Topic 2: Basic Modelling
(Version of 20th November 2020)

Pierre Flener and Jean-Noël Monette

Optimisation Group
Department of Information Technology
Uppsala University
Sweden

Course 1DL441:
Combinatorial Optimisation and Constraint Programming,
whose part 1 is Course 1DL451:
Modelling for Combinatorial Optimisation

1Many thanks to Guido Tack for feedback
Outline

1. The MiniZinc Language
2. Modelling
3. Set Variables & Constraints
4. Modelling Checklist
Outline

1. The MiniZinc Language

2. Modelling

3. Set Variables & Constraints

4. Modelling Checklist
MiniZinc Model

A MiniZinc model may comprise the following items:

- Parameter declarations
- Variable declarations
- Predicate and function definitions
- Constraints
- Objective
- Output
MiniZinc is strongly typed. The parameter types are:

- **int**: integer
- **bool**: Boolean
- **enum**: enumeration
- **float**: floating-point number
- **string**: string of characters
- **set of τ**: set of elements of type τ, which is int, bool, enum, float, or string
- **array[ρ] of τ**: possibly multidimensional array of elements of type τ, which is not an array; each range in ρ is an enumeration or an integer interval α..β

**Example**

The parameter declaration `int: n` declares an integer parameter of identifier n. One can also write `par int: n` in order to emphasise that n is a parameter.
Types for Decision Variables

Decision variables are implicitly existentially quantified: the aim is to find feasible (and optimal) values in their domains. The variable types for decision variables are:

- **int**: integer
- **bool**: Boolean
- **enum**: enumeration
- **float**: floating-point number (not used in this course)
- **set of enum** and **set of int**: set (do not use)

A possibly multidimensional array can be declared to have variables of any variable type, but it is itself not a variable.

Example

The variable declaration `var int: n` declares a decision variable of domain `int` and identifier `n`.

Tight domains for variables may accelerate the solving: see the next slides for how to do that.
The following literals (or: constants) can be used:

- **Boolean:** `true` and `false`
- **Integers:** in decimal, hexadecimal, or octal format
- **Sets:** between curly braces, for example `{1, 3, 5}`, or as integer intervals, for example `10..30`
- **1d arrays:** between square brackets, say `[6, 3, 1, 7]`
- **2d arrays:** A vertical bar `|` is used before the first row, between rows, and after the last row; for example `[[11, 12, 13, 14] | 21, 22, 23, 24 | 31, 32, 33, 34]`
- **For higher-dimensional arrays,** see slide 11

Careful: The indices of arrays start from 1 by default.
Declarations of Parameters and Variables

1 int: n = 4;
2 par int: p;
3 p = 10;
4 set of int: Primes = {2,3,5,7,11,13};
5 var int: x;
6 var 0..23: hour;
7 var set of Primes: Taken; % no var set in this course

- A parameter must be instantiated, once, to a literal; its declaration can be separated from its instantiation in the model (p), a datafile, the command line, or the IDE.
- The domain of a decision variable can be tightened by replacing its type by a set of values of that type:
  - x must take an integer value.
  - hour must take an integer value between 0 and 23.
  - Taken must be a subset of \{2,3,5,7,11,13\}.
Array and Set Comprehensions

An array or set can be built by a comprehension, using the notation \([\sigma | \gamma]\) or \(\{\sigma | \gamma\}\), where \(\sigma\) is an expression evaluated for each element generated by the generator \(\gamma\): a generator introduces one or more identifiers with values drawn from integer sets, optionally under a where test.

Examples

1. \([i\times2 | i \text{ in } 1..8]\)
   - evaluates to \([2, 4, 6, 8, 10, 12, 14, 16]\)
2. \([i\times j | i, j \text{ in } 1..3 \text{ where } i<j]\) \% both \(i\) and \(j\) in \(1..3\)
   - evaluates to \([2, 3, 6]\)
3. \([i + 2\times j | i \text{ in } 1..3, j \text{ in } 1..4]\)
   - evaluates to \([3, 5, 7, 9, 4, 6, 8, 10, 5, 7, 9, 11]\)
4. \(\{i + 2\times j | i \text{ in } 1..3, j \text{ in } 1..4\}\)
   - evaluates to \(\{3, 4, 5, 6, 7, 8, 9, 10, 11\}\)
5. Sudoku[row,..] \% slicing
   - is syntactic sugar for \([\text{Sudoku[row,col]} | col \text{ in } 1..9]\)
Indexing: Syntactic Sugar

