

Savile Row Manual

Peter Nightingale

Abstract

We describe the constraint modelling tool Savile Row, its input language and its main features. Savile Row translates a solver-independent constraint modelling language to the input languages for various solvers including constraint, SAT, and SMT solvers. After a brief introduction, the manual describes the Essence Prime language, which is the input language of Savile Row. Then we describe the functions of the tool, its main features and options and how to install and use it.

Contents

1	Introduction	2
1.1	Problem Classes and Instances	3
1.2	Solver Backends	3
2	The Essence Prime Language	4
2.1	The Essence Prime Expression Language	5
2.1.1	Types and Domains	5
2.1.2	Domain Expressions	6
2.1.3	Literals	6
2.1.4	Variables	7
2.1.5	Expression Types	7
2.1.6	Type Conversion	7
2.1.7	Matrix Indexing and Slicing	7
2.1.8	Integer and Boolean Expressions	8
2.1.9	Integer Operators	8
2.1.10	Boolean Operators	9
2.1.11	Integer and Boolean Functions	9
2.1.12	Numerical Comparison Operators	9
2.1.13	Matrix Comparison Operators	9
2.1.14	Set Operator	9
2.1.15	The Quantified Sum Operator	10
2.1.16	Universal and Existential Quantification	10
2.1.17	Quantification over Matrix Domains	11
2.1.18	Global Constraints	11
2.1.19	Table Constraints	11
2.1.20	Matrix Comprehensions	14
2.1.21	Matrix Comprehensions over Matrix Domains	14
2.1.22	Functions on Matrices	15
2.2	Model Structure	17
2.2.1	Parameter Declarations with <code>given</code>	18
2.2.2	Constant Definitions with <code>letting</code>	18
2.2.3	Constant Domains	19
2.2.4	Decision Variable Declaration with <code>find</code>	20

2.3	Constraints on Parameters with <code>where</code>	20
2.4	Objective	21
2.5	Solver Control	21
2.6	Constraints	22
3	Undefinedness in ESSENCE PRIME	22
4	Installing and Running SAVILE ROW	23
4.1	Running SAVILE ROW on Linux and Mac	24
4.2	Running SAVILE ROW on Windows	24
4.3	Solution Files	24
4.4	File Names	24
4.5	Command Line Options	25
4.5.1	Specifying input files	25
4.5.2	Specifying output format	25
4.5.3	Optimisation Levels	26
4.5.4	Symmetry-Breaking Levels	27
4.5.5	Translation Options	27
4.5.6	SAT Encoding Options	29
4.5.7	SMT Encoding Options	31
4.5.8	Warning Flags	31
4.5.9	Controlling SAVILE ROW	32
4.5.10	Solver Control	32
4.5.11	Solver Control – Minion	33
4.5.12	Solver Control – Gecode	33
4.5.13	Solver Control – Chuffed	33
4.5.14	Solver Control – OR Tools	33
4.5.15	Solver Control – Standard FlatZinc	33
4.5.16	Solver Control – SAT Solvers	33
4.5.17	Solver Control – SMT Solvers	34
4.5.18	ReadSolution Mode	34
4.5.19	Format Conversion	35
A	SAT Encoding	35
A.1	Decomposition	35
A.2	Encoding of CSP variables	35
A.3	Encoding of Constraints	36
B	Operator Precedence in ESSENCE PRIME	36
C	Reserved Words	36

1 Introduction

SAVILE ROW is a constraint modelling tool. It provides a high-level language for the user to specify their constraint problem, and automatically translates that language to the input languages of constraint solvers as well as SAT, MaxSAT, and SMT solvers. SAVILE ROW is a research tool, designed to enable research into reformulation of constraint models, therefore it is designed for flexibility. It is designed to be easy to add new transformation rules and develop new translation pipelines.

This manual is for version 1.11.1 of SAVILE ROW.

This manual covers the basics of using SAVILE ROW. It does not cover adding new rules or translation pipelines.

SAVILLE ROW reads the ESSENCE PRIME modelling language, which is described in Section 2 below. SAVILE ROW converts constraint models expressed in ESSENCE PRIME into the solver input format, in a process that has some similarities to compiling a high-level programming language. Like a compiler, SAVILE ROW applies some optimisations to the model (the most basic being partial evaluation and common subexpression elimination).

1.1 Problem Classes and Instances

The distinction between *problem classes* and *problem instances* will be important in this manual. It is easiest to start with an example. Sudoku in general is a problem class, and a particular Sudoku puzzle is an instance of that class. The problem class is described by writing down the rules of Sudoku, i.e. that we have a 9×9 grid and that each row, column and subsquare must contain all values 1..9. A particular instance is described by taking the class and adding the clues, i.e. filling in some of the squares. The set of clues is a *parameter* of the class – in this case the only parameter. Adding the parameters to a class to make a problem instance is called *instantiation*.

In typical use, SAVILE ROW will read a problem class file and also a parameter file, both written in ESSENCE PRIME. It will instantiate the problem class and unroll all quantifiers and matrix comprehensions, then perform reformulations and flattening of nested expressions before producing output for a constraint solver.

1.2 Solver Backends

Output produced by SAVILE ROW may be entirely flattened or it may contain some nested expressions depending on the backend solver or output language. Also, global constraints will be decomposed when not supported by the backend solver, another source of difference between the different backends. The SAT, MaxSAT, and SMT backends have a final encoding step to produce their lower-level representations. The backends are as follows:

- **Minion** – Output is produced in the Minion 3 language for Minion 1.9.1 or later. The model produced is not entirely flat, it makes use of nested `watched-or` and `watched-and` constraints.
- **Chuffed** – Output is produced for Chuffed in the entirely flat, instance-level language FlatZinc for use by the `fzn-chuffed` tool.
- **Gecode** – Similar to Chuffed with some differences in the set of constraints that are decomposed.
- **OR-Tools** – Similar to Chuffed and Gecode, with some differences in the set of constraints that are decomposed.
- **Standard FlatZinc** – Output for any solver with a FlatZinc parser.
- **SAT** – Output is produced in DIMACS format for use with any SAT solver. An overview of the SAT encoding is given below.
- **SMT** – Output is produced in SMT-LIB 2 format for use with SMT solvers such as Boolector, Z3, and Yices2.
- **MaxSAT** – Produces Weighted Partial MaxSAT for use with MaxSAT solvers such as Open-WBO and MaxHS.
- **MiniZinc** – Output is produced in an instance-level, almost flat subset of the MiniZinc modelling language. The MiniZinc output follows the FlatZinc outputs as closely as possible.

The Minion, Chuffed, OR-Tools, Gecode, standard FlatZinc, SAT, SMT, and MaxSAT backends are able to run the solver and parse the solution. For these backends, SAVILE ROW will produce a solution file (in ESSENCE PRIME) and a file with some statistics. The SAT and SMT backends make multiple calls to solve optimisation problems (using a bisecting strategy by default), unless the target solver natively supports optimisation. All constraints (i.e. the entire ESSENCE PRIME language) are supported with each backend. For the bundled solvers (Minion, Kissat (SAT), Chuffed), paths are set so that SAVILE ROW always uses the bundled solver by default.

MINION is well supported by SAVILE ROW so we will use MINION as the reference in this document. Modelling problems directly in MINION's input language is time-consuming and tedious because of its primitive structure (it can be compared to writing a complex program in assembly language), so a tool like SAVILE ROW takes much of the tedium out of modelling constraint problems for MINION.

5	3			7						5	3	4	6	7	8	9	1	2
6			1	9	5					6	7	2	1	9	5	3	4	8
	9	8						6		1	9	8	3	4	2	5	6	7
8				6					3	8	5	9	7	6	1	4	2	3
4			8		3				1	4	2	6	8	5	3	7	9	1
7				2					6	7	1	3	9	2	4	8	5	6
	6						2	8		9	6	1	5	3	7	2	8	4
			4	1	9				5	2	8	7	4	1	9	6	3	5
				8			7	9		3	4	5	2	8	6	1	7	9

Figure 1: On the left is a Sudoku puzzle. The objective is to fill in all blank squares using only the digits 1-9 such that each row contains all the digits 1-9, each column also contains all the digits 1-9, and each of the 3 by 3 subsquares outlined in thicker lines also contains all digits 1-9. On the right is the solution. (Images are public domain from Wikipedia.)

2 The Essence Prime Language

The purpose of this section is to describe the ESSENCE PRIME language and to be a reference for users of ESSENCE PRIME— not to be a formal description of the language. ESSENCE PRIME is a constraint modelling language, therefore it is mainly designed for describing \mathcal{NP} -hard decision problems. It is not the only (or the first) constraint modelling language. ESSENCE PRIME began as a subset of ESSENCE [8] and has been extended from there. It is similar to the earlier Optimization Programming Language (OPL) [16]. ESSENCE PRIME is implemented by the tool SAVILE ROW [13, 14].

ESSENCE PRIME is considerably different to procedural programming languages, it does not specify a procedure to solve the problem. The user specifies the problem in terms of decision variables and constraints, and the solver automatically finds a value for each variable such that all constraints are satisfied. This means, for example, that the order the constraints are presented in ESSENCE PRIME is irrelevant.

ESSENCE PRIME allows the user to solve *constraint satisfaction problems* (CSPs). A simple example of a CSP is a Sudoku puzzle (Figure 1). To convert a single Sudoku puzzle to CSP, each square can be represented as a decision variable with domain $\{1 \dots 9\}$. The clues (filled in squares) and the rules of the puzzle are easily expressed as constraints.

We will use Sudoku as a running example. A simple first attempt of modelling Sudoku in ESSENCE PRIME is shown in Figure 2. In this case the clues (for example $M[1, 1]=5$) are included in the model. We have used not-equal constraints to state that all digits must be used at most once in each row and column. We have also omitted the sub-square constraints for now.

In this example, some CSP decision variables are declared using a `find` statement. It is also worth noting that other variables exist that are not decision variables, for example, `row` is a *quantifier* variable that exists only to apply some constraints to every row.

An ESSENCE PRIME model usually describes a *class* of CSPs. For example, it might describe the class of all Sudoku puzzles. In order to solve a particular instance of Sudoku, the instance would be specified in a separate *parameter* file (also written in ESSENCE PRIME). The model would have parameter variables (of type integer, boolean or matrix), and the parameter file would specify a value for each of these variables.

Since ESSENCE PRIME is a constraint modelling language, we will define the constraint satisfaction problem (CSP). A CSP $\mathcal{P} = \langle \mathcal{X}, \mathcal{D}, \mathcal{C} \rangle$ is defined as a set of n *decision variables* $\mathcal{X} = \langle x_1, \dots, x_n \rangle$, a set of domains $\mathcal{D} = \langle D(x_1), \dots, D(x_n) \rangle$ where $D(x_i) \subseteq \mathbb{Z}$, $|D(x_i)| < \infty$ is the finite set of all potential values of x_i , and a conjunction $\mathcal{C} = C_1 \wedge C_2 \wedge \dots \wedge C_e$ of constraints.

For CSP $\mathcal{P} = \langle \mathcal{X}, \mathcal{D}, \mathcal{C} \rangle$, a constraint $C_k \in \mathcal{C}$ consists of a sequence of $m > 0$ variables $\mathcal{X}_k = \langle x_{k_1}, \dots, x_{k_m} \rangle$ with domains $\mathcal{D}_k = \langle D(x_{k_1}), \dots, D(x_{k_m}) \rangle$ s.t. \mathcal{X}_k is a subsequence¹ of \mathcal{X} , \mathcal{D}_k is a subsequence of \mathcal{D} , and each variable x_{k_i}

¹I use subsequence in the sense that $\langle 1, 3 \rangle$ is a subsequence of $\langle 1, 2, 3, 4 \rangle$.

```

language ESSENCE' 1.0

find M : matrix indexed by [int(1..9), int(1..9)] of int(1..9)

such that

M[1,1]=5,
M[1,2]=3,
M[1,5]=7,
...
M[9,9]=9,

forall row : int(1..9) .
    forall col1 : int(1..9) .
        forall col2: int(col1+1..9) . M[row, col1]!=M[row, col2],

forall col : int(1..9) .
    forall row1 : int(1..9) .
        forall row2: int(row1+1..9) . M[row1, col]!=M[row2, col]

```

Figure 2: First example of a (partial) Sudoku model in ESSENCE PRIME

and domain $D(x_{k_i})$ matches a variable x_j and domain $D(x_j)$ in \mathcal{P} . C_k has an associated set $C_k^S \subseteq D(x_{k_1}) \times \dots \times D(x_{k_m})$ of tuples which specify allowed combinations of values for the variables in \mathcal{X}_k .

