Deciding Set and Multi-set Constraints

Etienne Toussaint

Abstract

Set and multi-set constraints are very common in specification of data structures such as arrays or binary search trees. Then, to verify the correctness of the implementation of such data structures, the verification tools need decision procedures for checking the satisfiability of such constraints. This report provides a survey of the existing decision procedures for the constraints over set and multi-set of integers. Moreover, it reports on an implementation of such decision procedure on the top of existing solvers for arithmetical constraints.

Keywords: Set and multi-set constraints; decision procedures; SMT solvers

1 Introduction

Multi-set (aka bag) constraints are very useful for the specification of constraints over data structures. For example, the inductive definition of a binary search tree storing integer values rooted at E and having the multi-set of values M is specified by the following inductive definition in Separation Logic:

$$\begin{aligned}
\mathtt{bst}(E, M) &::= E = nil \wedge M = [\![]\!] && (1)\\
\mathtt{bst}(E, M) &::= \exists L, R, d.\ E \mapsto \{(\mathtt{left}, L), (\mathtt{right}, R), (\mathtt{data}, d)\} && (2)\\
&\quad \star\, \mathtt{bst}(L, M_L) \star \mathtt{bst}(R, M_R)\\
&\quad \wedge E \neq nil \wedge M_L \leq [\![d]\!] < M_R \wedge M = M_L \oplus [\![d]\!] \oplus M_R
\end{aligned}$$

The first equation specifies the case of an empty tree, where the root E is null and the parameter M is an empty bag. The second equation specifies a non null node at location E where the fields `left`, `right`, and `data` contain as values L, R, resp. d (atom $E \mapsto \{(\mathtt{left}, L), (\mathtt{right}, R), (\mathtt{data}, d)\}$). The locations L and R are roots of binary search trees of values in M_L resp. M_R (atoms $\mathtt{bst}(L, M_L)$ resp. $\mathtt{bst}(R, M_R)$), disjoint pairwise and from the node at E (due to the use of the separation conjunction $\star$). The constraint between the multi-sets of values and the value d, $M_L \leq [\![d]\!] < M_R$, specifies that all values in the left sub-tree are less than d and d is strictly less than all values in the right sub-tree.

This report provides a detailed presentation of two decision procedures for checking satisfiability of quantifier-free formulas including linear integer constraints and constraints over multi-sets of integers. The multi-set constraints include classical atoms, like multi-set inclusion or membership, as well as constraints ordering multi-sets with respect to their elements or constraining the minimum or maximum of a multi-set.

The decision procedure is based on the rewriting of a multi-set formula φ into an equi-satisfiable formula $\tilde{\varphi}$ in quantifier free logic of linear arithmetics. The latter logic is called QFLIA theory in SMTLIB format [2] and it is the input theory of very efficient solvers, e.g., CVC4 [1] or Z3 [3]. Moreover, because these solvers support also an extension of this theory with uninterpreted functions, we provide an alternative rewriting of φ into this theory.

First, we define the logic QFBILIA and its semantic [5]. Second, we explain the procedure reducing QFBILIA formula to QFLIA [6]. Third, we explain the procedure reducing QFBILIA formula to QFUFLIA. Finally, we discuss the implementation choices and some implemented optimisations.

2 Logic QFBILIA

This section presents the logic fragment of *quantifier free logic over bags of integers and linear arithmetics*, QFBILIA. We provide its syntax, its semantics, and the sub-fragments used in the decision procedures described in Sections 3 and 4.

2.1 Syntax

We denote by $\bot$, $\top$ the extremum term such as $\forall k \in \mathbb{Z}, k > \bot$, resp. $k < \top$, we have $\bot < \top$ [5]. We donote by $\mathbb{Z}^{\prec}$ the set $\mathbb{Z} \cup \{\bot, \top\}$. We use the classic notations for operations over $\mathbb{Z}$. Integer constants are denoted by k, natural ones by n. Let $V_{int} = \{a, b, c, \ldots\}$ be a finite set of symbols denoting integer variables, i.e., variables with values in $\mathbb{Z}$. Let $V_{ext} = \{m, n, p, \ldots\}$ be a finite set of symbols denoting extremum variables, i.e., variables with values in $\mathbb{Z}^{\prec}$. We denote by $\mathbb{M}[\mathbb{Z}]$ the domain of bags over integers, i.e., the set of functions $\mathbb{Z} \to \mathbb{N}$. Let $V_{bag} = \{x, y, z, \ldots\}$ be a finite set of symbols denoting bag variables, i.e., variables with values in $\mathbb{M}[\mathbb{Z}]$. We suppose that V_{int}, V_{ext} and V_{bag} are disjoint and we don't write explicitly the type ($\mathbb{Z}$, $\mathbb{Z}^{\prec}$ or $\mathbb{M}[\mathbb{Z}]$) of each variable.

Definition 1. *A QFBILIA formula F is defined by the following grammar:*

$$
\begin{aligned}
F &::= L \mid F \vee F \mid F \wedge F \mid \neg F \mid F \Rightarrow F && \textit{formula}\\
L &::= L_{int} \mid L_{bag} \mid L_{mix} \mid L_{ext} && \textit{boolean atom}\\
L_{int} &::= T_{int} = T_{int} \mid T_{int} \neq T_{int} \mid T_{int} < T_{int} \mid T_{int} \geq T_{int} &&\\
L_{ext} &::= T_{ext} = T_{ext} \mid T_{ext} \neq T_{ext} \mid T_{ext} < T_{ext} \mid T_{ext} \geq T_{ext} &&\\
L_{bag} &::= T_{bag} = T_{bag} \mid T_{bag} \neq T_{bag} \mid T_{bag} \subseteq T_{bag} \mid T_{bag} \not\subseteq T_{bag} \mid &&\\
&\quad\ \ T_{bag} < T_{bag} \mid T_{bag} \geq T_{bag} &&\\
L_{mix} &::= a \in T_{bag} \mid a \notin T_{bag} \mid a \in^{n} T_{bag} \mid a \notin^{n} T_{bag} \mid &&\\
&\quad\ \ m \in T_{bag} \mid m \notin T_{bag} \mid m \in^{n} T_{bag} \mid m \notin^{n} T_{bag} &&\\
T_{int} &::= k \mid a \mid T_{int} + T_{int} \mid T_{int} - T_{int} \mid && \textit{integer term}\\
&\quad\ \ \max(T_{int}, T_{int}) \mid \min(T_{int}, T_{int}) \mid \textit{ite}(F, T_{int}, T_{int}) &&\\
T_{ext} &::= k \mid m \mid \min(T_{bag}) \mid \max(T_{bag}) \mid \textit{ite}(F, T_{ext}, T_{ext}) && \textit{extremum term}\\
T_{bag} &::= \llbracket\rrbracket \mid \llbracket a \rrbracket \mid \llbracket m \rrbracket \mid x \mid T_{bag} \cup T_{bag} \mid T_{bag} \cap T_{bag} \mid && \textit{bag term}\\
&\quad\ \ T_{bag} \setminus T_{bag} \mid T_{bag} \uplus T_{bag} \mid \textit{ite}(F, T_{bag}, T_{bag}) &&
\end{aligned}
$$

We denote by $\mathcal{F}$ the set of formulas in QFBILIA, by $\mathcal{T}_{int}$ the set of integer terms, by $\mathcal{T}_{ext}$ the set of extremum terms, and by $\mathcal{T}_{bag}$ the set of multi-set terms.

For a formula F, we denote by $V_{int}(F)$, $V_{ext}(F)$ and $V_{bag}(F)$ the set of integer, resp. extremum, resp. multi-set variables used in F. We denote by $V(F) = V_{int}(F) \cup V_{ext}(F) \cup V_{bag}(F)$ the set of variables used in F. The same notation is overloaded for atoms and terms in QFBILIA. For a formula F, we denote by $L_{int}(F)$, $L_{bag}(F)$, $L_{ext}(F)$ and $L_{mix}(F)$ the multi-set of integer, resp. multi-set, resp. mix literals in F. We denote by $L(F) = L_{int}(F) \uplus L_{bag}(F) \uplus L_{mix}(F) \uplus L_{ext}(F)$ the multi-set of literals in F.

2.2 Semantics

A valuation I_{int} of variables in V_{int} is a function mapping variables in V_{int} to values in $\mathbb{Z}$, i.e., $I_{int} : V_{int} \to \mathbb{Z}$. Similarly, $I_{ext} : V_{ext} \to \mathbb{Z}^{\prec}$ denotes a valuation of variables in V_{ext} and $I_{bag} : V_{bag} \to \mathbb{M}[\mathbb{Z}]$ denotes a valuation of variables in V_{bag}. A valuation I of variables in $V_{int} \cup V_{bag} \cup V_{ext}$ is a tuple of valuations $(I_{int}, I_{bag}, I_{ext})$; we denote by I_{int}, I_{bag} and I_{ext} the first, resp. second, resp. third component of a valuation I. Let $\mathcal{I}$ be the set of valuations over variables in $V_{int} \cup V_{bag} \cup V_{ext}$. Two valuations I and I' *agree* on a set of variables V iff $\forall v \in V.\ I(v) = I'(v)$. For a valuation I_{bag} of variables in V_{bag} and a bag variable x we denote by $\mathcal{D}(I_{bag}(x)) = \{v \in \mathbb{Z} | I_{bag}(x)(v) > 0\}$ the integers occuring in the bag.

The semantics of QFBILIA is defined by

- a function $^{*} : \mathcal{I} \to \mathcal{T}_{int} \to \mathbb{Z}$ mapping a valuation and an integer term to an integer value,

- a function ${}^{\bullet} : \mathcal{I} \to \mathcal{T}_{ext} \to \mathbb{Z} \cup \{\bot, \top\}$ mapping a valuation and an integer term to an extremum value,
- a function ${}^{\circ} : \mathcal{I} \to \mathcal{T}_{bag} \to \mathbb{M}[\mathbb{Z}]$ mapping a valuation and a multi-set term to a multi-set value,
- a relation $\models \subseteq \mathcal{I} \times \mathcal{F}$ between valuations and formulas.

The components above are defined inductively on the syntax of terms and formulas in a mutually recursive way. This is due to the presence of ite terms.

Given a valuation I and an integer term T_{int}, the valuation of T_{int} in $\mathbb{Z}$, denoted by $I^*(T_{int})$, is defined as follows:

$$\begin{aligned}
I^*(k) &\triangleq k \\
I^*(a) &\triangleq I_{int}(a) \\
I^*(T^1_{int} + T^2_{int}) &\triangleq I^*(T^1_{int}) + I^*(T^2_{int}) \\
I^*(T^1_{int} - T^2_{int}) &\triangleq I^*(T^1_{int}) - I^*(T^2_{int}) \\
I^*(\max(T^1_{int}, T^2_{int})) &\triangleq \max(I^*(T^1_{int}), I^*(T^2_{int})) \\
I^*(\min(T^1_{int}, T^2_{int})) &\triangleq \min(I^*(T^1_{int}), I^*(T^2_{int})) \\
I^*(\mathsf{ite}(F, T^1_{int}, T^2_{int})) &\triangleq \begin{cases} I^*(T^1_{int}) & \text{if } I \models F \\ I^*(T^2_{int}) & \text{otherwise} \end{cases}
\end{aligned}$$

where $\max(k_1, k_2)$ and $\min(k_1, k_2)$ are the minimum and maximum operations over $\mathbb{Z}$.