For example,

\[
\text{sum}(i, j \text{ in } 1..n \text{ where } i<j)(X[i] \cdot X[j])
\]

is syntactic sugar for

\[
\text{sum}([X[i] \cdot X[j] \mid i, j \text{ in } 1..n \text{ where } i<j])
\]

This works for any function or predicate that takes an array as sole argument. In particular:

\[
\text{forall}(i \text{ in } 1..n)(Z[i] = X[i] + Y[i]);
\]

is syntactic sugar for

\[
\text{forall}([Z[i] = X[i] + Y[i] \mid i \text{ in } 1..n]);
\]

where a \text{forall} (array[int] of var bool: B) constraint holds if and only if (iff) all the expressions in B hold: it generalises the 2-ary logical-and connective (\&\&).
Array Manipulation

- Changing the number of dimensions and their ranges, provided the numbers of elements match:

  \[
  \text{array1d}(5..10, [|3,2|5,4|6,1|])
  \]

  \[
  \text{array2d}(1..2,1..3,[2,7,3,7,4,9])
  \]

  and so on, until \text{array6d}.

Try and keep your ranges starting from \textit{1}:

- It is easier to read a model under this usual convention.

- Subtle errors may occur otherwise.

- Concatenation: for example, \([1,2] ++ [3,4]\).
Subtyping

A parameter can be used wherever a variable is expected. This extends to arrays: for example, a predicate or function expecting an argument of type `array[int] of var int` can be passed an argument of type `array[int] of int`.

The type `bool` is a subtype of the type `int`. One can coerce from `bool` to `int` using the `bool2int` function: `bool2int(true) = 1` and `bool2int(false) = 0`. This coercion is automatic when needed.

In mathematics we use the Iverson bracket for this purpose: we define \([\phi] = 1\) iff formula \(\phi\) is true, and \([\phi] = 0\) otherwise.
Option Variables

An option variable is a decision variable that can also take the special value <> indicating the absence of the variable.

A variable is declared optional with the keyword opt.

For example, var opt 1..4: x declares a variable x of domain \{1, 2, 3, 4, <>\}.

Do not use explicit option variables in this course. However, one can see them:

- In the documentation:
  for example, var int is a subtype of var opt int.

- In error messages, due to implicit option variables being made explicit while flattening, but things getting too complex: see the symptomatic example at slide 21.
Constraints

A **constraint** is the keyword `constraint` followed by a Boolean expression that must be true in every solution.

### Examples

1. `constraint x < y;`
2. `constraint sum(Q) = 0 /\ alldifferent(Q);`

Constraints separated by a semi-colon (;) are implicitly connected by the 2-ary logical-and connective (`/\`).

What does `constraint x = x + 1` mean? MiniZinc is declarative and has no destructive assignment: this equality `constraint` is *not* satisfied by any value for `x`. MiniZinc allows the syntax `constraint x == x + 1`, but note that MiniZinc is syntax for mathematics and logic!
Objective

The `solve` item gives the objective of the problem:

- `solve satisfy;`
  The objective is to solve a satisfaction problem.

- `solve minimize x;`
  The objective is to minimise the value of variable `x`.

- `solve maximize abs(x)*y;`
  The objective is to maximise the value of the objective function `abs(x)*y`.

MiniZinc does not support multi-objective optimisation yet: multiple objective functions must either be aggregated into a weighted sum, or be handled outside a MiniZinc model.
Output

The \texttt{output} item prescribes what to print upon finding a solution: the keyword \texttt{output} is followed by a string array.

\begin{verbatim}
output [show(x)];
\end{verbatim}

The function \texttt{show} returns a string representing the value of its argument expression.

\begin{verbatim}
output ["Solution:""] ++ [if X[i]>0 then show(2*X[i]) ++ "", " else "", " endif | i in 1..n];
\end{verbatim}

The operator \texttt{++} concatenates two strings or two arrays.

"x = \( (x) , " is equivalent to "x = "++show(x)++", ".

The search strategy of the CP backend Gecode depends on the decision variables mentioned in the \texttt{output} statement.
Operators and Functions

- **Booleans:** `not, \/, \/, <\to, \to, <\to, xor, forall, exists, xorall, iffall, clause, bool2int`
  Beware of arbitrarily nested logical quantifications, such as `forall(...exists(...forall(...)))`!