2.1 The Essence Prime Expression Language

In ESSENCE PRIME *expressions* are built up from variables and literals using operators (such as +). We will start by describing the types that expressions may take, then describe the simplest expressions and build up from there.

2.1.1 Types and Domains

Types and domains play a similar role; they prescribe a set of values that an expression or variable can take. Types denote non-empty sets that contain all elements that have a similar structure, whereas domains denote possibly empty sets drawn from a single type. In this manner, each domain is associated with an underlying type. For example integer is the type underlying the domain comprising integers between 1 and 10. The type contains all integers, and the domain is a finite subset.

ESSENCE PRIME is a strongly typed language; every expression has a type, and the types of all expressions can be inferred and checked for correctness. Furthermore, ESSENCE PRIME is a finite-domain language; every decision variable is associated with a finite domain of values.

The atomic types of ESSENCE PRIME are `int` (integer) and `bool` (boolean). There is also a compound type, `matrix`, that is constructed of atomic types.

There are three different types of domains in ESSENCE PRIME: boolean, integer and matrix domains. Boolean and integer domains are both atomic domains; matrix domains are built from atomic domains.

Boolean Domains `bool` is the Boolean domain consisting of `false` and `true`.

Integer Domains An integer domain represents a finite subset of the integers, and is specified as a sequence of ranges or individual elements, for example `int(1..3, 5, 7, 9..20)`. Each range includes the endpoints, so the meaning of a range `a..b` is the set $\{i \in \mathbb{Z} | a \leq i \leq b\}$. A range `a..b` would normally be in order, i.e. $a \leq b$ but this is not strictly required. Out-of-order ranges correspond to the empty set of integers.

The meaning of an integer domain is the union of the ranges and individual elements in the domain. For example, `int(10, 1..5, 4..9)` is equivalent to `int(1..10)`.

Finally, the elements and endpoints of ranges may be expressions of type `int`. The only restriction is that they may not contain any CSP decision variables. The integer expression language is described in the following sections.

Matrix Domains A matrix is defined by the keyword `matrix`, followed by its dimensions and the base domain. The dimensions are a list, in square brackets, of domains. For instance,

```
Matrix1 : matrix indexed by [int(1..10),int(1..10)] of int(1..5)
```

is the domain of a two-dimensional square matrix, indexed by 1..10 in both dimensions, where each element of the matrix has the domain `int(1..5)`. Elements of this matrix would be accessed by `Matrix1[A,B]` where A and B are integer expressions.

Matrices may be indexed by any integer domain or the boolean domain. For example,

```
Matrix2 : matrix indexed by [int(-2..5),int(-10..10,15,17)] of int(1..5)
```

is a valid matrix domain.

2.1.2 Domain Expressions

ESSENCE PRIME contains a small expression language for boolean and integer domains. This language consists of three binary infix operators `intersect`, `union` and `-`. All three are left-associative and the precedences are given in Appendix B. The language also allows bracketed subexpressions with `(` and `)`.

For example, the first and second lines below are exactly equivalent.

```
letting dom be domain int(1..5, 3..8)
letting dom be domain int(1..5) union int(3..8)
```

2.1.3 Literals

Each of the three types (integer, boolean and matrix) has a corresponding literal syntax in ESSENCE PRIME. Any value of any type may be written as a literal. Sets and real numbers are not (as yet) part of the language. Integer and boolean literals are straightforward:

```
1 2 3 -5
true false
```

There are two forms of matrix literals. The simpler form is a comma-separated list of expressions surrounded by square brackets. For example, the following is a matrix literal containing four integer literals.

```
[ 1, 3, 2, 4 ]
```

Matrix literals may contain any valid expression in ESSENCE PRIME. For example a matrix literal is allowed to contain other matrix literals to build up a matrix with two or more dimensions. The types of the expressions contained in the matrix literal must all be the same.

The second form of matrix literal has an explicit index domain that specifies how the literal is indexed. This is specified after the comma-separated list of contents using a `;` as follows.

```
[ 1, 3, 2, 4 ; int(4..6,8) ]
```

The index domain must be a domain of type `bool` or `int`, and must contain the same number of values as the number of items in the matrix literal. When no index domain is specified, a matrix of size `n` is indexed by `int(1..n)`.

Finally a multi-dimensional matrix value can be expressed by nesting matrix literals. Suppose we have the following domain:

```
matrix indexed by [int(-2..0),int(1,2,4)] of int(1..5)
```

One value contained in this domain is the following:

```
[ [ 1,2,3 ; int(1,2,4) ],
  [ 1,3,2 ; int(1,2,4) ],
  [ 3,2,1 ; int(1,2,4) ]
; int(-2..0) ]
```

2.1.4 Variables

Variables are identified by a string. Variable names must start with a letter `a-z` or `A-Z`, and after the first letter may contain any of the following characters: `a-z A-Z 0-9 _`. A variable may be of type integer, boolean or matrix.

Scoping of variables depends on how they are declared and is dealt with in the relevant sections below. As well as a type, variables have a category. The category is *decision*, *quantifier* or *parameter*. Decision variables are CSP variables, and the other categories are described below.

Expressions containing decision variables are referred to as *decision expressions*, and expressions containing no decision variables as *non-decision expressions*. This distinction is important because expressions in certain contexts are not allowed to contain decision variables.

2.1.5 Expression Types

Expressions can be of any of the three basic types (integer, boolean or matrix). Integer expressions range over an integer domain, for instance `x + 3` (where `x` is an integer variable) is an integer expression ranging from $lb(x) + 3$ to $ub(x) + 3$. Boolean expressions range over the Boolean domain, for instance the integer comparison `x = 3` can either be `true` or `false`.

2.1.6 Type Conversion

Boolean expressions or literals are automatically converted to integers when used in a context that expects an integer. As is conventional `false` is converted to 0 and `true` is converted to 1. For example, the following are valid ESSENCE PRIME boolean expressions.

```
1+2-3+true-(x<y)=5
false < true
```

Integer expressions are not automatically converted to boolean. Matrix expressions cannot be converted to any other type (or vice versa).

2.1.7 Matrix Indexing and Slicing

Suppose we have a three-dimensional matrix `M` with the following domain:

```
matrix indexed by [int(1..3), int(1..3), bool] of int(1..5)
```

`M` would be indexed as follows: `M[x,y,z]`, where `x` and `y` may be integer or boolean expressions and `z` must be boolean. Because the matrix has the base domain `int(1..5)`, `M[x,y,z]` will be an integer expression. Matrix indexing is a partial function: when one of the indices is out of bounds then the expression is undefined. ESSENCE PRIME has the relational semantics, in brief this means that the boolean expression containing the undefined expression

is false. So for example, $M[1, 1, \text{false}] = M[2, 4, \text{true}]$ is always false because the 4 is out of bounds. The relational semantics are more fully described in Section 3 below.

Parts of matrices can be extracted by *slicing*. Suppose we have the following two-dimensional matrix named N:

```
[ [ 1, 2, 3 ; int(1, 2, 4) ],
  [ 1, 3, 2 ; int(1, 2, 4) ],
  [ 3, 2, 1 ; int(1, 2, 4) ]
; int(-2..0) ]
```

We could obtain the first row by writing $N[-2, \dots]$, which is equal to $[1, 2, 3 ; \text{int}(1..3)]$. Similarly the first column can be obtained by $N[\dots, 1]$ which is $[1, 1, 3 ; \text{int}(1..3)]$. In general, the indices in a matrix slice may be \dots or an integer or boolean expression that does not contain any decision variables. Matrix slices are always indexed contiguously from 1 regardless of the original matrix domain.

When a matrix slice has an integer or boolean index that is out of bounds, then the expression is undefined and this is handled as described in Section 3.

2.1.8 Integer and Boolean Expressions

ESSENCE PRIME has a range of binary infix and unary operators and functions for building up integer and boolean expressions, for example:

- Integer operators: $+$ $-$ $*$ $**$ $/$ $\%$ $|$ \min \max
- Boolean operators: $\setminus/$ $\setminus\setminus$ \rightarrow \leftrightarrow $!$
- Quantified sum: sum
- Logical quantifiers: forAll exists
- Numerical comparison operators: $=$ \neq $>$ $<$ \geq \leq
- Matrix comparison operators: \leq_{lex} $<_{\text{lex}}$ \geq_{lex} $>_{\text{lex}}$
- Set operator: in
- Global constraints, for example: allDiff gcc cumulative disjunctive diffn circuit
- Table constraint: table

These are described in the following subsections.

2.1.9 Integer Operators

ESSENCE PRIME has the following binary integer operators: $+$ $-$ $*$ $/$ $\%$ $**$. $+$, $-$ and $*$ are the standard integer operators.

The operators $/$ and $\%$ are integer division and modulo functions. a/b is defined as $\lfloor a/b \rfloor$ (i.e. it always rounds down). This does not match some programming languages, for example C99 which rounds towards 0.

The modulo operator $a\%b$ is defined as $a - b\lfloor a/b \rfloor$ to be complementary to $/$.

```
3/2 = 1
(-3)/2 = -2
3/(-2) = -2
(-3)/(-2) = 1
```

```
3 % 2 = 1
(-3) % 2 = 1
3 % (-2) = -1
(-3) % (-2) = -1
```

$**$ is the power function: $x**y$ is defined as x^y . There are two unary functions: absolute value (where $|x|$ is the absolute value of x), and negation (unary $-$).

2.1.10 Boolean Operators

ESSENCE PRIME has the $/\backslash$ (and) and $\backslash/$ (or) operators defined on boolean expressions. There are also \rightarrow (implies), \leftrightarrow (if and only if), and $!$ (negation). These operators all take boolean expressions and produce a new boolean expression. They can be nested arbitrarily.

The comma $,$ in ESSENCE PRIME is also a binary boolean operator, with the same meaning as $/\backslash$. However it has a different precedence, and is used quite differently. $/\backslash$ is normally used within a constraint, and $,$ is used to separate constraints (or separate groups of constraints constructed using a `forall`). Consider the following example.

```
forall i : int(1..n) . x[i]=y[i] /\ x[i]!=y[i+1],
exists i : int(1..n) . x[i]=1 /\ y[i]!=y[i+1]
```

Here we have two quantifiers, both with an $/\backslash$ inside. The comma is used to separate the `forall` and the `exists`. The comma has the lowest precedence of all binary operators.

2.1.11 Integer and Boolean Functions

ESSENCE PRIME has named functions `min(X, Y)` and `max(X, Y)` that both take two integer expressions X and Y . `min` and `max` can also be applied to one-dimensional matrices to obtain the minimum or maximum of all elements in the matrix (see Section 2.1.22). `factorial(x)` returns the factorial of values from 0 to 20 where the result fits in a 64-bit signed integer. It is undefined for other values of x and the expression x is not allowed to contain decision variables. `popcount(x)` returns the bit count of the 64-bit two's complement representation of x , and x may not contain decision variables.

The function `toInt(x)` takes a boolean expression x and converts to an integer 0 or 1. This function is included only for compatibility with ESSENCE: it is not needed in ESSENCE PRIME because booleans are automatically cast to integers.

The function `if(cond, a, b)` evaluates to a if `cond` is `true`, otherwise it evaluates to b . All three arguments must be constant, parameter or quantifier expressions (i.e. must not contain decision variables), `cond` must be Boolean, and a, b must be Boolean or integer.

