# Model-Theoretic Minimal Change Operators for Constraint Databases*

Peter Z. Revesz

Department of Computer Science and Engineering
University of Nebraska, Lincoln, NE 68588
revesz@cse.unl.edu, (402) 472-3488

**Abstract:** Database systems should allow users to insert new information (described by a first-order sentence) into a database without specifying exactly how. The database systems should be able to figure out what tuples to add or delete from the current database to satisfy fully the user's request. The guiding principle of accomplishing such insertions is the concept of model-theoretic minimal change. This paper shows that this concept can be applied to constraint databases. In particular, any constraint database change operator that satisfies the axioms for revision [AGM85], update [KM92], or arbitration [Rev96] accomplishes a model-theoretic minimal change in a well-defined sense. The paper also presents concrete operators for revision, update, and arbitration for constraint databases with real polynomial inequality constraints.

## 1 Introduction

Most change operators in current database systems require the users to know the exact contents of the database. Users are expected to know what tuples are in the database and specifically command to delete, modify or add specific tuples. However, it is very difficult to know exactly what is in a complex database and database users should be freed from that burden. This is especially true in constraint databases [KKR95]. Users should be able to change a database by simply telling the database system what new information to incorporate into it. The database system should be able to figure out by itself how to incorporate the new information.

The idea of automatic change requires a solution to the well-known frame problem of artificial intelligence: if some things are known to change, what other things must change with them and what things must stay the same? A nice solution to this problem is the principle of model-theoretic minimal change. Let's call the set of models of the world that are currently thought possible the database. Each new information, described in some logical form, allows several models of the world. The principle of minimal change states that the result of adding the new information to a database should be the set of models of the new information that are closest to some possible models in the current

* This work was supported by NSF grants IRI-9625055 and IRI-9632871. This work appears in *Proc. Sixth International Conference on Database Theory*, Springer-Verlag LNCS 1186, pp. 447-460, Delphi, Greece, January 1997.

database. Hence the database change problem can be largely reduced to finding good measures for distance between models of the world.

While the principle of minimal change is agreed upon by most people, to say precisely what are good operators and distance measures between models is difficult. Katsuno and Mendelzon [KM91, KM92] pointed out that there are in fact two very different contexts for database change. They divided change operators into two broad classes: revision and update. They characterized these two classes of change operators by a set of axioms that they have to satisfy and proved that members in each class accomplish a model-theoretic minimal change, in different ways. To these classes, [Rev96] added a third class of change operators: arbitration, which is also defined by a set of axioms. [Rev96] proved that members in this class also accomplish a model-theoretic minimal change, in a way distinct from both revision and update. Arbitration is motivated by heterogeneous database systems which often require combining information from various sources before answering queries (see also [BKMS92, BNR95, TSIMMIS, LS95, Sub94, Wie92]).

The model-theoretic characterizations in [KM91, KM92, Rev96] were shown only for propositional knowledgebases. It seems much more difficult to define operators for first-order knowledgebases that can be similarly characterized. Fagin et al. [FKUV86] present a revision operator for first-order logical databases but it violates an important axiom in [AGM85] on which the characterization depends, namely it violates the Principle of Irrelevance of Syntax. Grahne et al. [GMR92] present an update operator but it uses the active domain semantics, which may be unnatural for constraint databases.

In this paper we present for a simple type of first-order database, namely constraint (or generalized) databases with rational order constraints [KKR95], revision, update, and arbitration operators that satisfy appropriate generalizations of the axioms in [KM91, KM92] and [Rev96]. We also show that all operators satisfying the relevant axioms have a model-theoretic characterization.

Potential applications of the paper occur mainly in manipulating spatial data described by constraint databases. For example, a Geographic Information System (GIS) for agriculture would need update after overflooding of a river, or a new crop being planted. Revision may be needed in light of new data, e.g. this is a corn field, not wheat, and arbitration would solve many sensor fusion problems, e.g., merge of two different soil maps derived from two different satellite measurements.

This paper is organized as follows. Section 2 reviews basic concepts in constraint databases and database change operators. Section 3 presents a syntax independent way of measuring the size of constraint databases. Section 4 presents a syntax independent way of measuring distance. Sections 5-7 give model-theoretic characterizations of revision, update, and arbitration. Section 8 considers the case when the new information is a first-order sentence. Section 9 gives a conclusion and mentions open problems.

## 2 Basic Concepts

The following are basic definitions adopted from [KKR95].

**Definition 2.1** Let $\Phi$ be the set of atomic constraints of some constraint theory. A *generalized k-tuple* over variables $x_1, \ldots, x_k$ is of the form: $r(x_1, \ldots, x_k)$ :— $\phi_1 \wedge \ldots \wedge \phi_n$ where $r$ is a relation symbol, and $\phi_i \in \Phi$ for $1 \leq i \leq n$ and uses only the variables $x_1, \ldots, x_k$.
A *generalized relation $r$ with arity $k$* is a finite set of generalized k-tuples with symbol $r$ on left hand side.
A *generalized database* is a finite set of generalized relations.
A *generalized knowledgebase* is a finite set of generalized databases. □

**Definition 2.2** Let $D$ be the domain over which variables are interpreted. Then the *model of a generalized k-tuple $t$* with variables $x_1, \ldots, x_k$ is the unrestricted k-ary relation $\{(a_1, \ldots, a_k) : (a_1, \ldots, a_k) \in D^k$ and the substitution of $a_i$ for $x_i$ satisfies the right hand side$\}$.
The *model of a generalized relation* is the union of the models of its generalized tuples.
The *model of a generalized database* is the set of the models of its generalized relations.
The *model of a generalized knowledgebase* is the set of the models of its generalized databases. □

In this paper we will denote the models of a $A$ by $Mod(A)$ where $A$ is a generalized relation, database, or knowledgebase.