Given a valuation I and an extremum term T_{ext}, the valuation of T_{ext} in $\mathbb{Z} \cup \{\bot, \top\}$, denoted by $I^\bullet(T_{ext})$, is defined as follows:

$$\begin{aligned}
I^\bullet(k) &\triangleq k \\
I^\bullet(m) &\triangleq I_{ext}(m) \\
I^\bullet(\min(T_{bag})) &\triangleq \begin{cases} \top \text{ if } \forall k', I^\circ(T_{bag})(k') = 0 \\ k \text{ s.t. } I^\circ(T_{bag})(k) > 0 \text{ and } \forall k' < k.\ I^\circ(T_{bag})(k') = 0 \end{cases} \\
I^\bullet(\max(T_{bag})) &\triangleq \begin{cases} \bot \text{ if } \forall k', I^\circ(T_{bag})(k') = 0 \\ k \text{ s.t. } I^\circ(T_{bag})(k) > 0 \text{ and } \forall k' > k.\ I^\circ(T_{bag})(k') = 0 \end{cases} \\
I^\bullet(\mathsf{ite}(F, T^1_{ext}, T^2_{ext})) &\triangleq \begin{cases} I^\bullet(T^1_{ext}) & \text{if } I \models F \\ I^\bullet(T^2_{ext}) & \text{otherwise} \end{cases}
\end{aligned}$$

Given a valuation I and a bag term T_{bag}, the valuation of T_{bag} in $\mathbb{M}[\mathbb{Z}]$, denoted by $I^\circ(T_{bag})$, is defined as follows:

$$\begin{aligned}
I^\circ(\llbracket\rrbracket) &\triangleq M \text{ s.t. } \forall k.\ M(k) = 0 \\
I^\circ(\llbracket a\rrbracket) &\triangleq M \text{ s.t. } M(a) = 1 \text{ and } \forall k \neq a.\ M(k) = 0 \\
I^\circ(\llbracket m\rrbracket) &\triangleq \begin{cases} M \text{ s.t. } M(m) = 1 \text{ and} & \forall k \neq m.\ M(k) = 0 \\ & \text{if } I_{ext}(m) \notin \{\bot, \top\} \\ M \text{ s.t. } \forall k.\ M(k) = 0 & \text{otherwise} \end{cases} \\
I^\circ(x) &\triangleq I_{bag}(x) \\
I^\circ(T^1_{bag} \cup T^2_{bag}) &\triangleq M \text{ s.t. } \forall k.\ M(k) = \max(I^\circ(T^1_{bag})(k), I^\circ(T^2_{bag})(k)) \\
I^\circ(T^1_{bag} \cap T^2_{bag}) &\triangleq M \text{ s.t. } \forall k.\ M(k) = \min(I^\circ(T^1_{bag})(k), I^\circ(T^2_{bag})(k)) \\
I^\circ(T^1_{bag} \setminus T^2_{bag}) &\triangleq M \text{ s.t. } \forall k.\ M(k) = \max(0, I^\circ(T^1_{bag})(k) - I^\circ(T^2_{bag})(k)) \\
I^\circ(T^1_{bag} \uplus T^2_{bag}) &\triangleq M \text{ s.t. } \forall k.\ M(k) = I^\circ(T^1_{bag})(k) + I^\circ(T^2_{bag})(k) \\
I^\circ(\mathsf{ite}(F, T^1_{bag}, T^2_{bag})) &\triangleq \begin{cases} I^\circ(T^1_{bag}) & \text{if } I \models F \\ I^\circ(T^2_{bag}) & \text{otherwise} \end{cases}
\end{aligned}$$

Given a valuation I and an integer atom L_{int}, the satisfiability relation $I \models L_{int}$ is defined by structural induction as follows:

$$\begin{aligned}
I \models T^1_{int} = T^2_{int} &\text{ iff } I^*(T^1_{int}) = I^*(T^2_{int}) \\
I \models T^1_{int} \neq T^2_{int} &\text{ iff } I^*(T^1_{int}) \neq I^*(T^2_{int}) \\
I \models T^1_{int} < T^2_{int} &\text{ iff } I^*(T^1_{int}) < I^*(T^2_{int}) \\
I \models T^1_{int} \geq T^2_{int} &\text{ iff } I^*(T^1_{int}) \geq I^*(T^2_{int})
\end{aligned}$$

Given a valuation I and an integer atom L_{ext}, the satisfiability relation $I \models L_{ext}$ is defined by structural induction as follows:

$$\begin{aligned}
I \models T^1_{ext} = T^2_{ext} &\text{ iff } I^\bullet(T^1_{ext}) = I^\bullet(T^2_{ext}) \\
I \models T^1_{ext} \neq T^2_{ext} &\text{ iff } I^\bullet(T^1_{ext}) \neq I^\bullet(T^2_{ext}) \\
I \models T^1_{ext} < T^2_{ext} &\text{ iff } I^\bullet(T^1_{ext}) < I^\bullet(T^2_{ext}) \\
I \models T^1_{ext} \geq T^2_{ext} &\text{ iff } I^\bullet(T^1_{ext}) \geq I^\bullet(T^2_{ext})
\end{aligned}$$

Given a valuation I and a bag atom L_{bag}, the satisfiability relation $I \models L_{bag}$ is defined by structural induction as follows:

$$\begin{aligned}
I \models T^1_{bag} = T^2_{bag} &\text{ iff } \forall k.\ I^\circ(T^1_{bag})(k) = I^\circ(T^2_{bag})(k) \\
I \models T^1_{bag} \neq T^2_{bag} &\text{ iff } \exists k.\ I^\circ(T^1_{bag})(k) \neq I^\circ(T^2_{bag})(k) \\
I \models T^1_{bag} \subseteq T^2_{bag} &\text{ iff } \forall k.\ I^\circ(T^1_{bag})(k) \leq I^\circ(T^2_{bag})(k) \\
I \models T^1_{bag} \nsubseteq T^2_{bag} &\text{ iff } \exists k.\ I^\circ(T^1_{bag})(k) > I^\circ(T^2_{bag})(k) \\
I \models T^1_{bag} < T^2_{bag} &\text{ iff } I^\circ(\max(T^1_{bag})) < I^\circ(\min(T^2_{bag})) \\
I \models T^1_{bag} \geq T^2_{bag} &\text{ iff } I^\circ(\min(T^1_{bag})) \geq I^\circ(\max(T^2_{bag}))
\end{aligned}$$

Given a valuation I and a mixed atom L_{mix}, the satisfiability relation $I \models L_{mix}$ is defined by structural induction as follows:

$$
\begin{aligned}
I \models a \in T^2_{bag} &\text{ iff } I^\circ(T^2_{bag})\big(I_{int}(a)\big) \geq 1\\
I \models a \notin T^2_{bag} &\text{ iff } I^\circ(T^2_{bag})\big(I_{int}(a)\big) = 0\\
I \models a \in^n T^2_{bag} &\text{ iff } I^\circ(T^2_{bag})\big(I_{int}(a)\big) \geq n\\
I \models a \notin^n T^2_{bag} &\text{ iff } I^\circ(T^2_{bag})\big(I_{int}(a)\big) < n\\
I \models m \in T^2_{bag} &\text{ iff } \begin{cases} I^\circ(T^2_{bag})\big(I_{ext}(m)\big) \geq 1 & \text{if } I_{ext}(m) \notin \{\bot, \top\}\\ True & \text{if } \forall k.\ I^\circ(T^2_{bag})\big(k\big) = 0\\ & \quad \wedge\ (I_{ext}(m) \in \{\bot, \top\})\\ False & \text{otherwise} \end{cases}\\
I \models m \notin T^2_{bag} &\text{ iff } \begin{cases} I^\circ(T^2_{bag})\big(I_{ext}(m)\big) = 0 & \text{if } I_{ext}(m) \notin \{\bot, \top\}\\ False & \text{if } \forall k.\ I^\circ(T^2_{bag})\big(k\big) = 0\\ & \quad \wedge\ (I_{ext}(m) \in \{\bot, \top\})\\ True & \text{otherwise} \end{cases}\\
I \models m \in^n T^2_{bag} &\text{ iff } \begin{cases} I^\circ(T^2_{bag})\big(I_{ext}(m)\big) \geq n & \text{if } I_{ext}(m) \notin \{\bot, \top\}\\ True & \text{if } \forall k.\ I^\circ(T^2_{bag})\big(k\big) = 0 \wedge n = 1\\ & \quad \wedge\ (I_{ext}(m) \in \{\bot, \top\})\\ False & \text{otherwise} \end{cases}\\
I \models m \notin^n T^2_{bag} &\text{ iff } \begin{cases} I^\circ(T^2_{bag})\big(I_{ext}(m)\big) < n & \text{if } I_{ext}(m) \notin \{\bot, \top\}\\ False & \text{if } \forall k.\ I^\circ(T^2_{bag})\big(k\big) = 0 \wedge n = 1\\ & \quad \wedge\ (I_{ext}(m) \in \{\bot, \top\})\\ True & \text{otherwise} \end{cases}
\end{aligned}
$$

Given a valuation I and a formula F, the satisfiability relation $I \models F$ is defined by induction on the form of the formula as follows:

$$
\begin{aligned}
I \models L &\text{ iff } I \models L\\
I \models F_1 \vee F_2 &\text{ iff } I \models F_1 \text{ or } I \models F_2\\
I \models F_1 \wedge F_2 &\text{ iff } I \models F_1 \text{ and } I \models F_2\\
I \models \neg F &\text{ iff } I \not\models F\\
I \models F_1 \Rightarrow F_2 &\text{ iff } I \models \neg F_1 \text{ or } I \models F_2
\end{aligned}
$$

Definition 2. *A QFBILIA formula F is* satisfiable *if there exists a valuation I, called also its model, such that $I \models F$.*

Let denote by $[\![F]\!]$ the set of models of F. Thus, a formula F is satisfiable if $[\![F]\!] \neq \emptyset$.

