

# Enhancing Linear Algebraic Computation of Logic Programs Using Sparse Representation

Tuan Nguyen Quoc

Katsumi Inoue

Chiaki Sakama

National Institute of Informatics, Tokyo, Japan

Wakayama University, Wakayama, Japan

The Graduate University for Advanced Studies (SOKENDAI)

{tuannq, inoue}@nii.ac.jp

sakama@wakayama-u.ac.jp

Algebraic characterization of logic programs has received increasing attention in recent years. Researchers attempt to exploit connections between linear algebraic computation and symbolic computation in order to perform logical inference in large scale knowledge bases. This paper proposes further improvement by using sparse matrices to embed logic programs in vector spaces. We show its great power of computation in reaching the fixpoint of the immediate consequence operator from the initial vector. In particular, performance for computing the least models of definite programs is dramatically improved in this way. We also apply the method to the computation of stable models of normal programs, in which the guesses are associated with initial matrices, and verify its effect when there are small numbers of negation. These results show good enhancement in terms of performance for computing consequences of programs and depict the potential power of tensorized logic programs.

## 1 Introduction

For decades, Logic Programming (LP) representation has been considered mainly in the form of symbolic logic [9], which is useful for declarative problem solving and symbolic reasoning. Logic programming starts gaining more attention recently in order to build explainable learning models, whereas it still has some limitations in terms of computation. In other words, symbolic computation is not an efficient way when we need to combine it with other numerical learning models such as Artificial Neural Network (ANN). Recently, several studies have been done on embedding logic programs to numerical spaces so that we can exploit great computing resources ranging from multi-threaded CPU to GPU. The linear algebraic approach is a robust way to manipulate logic programs in numerical spaces. Because linear algebra is at the heart of many applications of scientific computation, this approach is promising to develop scalable techniques to process huge relational Knowledge Base (KB) [18], [12]. In addition, it enables the ability to use efficient parallel algorithms of numerical linear algebra for computing LP.

In [5], Cohen described a probabilistic deductive database system in which reasoning is performed by a differentiable process. With this achievement, they can enable novel gradient-based learning algorithms. In [14], Sato presented the use of first-order logic in vector spaces for Tarskian semantics. He demonstrates how tensorization realizes the efficient computation of Datalog. In [15], Sato proposed linear algebraic approach to datalog evaluation. In this work, the least Herbrand model of DB, is computed via adjacency matrices. He also provided theoretical proofs for translating a program into a system of linear matrix equations. This approach achieves  $O(N^3)$  time complexity where  $N$  is the number of variables in a clause. Continuing to this direction, Sato et al. developed linear algebraic abduction to abductive inference in Datalog [16]. They did empirical experiments on linear and recursive cases and indicated that this approach can derive base relations.

Using a linear algebraic method, Sakama et al. [13] define relations between LP and multi-dimensional array (tensor) then propose algorithms for computation of LP models. The representation is done by defining a series of conversions from logical rules to vectors and then the computation is done by applying matrix multiplication. Later, elimination techniques are applied to reduce the matrix size [11] and gain impressive performance. In [3], a similar idea using 3D-tensor was employed to compute solutions of abductive Horn propositional tasks. In addition, Aspis built upon previous works on matrix characterization of Horn propositional logic programs to explore how inference from logic programs can be done by linear algebraic algorithms [2]. He also proposed a new algorithm for non-monotonic deduction, based on linear algebraic reducts and differentiable deduction. These works prove that the linear algebraic method is promising for logic inference on large scale but has not yet been done much in experiments, to the best of our knowledge.

In this paper, we continue Sakama et al.'s idea of representing logic programs by tensors [13]. Although the method is well-defined, there are some problems, which limit the performance of the approach and have not been solved. First, the obtained matrix after conversion is sparse but sparsity analysis was not considered. Second, the experiments were limited with small-size logic programs that are not sufficient to prove the robustness of matrix representation. In this research, we further raise the bar of computing performance by using sparse representation for logic programs in order to reach the fixpoint of the immediate consequence operator from the initial vector. We are able to do experiments on larger size logic programs to demonstrate the performance for computing least models of definite programs. Further, we also conduct experiments on computation of stable models of normal programs with a small number of negations.

Accordingly, the rest of this paper is organized as follows: Section 2 reviews and summaries some definitions and computation algorithms for definite and normal programs, Section 3 discusses sparsity problem in tensorized logic programs and proposes a method to represent LP, Section 4 demonstrates experimental results with definite and normal programs, Section 5 gives final conclusions and future works.

## 2 Preliminaries

### 2.1 Definite programs

We consider a language  $\mathcal{L}$  that contains a finite set of propositional variables. Given a logic program  $P$ , the set of all propositional variables appearing in  $P$  is called the *Herbrand base* of  $P$  (written  $B_P$ ). A *definite program* is a finite set of *rules* of the form:

$$h \leftarrow b_1 \wedge \dots \wedge b_m \quad (m \geq 0) \quad (1)$$

where  $h$  and  $b_i$  are propositional variables (atoms) in  $\mathcal{L}$ .

A rule  $r$  is called an *OR-rule* if  $r$  is the form:

$$h \leftarrow b_1 \vee \dots \vee b_m \quad (m \geq 0) \quad (2)$$

where  $h$  and  $b_i$  are propositional variables in  $\mathcal{L}$ .

A *standardized program* is a finite set of rules that are either (1) or (2). Note that the rule (2) is a shorthand of  $m$  rules:  $h \leftarrow b_1, \dots, h \leftarrow b_m$ , so a standardized program is considered a definite program.

For each rule  $r$  of the form (1) or (2), define  $head(r) = h$  and  $body(r) = \{b_1, \dots, b_m\}$ .<sup>1</sup> A rule  $r$  is called a *fact* if  $body(r) = \emptyset$ .