- **Integers:** `+, -, *, div (\ is for float), mod, abs, pow, min, max, sum, product, = (or ==), <, <=, =>, >, !=`
  Beware of `div`, `mod`, and `pow` on variables!

- **Sets:** `.., in, card, subset, superset, union, array_union, intersect, array_intersect, diff, symdiff, set2array`

- **Strings:** `++, concat, join`

- **Arrays:** `length, index_set, index_set_1of2, index_set_2of2, ..., index_set_6of6, array1d, array2d, ..., array6d`
Predicates and Functions

MiniZinc offers a large collection of predefined predicates and functions to enable a high level at which models can be formulated. See Topic 3: Constraint Predicates.

Each predefined constrained function is defined by the use of the corresponding constraint predicate, possibly upon introducing a new variable.

Example

\[ \text{count} (A, v) > m \text{ is defined by } \text{count} (A, v, c) \land c > m. \]

It is also possible for modellers to define their own functions and predicates, as discussed at slide 25.
Reification

Reification enables the reasoning about the truth of a constraint or a Boolean expression.

Example

```
constraint x < y;
```

requires that \( x \) be smaller than \( y \).

```
constraint b <-> x < y;
```

requires that the Boolean variable \( b \) take the value \( \text{true} \) if and only if \( x \) is smaller than \( y \): the constraint \( x < y \) is said to be reified, and \( b \) is called its reification.

Reification is a powerful mechanism that enables:

- higher-level modelling;
- easier implementation of the logical connectives.
The expression `bool2int(\(\phi\))`, for a Boolean expression \(\phi\), denotes the integer 1 if \(\phi\) is true, and 0 if \(\phi\) is false.

**Example (Cardinality constraint)**

Constrain one or two of three constraints \(\gamma_1, \gamma_2, \gamma_3\) to hold:

\[
\text{bool2int}(\gamma_1) + \text{bool2int}(\gamma_2) + \text{bool2int}(\gamma_3) \text{ in } \{1, 2\}
\]

As `bool2int` coercion is automatic, one can actually write:

\[
\gamma_1 + \gamma_2 + \gamma_3 \text{ in } \{1, 2\}
\]

However, as a coding convention, we recommend to write:

\[
(\gamma_1) + (\gamma_2) + (\gamma_3) \text{ in } \{1, 2\}
\]

thereby mimicking the Iversion bracket (see slide 12).

Reification (implicit via `bool2int` and \((\ldots)\)) has pitfalls:

- Inference and relaxation may be poor: slow solving.
- Not all constraints can be reified in MiniZinc, such as some of those in Topic 3: Constraint Predicates.
A conditional expression can be formulated as follows:

- **Conditional**: `if θ then φ₁ else φ₂ endif`
- **Comprehension**: `[i | i in σ where θ]`

The expressions `φ₁` and `φ₂` must have the same type.

The test `θ` after `if` or `where` may depend on variables, but this can be a source of inefficiency, unexpected behaviour (see documentation Section 2.4.2), or impossible flattening!

### Example

1. `enum I; set of int: T; array[I] of var T: X;`
2. `array[I] of var T: Y = [X[i] | i in I where X[i]>0]; constraint sum(Y) < 7;`

This yields an error message with `var opt` (see slide 13) as the indices of `Y` cannot be determined when flattening and cannot just be set to `I`. But the following works:

2. `constraint sum([X[i] | i in I where X[i]>0]) < 7;`

and so does the use of implicit reification, possibly better:

2. `constraint sum([(X[i]>0) * X[i] | i in I]) < 7;`
Example (Soft Constraints: Photo Problem)

An enumeration Persons of n people lines up for a photo.
enum PrefRoles = {who, whom}; % do not use 1..2
array[1..q, PrefRoles] of Persons: Pref;

Preference k in 1..q denotes that person Pref[k, who] wants to be next to Pref[k, whom].
Maximise the number of satisfied preferences.

Let decision variable Pos[p] denote the position in 1..n, in left-to-right order, of person p in Persons on the photo.
The array Pos must form a permutation of the positions:
constraint alldifferent(Pos);

The objective, formulated using implicit reification, is:
solve maximize sum(k in 1..q) ( abs(Pos[Pref[k,1]]-Pos[Pref[k,2]])=1 );
Example (Soft Constraints: Weighted Photo Problem)

An enumeration Persons of n people lines up for a photo.

```plaintext
enum PrefRoles = {who,whom}; % do not use 1..2
array[1..q,PrefRoles] of Persons: Pref;
array[1..q] of int: Weight;
```

Preference k in 1..q denotes that person Pref[k, who] wants to pay Weight[k] to be next to Pref[k, whom]. Maximise the weighted number of satisfied preferences.

