Confluence in Constraint Handling Rules: A retrospective overview*

Extended abstract of an invited talk given at IWC 2018
7th International Workshop on Confluence, July 7th 2018, Oxford, United Kingdom

Henning Christiansen and Maja H. Kirkeby
Computer Science, Roskilde University, Denmark
henning@ruc.dk and majaht@ruc.dk

1 Introduction

Constraint Handling Rules, CHR, is a nondeterministic programming language whose programs consist of rewrite rules over program states, and being able to show confluence may be an important part of a program correctness proof. Confluence of CHR programs has been studied from the first introduction of the language [8, 9]. The first essential results on proving confluence in CHR [1–3], developed during the 1990s, were formulated with respect to a logic based semantics. This choice gave elegant proofs for terminating systems based on the subsumption principle (see Section 2 below). More recent work extends the previous methods for proving confluence to include invariants and confluence modulo equivalence. Furthermore, these new results were developed for a more realistic semantics that reflects the de-facto standard implementations of CHR upon Prolog, including a correct treatment of Prolog’s non-logical devices (e.g., var/1, nonvar/1, is/2) and runtime errors. In the following we give an overview of confluence results for Constraint Handling Rules, from early, fundamental results to recent extensions including state invariants and confluence modulo equivalence.

2 Preliminaries

We rephrase the following standard definitions and properties. A transition system \( D = \langle S, \rightarrow \rangle \) consists of a set of states \( S \); a transition is an element of \( \rightarrow : S \rightarrow S \), written \( s_1 \rightarrow s_2 \) or, alternatively, \( s_2 \leftarrow s_1 \), and \( \rightarrow^* \) is the reflexive transitive closure of \( \rightarrow \). An object corner is a structure of the form \( s_1 \leftarrow s \rightarrow s_2 \) in which the indicated relationships hold. A system is confluent whenever, for all states \( s, s_1, s_2 \) with \( s_1 \leftarrow* s \rightarrow* s_2 \), that \( s_1, s_2 \) are joinable, i.e., there exists a state \( t \) such that \( s_1 \rightarrow* t \leftarrow* s_2 \); and it is locally confluent whenever any object corner is joinable. The fundamental result is Newman’s lemma [17]: A terminating system is confluent if and only if it is locally confluent.

Proving different types of systems confluent has been facilitated by constructions of sets of critical pairs – or critical corners as we refer to, including the common ancestor state. Typically, these critical corners are selected object corners of the same system (but we will relax this later) and are accompanied by a notion of subsumption (that may vary depending on the type of system), i.e., a given critical corner subsumes a set object corners.

The set of critical corners should satisfy the following properties.

---

*This work is supported by The Danish Council for Independent Research, Natural Sciences, grant no. DFF 4181-00442.
whenever a critical corner is joinable, any object corner that it subsumes is joinable,

– an object corner not subsumed by a critical corner is known to be joinable in some way or another (and thus referred to as trivial corners).

Thus, a proof of confluence for a terminating system may a matter of 1) describe a set of critical corners, 2) show each of them joinable; we refer to this proof strategy as the subsumption principle. Ideally a set of critical corners is finite, but this may not always be the case.

3 Constraint Handling Rules

Constraint Handling Rules [8, 10, 11], CHR, is a nondeterministic programming language based on rewriting rules. It was originally designed as a logical language for implementation of traditional constraint solvers, but CHR has become important as a general-purpose language for representing knowledge and expressing algorithms in a high-level fashion; today it is applied in many areas, for instance analysis of types, multi-agent systems, scheduling, and abductive reasoning; see, e.g., [12] for an overview. While most common CHR implementations are deterministic rather than nondeterministic, it is still useful to consider CHR programs as nondeterministic as it allows the programmer to disregard the execution model of the specific implementation.

CHR relates to the logic programming tradition; constraints are first-order atoms and the language has a declarative semantics [11] based on a logical reading of the rules. CHR programs consist of (a finite set of) guarded rewrite rules over multi-sets of constraints, called constraint stores.

Example 1 ([4, 5]). The following CHR program, consisting of a single rule (without guard), collects a number of separate items into a set represented as a list of items.

```
set(L), item(A) <=> set([A|L]).
```

This rule will apply repeatedly, replacing constraints matched by the left hand side by those indicated to the right. With a nondeterministic semantics, the query

```
?- item(a), item(b), set([]).
```

may lead to two different final states, `{set([a,b])}` and `{set([b,a])}`, both representing the same set.