2.1.12 Numerical Comparison Operators

ESSENCE PRIME provides the following integer comparisons with their obvious meanings: $=$ \neq $>$ $<$ \geq \leq . These operators each take two integer expressions and produce a boolean expression.

2.1.13 Matrix Comparison Operators

ESSENCE PRIME provides a way of comparing one-dimensional matrices. These operators compare two matrices using the dictionary order (*lexicographical* order, or `lex` for short).

There are four operators. $A <_{lex} B$ ensures that A comes before B in `lex` order, and $A \leq_{lex} B$ which ensures that $A <_{lex} B$ or $A=B$. \geq_{lex} and $>_{lex}$ are identical but with the arguments reversed. For all four operators, A and B may have different lengths and may be indexed differently, but they must be one-dimensional. Multi-dimensional matrices may be flattened to one dimension using the `flatten` function described in Section 2.1.22 below.

2.1.14 Set Operator

The operator `in` states that an integer expression takes a value in a set expression. The set expression may not contain decision variables. The set may be a domain expression (Section 2.1.2) or the `toSet` function that converts a one-dimensional matrix to a set, as in the examples below.

```
x+y in (int(1,3,5) intersect int(3..10))
x+y in toSet([ i | i : int(1..n), i%2=0])
```

2.1.15 The Quantified Sum Operator

The `sum` operator corresponds to the mathematical \sum and has the following syntax:

$$\text{sum } i : D . E$$

where i is a quantifier variable, D is a domain, and E is the expression contained within the `sum`. More than one quantifier variable may be created by writing a comma-separated list i, j, k .

For example, if we want to take the sum of all numbers in the range 1 to 10 we write

```
sum i : int(1..n) . i
```

which corresponds to $\sum_{i=1}^n i$. n cannot be a decision variable.

Quantified sum has several similarities to the `forall` and `exists` quantifiers (described below in Section 2.1.16): it introduces new local variables (named quantifier variables) that can be used within E , and the quantifier variables all have the same domain D . However `sum` has one important difference: a `sum` is an integer expression.

A quantified sum may be nested inside any other integer operator, including another quantified sum:

```
sum i, j : int(1..10) .
  sum k : int(i..10) .
    x[i, j] * k
```

2.1.16 Universal and Existential Quantification

Universal and existential quantification are powerful means to write down a series of constraints in a compact way. Quantifications have the same syntax as `sum`, but with `forall` and `exists` as keywords:

```
forall i : D . E
exists i : D . E
```

For instance, the universal quantification

```
forall i : int(1..3) . x[i] = i
```

corresponds to the conjunction:

```
x[1] = 1 /\ x[2] = 2 /\ x[3] = 3
```

An example of existential quantification is

```
exists i : int(1..3) . x[i] = i
```

and it corresponds to the following disjunction:

```
x[1] = 1 \/ x[2] = 2 \/ x[3] = 3
```

Quantifications can range over several quantified variables and can be arbitrarily nested, as demonstrated with the `sum` quantifier.

In the running Sudoku example, `forall` quantification is used to build the set of constraints. The expression:

```
forall row : int(1..9) .
  forall col1 : int(1..9) .
    forall col2 : int(col1+1..9) . M[row, col1] != M[row, col2]
```

is a typical use of universal quantification.

2.1.17 Quantification over Matrix Domains

All three quantifiers are defined on matrix domains as well as integer and boolean domains. For example, to quantify over all permutations of size n :

```
forall perm : matrix indexed by [int(1..n)] of int(1..n) .
    allDiff(perm) -> exp
```

The variable `perm` represents a matrix drawn from the matrix domain, and the `allDiff` constraint evaluates to `true` when `perm` is a permutation of $1 \dots n$. Hence the expression `exp` is quantified for all permutations of $1 \dots n$.

If n is a constant, the example above could be written as a set of n nested `forall` quantifiers. However if n is a parameter of the problem class, it is very difficult to write the example above using other (non-matrix) quantifiers.

2.1.18 Global Constraints

ESSENCE PRIME provides a set of global constraints such as `allDiff` (which is satisfied when a vector of variables each take different values). Global constraints are all boolean expressions in ESSENCE PRIME. Typically it is worth using these in models because the solver often performs better with a global constraint compared to a set of simpler constraints.

For example, the following two lines are semantically equivalent (assuming `x` is a matrix indexed by $1 \dots n$).

```
forall i, j : int(1..n) . i < j -> x[i] != x[j]
allDiff(x)
```

Both lines will ensure that the variables in `x` take different values. However the `allDiff` will perform better in most situations.²

Tables 1 and 2 summarise the global constraints available in ESSENCE PRIME.

To give another example, the `cumulative` constraint for scheduling applies a bound on the amount of a resource that may be used at each time step in a schedule. It has arguments `X`, `Dur`, `Res`, and `Bound` (see Table 2) representing the start time of each task, the duration of each task, the resource required for each task while it is running, and the upper limit on the resource usage. Suppose there are three tasks, each task takes 3 time units to execute, each task consumes 2 units of resource when it is running, and the resource limit is 5:

```
Res=[2, 2, 2]
Bound=5
Dur=[3, 3, 3]
```

If all three tasks are running at any time step, they would exceed the resource bound. $X=[0, 1, 2]$ is not a solution, but $X=[0, 1, 3]$ is (the first task finishes before the third task starts).

Now we have the `allDiff` global constraint, the Sudoku example can be improved and simplified as shown in Figure 3.

Global constraints are boolean expressions like any other, and are allowed to be used in any context that accepts a boolean expression.

2.1.19 Table Constraints

In a table constraint the satisfying tuples of the constraint are specified using a matrix. This allows a table constraint to theoretically implement any relation, although practically it is limited to relations where the set of satisfying tuples is small enough to store in memory and efficiently search.

The first argument specifies the variables in the scope of the constraint, and the second argument is a two-dimensional matrix of satisfying tuples. For example, the constraint $a+b=c$ on boolean variables could be written as a table as follows.

²In the current version of Savile Row, with default settings, the `x[i] != x[j]` constraints would be aggregated to create `allDiff(x)` therefore there is no difference in performance between these two statements. There would be a difference in performance when aggregation is switched off (for example by using the `-O1` flag).

Global Constraint	Arguments	Description
<code>allDiff(X)</code>	<code>X</code> is a matrix	Ensures expressions in <code>X</code> take distinct values in any solution.
<code>atleast(X, C, Vals)</code>	<code>X</code> is a matrix, <code>Vals</code> is a matrix of non-decision expressions, <code>C</code> is a matrix of non-decision expressions	For each non-decision expression <code>Vals[i]</code> , the number of occurrences of <code>Vals[i]</code> in <code>X</code> is at least <code>C[i]</code> .
<code>atmost(X, C, Vals)</code>	<code>X</code> is a matrix, <code>Vals</code> is a matrix of non-decision expressions, <code>C</code> is a matrix of non-decision expressions	For each non-decision expression <code>Vals[i]</code> , the number of occurrences of <code>Vals[i]</code> in <code>X</code> is at most <code>C[i]</code> .
<code>gcc(X, Vals, C)</code>	<code>X</code> is a matrix, <code>Vals</code> is a matrix of non-decision expressions, <code>C</code> is a matrix	For each non-decision expression <code>Vals[i]</code> , the number of occurrences of <code>Vals[i]</code> in <code>X</code> equals <code>C[i]</code> .
<code>alldifferent_except(X, Value)</code>	<code>X</code> is a matrix, <code>Value</code> is a non-decision expression	Ensures variables in <code>X</code> take distinct values, except that <code>Value</code> may occur any number of times.
<code>inverse(X, Y)</code>	<code>X</code> is a matrix, <code>Y</code> is a matrix	Interpreting <code>X</code> and <code>Y</code> as functions, <code>inverse(X, Y)</code> ensures that <code>Y</code> is the inverse of <code>X</code> . For all relevant values <code>a, b</code> : $X[a]=b \leftrightarrow Y[b]=a$. The constraint <code>inverse(X, Y)</code> implies <code>allDiff(x)</code> and <code>allDiff(y)</code> .
<code>circuit(X)</code>	<code>X</code> is a matrix	Ensures that <code>X</code> represents a hamiltonian circuit, where <code>X[i]</code> is the successor of <code>i</code> . The constraint <code>circuit(X)</code> implies <code>allDiff(X)</code> , and $X[i] \neq i$ for all valid indices <code>i</code> .

Table 1: Global constraints in ESSENCE PRIME (except scheduling and packing constraints). Each may be nested within expressions and have arbitrary expressions nested within them. Each matrix parameter of a global constraint must be a one-dimensional matrix. In some cases the parameter is a matrix of *non-decision expressions* – that is, integer or boolean expressions that contain no decision variables. Quantifier and parameter variables are allowed in non-decision expressions.

cumulative (X, Dur, Res, Bound)	X, Dur, and Res are matrices, Bound is a scalar.	The cumulative constraint enforces a resource limit on a set of tasks. The start time of a task i is $X[i]$, the duration of i is $Dur[i]$, and the resource required by task i is $Res[i]$. At each time step, the sum of the resource in use (i.e. the sum of the resource required by tasks that are running at that time step) cannot be more than Bound.
disjunctive(X, Dur)	X and Dur are matrices.	The disjunctive constraint enforces non-overlap of a set of tasks. The start time of a task i is $X[i]$, the duration of i is $Dur[i]$. Tasks with duration 0 may be scheduled at any time, including when another task is running.
diffn(X, Y, dx, dy)	X, Y, dx, dy are matrices.	The diffn constraint enforces non-overlap of a set of rectangles. Each rectangle i is defined by its corner co-ordinates $X[i], Y[i]$ and dimensions $dx[i], dy[i]$. Rectangles of size 0 (in either dimension or both) still cannot overlap with any other rectangle.

Table 2: Global scheduling and packing constraints in ESSENCE PRIME. Each may be nested within expressions and have arbitrary expressions nested within them. Each matrix parameter of a global constraint must be a one-dimensional matrix. In some cases the parameter is a matrix of *non-decision expressions* – that is, integer or boolean expressions that contain no decision variables. Quantifier and parameter variables are allowed in non-decision expressions.

```

language ESSENCE' 1.0

find M : matrix indexed by [int(1..9), int(1..9)] of int(1..9)

such that

M[1,1]=5,
M[1,2]=3,
M[1,5]=7,
....
M[9,9]=9,

forall row : int(1..9) .
    allDiff(M[row,..]),

forall col : int(1..9) .
    allDiff(M[..,col])

```

Figure 3: Second version of Sudoku, using global constraints.

```
table( [a,b,c], [[0,0,0], [0,1,1], [1,0,1]] )
```

The first argument of `table` is a one-dimensional matrix expression. It may contain both decision variables and constants. The second argument is a two-dimensional matrix expression containing no decision variables. The second argument can be stated as a matrix literal, or an identifier, or by slicing a higher-dimension matrix, or by constructing a matrix using matrix comprehensions (see Section 2.1.20).

If the same matrix of tuples is used for many table constraints, a `letting` statement can be used to state the matrix once and use it many times. Lettings are described in Section 2.2.2 below.

2.1.20 Matrix Comprehensions

We have seen that matrices may be written explicitly as a matrix literal (Section 2.1.3), and that existing matrices can be indexed and sliced (Section 2.1.7). Matrices can also be constructed using *matrix comprehensions*. This provides a very flexible way to create matrices of variables or values. A single matrix comprehension creates a one-dimensional matrix, however they can be nested to create multi-dimensional matrices. There are two syntactic forms of matrix comprehension:

```
[ exp | i : domain1, j : domain2, cond1, cond2 ]
[ exp | i : domain1, j : domain2, cond1, cond2 ; indexedomain ]
```

where `exp` is any integer, boolean or matrix expression. This is followed by any number of comprehension variables, each with a domain. After the comprehension variables we have an optional list of conditions: these are boolean expressions that constrain the values of the comprehension variables. Finally there is an optional index domain. This provides an index domain to the constructed matrix.