Let decision variable Pos[p] denote the position in 1..n, in left-to-right order, of person p in Persons on the photo. The array Pos must form a permutation of the positions:

```plaintext
constraint alldifferent(Pos);
```

The objective, formulated using implicit reification, is:

```plaintext
solve maximize sum(k in 1..q) ( abs(Pos[Pref[k,1]]-Pos[Pref[k,2]])=1 ) ;
```
Example (Soft Constraints: Weighted Photo Problem)

An enumeration Persons of \( n \) people lines up for a photo.

```mini
enum PrefRoles = {who, whom}; % do not use 1..2
array[1..q,PrefRoles] of Persons: Pref;
array[1..q] of int: Weight;
```

Preference \( k \) in \( 1..q \) denotes that person \( \text{Pref}[k, \text{who}] \) wants to pay \( \text{Weight}[k] \) to be next to \( \text{Pref}[k, \text{whom}] \).

Maximise the weighted number of satisfied preferences.

Let decision variable \( \text{Pos}[p] \) denote the position in \( 1..n \), in left-to-right order, of person \( p \) in Persons on the photo.

The array \( \text{Pos} \) must form a permutation of the positions:

```mini
constraint alldifferent(Pos);
```

The objective, formulated using implicit reification, is:

```mini
solve maximize sum(k in 1..q) (Weight[k]*(abs(Pos[Pref[k,1]]-Pos[Pref[k,2]]))=1));
```
Example (Sum of unweighted reified constraints)

The expression \( \text{sum}(i \text{ in } \text{index_set}(A))(A[i]=v) \) denotes the count of occurrences in array \( A \) of \( v \).

This idiom is very common in constraint-based models. So:

**Definition (The count constraint predicate)**

The constraint \( \text{count}(A, v, c) \) holds iff variable \( c \) has the number of variables of array \( A \) that are equal to variable \( v \).

For other predicates, see Topic 3: Constraint Predicates.

**Definition (The count constrained function)**

The expression \( \text{count}(A, v) \) denotes the number of variables of array \( A \) that are equal to variable \( v \).

**Example (Unweighted Photo Problem, revisited)**

\[
\text{solve maximize } \text{count}([\text{abs}(\text{Pos}[\text{Pref}[k,1]]-\text{Pos}[\text{Pref}[k,2]]) \mid k \text{ in } 1..q], 1);
\]

Functional constraint predicates are available as functions.
Predicate and Function Definitions

Examples

1. function int: double(int: x);
2. function var int: double(var int: x);
3. predicate pos(var int: x);
4. function var bool: neg(var int: x);

A predicate is a function denoting a \texttt{var bool}:

Examples

3. function var bool: pos(var int: x);
4. predicate neg(var int: x);

Function and predicate names can be overloaded.
The body of a predicate or function definition is an expression of the same type as the denoted value.

Examples

1. function int: double(int: x) = 2*x;
2. function var int: double(var int: x) = 2*x;
3. predicate pos(var int: x) = x > 0;
4. function var bool: neg(var int: x) = x < 0;

One can use if ... then ... else ... endif, predicates and functions, such as forall and exists, as well as let expressions (see the next slide) in the body of a predicate or function definition.
Let Expressions

One can introduce local identifiers with a let expression.

Examples

```plaintext
function int: double(int: x) =
    let { int: y = 2 * x } in y;

function var int: double(var int: x) =
    let { var int: y = 2 * x } in y;

function var int: double(var int: x) =
    let { var int: y;
        constraint y = 2 * x
    } in y;
```

The second and third functions are equivalent: each use adds a decision variable to the model.
What is the difference between the next two definitions?

