An Agda Formalization of Friedman's "Extended Kruskal Theorem"

Makoto Kanazawa, Hosei University

Abstract

Friedman's Extended Kruskal Theorem states that a certain embedding relation between trees with a "gap condition" is a WQO. It is an important theorem in reverse mathematics in that it is unprovable in a very strong logical system. I formalize Friedman's proof in the proof assistant Agda (assuming excluded middle).

Rather than reasoning about binary relations being WQOs, I state all necessary lemmas in terms of sequences being *very good* with respect to a given binary relation. This simplifies the proof somewhat, at the cost of slightly complicating (and generalizing) the "minimal bad sequence" argument.

Well-Quasi-Ordering

Let $\preceq$ be a binary relation on a set A.

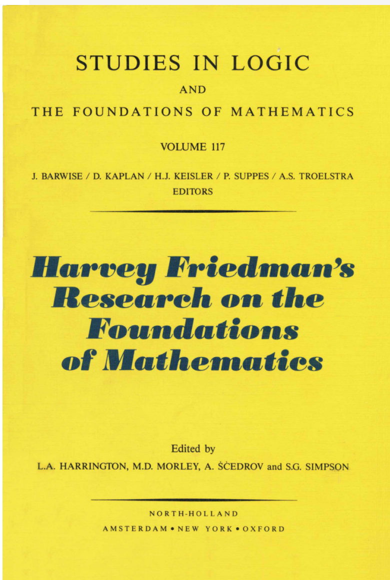
- An infinite sequence $f : \mathbb{N} \to A$ is *good* (w.r.t. $\preceq$) if there exist i, j such that $i < j$ and $f\,i \preceq f\,j$.
- $\preceq$ is *almost full* if every infinite sequence $f : \mathbb{N} \to A$ is good.
- $\preceq$ is a *well-quasi-ordering* if it is reflexive, transitive, and almost full.

Tree Embedding

Let t_1, t_2 be finite rooted trees. t_1 *embeds* into t_2 if there is an injective mapping h from the nodes of t_1 to the nodes of t_2 such that

- h preserves the ancestor relation, and
- if v_1, v_2 are distinct children of v_0, then any path between $h\,v_1$ and $h\,v_2$ must pass through $h\,v_0$.

Kruskal's Tree Theorem. The tree embedding relation is a WQO.



Gap Conditions

Let Tree A n be the set of all finite rooted trees whose interior nodes are labeled by natural numbers $< n$ and whose leaves are labeled by elements of A.

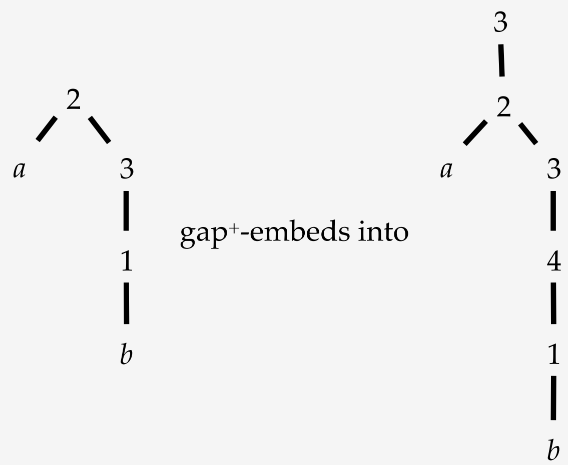
Let $t_1, t_2 \in$ Tree A n and let t_1°, t_2° be the unlabeled trees underlying t_1, t_2.

t_1 *embeds* into t_2 if there exists an embedding h of t_1° into t_2° such that

- h maps an interior node to an interior node with the same label, and
- if v is a leaf, then $h\,v$ is a leaf and $v \preceq h\,v$.

An embedding h of t_1 into t_2 satisfies

- the gap condition if label$(u) \geq$ label$(h\,v_2)$ whenever v_1 is the parent of an interior node v_2 and u lies between $h\,v_1$ and $h\,v_2$
- the gap⁺ condition if moreover label$(u) \geq$ label$(h(\text{root}(t_1))$ whenever u is an ancestor of $h(\text{root}(t_1))$.



Friedman's Extended Kruskal Theorem. For every n, the gap-embedding relation on Tree A n is a WQO.