There are three rule types in CHR; the one in the example above is a simplification rule that replaces constraints by new constraints; another rule type called propagation rule adds new constraints (such a rule is written using a different arrow `==>`); and the last type, called a simpagation, is a generalization of these two. CHR applies a bookkeeping mechanism to avoid trivial looping that otherwise arises with propagation rules. This is easily incorporated into a formal semantics, but for simplicity we of notation, ignore this in the present paper. For an in-depth introduction of all rule types, see, e.g., [11].

There are two sorts of constraints: user-defined constraints that are those appearing in the head of the rules, and built-in constraints with fixed meanings. The exact set of available built-in constraints and the modeling of their meaning varies with the CHR semantics of interest. Several CHR semantics have been proposed [1–7, 11]. The semantics used for confluence considerations has traditionally only allowed logical built-ins [1–3, 6, 7, 11], but a recent semantics [4, 5] reflects the de-facto standard implementations of CHR upon Prolog, including a correct treatment of Prolog’s non-logical devices (e.g., var/1, nonvar/1, is/2) and runtime errors.
The shape of a CHR state varies with the semantics: in a simple logic-based semantics, a state is of the form \(\langle S, B \rangle\) where \(S\) is a constraint store and \(B\) is a built-in store containing a conjunction of built-ins where each built-in has a logical meaning; and in a Prolog-based semantics, a state consists of a constraint store \(S\), and each built-in has an operational meaning producing a special substitution that is applied to \(S\). For instance \(t_1 \succ t_2\) evaluates to an empty substitution if \(t_1\) and \(t_2\) are ground arithmetic expressions with values \(v_1\) and \(v_2\), and \(v_1 \succ v_2\) holds, it evaluates to a failure substitution if instead \(v_1 \preceq v_2\), and to an error substitution, otherwise. Applying an error (a failure) substitution to a state change the state to a special state, namely an error-state (a failure-state); if such a state is reached no further transitions are allowed; therefore, the Prolog-based semantics is sensitive to the execution order of built-ins.

A built-in constraint may occur in the constraint store either introduced in the start query or by a rule body, and in rule-guards; user-defined constraints may not occur in the guards and built-in constraints may not occur on the rules left-hand sides. In the logic based semantics, built-ins in the constraint store are transferred to the built-in store, and in the Prolog based semantics they are evaluated by their operational meaning and the state is updated accordingly. For instance, considering the logic based semantics \(\langle \{a=X, p(X)\}, \text{true} \rangle\) is transformed into \(\langle \{p(X)\}, X=a \rangle\). Considering instead the Prolog based semantics \(\langle \{a=X, p(X)\} \rangle\) is transformed into \(\langle \{p(a)\} \rangle\); see, e.g., [11] and respectively [5] for definitions.

A guard is a sequence of built-in constraints; in the above example there are no guards, but the following example includes rule-guards on the right-hand side, e.g., \((X>0 \mid)\). For the logic based semantics a rule may be applied if the current built-in store implies the guard; and in the Prolog based semantics it may be applied if a rule-guard evaluates to neither failure nor an error and it does not instantiate existing constraint store variables.

Example 2. Consider the following CHR program with four rules, \(r_1\)–\(r_4\).

\[
\begin{align*}
    r_1: & \quad p(X) \leftrightarrow q(X) \quad r_3: \quad q(X) \leftrightarrow X>0 \mid r(X) \\
    r_2: & \quad p(X) \leftrightarrow r(X) \quad r_4: \quad r(X) \leftrightarrow X \leq 0 \mid q(X)
\end{align*}
\]

In both semantics \(r_3\) applies to \(q(n)\) constraints when \(n\) is a positive number, e.g., \(q(1)\). Under the Prolog based semantics the rule cannot apply to the state \(\langle \{q(X)\} \rangle\) since the variable \(X\) causes the guard \(X>0\) to result in an error. Under the logic based semantics the \(r_3\) may transform a state \(\langle \{q(X)\}, X>2 \rangle\) to \(\langle \{r(X)\}, X>2 \rangle\) because \(X>0\) is a logical consequence of \(X>2\), whereas \(r_3\), for instance, does not transform the state \(\langle \{q(X)\}, \text{true} \rangle\).

4 Confluence in CHR

The results on confluence for CHR are similar to those for term rewriting systems; critical corners appear when two instances of rules can apply to overlapping constraints in the constraint store, see, e.g., following example.

Example 3 (Ex. 1 continued). The set-program of Example 1 is not confluent under neither semantics, as both of its critical corners shown below are not joinable.

---

1In the original papers [1–3], the definition of states includes several extra components, that were later shown redundant [4, 5].

2In the logic based semantics it is seen as a conjunction. In the Prolog based semantics they are evaluated from left to right, each influencing the whole state including the subsequent series of the built-ins; if the guard evaluates to failure or error these updates are discarded, see [5] for details.

