# Rule-based Graph Repair using Minimally Restricted Consistency-Improving Transformations

Alexander Lauer

October 3, 2022

**Abstract**

# Contents

# 1 Preliminaries

**Definition 1.1** (**subgraph**)**.** *Let $G_1$ and $G_2$ be graphs. The graph $G_2$ is called a* subgraph *of $G_1$ if an injective morphism $f : G_2 \to G_1$ exists. We use the notation $G_2 \subseteq G_1$ if $G_2$ is a subgraph of $G_1$ and $G_2 \subset G_1$ if $f$ is not bijective.*

**Definition 1.2** (**minimal uppergraph**)**.** *Let $G_1$ and $G_2$ be graphs with $G_1 \subseteq G_2$. A graph $C$ is called an* minimal uppergraph *of $G_1$ w.r.t $G_2$, if $G_1 \subset C \subseteq G_2$ and no graph $C' \subset C$ with $G_1 \subset C \subseteq G_2$ exists. The set of minimal uppergraphs of $G_1$ w.r.t. $G_2$ is denoted by $\mathcal{U}(G_1, G_2)$. If $G_1 = G_2$, we set $\mathcal{U}(G_1, G_2) = \{G_1\}$.*

**Definition 1.3** (**overlap**)**.** *Let $G$ and $G'$ be graphs. A graph $H$ is called an* overlap *of $G$ and $G'$ if morphisms $p : G \hookrightarrow H$ and $p' : G' \hookrightarrow H$ such that $p$ and $p'$ are jointly surjective. The set of all overlaps of $G$ and $G'$ is denoted by $\mathrm{ol}(G, G')$.*

**Definition 1.4** (**overlap at morphism**)**.** *Let $C, G$ and $C'$ with $C \subset C'$ be graphs and $p : C \hookrightarrow G$ a morphism. A graph $H$ is called an* overlap of $G$ and $C'$ at $p$ *if a morphism $p' : C' \hookrightarrow H$ with $p'|_C = p$ exists. The set of all overlaps of $G$ and $C'$ at $p$ is denoted by $\mathrm{ol}_p(G, C')$.*

**Definition 1.5** (**partial morphism**)**.** *Let $f : G_1 \to G_2$ and $g : G_3 \to G_4$ be morphisms. The morphism $g$ is called a* partial morphism *of $f$ if $G_3 \subseteq G_1$, $G_4 \subseteq G_2$ and $f|_{G_3} = g$.*

**Definition 1.6** (**nested graph condition**)**.** *A* graph condition *over a graph $C_0$ is inductively defined as follows:*

- *true is a graph condition over every graph.*

- *$\exists(a : C_0 \hookrightarrow C_1, d)$ is a graph condition over $C_0$ if $a$ is a injective graph morphism and $d$ is a graph condition over $C_1$.*

- *$\neg d$ is a graph condition over $C_0$ if $d$ is a graph condition over $C_0$.*

- *$d_1 \wedge d_2$ and $d_1 \vee d_2$ are graph conditions over $C_0$ if $d_1$ and $d_2$ are graph conditions over $C_0$.*

*Conditions over the empty graph $\emptyset$ are called* constraints*. Every injective morphism $p : C_0 \hookrightarrow G$ satisfies* true*. An injective morphism $p$ satisfies $\exists(a : C_0 \hookrightarrow C_1, d)$ if there exists an injective morphism $q : C_1 \hookrightarrow G$ such that $q \circ a = p$ and $q$ satisfies $c$. An injective morphism satisfies $\neg d$ if it does not satisfy $d$, it satisfies $d_1 \wedge d_2$ if it satisfies $d_1$ and $d_2$ and it satisfies $d_1 \wedge d_2$ if it satisfies $d_1$ or $d_2$. A graph $G$ satisfies a constraint $c$, $G \models c$, if $p : \emptyset \hookrightarrow G$ satisfies $c$. We use the abbreviation $\forall(a : C_0 \hookrightarrow C_1, d) := \neg\exists(a : C_0 \hookrightarrow C_1, \neg d)$.*

*The* nesting level nl *of a condition is defined as $\mathrm{nl}(true = 0$ and $\mathrm{nl}(\exists(a : P \to Q, d)) := \mathrm{nl}(d) + 1$.*

**Definition 1.7** (**alternating quantifier normal form (ANF)**[1]). *A graph condition* $c$ *is in* alternating normal form *(ANF) if it is of the form*

$$c = Q(a_1 : C_0 \hookrightarrow C_1, \overline{Q}(a_2 : C_1 \hookrightarrow C_2, Q(a_3 : C_2 \hookrightarrow C_3, \overline{Q}(a_4 : C_3 \hookrightarrow C_4, \ldots))))$$

*with* $Q \in \{\exists, \forall\}$ *and* $\overline{Q} = \exists$ *if* $Q = \forall$, $\overline{Q} = \forall$ *if* $Q = \exists$.

## 2 partial consistency improving

### 2.1 extended alternating quantifier normal form

**Definition 2.1** (**extended alternating quantifier normal form**). *A condition* $c$ *is in* extended alternating quantifier normal form *(EANF) if it is in ANF, universally bound and ends with a condition of the form* $\exists(a_k : C_k \hookrightarrow C_{k+1}, e)$ *with* $e \in \{\text{true}, \text{false}\}$.

**Lemma 2.2.** *Any constraint in ANF can be transformed into an equivalent constraint in EANF.*

*Proof.* Let $c$ be a constraint in ANF. If $c$ is universally bound and ends with a condition of the form $\exists(a_k : C_k \hookrightarrow C_{k+1}, e)$ with $e \in \{\text{true}, \text{false}\}$, $c$ is already in EANF. We construct the equivalent constraint in EANF by two steps and show that, after each step, the derived constraint is equivalent to $c$.

1. If $c = \exists(a_1 : \emptyset \hookrightarrow C_0, e)$ is existentially bound, we show that $c$ is equivalent to $d := \forall(a_0 : \emptyset \hookrightarrow \emptyset, c)$. Let $G$ be a graph.

   "$\Longrightarrow$": Let $p : \emptyset \hookrightarrow G$ be a morphism with $p \models c$, therefore a morphism $q : C_0 \to G$ with $q \models e$ and $p = q \circ a_0$ exists. Then, $p \models d$, since $p$ is the only morphism from $\emptyset$ to $G$ and $p = p \circ a_1$ and $p \models c$.