To expand the comprehension, each assignment to the comprehension variables that satisfies the conditions is enumerated in lexicographic order. For each such assignment, the values of the comprehension variables are substituted into `exp`. The resulting expression then becomes one element of the constructed matrix.

The simplest matrix comprehensions have only one comprehension variable, as in the example below.

```
[ num**2 | num : int(1..5) ] = [ 1,4,9,16,25 ; int(1..5) ]
```

The matrix constructed by a comprehension is one-dimensional and is either indexed from 1 contiguously, or has the given index domain. The given domain must have a lower bound but is allowed to have no upper bound. For example the first line below produces the matrix on the second line.

```
[ i+j | i : int(1..3), j : int(1..3), i<j ; int(7..) ]
[ 3, 4, 5 ; int(7..9) ]
```

Matrix comprehensions allow more advanced forms of slicing than the matrix slice syntax in Section 2.1.7. For example it is possible to slice an arbitrary subset of the rows or columns of a two-dimensional matrix. The following two nested comprehensions will slice out the entries of a matrix `M` where both rows and columns are odd-numbered, and build a new two-dimensional matrix.

```
[ [ M[i,j] | j : int(1..n), j%2=1 ] | i : int(1..n), i%2=1 ]
```

Now we have matrix comprehensions, the Sudoku example can be completed by adding the constraints on the 3×3 subsquares (as shown in Figure 4). A comprehension is used to create a matrix of variables and the matrix is used as the parameter of an `allDiff` constraint.

In this example, the matrix constructed by the comprehension depends on the values of `i` and `j` from the `forall` quantifier. The comprehension variables `k` and `l` each take one of three values, to cover the 9 entries `M[k,l]` in the subsquare.

2.1.21 Matrix Comprehensions over Matrix Domains

Similarly to quantifiers, matrix comprehension variables can have a matrix domain. For example, the following comprehension builds a two-dimensional matrix where the rows are all permutations of $1..n$.

```
[ perm | perm : matrix indexed by [int(1..n)] of int(1..n), allDiff(perm) ]
```

```

language ESSENCE' 1.0

find M : matrix indexed by [int(1..9), int(1..9)] of int(1..9)

such that

M[1,1]=5,
M[1,2]=3,
M[1,5]=7,
....
M[9,9]=9,

forall row : int(1..9) .
    allDiff(M[row,..]),

forall col : int(1..9) .
    allDiff(M[..,col]),

$ all 3x3 subsquare have to be all-different
$ i,j indicate the top-left corner of the subsquare.
forall i,j : int(1,4,7) .
    allDiff([ M[k,l] | k : int(i..i+2), l : int(j..j+2)])

```

Figure 4: Third version of Sudoku, using global constraints, a matrix comprehension and quantifiers.

2.1.22 Functions on Matrices

Table 3 lists the matrix functions available in ESSENCE PRIME.

The functions `sum`, `product`, and `and or` were originally intended to be used with matrix comprehensions, but can be used with any matrix. The quantifiers `sum`, `forall` and `exists` can be replaced with `sum`, `and` and `or` containing matrix comprehensions. For example, consider the `forall` expression below (from the Sudoku model). It can be replaced with the second line below.

```

forall row : int(1..9) . allDiff(M[row,..])

and([ allDiff(M[row,..]) | row : int(1..9) ])

```

In fact, matrix functions combined with matrix comprehensions are strictly more powerful than quantifiers. Also, the function `product` has no corresponding quantifier. Quantifiers are retained in the language because they can be easier to read.

As a more advanced example, given an $n \times n$ matrix `M` of decision variables, the sum below is the determinant of `M` using the Leibniz formula. The outermost comprehension constructs all permutations of $1 \dots n$ using a matrix domain and an `allDiff`. Lines 3 and 4 contain a comprehension that is used to obtain the number of *inversions* of `perm` (the number of pairs of values that are not in ascending order). Finally line 5 builds a product of some of the entries of the matrix. Without the `product` function, it is difficult (perhaps impossible) to write the Leibniz formula in ESSENCE PRIME for a matrix of unknown size n .

```

sum([
    $ calculate the sign of perm from the number of inversions.
    ( (-1) ** sum([ perm[idx1]>perm[idx2]
                    | idx1 : int(1..n), idx2 : int(1..n), idx1<idx2 ]) ) *

```

Function	Arguments	Description
<code>sum(X)</code>	X is a one-dimensional matrix	Constructs the sum of elements in X
<code>product(X)</code>	X is a one-dimensional matrix	Constructs the product of elements in X
<code>and(X)</code>	X is a one-dimensional matrix of booleans	Constructs the conjunction of X
<code>or(X)</code>	X is a one-dimensional matrix of booleans	Constructs the disjunction of X
<code>min(X)</code>	X is a one-dimensional matrix	The integer minimum of elements in X
<code>max(X)</code>	X is a one-dimensional matrix	The integer maximum of elements in X
<code>flatten(X)</code>	X is a matrix	Constructs a one-dimensional matrix (indexed contiguously from 1) with the same contents as X
<code>flatten(n, X)</code>	X is a matrix	The first n+1 dimensions of X are flattened into one dimension that is indexed contiguously from 1. Therefore <code>flatten(n, X)</code> produces a new matrix with n fewer dimensions than X. The first argument n must be positive.
<code>cat(A, B, C, ...)</code>	A, B, C etc are matrices, each with the same number of dimensions.	Concatenate the given matrices into a single matrix by combining their first dimension. For example, if <code>A=[[1,2],[3,4]]</code> and <code>B=[[5,6]]</code> , then <code>cat(A,B)=[[1,2],[3,4],[5,6]]</code> . The first dimension of the resulting matrix is always indexed contiguously from 1.
<code>list(A, B, C, ...)</code>	A, B, C etc are scalars or matrices of type integer or Boolean.	Constructs a one-dimensional matrix (indexed contiguously from 1) with the same contents as A, B, C, ... Each matrix in the list of arguments A, B, C, ... is flattened (as in <code>flatten</code>) and each scalar becomes a one-dimensional matrix with a single element. The resulting matrices are concatenated into a single one-dimensional matrix (as in <code>cat</code>).
<code>toSet(X)</code>	X is a one-dimensional matrix of non-decision expressions	The set of elements in X
<code>indexOf(X)</code>	X is a one-dimensional matrix	The index set of the outer dimension of X. Useful for quantifying over a matrix of unknown size.

Table 3: Matrix Functions

```
product([ M[j, perm[j]] | j : int(1..n) ])
| perm : matrix indexed by [int(1..n)] of int(1..n), allDiff(perm)]])
```

The `flatten` function is typically used to feed the contents of a multi-dimensional matrix expression into a constraint that accepts only one-dimensional matrices. For example, given a three-dimensional matrix `M`, the following example is a typical use of `flatten`.

```
allDiff( flatten(M[1,..,..]) )
```

When flattening a matrix `M` to create a new matrix `F`, the order of elements in `F` is defined as follows. Suppose `M` were written as a matrix literal (as in Section 2.1.3) the order elements are written in the matrix literal is the order the elements appear in `F`. The following example illustrates this for a three-dimensional matrix.

```
flatten([ [ [1,2], [3,4] ], [ [5,6], [7,8] ] ]) = [1,2,3,4,5,6,7,8]
```

Finally, the `indexOf` function is useful for quantifying over a matrix (or part of a matrix, such as one row) of unknown size. For example, the following snippet uses `indexOf` to access the index domains of the rows of a matrix where rows have different lengths.

```
letting M = [ [1,2,3], [4,5], [6] ]
find X : matrix [int(1..6)] of int(1..6)
such that
forall i : int(1..3).
    forall j : indexOf(M[i,..]).
        X[M[i,j]] = M[i,j]
```

2.2 Model Structure

An ESSENCE PRIME model is structured in the following way:

1. Header with version number: `language ESSENCE' 1.0`
2. Parameter declarations (optional)
3. Constant definitions (optional)
4. Decision variable declarations (optional)
5. Constraints on parameter values (optional)
6. Objective (optional)
7. Solver Control (optional)
8. Constraints

Parameter declarations, constant definitions, decision variable declarations, and constraints on parameter values can be interleaved, but for readability we suggest to put them in the order given above. Comments are preceded by '\$' and run to the end of the line.

Parameter values are defined in a separate file, the *parameter file*. Parameter files have the same header as problem models and hold a list of parameter definitions. Table 4 gives an overview of the model structure of problem and parameter files. Each model part will be discussed in more detail in the following sections.

Problem Model Structure	Parameter File Structure
language ESSENCE' 1.0	language ESSENCE' 1.0
<i>\$ parameter declaration</i> given n : int	<i>\$ parameter instantiation</i> letting n=7
<i>\$ constant definition</i> letting c=5	
<i>\$ variable declaration</i> find x,y : int(1..n)	
<i>\$ constraints</i> such that x + y >= c, x + c*y = 0	

Table 4: Model Structure of problem files and parameter files in ESSENCE PRIME. '\$' denotes comments.

2.2.1 Parameter Declarations with **given**

Parameters are declared with the `given` keyword followed by a domain the parameter ranges over. Parameters are allowed to range over the infinite domain `int`, or domains that contain an open range such as `int(1..)` and `int(..10)`. For example,

```
given n : int(0..)
```

```
given d : int(0..n)
```

declares two parameters, and the domain of the second depends on the value of the first. Parameters may have integer, boolean or matrix domains.

Now we have parameters we can generalise the Sudoku model of Figure 3 to represent the problem class of all Sudoku puzzles. The generalised model is shown in Figure 5. The parameter is the `clues` matrix, where blank spaces are represented as 0 and non-zero entries in `clues` are copied to `M`.

2.2.2 Constant Definitions with **letting**

In most problem models there are re-occurring constant values and it can be useful to define them as constants. The `letting` statement allows to assign a name with a constant value. The statement

```
letting NAME = A
```

introduces a new identifier `NAME` that is associated with the expression `A`. Every subsequent occurrence of `NAME` in the model is replaced by the value of `A`. Please note that `NAME` cannot be used in the model *before* it has been defined. `A` may be any integer, boolean or matrix expression that does not contain decision variables. Some integer examples are shown below.

```
given n : int(0..)
letting c = 10
letting d = c*n*2
```

Here the second integer constant depends on the first. As well as integer and boolean expressions, lettings may contain matrix expressions, as in the example below. When using a matrix literal the domain is optional – the two lettings below are equivalent. The version with the matrix domain may be useful when the matrix is not indexed from 1.

```

language ESSENCE' 1.0

given clues : matrix indexed by [int(1..9), int(1..9)] of int(0..9)

find M : matrix indexed by [int(1..9), int(1..9)] of int(1..9)

such that

forall row : int(1..9) .
  forall col : int(1..9) .
    (clues[row, col] != 0) -> (M[row, col] = clues[row, col]),

forall row : int(1..9) .
  allDiff(M[row, ..]),

forall col : int(1..9) .
  allDiff(M[.., col]),

$ all 3x3 subsquare have to be all-different
$ i,j indicate the top-left corner of the subsquare.
forall i,j : int(1,4,7) .
  allDiff([ M[k,l] | k : int(i..i+2), l : int(j..j+2)])

```

Figure 5: Fourth version of Sudoku, adding the `clues` parameter to the version in Figure 4.

```

letting cmatrix = [ [2,8,5,1], [3,7,9,4] ]
letting cmatrix2 : matrix indexed by [ int(1..2), int(1..4) ] of int(1..10)
  = [ [2,8,5,1], [3,7,9,4] ]

```

Finally new matrices may be constructed using a slice or comprehension, as in the example below where the `letting` is used for the table of a table constraint.

```

letting xor_table = [ [a,b,c] | a : bool, b : bool, c : bool,
                      (a /\ b) \/ (!a /\ !b) <-> c ]

find x,y,z : bool
such that
table([x,y,z], xor_table)

```

2.2.3 Constant Domains

Constant domains are defined in a similar way using the keywords `be` `domain`.

```

letting c = 10
given n : int(1..)
letting INDEX be domain int(1..c*n)

```

In this example `INDEX` is defined to be an integer domain, the upper bound of which depends on a parameter `n` and another constant `c`.