Katsuno and Mendelzon studied propositional knowledgebases. Each propositional knowledgebase is described by a single propositional formula. The models of the formula are interpretations (truth assignments to propositional variables) that make the formula true. Contrast that with our definition of a generalized knowledgebase: instead of a set of interpretations we have a set of generalized databases. This change is important because a crucial issue for [KM91] is to define the distance between pairs of interpretations. For us an important task will be to define distances between pairs of generalized databases.

Let $\mathcal{M}$ be the set of all possible generalized databases. (In [KM91] $\mathcal{M}$ is the set of interpretations.) A *pre-order* $\leq$ over $\mathcal{M}$ is a reflexive and transitive relation on $\mathcal{M}$. A pre-order is *total* if for every pair $I, J \in \mathcal{M}$, either $I \leq J$ or $J \leq I$ holds. We define the relation $<$ as $I < J$ if and only if $I \leq J$ and $J \not\leq I$. The set of *minimal elements* of a subset $\mathcal{S}$ of $\mathcal{M}$ with respect to a pre-order $\leq_\psi$ is defined as:

$$Min(\mathcal{S}, \leq_\psi) = \{I \in \mathcal{S} : \not\exists I' \in \mathcal{S} \text{ where } I' <_\psi I\}$$

Katsuno and Mendelzon gave the following model-theoretic characterization of revision and update when the knowledge base is represented by a single propositional formula. Let the symbol $\circ$ denote revision and the symbol $\diamond$ denote update operators.

Suppose we have for each knowledge base $\psi$ a total pre-ordering $\leq_\psi$ of interpretations for closeness to $\psi$, where the pre-order $\leq_\psi$ satisfies certain conditions [KM91]. Revision operators that satisfy the AGM postulates are exactly those that select from the models of the new information $\phi$ the closest models to the propositional knowledge base $\psi$. That is,

$$Mod(\psi \circ \phi) = Min(Mod(\phi), \leq_\psi)$$

For updates assume for each interpretation $I$ some partial pre-ordering $\leq_I$ of interpretations for closeness to $I$. Update operators select for each model $I$ in $Mod(\psi)$ the set of models from $Mod(\phi)$ that are closest to $I$. The new theory is the union of all such models. That is,

$$Mod(\psi \diamond \phi) = \bigcup_{I \in Mod(\psi)} Min(Mod(\phi), \leq_I)$$

A third type of theory change that is axiomatically defined in [Rev96] is called arbitration. Let $\triangleright$ denote arbitration. Then arbitration operators will be characterized similarly to revision above, but the pre-order $\leq_\psi$ has to satisfy a different set of conditions.

## 3 A Multi-Measure for Real Polynomial Constraint Relations

In the context of aggregate operators, Kuper [Kup93] suggested area to measure generalized databases, and Chomicki and Kuper [ChK95] suggested as a measure the asymptotic probability that an arbitrary point belongs to the model of the generalized database. This section presents another measure for the area of constraint databases assuming that the atomic constraints are real polynomial inequalities.

The reason for introducing a multi-measure is that even a very large difference between two regions in dimension $i$ is unimportant compared to the smallest difference in dimension $i+1$. For example, if two drawings in the $\mathcal{R}^2$ plane differ in any small line segment, then it should be considered more important than that they also differ on any finite number of points. However, in case they agree on all line segments, then the number of point differences can be very useful to know. Hence we need a multi-measure that records all dimensional volumes simultaneously, i.e. some vector of dimension $k+1$ for measuring $k$ dimensional regions.

**Definition 3.1** A region is *elementary* if and only if it is one of (a) a point (b) a line without endpoints or (c) an open region. □

**Definition 3.2** A *partition* of a region $R$ is a disjoint set of elementary regions $P_1, \ldots, P_n$ such that $Mod(R) = Mod(P_1 \cup \ldots \cup P_n)$. □

Since the elements of a partition are disjoint, no two lines may cross each other and no region may contain other elements in the partition. We call a point *extensional* if it is within the partition but is not within any line or region of the partition. We also call the number of line elements incident on a point within a partition the *degree* of the point.

In 3-dimensional space the degree of each edge is two, nd the degree of each corner is three. For any higher dimension we take the degree of a line (hyperplane facet) to be the number of surfaces (higher dimensional hyperplane facets) that have the whole line (hyperplane facet) as a boundary.

**Definition 3.3** Let $P$ be any partition of k-dimensional region $R$. For each $1 \leq i \leq n$, the *multi-measure* of $P_i$, denoted $m(P_i)$ is the following $(a_0, \ldots, a_k)$ form vector:
(1) $(0, \ldots, 0, 1)$ if $P_i$ is an extensional point.
(2) $(0, \ldots, 0, degree(P_i) \times length(P_i), -(a+b))$ if $P_i$ is a line without endpoints where $P_i$ is incident upon points $A$ and $B$ and $a = 1/degree(A)$ if $A$ is extensional otherwise $a = 0$, and $b = 1/degree(B)$ if $B$ is extensional otherwise $b = 0$.
(3) $(0, \ldots, a_{k-j}, a_{k-(j+1)}, \ldots, 0)$ where $a_{k-j} = degree(P_i) \times volume(P_i)$ and $a_{k-(j+1)} =$ -boundary$(P_i)$ if $P_i$ is any $j \geq 2$ dimensional open region.