   "$\Longleftarrow$": Let $p : \emptyset \hookrightarrow G$ be a morphism with $p \models d$, therefore all morphisms $q : \emptyset \hookrightarrow G$ with $p = q \circ a_0$ satisfy $c$. With $p = p \circ a_0$, $p \models c$ follows immediately.

2. If $c$ ends with a condition of the form $d = \forall(a_k : C_k \hookrightarrow C_{k+1}, e)$ with $e \in \{\text{true}, \text{false}\}$. We show that $d$ is equivalent to $d' = \forall(a_k : C_k \hookrightarrow C_{k+1}, \exists(a_{k+1} : C_{k+1} \hookrightarrow C_{k+1}, e))$. Let $G$ be a graph.

   (a) If $e = \text{true}$. Let $p : C_k \hookrightarrow G$ be a morphism, we show that $p \models d$ and $p \models d'$. Since every morphism satisfies $\text{true}$, a morphism $q : C_k \hookrightarrow G$ with $p = a_k \circ q$ and $p \not\models e$ cannot exist, therefore $p \models d$. The morphism $a_{k+1}$ has to be the identity on $C_{k+1}$ and therefore for every $q : C_k \hookrightarrow G$ with $p = a_k \circ q$ is holds that $q = a_{k+1} \circ q$ and $q \models e$. It follows that $p \models d'$.

   (b) If $e = \text{false}$:

   "$\Longrightarrow$": Let $p : C_k \hookrightarrow G$ be a morphism with $p \models d$. Since $e = \text{false}$, no morphism $q : C_{k+1} \hookrightarrow G$ with $p = a_k \circ q$ exists. Therefore, no morphism $q : C_{k+1} \hookrightarrow G$ with $p = a_k \circ q$ and $q \not\models \exists(a_{k+1} : C_{k+1} \hookrightarrow C_{k+1}, e)$ exists. It follows, that $p \models d'$.

"$\Longleftarrow$": Let $p : C_k \hookrightarrow G$ be a morphism with $p \models d'$. Since no morphism satisfies false, no morphism $q : C_{k+1} \hookrightarrow G$ satisfies $\exists(a_{k+1} : C_{k+1} \hookrightarrow C_{k+1}, e)$. Hence, there does not exist a morphism $q : C_{k+1} \hookrightarrow G$ with $p = q \circ a_k$ and $p \models d$ follows.

$\square$

## 2.2 conditions up to layer

**Definition 2.3** (**Layer of a subcondition**). *Let $c$ be a condition and $d$ a subcondition of $c$. The layer of $d$ is defined as* $\mathrm{lay}(d) := \mathrm{nl}(c) - \mathrm{nl}(d) - 1$.

**Definition 2.4** (**substitution at layer**). *Let $c = Q(a : C_0 \hookrightarrow C_1, d)$ be a condition in ANF, such that the subcondition of $c$ with layer $0 \leq k \leq \mathrm{nl}(c)$ is a condition over $C_k$. Let $e$ be a condition over $C_k$. The substitution in $c$ at layer $k$ with $e$, $\mathrm{sub}(k, c, e)$, is recursively defined as:*

1. *If $k = 0$:*
$$\mathrm{sub}(0, c, e) := e$$

2. *If $k > 0$:*
$$\mathrm{sub}(k, c, e) := Q\big(a : C_0 \hookrightarrow C_1, \mathrm{sub}(k - 1, d, e)\big)$$

**Definition 2.5** (**Condition up to layer**). *Let $c$ be a condition in ANF and $d$ be the subcondition of $c$ at layer $0 \leq k \leq \mathrm{nl}(c)$. The condition up to layer $k$ of $c$, $\mathrm{cond}(k, c)$, is defined as*

$$\mathrm{cond}(k, c) := \begin{cases} \mathrm{sub}(k, c, \mathsf{true}) & \text{, if } k = 0 \vee d \text{ is existentially bound} \\ \mathrm{sub}(k, c, \mathsf{false}) & \text{, if } d \text{ is universally bound.} \end{cases}$$

**Definition 2.6** (**Satisfaction up to layer**). *Let $G$ be a graph and $c$ be a condition over $C_0$. A morphism $p : C_0 \hookrightarrow G$ satisfies $c$ up to layer $k$, $p \models_k c$, if*

$$p \models \mathrm{cond}(k, c).$$

*A graph $G$ satisfies a constraint $c$ up to layer $k$, $G \models_k c$, if $q : \emptyset \hookrightarrow G$ satisfies $\mathrm{cond}(k, c)$. The biggest $k$ with $G \models_k c$ such that no $j > k$ with $G \models_j c$ exists is denoted by $c_{\max}$.*

**Lemma 2.7.** *Let $G$ be a graph $p : C_0 \hookrightarrow G$ a morphism and $c$ a condition over $C_0$ in ANF with $p \models_k c$. If the subcondition $d = Q(a_k : C_{k-1} \hookrightarrow C_k, e)$ of $c$ at layer $k$ is universally bound, then for any condition $f$ over $C_k$ it holds that*

$$p \models \mathrm{sub}(k, c, f).$$

5

*Proof.* Let $k$ be the smallest number such that $p \models_k c$ and the subcondition of $c$ with layer $k$ is universally bound, let $d = \forall(a_k : C_{k-1} \hookrightarrow C_k, e)$ be this subcondition. Let $q : G_{k-1} \to G$ be a morphism such that $q \models \forall(a_k : C_{k-1} \hookrightarrow C_k, \mathsf{false})$. This must exist, since $p \models_k c$ and $k$ is the smallest number such that $p \models_k v$ and the subcondition of $c$ with layer $k$ is universally bound.