---

<sup>1</sup>We assume  $b_i \neq b_j \forall i \neq j$ .

**Definition 1 *Singly-Defined (SD) program*:** A definite program  $P$  is called a SD program if  $\text{head}(r_1) \neq \text{head}(r_2)$  for any two rules  $r_1$  and  $r_2$  ( $r_1 \neq r_2$ ) in  $P$ .

Any definite program can be transformed to a standardized program by introducing new propositional variables. That is, if there are several rules with the same head but different bodies:  $h \leftarrow \text{body}(r_1), \dots, h \leftarrow \text{body}(r_k)$ , then we replace all these rules by  $h \leftarrow b_1 \vee \dots \vee b_k, b_1 \leftarrow \text{body}(r_1), \dots, b_k \leftarrow \text{body}(r_k)$ . In this paper, a program means a standardized program unless stated otherwise.

A set  $I \subseteq B_P$  is an *interpretation* of  $P$ . An interpretation  $I$  is a *model* of a standardized program  $P$  if  $\{b_1, \dots, b_m\} \subseteq I$  implies  $h \in I$  for every rule (1) in  $P$ , and  $\{b_1, \dots, b_m\} \cap I \neq \emptyset$  implies  $h \in I$  for every rule (2) in  $P$ . A model  $I$  is the *least model* of  $P$  if  $I \subseteq J$  for any model  $J$  of  $P$ . A mapping  $T_P : 2^{B_P} \rightarrow 2^{B_P}$  (called a  $T_P$ -operator) is defined as:  $T_P(I) = \{h \mid h \leftarrow b_1 \wedge \dots \wedge b_m \in P \text{ and } \{b_1, \dots, b_m\} \subseteq I\} \cup \{h \mid h \leftarrow b_1 \vee \dots \vee b_n \in P \text{ and } \{b_1, \dots, b_n\} \cap I \neq \emptyset\}$ .

The powers of  $T_P$  are defined as:  $T_P^{k+1}(I) = T_P(T_P^k(I))$  ( $k \geq 0$ ) and  $T_P^0(I) = I$ . Given  $I \subseteq B_P$ , there is a fixpoint  $T_P^{n+1}(I) = T_P^n(I)$  ( $n \geq 0$ ). For a definite program  $P$ , the fixpoint  $T_P^n(\emptyset)$  coincides with the least model of  $P$  [17].

**Definition 2 *Matrix representation of standardized programs*** [13]: Let  $P$  be a standardized program and  $B_P = \{p_1, \dots, p_n\}$ . Then  $P$  is represented by a matrix  $M_P \in \mathbb{R}^{n \times n}$  such that for each element  $a_{ij}$  ( $1 \leq i, j \leq n$ ) in  $M_P$ ,

1.  $a_{ijk} = \frac{1}{m}$  ( $1 \leq k \leq m; 1 \leq i, j, k \leq n$ ) if  $p_i \leftarrow p_{j_1} \wedge \dots \wedge p_{j_m}$  is in  $P$ ;
2.  $a_{ijk} = 1$  ( $1 \leq k \leq l; 1 \leq i, j, k \leq n$ ) if  $p_i \leftarrow p_{j_1} \vee \dots \vee p_{j_l}$  is in  $P$ ;
3.  $a_{ii} = 1$  if  $p_i \leftarrow$  is in  $P$ ;
4.  $a_{ij} = 0$ , otherwise.

$M_P$  is called a *program matrix*. We write  $\text{row}_i(M_P) = p_i$  and  $\text{col}_j(M_P) = p_j$  ( $1 \leq i, j \leq n$ ).

To better understand Definition 2, let's consider a concrete example.

**Example 1** Consider the definite program  $P = \{p \leftarrow q \wedge r, p \leftarrow s \wedge t, r \leftarrow s, q \leftarrow t, s \leftarrow, t \leftarrow\}$ .

$P$  is not an SD program because there are two rules  $p \leftarrow q \wedge r$  and  $p \leftarrow s \wedge t$  having the same head, then  $P$  is transformed to the standardized program  $P'$  by introducing new atoms  $u$  and  $v$  as follows:  $P' = \{u \leftarrow q \wedge r, v \leftarrow s \wedge t, p \leftarrow u \vee v, r \leftarrow s, q \leftarrow t, s \leftarrow, t \leftarrow\}$ . Then by applying Definition 2, we obtain:

$$\begin{matrix} & p & q & r & s & t & u & v \\ \begin{matrix} p \\ q \\ r \\ s \\ t \\ u \\ v \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \end{pmatrix} \end{matrix}$$

Sakama et al. further define representation of interpretation by using *interpretation vector* (Definition 3). This vector is used to store the truth value of all propositions in  $P$ . The starting point of *interpretation vector* is defined as the *initial vector* (Definition 4).

**Definition 3 *Interpretation vector*** [13]: Let  $P$  be a program and  $B_P = \{p_1, \dots, p_n\}$ . Then an interpretation  $I \subseteq B_P$  is represented by a vector  $v = (a_1, \dots, a_n)^T$  where each element  $a_i$  ( $1 \leq i \leq n$ ) represents the truth value of the proposition  $p_i$  such that  $a_i = 1$  if  $p_i \in I$ ; otherwise,  $a_i = 0$ . We write  $\text{row}_i(v) = p_i$ .

**Definition 4 Initial vector:** Let  $P$  be a program and  $B_P = \{p_1, \dots, p_n\}$ . Then the initial vector of  $P$  is an interpretation vector  $v_0 = (a_1, \dots, a_n)^T$  such that  $a_i = 1$  ( $1 \leq i \leq n$ ) if  $\text{row}_i(v_0) = p_i$  and a fact  $p_i \leftarrow$  is in  $P$ ; otherwise,  $a_i = 0$ .