Definition 3. *Two formula F and F' are* semi-equivalent*, denoted $F \sim F'$, if $V(F) \subseteq V(F')$ or $V(F') \subseteq V(F)$ and $\forall I \in [\![F]\!], \exists I' \in [\![F']\!]$ that agrees on $V(F) \cap V(F')$, and $\forall I' \in [\![F']\!], \exists I \in [\![F]\!]$ that agrees on $V(F) \cap V(F')$.*

Definition 4. *Two formula F and F' are* equi-satisfiable*, denoted $F \sim_{sat} F'$, if $[\![F]\!] \neq \emptyset$ if and only if $[\![F']\!] \neq \emptyset$.*

2.3 Examples

Example 1. *If an integer a is in a multi-set x, the multi-set $[\![a]\!]$ is included in the union of multi-sets x and y:*

$$(a \in x) \Rightarrow ([\![a]\!] \subseteq (x \cup y)) \tag{3}$$

Example 2. *If 1 is the smallest element of a multi-set x then x is less than y otherwise x is greater than z:*

$$\big((\min(x) = 1) \Rightarrow (x < y)\big) \wedge \big((\min(x) \neq 1) \Rightarrow (x < z)\big) \tag{4}$$

Example 3. *If the maximum of a multi-set x is 0 then x is the empty bag or the minimum of x is not 0:*

$$(\max(x) = 0) \Rightarrow \big((x = [\![]\!]) \vee (\min(x) \neq 0)\big) \tag{5}$$

2.4 Main Fragments

The QFBILIA logic contains several fragments that are interesting for its reduction to QFLIA theory because they are simpler in syntax but they are, in general, as expressive as QFBILIA.

2.4.1 Fragment QFBILIA$_{\mathsf{nnf}}$

The QFBILIA$_{\mathsf{nnf}}$ fragment contains only formulas in the negated normal form, i.e., the logical negation is pushed to the level of literals $L_{int}, L_{bag}, L_{mix}, L_{ext}$ and ite terms are not present.

Definition 5. *A QFBILIA$_{\mathsf{nnf}}$ formula F is defined by the following grammar:*

$$\begin{aligned}
F &::= L \mid F \vee F \mid F \wedge F \\
L &::= L_{int} \mid L_{bag} \mid L_{mix} \mid L_{ext} \\
L_{int} &::= T_{int} = T_{int} \mid T_{int} \neq T_{int} \mid T_{int} < T_{int} \mid T_{int} \geq T_{int} \\
L_{ext} &::= T_{ext} = T_{ext} \mid T_{ext} \neq T_{ext} \mid T_{ext} < T_{ext} \mid T_{ext} \geq T_{ext} \\
L_{bag} &::= T_{bag} = T_{bag} \mid T_{bag} \neq T_{bag} \mid T_{bag} \subseteq T_{bag} \mid T_{bag} \nsubseteq T_{bag} \mid \\
&\quad\; T_{bag} < T_{bag} \mid T_{bag} \geq T_{bag} \\
L_{mix} &::= T_{int} \in T_{bag} \mid T_{int} \notin T_{bag} \mid T_{int} \in^{n} T_{bag} \mid T_{int} \notin^{n} T_{bag} \\
T_{int} &::= k \mid a \mid T_{int} + T_{int} \mid T_{int} - T_{int} \mid \\
&\quad\; \max(T_{int}, T_{int}) \mid \min(T_{int}, T_{int}) \\
T_{ext} &::= k \mid m \mid \min(T_{bag}) \mid \max(T_{bag}) \\
T_{bag} &::= [\![]\!] \mid [\![a]\!] \mid x \mid T_{bag} \cup T_{bag} \mid T_{bag} \cap T_{bag} \mid T_{bag} \setminus T_{bag} \mid T_{bag} \uplus T_{bag}
\end{aligned}$$

We denote by $\mathcal{F}_{\mathsf{nnf}}$, $\mathcal{T}_{int,\mathsf{nnf}}$, $\mathcal{T}_{ext,\mathsf{nnf}}$, and $\mathcal{T}_{bag,\mathsf{nnf}}$ the set of formulas, integer terms, resp. extremum terms, resp. multi-set terms in QFBILIA$_{\mathsf{nnf}}$.

Thus, the examples (3), (6), and (5) are rewritten in $\mathsf{QFBILIA}_{\mathsf{nnf}}$ as follows:

$$(a \notin x) \vee (\llbracket a \rrbracket \subseteq (x \cup y)) \tag{6}$$

$$\big((\min(x) \neq 1) \vee (x < y)\big) \wedge \big((\min(x) = 1) \vee (x < z)\big) \tag{7}$$

$$(\max(x) \neq 0) \vee \big((x = \llbracket\rrbracket) \vee (\min(x) \neq 0)\big) \tag{8}$$

Proposition 1 (Section 3.1) states that $\mathsf{QFBILIA}$ and $\mathsf{QFBILIA}_{\mathsf{nnf}}$ have the same expressive power, i.e., for any formula F in $\mathsf{QFBILIA}$, there exists an equivalent formula $\tilde{F}$ in $\mathsf{QFBILIA}_{\mathsf{nnf}}$.

2.4.2 Fragment $\mathsf{QFBILIA}_{pure}$

The $\mathsf{QFBILIA}_{pure}$ fragment allows only a restricted syntax for multi-set atoms and terms. Intuitively, we restrict the syntax to essential terms, where no syntactic sugar can be applied to simplify them. For example, the multi-set terms of the form $x \cup y \cup z$ are equivalent (semantically speaking) to a term $x \cup y'$ and the atom $y' = y \cup z$, where y' is a fresh multi-set variable. Also, the atom $x \leq y$ is equivalent (wrt the semantics, i.e., allows same set of models) to the term $\max(x) \leq \min(y)$. The integer atoms using multi-set variables are reduced to only two equality constraints between an integer variable and the min or max of some multi-set variable. Thus, the fragment $\mathsf{QFBILIA}_{pure}$ contains only formulas with the simpler multi-set terms which are furthermore in negative normal form.

Definition 6. *A $\mathsf{QFBILIA}_{pure}$ formula F is defined by the following grammar:*

$$\begin{aligned}
F &::= L \mid F \vee F \mid F \wedge F \\
L &::= L_{int} \mid L_{bag} \mid L_{mix} \mid L_{ext} \\
L_{int} &::= T_{int} = T_{int} \mid T_{int} \neq T_{int} \mid T_{int} < T_{int} \mid T_{int} \geq T_{int} \\
L_{ext} &::= T_{ext} = T_{ext} \mid T_{ext} \neq T_{ext} \mid T_{ext} < T_{ext} \mid T_{ext} \geq T_{ext} \\
L_{bag} &::= x = T_{bag} \mid x \neq y \mid x \subseteq y \mid x \nsubseteq y \\
L_{mix} &::= a \in x \mid a \notin x \mid a \in^n x \mid a \notin^n x \\
T_{int} &::= k \mid a \mid T_{int} + T_{int} \mid T_{int} - T_{int} \mid \max(T_{int}, T_{int}) \mid \min(T_{int}, T_{int}) \\
T_{ext} &::= k \mid m \mid min(x) \mid max(x) \\
T_{bag} &::= \llbracket\rrbracket \mid \llbracket a \rrbracket \mid x \mid x \cup y \mid x \cap y \mid x \setminus y \mid x \uplus y
\end{aligned}$$

The examples (6)–(8) are semi-equivalent with the following formulas in $\mathsf{QFBILIA}_{pure}$:

$$\big((a \notin x) \vee (z_1 \subseteq z_2)\big) \wedge (\llbracket a \rrbracket = z_1) \wedge (z_2 = (x \cup y)) \tag{9}$$

$$\begin{aligned}
&\big((b_1 \neq 1) \vee (b_2 < b_3)\big) \wedge \big((b_1 = 1) \vee (b_2 < b_4)\big) \\
&\wedge (b_1 = \min(x)) \wedge (b_2 = \max(x)) \wedge (b_3 = \min(y)) \wedge (b_4 = \min(z))
\end{aligned} \tag{10}$$

$$(b_1 \neq 0) \vee \big((x = \llbracket\rrbracket) \vee (b_1 \neq 0)\big) \wedge (b_1 = \min(x)) \tag{11}$$

We might loose equivalency between formula due to the creation of variable, however semi-equivalency is preserved.

Proposition 2 (Section 3.2) states that $\mathsf{QFBILIA}_{pure}$ and $\mathsf{QFBILIA}_{\mathsf{nnf}}$ have the same expressive power.

3 Decision procedure by reduction to QFLIA

The decision procedure for checking satisfiability of a QFBILIA formula F proceeds in four steps. These steps are removing gradually the multi-set atoms to obtain an equi-satisfiable formula in quantifier free linear arithmetics (QFLIA theory).

The first step translates F into a $\mathsf{QFBILIA}_{\mathsf{nnf}}$ formula $\tilde{F}$ which is equivalent with F, from Proposition 1. The second step transforms $\tilde{F}$ into an semi-equivalent formula $\tilde{F}^2$ in $\mathsf{QFBILIA}_{pure}$, from Proposition 2. The third step introduces two extremum variables to represent $\perp$ and $\top$, two sets of fresh integer variables to represent (i) the elements that validates atoms $x \neq y$ and $x \not\subseteq y$ and (ii) for each multi-set variable $x \in V_{bag}(\tilde{F}^2)$ and integer variable $a \in V_{int}(\tilde{F}^2)$, resp. extremum variable $m \in V_{ext}(\tilde{F}^2)$, an integer variable $w_{a,x}$, resp. $w_{m,x}$ counting the occurrences of values of a, resp. m in x. This step only adds to $\tilde{F}^2$ a formula F_3 over integer variables. The formula $\tilde{F}^2 \wedge F_3$ is then equi-satisfiable to F. The fourth step removes multi-set terms from $\tilde{F}^3$ using the fresh integer variables introduced by the counting abstraction and produces a formula $\tilde{F}^4 \wedge F_3$ equi-satisfaisable to F.

3.1 First Step: Translation to $\mathsf{QFBILIA}_{\mathsf{nnf}}$

The correctness of this step is stated by the following proposition whose proof gives a procedure to build a formula $\tilde{F}$ in $\mathsf{QFBILIA}_{\mathsf{nnf}}$ from a formula F in QFBILIA.

Proposition 1. *For any F formula in QFBILIA, there exists a formula $\tilde{F}$ in $\mathsf{QFBILIA}_{\mathsf{nnf}}$ over a set of variables $V_{int}(F) \cup V_{int}^{nnf}$ and $V_{bag}(F) \cup V_{bag}^{nnf}$ such that (i) any model of $\tilde{F}$ is a model of F and (ii) for any model of F, there is a model of $\tilde{F}$ that agrees on $V_{int}(F)$ and $V_{bag}(F)$.*

Proof. The proof is given by construction of $\tilde{F}$ from F.