Therefore, there does not exist a morphism $q' : C_k \to G$ with $q = q' \circ a_k$. Hence, for every condition $f$ over $C_k$ a morphism $q' : C_k \to G$ with $q \not\models f$ and $q = q' \circ a_k$ cannot exist. It follows immediately that $q \models \forall(a_k : C_{k-1} \hookrightarrow C_k, f)$. $\qquad\square$

**Lemma 2.8.** *Let $G$ be a graph, $p : C_0 \to G$ a morphism and $c$ a condition over $C_0$ in ANF with $p \models_k c$. If the subcondition $d$ of $c$ with $\mathrm{lay}(d) = k$ is universally bound,*

$$p \models_k c \implies p \models c.$$

*Proof.* Follows immediately by using lemma 2.7 and setting $f$ to the subcondition of $c$ with layer $k+1$. $\qquad\square$

**Lemma 2.9.** *Let $c$ be a condition in ANF over $C_0$ and $p : C_0 \hookrightarrow G$ a morphism with $p \models_k c$. Let $d = Q(a_{k+2} : C_{k+1} \hookrightarrow C_{k+2}, e)$ be the subcondition of $c$ with layer $k+2$. There does exist a graph $C_{k+1} \subseteq C' \subseteq C_{k+2}$ such that*

$$p \models \mathrm{sub}\big(k+1, c, Q(a_{k+2} : C_{k+1} \hookrightarrow C', f)\big)$$

*with $f$ being a $\overline{Q}$ bound condition over $C'$.*

*Proof.* If $p \models c$, we can choose $C' = C_{k+2}$ and $f = e$.

If $p \not\models c$, there does not exists a $j$ with $p \models_j c$ and the subcondition of $c$ with layer $j$ is universally bound and $Q = \exists$ follows immediately. We choose $C' = C_{k+1}$ and $f = \mathsf{true}$. Let $q : C_k \to G$ with $p = q \circ a_k \circ \ldots \circ a_1$ and $q \circ \ldots \circ a_\ell$ satisfying the condition up to $\ell - k$ of the subcondition of $c$ at layer $\ell$ for all $0 \le \ell \le k$. This morphism must exists since $p \models_k c$ and $p \not\models c$. Let $q' : C_{k+1} \to G$ be a morphism with $q = q' \circ a_{k+1}$. Since $C' = C_{k+1}$, the morphism $a'_{k+2}$ has to be the identity and therefore $q' = q' \circ a'_{k+2}$. It follows that $q' \models \exists(a'_{k+2} : C_{k+1} \hookrightarrow C', \mathsf{true})$ and therefore $p \models \mathrm{sub}\big(k+1, c, Q(a_{k+2} : C_{k+1} \hookrightarrow C', f)\big)$. $\qquad\square$

**Definition 2.10** (**partial condition**). *Let $c$ be a condition in ANF over $C_0$. Let $d$ be the subcondition of $c$ at layer $k+1$. The* partial condition *of $c$ at layer $k$ with $C'$, $\mathrm{part}(k, c, C')$ is defined as:*

1. *If $d$ is universally bound, let $e = \exists(a : C_{k+1} \hookrightarrow C_{k+2}, f)$ be the subcondition of $c$ at layer $k+2$ with $C_{k+1} \subseteq C' \subseteq C_{k+2}$:*

$$\mathrm{part}(k, c, C') := \mathrm{sub}\big(k+2, c, \exists(a : C_{k+1} \hookrightarrow C', \mathit{true})\big)$$

2. *If $d = \exists(a : C_k \hookrightarrow C_{k+1}, f)$ is existentially bound with $C_k \subseteq C' \subseteq C_{k+1}$:*

$$\mathrm{part}(k, c, C') := \mathrm{sub}\big(k+1, c, \exists(a : C_{k+1} \hookrightarrow C', \mathit{true})\big)$$

**Definition 2.11 (biggest partially satisfying condition).** *Let $G$ be a graph, $c$ a condition over $C_0$ and $p : C_0 \hookrightarrow G$ a morphism with $p \models_k c$.*

*A partial condition $c = \mathrm{part}(c_{\max}, c, C')$ with $p \models c$ is a* biggest partially satisfying condition *if there does not exist a graph $C' \subset C''$ with $p \models \mathrm{part}(c_{max}, c, C'')$. The graph $C'$ is called a* biggest partially satisfying graph.

*The set of biggest partially satisfying conditions of $c$ is denoted by $\mathcal{P}_c^G$.*

*The set of all biggest partially satisfying graphs is denoted by $\mathcal{G}_c^G$.*

## 2.3 minimal consistency improving

**Definition 2.12 (number of violations).** *Let $G$ be a graph and $c$ a constraint in EANF. The* number of violations $\mathrm{nvc}(j, G)$ *at layer $j$ in $G$ is defined as:*

1. *If $j < c_{\max}$:*
$$\mathrm{nvc}(j, G) := 0$$

2. *If $j = c_{\max}$, let $d = \forall(a_k : C_j \hookrightarrow C_{j+1}, e)$ be the subcondition of $c$ at layer $j + 1$.*

$$\mathrm{nvc}(j, G) := \sum_{C \in \mathcal{G}_c} \sum_{C' \in \mathcal{U}(C, C_{j+1})} |\{q \mid q : C_{j+1} \hookrightarrow G \wedge q \not\models \mathrm{part}(1, e, C')\}|$$

3. *If $j > c_{\max}$:*
$$\mathrm{nvc}(j, G) := \infty$$

**Definition 2.13 (minimal consistency improving).** *Let a graph $G$, a rule $r$ and a constraint $c$ in ANF be given.*

*A transformation $G \Longrightarrow_{r,m} H$ is called* minimal consistency improving, *if*

$$\mathrm{nvc}(k, H) < \mathrm{nvc}(k, G)$$

*for any $0 \leq k \leq \mathrm{nl}(c)$. A rule $r$ is called* minimal consistency improving, *if all of its applications to graphs $G$ with $G \not\models c$ are.*

**Lemma 2.14.** *Let a graph $G$, a morphism $p : C_0 \to G$ and a constraint $c$ in ANF over $C_0$ with $p \models_k c$ be given. Then, $p \models_j c$ for all $j < k$ such that the subcondition of $c$ at layer $j$ is existentially bound.*

*Proof.* 1. The subcondition of $c$ at layer $k$ is existentially bound: If an $j < k$ with $p \models_j c$ exists such that the subcondition of $c$ at layer $j$ is universally bound, let $j_1$ be the smallest of these. With lemma 2.7 follows that $p \models_{j_2} c$ for all $j_1 < j_2$. Let $\ell < j_1$, such that the subcondition of $c$ at layer $\ell$ is existentially bound and let $d = \exists(a_\ell : C_\ell \hookrightarrow C_{\ell+1}, e)$ be the condition up to layer $j_1 - \ell$ of the subcondition of $c$ at layer $\ell$. Since $\ell < j_1$, a morphism $q : C_\ell \to G$ with $q \models d$ must exists and therefore a morphism $q' : C_{\ell+1} \to G$ with $q = q' \circ a_k$ must exists. If follows that $q \models \exists(a_\ell : C_\ell \hookrightarrow C_{\ell+1}, \mathsf{true})$ and with that $p \models_\ell c$.