In order to compute the least model in vector space, Sakama et al. proposed an algorithm which is equivalent to the result of computing least models by the  $T_P$ -operator. This algorithm is presented in Algorithm 1.

**Definition 5  $\theta$ -thresholding:** Given a value  $x$ , define  $\theta(x) = x'$  where  $x' = 1$  if  $x \geq 1$ ; otherwise,  $x' = 0$ . Similarly, the  $\theta$ -thresholding is extended in an element-wise way to vectors and matrices.

---

**Algorithm 1** Matrix computation of least model

---

**Input:** a definite program  $P$  and its Herbrand base  $B_P = \{p_1, p_2, \dots, p_n\}$

**Output:** a vector  $v$  representing the least model

- 1: transform  $P$  to a standardized program  $P^\delta = Q \cup D$  with  $B_{P^\delta} = \{p_1, p_2, \dots, p_n, p_{n+1}, \dots, p_m\}$  where  $Q$  is an SD program and  $D$  is a set of OR-rules.
  - 2: create matrix  $M_{P^\delta} \in \mathbb{R}^{m \times m}$  representing  $P^\delta$
  - 3: create initial vector  $v_0 = (v_1, v_2, \dots, v_m)^T$  of  $P^\delta$
  - 4:  $v = v_0$
  - 5:  $u = \theta(M_{P^\delta} v)$  {refer to Definition 5}
  - 6: **while**  $u \neq v$  **do**
  - 7:    $v = u$
  - 8:    $u = \theta(M_{P^\delta} v)$  {refer to Definition 5}
  - 9: **end while**
  - 10: **return**  $v$
- 

## 2.2 Normal programs

Normal programs can be transformed to definite programs as introduced in [1]. Therefore, we transform normal programs to definite programs before encoding them in matrices.

**Definition 6 Normal program:** A normal program is a finite set of normal rules:

$$h \leftarrow b_1 \wedge b_2 \wedge \dots \wedge b_l \wedge \neg b_{l+1} \wedge \dots \wedge \neg b_m \quad (m \geq l \geq 0) \quad (3)$$

where  $h$  and  $b_i$  ( $1 \leq i \leq m$ ) are propositional variables (atoms) in  $\mathcal{L}$ .

$P$  is transformed to a definite program by rewriting the above rule to the following form:

$$h \leftarrow b_1 \wedge b_2 \wedge \dots \wedge b_l \wedge \bar{b}_{l+1} \wedge \dots \wedge \bar{b}_m \quad (m \geq l \geq 0) \quad (4)$$

where  $\bar{b}_i$  is a new proposition associated with  $b_i$ .

In this part, we denote  $P$  as a normal program with an interpretation  $I \subseteq B_P$ . The positive form  $P^+$  of  $P$  is obtained by applying the above transformation. Since a definite program  $P^+$  is transformed to its standardized program, then we can apply Algorithm 1 to compute the least model. [1] proved that if  $P$  is a normal program,  $I$  is a stable model of  $P$  iff  $I^+$  is the least model of  $P^+ \cup \bar{I}$ . We should note that  $I^+$  is interpretation of  $P^+$  which is a definite program. We can obtain  $P^+$  by applying Algorithm 1 to the transformed program  $P^+$ . Define  $\bar{I} = \{\bar{p} \mid p \in B_P \setminus I\}$ , then  $I^+ = I \cup \bar{I}$ .

**Definition 7 Matrix representation of normal programs** [11]: Let  $P$  be a normal program and  $B_P = \{p_1, \dots, p_n\}$  and its positive form  $P^+$  with  $B_{P^+} = \{p_1, \dots, p_n, \bar{q}_{n+1}, \dots, \bar{q}_m\}$ .

Then  $P^+$  is represented by a matrix  $M_P \in \mathbb{R}^{m \times m}$  such that for each element  $a_{ij}$  ( $1 \leq i, j \leq m$ ):

1.  $a_{ii} = 1$  for  $n+1 \leq i \leq m$ ;
2.  $a_{ij} = 0$  for  $n+1 \leq i \leq m$  and  $1 \leq j \leq m$  such that  $i \neq j$ ;
3. Otherwise,  $a_{ij}$  ( $1 \leq i \leq n$ ;  $1 \leq j \leq m$ ) is encoded as in Definition 2.

$M_P$  is called a *program matrix*. We write  $\text{row}_i(M_P) = p_i$  and  $\text{col}_j(M_P) = p_j$  ( $1 \leq i, j \leq n$ ).

**Example 2** Consider a program  $P = \{p \leftarrow q \wedge s, q \leftarrow p \wedge t, s \leftarrow \neg t, t \leftarrow, u \leftarrow v\}$ .

First, we need to transform  $P$  to  $P^+$  such that  $P^+ = \{p \leftarrow q \wedge s, q \leftarrow p \wedge t, s \leftarrow \bar{t}, t \leftarrow, u \leftarrow v\}$ . Then applying Definition 7, we obtain:

$$\begin{array}{c} p \\ q \\ s \\ t \\ u \\ v \\ \bar{t} \end{array} \begin{pmatrix} p & q & s & t & u & v & \bar{t} \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Instead of the initial vector in the case of definite programs, the notion of an initial matrix is introduced to encode positive and negative facts in a program.

**Definition 8 Initial matrix** [11]: Let  $P$  be a normal program and  $B_P = \{p_1, \dots, p_n\}$  and its positive form  $P^+$  with  $B_{P^+} = \{p_1, \dots, p_n, \bar{q}_{n+1}, \dots, \bar{q}_m\}$ .