First, every integer term $\mathsf{ite}(F', T^1, T^2)$ is replaced by a fresh integer variable a (i.e., not in $V_{int}(F)$) and F is rewritten in $F \wedge (F' \Rightarrow a = T^1) \wedge \big((\neg F') \Rightarrow (a = T^2)\big)$. The same procedure is applied for ite terms in multi-set atoms. Using the semantics of ite terms, we show that the formula $\tilde{F}$ obtained by this rewriting satisfies the conclusion of the theorem.

Second, we apply to $\tilde{F}$ the de Morgan's rules to eliminate the negation by pushing them at the level of literals. We definite inductively on the formula syntax a function nnf of $\mathsf{QFBILIA} \to \mathsf{QFBILIA}_{\mathsf{nnf}}$, for any $F_1, F_2 \in \mathsf{QFBILIA}$ and

$\ell \in L$, as follows:

$$\begin{aligned}
\mathsf{nnf}(F_1 \Rightarrow F_2) &= \mathsf{nnf}(\neg F_1) \vee \mathsf{nnf}(F_2) && (12)\\
\mathsf{nnf}(\neg(F_1 \Rightarrow F_2)) &= \mathsf{nnf}(F_1) \wedge \mathsf{nnf}(\neg F_2) && (13)\\
\mathsf{nnf}(F_1 \vee F_2) &= \mathsf{nnf}(F_1) \vee \mathsf{nnf}(F_2) && (14)\\
\mathsf{nnf}(\neg(F_1 \vee F_2) &= \mathsf{nnf}(\neg F_1) \wedge \mathsf{nnf}(\neg F_2) && (15)\\
\mathsf{nnf}(F_1 \wedge F_2) &= \mathsf{nnf}(F_1) \wedge \mathsf{nnf}(F_2) && (16)\\
\mathsf{nnf}(\neg(F_1 \wedge F_2) &= \mathsf{nnf}(\neg F_1) \vee \mathsf{nnf}(\neg F_2) && (17)\\
\mathsf{nnf}(\neg\neg F) &= \mathsf{nnf}(F) && (18)\\
\mathsf{nnf}(\ell) &= \ell && (19)\\
\mathsf{nnf}(\neg\ell) &= \tilde{\ell} \text{ i.e., the opposite atom of } \ell && (20)\\
& && (21)
\end{aligned}$$

Since any atom has its opposice in $\mathsf{QFBILIA}_{\mathsf{nnf}}$, and since $(F \wedge F)$, $(F \vee F)$ are also formulas in $\mathsf{QFBILIA}_{\mathsf{nnf}}$, then $\mathsf{nnf}(F) \in \mathsf{QFBILIA}_{\mathsf{nnf}}$, (for instace $\neg(T^1_{int} < T^2_{int})$ is replaced by $(T^1_{int} \geq T^2_{int})$). □

3.2 Second Step: Translation to $\mathsf{QFBILIA}_{pure}$

The step applies the following sequence of transformations on $\tilde{F}$ in $\mathsf{QFBILIA}_{\mathsf{nnf}}$:

S2.1: Every bag atom $T^1 \prec T^2$ constraint (with $\prec \in \{<, \geq\}$) is rewritten using the min and max extremum terms as follows:

- $T^1 \geq T^2$ becomes $\min(T^1) \geq \max(T^2)$
- $T^1 < T^2$ becomes $\max(T^1) < \min(T^2)$

Thus, the bag atom becomes an extremum atom. Let denote by $\tilde{F}_{2.1}$ the result of this process.

S2.2: Every T_{bag} term used in *integer, mixed, extremum* atoms which is not a variable is replaced by a fresh multi-set variable x and the multi-set atom $x = T_{bag}$ is conjuncted to $\tilde{F}_{2.1}$. Let us denote by $\tilde{F}_{2.2}$ the result of this process and by $V^{2.2}_{bag}$ the set of fresh bag variables.

S2.3: Every extremum term of the form $\min(x)$ or $\max(x)$ is replaced by a fresh extremum variable $\min_x$ rep. $\max_x$ and the mixed atom $\min(x) = \min_x$ is conjuncted to $\tilde{F}_{2.2}$. Let us denote by $\tilde{F}_{2.3}$ the result of this process and by $V^{2.3}_{ext}$ the set of fresh extremum variables.

S2.4: Every bag atom L_{bag} of $\tilde{F}_{2.3}$ using more than two bag variables for operations $\neq, \subseteq, \nsubseteq$ and more than three variables for operation $=$ are iteratively rewritten to be reduced to the bag atoms in $\mathsf{QFBILIA}_{pure}$. Notice that integer, mixed and extremum atoms already use only bag variables and not bag terms due to the step S2.2. So, at the end of this step, the resulting formula $\tilde{F}_{2.4}$ is in the $\mathsf{QFBILIA}_{pure}$ fragment. The set of fresh bag variables is denoted by $V^{2.4}_{bag}$

The formula $\tilde{F}_{2.4}$ has the set of integer variables $V_{int}(\tilde{F})$, the set of extremum variables $V_{ext}(\tilde{F}) \cup V_{ext}^{2.3}$ and the set of bag variables $V_{bag}(\tilde{F}) \cup V_{bag}^{2.2} \cup V_{bag}^{2.4}$.

Example 4. *We explain the previously mentioned steps on the following* $\mathsf{QFBILIA}_{\mathsf{nnf}}$ *formula:*

$$\min(x \cup y \cup z) = \min(x \cap y \cap z).$$

It is translated in the following $\mathsf{QFBILIA}_{pure}$ *formula over the set of bag variables* $\{x, y, z\} \cup \{xunionyunionz, yunionz, xinteryinterz, yinterz\}$ *and the set of integer variables* $\{min_{xunionyunionz}, min_{xinteryinterz}\}$ *:*

$$\begin{aligned} min_{xunionyunionz} = min_{xinteryinterz} \ \wedge \\ min_{xunionyunionz} = min(xunionyunionz) \ \wedge \\ min_{xinteryinterz} = min(xinteryinterz) \ \wedge \\ xunionyunionz = x \cup yunionz \ \ \wedge yunionz = y \cup z \ \wedge \\ xinteryinterz = x \cap yinterz \ \ \wedge yinterz = y \cap z \end{aligned}$$

The correctness of this transformation with respect to the semi-equivalence is stated by the following result.

Proposition 2. *For any* $\tilde{F}$ *formula in* $\mathsf{QFBILIA}_{\mathsf{nnf}}$*, there exists a formula* $\tilde{F}^2$ *in* $\mathsf{QFBILIA}_{pure}$ *over a set of variables* $V_{int}(\tilde{F})$*,* $V_{ext}(\tilde{F}) \cup V_{ext}^p$ *and* $V_{bag}(\tilde{F}) \cup V_{bag}^p$ *such that (i) any model of* $\tilde{F}^2$ *is a model of* $\tilde{F}$ *and (ii) for any model of* $\tilde{F}$*, there is a model of* $\tilde{F}^2$ *that agrees on* $V_{int}(\tilde{F})$ *and* $V_{bag}(\tilde{F})$*.*

Proof. The proof is given by construction of $\tilde{F}^2$ from $\tilde{F}$ following the steps S2.1–4 above.

The formula $\tilde{F}_{2.1}$ is equivalent to $\tilde{F}$ due to the definition of the semantics of comparison operations over multi-sets.

The formula $\tilde{F}_{2.2}$ is built over the set of variables $V_{int}(\tilde{F})$ and $V_{bag}(\tilde{F}) \cup V_{bag}^{2.2}$ where $V_{bag}^{2.2}$ is the set of fresh variables used to replace multi-set terms which are not variables in integer, extremum or mixed atoms. The semantics of equality between bags ensures that the equi-satisfiability is preserved.

The formula $\tilde{F}_{2.3}$ is built over the set of variables $V_{int}(\tilde{F}_{2.2})$, $V_{ext}(\tilde{F}_{2.2}) \cup V_{ext}^{2.3}$ and $V_{bag}(\tilde{F}_{2.2})$ where $V_{ext}^{2.3}$ is the set of fresh variables used to replace extremum term of the form $\min(x)$ or $\max(x)$. The semantics of equality between integer ensures that the equi-satisfiability is preserved.

The formula $\tilde{F}_{2.4}$ is built over the set of variables $V_{int}(\tilde{F}_{2.3})$ and $V_{bag}(\tilde{F}_{2.3}) \cup V_{bag}^{2.4}$ where $V_{bag}^{2.4}$ is the set of fresh variables used to replace multi-set terms in bag atoms which don't fit in the $\mathsf{QFBILIA}_{pure}$ fragment. The semantics of equality between bags ensures that the equi-satisfiability is preserved, i.e., the two conclusions are valid.

The formula $\tilde{F}^2 = \tilde{F}_{2.4}$ is in $\mathsf{QFBILIA}_{pure}$ and the two conclusions are valid as a consequence of steps S2.1–4.

□

3.3 Third Step: Introducing the Counting Abstraction

This step adds to $\tilde{F}^2$ a formula F_3 that introduces the integer variables which will replace the multi-set variables. The transformation has three steps:

S3.1: Build the set V_{int}^{31} as a set of fresh variables, one variable for each atom $(x \neq y)$ or $(x \nsubseteq y)$ in $\tilde{F}^2$. We denote these variables by $a_{x \neq y}$, resp. $a_{x \nsubseteq y}$. Intuitively, these variables are introduced to be able to express the fact that there is a value on which x and y differ, resp. x has more copies than y (see subsection 3.4).

S3.2: We introduce two fresh extremum variables $m_\perp$ and $m_\top$. Intuitively, these variables are introduced to be able to express $\perp$, resp. $\top$.

S3.3: Build the set

$$V_{int}^{33} = \bigcup_{x \in V_{bag}(\tilde{F}^2), a \in V_{int}^{31} \cup V_{int}(\tilde{F}^2)} w_{a,x},$$

where an integer variable is added for each pair of bag variable x and integer variable a in order to represent the number of element a in x.

$$V_{ext}^{33} = \bigcup_{x \in V_{bag}(\tilde{F}^2), m \in V_{ext}(\tilde{F}^2)} w_{m,x},$$

where an integer variable is added for each pair of bag variable x and extremum variable m in order to represent the number of element m in x. Let $V_{int}^3 = V_{int}(\tilde{F}^2) \cup V_{int}^{31}$. Let $V_{ele}^3 = V_{int}^3 \cup V_{ext}(\tilde{F}^2)$.

The formula $\tilde{F}^3$ is built as follows:

$$\tilde{F}^3 \triangleq \tilde{F}^2 \wedge F_3 \tag{22}$$

$$F_3 \triangleq \left(\bigwedge_{a \in V_{int}^3} (m_\perp < a < m_\top) \right) \tag{23}$$

$$\wedge \left(\bigwedge_{m \in V_{ext}(\tilde{F}^2)} (m_\top \geq m \geq m_\perp) \right) \tag{24}$$

$$\wedge \left(\bigwedge_{w \in V_{int}^{33} \cup V_{ext}^{33}} (w \geq 0) \right) \tag{25}$$

$$\wedge \left(\bigwedge_{t,u \in V_{ele}^3} \left(\bigwedge_{x \in V_{bag}(\tilde{F}^2)} (t \neq u) \vee (w_{t,x} = w_{u,x}) \right) \right) \tag{26}$$

$$\tag{27}$$

The set of variables of $\tilde{F}^3$ are $V_{int}(\tilde{F}^3)$, resp. $V_{ext}(\tilde{F}^3)$, resp. $V_{bag}(\tilde{F}^2)$.