The multi-measure of $P$ is the sum of the vectors $m(P_i)$ for $1 \leq i \leq n$. □

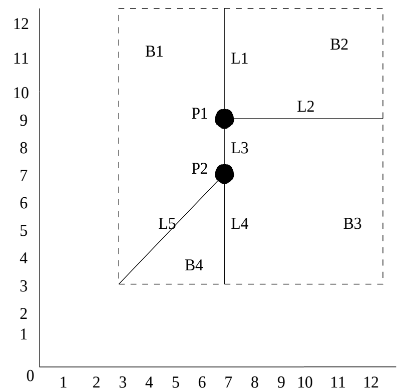
There are certain balances built into the above definition to make it partition independent. For example, if we have a square region in the plane and cut it diagonally by a line, then we introduced a new line segment (the diagonal). This would increase the sum of measures considered under condition (2) above in a positive direction, and those considered under (3) above in a negative direction by the same amount. In general, we can prove the following.

**Theorem 3.1** Let $R$ be any region, then all partitions of $R$ have the same multi-measure. □

**Definition 3.4** The multi-measure of a region $R$ is $m(P)$ where $P$ is any partition of $R$. □

**Example 3.1** Let $R$ be a region defined as follows: $A(x,y)$ :— $3 < x \wedge x < 12 \wedge 3 < y \wedge y < 12$

Note that $R$ is elementary because it is an open square region. Hence a trivial partition of $R$ is just itself. The measure of $R$ is $m(R) = (162, -36, 0)$ applying the definition. Another partition of $R$ is shown in the figure below. Hence another way of finding the measure of $R$ is to sum up the measure of the disjoint regions in the figure.




Here B1-B4 are open regions and L1-L5 are lines without endpoints and $P1, P2$ are extensional points. We find that $m(B_1) = (56, -(4\sqrt{2} + 18), 0)$, $m(B_2) = (30, -16, 0)$, $m(B_3) = (60, -22, 0)$, $m(B_4) = (16, -(4\sqrt{2} + 8), 0)$, $m(L_1) = (0, 6, -1/3)$, $m(L_2) = (0, 10, -1/3)$, $m(L_3) = (0, 4, -2/3)$, $m(L_4) = (0, 8, -1/3)$, $m(L_5) = (0, 8\sqrt{2}, -1/3)$, $m(P_1) = (0, 0, 1)$, $m(P_2) = (0, 0, 1)$, The sum of these measures is $(162, -36, 0)$ matching the value we found using the other partition. □

Next, let's see an example in 3-dimension.

**Example 3.2** The open cube with opposite corners $(0, 0, 0)$ and $(9, 9, 9)$ has multi-measure $(729, -486, 0, 0)$. Suppose that the cube is partitioned into an open cube with opposite corners $(3, 3, 3)$ and $(6, 6, 6)$, a volume surrounding the cube, and the sides, edges, and corners of the little cube. The surrounding region will have multi-measure $(702, -540, 0, 0)$, little cube will have $(27, -54, 0, 0)$, each of its six sides will have $(0, 2 \times 9, -12, 0)$, each of its twelve edges will have $(0, 0, 2 \times 3, -2/3)$ and its eight corners will have $(0, 0, 0, 1)$. The sum of the multi-measures of the partition is still $(729, 0, 0, 0)$. □

## 4 A Distance Measure between Relations

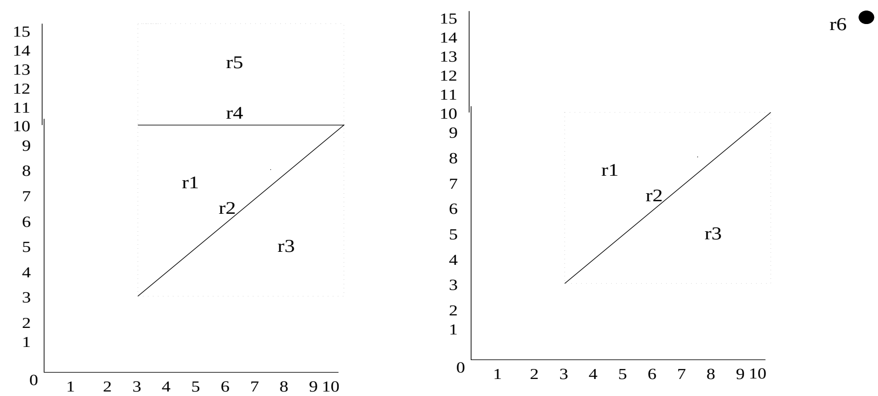
The *distance measure* between two regions $R_1$ and $R_2$, denoted $dist(R_1, R_2)$, is $m((R_1 \cup R_2) \setminus (R_1 \cap R_2))$. It follows from Theorem 3.1 that for any two regions $R_1$ and $R_2$, $dist(R_1, R_2)$ is a unique multi-measure. The distance between two constraint relations with the same relation symbol is the distance between regions associated with them. The distance between two constraint databases is the sum of the distances between the corresponding constraint relations in them, and the measure of the constraint relations whose symbol occurs in only one of them.