The initial matrix  $M_0 \in \mathbb{R}^{m \times h}$  ( $1 \leq h \leq 2^{m-n}$ ) is defined as follows:

1. each row of  $M_0$  corresponds to each element of  $B_P$  in a way that  $\text{row}_i(M_0) = p_i$  for  $1 \leq i \leq n$  and  $\text{row}_i(M_0) = \bar{q}_i$  for  $n+1 \leq i \leq m$ ;
2.  $a_{ij} = 1$  ( $1 \leq i \leq n, 1 \leq j \leq h$ ) iff a fact  $q_i \leftarrow$  is in  $P$ ; otherwise  $a_{ij} = 0$ ;
3.  $a_{ij} = 0$  ( $n+1 \leq i \leq m, 1 \leq j \leq h$ ) iff a fact  $q_i$  is in  $P$ ; otherwise, there are two possibilities 0 and 1 for  $a_{ij}$ , so it is either 0 or 1.

Each column of  $M_0$  is a potential stable model in the first stage. We update  $M_0$  by applying matrix multiplication with the matrix representation obtained by Definition 7 as  $M_{k+1} = \theta(M_P M_k)$ . Then, the algorithm for computing the stable models is presented in Algorithm 2.

This method requires extra steps on transforming and finding stable models of a program. In addition, the initial matrix size grows exponentially by the number of negations  $m - n$ . Therefore this representation requires a lot of memory and the algorithm performs considerably slower than the method for definite programs if there are many negations appear in the program. But we will later show that this method still has the advantage when there are a small number of negations.

**Algorithm 2** Matrix computation of stable models**Input:** a normal program  $P$  and its Herbrand base  $B_P = \{p_1, p_2, \dots, p_n\}$ **Output:** a set of vectors  $V$  representing the stable models of  $P$ 

- 1: transform  $P$  to a *standardized program*  $P^+$  with  $B_{P^+} = \{p_1, \dots, p_n, \bar{q}_{n+1}, \dots, \bar{q}_m\}$ .
- 2: create the matrix  $M_P \in \mathbb{R}^{m \times m}$  representing  $P^+$
- 3: create the initial matrix  $M_0 \in \mathbb{R}^{m \times h}$
- 4:  $M = M_0, U = \theta(M_P M)$  {refer to Definition 5}
- 5: **while**  $U \neq M$  **do**
- 6:    $M = U, U = \theta(M_P M)$  {refer to Definition 5}
- 7: **end while**
- 8:  $V =$  find stable models of  $P$  {refer to Algorithm 3}
- 9: **return**  $V$

### 3 Sparse representation of logic programs

The idea of representing logic programs in vector spaces could minimize the work with symbolic computation and utilize better computing performance. Besides that, this method copes with the curse of dimension when a matrix representing logic programs becomes very large. Previous works on this topic only consider dense matrices for their implementation and it seems not very impressive in terms of performance even on small datasets [11]. In order to solve this problem, this paper focuses on analyzing the sparsity of logic programs in vector spaces and proposes improvement using sparse representation for logic programs.

#### 3.1 Sparsity of logic programs in vector spaces

A sparse matrix is a matrix in which most of the elements are zero. The level of sparseness is measured by sparsity which equals the number of zero-valued elements divided by the total number of elements [4]. Because there are a large number of zero elements in sparse matrices, we can save the computation by ignoring these zero values [8]. According to the conversion method of linear algebraic approach, we can calculate the sparsity of a program  $P$ <sup>2</sup>. This calculation is done by counting the number of nonzero-valued elements of each rule in  $P$ , then let 1 minus the fraction of the number of nonzero-valued elements and the matrix size.

By definition, the sparsity of a program  $P$  is computed by the following equation:

$$sparsity(P) = 1 - \frac{\sum_{r \in P} |body(r)|}{n^2} \quad (5)$$

where  $n$  is the number of elements in  $B_P$  and  $|B_r|$  is the length of body of rule  $r$ .

Accordingly, the representation matrix becomes a high level of sparsity if matrix size becomes larger while the length of body rule is insignificant. In fact, a rule  $r$  in a logic program rarely has a body length approx  $n$ , therefore,  $|B_r| \ll n$ . In short, we can say that the matrix representation of a logic program according to the linear algebraic approach is highly sparse.

<sup>2</sup>We only consider the programs in Definition 2 and Definition 7.

**Algorithm 3** Find stable models of  $P$ **Input:** program matrix  $M$ **Output:** a set of vectors  $V$  representing the stable models of  $P$ 

```

1:  $V = \emptyset$ 
2: for  $i$  from 1 to  $h$  do
3:    $v = (a_1, \dots, a_n, a_{n+1}, \dots, a_m)^T$  ( $i^{\text{th}}$ -column of  $M$ )
4:   for  $j$  from  $n+1$  to  $m$  do
5:      $\bar{q}_j = \text{row}_j(M)$ 
6:     for  $l$  from 1 to  $n$  do
7:       if  $\text{row}_l(M) = q_j$  then
8:         if  $a_l + a_j \neq 1$  then break;
9:       end if
10:    end for
11:    if  $l \leq n$  then break;
12:  end for
13:  if  $j \leq m$  then break;
14:  else  $V = V \cup \{v\}$ 
15: end for
16: return  $V$ 

```

**3.2 Converting logic programs to sparse matrices**

There are several sparse matrix representations. Compressed Sparse Row (CSR) is one of the most efficient, robust and widely adopted by many programming libraries among those [4]. Hence, in this paper, we propose CSR for representing logic programs.

In order to understand the CSR format, firstly we need to mention the Coordinate (COO) format which is simple using 2 arrays of coordinates and 1 array of values. The length of these arrays is equal to the number of nonzero elements. The first array stores the row index of each value, and the second array stores the row and column indices of each value, while the third array stores the values in the original matrix. We can imagine that the  $i^{\text{th}}$  nonzero element in a matrix is represented by a 3-tuple extracted from these 3 arrays at index  $i$ . Example 3 illustrates sparse representation in COO format for the program  $P$  in Example 1. We should note that in Example 3, zero-based indexing is used.