The following property states that $\tilde{F}^3 \sim_{sat} \tilde{F}^2$:

Proposition 3. *The formula $\tilde{F}^3$ is in $\mathsf{QFBILIA}_{pure}$ and (i) any model of $\tilde{F}^3$ is a model of $\tilde{F}^2$ and (ii) for any model of $\tilde{F}^2$, there is a model of $\tilde{F}^3$ that agrees on $V_{int}(\tilde{F}^2)$, $V_{ext}(\tilde{F}^2)$ and $V_{bag}(\tilde{F}^2)$.*

Proof. Clearly $\tilde{F}^3$ is still in $\mathsf{QFBILIA}_{pure}$.

Moreover, because $\tilde{F}^3 \triangleq \tilde{F}^2 \wedge F_3$, any model I of $\tilde{F}^3$ is also a model of $\tilde{F}^2$.

We prove now the point (ii):

A.1: I be a model of $\tilde{F}^2$

C.1: There is I' which agrees on $V(\tilde{F}^2)$ to I and I' is a model of $\tilde{F}^3$.

Proof:

1: Let $I' = (I_{bag}, I'_{int}, I'_{ext})$ such that for any $a \in V_{int}(\tilde{F}^3)$ we have

$$I'_{int}(a) \triangleq \begin{cases} I_{int}(a) & \text{if } a \in V_{int}(\tilde{F}^2) \\ v & \text{if } a_{x \neq y} \in V^{31}_{int} \text{ and } I_{bag}(x)(v) \neq I_{bag}(y)(v) \\ v & \text{if } a_{x \not\subseteq y} \in V^{31}_{int} \text{ and } I_{bag}(x)(v) > I_{bag}(y)(v) \\ I_{bag}(x)(b) & \text{if } a \equiv w_{b,x} \in V^{32}_{int} \\ 0 & \text{otherwise} \end{cases}$$

for any $m \in V_{ext}(\tilde{F}^3)$ we have

$$I'_{ext}(m) \triangleq \begin{cases} I_{ext}(m) & \text{if } m \in V_{ext}(\tilde{F}^2) \\ k & \text{if } m_\perp \text{ and } \forall a \in V_{int}(\tilde{F}^3), k < I'_{int}(a) \\ & \qquad \text{and } \forall m \in V_{ext}(\tilde{F}^3), k \leq I'_{ext}(m) \\ k & \text{if } m_\top \text{ and } \forall a \in V_{int}(\tilde{F}^3), k > I'_{int}(a) \\ & \qquad \text{and } \forall m \in V_{ext}(\tilde{F}^3), k \geq I'_{ext}(m) \end{cases}$$

This valuation of integer and extremum variables satisfies the part $\tilde{F}^2$ of $\tilde{F}^3$ by (A.1) and it satisfies the F_3 part by the properties of I_{bag}. Thus, C.1. is satisfied.

□

3.4 Fourth Step: Removing Multi-set Constraints

This step rewrites the bag atoms and the mixed atoms in order to remove the bag terms and so transforms $\tilde{F}^3$ to a formula in QFLIA. The transformation preserves the boolean structure of the initial formulas and the integer atoms. It is given by the $S4$ function defined in Figure 1.

Proposition 4. *The formula $\tilde{F}^4 = S4(\tilde{F}^3)$ is in QFLIA and (i) for any model of I of $\tilde{F}^3$ there is a model I' of $\tilde{F}^4$ such that I and I' agree on $V^3(\int)$. (ii) for any model I of $\tilde{F}^4$, there is a model I' of $\tilde{F}^3$ such that I and I' agree on $V_{int}(\tilde{F}^3)$.*

Translation of mixed atoms:

$$
\begin{aligned}
S4(m = max(x)) &\triangleq \left((m = m_{\bot}) \wedge \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = 0) \right) \vee \\
&\left((m \neq m_{\bot}) \wedge (m \neq m_{\top}) \wedge (w_{m,x} \geq 1) \wedge \bigwedge_{t \in V_{ele}^{3}} ((t \leq m) \vee (w_{t,x} = 0)) \right) \\
S4(m = min(x)) &\triangleq \left((m = m_{\top}) \wedge \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = 0) \right) \vee \\
&\left((m \neq m_{\bot}) \wedge (m \neq m_{\top}) \wedge (w_{m,x} \geq 1) \wedge \bigwedge_{t \in V_{ele}^{3}} ((t \geq m) \vee (w_{t,x} = 0)) \right) \\
S4(a \in x) &\triangleq w_{a,x} \geq 1 \\
S4(a \in^{n} x) &\triangleq w_{a,x} \geq n \\
S4(a \notin x) &\triangleq w_{a,x} = 0 \\
S4(a \notin^{n} x) &\triangleq w_{a,x} < n \\
S4(m \in x) &\triangleq (m = m_{\bot} \vee m = m_{\top}) \wedge \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = 0) \vee w_{m,x} \geq 1 \\
S4(m \in^{n} x) &\triangleq m \neq m_{\bot} \wedge m \neq m_{\top} \wedge w_{m,x} \geq n \\
S4(m \notin x) &\triangleq (m \neq m_{\bot} \wedge m \neq m_{\top}) \vee \bigvee_{t \in V_{ele}^{3}} (w_{t,x} \neq 0) \wedge w_{m,x} = 0 \\
S4(m \notin^{n} x) &\triangleq m = m_{\bot} \vee m = m_{\top} \vee w_{m,x} < n
\end{aligned}
$$

Translation of bag atoms:

$$
\begin{aligned}
S4(x = y) &\triangleq \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = w_{t,y}) \\
S4(x \subseteq y) &\triangleq \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} \leq w_{t,y}) \\
S4(x \neq y) &\triangleq w_{a_{x \neq y},x} \neq w_{a_{x \neq y},y} \\
S4(x \nsubseteq y) &\triangleq w_{a_{x \nsubseteq y},x} > w_{a_{x \nsubseteq y},y} \\
S4(x = [\![]\!]) &\triangleq \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = 0) \\
S4(x = [\![a]\!]) &\triangleq (w_{a,x} = 1) \wedge \bigwedge_{t \in V_{ele}^{3}} ((a = t) \vee (w_{t,x} = 0)) \\
S4(x = y \cup z) &\triangleq \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = max(w_{t,y}, w_{t,z})) \\
S4(x = y \cap z) &\triangleq \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = min(w_{t,y}, w_{t,z})) \\
S4(x = y \uplus z) &\triangleq \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = w_{t,y} + w_{t,z}) \\
S4(x = y \setminus z) &\triangleq \bigwedge_{t \in V_{ele}^{3}} (w_{t,x} = \max(0, (w_{t,y} - w_{t,z})))
\end{aligned}
$$

Figure 1: Translation of mixed and bags atoms

Proof. Clearly $\tilde{F}^4$ is in QFLIA from the definition of $S4$.

Point (i): Because $S4$ does not rewrite integer atoms, a model I of $\tilde{F}^3$ satisfies, with its integer valuation I_{int}, the integer atoms not rewritten by $S4$. The integer atoms introduced by $S4$ are conform with the semantics of formulas while interpreting the variables $w_{*,x}$ as counting abstractions of a bag x. Then if I does not interpret the variables $w_{*,x}$ as counting abstractions of a bag x, we construct I' such that it does and agree on V^3_{int} with I. If $x \neq y$ is satisfied then there exists a value v in $\mathbb{Z}$ such as $I_{bag}(x)(v) \neq I_{bag}(y)(v)$. By the choice of I_{int} in the 3rd step (i.e., it satisfies F_3), we have that $I_{int}(a_{x\neq y}) = v$ If $m = max(x)$ is satisfied then either $I_{bag}(x) = [\![]\!]$ and $I_{ext}(m) = I_{ext}(m_\bot)$, or $I_{ext}(m)$ is the greatest integer of $\mathcal{D}(I_{bag}(x))$ but different from $I_{ext}(m_\bot)$ and $I_{ext}(m_\top)$.

Point (ii):

A.1 I is a model of $S4(\tilde{F}^3)$

C.1 there is a model I' of $\tilde{F}^3$ that agrees on $V_{int}(\tilde{F}^3)$ with I

Proof:

1. I_{int} satisfies the F_3 component of $\tilde{F}^3$ which is preserved as it is by $S4$. By the definition of $S4$, $V_{int}(\tilde{F}^3) = V_{int}(S4(\tilde{F}^3))$.

2. We define a valuation of bag variables in $\tilde{F}^3$, I'_{bag} as follows:

$$I'_{bag}(x)(v) \triangleq \begin{cases} I_{int}(w_{a,x}) & \text{if } v = I_{int}(a) \text{ for some } a \in V^3_{int} \\ I_{int}(w_{m,x}) & \text{if } v = I_{ext}(m) \text{ for some } m \in V^3_{ext} \\ 0 & \text{otherwise} \end{cases}$$

   Then I'_{bag} is a well funded function.

   Proof: By the fact that I_{int} satisfies the constraint F_3, which means that values of $w_{a,x}$ and $w_{b,x}$ are the same if $I_{int}(a) = I_{int}(b)$.

3. $I' = (I_{int}(\tilde{F}^3), I'_{bag})$ is a model of $S4(\tilde{F}^3)$.

   Proof: The proof proceeds by induction on the form of atoms transformed by $S4$. We show only the main points of the induction, the others are done similarly.

   3.1 for $L_{mix} ::= \ m = \max(x)$, if $I \models S4(L_{mix})$ then $I' \models L_{mix}$ (and vice versa).

   3.1.1 If $I_{bag}(x) \neq [\![]\!]$:

   3.1.1.1 From definition of $S4(L_{mix})$, $I_{int}(w_{m,x}) \geq 1$ and $I_{ext}(m) \neq I_{ext}(m_\bot)$ and $I_{ext}(m) \neq I_{ext}(m_\bot)$.

   3.1.1.2 From 3.1.1.1 and the definition of I'_{bag}, $I_{ext}(m) \in \mathcal{D}(I_{bag}(x))$.

3.1.1.3 There is no value v in $\mathcal{D}(I_{bag}(x))$ greater than $I_{ext}(m)$
By absurd, suppose that such a value exists, i.e., $v > I_{ext}(m)$ and $v \in \mathcal{D}(I_{bag}(x))$.
From the definition of I'_{bag}, there exists $t \in V^3_{ele}$ such as $I_{int}(t) = v > I_{ext}(m)$ or $I_{ext}(t) = v > I_{ext}(m)$ and $v = I_{int}(w_{t,x}) > 0$.
This is in contradiction with the $S4(L_{mix})$ conjunct saying that $t \leq a \vee w_{t,x} = 0$.