2. The subcondition of $c$ at layer $k$ is universally bound: With lemma 2.7 follows that $p \models_{k+1} c$. Since $c$ is in ANF, 1. can be applied to $k+1$. $\qquad\square$

**Theorem 2.1.** *Let a graph $G$, a rule $r$ and a constraint $c$ in ANF be given. Let $k < \mathrm{nl}(c)$ be the biggest number, such that $G \models_k c$. A transformation $G \Longrightarrow_{r,m} H$ is minimal consistency improving if $G \models_j c$ and $k < j$.*

*Proof.* No $\ell > k$ with $G \models_\ell c$ exists and $G \models_k c$. Hence, $\mathrm{nvc}(k, G) > 0$ and $\mathrm{nvc}(k, G) \neq \infty$. Since $j > k$, $\mathrm{nvc}(k, H) = 0$ and it follows immediately that the transformation is minimal consistency improving. $\qquad\square$

**Definition 2.15** (**direct minimal consistency improving**). *Let $G$ be a graph, $r$ a plain rule and $c$ a constraint in EANF. Let $d = \forall(a_k : C_{k-1} \hookrightarrow C_k, e)$ be the condition at layer $k = \mathrm{c_{max}} + 1 \leq \mathrm{nl}(c)$ of $c$ and*

$$\mathbf{G} := \bigcup_{C' \in \mathcal{G}_c^G} \mathcal{U}(C', C_{k+1})$$

*be the set of all minimal upper-graphs of all biggest partially satisfying graphs. A transformation $t : G \Longrightarrow_{r,m} H$ is called* direct minimal consistency improving *if $G \models_{k-1} c$ and equations (2.1), (2.2) and (2.3) hold.*

*Every occurrence of $C_k$ in $G$ that satisfies $\mathrm{part}(1, e, C')$ for any $C' \in \mathbf{G}$ still satisfies $\mathrm{part}(1, e, C')$ in $H$.*

$$\forall p : C_k \hookrightarrow G\Big( \bigwedge_{C' \in \mathbf{G}} \big(p \models \mathrm{part}(1, e, C') \wedge \mathrm{tr}_t \circ p \text{ is total}\big)$$
$$\implies \mathrm{tr}_t \circ p \models \mathrm{part}(1, e, C')\Big) \tag{2.1}$$

*Every new inserted occurrence of $C_k$ by $t$ satisfies $\mathrm{part}(1, e, C')$ for all $C' \in \mathbf{G}$.*

$$\forall p' : C_k \hookrightarrow H\Big(\neg \exists p : C_k \hookrightarrow G(p' = \mathrm{tr}_t \circ p) \implies \big( \bigwedge_{C' \in \mathbf{G}} p' \models \mathrm{part}(1, e, C')\big)\Big) \tag{2.2}$$

*At least one occurrence of $C_k$ in $G$ that does not satisfy $\mathrm{part}(1, e, C')$, for any $C' \in \mathbf{G}$, either has been destroyed by $t$ or satisfies $\mathrm{part}(1, e, C')$ in $H$.*

$$\exists p : C_k \hookrightarrow G\Big( \bigvee_{C' \in \mathbf{G}} \big(p \not\models \mathrm{part}(1, e, C') \wedge \mathrm{tr}_t \circ p \text{ is not total}$$
$$\vee (\mathrm{tr}_t \circ p \text{ is total } \wedge \mathrm{tr}_t \circ p \models \mathrm{part}(1, e, C')))\Big) \tag{2.3}$$

**Lemma 2.16.** *Let a graph $G$, a constraint $c$ and a direct minimal improving transformation $t : G \Rightarrow_{r,m} H$ w.r.t. $c$ be given. Then, $t$ is also a minimal improving transformation.*

*Proof.* Let $G$ be a graph with $k = \mathrm{c_{max}}$ and $G \models \mathrm{part}(k, c, C)$ with $\mathrm{part}(k, c, C) \in \mathcal{P}_c^G$. Let $d$ be the subcondition of $c$ at layer $k + 1$.

1. We show that equations (2.1) and (2.2) imply that $\mathrm{nvc}(k, H) \leq \mathrm{nvc}(k, G)$. Assume that $\mathrm{nvc}(k, H) > \mathrm{nvc}(k, G)$. Therefore, a morphism $p : C_k \hookrightarrow G$ with $p \not\models \mathrm{part}(1, d, C')$ for any $C' \in \mathcal{U}(C, C_{k+1})$ exists, such that either 1a or 1b is satisfied.

   (a) There does exist a morphism $q' : C_k \hookrightarrow G$ with $q' \models \mathrm{part}(1, d, C')$ and $p = \mathrm{tr}_t \circ q'$.

   (b) There does not exist a morphism $q : C_k \hookrightarrow G$, such that $p = \mathrm{tr}_t \circ q$.

   This is a contradiction, if 1a is satisfied, $q'$ does not satisfy equation (2.1) and if 1b is satisfied $q$ does not satisfy equation (2.2).

2. Since (2.3) is satisfied, a morphism $p : C_k \hookrightarrow G$ with $p \not\models \mathrm{part}(1, d, C')$, such that either $\mathrm{tr} \circ p$ is total and $p \models \mathrm{part}(1, d, C')$ or $\mathrm{tr} \circ p$ is not total exists, for any $C' \in \mathbf{G}$. In both cases the following holds

$$p \in \{q \mid q : C_k \hookrightarrow G \wedge q \not\models \mathrm{part}(1, e, C')\} \wedge$$
$$\mathrm{tr} \circ p \notin \{q \mid q : C_{k+2} \hookrightarrow H \wedge q \not\models \mathrm{part}(1, e, C')\}.$$

   With that and 1 it follows that

$$|\{q \mid q : C_k \hookrightarrow G \wedge q \not\models \mathrm{part}(1, e, C')\}| < |\{q \mid q : C_{k+2} \hookrightarrow H \wedge q \not\models \mathrm{part}(1, e, C')\}|.$$