1. `predicate posProd(var int: x, var int: y) =`  
2. `let { var int: z; constraint z = x * y } in z > 0;`

5. `predicate posProd(var int: x, var int: y) =`  
6. `let { var int: z } in z = x * y /\ z > 0;`

Their behaviour is different in a negative context, such as `not posProd(a,b)`:

- The 1st one then ensures `a * b = z /\ z <= 0`.
- The 2nd one then ensures `a * b != z \/ z <= 0` and leaves `a` and `b` unconstrained.
Using Predicates and Functions

Advantages of using predicates and functions in a model:

- Software engineering good practice:
  - Reusability
  - Readability
  - Modularity

- The model might be solved more efficiently:
  - Better common-subexpression elimination.
  - The definitions can be technology- or solver-specific. If a predefined constraint predicate is a built-in of a solver, then its solver-specific definition is identity!
Remarks

- The order of model items does not matter.
- One can include other files. Example: `include "globals.mzn"`.
- The following functions are useful for debugging:
  - `assert(θ,"error message")`
    If the Boolean expression $\theta$ evaluates to `false`, then abort with the error message, otherwise denote `true`.
  - `trace("message", φ)`
    Print the message and denote the expression $\phi$.
  - ...

Other Modelling Languages

- OPL: https://www.ibm.com/analytics/optimization-modeling-interfaces
- Comet: https://mitpress.mit.edu/books/constraint-based-local-search
- Essence and Essence’: https://constraintmodelling.org
- Zinc: https://dx.doi.org/10.1007/s10601-008-9041-4
- AIMMS: https://aimms.com
- AMPL: https://AMPL.com
- GAMS: https://gams.com
- SMT-lib: http://www.smt-lib.org
- …
Outline

1. The MiniZinc Language

2. Modelling

3. Set Variables &Constraints

4. Modelling Checklist
From a Problem to a Model

What is a good model for a constraint problem?

- A model that correctly represents the problem

- A model that is easy to understand and maintain

- A model that is solved efficiently, that is:
  - short solving time to find one, all, or best solution(s)
  - good solution within a limited amount of time
  - small search space (under systematic search)

Food for thought: What is correct, easy, short, good, . . . ?
Modelling Issues

Modelling is still more an Art than a Science:

- Choice of the decision variables and their domains
- Choice of the constraint predicates, in order to model the objective function, if any, and the constraints
- Optional for CP and LCG:
  - Choice of the consistency for each constraint
  - Choice of the variable selection strategy for search
  - Choice of the value selection strategy for search

See Topic 8: Inference & Search in CP & LCG.

Make a model correct before making it efficient!
## Choice of the Decision Variables

### Examples (Alphametic Problems)

**SEND + MORE = MONEY:**
Model without carry variables: 19 of 23 CP nodes visited:

\[
1000 \cdot (S + M) + 100 \cdot (E + O) + 10 \cdot (N + R) + (D + E) \\
= 10000 \cdot M + 1000 \cdot O + 100 \cdot N + 10 \cdot E + Y
\]

Model with carry variables: 23 of 29 CP nodes are visited:

\[
\begin{align*}
D + E &= 10 \cdot C_1 + Y \\
\wedge &\quad N + R + C_1 = 10 \cdot C_2 + E \\
\wedge &\quad E + O + C_2 = 10 \cdot C_3 + N \\
\wedge &\quad S + M + C_3 = 10 \cdot M + O
\end{align*}
\]

**GERALD + DONALD = ROBERT:**
The model with carry variables is more effective in CP: only 791 of 869 nodes are visited, rather than 13,795 of 16,651 search nodes for the model without carry variables.
Choice of the Constraint Predicates

Example (The alldifferent constraint predicate)

The constraint alldifferent(A) for an array A of size n usually leads to faster solving than its definition by a conjunction of \( \frac{n \cdot (n-1)}{2} \) disequality constraints:

\[
\forall (i, j \text{ in index_set}(A) \text{ where } i < j) (A[i] \neq A[j])
\]

For more examples, see Topic 3: Constraint Predicates.
Guidelines: Reveal Problem Structure

- Use few decision variables, and declare tight domains
- Beware of nonlinear and power constraints: \texttt{pow}
- Beware of division constraints: \texttt{div} and \texttt{mod} (avoid \texttt{/})
- Beware of disjunction & negation: \texttt{/}, \texttt{<-}, \texttt{->}, \texttt{<->}, \texttt{not}
- Express the problem concisely
  (Topic 3: Constraint Predicates)
- Precompute solutions to a sub-problem into a table
  (Topic 3: Constraint Predicates; Topic 4: Modelling)
- Use implied constraints (Topic 4: Modelling)
- Use different viewpoints (Topic 4: Modelling)
- Exploit symmetries (Topic 5: Symmetry)

Careful: These guidelines of course have their exceptions!
It is important to test empirically several combinations of model, solver, and solving technology.
Use Few Decision Variables

When appropriate, use a single integer variable instead of an array of Boolean variables:

Example

Assume Joe must be assigned to exactly one task in 1..n:

- Use a single integer variable, \( \text{var } 1..n: \text{joesTask} \), representing which task Joe is assigned to.