**Example 4.1** Suppose that we are given two constraint databases $I_1$ and $I_2$ describing land areas. $I_1$ is the constraint database:
$Wood(x, y) :\!\!— 3 < x \wedge x < 10 \wedge 3 < y \wedge y < 15$

and $I_2$ is the constraint database:

$Wood(x, y) :- 3 < x \land x < 10 \land 3 < y \land y < 10$
$Wood(x, y) :- x = 15 \land y = 15$

Let's call $\phi_1$, and $\phi_2$ respectively the (disjunction) of the formulas on the right hand side of $I_1$ and $I_2$. Let us use the partition shown in the figure below, where $r_1, r_3, r_5$ are open regions, $r_2, r_4$ are lines without endpoints, and $r_6$ is a point. Clearly, it is possible to express each $r_i$ as a conjunction of atomic constraints and $\phi_1$ as $r_1 \lor r_2 \lor r_3 \lor r_4 \lor r_5$ and $\phi_2$ as $r_1 \lor r_2 \lor r_3 \lor r_6$. Now we can calculate the distance between the two regions as $dist(I_1, I_2) = m((I_1 \cup I_2) \setminus (I_1 \cap I_2)) = m(r_4 \cup r_5 \cup r_6) = (0, 14, 0) + (70, -24, 0) + (0, 0, 1) = (70, -10, 1)$. $\square$




Jagadish [Jag91] considered as a distance measure between two-dimensional regions composed of a set of rectangles the difference in their areas. This measure cannot distinguish between the picture of a rectangle and the picture of the same rectangle with a dot on top of it. Nor can it distinguish between an open and a closed rectangle. In contrast, the distance measure introduced in this paper does distinguish in the above cases.

## 5 Revision

Katsuno and Mendelzon [KM91] translated the AGM [AGM85] postulates into six equivalent axioms on propositional knowledgebases. Since we use generalized knowledgebases, we translate the AGM axioms into a slightly different set of axioms. We say that $\circ$ is a revision operator on generalized knowledgebases if for each generalized knowledgebase $\psi, \mu$ and $\phi$ the following hold:

**(R1)** $Mod(\psi \circ \mu) \subseteq Mod(\mu)$.
**(R2)** If $\psi \sqcap \mu$ is nonempty then $Mod(\psi \circ \mu) = Mod(\psi \sqcap \mu)$.
**(R3)** If $\mu$ is nonempty then $\psi \circ \mu$ is nonempty.
**(R4)** If $Mod(\psi_1) = Mod(\psi_2)$ and $Mod(\mu_1) = Mod(\mu_2)$ then $Mod(\psi_1 \circ \mu_1) = Mod(\psi_2 \circ \mu_2)$.
**(R5)** $Mod((\psi \circ \mu)) \sqcap \phi) \subseteq Mod(\psi \circ (\mu \sqcap \phi))$.
**(R6)** If $(\psi \circ \mu) \sqcap \phi$ is nonempty then $Mod(\psi \circ (\mu \sqcap \phi)) \subseteq Mod((\psi \circ \mu) \sqcap \phi)$.

In the above $\mu \sqcap \phi = \{I \in \mu : \exists J \in \phi \text{ such that } Mod(I) = Mod(J)\}$. Note that $Mod(\mu \sqcap \phi) = Mod(\mu) \cap Mod(\phi)$. Next we define a concrete revision operator based on the distance measure in Section 4. We define the distance between a generalized knowledgebase $\psi$ and a generalized database $I$ as follows:

$$dist(\psi, I) = \min_{J \in \psi} dist(I, J)$$

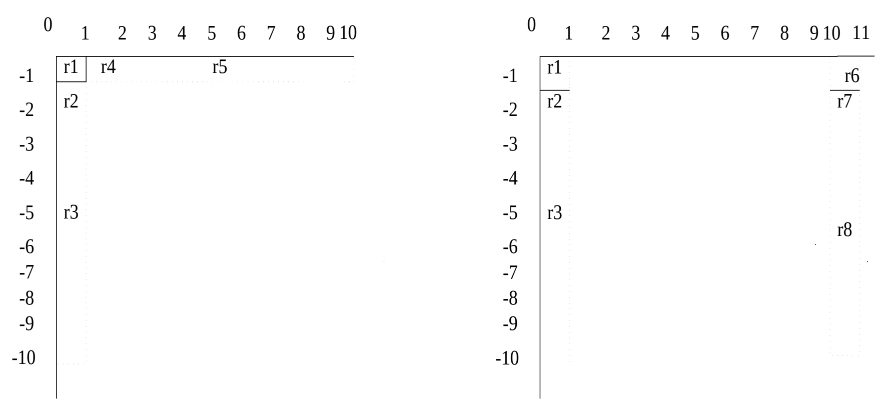
In comparing multi-measures, we consider the elements from left to right to be the most to least significant. Hence for example, $(1, 0, 5) \leq (2, 1, 6)$ and $(5, -8, 7) \leq (5, -7, 2)$. Next we define with respect to any generalized knowledgebase $\psi$ a total pre-order $\leq_\psi$ as follows. For each pair of generalized databases $I$ and $J$ let $I \leq_\psi J$ if and only if $dist(\psi, I) \leq dist(\psi, J)$. Now the revision operator $\circ$ can be defined as:

$$\psi \circ \mu = Min(\mu, \leq_\psi)$$

**Example 5.1** Suppose that there are two options for landscaping an empty area $I_1$. (Note that $I_1$ is described by an empty database.) The first option $J_1$ is:
$Wood(x, y)$ :— $0 < x \wedge x < 1 \wedge -10 < y \wedge y < 0$
$Wood(x, y)$ :— $1 < x \wedge x < 10 \wedge -1 < y \wedge y < 0$.