3The built-in store under the logic based semantics is true for all indicated states.
The program of Example 2 is not confluent under neither semantics\(^3\), as its single critical corner \(\{q(X)\} \leftarrow \{p(X)\} \rightarrow \{r(X)\}\) is not joinable.

When a critical corner is formed by rules with guards, their satisfaction is incorporated into the common ancestor state, which works nicely under the logic based semantics when only logical built-ins are assumed. Here subsumption of an object corner by a critical corner is defined by applying substitution and adding more constraints; the inherent monotonicity ensures joinability of any object corner subsumed by a critical corner.

However, the subsumption principle as explained so far as well as the use of the logic based semantics cannot handle non-logical built-in predicates that are available — and extensively used in practice — in standard implementations of CHR. Consider, for example, the CHR rule \(p(X) \iff \text{var}(X) \mid q(X)\), whose guard consists of Prolog’s test for whether its argument is an uninstantiated variable. While the rule may apply to a state containing \(p(X)\), it does not apply to a more specific state containing \(p(1)\). The other way round, if the guard is instead \(\text{nonvar}(X)\), the rule may apply to a lot of subsumed instances with \(p(1)\), \(p(2)\), ..., but this cannot be “verified” by investigating a most general state including \(p(X)\) to which the \text{nonvar} version of the rule does not apply.

In order to restore a subsumption principle aiming at finite proofs, we have taken the consequence in our own work to describe critical corners in a different system with higher expressibility than the object system. In the informal example considered above we can formally characterize — as a meta-level state — expressions like “\(p(x)\) where \(x\) is a variable”, as well as “... \(x\) is a constant”, and perform meta-level transitions. As it is shown below, the use of such a meta-level representation is also a powerful tool when confluence under invariant and/or modulo equivalence is considered, which otherwise has been problematic, even for the logical subset of CHR under a logic based semantics.

## 5 Invariants and modulo equivalence

It can be argued that invariants and confluence modulo equivalence are important from a practical point of view. In this section we give definitions and examples and later we consider how to prove the properties. Most CHR programs are developed with a particular set of initial queries in mind, which reduces the set of reachable states. In 2007, Duck et al. [7] suggested to take such an induced invariant into account and, thus, make a much larger class of programs confluent.

**Definition 1.** A set \(I\) is a state invariant for a relation \(\rightarrow\) if \(x \in I \land x \rightarrow y\) implies \(y \in I\).

Such an invariant may be induced by a set of reachable states from a set of (initial) states \(Q\), i.e., \(I = \{s' \mid s \in Q \land s \rightarrow^* s'\}\).

**Example 5** (Ex. 1 continued). The set-program reflects a tacitly assumed state invariant: only one set-constraint is allowed. If we open up for a query such as

\[
?- \text{item(a), item(b), set([]), set([c])}.
\]
we obtain a collection of different final states, representing different ways of partitioning \{a, b, c\} into two sets. However, this may not be intended, and the relevant invariant \(I_{set}\) must specify that a state includes exactly one set/1 constraint and a series of item/1 constraints.

**Definition 2.** A relation \(\rightarrow\) is **observable confluence** (under invariant \(I\)) if and only if \(\forall x, y, y' \in I: y' \leftarrow^* x \rightarrow^* y' \Rightarrow \exists z \in I: y' \rightarrow^* z \leftarrow^* y'\). We may write \(I\)-observable confluence meaning observably confluent (under invariant \(I\)).

**Example 6** (Ex. 2 continued). Consider the program of Example 2 together with an invariant \(I_{>0}\) induced by initial states with a single atom \(p(n)\) where \(n\) is a positive number (not a variable). The program is \(I_{>0}\)-observable confluent, since each forked state \(q(n) \leftarrow p(n) \rightarrow r(n)\) is joinable by rule \(r_3\): \(r(n) \rightarrow q(n)\).

Confluence modulo equivalence is a generalization where forked states must reach equivalent states, rather than a common state. For instance, a program may produce redundant data structures such as representing sets as lists, and the equivalence states that the order of the elements does not matter. Confluence modulo equivalence was first considered for CHR in 2014 by Christiansen and Kirkeby [4].

**Definition 3.** A relation \(\rightarrow\) is **confluent modulo an equivalence** \(\approx\) if and only if \(\forall x, y, x', y': y' \leftarrow^* x' \approx x \rightarrow^* y' \Rightarrow \exists z, z': y' \leftarrow^* z' \approx z \leftarrow^* y'\).