3.1.2 If $I_{bag}(x) = [\![]\!]$:

3.1.2.1 From definition of $S4(L_{mix})$, $I_{ext}(m) = I_{ext}(m_\perp)$.

The reverse direction is shown similarly.

Hence, 3.1 is valid.

3.2 for $L_{bag} ::= \; x = y \cup z$, if $I \models S4(L_{bag})$ then $I' \models L_{bag}$ (and vice versa).

3.2.1 From definition of $S4(L_{bag})$,

$$\bigwedge_{t \in V^3_{ele}} (I_{int}(w_{t,x}) = max(I_{int}(w_{t,y}), I_{int}(w_{t,z}))).$$

3.2.2 From 3.2.1 and the definition of I'_{bag},

$$\bigwedge_{t \in V^3_{ele}} (I'_{bag}(x)(I_{int}(t)) = max(I'_{bag}(y)(I_{int}(t)), I'_{bag}(z)(I_{int}(t)))).$$

or

$$\bigwedge_{t \in V^3_{ele}} (I'_{bag}(x)(I_{ext}(t)) = max(I'_{bag}(y)(I_{ext}(t)), I'_{bag}(z)(I_{ext}(t)))).$$

3.2.3 From 3.2.2 and the definition of I'_{bag},

$$\bigwedge_{k \in \mathbb{Z}} (I'_{bag}(x)(k) = max(I'_{bag}(y)(k), I'_{bag}(z)(k))).$$

3.2.4 From 3.2.3 the semantic 2.2 of T_{bag} and L_{bag},

$$I'_{bag}(x) = I'_{bag}(y) \cup I'_{bag}(z)$$

.

The reverse direction is shown similarly.

Hence, 3.2 is valid.

4. $I' = (I_{int}, I'_{bag}) \models S4(\tilde{F}^3)$
   Proof: by 1, 2, and 3 above.

5. C.1
   Proof: by 4 and the definition of I'.

□

4 Decision procedure by reduction to QFUFLIA

This section describes a reduction to QFBILIA formulas to an equi-satisfiable formula in quantifier free linear arithmetics with uninterpreted functions QFUFLIA. The first two steps of the procedure are identical to the previous one as a consequence we consider a formula $\tilde{F}^2$ as input. The third step introduces two extremum variables to represent $\bot$ and $\top$, a set of fresh integer variables and a set of uninterpreted function variables to represent (i) the elements that validates atoms $x \neq y$ and $x \nsubseteq y$ and (ii) for each multi-set variable $x \in V_{bag}(\tilde{F}^2)$ uninterpreted function variables G_x such as for each integer variable $a \in V_{int}(\tilde{F}^2)$, $G_x(a)$ is counting the occurrences of values of a in x. This step only adds to $\tilde{F}^2$ a formula F_3 over integer variables and uninterpreted function. The formula $\tilde{F}^2 \wedge F_3$ is then equi-satisfiable to F. The fourth step removes multi-set terms from $\tilde{F}^3$ using the fresh uninterpreted function variables introduced by the counting abstraction and produces a formula $\tilde{F}^4 \wedge F_3$ equi-satisfaisable to F.

4.1 QFBILIAUF Syntax

Let $V_{uf} = \{G, H, \ldots\}$ be a finite set of symbols denoting uninterpreted function variables, i.e., variables with values in $\mathbb{Z} \to \mathbb{Z}$. We suppose that V_{int}, V_{ext}, V_{uf} and V_{bag} are disjoint and we do not write explicitly the type ($\mathbb{Z}$, $\mathbb{Z}^{\prec}$, $(\mathbb{Z} \to \mathbb{Z})$ or $\mathbb{M}[\mathbb{Z}]$) of each variable, where V_{int}, V_{ext} and V_{bag} are defined as in the previous section.

Definition 7. *A QFBILIAUF formula F is defined by the following grammar:*

$$
\begin{array}{rll}
F & ::= L \mid F \vee F \mid F \wedge F \mid \neg F \mid F \Rightarrow F & \textit{formula} \\
L & ::= L_{int} \mid L_{bag} \mid L_{mix} \mid L_{ext} & \textit{boolean atom} \\
L_{int} & ::= T_{int} = T_{int} \mid T_{int} \neq T_{int} \mid T_{int} < T_{int} \mid T_{int} \geq T_{int} & \\
L_{ext} & ::= T_{ext} = T_{ext} \mid T_{ext} \neq T_{ext} \mid T_{ext} < T_{ext} \mid T_{ext} \geq T_{ext} & \\
L_{bag} & ::= T_{bag} = T_{bag} \mid T_{bag} \neq T_{bag} \mid T_{bag} \subseteq T_{bag} \mid T_{bag} \nsubseteq T_{bag} \mid & \\
 & T_{bag} < T_{bag} \mid T_{bag} \geq T_{bag} & \\
L_{mix} & ::= a \in T_{bag} \mid a \notin T_{bag} \mid a \in^n T_{bag} \mid a \notin^n T_{bag} & \\
T_{int} & ::= k \mid a \mid T_{uf} \mid T_{int} + T_{int} \mid T_{int} - T_{int} \mid & \textit{integer term} \\
 & \max(T_{int}, T_{int}) \mid \min(T_{int}, T_{int}) \mid \mathit{ite}(F, T_{int}, T_{int}) & \\
T_{ext} & ::= k \mid m \mid \min(T_{bag}) \mid \max(T_{bag}) \mid \mathit{ite}(F, T_{ext}, T_{ext}) & \textit{extremum term} \\
T_{uf} & ::= G(T_{int}) & \textit{UF term} \\
T_{bag} & ::= [\![]\!] \mid [\![a]\!] \mid x \mid T_{bag} \cup T_{bag} \mid T_{bag} \cap T_{bag} \mid & \textit{bag term} \\
 & T_{bag} \setminus T_{bag} \mid T_{bag} \uplus T_{bag} \mid \mathit{ite}(F, T_{bag}, T_{bag}) &
\end{array}
$$

We denote by by $\mathcal{T}_{uf}$ the set of uninterpreted function terms.

For a formula F, we denote by $V_{uf}(F)$ the set of uninterpreted function variables used in F.

4.2 QFBILIAUF Semantics

A valuation I_{uf} of variables in V_{uf} is a function mapping variables in V_{uf} to values in $(\mathbb{Z} \to \mathbb{Z})$, i.e., $I_{uf} : V_{uf} \to (\mathbb{Z} \to \mathbb{Z})$. A valuation I of variables in $V_{int} \cup V_{uf} \cup V_{bag} \cup V_{ext}$ is a tuple of valuations $(I_{int}, I_{uf}, I_{bag}, I_{ext})$; we denote by I_{int}, I_{uf}, I_{bag} and I_{ext} the first, resp. second, resp. third, resp. fourth component of a valuation I. Let $\mathcal{I}$ be the set of valuations over variables in $V_{int} \cup V_{uf} \cup V_{ext} \cup V_{bag}$. The semantics of QFBILIAUF is defined by

- a function $^* : \mathcal{I} \to \mathcal{T}_{int} \to \mathbb{Z}$ mapping a valuation and an integer term to an integer value,
- a function $^\square : \mathcal{I} \to \mathcal{T}_{uf} \to \mathbb{Z}$ mapping a valuation and an uninterpreted function term to an integer value,
- a function $^\circ : \mathcal{I} \to \mathcal{T}_{bag} \to \mathbb{M}[\mathbb{Z}]$ mapping a valuation and a multi-set term to a multi-set value,
- a function $^\bullet : \mathcal{I} \to \mathcal{T}_{ext} \to \mathbb{Z} \cup \{\bot, \top\}$ mapping a valuation and an integer term to an extremum value,
- a relation $\models \subseteq \mathcal{I} \times \mathcal{F}$ between valuations and formulas.

I^*, $I^\bullet$ and I° have the same definition as in QFBILIA, see section 2.2.
Given a valuation I and an uninterpreted function term T_{uf}, the valuation of T_{uf} in $\mathbb{Z}$, denoted by $I^\square(T_{uf})$, is defined as follows:

$$I^\square(G(T_{int})) \triangleq I_{uf}(G)(I^*(T_{int}))$$

Definition 8. *A QFBILIAUF formula F is* satisfiable *if there exists a valuation I, called also its model, such that $I \models F$.*

4.3 Third Step: Introducing the Counting Abstraction

This step adds to $\tilde{F}^2$ a formula F_3 that allow to introduce the uninterpreted functions variable that will replace the multi-set ones. The transformation has three steps:

S_u3.1: Build the set V^{31}_{int} as a set of fresh variables, one variable for each atom $(x \neq y)$ or $(x \nsubseteq y)$ in $\tilde{F}^2$. We denote these variables by $a_{x \neq y}$, resp. $a_{x \nsubseteq y}$. Intuitively, these variables are introduced to be able to express the fact that there is a value on which x and y differ, resp. x has more copies than y. see subsection 4.4.

S_u3.2: We introduce two fresh extremum variables $m_\perp$ and $m_\top$. Intuitively, these variables are introduced to be able to express $\perp$, resp. $\top$.

S_u3.3: Build the set

$$V_{uf} = \bigcup_{x \in V_{bag}(\tilde{F}^2)} G_x,$$

where an uninterpreted variable G_x is added for each bag variable x in order to represent the bag x. Let $V^3_{int} = V_{int}(\tilde{F}^2) \cup V^{31}_{int}$. Let $V^3_{ele} = V^3_{int} \cup V_{ext}(\tilde{F}^2)$.

The formula $\tilde{F}^3$ is built as follows:

$$\tilde{F}^3 \triangleq \tilde{F}^2 \wedge F_3 \tag{28}$$

$$F_3 \triangleq \left(\bigwedge_{a \in V^3_{int}} (m_\perp < a < m_\top) \right) \tag{29}$$

$$\wedge \left(\bigwedge_{m \in V_{ext}(\tilde{F}^2)} (m_\top \geq m \geq m_\perp) \right) \tag{30}$$

$$\wedge \left(\bigwedge_{t \in V^3_{ele}, x \in V_{bag}} (G_x(t) \geq 0) \right) \tag{31}$$

$$\tag{32}$$

The set of variables of $\tilde{F}^3$ are $V_{int}(\tilde{F}^3)$, resp. V_{uf}, resp. $V_{bag}(\tilde{F}^2)$.