With 1 and 2 follows that $\mathrm{nvc}(k, G) < \mathrm{nvc}(k, G)$ and therefore $t$ is a minimal improving transformation. $\qquad \square$

## 3 application condition

**Definition 3.1 (extended overlap).** *Let $G$ and $C_0 \subseteq C_1$ be graphs. Let $C$ be an overlap of $C_0$ and $G$ with the overlap morphism: $q : C_0 \hookrightarrow C$.*

*Let $p = L \hookleftarrow K \hookrightarrow R$ be a plain rule with $K = C_0$, $L = K$ and $R = C_1$. The graph $H$, derived by the transformation*

$$C \Longrightarrow_{p,q} H$$

*is called the* extended overlap *of $C$ with $C_1$. The extended overlap of an overlap $C$ with an graph $C_1$ is denoted by $\mathrm{eol}(C, C_1)$.*

**Lemma 3.2.** *Let graphs $G$, $C_0 \subset C_1$ and an overlap $C$ of $G$ and $C_0$ be given. Then, $\mathrm{eol}(C, C_1)$ is an overlap of $G$ and $C_1$ and $C \subset \mathrm{eol}(C, C_1)$.*

*Proof.* The graph $\mathrm{eol}(C, C_1)$ is constructed by the transformation $C \Longrightarrow_{p,q} \mathrm{eol}(C, C_1)$ with $p = C_0 \hookleftarrow C_0 \hookrightarrow C_1$. Since the comatch $n : C_1 \hookrightarrow \mathrm{eol}(C, C_1)$ exists, $\mathrm{eol}(C, C_1)$ is an overlap of $G$ and $C_1$. Because $L = K$, the transformation does not delete any elements and $C \subset \mathrm{eol}(C, C_1)$ follows. $\qquad \square$

9

**Definition 3.3 (overlap shift).** *Let $r = L \hookleftarrow K \hookrightarrow R$ be a plain rule, $C$ a graph and $C'$ an overlap of $C$ and $L$ with morphisms $p : L \hookrightarrow C'$, $k : K \hookrightarrow C'$, $c : C \hookrightarrow C'$ and the partial morphism $q : R \hookrightarrow C'$. We define*

$$
\begin{aligned}
D := \{e \in C' \,|&(\exists e' \in L : p(e') = e \\
&\vee \exists e' \in R : q(e') = e) \\
&\wedge \exists e' \in C : c(e') = e\}
\end{aligned} \tag{3.1}
$$

*Let $r = L \hookleftarrow K' \hookrightarrow R$ be the rule with*

$$
K' := K \cup D
$$

*The graph $H$ derived by the transformation $G \Rightarrow_{r,p} H$ is called the* overlap shifted graph *of $C'$. The overlap shifted graph of an graph $C$ is denoted by $\mathrm{ols}(C)$.*

**Definition 3.4.** *Let $r = L \hookleftarrow K \hookrightarrow R$ be a plain rule and $c$ a constraint in EANF. Let $d = \forall(a_k : C_{k-1} \hookrightarrow C_k, \exists(a_{k+1} : C_k \hookrightarrow C_{k+1}, e))$ be the subcondition of $c$ at layer $k$ with $k = 2i - 1$ for an $i \in \mathbb{N}$. The application condition $\mathrm{ap}_k$ of the condition at layer $k$ of $c$ with $C' \in \mathcal{U}(C_k, C_{k+1})$ is defined as:*

$$
\begin{aligned}
\mathrm{ap}(k, C') := &\Big( \bigvee_{P \in \mathrm{ol}(L, C_k)} \mathrm{nex}(P, C') \wedge (\mathrm{rep}(P, C') \vee \mathrm{del}(P, C'))\Big) \wedge \\
&\Big( \bigwedge_{P \in \mathrm{ol}(L, C_k)} \mathrm{ex}(P, C')\Big) \wedge \\
&\Big( \bigwedge_{P \in \mathrm{ol}(R, C_{k+1})} \mathrm{rem}(P, C')\Big)
\end{aligned} \tag{3.2}
$$

with

1.

$$
\mathrm{nex}(P, C') := \exists(a : L \hookrightarrow P, \neg\exists(b : P \hookrightarrow \mathrm{eol}(P, C'), \textit{true}))
$$

2.

$$
\mathrm{rep}(P, C') := \mathrm{Left}(\forall(a : R \hookrightarrow \mathrm{ols}(P), \exists(b : \mathrm{ols}(P) \hookrightarrow \mathrm{ols}(\mathrm{eol}(P, C')), \textit{true}), r)
$$

3. *Let $i_1 : L \hookrightarrow P$ and $i_2 : C_k \hookrightarrow P$ be the overlap morphisms of $P$:*

$$
\mathrm{del}(P, C') := \begin{cases} \exists(L \hookrightarrow P, \textit{true}) & , \text{if } i_1(L \setminus K) \cap i_2(C_k \setminus C_{k-1}) \neq \emptyset \\ \textit{false} & , \text{otherwise} \end{cases}
$$

4. *Let $i' : L \hookrightarrow P$ and $i_j : C_j \hookrightarrow P$ be the inclusion morphisms for all $j \leq k$, let $E$ be the set of all existentially bound graphs $C_j$ with $j \leq k$:*

$$
\mathrm{ex}(P, C') := \begin{cases} \neg\exists(L \hookrightarrow P, \textit{true}) & , \text{if } \bigcup_{C_j \in E} \big(i_j(C_j \setminus C_{j-1}) \cap i'(L \setminus K)\big) \neq \emptyset \\ \textit{true} & , \text{otherwise} \end{cases}
$$

5. Let $i' : R \hookrightarrow P$ and $i_j : C_j \hookrightarrow P$ be the inclusion morphisms for all $j \leq k$, let $U$ be the set of all universally bound graphs $C_j$ with $j \leq k$:

$$\mathrm{rem}(P, C') := \begin{cases} \mathrm{Left}(\neg\exists(R \hookrightarrow P, \textit{true}), r) & , \text{ if } \bigcup_{C_j \in U}\big(i_j(C_j \setminus C_{j-1}) \cap i'(R \setminus K)\big) \neq \emptyset \\ \textit{true} & , \text{ otherwise} \end{cases}$$