The second option $J_2$ is:
$Wood(x, y)$ :— $0 < x \wedge x < 1 \wedge -10 < y \wedge y < 0$
$Wood(x, y)$ :— $10 < x \wedge x < 11 \wedge -10 < y \wedge y < 0$.

Which landscape option is better to choose assuming that we want to do minimal work, i.e. to forest a minimal area? In this example let $\mu = \{J_1, J_2\}$. We need to find out which options in $\mu$ are closest to $I_1$. That is, we need to find $\{I_1\} \circ \mu$. Let $\phi_1$ and $\phi_2$ be the disjunctions of the right hand sides in $J_1$ and $J_2$. Let us use in this example the partition shown in the figure below. Here $\phi_1 = r_1 \vee r_2 \vee r_3 \vee r_4 \vee r_5$ and $\phi_2 = r_1 \vee r_2 \vee r_3 \vee r_6 \vee r_7 \vee r_8$.




Since the partition of $I_1$ is empty $dist(I_1, J_1) = m((I_1 \cup J_1) \setminus (I_1 \cap J_1)) = m(J_1) = m(r_1 \cup r_2 \cup r_3 \cup r_4 \cup r_5) = (2, -4, 0) + (0, 2, 0) + (18, -20, 0) + (0, 2, 0) +$

$(18, -20, 0) = (38, -40, 0)$. Similarly, we calculate that $dist(I_1, J_2) = (40, -44, 0)$. Note that $dist(I_1, J_1) < dist(I_1, J_2)$, hence $J_1 <_{\{I_1\}} J_1$. Therefore $\{I_1\} \circ \mu = \{J_1\}$. That means that we should choose the first landscape option, because it will require less work to realize. $\Box$

Our next example is more complex in that the knowledgebase to be revised contains several possible databases.

**Example 5.2** Suppose that we have the option of purchasing the land either completely cleared $I_1$ or with a patch of wooded area remaining $I_2$. The price of the two options is the same. Here the second purchase option can be described as:

$Wood(x, y) :\!\!-\; 10 < x \wedge x < 11 \wedge -10 < y, y < 0$

What is the best landscape option to choose in this case? Let $\psi = \{I_1, I_2\}$. To answer the question we need to do revision.

Calculating distances we find that $dist(I_2, J_1) = m(\cup_{1 \leq i \leq 8} r_i) = (58, -62, 0)$ and $dist(I_2, J_2) = (20, -22, 0)$. We see that the minimum distance among pairs of I's and J's is between $I_2$ and $J_2$. Hence $\psi \circ \mu = \{J_2\}$, that is, in this case we should choose the second landscape option (and implicitly buy the lot with the wooded patch in it.). $\Box$

A generalized *faithful* assignment is a function that assigns for each generalized knowledgebase $\psi$ a pre-order $\leq_\psi$ such that the following conditions hold. For each $I, J \in \mathcal{M}$ and generalized knowledgebases $\psi, \psi_1, \psi_2$:

(1) If $Mod(I), Mod(J) \in Mod(\psi)$ then $I <_\psi J$ does not hold.
(2) If $Mod(I) \in Mod(\psi)$ and $Mod(J) \notin Mod(\psi)$ then $I <_\psi J$.
(3) If $Mod(\psi_1) = Mod(\psi_2)$ then $\leq_{\psi_1} = \leq_{\psi_2}$.

The revision operator defined above is faithful. It is also possible to show that it satisfies the axioms (R1-R6). In general, we can extend the characterization theorem of Katsuno and Mendelzon as follows.

**Theorem 5.1** A revision operator satisfies axioms (R1-R6) if and only if there exists a generalized faithful assignment that maps each generalized knowledgebase $\psi$ to a total pre-order $\leq_\psi$ such that for every other generalized knowledgebase $\mu$, $Mod(\psi \circ \mu) = Mod(Min(\mu, \leq_\psi))$. $\Box$

# 6 Update

We say that $\diamond$ is an update operator on generalized knowledgebases if for each generalized knowledgebase $\psi$ and $\mu$ and generalized database $I$ the following hold:

**(U1)** $Mod(\psi \diamond \mu) \subseteq Mod(\mu)$.
**(U2)** If $Mod(\psi) \subseteq Mod(\mu)$ then $Mod(\psi \diamond \mu) = Mod(\psi)$.
**(U3)** If $\psi$ and $\mu$ nonempty, then $\psi \diamond \mu$ is nonempty.
**(U4)** If $Mod(\psi_1) = Mod(\psi_2)$ and $Mod(\mu_1) = Mod(\mu_2)$ then $Mod(\psi_1 \diamond \mu_1) = Mod(\psi_2 \diamond \mu_2)$.