Constant domains are convenient when a domain is reused several times. In the Sudoku running example, we could use a `letting` for the domain `int(1..9)`, as shown in Figure 6.

```

language ESSENCE' 1.0
letting range be domain int(1..9)

given clues : matrix indexed by [range, range] of int(0..9)

find M : matrix indexed by [range, range] of range

such that

forall row : range .
  forall col : range .
    (clues[row, col] != 0) -> (M[row, col] = clues[row, col]),

forall row : range .
  allDiff(M[row, ..]),

forall col : range .
  allDiff(M[.., col]),

$ all 3x3 subsquare have to be all-different
$ i, j indicate the top-left corner of the subsquare.
forall i, j : int(1, 4, 7) .
  allDiff([ M[k, l] | k : int(i..i+2), l : int(j..j+2)])

```

Figure 6: Fifth version of the Sudoku model. Compared to Figure 5 the `range` letting statement has been added.

2.2.4 Decision Variable Declaration with `find`

Decision variables are declared using the `find` keyword followed by a name and their corresponding domain. The domain must be finite. The example below

```
find x : int(1..10)
```

defines a single decision variable with the given domain. It is possible to define several variables in one `find` by giving multiple names, as follows.

```
find x, y, z : int(1..10)
```

Matrices of decision variables are declared using a matrix domain, as in the following example.

```
find m : matrix indexed by [ int(1..10) ] of bool
```

This declares `m` as a 1-dimensional matrix of 10 boolean variables. Simple and matrix domains are described in Section 2.1.1.

In the Sudoku running example, we have been using the following two-dimensional matrix domain.

```
find M : matrix indexed by [int(1..9), int(1..9)] of int(1..9)
```

2.3 Constraints on Parameters with `where`

In some cases it is useful to restrict the values of the parameters. This is achieved with the `where` keyword, which is followed by a boolean expression containing no decision variables. In the following example, we require the first parameter to be less than the second.

```
given x : int(1..)
given y : int(1..)
where x < y
```

2.4 Objective

The objective of a problem is either maximising or minimising an integer or boolean expression. For instance,

```
minimising x
```

states that the value assigned to variable x will be minimised. Only one objective is allowed, and it is placed after all `given`, `find` and `letting` statements.

2.5 Solver Control

In addition to instructing the solver to minimise or maximise some expression, ESSENCE PRIME also allows a variable order to be specified for search. The `branching on` statement specifies a sequence of variables for the solver to branch on. Each element of the `branching on` list may be an individual variable or a matrix with any number of dimensions. Decision variables in matrices are enumerated in the order produced by the `flatten` function. Any matrix expression is allowed within the `branching on` list, including matrix comprehensions. A simple use of `branching on` would be to list some individual variables:

```
find w, x, y, z : int(1..10)

branching on [w, x]
```

The `branching on` list may contain the same decision variable more than once. This can be useful to pick some variables from a matrix to branch on first, then include the rest of the matrix simply by including the entire matrix. In the following example the diagonal of M is branched first, then the rest of M is included.

```
find M : matrix indexed by [int(1..9), int(1..9)] of int(1..9)

branching on [ [ M[i,i] | i : int(1..9) ], M ]
```

In some cases a good search order does not follow the natural order of matrices, and might even alternate between matrices. The following example uses a comprehension to alternate between two matrices that are both indexed by `timestep`:

```
find arm : matrix indexed by [int(1..timesteps)] of int(1..10)
find action : matrix indexed by [int(1..timesteps)] of int(1..10)

branching on [ [ [arm[i], action[i]] | i : int(1..timesteps) ] ]
```

Some solver types (e.g. SAT, MaxSAT, and SMT solvers) ignore the search order, while Chuffed (with its free search option) uses the search order as part of a hybrid search.

When searching for all solutions (`-all-solutions`) or a given number of solutions (`-num-solutions`) with a `branching on` statement, each solution produced will have a distinct assignment to the `branching on` variables, thus using a `branching on` statement can reduce the number of solutions found.

When using Minion as the backend solver, a `heuristic` statement may be used in the preamble to specify a dynamic search heuristic to apply to the variables in the `branching on` list. `heuristic` is followed by `static` (the default), `sdf`, `conflict` or `srf` (refer to the Minion documentation for a description of these heuristics). If any other solver is used then the `heuristic` statement is ignored.

2.6 Constraints

After defining constants and declaring decision variables and parameters, constraints are specified with the keywords `such that`. The constraints in ESSENCE PRIME are boolean expressions as described in Section 2.1.8.

Typically the constraints are written as a list of boolean expressions separated by the `,` operator.

3 Undefinedness in ESSENCE PRIME

Since the current version of ESSENCE PRIME is a closed language, there are a finite set of partial functions in the language. For example, x/y is a partial function because it is not defined when $y=0$. In its current version ESSENCE PRIME implements the relational semantics as defined by Frisch and Stuckey [9]. The relational semantics has the advantage that it can be implemented efficiently.

The relational semantics may be summarised as follows:

- Any integer or matrix expression directly containing an undefined expression is itself undefined.
- Any domain or domain expression directly containing an undefined expression is itself undefined.
- Any statement in the preamble (`find`, `letting` etc) directly containing an undefined expression is undefined.
- Any boolean expression that directly contains an undefined expression is `false`.

Informally, the relational semantics confines the effect of an undefined expression to a small part of the problem instance (which becomes `false`), in many cases avoiding making the entire problem instance `false`.

Consider the four examples below. Each contains a division by zero which is an undefined integer expression. In each case the division is contained in a comparison. Integer comparisons are boolean expressions.

```
(x/0 = y) = false
(x/0 != y) = false
! (x/0 = y) = true
! (x/0 != y) = true
```

Applying the rules of the relational semantics results in each of the comparisons inside the brackets becoming `false`:

```
(false) = false
(false) = false
! (false) = true
! (false) = true
```

In the relational semantics, $(x/0 \neq y)$ is not semantically equivalent to $!(x/0 = y)$, which is somewhat counter-intuitive.

Another counter-intuitive case arises with matrix indexing. In the following example, the expression $M[0]$ is undefined because 0 is not in the index domain. If the matrix is boolean (i.e. `DOM` is `bool`) then $M[0]$ becomes `false`, and the model has a solution when $M[1]=\text{false}$. However, if the matrix contains integer variables (i.e. `DOM` is `int(0..1)`) then the constraint $M[0] = M[1]$ becomes `false` and the model has no solutions.

```
find M : matrix indexed by [int(1)] of DOM
such that
M[0] = M[1]
```

In the SAVILE ROW implementation of ESSENCE PRIME, all partial functions are removed in a two-step process, before any other transformations are applied. The first step is as follows. For each partial function a boolean expression is created that is `true` when the partial function is defined and `false` when it is undefined. There are six operators that may be partial: division, modulo, power, factorial, matrix indexing and matrix slicing. Table 5 shows the generated

Partial Function	Defined When
X/Y	$Y \neq 0$
$X \% Y$	$Y \neq 0$
$X ** Y$	$(X \neq 0 \ \wedge \ Y \neq 0) \ \wedge \ Y \geq 0$
<code>factorial(X)</code>	$X \geq 0$
Matrix indexing: $M[I, J, K]$ where domain of M is matrix indexed by $[D1, D2, D3]$ of $DBase$	$I \in D1 \ \wedge$ $J \in D2 \ \wedge$ $K \in D3$
Matrix slicing: $M[I, \dots, K]$ where domain of M is matrix indexed by $[D1, D2, D3]$ of $DBase$	$I \in D1 \ \wedge$ $K \in D3$

Table 5: Partial functions in ESSENCE PRIME. X , Y , I , J and K are arbitrary expressions of the correct type.

boolean expression for each operator. The boolean expression is then added to the model by connecting it (with \wedge) onto the closest boolean expression above the partial function in the abstract syntax tree.

The second step is to replace the partial function P with a total function SP . For each input where P is defined, SP is defined to the same value. For inputs where P is undefined, SP takes a default value (typically 0 for integer expressions).

Once both steps have been applied to each partial function, the model is well defined everywhere. This is done first, before any other model transformations, and thus allows all subsequent transformations to be simpler because there is no need to allow for undefinedness. For example, the expression $x/y = x/y$ may not be simplified to `true`, because it is `false` when $y=0$. However, after replacing the partial division function, the resulting equality can be simplified to `true`. In general any equality between two syntactically identical expressions can be simplified to `true` once there are no partial functions.

4 Installing and Running SAVILE ROW

SAVILE ROW is distributed as an archive with the following contents:

- The Java source code (in `src`) licensed with GPL 3.
- The compiled classes in a JAR file named `savilerow.jar`.
- Executable script `savilerow` for running SAVILE ROW.
- This document in `doc`.
- Various example ESSENCE PRIME files and parameter files, in `examples`.
- Required Java libraries in `lib`
- Executables for backend solvers Minion, Kissat, and Chuffed, and the graph automorphism solver Ferret in `bin`
- Build script `compile.sh` to rebuild `savilerow.jar`³.

Archives are provided for Linux and Mac, and the Linux archive works well on Windows 10 in the Ubuntu app. The two distributions are largely the same, with the only difference being the packaged executables in `bin`.

A recent version of Java is required on all platforms. SAVILE ROW was compiled to run on Java 8 or later.

³Requires a JDK to be installed and `javac` to be in the path. It compiles for Java 8 backward compatibility by using the `--release 8` flag. If compiling using Java 8 the `--release 8` flag needs to be removed from the `javac` command-line.

4.1 Running SAVILE ROW on Linux and Mac

Download the appropriate archive and unpack it somewhere convenient. Open a terminal and navigate to the SAVILE ROW directory. Use the script named `savilerow`. One of the simplest ways of running SAVILE ROW is given below.

```
./savilerow problemFile.eprime parameterFile.param
```

The first argument is the problem class file. This is a plain text file containing the constraint problem, expressed in the ESSENCE PRIME language.

The second argument (`parameterFile.param`) is the parameter file (again in the ESSENCE PRIME language). This contains the data for the problem instance. The parameter file can be omitted if the problem has no parameters (no given statements in the preamble).

4.2 Running SAVILE ROW on Windows

The current (1.11.1) version of SAVILE ROW does not have a release for Windows specifically, however the Linux release works well in the Ubuntu 20.04 app on Windows 10. To install, first enable the Windows Subsystem for Linux. Second, install the Ubuntu 20.04 app and update packages in it (for example with `sudo apt update`; `sudo apt upgrade`). Third, install Java (for example, `sudo apt install default-jre`). Then follow the instructions for installing and running on Linux.

4.3 Solution Files

For the Minion, SAT, SMT, MaxSAT, Standard FlatZinc, Gecode, and Chuffed backends, SAVILE ROW is able to run the solver and parse the solution (or set of solutions) produced by the solver. These solutions are translated back into ESSENCE PRIME. For each `find` statement in the model file (i.e. each statement that declares decision variables), the solution file contains a matching `letting` statement. For example, if the model file contains the following `find` statement:

```
find M : matrix indexed by [int(1..2), int(1..2)] of int(1..5)
```

The solution file could contain the following `letting` statement.

```
letting M = [[2,3],
             [1,2]]
```

4.4 File Names

The input files for SAVILE ROW typically have the extension `.eprime`, `.param`, or `.eprime-param`. These extensions allow SAVILE ROW to identify the model and parameter file on the command line. If these files have a different extension (or no extension) then SAVILE ROW must be called in a slightly different way:

```
./savilerow -in-eprime problemFile -in-param parameterFile
```

Given input file names `<problemFile>` and `<parameterFile>`, output files have the following names by default.