The following property states that $\tilde{F}^3 \sim_{sat} \tilde{F}^2$:

Proposition 5. *(i) any model of $\tilde{F}^3$ is a model of $\tilde{F}^2$ and (ii) for any model of $\tilde{F}^2$, there is a model of $\tilde{F}^3$ that agrees on $V_{int}(\tilde{F}^2)$ and $V_{bag}(\tilde{F}^2)$.*

Proof. Because $\tilde{F}^3 \triangleq \tilde{F}^2 \wedge F_3$, any model I of $\tilde{F}^3$ is also a model of $\tilde{F}^2$.

We prove now the point (ii):

A.1: I be a model of $\tilde{F}^2$

C.1: There is I' which agrees on $V(\tilde{F}^2)$ to I and I' is a model of $\tilde{F}^3$.

Proof:

1: Let $I' = (I_{bag}, I_{int}, I'_{uf})$ such that for any $G_x \in V_{uf}(\tilde{F}^3)$ and for any $k \in \mathbb{Z}$ we have

$$I'_{uf}(G_x)(k) \triangleq \begin{cases} I_{bag}(x)(a) & \text{if } I_{int}(a) = k \text{ for some } a \in V_{int}(\tilde{F}^2) \\ I_{bag}(x)(m) & \text{if } I_{ext}(m) = k \text{ for some } m \in V_{ext}(\tilde{F}^2) \\ I_{bag}(x)(v) & \text{if } I_{int}(a_{x \neq y}) = k \text{ for some } a_{x \neq y} \in V^{31}_{int} \\ & \quad \text{and } I_{bag}(x)(v) \neq I_{bag}(y)(v) \\ I_{bag}(x)(v) & \text{if } I_{int}(a_{x \not\subseteq y}) = k \text{ for some } a_{x \not\subseteq y} \in V^{31}_{int} \\ & \quad \text{and } I_{bag}(x)(v) > I_{bag}(y)(v) \\ 0 & \text{otherwise} \end{cases}$$

for any $m \in V_{ext}(\tilde{F}^3)$ we have

$$I'_{ext}(m) \triangleq \begin{cases} I_{ext}(m) & \text{if } m \in V_{ext}(\tilde{F}^2) \\ k & \text{if } m_\perp \text{ and } \forall a \in V_{int}(\tilde{F}^3), k < I'_{int}(a) \\ & \qquad \text{and } \forall m \in V_{ext}(\tilde{F}^3), k \leq I'_{ext}(m) \\ k & \text{if } m_\top \text{ and } \forall a \in V_{int}(\tilde{F}^3), k > I'_{int}(a) \\ & \qquad \text{and } \forall m \in V_{ext}(\tilde{F}^3), k \geq I'_{ext}(m) \end{cases}$$

This valuation of integer and extremum variables satisfies the part $\tilde{F}^2$ of $\tilde{F}^3$ by (A.1) and it satisfies the F_3 part by the properties of I_{bag}. Thus, C.1. is satisfied.

□

4.4 Fourth Step: Removing Multi-set Constraints

This step rewrites the bag atoms and the mixed atoms in order to remove the bag terms as so transform $\tilde{F}^3$ to a formula in QFUFLIA. The transformation preserves the boolean structure of the initial formulas and the integer atoms. It is given by the $S4_{uf}$ function defined in Figure 2.

Proposition 6. *The formula $\tilde{F}^4 = S_u4(\tilde{F}^3)$ is in QFUFLIA and (i) any model of $\tilde{F}^3$ is a model of $\tilde{F}^4$ and (ii) for any model of $\tilde{F}^4$, there is a model of $\tilde{F}^3$ that agrees on $V_{int}(\tilde{F}^3)$.*

Proof. Clearly $\tilde{F}^4$ is in QFLIA from the definition of S_u4.

Point (i): Because S_u4 does not rewrite integer atoms and the F_3 conjunct, a model I of $\tilde{F}^3$ satisfies, with its integer valuation I_{int}, the integer atoms not rewritten by S_u4. The uninterpreted function atoms introduced by S_u4 are conform with the semantics of formulas while interpreting the variables $X(*)$ as counting abstractions of a bag x. If $x \neq y$ is satisfied then there exists a value v in $\mathbb{Z}$ such as $I_{bag}(x)(v) \neq I_{bag}(y)(v)$. By the choice of I_{int} in the 3rd step (i.e., it satisfies F_3), we have that $I_{int}(a_{x \neq y}) = v$ If $m = max(x)$ is satisfied then either $I_{bag}(x) = [\![]\!]$ and $I_{ext}(m) = I_{ext}(m_\perp)$, or $I_{ext}(m)$ is the greatest integer of $\mathcal{D}(I_{bag}(x))$ but different from $I_{ext}(m_\perp)$ and $I_{ext}(m_\top)$.

Point (ii):

A.1 I is a model of $S_u4(\tilde{F}^3)$

C.1 there is a model I' of $\tilde{F}^3$ that agrees on $V_{int}(\tilde{F}^3)$

Proof:

1. I_{int} satisfies the F_3 component of $\tilde{F}^3$ which is preserved as it is by S_u4. By the definition of S_u4, $V_{int}(\tilde{F}^3) = V_{int}(S_u4(\tilde{F}^3))$.

Translation of mixed atoms:

$$S4_{uf}(m = max(x)) \triangleq \left((m = m_{\perp}) \wedge \bigwedge_{t \in V^3_{ele}} (G_x(t) = 0)\right) \vee \left((m \neq m_{\perp}) \wedge (m \neq m_{\top}) \wedge (G_x(m) \geq 1) \wedge \bigwedge_{t \in V^3_{ele}} ((t \leq m) \vee (G_x(t) = 0))\right)$$

$$S4_{uf}(m = min(x)) \triangleq \left((m = m_{\top}) \wedge \bigwedge_{t \in V^3_{ele}} (G_x(t) = 0)\right) \vee \left((m \neq m_{\perp}) \wedge (m \neq m_{\top}) \wedge (G_x(m) \geq 1) \wedge \bigwedge_{t \in V^3_{ele}} ((t \geq m) \vee (G_x(t) = 0))\right)$$

$$S4_{uf}(a \in x) \triangleq G_x(a) \geq 1$$

$$S4_{uf}(a \in^n x) \triangleq G_x(a) \geq n$$

$$S4_{uf}(a \notin x) \triangleq G_x(a) = 0$$

$$S4_{uf}(a \notin^n x) \triangleq G_x(a) < n$$

$$S4_{uf}(m \in x) \triangleq (m = m_{\perp} \vee m = m_{\top}) \wedge \bigwedge_{t \in V^3_{ele}} (G_x(t) = 0) \vee G_x(m) \geq 1$$

$$S4_{uf}(m \in^n x) \triangleq m \neq m_{\perp} \wedge m \neq m_{\top} \wedge G_x(m) \geq n$$

$$S4_{uf}(m \notin x) \triangleq (m \neq m_{\perp} \wedge m \neq m_{\top}) \vee \bigvee_{t \in V^3_{ele}} (G_x(t) \neq 0) \wedge G_x(m) = 0$$

$$S4_{uf}(m \notin^n x) \triangleq m = m_{\perp} \vee m = m_{\top} \vee G_x(m) < n$$

Translation of bag atoms:

$$S4_{uf}(x = y) \triangleq \bigwedge_{t \in V^3_{ele}} (G_x(t) = G_y(t))$$

$$S4_{uf}(x \subseteq y) \triangleq \bigwedge_{t \in V^3_{ele}} (G_x(t) \leq G_y(t))$$

$$S4_{uf}(x \neq y) \triangleq G_x(a_{x \neq y}) \neq G_y(a_{x \neq y})$$

$$S4_{uf}(x \nsubseteq y) \triangleq G_x(a_{x \nsubseteq y}) > G_y(a_{x \nsubseteq y})$$

$$S4_{uf}(x = [\![]\!]) \triangleq \bigwedge_{t \in V^3_{ele}} (G_x(t) = 0)$$

$$S4_{uf}(x = [\![a]\!]) \triangleq (G_x(a) = 1) \wedge \bigwedge_{t \in V^3_{ele}} ((a = t) \vee (G_x(t) = 0))$$

$$S4_{uf}(x = y \cup z) \triangleq \bigwedge_{t \in V^3_{ele}} (G_x(t) = max(G_y(t), G_z(t)))$$

$$S4_{uf}(x = y \cap z) \triangleq \bigwedge_{t \in V^3_{ele}} (G_x(t) = min(G_y(t), G_z(t)))$$

$$S4_{uf}(x = y \uplus z) \triangleq \bigwedge_{t \in V^3_{ele}} (G_x(t) = G_y(t) + G_z(t))$$

$$S4_{uf}(x = y \setminus z) \triangleq \bigwedge_{t \in V^3_{ele}} (G_x(t) = \max(0, (G_y(t) - G_z(t))))$$

Figure 2: Translation of mixed and bags atoms

2. We define a valuation of bag variables in $\tilde{F}^3$, I'_{bag} as follows:

$$I'_{bag}(x)(v) \triangleq \begin{cases} I_{uf}(G_x)(a) & \text{if } v = I_{int}(a) \text{ for some } a \in V^3_{int} \\ I_{uf}(G_x)(m) & \text{if } v = I_{ext}(m) \text{ for some } m \in V^3_{ext} \\ 0 & \text{otherwise} \end{cases}$$

   Then I'_{bag} is a well funded function.

   Proof: By the fact that I_{uf} satisfies the constraint F_3, which means that values of $I_{uf}(G_x)(*) \geq 0$.

3. $I' = (I_{int}(\tilde{F}^3), I'_{bag})$ is a model of $S_u4(\tilde{F}^3)$.

   Proof: The proof proceeds by induction on the form of atoms transformed by S_u4. We show only the main points of the induction, the others are done similarly.

   3.1 for $L_{bag} ::= \ x = [\![a]\!]$, if $I \models S_u4(L_{bag})$ then $I' \models L_{bag}$ (and vice versa).

   3.1.1 From definition of $S_u4(L_{bag})$,

$$(I_{uf}(G_x)(I_{int}(a)) = 1) \wedge \bigwedge_{b \in V^3_{int}} ((I_{int}(a) = I_{int}(b)) \vee (I_{uf}(G_x)(I_{int}(b)) = 0)) \wedge \bigwedge_{m \in V^3_{ext}} ((I_{int}(a) = I_{ext}(m)) \vee (I_{uf}(G_x)(I_{ext}(m)) = 0)).$$

   3.2.2 From 3.1.1 and the definition of I'_{bag},

$$(I_{bag}(x)(I_{int}(a)) = 1) \wedge \bigwedge_{b \in V^3_{int}} ((I_{int}(a) = I_{int}(b)) \vee (I_{bag}(x)(I_{int}(b)) = 0)) \wedge \bigwedge_{m \in V^3_{ext}} ((I_{int}(a) = I_{ext}(m)) \vee (I_{bag}(x)(I_{ext}(m)) = 0)).$$

   3.3.3 From 3.1.2 and the definition of I'_{bag},

$$(I_{bag}(x)(I_{int}(a)) = 1) \wedge \bigwedge_{k \in \mathbb{Z}} ((I_{int}(a) = k) \vee (I_{bag}(x)(k) = 0)).$$

   3.4.4 From 3.1.3 the semantic 2.2 of T_{bag} and L_{bag},

$$I'_{bag}(x) = [\![I_{int}(a)]\!]$$

   .