**Example 3** COO representation for  $P$  in Example 1

Row index	0	0	1	2	3	4	5	5	6	6
Col index	5	6	4	3	3	4	1	2	3	4
Value	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5	0.5	0.5

Noticeably, in the row index array, it is possible for a value to be repeated consecutively because the nonzero elements may appear in the same row for many times. We may reduce the size of the row index array by considering CSR format. In this format, while the column index and the value arrays remain the same, we compress the row index array by storing the index of the row only where nonzero elements appear. That means we do not need to store 2 consecutive 0 and 2 consecutive 5 as in Example 3. Instead, we store the index of the next row, then finally point the last index to the end of the row (which equals to the number of nonzero elements). Concretely in the row index array, the first element is starting index

which is 0. The last element is an extra element to indicate the end of this array which is equal to the number of nonzero elements. We need 2 consecutive values in row index array to extract the nonzero elements in this row. To be specific, we need to interpret  $row\_start$  and  $row\_end$  of the  $i^{th}$  row from the compressed value in  $row\_index$  array:  $row\_start_i = row\_index[i]$ ,  $row\_end_i = row\_index[i + 1]$ .

Example 4 illustrates this method. For the first row ( $i = 0$ ), we have  $row\_start_0 = 0$ ,  $row\_end_0 = 2$ , then we extract 2 values 0 and 1 for the nonzero element in the first row. These  $start$  and  $end$  will be used to extract column index and value of nonzero elements. Similarly, the second row ( $i = 1$ ), we have  $row\_start_1 = 2$ ,  $row\_end_1 = 3$  then we have only one nonzero element at index 2. Continue this interpretation until we reach the final row ( $i = 6$ ), we have  $row\_start_6 = 8$ ,  $row\_end_6 = 10$  then we extract last two nonzero elements at index 8 and 9 for the final row.

**Example 4** CSR representation for  $P$  in Example 1

Row index	0	2	3	4	5	6	8	10		
Col index	5	6	4	3	3	4	1	2	3	4
Value	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5	0.5	0.5

As we can see in Example 4, the row index array now has only 8 indices rather than 10 in Example 3. We save storing repeatedly indices in the row index array by storing only the position where it starts and ends. This phenomenon does not always encourage the reduction in terms of array size. Imagine the situation where each row has only one nonzero element, then we save nothing with this representation. Actually for a sparse matrix of the size  $m \times n$ , the CSR format saves on memory only when  $nnz < (m(n - 1) - 1)/2$  (where  $nnz$  is number of nonzero elements). Fortunately, in case of linear algebraic method, each rule normally has many atoms in the body, therefore the matrix representation has many nonzero elements in a single row. That is why CSR format could considerably save memory. In fact, we can save up to 20% of the size of the row index array using CSR format. Accordingly, in our method, we propose to implement CSR rather than COO not only because it saves more memory, but also because it is widely adopted by many programming libraries.

Because a logic program  $P$  is highly sparse, applying Algorithm 1 and Algorithm 2 on sparse representation is remarkably faster than the dense matrix. Moreover, sparse representation saves the memory space as well, therefore enabling the ability to deal with large scale KBs. We are going to reveal the performance gain by using sparse representation in the next section.

Note that only the matrix representation of the program is sparse, while the initial vector and the initial matrix are not sparse. Thus, in our methods, we will keep the dense format for interpretation matrices.

## 4 Experimental results

In this section, we conduct two experiments on finding the least models of definite programs and computing stable models of normal programs. In order to evaluate the performance of linear algebraic methods, we complete the implementations of Algorithm 1 and Algorithm 2 with (i)  $T_P$ -operator and (ii) Clasp (Clingo v5.4.1 running with flag `--method=clasp`). Our implementations are done with (iii) dense matrices and (iv) sparse matrices. Except Clasp, all implementations are implemented on C++ with CPU x64 as a targeted device (we do not use GPU accelerated code). In terms of matrix representations and operators, we use Eigen 3 library [7]. The computer running experiments has the following con-

figurations: CPU: Intel Cote i7-4770 (4 cores, 8 threads) @3.4GHz; RAM: 16GB DDR3 @1333MHz; Operating system: Ubuntu 18.04 LTS 64bit.

Focusing on analyzing the performance of sparse representation, we first evaluate our method by conducting experiments on randomized logic programs. We use the same method of LP generation conducted in [11] that the size of logic program defined by the size  $n = |B_P|$  of the Herbrand base  $B_P$  and the number of rules  $m = |P|$  in  $P$ . The rules are uniformly generated based on the length (maximum length is 8) of rule body according to Table 1.

Table 1: Proportion of rules in  $P$  based on the number of propositional variables in their bodies.

Body length	0	1	2	3	4	5	6	7	8
Allocated proportion	$< n/3$ <sup>3</sup>	4%	4%	10%	40%	35%	4%	2%	1%

We further generate denser matrices in order to analyze the efficacy of the sparse method. While keeping the same proportion of facts and rules with body length are 1 and 2, we generate the rest 70 ~ 80% rules such that their body length is around 5% of the number of propositions. This method leads to the lower sparsity level of generated matrices with approximate 0.95.

Also based on the generation method for definite programs, we generate normal programs by randomly changing  $k$  ( $4 \leq k \leq 8$ ) literals to negations. The important difference from [11] is we do experiments on much larger  $n$  and  $m$ , because our method, which is implemented on C++, is dramatically more efficient than Nguyen et al.’s implementation using Maple. The largest size of the logic program in this experiment reaches thousands of propositions and hundreds of thousands of rules. Further, we also compare our method with one of the best Answer Set Programming (ASP) solvers - Clasp [6] running in the same environment. All methods are conducted 30 times on each LP to obtain mean values of execution time.