- For Minion output, `<parameterFile>.minion` (or if there is no parameter file, `<problemFile>.minion`).
- For Gecode or Chuffed output, `<parameterFile>.fzn` (or if there is no parameter file, `<problemFile>.fzn`).
- For SAT or MaxSAT output, `<parameterFile>.dimacs` (or if there is no parameter file, `<problemFile>.dimacs`).
- For Minizinc output, `<parameterFile>.mzn` (or if there is no parameter file, `<problemFile>.mzn`).

When SAVILE ROW parses a single solution from the output of a solver, it produces a file named `<parameterFile>.solution` (or if there is no parameter file, `<problemFile>.solution`). When there are multiple solutions (e.g. when using the `-all-solutions` flag) the solution files are numbered (for example, `nurses.param.solution.000001` to `nurses.param.solution.000871`).

If SAVILE ROW runs Minion or a SAT solver it produces a file `<parameterFile>.info` (or if there is no parameter file, `<problemFile>.info`) containing solver statistics.

Finally, a file named `<parameterFile>.aux` (or `<problemFile>.aux`) is also created. This contains the symbol table and is read if SAVILE ROW is called a second time to parse a solution.

4.5 Command Line Options

The command-line help text is produced with the following flag.

```
-help
```

4.5.1 Specifying input files

The options `-in-eprime` and `-in-param` specify the input files as in the example below.

```
savilerow -in-eprime sonet.eprime -in-param sonet1.param
```

These flags may be omitted for any filename that ends with `.eprime`, `.param`, or `.eprime-param`.

The option `-params` may be used to specify the parameters on the command line. For example, suppose the `nurse.eprime` model has two parameters. We can specify them on the command line as follows. The format of the parameter string is identical to the format of a parameter file (where, incidentally, the `language` line is optional and line breaks are not significant).

```
savilerow nurse.eprime -params "letting n_nurses=4 letting Demand=[[1,0,1,0],[0,2,1,0]]"
```

4.5.2 Specifying output format

SAVILE ROW produces output for various solvers and classes of solver as described in Section 1.2. The output format is specified by using one of the following command-line options.

```
-minion
-gecode
-chuffed
-or-tools
-flatzinc
-minizinc
-sat
-smt
-maxsat
```

The output filename may be specified as follows. In each case there is a default filename so the flag is optional. Default filenames are described in Section 4.4. The flag `-out-sat` also sets the MaxSAT filename.

```
-out-minion <filename>
-out-flatzinc <filename>
-out-minizinc <filename>
-out-sat <filename>
-out-smt <filename>
```

The flag `-out-flatzinc` sets the FlatZinc filename.

In addition, the file names for solution files, solver statistics files and symbol table files (called aux files) may be specified as follows. Once again there are default filenames described in Section 4.4.

```
-out-solution <filename>
-out-info <filename>
-out-aux <filename>
```

The following flag sets each of the output files to `prefix.extension` where `extension` is the conventional one for the file type (i.e. `.minion`, `.fzn`, `.dimacs`, etc).

```
-out-prefix <prefix>
```

The symbol table is not saved by default. To obtain the symbol table file, use the following flag. The symbol table is used in ReadSolution mode, described below.

```
-save-symbols
```

4.5.3 Optimisation Levels

The optimisation levels (`-O0` to `-O3`) provide an easy way to control how much optimisation SAVILE ROW does, without having to switch on or off individual optimisations. The default is `-O2`, which is intended to provide a generally recommended set of optimisations. The rightmost `-O` flag on the command line is the one that takes precedence.

`-O0`

The lowest optimisation level, `-O0`, turns off all optional optimisations. SAVILE ROW will still simplify expressions (including constraints). Any expression containing only constants will be replaced with its value. Some expressions have quite sophisticated simplifiers that will run even at `-O0`. For example, `allDiff([x+y+z, z+y+x, p, q])` would simplify to `false` because the first two expressions in the `allDiff` are symbolically equal after normalisation.

`-O0` will do no common subexpression elimination, will not unify equal variables, and will not filter the domains of variables.

`-O1`

`-O1` does optimisations that are very efficient in both space and time. Variables that are equal are unified, and a form of common subexpression elimination is applied (Active CSE, described below). `-O1` is equivalent to `-deletevars` and `-active-cse`.

`-O2`

In addition to the optimisations performed by `-O1`, `-O2` performs filtering of variable domains (both `find` and auxiliary variables) and aggregation (both of which are described in the following section). `-O2` is the default optimisation level. `-O2` adds `-reduce-domains-extend` and `-aggregate` compared to `-O1`.

`-O3`

In addition to `-O2`, `-O3` enables TabID (`-tabid-light`) and associative-commutative common subexpression elimination (`-ac-cse`) (both described below).

4.5.4 Symmetry-Breaking Levels

The options `-S0`, `-S1` and `-S2` control optional optimisations that can change the number of solutions. The default is `-S1` which applies only very simple and fast optimisations.

`-S0`

Preserve the number of solutions, i.e. perform no optimisations that might change the number of solutions.

`-S1`

Remove variables that appear in no constraints, and allow SAVILE ROW to introduce auxiliary variables that are not functionally defined by the primary ('find') variables. At present, non-functional auxiliary variables are only used with the SAT, MaxSAT and SMT backends. Equivalent to `-remove-redundant-vars` and `-aux-non-functional`.

`-S2`

In addition to `-S1`, apply a graph automorphism solver to find symmetries among the decision variables. The symmetries are then broken using the standard lex-leader method. Equivalent to adding the flag `-var-sym-breaking` to `-S1`.

4.5.5 Translation Options

Common Subexpression Elimination

SAVILE ROW currently implements four types of common subexpression elimination (CSE). Identical CSE finds and eliminates syntactically identical expressions. This is the simplest form of CSE, however it can be an effective optimisation. Active CSE performs some reformulations (for example applying De Morgan's laws) to reveal expressions that are semantically equivalent but not syntactically identical. Active CSE subsumes Identical CSE. Active CSE is enabled by default as part of `-O2`. Associative-Commutative CSE (AC-CSE) works on the associative and commutative (AC) operators `+`, `*`, `/\` (and) and `\/` (or). It is able to rearrange the AC expressions to reveal common subexpressions among them. AC-CSE is not enabled by default. It would normally be used in conjunction with Identical or Active CSE. Finally, Active AC-CSE combines one active reformulation (integer negation) with AC-CSE, so for example Active AC-CSE is able to extract $y - z$ from the three expressions $x + y - z$, $w - x - y + z$, and $10 - y + z$, even though the sub-expression occurs as $y - z$ in one and $-y + z$ in the other two. Active AC-CSE is identical to AC-CSE for And, Or and Product, it differs only on sum.

The following flags control CSE. The first, `-no-cse`, turns off all CSE. The other flags each turn on one type of CSE.

```
-no-cse
-identical-cse
-active-cse
-ac-cse
-active-ac-cse
```

Variable Deletion

SAVILE ROW can remove a decision variable (either variables declared with `find` or auxiliary variables introduced during tailoring) when the variable is equal to a constant, or equal to another variable. This is often a useful optimisation. It can be enabled using the following flag.

```
-deletevars
```

Domain Filtering

It can be useful to filter the domains of variables. In SAVILE ROW this is done by running the translation pipeline to completion and producing a Minion file, then running Minion usually with the options `-preprocess SACBounds.limit -outputCompressedDomains`. With these options Minion performs conventional propagation plus SAC on the lower and upper bound of each variable (SACBounds, also known as shaving) with a limit on the number of iterations. The filtered domains are then read back in to SAVILE ROW. The translation process is started again at the beginning. Thus domain filtering can benefit the entire translation process: variables can be removed (with `-deletevars`), constraint expressions can be simplified or even removed (if they are entailed), the number of auxiliary variables may be reduced. In some cases domain filtering can enable another optimisation, for example on the BIBD problem it enables associative-commutative CSE to do some very effective reformulation.

Domain filtering can be used with any target solver (but Minion is always used to perform the domain filtering regardless of the target solver). It is switched on using the following flag.

```
-reduce-domains
```

Standard domain filtering affects the decision variables defined by `find` statements. Auxiliary variables created by SAVILE ROW are not filtered by default. To enable filtering of both `find` variables and auxiliary variables, use the following flag:

```
-reduce-domains-extend
```

Aggregation of Constraints

Aggregation collects sets of constraints together to form global constraints that typically propagate better in the target solver. At present there are two aggregators for `allDifferent` and `GCC`. SAVILE ROW constructs `allDifferent` constraints by collecting not-equal, less-than and other shorter `allDifferent` constraints. `GCC` is aggregated by collecting `atmost` and `atleast` constraints over the same scope.

```
-aggregate
```

TabID

TabID converts constraint expressions into table constraints to improve propagation. A set of heuristics are used to identify candidate expressions for tabulation. Various limits prevent the generation of excessively large tables, or spending excessive amounts of time generating tables. Refer to the JAIR paper for more detail [1].

Full TabID is enabled with the following flag:

```
-tabid
```

We found the best results for TabID with SAT, OR-Tools CP-SAT, and Chuffed when the constraint arity was 2 or 1 [1]. For these solvers, tables of arity 2 are encoded without introducing any additional variables, but extra variables are introduced for arity 3 or higher. Therefore we have the following flag that enables TabID but limits the arity to 2 for the following backends: SAT, MaxSAT, SMT, OR-Tools CP-SAT, and Chuffed.

```
-tabid-light
```

The following flag limits the increase in SAT encoding size (measured by number of clauses) caused by TabID to no more than 2 times larger. For example, if the original constraint would encode with 100 clauses, the encoding of the replacement table cannot have more than 200 clauses.

```
-tab-satsizelimit
```

Finally, the following flag switches on diagnostic output for TabID.

```
-tab-diagnostics
```

Factor Encoding

The factor encoding of table constraints [6] may be useful to strengthen propagation when a problem class has overlapping table constraints (or tabulation produces overlapping table constraints). It is enabled with the following flag.

```
-factor-encoding
```

Integer Variable Types

When creating Minion output SAVILE ROW will by default use the `DISCRETE` or `BOOL` variable type when the variable has fewer than 10,000 values and `BOUND` otherwise. `DISCRETE` and `BOOL` represent the entire domain and `BOUND` only stores the upper and lower bound, thus propagation may be lost when using `BOUND`. The following flag prevents SAVILE ROW using `BOUND` variables.

```
-no-bound-vars
```

Remove Redundant Variables

The following flag causes SAVILE ROW to remove any decision variables that are not mentioned in a constraint or in the objective function. It is enabled by default as part of `-S1`.

```
-remove-redundant-vars
```

Non-functional Auxiliary Variables

The following flag allows SAVILE ROW to create auxiliary variables that are not functionally defined on the primary variables. May change the number of solutions. Currently only affects the SAT, SMT, MaxSAT backends. Typically a non-functional variable would be \leq or \geq an expression rather than equal to the expression, thus avoiding encoding one side of the equality.

```
-aux-non-functional
```

Variable Symmetry Breaking

If variable symmetry breaking is enabled, SAVILE ROW will apply a graph automorphism solver based on Ferret to find symmetries among the decision variables. Any symmetries discovered will be broken by the standard lex-leader method. It is disabled by default because it can be expensive. Variable symmetry breaking is part of `-S2`.

```
-var-sym-breaking
```

4.5.6 SAT Encoding Options

The SAT backend in general is described in Appendix A. There are several options for encoding constraints into SAT, in particular sums and tables.

Sum constraints are divided into three groups and there are flags to control the encoding of each group. The simplest group is at-most-one (AMO) and exactly-one (EO) constraints containing only Boolean terms (not just Boolean variables but any expression of type Boolean) and where all coefficients are 1. The encoding of these constraints is controlled by `-sat-amo` flags. Pseudo-Boolean (PB) sums also contain only Boolean terms but can have coefficients other than 1. The encoding of PB constraints is controlled by `-sat-pb` flags. Finally, the encoding of general sums (containing at least one integer term) is controlled by `-sat-sum` flags.

