

2.6 Optimization Function

While the optimization problems from previous section are applicable to an arbitrary monotonic objective function, we now consider a specific optimization criterion in this section in more detail: *maximizing the probability of not violating local constraints*.

We provide an analytical expression of this probability function in terms of a parametric characterization of compact split decompositions, and polyhedron volume function (V) [CH79, Las83, Bea96].

2.6.1 Uniformity Assumptions

Assumptions described in this section are related to how the database instances are distributed over the space defined by Ω . We consider this distribution as a uniform one, i.e., each database instance on Ω has the same probability of occurrence. More formally, our assumptions are as follows:

- Ω is fully dimensional, and therefore $Volume(\Omega) \neq 0$.
- Not using local consistency (lc) property (f_{no-lc}): updates \vec{x} of the database are uniformly distributed on the space defined by constraint Ω . Thus, if $Volume(\Omega) \neq 0$, then

$$prob[\vec{x} \text{ satisfies } \mathbb{C} / \vec{x} \text{ satisfies } \Omega] = \frac{Volume(\mathbb{C})}{Volume(\Omega)}$$

- Using local consistency (lc) property (f_{lc}): we define the following two predicates
 α : \vec{x} satisfy Ω and site i is being updated, and β : \vec{x} satisfies Ω and one of the sites $k+1, \dots, M$ is updated.

1. The probability p_i that a site i is being updated is given, for every $1 \leq i \leq M$. Therefore, for full decomposition:

$$prob[\vec{x} \text{ satisfies } \mathbb{C} / \vec{x} \text{ satisfies } \Omega] = \sum_{i=1}^M p_i \times prob[\vec{x} \text{ satisfies } \mathbb{C} / \alpha]$$

and for θ -decomposition:

$$prob[\vec{x} \text{ satisfies } \mathbb{C} / \beta] = \sum_{i=k+1}^M \frac{p_i}{\sum_{j=k+1}^M p_j} \times prob[\vec{x} \text{ satisfies } \mathbb{C} / \beta]$$

2. The distribution of updates at site i (on variables \vec{y}_i) is uniformly distributed on Ω , when values for all variables, except \vec{y}_i , are fixed. We denote by \vec{z}_i all variables on \vec{x} except those on \vec{y}_i . Then for full decompositions:

$$\begin{aligned} prob[\vec{x} \text{ satisfies } \mathbb{C} / \alpha] &= \frac{Volume(\mathbb{C}[\vec{z}_i \setminus \vec{z}_i^0])}{Volume(\Omega[\vec{z}_i \setminus \vec{z}_i^0])} \\ &= \frac{Volume(C_i)}{Volume(\Omega[\vec{z}_i \setminus \vec{z}_i^0])} \end{aligned}$$

where \vec{z}_i^0 are the values for \vec{z} before the update, $\mathbb{C}[\vec{z}_i \setminus \vec{z}_i^0]$ and $\Omega[\vec{z}_i \setminus \vec{z}_i^0]$ denote the formulas after \vec{z}_i is replaced with \vec{z}_i^0 values.

For θ -compact split decompositions, we also assume that updates on the space defined by $D(\vec{r}_1, \dots, \vec{r}_k)$ are uniformly distributed, and, therefore,

$$\text{prob}[\vec{x} \text{ satisfies } \mathbb{C}/\beta] = \int_{D(\vec{r}_1, \dots, \vec{r}_k)} \frac{\text{Volume}(C_i)}{\text{Volume}(\Omega[\vec{z}_i \setminus \vec{z}_i^0])} d\vec{y}_1^0 \dots d\vec{y}_k^0$$

2.6.2 Parametric Representation

The following is a parametric description of the optimization criteria for the probability of not violating local constraints.

Proposition 10. *Let $\Omega = A\vec{x} \leq \vec{b}$ be a constraint, $\vec{x}^0 = (\vec{y}_1^0, \dots, \vec{y}_M^0)$ be a database instance, p_i be the probability that an update arrives at site i , θ be a subset of sites (say $\{k+1, \dots, M\}$) and $\bar{\theta}$ be its complement, $D(\vec{r}_1, \dots, \vec{r}_k)$ be a (partial) $\bar{\theta}$ -split of Ω , and f_{no-lc} and f_{lc} are the probability of not violating local constraints as defined in Subsection 2.6.1. Then, under the uniformity assumptions of Subsection 2.6.1, the following holds:*

1. *Not using local consistency (lc) property: the function f_{no-lc} is monotonic and has a θ -localizer as follows:*

$$f_{no-lc}^\theta(\vec{r}_{k+1}, \dots, \vec{r}_M) = \prod_{i=k+1}^M V(n_i, A_i, \vec{r}_i)$$

2. *Using local consistency (lc) property: the function f_{lc} is monotonic and has a*

θ -localizer as follows:

$$f_{lc}^\theta(\vec{r}_{k+1}, \dots, \vec{r}_M) = \sum_{i=k+1}^M \frac{p_i}{P} V(n_i, A_i, \vec{r}_i) \times I$$

where $\vec{b}_i = (\vec{b} - \sum_{(j=1, j \neq i)}^M \vec{A}_j \vec{y}_j^0)$, $P = \sum_{j=k+1}^M p_j$, and

$$I = \int_{D(\vec{r}_1, \dots, \vec{r}_k)} \frac{1}{V(n_i, A_i, \vec{b}_i)} d\vec{y}_k^0 \dots d\vec{y}_1^0$$

Proof. First, f is monotonic since it is a probability function. Now, we prove that f has θ -localizers.