**Lemma 3.5.** *Let $G$ be a graph, $c$ a constraint in EANF, with $G \not\models c$, and $r$ a plain rule. Let $d = \forall(a_k : C_{k-1} \hookrightarrow C_k, \exists(b : C_k \hookrightarrow C_{k+1}, e))$ be the subcondition of $c$ at layer $\mathrm{c_{max}}$. Then, every application of $(r, \mathrm{ap}(\mathrm{c_{max}}, C'))$ with*

$$C' \in \bigcap_{C \in \mathcal{G}_c^G} \mathcal{U}(C, C_{k+1})$$

*is direct minimal consistency improving.*

*Proof.* $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

**Lemma 3.6.** *Let $G$ be a graph, $c$ a constraint in EANF, with $\mathrm{c_{max}} < \mathrm{nl}(c)$, and $r = L \hookleftarrow K \hookrightarrow R$ a plain rule. Let $d = \forall(a_k : C_{k-1} \hookrightarrow C_k, \exists(b : C_k \hookrightarrow C_{k+1}, e))$ be the subcondition of $c$ at layer $\mathrm{c_{max}} + 1$ and $\mathrm{ap}(\mathrm{c_{max}}, C')$ the application condition constructed by definition 3.4 with $C' \in \mathcal{U}(C, C_{k+1})$ for any $C \in \mathcal{G}_c^G$. The following simplifications can be applied:*

1. *Let $P \in \mathrm{ol}(L, C_k)$. If an injective morphism $p : P \hookrightarrow G$ does not exist, $\mathrm{nex}(P, C')$ can be replaced by false.*

2. *If $(L \setminus K) \cap C_k = \emptyset$, every $\mathrm{del}(P, C')$ can be replaced by false and $\mathrm{ex}(P, C')$ can be replaced by true.*

3. *If $(R \setminus K) \cap C' = \emptyset$, every $\mathrm{rep}(P, C')$ can be replaced by false.*

4. *If 2. and 3. apply, $\mathrm{ap}(k, C')$ can be replaced by false.*

5. *If $(R \setminus K) \cap C_{k+1} = \emptyset$, every $\mathrm{rem}(P, C')$ can be replaced by true.*

*Proof.* Let $m : L \hookrightarrow G$ be the match of the application of $r$. We show the simplifications by showing that the replaced parts of $\mathrm{ap}(\mathrm{c_{max}}, C')$ will be evaluated to the values they have been replaced with, if the required conditions are met. Note that $C_j \subseteq C_{k'}$ for all $0 \leq j \leq k'$ and $k' \in \{0, \ldots, \mathrm{nl}(c)\}$.

- Proof of 1.: Let $P \in \mathrm{ol}(L, C_k)$ with $a : L \hookrightarrow P$, such that no morphism $p : P \hookrightarrow G$ exists. Therefore, no morphism $q : P \hookrightarrow G$ with $m = p \circ a$ exists and $\mathrm{nex}(P, C')$ will be evaluated to false.

- Proof of 2.: Let $(L \setminus K) \cap C_k = \emptyset$ and $P \in \mathrm{ol}(L, C_k)$ with $i' : L \hookrightarrow P$ and $i_j : C_j \hookrightarrow P$ for all $j \leq k$. If follows that $i'(L \setminus K) \cap i_k(C_k \setminus C_{k-1}) = \emptyset$ and with that $\mathrm{del}(P, C') = \mathsf{false}$ for every $P \in \mathrm{ol}(L, C_k)$. Similar it follows that

$$\bigcup_{j \leq k} i_j(C_j \setminus C_{j-1}) \cap i'(L \setminus K) = \emptyset$$

and with that $\mathrm{ex}(P, C') = \mathsf{true}$ for every $P \in \mathrm{ol}(L, C_k)$.

11

- Proof of 3.: work in progress.

- Proof of 4.: Let the simplifications 2. and 3. apply. Every $\mathrm{del}(P, C')$ and every $\mathrm{rep}(P, C')$ have been replaced by false. Therefore, the expression

$$\bigvee_{P \in \mathrm{ol}(L, C_k)} \mathrm{nex}(P, C') \wedge (\mathrm{rep}(P, C') \vee \mathrm{del}(P, C'))$$

  will be evaluated to false and with that $\mathrm{ap}(k, C')$ will also be evaluated to false.

- Proof of 5.: Let $(R \setminus K) \cap C_{k+1} = \emptyset$ and $P \in \mathrm{ol}(R, C_{k+1})$ with the morphisms $i' : R \hookrightarrow P$ and $i_j : C_j \hookrightarrow P$ for all $j \leq k$. If follows that

$$\bigcup_{j \leq k} i'(R \setminus K) \cap i_j(C_j \setminus C_{j-1}) = \emptyset.$$

  Therefore $\mathrm{rem}(P, C') = \mathsf{true}$ for all $P \in \mathrm{ol}(R, C_{k+1})$.

$\square$

## 3.1 potentially minimal improving rules

**Definition 3.7** (**potentially minimal improving rule**). *Let a constraint $c$ and a plain rule $r = L \hookleftarrow K \hookrightarrow R$ be given. The rule $r$ is called* potentially minimal improving *w.r.t $c$ at layer $k$ with $C_k \subset P \subseteq C_{k+1}$ and $k \in \{1, 3, \ldots, \mathrm{nl}(c)\}$, if*

$$(L \setminus K) \cap (C_{k-1} \cup (C_{k+1} \setminus C_k)) = \emptyset \tag{3.3}$$

*and*

$$(R \setminus K) \cap C_k = \emptyset \tag{3.4}$$

*and either 1. or 2. applies.*

1. *The rule $r$ deletes elements of $C_k \setminus C_{k-1}$:*

$$L \subseteq C_k \wedge L \setminus K \neq \emptyset \tag{3.5}$$

   *Then, $r$ is called a* deleting potentially minimal improving rule.

2. *The rule $r$ creates an instance of $P$:*

$$L = C_k \wedge P \subseteq R \tag{3.6}$$

   *Then, $r$ is is called an* inserting potentially minimal improving rule.