**(U5)** $Mod((\psi \diamond \mu) \sqcap I) \subseteq Mod(\psi \diamond (\mu \sqcap I))$.
**(U6)** If $Mod(\psi \diamond \mu_1) \subseteq Mod(\mu_2)$ and $Mod(\psi \diamond \mu_2) \subseteq Mod(\mu_1)$ then $Mod(\psi \diamond \mu_1) = Mod(\psi \diamond \mu_2)$.
**(U7)** $(Mod((\{I\} \diamond \mu_1) \sqcap (\{I\} \diamond \mu_2))) \subseteq Mod(\{I\} \diamond (\mu_1 \cup \mu_2))$.
**(U8)** $Mod((\psi_1 \cup \psi_2) \diamond \mu) = Mod((\psi_1 \diamond \mu) \cup (\psi_2 \diamond \mu))$.

Note that axioms (U1) and (U4-U5) are the same as axioms (R1) and (R4-R5). Axiom (U2) is a weakening of axiom (R2) in the case when $\psi$ is satisfiable. Axiom (U3) is a weakening of axiom (R3) that is needed to avoid defining the update of an empty knowledgebase. Axioms (U6-U7) replace axiom (R6). They generalize (R6) slightly by admitting orderings where some pair of models of the new information are not comparable as to closeness to the knowledgebase. Axiom (U8) guarantees that each model in the knowledgebase is updated independently.

Next we define with respect to any generalized database $I$ a pre-order $\leq_I$ as follows. For each pair of generalized databases $J$ and $K$ let $J \leq_I K$ if and only if $dist(I, J) \leq dist(I, K)$. Next we define a concrete update operator $\diamond$ as follows:

$$\psi \diamond \mu = \bigcup_{I \in \psi} Min(\mu, \leq_I)$$

**Example 6.1** Let us return to Example 5.2. Suppose the neighbor tries to figure out what the new land will be like. The neighbor knows both the landscape options and the two ways of purchasing the land. However, suppose that the neighbor does not know that the price of the two purchase options is the same. What can the neighbor conclude?

This would require calculating $\psi \diamond \mu$, which turns out to be $\mu$. This is because to $I_1$ the closest is $J_1$ and to $I_2$ the closest is $J_2$. Hence the neighbor can expect that if $I_1$ is purchased then $J_1$ and if $I_2$ is purchased then $J_2$ will be the landscape chosen. However, even though $dist(I_2, J_2) \leq dist(I_1, J_1)$ as far as the neighbor knows the price of $I_1$ may be much lower than the price of $I_2$ to offset the extra landscaping work required. Therefore, the neighbor can conclude only that one of the two landscape options will be chosen, but cannot say for sure which one. $\square$

A generalized *faithful* assignment for updates satisfies the following condition:

For any generalized database $I$ if $I \neq J$ then $I <_I J$.

The update operator $\diamond$ defined above is faithful for updates and satisfies the axioms (U1-U8). In general, we have that:

**Theorem 6.1** An update operator satisfies axioms (U1-U8) if and only if there exists a generalized faithful assignment that maps each generalized database $I$ to a pre-order $\leq_I$ such that for every generalized knowledgebases $\psi, \mu$, $Mod(\psi \diamond \mu) = \bigcup_{I \in Mod(\psi)} Mod(Min(\mu, \leq_I))$. $\square$

## 7 Arbitration

We call generalized arbitration operators, denoted by $\triangleright$, those operators that satisfy axioms (R1) and (R3-R6) and axioms (A2) and (A7) below:

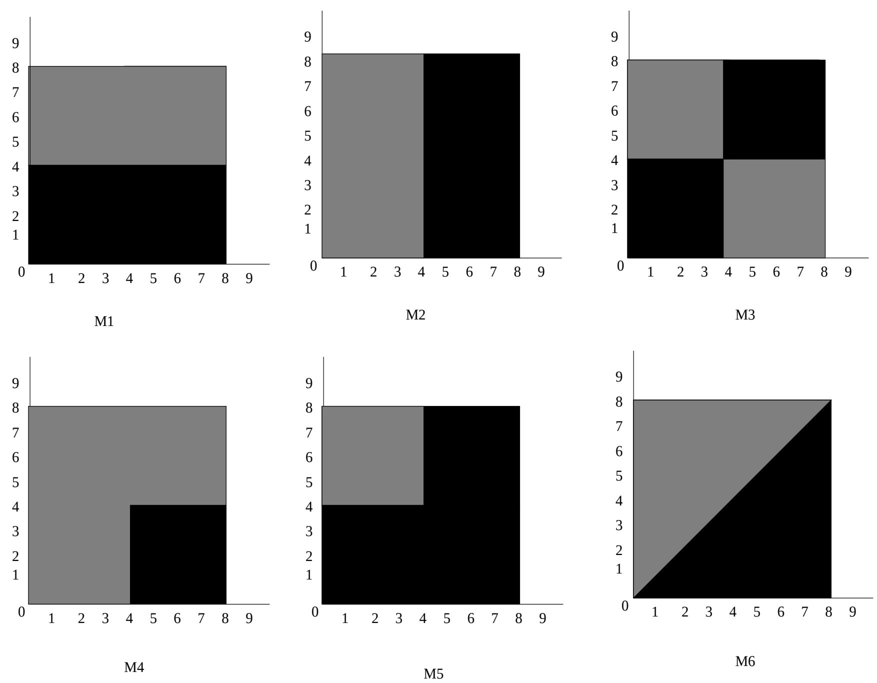
**(A2)** If $\psi$ is empty, then $\psi \triangleright \mu$ is empty.
**(A7)** $Mod((\psi_1 \triangleright \mu) \sqcap (\psi_2 \triangleright \mu)) \ \subseteq \ Mod((\psi_1 \cup \psi_2) \triangleright \mu)$.