   The reverse direction is shown similarly.

   Hence, 3.1 is valid.

   3.2 for $L_{bag} ::= \ x \neq y$, if $I \models S_u4(L_{bag})$ then $I' \models L_{bag}$ (and vice versa).

3.2.1 From definition of $S_u4(L_{bag})$,

$$I_{uf}(G_x)(I_{int}(a_{x\neq y})) \neq I_{uf}(G_y)(I_{int}(a_{x\neq y}))$$

.

3.2.2 From 3.2.1 and the definition of I'_{bag},

$$I_{bag}(x)(I_{int}(a_{x\neq y})) \neq I_{bag}(y)(I_{int}(a_{x\neq y}))$$

.

3.2.3 From 3.2.2 and the definition of I'_{bag},

$$\exists k \in \mathbb{Z} I_{bag}(x)(k) \neq I_{bag}(y)(k)$$

.

3.2.4 From 3.2.3 and the semantic 2.2 of L_{bag},

$$I'_{bag}(x) \neq I'_{bag}(y)$$

.

The reverse direction is shown similarly.
Hence, 3.2 is valid.

4. $I' = (I_{int}, I'_{bag}) \models S_u4(\tilde{F}^3)$
   Proof: by 1, 2, and 3 above.

5. C.1
   Proof: by 4 and the definition of I'.

□

5 Implementation

To obtain efficient solving times, we implemented several optimisation of the procedures described in the previous section. We describe these optimisations in the next subsections. It is important to understand that our decision procedure being based upon an SMT solver, we want our output formula to be as friendly as possible. However, we try to keep our computation time as small as possible too. The challenge is to understand when we should increase our computation time in order to decrease the computation time of SMT solvers. As shown by our benchmark, the SMT solvers are very efficient when dealing with standards operations (for instance, convert a formula in CNF), hence we do not modify our formula form. However SMT solvers computation time increases with the number of atoms in the input formula. As a consequence it is interesting to reduce this number by our reduction procedure, if it does not increase too much our computation time (see Section 5.4).

5.1 Unitary Bag Optimisation

Many atoms are added in steps 2-4 due to L_{bag} and L_{mix}; more precisely, their number is linear in the number of variable in V_{int} and V_{bag}. To reduce this number, we rewrite the formulas in the left side below to the right side formulas which add less variables and conjunctions in steps 2-4:

$$\begin{array}{lcl} [\![a]\!] \subseteq x & \sim_{sat} & a \in x \\ [\![a]\!] \not\subseteq x & \sim_{sat} & a \notin x \\ max([\![a]\!]) & = & a \\ min([\![a]\!]) & = & a \end{array}$$

In $S4$, $a \in x$ does not add any conjunction to the formula, hence the output have fewer atoms than the equi-satisfiable one with $[\![a]\!] \subseteq x$. As $a \notin x$ does not add any variable to V_{int} fewer atoms will be conjuncted during $S4$ than with $[\![a]\!] \not\subseteq x$ Same can be said about $max([\![a]\!])$ and $min([\![a]\!])$ where step 2 adds a fresh variable to V_{int} and step 4 conjuncts many atoms. Notice that the rewriting can be done without any increase of computation time, as it is just a different rewriting of an already existing literal.

5.2 Bag variables creation

In step 2.4, every bag atom L_{bag} of $\tilde{F}_{2.3}$ using more than two bag variables for operations $\neq, \subseteq, \not\subseteq$ and more than three variables for operation $=$ are iteratively rewritten to be reduced to the bag atoms in $\mathsf{QFBILIA}_{pure}$. However, this rewriting adds many bag variable which cost a lot of computation time, as we will have to deal with more bags. Let define a function $apply : (T_{bag}, V_{int}) \rightarrow T_{int}$ by:

$$apply(x, a) = w_{a,x} \tag{33}$$

$$apply(T^1_{bag} \cup T^2_{bag}, a) = max(apply(T^1_{bag}, a), apply(T^2_{bag}, a)) \tag{34}$$

$$apply(T^1_{bag} \cap T^2_{bag}, a) = min(apply(T^1_{bag}, a), apply(T^2_{bag}, a)) \tag{35}$$

$$apply(T^1_{bag} \uplus T^2_{bag}, a) = apply(T^1_{bag}, a) + apply(T^2_{bag}, a) \tag{36}$$

$$apply(T^1_{bag} \setminus T^2_{bag}, a) = max(0, apply(T^1_{bag}, a) - apply(T^2_{bag}, a)) \tag{37}$$

$$\tag{38}$$

Then we can extend S4:

$$S4(T^1_{bag} = T^2_{bag}) \triangleq \bigwedge_{t \in V^3_{ele}} (apply(T^1_{bag}, t) = apply(T^2_{bag}, t))$$

while preserving equi-satisfiability. Same can be done for $\neq, \subseteq, \not\subseteq$ operators. This rewriting allows us to skip the step 2.4.

5.3 Integer Element Optimisation

A QFBILIA formula might have pure integer variables, i.e., integer variables that are not involve in any bag, mix or extremum literals. Such variable does not have to be consider during the counting abstraction as they do not constrain any bag, hence we can reduce the cardinal of V^3_{ele}

Definition 9. *Let a formula F in QFBILIA let two variable u, v in $V_{int}(F) \cup V_{bag}(F)$ then u operate with v, denote $u \mathrel{\underline{\frown}} v$, iff exists l in $L(F)$ where u and v are in $V_{int}(l) \cup V_{bag}(l)$.*

$$u \mathrel{\underline{\frown}} v \iff \exists l \in L(F), u, v \in V_{int}(l) \cup V_{bag}(l)$$

Definition 10. *Let a formula F in QFBILIA , $V_{ope}(F)$ is the set of integer variable:*

$$V_{ope}(F) = \{a \in V_{int}(F) | \exists x \in V_{bag}(F), a \mathrel{\underline{\frown}} y\} \cup V_{ext}$$

and

$$< V_{ope}(F) >= \{a \in V_{ope}(F) | \exists b \in V_{ope}(F), a \mathrel{\underline{\frown}} b\}$$

Proposition 7. *If V^3_{ele} is redefined (step 3-4) as :$V^3_{ele} =< V_{ope}(F) >$ then (i) Property 3 is still valid, and (ii) Property 4 is still valid.*

Proof. The proof is the same as before because the only difference is the absence of useless w. □

5.4 Benchmark

Our solvers takes as input problems written in the SMTLIB format [2] for the theory of bags and sets.

We applied it on a benchmark of 295 problems generated from verification conditions generated by [4] for programs manipulating data structures like lists, binary search trees, AVL, skip lists,. . . . All problems take less than 1 sec to be reduced to the base theory (QFLIA or QFUFLIA).

6 Conclusion

We defined the logic QFBILIA to express constraints over bags of integers. We then presented two decision procedures for QFBILIA. The main flow of our work is that we were only able to be determinist because we were working with bags of integers. Moreover the efficiency of our decision procedures rely on the SMT solvers upon which we are working.

Acknowledgments

I would like to express my special thanks of gratitude to my internship tutors Arnaud Sangnier and Mihaela Sighireanu who gave me opportunity to work on multi-set constraints, helped me in doing research, and corrected my mistakes. I am really thankful to them.

Appendices

SMTLIB2 Syntax

Constraints	SMT-LIB2
$\neg p$	$(not\ p)$
$p \wedge q$	$(and\ p\ q)$
$p \vee q$	$(or\ p\ q)$
$p \Rightarrow q$	$(=>\ p\ q)$
$a + b$	$(+\ a\ b)$
$a - b$	$(-\ a\ b)$
$ite\ p\ a\ b$	$(ite\ p\ a\ b)$
$max(a, b)$	$(max\ a\ b)$
$min(a, b)$	$(min\ a\ b)$
$a \leq b$	$(<=\ a\ b)$
$a < b$	$(<\ a\ b)$
$a \geq b$	$(>=\ a\ b)$
$a > b$	$(>\ a\ b)$
$a = b$	$(=\ a\ b)$
$a \neq b$	$(neq\ a\ b)$
$\llbracket a \rrbracket$	$(bag\ a)$
$\llbracket a \rrbracket^n$	$(bagn\ a\ n)$
$\llbracket\rrbracket$	$emptybag$
$x \cup y$	$(bagunion\ x\ y)$
$x \cap y$	$(baginter\ x\ y)$
$x \uplus y$	$(bagsum\ x\ y)$
$x \setminus y$	$(bagminus\ x\ y)$
$x = y$	$(=\ x\ y)$
$x \neq y$	$(neq\ x\ y)$
$x \subset y$	$(subset\ x\ y)$
$x \subseteq y$	$(subseteq\ x\ y)$
$x \nsubseteq y$	$(nsubseteq\ x\ y)$
$x \in y$	$(in\ x\ y)$
$x \in^n y$	$(inn\ x\ y\ n)$
$x \notin y$	$(ni\ x\ y)$
$ite\ p\ x\ y$	$(ite\ p\ x\ y)$
$max(x)$	$(bagmax\ x)$
$min(x)$	$(bagmin\ x)$
$x \leq y$	$(<=\ x\ y)$
$x \geq y$	$(>=\ x\ y)$
$x > y$	$(>\ x\ y)$
$x < y$	$(<\ x\ y)$

References

[1] Clark Barrett, Christopher L Conway, Morgan Deters, Liana Hadarean, Dejan Jovanović, Tim King, Andrew Reynolds, and Cesare Tinelli. CVC4. In *Computer Aided Verification*, LNCS, pages 171–177. Springer, 2011.

[2] Clark Barrett, Aaron Stump, and Cesare Tinelli. The Satisfiability Modulo Theories Library (SMT-LIB). `www.SMT-LIB.org`, 2010.

[3] Leonardo De Moura and Nikolaj Bjørner. Z3: An efficient SMT solver. In *Tools and Algorithms for the Construction and Analysis of Systems*, LNCS, pages 337–340. Springer, 2008.

[4] Constantin Enea, Ondřej Lengál, Mihaela Sighireanu, and Tomáš Vojnar. Compositional entailment checking for a fragment of Separation Logic. In *Asian Programming Languages and Systems*, LNCS, pages 314–333. Springer, 2014.

[5] Ruzica Piskac, Philippe Suter, and Viktor Kuncak. On decision procedures for ordered collections. Technical report, 2010.

[6] Calogero G Zarba. Combining multisets with integers. In *Automated Deduction—CADE-18*, LNCS, pages 363–376. Springer, 2002.