT '39: Proceedings of the sixteenth Australasian conference on Database technologies', 175–184, Darlinghurst, Australia, Australia, (2005). Australian Computer Society, Inc.

[5] S. Kaci and L. W. N. van der Torre. Non-monotonic reasoning with various kinds of preferences. In 'IJCAI-05 Multidisciplinary Workshop on Advances in Preference Handling', Ronen I.Brafman and U. Junker, (eds.), 112–117, (August 2005).

[6] W. Kießling. Foundations of Preferences in Database Systems. In 'Proceedings of the 28th VLDB Conference', 311–322, Hong Kong, China, (2002).

[7] W. Kießling and B. Hafenrichter. Algebraic optimization of relational preference queries. Technical Report 2003-01, Institute of Computer Science, University of Augsburg, (February 2003).

[8] M. Lacroix and P. Lavency. Preferences; Putting More Knowledge into Queries. In 'VLDB', P. M. Stocker, W. Kent, and P. Hammersley, (eds.), 217–225. Morgan Kaufmann, (1987).

[9] R. Nedbal. *Relational Databases with Ordered Relations.* Logic Journal of the IGPL **13** (2005), 587–597.

[10] R. Nedbal. Non-monotonic reasoning with various kinds of preferences in the relational data model framework. In 'ITAT 2007, Information Technologies – Applications and Theory', P. Vojtáš, (ed.), 15–20, Poľana, (September 2007). PONT.

[11] G. von Wright. *The logic of preference.* Edinburgh University Press, Edinburgh, (1963).