**Definition 3.8** (**appl. conditions for potentially minimal improving rules**). *Let a constraint $c$ in EANF and a potentially minimal improving rule $r = L \hookleftarrow K \hookrightarrow R$ w.r.t $c$ at layer $k$ with $C_k \subseteq P \subseteq C_{k+1}$ be given. We define the application condition for $r$ as:*

1. *If $r$ is a deleting potentially minimal improving rule:*

$$\mathrm{ap_{pi}}(j,P) := \begin{cases} \exists\big(a_0 : L \hookrightarrow C_k, \neg\exists(a_1 : C_k \hookrightarrow C_{k+1}, \textsf{true})\big) & \text{, if } j = k \\ \textsf{false} & \text{, if } j \neq k. \end{cases}$$

2. *If $r$ is an inserting potentially minimal improving rule:*

$$\mathrm{ap_{pi}}(j,P) := \begin{cases} \neg\exists(b : L \hookrightarrow P, \textsf{true}) & \text{, if } j = k \\ \textsf{false} & \text{, if } j \neq k \end{cases}$$

**Theorem 3.1.** *Let a graph $G$, a constraint $c$ in EANF, with $G \models \mathrm{part}(c_{\max}, c, C)$ and $\mathrm{part}(c_{\max}, c, C) \in \mathcal{P}_c^G$, and a potentially minimal improving rule $r = L \hookleftarrow K \hookrightarrow R$ at layer $c_{\max}$ with $C_{c_{\max}+1} \subseteq P \subseteq C_{c_{\max}+2}$ be given. Then, $r' = (r, \mathrm{ap_{pi}}(c_{\max}+1, P))$ is a direct minimal consistency improving rule.*

*Proof.* Let $t : G \Longrightarrow_{r',m} H$ be a transformation, $k = c_{\max}+1$ and $e$ be the subcondition of $c$ at layer $k+1$. We show that $t$ is a direct minimal consistency improving transformation. Firstly, we show that equation (2.1) is satisfied. Let $p : C_k \hookrightarrow G$ be a morphism. If $r$ is a deleting potentially minimal improving rule, either 1. or 2. applies, if $r$ is an inserting and not a deleting potentially minimal improving rule, only 2. applies, because $r$ cannot destroy any occurrences of $C_k$ in $G$.

1. If $p(C_k) \cap m(L \setminus K) \neq \emptyset$, $\mathrm{tr}_t \circ p$ is not total, since at least one element of $p(C_k)$ has been deleted by $t$ and $p$ does satisfy $\bigwedge_{C' \in \mathbf{G}}(p \models \mathrm{part}(1, e, C') \wedge \mathrm{tr}_t \circ p$ is total) $\implies \mathrm{tr}_t \circ p \models \mathrm{part}(1, e, C'))$.

2. If $p(C_k) \cap m(L \setminus K) = \emptyset$, $\mathrm{tr}_t \circ p$ is total, since no element of $p(C_k)$ has been deleted by $t$. Because (3.3) holds, $t$ does not delete any elements of $C_{k+1} \setminus C_k$ and therefore $p \models \mathrm{part}(1, e, C') \implies \mathrm{tr}_t \circ p \models \mathrm{part}(1, e, C')$ for all $C_k \subseteq C' \subseteq C_{k+1}$.

With 1. and 2. follows that (2.1) is satisfied.

Secondly, we show that equation (2.2) is satisfied. Let $p' : C_k \hookrightarrow H$ be a morphism. Because (3.4) is satisfied, $t$ does not create any elements of $C_k$ and therefore, there must exist an morphism $p : C_k \hookrightarrow G$ with $\mathrm{tr}_t \circ p = p'$. If follows that (2.2) is satisfied.

Lasty, we show that (2.3) is satisfied. We consider the cases that $r$ is a deleting minimal potentially improving rule and that $r$ is an inserting minimal potentially improving rule.

1. If $r$ is a deleting potentially minimal improving rule, the condition $\mathrm{ap_{pi}}(k, P) = \exists\big(a_0 : L \hookrightarrow C_k, \neg\exists(a_1 : C_k \hookrightarrow C_{k+1}, \textsf{true})\big)$ is satisfied by $m$. Therefore a morphism $p : C_k \hookrightarrow G$ with $p \not\models \neg\exists(a_1 : C_k \hookrightarrow C_{k+1}, \textsf{true}) = \mathrm{part}(1, e, C_{k+1})$ and $m = p \circ a_0$ must exist. Since $r$ is a deleting rule, at least one element of $p(C_k)$ has been deleted by $t$ and therefore $\mathrm{tr}_t \circ p$ is not total. It follows that (2.3) is satisfied.

2. If $r$ is a inserting potentially minimal improving rule, $\text{ap}_{\text{pi}}(k, P) = \neg \exists (b : L \hookrightarrow P, \textsf{true})$ is satisfied by $m$. Because $L = C_k$, $m \models \neg \exists (b : C_k \hookrightarrow P, \textsf{true}) = \text{part}(1, e, P)$. Since (3.6) is satisfied, $\text{tr} \circ p$ is total and $\text{tr}_t \circ p \models \text{part}(1, e, P)$. Therefore, (2.3) is satisfied. In total follows that $t$ is a direct minimal consistency improving transformation and therefore $r'$ is direct minimal consistency improving rule.

$\square$

**Definition 3.9** (**repairing rule set**)**.** *Let a constraint $c$ in EANF and a set of rules $\mathcal{R}$ be given. Then, $\mathcal{R}$ is called a* repairing rule set *for $c$ at layer $k$ if for all graphs $G$ with $k = c_{\max}$ a sequence*

$$G = G_0 \Rightarrow_{r_0} \ldots \Rightarrow_{r_{n-1}} G_n = H$$

*exists, such that $r_j \in \mathcal{R}$ for all $j \in \{0, \ldots, n-1\}$ and $H \models_{k+2} c$.*

**Corollary 3.10.** *Let a constraint $c$ in EANF and a set of rules $\mathcal{R}$ be given. If $\mathcal{R}$ is a repairing rule set for $c$ at layer $k$, $\mathcal{R}$ is a repairing rule set w.r.t $c$ at layer $j$ for all $k < j \leq \text{nl}(c)$.*

**Corollary 3.11.** *Let a constraint $c$ in EANF and a repairing rule set $\mathcal{R}$ for $c$ at layer $k$, for all $k \in \{1, 3, \ldots, \text{nl}(c)\}$, be given. Then, for all graphs $G$, a sequence*

$$G = G_0 \Rightarrow_{r_0} \ldots \Rightarrow_{r_{n-1}} G_n = H$$

*exists, such that $r_j \in \mathcal{R}$ for all $j \in \{0, \ldots, n-1\}$ and $H \models c$.*