**Example 7** (Ex. 1, 5 continued). The set-program is supposed to produce one set representation, and we introduce a state equivalence \(\approx_{set}\) reflecting the redundant data structures. Two states are equivalent if they have the same item-constraints and their respective set-constraint \(set(L_1)\) and \(set(L_2)\) are such that \(L_1\) and \(L_2\) are permutations of each other.

We generalize confluence modulo equivalence and observable confluence as follows.

**Definition 4.** A relation \(\rightarrow\) is **I-observable confluence modulo an equivalence** \(\approx\) if and only if \(\forall x, y, x', y': y' \leftarrow^* x' \approx x \rightarrow^* y' \Rightarrow \exists z, z': y' \leftarrow^* z' \approx z \leftarrow^* y'\).

Both observable confluence, confluence modulo equivalence and classic confluence are special cases of this definition.

Huet [14] provided a pair of local properties for showing terminating programs confluent modulo equivalence; we extend these with an invariant as follows.

**Definition 5.** A rewriting system \(\rightarrow\) is **locally I-observable confluence modulo** \(\approx\) if and only if it has the following \(\alpha\) - and \(\beta\)-properties.

- \(\alpha: \forall x, y, y' \in I: y' \leftarrow x \rightarrow y' \Rightarrow \exists z, z' \in I: y' \rightarrow z' \approx z \leftarrow y'\)
- \(\beta: \forall x, y, y' \in I: y' \approx x \rightarrow y' \Rightarrow \exists z, z' \in I: y' \rightarrow z' \approx z \leftarrow y'\)

We refer to structures of the form \(y' \leftarrow x \rightarrow y'\) as \(\alpha\)-corners and those of the form \(y' \approx x \rightarrow y'\) as \(\beta\)-corners.

**Theorem 1** (obs. confl. mod. equivalence). *A terminating relation \(\rightarrow\) is I-observable confluency modulo \(\approx\) if and only if it is locally I-observable confluence modulo \(\approx\).*

In the special cases where equivalence is \('=\)', the \(\beta\)-property trivially holds and when, furthermore, the invariant is unrestricted the \(\alpha\)-property reduces to local confluence.

Both the invariant and the equivalence relation may be tailored for the individual program. By nature, they are meta level properties that in general cannot be expressed in its own system: the state itself is implicit and properties such as groundness (or certain arguments bound to be variables) cannot be expressed in a logic-based semantics for CHR.
6 Proving observable confluence

As mentioned, observable confluence was introduced by Duck et al. [7]. They suggested methods of proving this property for logic CHR programs using a logic based semantics and as a direct continuation of the logic subsumption principle.

Firstly, they construct the set of critical corners based from the program rules as explained in Section 4. Typically, the states in these corners do not satisfy the invariant (a rule typically includes variables, contradicting groundness), and the next step is to characterize a set of “minimal extensions” of each critical corner such that 1) the states of such an extension satisfies the invariant, 2) the set of all such extensions subsumes (by substitution and adding constraints) all non-trivial object corners. Proving observable confluence amounts to show joinability of such extension, using the standard transition relation for CHR.

There are two problems in this approach, first of all there is no formal representation of the invariant that allows to take it into account when reasoning formally about joinability, and – more importantly – as also noticed by Duck et al. [7], quite often there is an infinite number of such extensions. This happens even for an intuitively simple invariant such as requiring ground states. We can demonstrate this phenomenon for the program of Example 2.

Example 8. Consider the program of Example 2; it is not confluent as its single critical corner \( q(X) \leftarrow p(X) \rightarrow r(X) \) is not joinable (the built-in store is always true and thus omitted). However, adding an invariant “reachable from an initial state \( p(n) \) where \( n \) is an integer” makes it confluent. We indicate the smallest set of corners found by minimal extensions of the critical corner; the dotted transitions prove each of them joinable:

\[
\begin{array}{c}
\cdots \quad p(-1) \quad \cdots \quad p(0) \quad \cdots \quad p(1) \quad \cdots \\
q(-1) \cdots q(0) \cdots q(1) \cdots q(2) \cdots \\
\end{array}
\]

This set subsumes exactly the set of all non-trivial object level corners. These corners and their proofs of joinability obviously fall in two groups of similar shapes, but there is no way to construct a finite set (of, say, one or two elements) of critical corners in CHR, that subsumes all \( I \) object corner.