Axiom (A7) asserts that any generalized database that is closest to both $\psi_1$ in $\mu$ and to $\psi_2$ in $\mu$ must also be a closest generalized database to $\psi_1 \cup \psi_2$ in $\mu$. This is sometimes called the overall distance requirement. We define the *overall distance* between a generalized knowledgebase $\psi$ and a generalized database $I$ as follows:

$$odist(\psi, I) = \max_{J \in \psi} dist(I, J)$$

Note the change to max from min in the corresponding definition for revisions. Similarly to the previous cases, we define with respect to any generalized knowledgebase $\psi$ a total pre-order $\leq_\psi$ as follows. For each pair of generalized databases $I$ and $J$ let $I \leq_\psi J$ if and only if $odist(\psi, I) \leq odist(\psi, J)$. Then a concrete arbitration operator is the following:

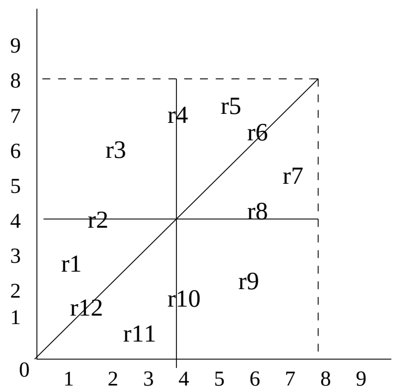
$$\psi \triangleright \mu = Min(\mu, \leq_\psi)$$




**Example 7.1** You are chosen to design the flag of a newly independent country that has two factious parties each suggesting a different flag, namely flags M1 and M2 shown in the figure above. Upon some consideration of the materials available, sewing techniques, dyes, aspect ratios and other esthetic reasons, you limited the choices to M1 and M2 and four other flags shown above. At this point you receive death threats from supporters of both parties "in case you don't choose the right flag". Which flag would you choose?

In this case we need to find out which one of the six possible flags would irk least both parties, i.e., be closest to their flag proposals. That of course we can find out using arbitration. We will calculate $\psi \triangleright \mu$ where $\psi = \{M1, M2\}$ and $\mu = \{M1, M2, M3, M4, M5, M6\}$.

We can represent each flag by a relation $Flag(x, y, z)$ where the $x$ and $y$ will be points in the area of the flag and $z$ its color. Let $z = 10$ be gray and $z = 20$ be black. Since we use $x, y, z$ coordinates we have a 3-dimensional problem. Each flag can be represented as the union of the elementary regions $r_i$ in the gray and $r_{ib}$ in the black plane for $1 \leq i \leq 12$, where the regions in the gray plane are shown below. The elementary regions in the black plane are like the ones in the gray plane but shifted up 10 units.




We give some examples of representing flags using the above partition. We assume that the boundary between any two differently colored regions is black, i.e., they are sewn together by black stitches.

$M1 = r_{1b} \vee r_{2b} \vee r_3 \vee r_4 \vee r_5 \vee r_6 \vee r_7 \vee r_{8b} \vee r_{9b} \vee r_{10b} \vee r_{11b} \vee r_{12b}$
$M2 = r_1 \vee r_2 \vee r_3 \vee r_{4b} \vee r_{5b} \vee r_{6b} \vee r_{7b} \vee r_{8b} \vee r_{9b} \vee r_{10b} \vee r_{11} \vee r_{12}$
$M4 = r_1 \vee r_2 \vee r_3 \vee r_4 \vee r_5 \vee r_6 \vee r_7 \vee r_{8b} \vee r_{9b} \vee r_{10b} \vee r_{11} \vee r_{12}$
$M6 = r_1 \vee r_2 \vee r_3 \vee r_4 \vee r_5 \vee r_{6b} \vee r_{7b} \vee r_{8b} \vee r_{9b} \vee r_{10b} \vee r_{11b} \vee r_{12b}$

Next, let's calculate the distance between some pair of flags.
$dist(M1, M4) = m((M1 \cup M4) \setminus (M1 \cap M4)) = m(r_1 \vee r_{1b} \vee r_2 \vee r_{2b} \vee r_{11} \vee r_{11b} \vee r_{12} \vee r_{12b}) = 2(m(r_1) + m(r_2) + m(r_{11}) + m(r_{12})) = 2((0, 16, -(8 + \sqrt{32}), 0) + (0, 0, 8, 0) + (0, 16, -(8 + \sqrt{32}), 0) + (0, 0, 2\sqrt{32}, 0)) = (0, 64, -16, 0)$

Similarly we can calculate that $dist(M2, M4) = dist(M1, M5) = dist(M2, M5) = (0, 64, -16, 0)$. Also, $dist(M1, M6) = dist(M2, M6) = (0, 64, -16\sqrt{2}, 0)$. After calculating the other distances, it is easy to see that the flags $M4$ and $M5$ give the smallest maximum distances to the two original proposals. Hence after arbitration the new knowledge base will contain these two flags, i.e., $\psi \triangleright \mu = \{M4, M5\}$. (Note that $M6$ will not be in the solution because although its area distance from $M1$ and $M2$ is as good as that of $M4$ and $M5$, its line distance is less than optimal.) □