**Theorem 3.2.** *Let a constraint $c$ in EANF and a set of rules $\mathcal{R}$ be given. Then, $\mathcal{R}$ is a repairing set of $c$ at layer $k \leq \text{nl}(c)$ if either 1. or 2. holds.*

1. *For any universally bound graph $C_j$ at layer $j \leq k$ of $c$, $(r, \text{ap}_{\text{pi}}(j, C_{j+1})) \in \mathcal{R}$ and $r = L \hookleftarrow K \hookrightarrow R$ is a deleting potentially minimal improving rule at layer $j$ with $C_{j+1}$, such that $r$ only deletes edges of $C_j$:*

$$(L \setminus K) \cap V_{C_k} = \emptyset$$

2. *Let $C_{k+1}$ be the existentially bound graph of $c$ at layer $k+1$. A set of graphs*

$$\mathbf{P} \subseteq \{P \mid C_k \subseteq P \subseteq C_{k+1}\},$$

*such that the inclusions $i_\ell : P_\ell \hookrightarrow C_{k+1}$ for all $P_\ell \in \mathbf{P}$ are jointly surjective, exists, and for all $P \in \mathbf{P}$, a rule $(r, \text{ap}_{\text{pi}}(k, P)) \in \mathcal{R}$ exists, such that $r = L \hookleftarrow K \hookrightarrow R$ is an inserting potentially minimal improving rule at layer $k$ with $P$ and $L = K$.*

*Proof.* Let a constraint $c$ in EANF, a rule set $\mathcal{R}$ and a graph $G$ with $k = c_{\max}$ and $c_{\max} < \text{nl}(c)$ be given. We show that a sequence $G = C_0' \Rightarrow \ldots \Rightarrow C_n' = H$ with rules of $\mathcal{R}$ exists, such that $H \models_{k+2} c$ if 1. or 2. of theorem 3.2 is satisfied.

1. Assume that 1. of theorem 3.2 holds. Let $(r, \mathrm{ap_{pi}}(j, C_{j+1})) \in \mathcal{R}$, such that $r = L \hookleftarrow K \hookrightarrow R$ is a deleting potentially minimal improving rule at layer $j \leq k$ with $C_{j+1}$ and $C_j$ is a universally bound graph of $c$. Then, $\mathrm{ap_{pi}}(j, C_{j+1}) = \exists\big(a_0 : L \hookrightarrow C_j, \neg\exists(a_1 : C_j \hookrightarrow C_{j+1}, \mathsf{true})\big)$. Let $q : C_j \hookrightarrow G$ be a morphism such that $q \not\models \exists(C_j \hookrightarrow C_{j+1}, \mathsf{true})$. Since $L \subseteq C_j$, we can construct a morphism $m_1 : L \hookrightarrow G$ with $m_1(L) = q(L)$. It holds that $m_1 = q \circ a_0$, therefore $m_1 \models \mathrm{ap_{pi}}(j, C_{j+1})$. Since $r$ only deletes edges, a transformation $t : G = G_0 \Rightarrow_{r,m_1} G_1$ exists and $\mathrm{tr}_t \circ p$ is not total. Because $r$ does not insert any elements of $C_j$:

$$|\{q : C_k \hookrightarrow G_0 \mid q \not\models d\}| < |\{q : C_k \hookrightarrow G_1 \mid q \not\models d\}|$$

   with $d = \exists(C_j \hookrightarrow C_{j+1}, \mathsf{true})$. By iteratively applying this construction, we generate a finite sequence of transformations

$$G = G_0 \Rightarrow_{r,m_1} G_1 \Rightarrow_{r,m_2} \ldots \Rightarrow_{r,m_n} G_n = H$$

   such that $|\{q : C_k \hookrightarrow G_n \mid q \not\models d\}| = 0$ and therefore $H \models_j c$. With lemma 2.8, $H \models_{k+2} c$ and $H \models c$ follows.

2. Assume that 2. of theorem 3.2 holds. Let $r_0 = (r, \mathrm{ap_{pi}}(k, P)) \in \mathcal{R}$ be an inserting potentially minimal improving rule of $c$ at layer $k$ with $P \in \mathbf{P}$. Then, $\mathrm{ap_{pi}}(k, P) = \neg\exists(b : L \hookrightarrow P, \mathsf{true})$. Let $q : C_k \hookrightarrow G$ be a morphism, such that $q \not\models \exists(C_k \hookrightarrow P, \mathsf{true})$. Since $L = C_k$, we set $m : C_k \hookrightarrow G$ with $m_0 = q$. It follows that $m_0 \models \neg\exists(C_k \hookrightarrow P, \mathsf{true}) = \mathrm{ap_{pi}}(k, P)$. Because $r$ does not delete any elements, a transformation $t_0 : G \Longrightarrow_{r_0,m_0} G_1$ exists and $\mathrm{tr}_t \circ q \models \exists(C_k \hookrightarrow P, \mathsf{true})$. By iteratively applying this construction with rules $(r, \mathrm{ap_{pi}}(k, P)) \in \mathcal{R}$ such that $r$ is a inserting potentially minimal improving rule of $c$ at layer $k$ with $P \in \mathbf{P}$ we can construct a finite sequence of transformations

$$G \Longrightarrow_{r_0,m_0} G_0 \Longrightarrow_{r_1,m_1} \ldots \Longrightarrow_{r_n,m_n} G_n$$

   such that $m_i = \mathrm{tr}_{t_{i-1}} \circ \ldots \circ \mathrm{tr}_{t_0} \circ m_0$ and $G_n \models \exists(C_k \hookrightarrow P_j, \mathsf{true})$ for all $P_j \in \mathbf{P}$. Because the inclusions $t_\ell : P_\ell \hookrightarrow C_{k+1}$ for all $P_\ell \in \mathbf{P}$ are jointly surjective, $\mathrm{tr}_{t_n} \circ \ldots \circ q \models \exists(C_k \hookrightarrow C_{k+1}$ follows. Therefore, $G_n \models_{k+2} c$.

$\square$

# References

[1] C. Sandmann and A. Habel. Rule-based graph repair. *arXiv preprint arXiv:1912.09610*, 2019.