In addition, we also conduct a further experiment using non-random problems with definite programs using transitive closure problem. The graph we use is selected from the Koblenz network collection [10]. This dataset contains binary tuples and we compute transitive closure of them using the following rules:  $path(X, Y) \leftarrow edge(X, Y)$  and  $path(X, Y) \leftarrow edge(X, Z) \wedge path(Z, Y)$

## 4.1 Definite programs

The final results on definite programs are illustrated in Table 2. We can see in the results, Dense matrix method is the slowest method and being unable to run with very large programs. Overall, Sparse matrix method is very efficient which is 10 ~ 15 faster than Clasp. We should mention that all the codes are executed on single-threaded CPU without using GPU boost or any other parallel computing techniques.

<sup>3</sup>This is the proportion of facts in  $P$ .

<sup>4</sup> $n'$  is the size of the Herbrand base of a standardized program

Table 2: Details of experimental results on definite programs of  $T_P$ -operator, Clasp and linear algebraic methods (with dense and sparse representation).  $n'$  indicates the actual matrix size after transformation.

No.	$n$	$m$	$n'$ <sup>4</sup>	Sparsity	$T_P$ -operator	Clasp	Dense matrix	Sparse matrix
1	1000	5000	5788	0.99	0.0402	0.1680	2.0559	<b>0.0071</b>
2	1000	10000	10799	0.99	0.1226	0.2940	17.9986	<b>0.0127</b>
3	1600	24000	25198	0.99	0.3952	1.8480	73.3541	<b>0.0357</b>
4	1600	30000	31285	0.99	0.4793	2.5360	116.1158	<b>0.0605</b>
5	2000	36000	37596	0.99	0.7511	3.1690	155.4312	<b>0.0692</b>
6	2000	40000	41936	0.99	0.9763	5.1610	187.6549	<b>0.0675</b>
7	10000	120000	127119	0.99	18.5608	9.0720	-	<b>0.3798</b>
8	10000	160000	167504	0.99	25.6532	15.7760	-	<b>0.4832</b>
9	16000	200000	211039	0.99	57.0223	19.9760	-	<b>0.8643</b>
10	16000	220000	231439	0.99	60.4486	24.7860	-	<b>0.9429</b>
11	20000	280000	297293	0.99	104.9978	30.5730	-	<b>0.9048</b>
12	20000	320000	337056	0.99	108.5883	34.4030	-	<b>1.0614</b>

Table 3: Details of experimental results on definite programs (with lower sparsity level) of  $T_P$ -operator, Clasp and linear algebraic methods (with dense and sparse representation).  $n'$  indicates the actual matrix size after transformation.

No.	$n$	$m$	$n'$	Sparsity	$T_P$ -operator	Clasp	Dense matrix	Sparse matrix
1	1000	5000	5876	0.95	0.1044	0.3970	2.3102	<b>0.0384</b>
2	1000	10000	10243	0.95	0.3613	0.9160	17.5917	<b>0.0519</b>
3	1600	24000	25712	0.95	0.9478	2.2540	70.0931	<b>0.1634</b>
4	1600	30000	31430	0.95	1.1817	3.0130	120.5195	<b>0.3772</b>
5	2000	36000	36612	0.95	1.7335	4.7810	152.9104	<b>0.5499</b>
6	2000	40000	41509	0.95	2.0378	6.3260	192.3609	<b>0.6284</b>
7	10000	120000	125692	0.95	27.8011	10.8930	-	<b>1.0816</b>
8	10000	160000	166741	0.95	47.2419	18.6050	-	<b>2.2907</b>
9	16000	200000	210526	0.95	89.5501	21.7110	-	<b>3.7931</b>
10	16000	220000	230178	0.95	108.1297	28.5370	-	<b>4.8605</b>
11	20000	280000	298582	0.95	144.8006	35.0920	-	<b>5.3361</b>
12	20000	320000	335918	0.95	183.5328	42.8420	-	<b>5.9182</b>

The benchmark on denser matrix are presented in Table 3. As can be seen in the results, denser matrices require more computation for Sparse matrix method while it does not affect the same scale on other competitors. Despite of that fact, Sparse matrix method still holds the first place in this benchmark. In terms of analyzing the sparseness level of logic programs, we hardly find a program in which the sparsity is less than 0.97. This observation strongly encourages the use of sparse representation for logic programs.

We next show the comparison for computing transitive closure. We assume that a dataset contains *edges* (tuples of *nodes*), then first perform grounding two rules of defining *path*. The obtained results are

demonstrated in Table 4. In this non-randomized problem, we can see that the matrix representations are very sparse. Therefore, it is no doubt that Sparse matrix method outperforms Dense matrix method. Accordingly, we only highlight the efficiency of sparse representation and omit the dense matrix approach. Surprisingly, Sparse matrix method surpasses Clasp once again in this experiment by a large margin.

Table 4: Details of experimental results on transitive closure problem of  $T_p$ -operator, Clasp and sparse representation approach.  $n'$  indicates the actual matrix size after transformation.

Data name ( $ V ,  E $ )	$n$	$m$	$n'$	Sparsity	$T_p$ -operator	Clasp	Sparse matrix
Club membership (65, 95)	1200	14492	15600	0.99	0.8397	0.3370	<b>0.0255</b>
Cattle (28, 217)	1512	20629	21924	0.99	0.9541	0.5060	<b>0.0365</b>
Windsurfers (43, 336)	4324	99788	103776	0.99	3.6453	3.3690	<b>0.1824</b>
Contiguous USA (49, 107)	4704	113003	117600	0.99	4.2975	3.8830	<b>0.1830</b>
Dolphins (62, 159)	7564	230861	238266	0.99	12.3067	9.3820	<b>0.4019</b>
Train bombing (64, 243)	8064	254259	262080	0.99	15.2257	10.6350	<b>0.4524</b>
Highschool (70, 366)	9660	333636	342930	0.99	19.9622	15.8010	<b>0.6618</b>
Les Miserables (77, 254)	11704	445006	456456	0.99	27.7931	21.9560	<b>0.8300</b>

As can be witnessed in the results, Dense matrix method is the slowest, even slower than  $T_p$ -operator, in terms of computation time due to wasting of computation on a huge amount of zero elements. This could be explained by the high level of sparsity of logic programs provided in Table 2, Table 3 and Table 4. Moreover, large dense matrices consume a huge amount of memory therefore the method is unable to run with large scale matrix size. Overall, sparse matrix method is effective in computing the fixpoint of definite programs. On the other hand, the performance would be improved if we use GPU accelerated code and exploit parallel computing power. The results indicates a potential for logical inference using an algebraic method.