A generalized assignment is *loyal* if it satisfies the following:

(1) If $Mod(\psi_1) = Mod(\psi_2)$ then $\leq_{\psi_1} = \leq_{\psi_2}$. (2) If $I \leq_{\psi_1} J$ and $I \leq_{\psi_2} J$ then $I \leq_{\psi_1 \cup \psi_2} J$.
The above arbitration operator $\triangleright$ is loyal and satisfies axioms (R1,A2,R3-R6,A7). In general:

**Theorem 7.1** A knowledgebase operator satisfies axioms (R1,A2,R3-R6,A7) if and only if there exists a loyal assignment that maps each knowledge base $\psi$ to a total pre-order $\leq_\psi$ such that $Mod(\psi \triangleright \mu) = Mod(Min(\mu, \leq_\psi))$. □

## 8 Changing a Knowledgebase by First-Order Sentences

The previous sections described revision, update, and arbitration when the new information is a set of models. Grahne et al [GMR92] described a method of updating knowledgebases composed of a set of relational databases by a first-order sentence. [BNR95, Rev96] described similar methods for revision and arbitration by first-order sentences. We can extend these ideas to generalized knowledgebases.

The idea is to allow new informations that are describable in a constraint-query language. For example, if the new information is described in the language of Datalog with dense order constraints, then it describes a query that is evaluable in closed-form on any dense-order constraint database. Hence the new information can be applied to any knowledgebase composed of a set of dense order constraint databases. For example, the updated knowledgebase will be the union of the constraint database outputs obtained when applying the new information to the constraint database inputs.

## 9 Conclusions and Further Work

It would be interesting to test the above operators for real data sets that have a fractal dimension [FK94]. It would be also interesting to compare the operators with other measures, (for example Hausdorff distance [Dou92]) and also to test how well people's assessments correlate with the proposed distance measure and operators.

## References

[AGM85] C. E. Alchourrón, P. Gärdenfors & D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, (50), pages 510–530, 1985.

[BKMS92] C. Baral, S. Kraus, J. Minker & V.S. Subrahmanian. Combining knowledge bases consisting of first-order theories. *Computational Intelligence*, vol. 8, no. 1, pages 45–71, 1992.

[BNR95] A. Benczur, A.B. Novak, & P.Z. Revesz. On Weighted Knowledgebase Transformations. *Proc. Fourth Symposium on Programming Languages and Software Tools*, pp. 404–415, Visegrad, Hungary, June 1995.

[TSIMMIS] S. Chawathe, H. Garcia-Molina, J. Hammer, K. Ireland, Y. Papakonstantinou, J. Ullman, & J. Widom. The TSIMMIS Project: Integration of Heterogeneous Information Sources, *Proc. IPSJ Conf.*, Tokyo, Japan, Oct 1994.

[ChK95] J. Chomicki & G. Kuper. Measuring Infinite Relations. *Fourteenth ACM Symposium on Principles of Database Systems*, pages 246–260, 1995.

[Dou92] E.R. Dougherty. An Introduction to Morphological Image Processing. *SPIE Press*, 1992.

[FKUV86] R. Fagin, G. Kuper, J. D. Ullman & M. Y. Vardi. Updating logical databases. In: *Advances in Computing Research*, Vol. 3, P. C. Kanellakis and F. Preparata eds., pages 1–18, JAI Press, 1986.

[FK94] C. Faloutsos & I. Kamel. Beyond Uniformity and Independence: Analysis of R-tree Using the Concept of Fractal Dimension. *Proc. 13th ACM Symposium on Principles of Database Systems*, pages 4–13, 1994.

[GMR92] G. Grahne, A. O. Mendelzon, & P. Z. Revesz. Knowledgebase transformations. *Journal of Computer and System Sciences*, to appear. (Preliminary version in *Proc. 11th ACM PODS*, 1992.)

[Jag91] H. V. Jagadish. A retrieval technique for similar shapes. *ACM SIGMOD Symposium on the Management of Data*, pages 208-217, 1991.

[KKR95] P. C. Kanellakis, G. M. Kuper, & P. Z. Revesz. Constraint Query Languages, *Journal of Computer and System Sciences*, vol. 51, pages 26-52, 1995. (Preliminary version in *Proc. 9th ACM PODS*, 1990.)

[KM91] H. Katsuno & A. O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52, pages 263–294, 1991.

[KM92] H. Katsuno & A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In: *Belief Revision*, Gärdenfors, ed., pages 183–203, Cambridge University Press, 1992.

[Kup93] G. Kuper. Aggregation in constraint databases. *First Workshop on Principles and Practice of Constraint Programming*, p. 176–183, 1993.

[LS95] P. Liberatore & M. Schaerf. Arbitration: A commutative operator for belief revision. *World Conference on the Fundamentals of Artificial Intelligence*, 1995.

[Rev96] P. Z. Revesz. On the semantics of arbitration. *International Journal of Algebra and Computation*, to appear. (Preliminary version in *Proc. 12th ACM PODS*, 1993.)

[Sub94] V.S. Subrahmanian. Amalgamating Knowledge bases. *ACM Transactions on Database Systems*, Vol 19, no. 2, p. 291–331, 1994.

[Wie92] G. Wiederhold. Mediators in the Architecture of Future Information Systems. *IEEE Computers*, Vol. 25, p. 38–49, 1992.