- Don’t use \( \text{array[1..n] of var bool:joesTask} \), each element \( \text{joesTask}[t] \) representing whether (true) or not (false) Joe is assigned to task \( t \), plus \( \text{count(joesTask, true)} = 1 \).

When appropriate (but not in this course), use a single set variable instead of an array of Boolean or integer variables: see slides 48 and 50.
Declare the Variables with Tight Domains

Manually tightening the domains of the variables might accelerate the solving. In particular, try and avoid \texttt{var int:}

Example (Bound the domains by using parameters)

If the variable $t$ denotes a time, then write \texttt{var 0..h: t}, where $h$ is a parameter, instead of \texttt{var int: t}.

Example (Derive tight domains from the parameters)

\begin{verbatim}
array[Apps] of 0..1000: Salary; % Salary[a]/job by a
var Apps: teacher; var 0..max(Salary): salaryTeacher;
constraint salaryTeacher = Salary[teacher] \and ...;
\end{verbatim}

Domain information may be exploited during flattening, so:

Counterexample (Do not set domains by constraints)

Do not reformulate \texttt{var 0..h: t} as \texttt{var int: t; constraint 0<=t \and t<=h}.
Beware of Nonlinear and Power Constraints

Constraining the product of two or more variables often makes the solving slow. Try and find a linear reformulation.

**Example**

```plaintext
array[1..n] of var 0..1: X;
array[1..n] of var 0..1: Y;
constraint count([X[i]*Y[i] | i in 1..n], 1) = b;
```

should be reformulated as:

```plaintext
array[1..n] of var 0..1: X;
array[1..n] of var 0..1: Y;
constraint count([X[i]+Y[i] | i in 1..n], 2) = b;
```
Beware of Division Constraints

The use of `div` and `mod` often makes the solving slow. Use `table` (see Topic 3: Constraint Predicates) or reformulate.

Example

The model snippet

```mini
solve minimize sum(X) div n; % minimise the average over n variables X[i] and parameter n should become:
```

```mini
solve minimize sum(X); % minimise the sum
output [show(sum(X) div n)]; % output the average
```
Beware of Disjunction and Negation

The disjunction of constraints (with \(\lor\), \(\text{xor}\), \(<-\), \(<-\), \(\exists\), \(\text{xorall}\), \(\text{if } \theta \text{ then } \phi \text{ else } \psi \text{ endif}\)) often makes the solving slow. Try and express disjunctive combinations of constraints otherwise.

**Example**

```plaintext
constraint x = 0 \(\lor\) (low <= x \(\lor\) x <= up);
```

with parameters `low` and `up`, should be reformulated as:

```plaintext
constraint x in {0} union low..up;
```

or, even better in this particular case, as:

```plaintext
var {0} union low..up: x;
```

Disjunction or other sources of slow solving may also be introduced by `not`, so try and avoid negation as well.
Example

```plaintext
constraint b -> x = 9;
constraint (not b) -> x = 0;
```

can be reformulated as (recall that `bool2int(true)=1`):

```plaintext
constraint x = 9 * b;
```

or as (note that array indexing starts by default at 1):

```plaintext
constraint x = [0,9][1+b];
```

But beware of such premature fine-tuning of a model!
The following versions are clearer and often good enough:

```plaintext
constraint x = if b then 9 else 0 endif;
```

and

```plaintext
constraint if b then x=9 else x=0 endif;
```
Express the Problem Concisely

Whenever possible, use a single predefined constraint predicate instead of a long-winded (quantified) formulation.