## 4.2 Normal programs

In our current method, the number of columns in the initial matrix (Definition 8) grows exponentially by the number of negations, we limit the number of negations in this benchmark by 8 as specified in the experiment setup.

Table 5: Details of experimental results on normal programs of  $T_P$ -operator, Clasp and linear algebraic methods (with dense and sparse representation).  $n'$  indicates the actual matrix size after transformation.

No.	$n$	$m$	$n'$	$k$ <sup>5</sup>	Sparsity	$T_P$ -operator	Clasp	Dense matrix	Sparse matrix
1	1000	5000	6379	8	0.99	0.0472	0.3070	3.9560	<b>0.0119</b>
2	1000	10000	12745	8	0.99	0.1838	1.0920	28.1806	<b>0.0178</b>
3	1600	24000	30061	8	0.99	0.5525	3.2760	105.4931	<b>0.0559</b>
4	1600	30000	36402	7	0.99	0.6801	4.3050	168.8044	<b>0.0832</b>
5	2000	36000	42039	5	0.99	1.2378	6.7180	203.2749	<b>0.0897</b>
6	2000	40000	48187	8	0.99	1.5437	7.1800	256.9701	<b>0.0991</b>
7	10000	120000	171967	6	0.99	27.3162	7.6820	-	<b>0.7124</b>
8	10000	160000	207432	7	0.99	32.5547	24.6990	-	<b>0.8424</b>
9	16000	200000	250194	5	0.99	70.3114	30.7180	-	<b>1.5603</b>
10	16000	220000	278190	6	0.99	86.5192	35.4050	-	<b>1.8314</b>
11	20000	280000	357001	4	0.99	133.7881	50.1970	-	<b>1.9170</b>
12	20000	320000	396128	4	0.99	150.3377	58.6090	-	<b>2.1066</b>

Table 6: Details of experimental results on normal programs (with lower sparsity level) of  $T_P$ -operator, Clasp and linear algebraic methods (with dense and sparse representation).  $n'$  indicates the actual matrix size after transformation.  $h$  indicates the column size of the initial matrix.

No.	$n$	$m$	$n'$	$k$	Sparsity	$T_P$ -operator	Clasp	Dense matrix	Sparse matrix
1	1000	5000	6385	7	0.95	0.1680	0.3680	3.7791	<b>0.1133</b>
2	1000	10000	12294	8	0.95	0.2453	1.4940	30.0642	<b>0.1867</b>
3	1600	24000	33172	7	0.95	0.6819	3.7830	102.5389	<b>0.2219</b>
4	1600	30000	35091	8	0.95	0.7741	5.9120	174.5192	<b>0.3462</b>
5	2000	36000	44145	8	0.95	2.3194	7.1020	197.3004	<b>0.4131</b>
6	2000	40000	49080	7	0.95	3.2665	8.6690	250.0876	<b>0.4895</b>
7	10000	120000	181550	8	0.95	36.9532	10.4530	-	<b>3.2504</b>
8	10000	160000	203576	6	0.95	54.1106	33.1920	-	<b>4.0186</b>
9	16000	200000	246159	4	0.95	86.3571	48.1860	-	<b>7.2193</b>
10	16000	220000	282734	5	0.95	106.0275	56.9150	-	<b>8.3059</b>
11	20000	280000	365190	4	0.95	163.0558	78.1790	-	<b>9.0177</b>
12	20000	320000	387094	4	0.95	202.5501	84.3270	-	<b>11.5203</b>

First, we perform benchmarks on normal programs which has 0.99 sparsity level. Table 5 illustrates the execution time in detail. As can be witnessed in the results, Sparse matrix method is still faster than Clasp but with a smaller scale it did in definite programs. It is needed to mention that the initial matrix size is remarkably larger due to the limitation of representation. We have to initialize all possible combinations of an atom which appeared with its negation form in the program. There is no doubt that with a larger number of negations, the space complexity of linear algebraic method is exponential.

---

<sup>5</sup>The number of negations.

Accordingly, the performance of Sparse matrix method is only better than Clasp with a small fraction of negations.

In the next experiments, we compare different methods on denser matrix. Table 6 presents the data for this benchmark. Once again, with a limited number of negations, Sparse matrix method holds the winner position.

Noticeably, execution time on normal programs is generally greater than that on definite programs. This is obvious because we have a larger size of initial matrices as well as the need of extra computation on transforming and finding the least models as described in Algorithm 2. Then the weakness of the linear algebraic method is that we have to deal with all combinations of truth assignments in order to compute the stable model. Accordingly, the column size of the initial matrix exponentially increases by the number of negations. Thus, in the benchmark on randomized programs, we limit the number of negations for all benchmarks so that the matrix can fit in memory. This limitation will become clearer in real problems which have many negations. This is a major problem that we are investigating to do further research.