Proof of 1 (not using local consistency). We know that the probability to satisfy \mathbb{C} given that Ω is satisfied, is given by:

$$\frac{Volume(\mathbb{C})}{Volume(\Omega)}$$

but, $Volume(\mathbb{C}) = \prod_{i=1}^M Volume(C_i)$, since different C_i 's are defined in disjoint set of variables. We denote $Volume(C_i)$ as $V(n_i, A_i, \vec{r}_i)$. Then,

$$f_{no-lc}(\vec{r}_1, \dots, \vec{r}_M) = \frac{\prod_{i=1}^M V(n_i, A_i, \vec{r}_i)}{V(n, A, \vec{b})}$$

Let $\mathbb{V} = \prod_{i=1}^k V(n_i, A_i, \vec{r}_i)$. Then, it is easy to see that for any two splits of Ω ,

$D(\vec{r}_1, \dots, \vec{r}_k, \vec{r}_{k+1}^{\vec{r}}, \dots, \vec{r}_M^{\vec{r}})$ and $D(\vec{r}_1, \dots, \vec{r}_k, \vec{r}_{k+1}^{\vec{r}'}, \dots, \vec{r}_M^{\vec{r}'})$

$$\begin{aligned} \frac{\mathbb{V} \times \prod_{i=k+1}^M V(n_i, A_i, \vec{r}_i^{\vec{r}})}{V(n, A, \vec{b})} &\geq \frac{\mathbb{V} \times \prod_{i=k+1}^M V(n_i, A_i, \vec{r}_i^{\vec{r}'})}{V(n, A, \vec{b})} \Leftrightarrow \\ \frac{\prod_{i=k+1}^M V(n_i, A_i, \vec{r}_i^{\vec{r}})}{V(n, A, \vec{b})} &\geq \frac{\prod_{i=k+1}^M V(n_i, A_i, \vec{r}_i^{\vec{r}'})}{V(n, A, \vec{b})} \end{aligned}$$

because \mathbb{V} is a non-negative constant. Therefore, f_{no-lc} is monotonic and its θ -localizer is equivalent to

$$f_{no-lc}^{\theta} = \prod_{i=k+1}^M V(n_i, A_i, \vec{r}_i)$$

This completes this part of the proof.

Proof of 2 (using local consistency). We know that the probability to satisfy \mathbb{C} given that Ω is satisfied by \vec{x}^0 , is given by:

$$\sum_{i=1}^M p_i \times \frac{Volume(C_i)}{Volume(\Omega[\vec{z}_i \setminus \vec{z}_i^0])}$$

but, $\Omega[\vec{z}_i \setminus \vec{z}_i^0]$ is equivalent to $A_i \vec{y}_i \leq (\vec{b} - \sum_{(j=1, j \neq i)}^M A_j \vec{y}_j^0)$, and denoting the right hand side by \vec{b}_i . Then, $Volume(\Omega[\vec{z}_i \setminus \vec{z}_i^0])$ is the volume $A_i \vec{y}_i \leq \vec{b}_i$, i.e., $V(n_i, A_i, \vec{b}_i)$. Let $\mathbb{V} = \sum_{i=1}^k p_i V(n_i, A_i, \vec{r}_i) / V(n_i, A_i, \vec{b}_i)$. Then, for any two splits of Ω , $D(\vec{r}_1, \dots, \vec{r}_k,$

$\vec{r}_{k+1}^{\prime}, \dots, \vec{r}_M^{\prime})$ and $D(\vec{r}_1, \dots, \vec{r}_k, \vec{r}_{k+1}^{\prime\prime}, \dots, \vec{r}_M^{\prime\prime})$,

$$\begin{aligned} \mathbb{V} + \sum_{i=k+1}^M p_i \times \frac{V(n_i, A_i, \vec{r}_i^{\prime})}{V(n_i, A_i, \vec{b}_i)} &\geq \mathbb{V} + \sum_{i=k+1}^M p_i \times \frac{V(n_i, A_i, \vec{r}_i^{\prime\prime})}{V(n_i, A_i, \vec{b}_i)} \Leftrightarrow \\ \sum_{i=k+1}^M p_i \times \frac{V(n_i, A_i, \vec{r}_i^{\prime})}{V(n_i, A_i, \vec{b}_i)} &\geq \sum_{i=k+1}^M p_i \times \frac{V(n_i, A_i, \vec{r}_i^{\prime\prime})}{V(n_i, A_i, \vec{b}_i)} \end{aligned} \quad (2.13)$$