Example (The `alldifferent` constraint predicate)

The constraint `alldifferent(A)` for an array `A` of size `n` usually leads to faster solving than its definition by a conjunction of \( \frac{n(n-1)}{2} \) disequality constraints:

\[
\text{forall}(i, j \text{ in index_set}(A) \text{ where } i < j) (A[i] \neq A[j])
\]

For more examples, see Topic 3: Constraint Predicates.
Outline

1. The MiniZinc Language
2. Modelling
3. Set Variables & Constraints
4. Modelling Checklist
### Motivating Example 1

#### Example (Agricultural experiment design, AED)

<table>
<thead>
<tr>
<th></th>
<th>plot1</th>
<th>plot2</th>
<th>plot3</th>
<th>plot4</th>
<th>plot5</th>
<th>plot6</th>
<th>plot7</th>
</tr>
</thead>
<tbody>
<tr>
<td>barley</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>corn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>millet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>oats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rye</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>spelt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wheat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Constraints to be satisfied:**

1. Equal growth load: Every plot grows 3 grains.
2. Equal sample size: Every grain is grown in 3 plots.
3. Balance: Every grain pair is grown in 1 common plot.

**Instance:** 7 plots, 7 grains, 3 grains/plot, 3 plots/grain, balance 1.
Motivating Example 1

Example (Agricultural experiment design, AED)

<table>
<thead>
<tr>
<th>Grain</th>
<th>plot1</th>
<th>plot2</th>
<th>plot3</th>
<th>plot4</th>
<th>plot5</th>
<th>plot6</th>
<th>plot7</th>
</tr>
</thead>
<tbody>
<tr>
<td>barley</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>corn</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>millet</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>oats</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>–</td>
</tr>
<tr>
<td>rye</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>spelt</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>wheat</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
</tr>
</tbody>
</table>

Constraints to be satisfied:

1. Equal growth load: Every plot grows 3 grains.
2. Equal sample size: Every grain is grown in 3 plots.
3. Balance: Every grain pair is grown in 1 common plot.

Instance: 7 plots, 7 grains, 3 grains/plot, 3 plots/grain, balance 1.
In a BIBD, the plots are **blocks** and the grains are **varieties**:

**Example (BIBD integer model: ✓ ⊸ 1 and ⊸ 0)**

-3 enum Varieties; enum Blocks;
-2 int: blockSize; int: sampleSize; int: balance;
-1 array[Varieties,Blocks] of var 0..1: BIBD;
0 solve satisfy;
1 constraint forall(b in Blocks)
   (blockSize = sum(BIBD[..,b]));
2 constraint forall(v in Varieties)
   (sampleSize = sum(BIBD[v,..]));
3 constraint forall(v, w in Varieties where v < w)
   (balance = sum([BIBD[v,b]*BIBD[w,b] | b in Blocks]));

**Example (Instance data for our AED)**

-3 Varieties = {barley,...,wheat}; Blocks = {plot1,...,plot7};
-2 blockSize = 3; sampleSize = 3; balance = 1;
Example (Idea for another BIBD model)

<table>
<thead>
<tr>
<th></th>
<th>barley</th>
<th>corn</th>
<th>millet</th>
<th>oats</th>
<th>rye</th>
<th>spelt</th>
<th>wheat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>{plot1, plot2, plot3}</td>
<td>{plot1, plot4, plot5}</td>
<td>{plot1, plot6, plot7}</td>
<td>{plot2, plot4, plot6}</td>
<td>{plot2, plot5, plot7}</td>
<td>{plot3, plot4, plot7}</td>
<td>{plot3, plot5, plot6}</td>
</tr>
</tbody>
</table>

**Constraints to be satisfied:**

1. Equal growth load: Every plot grows 3 grains.
2. Equal sample size: Every grain is grown in 3 plots.
3. Balance: Every grain pair is grown in 1 common plot.
Example (BIBD set model: a block set per variety)

```mini
-3 enum Varieties; enum Blocks;
-2 int: blockSize; int: sampleSize; int: balance;
-1 array[Varieties] of var set of Blocks: BIBD;
0 solve satisfy;
1 constraint forall(b in Blocks)
   (blockSize = sum(v in Varieties)(b in BIBD[v]));
2 constraint forall(v in Varieties)
   (sampleSize = card(BIBD[v]));
3 constraint forall(v, w in Varieties where v < w)
   (balance = card(BIBD[v] intersect BIBD[w]));
```

Example (Instance data for our AED)

```mini
-3 Varieties = {barley,...,wheat}; Blocks = {plot1,...,plot7};
-2 blockSize = 3; sampleSize = 3; balance = 1;
```
Motivating Example 2

Example (Hamming code: problem)

The **Hamming distance** between two same-length strings is the number of positions at which the two strings differ. Examples: \( h(10001, 01001) = 2 \) and \( h(11010, 11110) = 1 \).