Table Encodings

Binary table constraints use the support encoding. For non-binary tables, the default encoding introduces a SAT variable for each tuple and was proposed by Bacchus [3]. Another encoding based on MDDs is available for non-binary tables, enabled with the following flag:

```
-sat-table-mdd
```

AMO Encodings

The default encoding for AMO and EO constraints is Chen's 2-product, it is small and tends to be efficient. Other encodings are available: the Commander encoding of Klieber and Kwon (with group size 3); the ladder encoding; an encoding similar to the totalizer (named Tree); the pairwise encoding (with no auxiliary SAT variables); and the bimander encoding with group size 2. The flags to select the encoding are below:

```
-sat-amo-product    (default)
-sat-amo-commander
-sat-amo-ladder
-sat-amo-tree
-sat-amo-pairwise
-sat-amo-bimander
```

PB Encodings

There are nine encodings of Pseudo-Boolean constraints. Eight of them are based on Boffill et al [4] (with thanks to Jordi Coll for providing the implementations): Multi-valued Decision Diagrams (MDD); the Global Polynomial Watchdog (GPW); the Local Polynomial Watchdog (LPW); the Sequential Weighted Counter (SWC); the Generalized Generalized Totalizer (GGT); the Reduced Generalized Generalized Totalizer (RGGT); another version of GGT with a tree-construction heuristic (GGTH); and the Generalized n-Level Modulo Totalizer (GMTO). Each of these is implemented exactly as described in [4]. For these eight encodings, any pseudo-Boolean equality constraints are decomposed into \leq and \geq . Also these eight encodings support only top-level constraints (i.e. constraints that are not nested inside a logic expression). The flags are as follows:

```
-sat-pb-mdd
-sat-pb-gpw
-sat-pb-lpw
-sat-pb-swc
-sat-pb-ggt
-sat-pb-rggt
-sat-pb-ggth
-sat-pb-gmto
```

The final encoding (called Tree) is similar to the Generalized Totalizer. It can directly encode sum-equality, and can also encode non-top-level constraints. It is the default encoding and also the fallback for non-top-level constraints when using one of the other encodings.

```
-sat-pb-tree    (default)
```

The first eight encodings (MDD, GPW, LPW, SWC, GGT, RGGT, GGTH, GMTO) are able to take advantage of at-most-one (and exactly-one) relations on the Boolean variables to reduce the size of the encoding. All four of these encodings work with automatic AMO and EO detection as described in the CP 2019 paper [2], which uses Minion to find mutexes between variables then constructs maximal cliques of the mutex graph. The flag below is used to switch on the AMO and EO detection. It is often worthwhile for sufficiently challenging instances with PB constraints.

```
-amo-detect
```

Sum Encodings

The options for encoding general sums (any sum containing at least one integer variable) are the same as for PB constraints. The flags are as follows:

```
-sat-sum-mdd
-sat-sum-gpw
-sat-sum-lpw
-sat-sum-swc
-sat-sum-ggt
-sat-sum-rggt
-sat-sum-ggth
-sat-sum-gmto
-sat-sum-tree (default)
```

The Tree encoding natively handles integer terms. The other eight (MDD, GPW, LPW, SWC, GGT, RGGT, GGTH, GMTO) treat integer terms as a group of Boolean terms (e.g. $1 * (x = 1) + 2 * (x = 2) + \dots$) with an exactly-one relation on them. These eight encodings also can benefit from automatically detected AMO and EO relations with the `-amo-detect` flag.

4.5.7 SMT Encoding Options

The SMT backend, enabled with the `-smt` flag, is able to produce the SMT-LIB 2 language with a choice of four theories: QF.BV (the theory of bit vectors), QF.IDL (integer difference logic), QF.LIA (linear integer arithmetic), and QF.NIA (nonlinear integer arithmetic). The theory is selected using one of the four flags below:

```
-smt-bv (default)
-smt-idl
-smt-lia
-smt-nia
```

For each theory there is a choice of *nested* or *flat* encodings, where nested aims to maintain the original nesting of expressions (for example, a product within a sum) as far as possible, whereas flat removes nesting by introducing auxiliary variables. Nested is the default. The encoding may be selected using one of the flags below:

```
-smt-nested (default)
-smt-flat
```

The SMT backend is described and evaluated in a CP 2020 paper [7], including comparison of the four theories and two encodings.

Combined with `-run-solver`, the SMT backend will run Boolector, Z3, or Yices2. It will select a default solver for the chosen theory, or the solver can be set manually using one of the following flags:

```
-boolector (default for QF_BV theory)
-z3 (default for QF_NIA)
-yices2 (default for QF_IDL and QF_LIA)
```

Choosing a solver does not affect any other setting of the SMT backend. If the solver does not support the chosen theory (at the time of release) SAVILE ROW will output a warning and continue.

4.5.8 Warning Flags

The following flag produces a warning when the semantics of a model is affected by undefinedness (see Section 3).

```
-Wundef
```

4.5.9 Controlling SAVILE ROW

The following flag specifies a time limit in seconds. SAVILE ROW will stop when the time limit is reached, unless it has completed tailoring and is running a backend solver. The time measured is wallclock time not CPU time.

```
-timelimit <time>
```

Time limits may also be passed through to solvers, for example when using Minion as the backend solver, the following flag will limit Minion's run time (specified in seconds).

```
-solver-options "-cpulimit <time>"
```

A similar approach can be used to apply a time limit to other solvers.

In some cases the SAT or MaxSAT output can be very large. The following option allows the number of clauses to be limited. If the specified number of clauses is reached, SAVILE ROW will delete the partial SAT file and exit.

```
-cnflimit <numclauses>
```

Some reformulations use a pseudorandom number generator with a fixed seed. The following option may be used to set a different seed.

```
-seed <integer>
```

4.5.10 Solver Control

The following flag causes SAVILE ROW to run a solver and parse the solutions produced by it. This is currently implemented for Minion, Gecode, Chuffed, Standard FlatZinc, SAT, SMT, and MaxSAT backends.

```
-run-solver
```

The following two flags control the number of solutions. The first causes the solver to search for all solutions (and SAVILE ROW to parse all solutions). The second specifies a required number of solutions.

```
-all-solutions
-num-solutions <n>
```

When parsing solutions the default behaviour is to create one file for each solution. As an alternative, the following flag will send solutions to standard out, separated by lines of 10 minus signs.

```
-solutions-to-stdout
```

To prevent solutions being output at all, use the following flag.

```
-solutions-to-null
```

The following flag passes through command-line options to the solver. The string would normally be quoted.

```
-solver-options <string>
```

For example when using Minion `-solver-options "-cpulimit <time>"` may be used to impose a time limit, or when using Gecode `-solver-options "-p 8"` causes Gecode to parallelise to 8 cores.

4.5.11 Solver Control – Minion

The following flag specifies where the Minion executable is. The default value is `minion`.

```
-minion-bin <filename>
```

When Minion is run directly by SAVILE ROW, preprocessing is usually applied before search starts. The following flag allows the level of preprocessing to be specified.

```
-preprocess LEVEL
  where LEVEL is one of None, GAC, SAC, SAC_limit, SSAC, SSAC_limit,
  SACBounds, SACBounds_limit, SSACBounds, SSACBounds_limit
```

The default level of preprocessing is `SACBounds_limit`.

The `-preprocess` flag affects the behaviour of Minion in two cases: first when Minion is called to filter domains (the `-reduce-domains` option), and second when Minion is called to search for a solution.

4.5.12 Solver Control – Gecode

The following flag specifies where the Gecode executable is. The default value is `fzn-gecode`.

```
-gecode-bin <filename>
```

4.5.13 Solver Control – Chuffed

The following flag specifies where the Chuffed executable is. The default value is `fzn-chuffed`.

```
-chuffed-bin <filename>
```

4.5.14 Solver Control – OR Tools

The following flag specifies where the OR Tools executable is. The default value is `fzn-ortools`.

```
-or-tools-bin <filename>
```

By default OR Tools is called without its “free search” option. To enable free search, use the pass-through flag:

```
-solver-options "-f"
```

4.5.15 Solver Control – Standard FlatZinc

The following flag specifies a FlatZinc solver binary.

```
-fzn-bin <filename>
```

4.5.16 Solver Control – SAT Solvers

SAVILE ROW is able to run and parse the output of MiniSAT, Lingeling, Glucose, CaDiCaL, and kissat solvers as well as two MiniSAT-based solvers for finding all solutions. The following flag specifies which family of solvers is used. Kissat is the default.

```
-sat-family [ minisat | lingeling | glucose | cadical | kissat |
              nbc_minisat_all | bc_minisat_all ]
```

The ‘all’ values imply the `-all-solutions` flag.

The following flag specifies the SAT solver executable. The default value is `minisat`, `lingeling`, `glucose`, `cadical`, `kissat`, `nbc_minisat_all_release`, or `bc_minisat_all_release` depending on the SAT family.

```
-satsolver-bin <filename>
```

The following flag enables interactive usage of the supported solvers when SAVILE ROW is built with this feature. Supported SAT solvers are used incrementally via JNI calls. Currently supported: `glucose`, `cadical`, and `nbc_minisat_all`.

```
-interactive-sat
```

When using SAT or SMT solvers, optimisation is performed with multiple calls to the solver. There are three modes: *bisect* (default), *linear*, and *unsat*. Each places constraints on the (encoding of the) objective variable. Bisect splits the interval of the objective variable into two approximately equal sub-intervals, and attempts to find a solution in the better sub-interval (i.e. smaller if minimising). The interval is updated and the procedure is repeated until the interval becomes empty, at which point the last solution found is an optimal solution. Linear bounds the objective variable each time a solution is found, such that the next solution will be better. Unsat assigns the objective variable to its best remaining value (smallest value when minimising). If no solution is found, the best value is removed and the procedure iterates, continuing until a solution is found.

The optimisation strategy may be set with the following flag:

```
-opt-strategy [ bisect | linear | unsat ]
```

4.5.17 Solver Control – SMT Solvers

To set the SMT solver executable use the following flag. If it is not set, then SAVILE ROW will look for a binary named `z3`, `yices-smt2`, or `boolector` in the `bin` directory (depending which SMT solver was selected, `boolector` by default, see Section 4.5.7).

```
-smtsolver-bin <filename>
```

Note that the `-opt-strategy` flag is relevant when doing optimisation with an SMT solver.

4.5.18 ReadSolution Mode

SAVILE ROW has two modes of operation, Normal and ReadSolution. Normal is the default, and in this mode SAVILE ROW reads an ESSENCE PRIME model file and optional parameter file and produces output for a solver. In some cases it will also run a solver and parse the solution(s), producing ESSENCE PRIME solution files.

When using ReadSolution mode, SAVILE ROW will read a solution in Minion format and translate it back to ESSENCE PRIME. This allows the user to run Minion separately from SAVILE ROW but still retrieve the solution in ESSENCE PRIME format. When running Minion, the `-solsout` flag should be used to retrieve the solution(s) in a table format.

The mode is specified as follows.

```
-mode [Normal | ReadSolution]
```

When using ReadSolution mode, the name of the aux file previously generated by SAVILE ROW needs to be specified using the `-out-aux` flag, so that SAVILE ROW can read its symbol table. Also the name of the solution table file from Minion is specified using `-minion-sol-file`. `-out-solution` is required to specify where to write the solutions. `-all-solutions` and `-num-solutions <n>` are optional in ReadSolution mode. Below is a typical command.

```
savilerow -mode ReadSolution -out-aux <filename> -minion-sol-file <filename>
          -out-solution <filename>
```

When neither `-all-solutions` nor `-num-solutions <n>` is given, SAVILE ROW parses the last solution in the Minion solutions file. For optimisation problems, the last solution will be the optimal or closest to optimal.

4.5.19 Format Conversion

`-param-to-json`

Reads a parameter file or solution file in Essence Prime format and converts it into a simple JSON format. Special matrix indexing (anything other than $1..n$) will be lost in the conversion.

Appendices

A SAT Encoding

We have used standard encodings from the literature such as the order encoding for sums [15] and support encoding [11] for binary constraints. Also we do not attempt to encode all constraints in the language: several constraint types are decomposed before encoding to SAT.

