Formalizing Kruskal’s Tree Theorem in Isabelle/HOL

Christian Sternagel

JAIST

July 5, 2012

Kickoff Meeting of Austria-Japan Joint Project
Gamagori, Japan
Kruskal’s Tree Theorem

If the set $A$ is well-quasi-ordered then the set of finite trees over $A$ is well-quasi-ordered by homeomorphic embedding.

Comments

• proof structure as Nash-Williams 1963
• which claims
  
  A new and simple proof is given . . .
Overview

• Motivation

• Preliminaries

• Kruskal’s Tree Theorem - A Proof Sketch

• Formalization Challenges

• Conclusion
Bibliography

G. Higman.

J. B. Kruskal.

C. S. J. A. Nash-Williams.
doi:10.1017/S0305004100003844.

C. Sternagel.
Why?

- long-standing open problem in formalized mathematics
- Kruskal’s Tree Theorem is main ingredient to prove well-foundedness of simplification orders for first-order rewriting
- ultimately, we want to strengthen termination library of IsaFoR (Isabelle Formalization of Rewriting)

---

First, for constrained rewriting with forbidden patterns, we want to be able to certify the loop detection algorithm of \cite{78} which encompasses the algorithms for loops under the innermost and outermost strategy \cite{77,76}. To date no formal certification techniques for these highly interesting techniques are known. This is partly due to the fact that the correctness proof uses a very powerful and complex theorem: Kruskal’s tree theorem \cite{40}, whose formal verification is open.
Homeomorphically Embedding on Lists

- **empty list**, \([\text{ ]} \)  
- **adding element** \(x\) to finite list \(xs\), \(x \cdot xs\)  
- **append lists** \(xs\) and \(ys\), \(xs @ ys\)  
- **set of finite lists** over \(A\), \(A^*\):

  \[
  \begin{align*}
  [\text{ ]} & \in A^* \\
  x \in A & \quad xs \in A^* \\
  x \cdot xs & \in A^*
  \end{align*}
  \]

- **embedding relation** w.r.t. \(\leq\):

  \[
  \begin{align*}
  [\text{ ]} & \leq_{\text{emb}} ys \\
  xs & \leq_{\text{emb}} ys \\
  x & \leq y & xs & \leq_{\text{emb}} ys \\
  & x \cdot xs & \leq_{\text{emb}} y \cdot ys
  \end{align*}
  \]
Example - List Embedding

We can . . .

- drop elements \( \left( \frac{xs \preceq_{emb} ys}{xs \preceq_{emb} y \cdot ys} \right) \)
- replace elements by smaller elements \( \left( \frac{x \preceq y \quad xs \preceq_{emb} ys}{x \cdot xs \preceq_{emb} y \cdot ys} \right) \)
- note that empty list is embedded in any list \( (\) \preceq_{emb} ys) \)
Homeomorphically Embedding on Trees

- tree with node $f$ and list of direct subtrees $ts$, $f(ts)$
- root of tree $\text{root}(f(ts)) = f$, direct subtrees $\text{args}(f(ts)) = ts$
- set of finite trees over $A$, $\mathcal{T}(A)$:

$$f \in A \quad \forall t \in ts. \ t \in \mathcal{T}(A)$$
$$f(ts) \in \mathcal{T}(A)$$

- homeomorphically embedding relation w.r.t. $\preceq$:

$$t \in ts \quad t \preceq_{\text{emb}} f(ts)$$
$$f \preceq g \quad ss \ (\preceq_{\text{emb}})^{\text{emb}} ts$$
$$f(ss) \preceq_{\text{emb}} g(ts)$$

$$s \preceq_{\text{emb}} t \quad t \preceq_{\text{emb}} u$$
$$s \preceq_{\text{emb}} u$$

$$s \preceq_{\text{emb}} t$$
$$f(ss_1 @ s \cdot ss_2) \preceq_{\text{emb}} f(ss_1 @ t \cdot ss_2)$$
**Homeomorphic Embedding on Trees (cont’d)**

**Embedding TRS**

let $\text{Emb}(\preceq)$ be the infinite TRS

\[
\begin{align*}
    f(ts) & \rightarrow t & \text{if } t \in ts \\
    f(ts) & \rightarrow g(ss) & \text{if } g \preceq f \text{ and } ss =_{\text{emb}} ts
\end{align*}
\]

**Result**

$s \preceq_{\text{emb}} t$ iff $t \rightarrow^{+} \text{Emb}(\preceq) s$
**Well-Quasi-Orders - Definitions**

- let $A$ be a set and $\leq$ a binary relation
- $A$ is **wqo** by $\leq$ ($\preceq_A$ is a wqo, or wqo($\preceq_A$)):
  1. **transitive**: $\forall x \in A. \forall y \in A. \forall z \in A. x \preceq y \land y \preceq z \rightarrow x \preceq z$
  2. **all infinite sequences over $A$ are good**:

$$\forall f. (\forall i. f(i) \in A) \rightarrow (\exists j \; k. j < k \land f(j) \preceq f(k))$$

\[
\begin{array}{ccccccc}
  | & f(1) & | & f(2) & | & f(3) & \ldots & | & f(j) & \ldots & | & f(k) & \ldots \\
\end{array}
\]