ASCII has codewords of \( m = 8 \) bits for \( n = 2^m \) symbols, but the least Hamming distance is \( d = 1 \): no robustness!

Toward high robustness in data transmission, we want to generate a codeword of \( m \) bits for each of the \( n \) symbols of an alphabet, such that the Hamming distance between any two codewords is at least some given constant \( d \).

\(^2\)Based on material by Christian Schulte
Example (Hamming code: model)

We encode a codeword of $m$ bits as the set of positions of its unit bits, the least significant bit being at position 1. Example: 10001 is encoded as $\{1, 5\}$, and 01001 as $\{1, 4\}$. In general: $b_m \cdots b_1$ is encoded as $\{1 \cdot b_1, \ldots, m \cdot b_m\} \setminus \{0\}$. So the Hamming distance between two codewords is $u - i$, where $u$ is the size of the union of their encodings and $i$ is the size of the intersection of their encodings, that is the size of the symmetric difference of their encodings. Hence:

$$\text{array}[1..n] \text{ of var set of } 1..m: C;$$

$$\text{constraint } \forall (i, j \in 1..n \text{ where } i < j) \quad (\text{card}(C[i] \text{ symdiff } C[j]) \geq d);$$

Definition

A set (decision) variable takes a set as value, and has a set of sets as domain. For its domain to be finite, a set variable must be a subset of a given finite set.
Set-constraint predicates exist for the following semantics:

- **Cardinality:** $|S| = n$
- **Membership:** $n \in S$
- **Equality:** $S_1 = S_2$
- **Disequality:** $S_1 \neq S_2$
- **Subset:** $S_1 \subseteq S_2$
- **Union:** $S_1 \cup S_2 = S_3$
- **Intersection:** $S_1 \cap S_2 = S_3$
- **Difference:** $S_1 \setminus S_2 = S_3$
- **Symmetric difference:** $(S_1 \cup S_2) \setminus (S_1 \cap S_2) = S_3$
- **Order:** $S_1 \subseteq S_2 \lor \min((S_1 \setminus S_2) \cup (S_2 \setminus S_1)) \in S_1$
- **Strict order:** $S_1 \subset S_2 \lor \min((S_1 \setminus S_2) \cup (S_2 \setminus S_1)) \in S_1$

where the $S_i$ are set variables and $n$ is an integer variable. Avoid set variables in M4CO: few solvers support them well.
Outline

1. The MiniZinc Language
2. Modelling
3. Set Variables & Constraints
4. Modelling Checklist
Conventions of all Slides (recommended!)

- Scalar identifiers (\texttt{enum} items, \texttt{int}) start in lowercase
- Mass identifiers (\texttt{array}, \texttt{enum}, \texttt{set}) start in uppercase
- Arrays have self-explanatory total-function identifiers: a given/unknown total function $f : X \rightarrow Y$ can be modelled as $\texttt{array}[X]$ of \texttt{par}|\texttt{var} Y : \texttt{F}$
- Index identifiers are lowercase mnemonic: memory aid
- Comments about the \textit{next} line end in : like line 2 below

**Example**

1. \texttt{int: nQueens; \% the given number of queens}
2. \% Row[c] = the row number of the queen in column \texttt{c}:
3. \texttt{array[1..nQueens] of var 1..nQueens: Row;}

Variable \texttt{Row[c]} is like \texttt{Row(c)}, denoting the function \texttt{Row} applied to argument \texttt{c}. The array \texttt{Row} is \textit{not} a variable, but an \textit{array of variables}: it contains row numbers, but calling it \texttt{Rows} would make \texttt{Rows[c]} seem to denote a set of rows!
Checklist for Designing or Reading a Model

1. Use `enum` instead of $\alpha \ldots \beta$ for index ranges when fitting.
2. All decision variables have tight declared domains.
3. No explicit variables of type `opt \tau` are used (in M4CO).
4. No variables of type `set of \tau` are used (in M4CO).
5. The index ranges of all arrays start from 1, for clarity.
6. No `sum|forall(i in 1..x)` with `x` a var. is used.
7. Beware of `where \theta` and `if \theta` with \theta having variables.
9. Beware of negation and disjunction: `not, \land, \lor, exists, xor, xorall, if \theta then \phi else \psi endif, <-, ->, <->`.
10. Beware of arbitrarily nested logical quantifications, such as `forall(...exists(...forall(...)))`.