## 5 Conclusion

In this paper, we analyze the sparsity of matrix representation for LP and then propose an improvement for logic programming in vector space using sparse matrix representation. The experimental results on computing the least models of definite programs demonstrate very significant enhancement in terms of computation performance even when compared to Clasp. This improvement remarkably reduced the burden of computation in previous linear algebraic approaches for representing LP. Whereas in finding stable models of normal programs, the efficacy of linear algebraic method is limited due to the representation requires a huge amount of memory to store all possible combinations of negated atoms. In spite of the fact, our method is efficient when there are small numbers of negation. However, matrix computation could be more accelerated using GPU. We have tested our implementation in this way, and obtained expected results too.

Sato's linear algebraic method [15] is based on a completely different idea to represent logic programs, where each predicate is represented in one matrix and an approximation method is used to compute the extension of a target predicate of a recursive program. We should note that this approximation method is limited in a matrix size 10,000, while our exact method is comfortable with 320,000. Further comparison is a future research topic, yet we could expect that Sato's method can also be enhanced by sparse representation.

The encouraging results open up rooms for improvement and optimization. In an extended version of this paper, we will add more experimental results including a comparison on other benchmark programs for ASP. Potential future work is to apply a sampling method to reduce the number of guesses in the initial matrix for normal programs. An algorithm would be to prepare some manageable size of the initial matrix, and if all guesses fail we do some local search and replace column vectors by new assignments then repeat it until a stable model is found. Further research directions on implementing disjunctive LP and abductive LP should be considered in order to reveal the applicability of tensor-based approaches for LP. Additionally, more complex types of the program should be taken in to account to be represented in vector space, for instance 3-valued logic programs and answer set programs with aggregates and constraints.

## References

- [1] Jose Julio Alferes, Joao Alexandre Leite, Luis Moniz Pereira, Halina Przymusinska & Teodor C. Przymusinski (2000): *Dynamic updates of non-monotonic knowledge bases*. *The Journal of Logic Programming* 45(1-3), pp. 43–70, doi:10.1016/S0743-1066(99)00065-5.
- [2] Yaniv Aspis (2018): *A Linear Algebraic Approach to Logic Programming*. Master thesis at Imperial College London.
- [3] Yaniv Aspis, Krysia Broda & Alessandra Russo (2018): *Tensor-based abduction in horn propositional programs*. 2206, CEUR Workshop Proceedings, pp. 68–75, doi:10.1145/200836.200838.
- [4] James R Bunch & Donald J Rose (2014): *Sparse matrix computations*. Academic Press, doi:10.1016/C2013-0-10439-4.
- [5] William W Cohen (2016): *Tensorlog: A differentiable deductive database*. *arXiv preprint arXiv:1605.06523*.
- [6] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, Max Ostrowski, Torsten Schaub & Philipp Wanko (2016): *Theory solving made easy with clingo 5*. In: *OASICS-OpenAccess Series in Informatics*, 52, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, doi:10.4230/OASICS.ICLP.2016.2.
- [7] Gaël Guennebaud, Benoît Jacob et al. (2010): *Eigen v3*. <http://eigen.tuxfamily.org>.
- [8] Fred G Gustavson (1978): *Two fast algorithms for sparse matrices: Multiplication and permuted transposition*. *ACM Transactions on Mathematical Software (TOMS)* 4(3), pp. 250–269, doi:10.1145/355791.355796.
- [9] Robert Kowalski (1974): *Logic for problem solving*. Department of Computational Logic, Edinburgh University, doi:10.1145/1005937.1005947.
- [10] Jérôme Kunegis (2013): *Konect: the koblenz network collection*. In: *Proceedings of the 22nd International Conference on World Wide Web*, pp. 1343–1350, doi:10.1145/2487788.2488173.
- [11] Hien D. Nguyen, Chiaki Sakama, Taisuke Sato & Katsumi Inoue (2018): *Computing Logic Programming Semantics in Linear Algebra*. In: *Proceedings of the 12th International Conference on Multi-disciplinary Trends in Artificial Intelligence (MIWAI 2018)*, Springer International Publishing, Lecture Notes in Artificial Intelligence, Vol. 11248, pp. 32–48, doi:10.1007/978-3-030-03014-8\_3.
- [12] Tim Rocktäschel, Matko Bosnjak, Sameer Singh & Sebastian Riedel (2014): *Low-dimensional embeddings of logic*. In: *Proceedings of the ACL 2014 Workshop on Semantic Parsing*, pp. 45–49, doi:10.3115/v1/W14-2409.
- [13] Chiaki Sakama, Katsumi Inoue & Taisuke Sato (2017): *Linear Algebraic Characterization of Logic Programs*. In: *Proceedings of the 10th International Conference on Knowledge Science, Engineering and Management (KSEM 2017)*, Springer International Publishing, Lecture Notes in Artificial Intelligence, Vol. 10412, pp. 520–533, doi:10.1007/978-3-319-63558-3\_44.
- [14] Taisuke Sato (2017): *Embedding Tarskian semantics in vector spaces*. In: *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*.
- [15] Taisuke Sato (2017): *A linear algebraic approach to Datalog evaluation*. *Theory and Practice of Logic Programming* 17(3), pp. 244–265, doi:10.1017/S1471068417000023.
- [16] Taisuke Sato, Katsumi Inoue & Chiaki Sakama (2018): *Abducing Relations in Continuous Spaces*. In: *IJCAI*, pp. 1956–1962, doi:10.24963/ijcai.2018/270.
- [17] Maarten H Van Emden & Robert A Kowalski (1976): *The semantics of predicate logic as a programming language*. *Journal of the ACM (JACM)* 23(4), pp. 733–742, doi:10.1145/321978.321991.
- [18] Bishan Yang, Scott Wen-tau Yih, Xiaodong He, Jianfeng Gao & Li Deng (2015): *Embedding Entities and Relations for Learning and Inference in Knowledge Bases*. In: *Proceedings of the International Conference on Learning Representations (ICLR) 2015*. Available at <https://www.microsoft.com/en-us/research/publication/embedding-entities-and-relations-for-learning-and-inference-in-knowledge-bases/>.