- a sequence that is not good, is called **bad**

**Property**

- strict part of $\preceq$ is $x \prec y = x \preceq y \land y \npreceq x$
- let wqo($\preceq_A$), then $\prec_A$ is well-founded on $A$
Kruskal’s Tree Theorem - A Proof Sketch
**Proof Structure of** \( \text{wqo}(\preceq_F) \implies \text{wqo}(\preceq_T(F)) \)

1. **Assume** \( \preceq_F \) is wqo.
2. **Assume** \( \preceq_T(F) \) is not wqo.
3. \( \implies \) exists **minimal** bad sequence \( t_1, t_2, t_3, \ldots \) with \( t_i \in T(F) \).
4. \( \implies \) exists sequence \( f_1, f_2, f_3, \ldots \) with \( \text{root}(t_i) = f_i, \text{args}(t_i) = ts_i \).
5. \( \implies \) let \( T = \bigcup_i (\bigcup \text{args}(t_i)) \).
6. \( \implies \) \( \preceq \{f_i\} \) and \( \preceq_T^* \) are wqo.
7. \( \implies \) \( \preceq \{f_i\} \times T^* \) is wqo.
8. \( \implies \) exist \( j, k \) with \( j < k \) and \( (f_j, ts_j) \preceq \{f_i\} \times T^* (f_k, ts_k) \).
9. \( \implies \) \( t_j \preceq_T(F) t_k \).
10. \( \implies \) \( t_1, t_2, t_3, \ldots \) is good.
11. **Contradiction!**

*In what sense?*

**Prove it!**
Formalization Challenges
Existence of Minimal Bad Sequence - Nash-Williams 1963

Select an $t_1 \in \mathcal{T}(\mathcal{F})$ such that $t_1$ is the first term of a bad sequence of members of $\mathcal{T}(\mathcal{F})$ and $t_1$ is as small as possible. Then select an $t_2$ such that $t_1, t_2$ are the first two terms of a bad sequence of members of $\mathcal{T}(\mathcal{F})$ and $t_2$ is as small as possible [...]. Assuming the Axiom of Choice, this process yields a bad sequence $t_1, t_2, t_3, \ldots$. 
The Axiom of Choice in Isabelle

- \( \forall x. \exists y. P \times y \implies \exists f. \forall x. P \times (f \times) \)

Minimal in What Sense?

- subtree relation, \( t \) is (proper) subtree of \( s \), written \( t \triangleleft s \)

\[
(t \triangleleft s), \text{ iff: } \begin{cases} t \in ts & s \triangleleft t \\ t \triangleleft f(ts) & t \triangleleft f(ts) \end{cases}
\]

- proper subtree relation is well-founded (allowing for induction)

Auxiliary Definitions

- infinite sequence \( f \) is minimal at position \( n \) (min\(_n\)\( f \)), iff:

\[
\forall g. (\forall i < n. g(i) = f(i)) \land g(n) \triangleleft f(n) \land (\forall i \geq n. \exists j \geq n. g(i) \triangleleft f(j))
\]

\[\implies \text{good}_{\triangleleft\text{emb}}(g)\]

- replace elements of sequence \( f \) by those of sequence \( g \), starting at \( n \):

\[
(f\langle n \rangle g)(i) = \text{if } i \geq n \text{ then } g(i) \text{ else } f(i)
\]
Key Lemma

\[ \text{(1) } \min_n(f) \]
\[ \text{(2) } \text{bad}_{\preceq_{\text{emb}}}(f) \]

\[ \implies \exists g. \forall i \leq n. g(i) = f(i) \]
\[ \land g(n + 1) \preceq f(n + 1) \]
\[ \land \forall i \geq n + 1. \exists j \geq n + 1. g(i) \preceq f(j) \]
\[ \land \text{bad}_{\preceq_{\text{emb}}}(f \langle n + 1 \rangle g) \]
\[ \land \min_{n+1}(f \langle n + 1 \rangle g) \]

Construct Minimal Bad Sequence

- from AC and key lemma obtain function \( \nu \), s.t., given sequence satisfying (1) and (2) and index \( n \), returns sequence satisfying conclusion
- auxiliary sequence (of sequences)

\[ m'(n) = \begin{cases} 
    \nu(f, n) & \text{if } n = 0 \\
    m'(n - 1) \langle n \rangle \nu(m'(n - 1), n - 1) & \text{otherwise}
\end{cases} \]

- minimal bad sequence \( m(i) = m'(i)(i) \)
Conclusion
Related Work

- Murthy, Extracting Constructive Content from Classical Proofs, PhD, 1990 (Nuprl)
- Fridlender, Ramsey’s Theorem in Type Theory, Tech. Report, 1993 (ALF)
- Herbelin, A Program from an A-Translated Impredicative Proof of Higman’s Lemma, 1994 (2-letter alphabet, Coq)
- Fridlender, Higman’s Lemma in Type Theory, PhD, 1997 (ALF)
- Seisenberger, On the Constructive Content of Proofs, PhD, 2003 (Minlog)
- Berghofer, A Constructive Proof of Higman’s Lemma in Isabelle, TYPES 2003 (2-letter alphabet, Isabelle)
- Martín-Mateos et al., A Formal Proof of Higman’s Lemma in ACL2, JAR 2011 (ACL2)
Future Work

- investigate how Zorn’s Lemma could be of help
- reformulate proof using open induction

The End