A.1 Decomposition

The first step is decomposition of the constraints AllDifferent, GCC, Atmost and Atleast. All are decomposed into sums of equalities and we have one sum for each relevant domain value. For example to decompose AllDifferent($[x, y, z]$): for each relevant domain value a we have $(x = a) + (y = a) + (z = a) \leq 1$ (or $(x = a) + (y = a) + (z = a) = 1$ when the number of variables and values are equal). These decompositions are done after AC-CSE if AC-CSE is enabled (because the large number of sums generated hinders the AC-CSE algorithm) and before Identical and Active CSE.

The second step is decomposition of lexicographic ordering constraints. We use the decomposition of Frisch et al [10] (Sec.4) with implication rewritten as disjunction, thus the conjunctions of equalities in Frisch et al become disjunctions of disequalities. Also the common subsets of disequalities are factored out and replaced with auxiliary Boolean variables. This decomposition is also performed after AC-CSE and before Identical and Active CSE.

The third step occurs after all flattening is completed. The constraints min, max, and element are decomposed. For $\min(V) = z$ we have $V[1] = z \vee V[2] = z \dots$ and $z \leq V[1] \wedge z \leq V[2] \dots$. Max is similar to min with \leq replaced by \geq . The constraint $\text{element}(V, x) = z$ becomes $\forall i : (x \neq i \vee V[i] = z)$.

A.2 Encoding of CSP variables

The encoding of a CSP variable provides SAT literals for facts about the variable: $[x = a]$, $[x \neq a]$, $[x \leq a]$ and $[x > a]$ for a CSP variable x and value a . We will refer to $[x = a]$ and $[x \neq a]$ as direct literals, and $[x \leq a]$ and $[x > a]$ as order literals. If the variable has only two values, it is represented with a single SAT variable. All the above facts (for both domain values) map to the SAT variable, its negation, *true* or *false*.

When the CSP variable has more than two values, an encoding is chosen to provide the literals that are needed to encode the constraints. There are three encodings, with one providing the order literals, another providing the direct literals, and the third providing both. Encodings of the binary constraints $<$ and \leq use the order literals, and the tree encoding (see Section 4.5.6) of sums also uses the order literals. All other constraint encodings use the direct literals. If the objective function is a single variable, the order literals are required for that variable. Otherwise, if the objective function is a sum at encoding time (only possible with the MaxSAT backend) then direct literals are required for all variables in the sum.

If the order literals are not needed for a variable x , then the variable is encoded with one SAT variable per domain value (representing $[x = a]$). The 2-product encoding is used for at-most-one and a single clause for at-least-one.

If the direct literals are not needed, then using the language of the *ladder* encoding of Gent et al [12], we have only the ladder variables and the clauses in Gent et al formula (2). If both order and direct literals are required, then we use the full ladder encoding with the clauses in formulas (1), (2) and (3) of Gent et al. Also, for the maximum value $\max(D(x))$ the facts $[x \neq \max(D(x))]$ and $[x < \max(D(x))]$ are equivalent so one SAT variable is saved. For the minimum value $\min(D(x))$, the facts $[x = \min(D(x))]$ and $[x \leq \min(D(x))]$ are equivalent so one SAT variable is saved. Also, a variable may have gaps in its domain. Suppose variable x has domain $D(x) = \{1 \dots 3, 8 \dots 10\}$. SAT variables are

created only for the 6 values in the domain. Facts containing values $\{4 \dots 7\}$ are mapped appropriately (for example $[x \leq 5]$ is mapped to $[x \leq 3]$). The encoding has $2|D(x)| - 3$ SAT variables.

A.3 Encoding of Constraints

Now we turn to encoding of constraints. Some simple expressions such as $x = a$, $x \leq a$ and $\neg x$ (for CSP variable x and value a) are represented with a single SAT literal. We have introduced a new expression type named `SATLiteral`. Each expression that can be represented as a single literal is replaced with a `SATLiteral` in a final rewriting pass before encoding constraints. `SATLiterals` behave like boolean variables hence they can be transparently included in any constraint expression that takes a boolean subexpression.

At-most-one and exactly-one constraints (i.e. a sum of boolean expressions that is ≤ 1 or $= 1$) are by default encoded with the 2-product encoding, which is small and tends to perform better than alternatives. Other encodings can be selected by command-line flags (Section 4.5.6).

There is a choice of 5 encodings of pseudo-Boolean and other sum constraints, as described in Section 4.5.6. The MDD, GPW, SWC, and GGT encodings are implemented exactly as described in Bofill et al [5]. Sum-equal constraints are split into sum- \leq and sum- \geq before encoding. All four of these encodings work with automatic AMO and EO detection as described in the CP 2019 paper [2]. The flag `-amo-detect` is used to switch on the AMO and EO detection.

The Tree encoding for sums decomposes each sum constraint using a binary tree, similar to the generalized totalizer. Auxiliary integer variables are introduced for the internal nodes of the tree. A ternary sum constraint is used to connect each internal node to its two children. The ternary sums are then encoded using the order encoding [15]. The Tree encoding does not work with AMO and EO detection. Tree can be superior to MDD, GPW, SWC, and GGT in the case where the original constraint is a sum-equality, because Tree does not first break down the sum-equality constraint into sum- \leq and sum- \geq .

For other constraints we used the standard support encoding wherever possible [11]. Binary constraints such as $|x| = y$ use the support encoding, and ternary functional constraints $x \diamond y = z$ (eg $x \times y = z$) use the support encoding when z is a constant. Otherwise, $x \diamond y = z$ are encoded as a set of ternary SAT clauses: $\forall i \in D(x), \forall j \in D(y) : (x \neq i \vee y \neq j \vee z = i \diamond j)$. When $i \diamond j$ is not in the domain of z , the literal $z = i \diamond j$ will be false. Logical connectives such as $\wedge, \vee, \leftrightarrow$ have custom encodings. Binary table constraints use the support encoding. Non-binary tables use Bacchus' encoding [3] (Sec.2.1) by default. An MDD-based encoding is also implemented for non-binary table constraints and is enabled with a command-line flag (Section 4.5.6).

B Operator Precedence in ESSENCE PRIME

Table 6 shows the precedence and associativity of operators in ESSENCE PRIME. As you would expect, operators with higher precedence are applied first.

Left-associative operators are evaluated left-first, for example $2/3/4 = (2/3)/4$. The only operator with right associativity is `**`. This allows double exponentiation to have its conventional meaning: $2**3**4 = 2**(3**4)$

Unary operators usually have a higher precedence than binary ones. There is one exception to this rule: that `**` has a higher precedence than unary minus. This allows $-2**2**3$ to have its conventional meaning of $-(2**(2**3)) = -256$, as opposed to $(-2)**(2**3) = 256$.

C Reserved Words

The following words are keywords and therefore are not allowed to be used as identifiers.

```
forall, forAll, exists, sum,
such, that, letting, given, where, find, language,
int, bool, union, intersect, in, false, true
```

Operator	Functionality	Associativity	Precedence
!	Boolean negation		20
	Absolute value		20
**	Power	Right	18
-	Integer negation		15
*	Multiplication	Left	10
/	Division	Left	10
%	Modulo	Left	10
intersect	Domain intersection	Left	2
union	Domain union	Left	1
+	Addition	Left	1
-	Subtraction & Domain Subtraction	Left	1
=	Equality	None	0
!=	Disequality	None	0
<=	Less-equal	None	0
<	Less than	None	0
>=	Greater-equal	None	0
>	Greater than	None	0
<=lex	Lex less-equal	Left	0
<lex	Lex less than	Left	0
>=lex	Lex greater-equal	Left	0
>lex	Lex greater than	Left	0
in	Value in a set	Left	0
/\	And	Left	-1
\/	Or	Left	-2
->	Implication	None	-4
<->	If and only if	None	-4
forall, exists, sum	Quantifiers		-10
,	And	Left	-20

Table 6: Operator precedence in ESSENCE PRIME

References

- [1] Özgür Akgün, Ian Gent, Christopher Jefferson, Zeynep Kiziltan, Ian Miguel, Peter Nightingale, András Salamon, and Felix Ulrich-Oltean. Tabid: automatic identification and tabulation of subproblems in constraint models. *Journal of Artificial Intelligence Research*, 82:1999–2056, 2025.
- [2] Carlos Ansótegui, Miquel Bofill, Jordi Coll, Nguyen Dang, Juan Luis Esteban, Ian Miguel, Peter Nightingale, András Z. Salamon, Josep Suy, and Mateu Villaret. Automatic detection of at-most-one and exactly-one relations for improved sat encodings of pseudo-boolean constraints. In *Proceedings of the 25th International Conference on Principles and Practice of Constraint Programming (CP)*, pages 20–36, 2019.
- [3] Fahiem Bacchus. Gac via unit propagation. In *Principles and Practice of Constraint Programming–CP 2007*, pages 133–147. Springer, 2007.
- [4] Miquel Bofill, Jordi Coll, Peter Nightingale, Josep Suy, Felix Ulrich-Oltean, and Mateu Villaret. SAT encodings for Pseudo-Boolean constraints together with at-most-one constraints. *Artificial Intelligence*, 302:103604, January 2022.
- [5] Miquel Bofill, Jordi Coll, Josep Suy, and Mateu Villaret. Sat encodings of pseudo-boolean constraints with at-most-one relations. In *Proceedings of the 16th CPAIOR 2019: Thessaloniki, Greece*, pages 112–128, 2019.
- [6] Likitvivatanavong C., Xia W., and Yap R.H.C. Higher-order consistencies through gac on factor variables. In *Principles and Practice of Constraint Programming. CP 2014. Lecture Notes in Computer Science*, vol 8656, 2014.
- [7] Ewan Davidson, Özgür Akgün, Joan Espasa, and Peter Nightingale. Effective encodings of constraint programming models to smt. In *International Conference on Principles and Practice of Constraint Programming*, pages 143–159. Springer, 2020.
- [8] Alan M. Frisch, Warwick Harvey, Chris Jefferson, Bernadette Mart’inez Hernández, and Ian Miguel. ESSENCE: A constraint language for specifying combinatorial problems. *Constraints*, 13(3):268–306, 2008.
- [9] Alan M. Frisch and Peter J. Stuckey. The proper treatment of undefinedness in constraint languages. In *Proc. Principles and Practice of Constraint Programming - CP 2009*, pages 367–382, 2009.
- [10] A.M. Frisch, B. Hnich, Z. Kiziltan, I. Miguel, and T. Walsh. Global constraints for lexicographic orderings. In P. van Hentenryck, editor, *Proceedings of the Eighth International Conference on Principles and Practice of Constraint Programming*, pages 93–108, 2002.
- [11] Ian P. Gent. Arc consistency in SAT. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI 2002)*, pages 121–125, 2002.
- [12] Ian P. Gent and Peter Nightingale. A new encoding of alldifferent into sat. In *Proc. 3rd International Workshop on Modelling and Reformulating Constraint Satisfaction Problems (ModRef 2004)*, pages 95–110, 2004.
- [13] Peter Nightingale, Özgür Akgün, Ian P Gent, Christopher Jefferson, and Ian Miguel. Automatically improving constraint models in Savile Row through associative-commutative common subexpression elimination. In *20th International Conference on Principles and Practice of Constraint Programming (CP 2014)*, pages 590–605. Springer, 2014.
- [14] Peter Nightingale, Patrick Spracklen, and Ian Miguel. Automatically improving SAT encoding of constraint problems through common subexpression elimination in Savile Row. In *Proceedings of the 21st International Conference on Principles and Practice of Constraint Programming (CP 2015)*, pages 330–340, 2015.
- [15] Naoyuki Tamura, Akiko Taga, Satoshi Kitagawa, and Mutsunori Banbara. Compiling finite linear CSP into SAT. *Constraints*, 14(2):254–272, 2009.
- [16] Pascal Van Hentenryck. *The OPL Optimization Programming Language*. MIT Press, Cambridge, MA, USA, 1999.