because \mathbb{V} is a non-negative constant for $D(\vec{r}_1, \dots, \vec{r}_k, \vec{r}_{k+1}^{\prime}, \dots, \vec{r}_M^{\prime})$ and $D(\vec{r}_1, \dots, \vec{r}_k, \vec{r}_{k+1}^{\prime\prime}, \dots, \vec{r}_M^{\prime\prime})$. Now, we show which is the θ -localizer of f_{lc} . Since, $\vec{b}_i = \vec{b} - \sum_{(j=1, j \neq i)}^M A_j \vec{y}_j^0$, for all i , $1 \leq i \leq M$, function $V(n_i, A_i, \vec{b}_i)$ depends of values $(\vec{y}_1^0, \dots, \vec{y}_k^0)$ and $(\vec{y}_{k+1}^0, \dots, \vec{y}_{i-1}^0, \vec{y}_{i+1}^0, \dots, \vec{y}_M^0)$, for all i , $k+1 \leq i \leq M$. However, $(\vec{y}_1^0, \dots, \vec{y}_k^0)$ are values outside of θ , that satisfy \mathbb{C}_p . Then, for every $(\vec{y}_1^0, \dots, \vec{y}_k^0)$ with those properties,

$$\sum_{i=k+1}^M p_i \times V(n_i, A_i, \vec{r}_i^{\prime}) \int_{D(\vec{r}_1, \dots, \vec{r}_k)} \frac{1}{V(n_i, A_i, \vec{b}_i)} d\vec{y}_1^0 \dots d\vec{y}_k^0$$

Finally, dividing by the constant $\sum_{j=k+1}^M p_j$,

$$\sum_{i=k+1}^M \frac{p_i}{\sum_{j=k+1}^M p_j} \times V(n_i, A_i, \vec{r}_i^{\prime}) \int_{D(\vec{r}_1, \dots, \vec{r}_k)} \frac{1}{V(n_i, A_i, \vec{b}_i)} d\vec{y}_1^0 \dots d\vec{y}_k^0$$

This is the θ -localizer of f_{lc} . This completes the proof. \square

Proposition 10 characterizes the θ -localizer for the probability of not violating local constraints, based on uniformity update assumptions (Section 2.6.1), and convex polyhedron volume $(V(n_i, A_i, \vec{r}_i^{\prime}))$. This volume calculation has been addressed

in [CH79, Las83, Bea96]. Those papers prove that under certain conditions the volume exists, and provide algorithms to compute it.

For the individual partition case, the volume calculation is easy, since local constraints are on individual variables. Those constraints define a (multi-dimensional) rectangle. More specifically,

$$V(C_{k+1}, \dots, C_n) = \prod_{k+1}^n l_i \quad (2.14)$$

where l_i is the length of the i^{th} side of the rectangle. This simple formula gives Proposition 11 below, which is the simplification of Proposition 10 for the individual partition case.

Proposition 11. *Let $\Omega = A\vec{x} \leq \vec{b}$ be a constraint, $\vec{x}^0 = (\vec{x}_1^0, \dots, \vec{x}_n^0)$ be a database instance, p_i be the probability that an update arrives at site i , θ be a subset of sites (say $\{k+1, \dots, n\}$), $\mathbb{C}_p(u_{11}, u_{21}, \dots, u_{1k}, u_{2k})$ be a partial split of Ω , and f_{no-lc} and f_{lc} are the probability of not violating local constraints as defined in Subsection 2.6.1. Then, under the uniformity assumptions of Subsection 2.6.1 hold,*

1. *Not using local consistency (lc) property: the function f_{no-lc} is monotonic and its θ -localizer is as follows:*

$$f_{no-lc}^\theta(u_{1k+1}, u_{2k+1}, \dots, u_{1n}, u_{2n}) = \prod_{i=k+1}^n (u_{2i} - u_{1i})$$

2. Using local consistency (lc) property: the function f_{lc} is monotonic and its θ -localizer is as follows:

$$f_{lc}^{\theta}(u_{1k+1}, u_{2k+1}, \dots, u_{1n}, u_{2n}) = \sum_{i=k+1}^n \frac{p_i \times (u_{2i} - u_{1i})}{\sum_{j=k+1}^n p_j} \times I$$

where

$$I = \int_{x_1=u_{11}}^{u_{21}} \dots \int_{x_1=u_{1k}}^{u_{2k}} \frac{1}{(v_{2i} - v_{1i})} dx_1 \dots dx_k$$

$$v_{1i} = \text{Max}_{(b'_l/a_{li} < 0)} \left\{ \frac{b'_l}{a_{li}} \right\},$$

$$v_{2i} = \text{Min}_{(b'_l/a_{li} > 0)} \left\{ \frac{b'_l}{a_{li}} \right\},$$

$$\vec{b}' = (\vec{b} - \sum_{(j=1, j \neq i)}^n A_j x_j^0), \text{ and}$$

A_j is the j^{th} column of A .

Therefore, optimization problems defined in Theorems 4 and 5 can use f_{no-lc}^{θ} , f_{lc}^{θ} or equivalent objective functions. Note that the optimization problem has linear constraints, and a non-linear objective function. An algorithm to solve this problem is presented in Section 2.8. Now, we present our distributed protocol.