To avoid this problem, Christiansen and Kirkeby [15] suggests to describe critical corners in a more powerful meta-language rather than using CHR itself, inspired by earlier work on metaprogramming in logic programming. Each term of CHR is here named by a ground term, specifically, variables named by ground constants. A variable in such a term is thus a meta-variable, which may be covered by a meta-level constraint. For example, the meta-level term \( p(x) \) WHERE type(const, X) subsumes all object level (i.e., CHR) atoms whose predicate is \( p/1 \) and whose argument is a constant, i.e., \( p(a), p(b), \ldots, p(1), \ldots \). The authors define a notion of abstract simulation making precise what it means for meta-level transitions and corners to subsume \(^4\) sets of object level transitions and corners. Built-in predicates are reflected at the meta-level, such that, say \( n<0 \) restricts \( n \) to names of terms \( t \) that satisfy the object level condition \( t<0 \), i.e., \( n \) is limited names of number constraints less that zero.

Example 9. (Continuing Ex. 2, 8) The invariant is formalized at the meta-level as states of the form \( \{\{\text{pred}(n)\}, \text{true}\} \) WHERE type(int, n) where pred is one of \( p, q \) and \( r \). Below is shown the two joinable critical meta-level corners that can be shown to subsume all non-trivial object level corners.

\(^4\)[15] use the terminology of a meta-level term covering object level notions.
This example illustrates an additional technique called splitting: First a meta-level corner is produced in the classical way, considering how rules can overlap; this yields a common meta-level ancestor state p(x) where type(int,x) and the other states as above containing q(x), resp. r(x). This meta-level corner is in itself not joinable as no single rule can apply, but turning it into two meta-level corners, each joinable and together subsuming the same set of object-level corners, proves observable confluence.

7 Proving Confluence modulo equivalence

Confluence modulo equivalence has been studied since the first half of the 20th century in a variety of contexts; see, e.g., [5, 14] for an overview. It was introduced and motivated for CHR by [4], also arguing that invariants are important for specifying meaningful equivalences. An in-depth theoretical analysis, including the use of the ground representation, is given by [5] in relation the Prolog-related semantics mentioned above.

To show confluence modulo equivalence for terminating CHR programs, two types of critical corners must be constructed: critical α-corners which are the standard critical corners constructed by rule overlap as above, and critical β-corners of the form y′ ≈ x → y′, cf. Definition 5. As before the critical β-corners must subsume all non-trivial object-level β-corners. The meta-level language described above is also suitable for describing state equivalences [15].

Example 10 (Ex. 1 continued). The set-program of Example 1 is observably confluent modulo ≃_{set} (Ex. 7) under invariant I_{set} (Ex. 5) since the critical α-corner with two set-constraints does not subsume I_{set} corners and both the other critical α-corner and the critical β-corner are joinable modulo ≃_{set}, see below. The meta-level constraint its/1 constrains its argument to a set of item-constraints and perm/2 constrains the arguments to a pair of permuted lists.

The α-corner:
\[ \{\text{set}([X_1|L]), \text{item}(X_2)\} \cup C \text{ where } \text{its}(C) \quad \text{r} \rightarrow \quad \{\text{set}([X_2,X_1|L])\} \cup C \text{ where } \text{its}(C) \]
\[ \{\text{item}(X_1), \text{set}(L), \text{item}(X_2)\} \cup C \text{ where } \text{its}(C) \quad \text{r} \rightarrow \quad \{\text{set}([X_1,X_2|L])\} \cup C \text{ where } \text{its}(C) \]

The β-corner:
\[ \{\text{set}([L_2]), \text{item}(X)\} \cup C \text{ where } \text{perm}(L_1,L_2) \land \text{its}(C) \quad \text{r} \rightarrow \quad \{\text{set}([X|L_2])\} \cup C \text{ where } \text{perm}(L_1,L_2) \land \text{its}(C) \]
\[ \{\text{set}(X|L_1)\} \cup C \text{ where } \text{perm}(L_1,L_2) \land \text{its}(C) \quad \text{φ} \]
\[ \{\text{set}([X|L_1])\} \cup C \text{ where } \text{perm}(L_1,L_2) \land C \land \text{its}(C) \]

A recent paper [13] attempts to handle (observable) confluence modulo equivalence within the logic-based semantics, along the lines of Duck et al [7]. However, this implies the mentioned
problems of infinitely many proof cases, which seems to be inherent in relying on pure logic based subsumption without having the enhanced expressibility provided by a suitable meta-level representation.

8 Future work

We have given an overview of classic and recent results for confluence in CHR. The classic results provide a theoretical foundation and the recent results on observable confluence and modulo equivalence points towards more practical applications of these notions in a programming context.

There exist methods for automatic check of confluence for CHR [16] in a strictly logical setting, and in our own work we are developing similar methods for automatic or semi-automatic proofs of observable confluence modulo equivalence. Naturally, invariants and state equivalences may involve undecidability.

References