Very Good Sequences

An infinite sequence $f : \mathbb{N} \to A$ is *very good* if every subsequence of f is good. Many closure properties of WQOs can be stated in terms of very good sequences.

Dickson's Lemma. If $\preceq_1$ is a WQO on A_1 and $\preceq_2$ is a WQO on A_2, then $\preceq_1 \times \preceq_2$ is a WQO on $A_1 \times A_2$.

金沢　誠／法政大学理工学部　創生科学科

Web: makotokanazawa.ws.hosei.ac.jp　Email: kanazawa@hosei.ac.jp

Lemma. If $f_1 : \mathbb{N} \to A_1$ and $f_2 : \mathbb{N} \to A_2$ are very good, then so is $\lambda i \to (f_1 i, f_2 i)$.

If $\preceq$ is a binary relation on A, a finite sequence xs : List A *embeds* into a finite sequence ys : List A if there is a (scattered) subsequence zs of ys such that $xs[i] \preceq zs[i]$ for all positions i in xs.

Higman's Theorem. If $\preceq$ is a WQO on A, then the embedding relation on List A is a WQO.

Theorem. Let $f : \mathbb{N} \to$ List A. Suppose that for every infinite sequence $g : \mathbb{N} \to A$, if there is a subsequence $\hat{f}$ of f such that $g\,i$ occurs in $\hat{f}\,i$ for all $i \in \mathbb{N}$, then g is very good w.r.t. $\preceq$. Then f is very good w.r.t. the embedding relation on List A.

Reformulating Friedman's Lemmas

Lemma 4.6 (Friedman).
$\forall (A, \preceq)$ ($\preceq$ is a WQO on A $\to$ gap-embedding is a WQO on Tree A $(1+n)$) $\to$
$\forall (A, \preceq)$ ($\preceq$ is a WQO on A $\to$ gap⁺-embedding is a WQO on Tree A n) $\to$
$\forall (A, \preceq)$ ($\preceq$ is a WQO on A $\to$ gap⁺-embedding is a WQO on Tree A $(1+n)$)

(In the proof, A is instantiated to Tree A $(1+n)$!!)

Definition.
Gap⁺ n: For every infinite sequence $f : \mathbb{N} \to$ Tree A n, if every $g : \mathbb{N} \to A$ such that $g\,i$ labels a leaf of $f\,i$ is very good, then f is very good w.r.t. gap⁺-embedding.

Gap n: For every infinite sequence $f : \mathbb{N} \to$ Tree A n, if every $g : \mathbb{N} \to A$ such that $g\,i$ labels a leaf of $f\,i$ is very good, then f is very good w.r.t. gap-embedding.

Lemma. Gap $(1+n) \to$ Gap⁺ $n \to$ Gap⁺ $(1+n)$.

Lemma 4.7 (Friedman).
$\forall (A, \preceq)$ ($\preceq$ is a WQO on A $\to$ gap⁺-embedding is a WQO on Tree A n) $\to$
$\forall (A, \preceq)$ ($\preceq$ is a WQO on A $\to$ gap-embedding is a WQO on Tree A $(1+n)$)

Lemma. Gap⁺ $n \to$ Gap $(1+n)$.

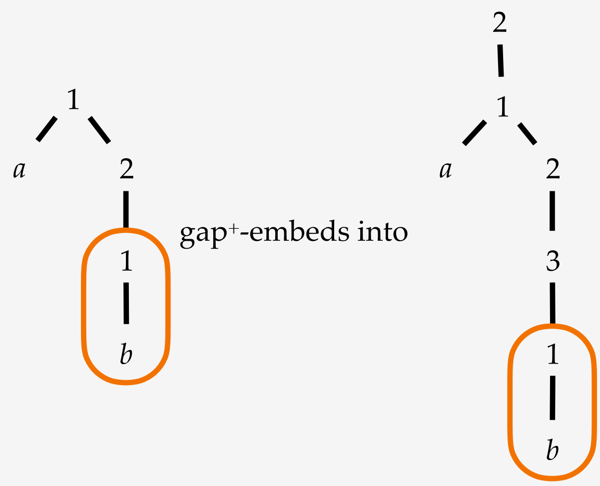
Friedman's Star Function

A tree $t \in$ Tree A $(1+n)$ is turned into a tree $t^* \in$ Tree (Tree A $(1+n)$) n by

- turning each maximal subtree t' rooted at a node labeled by m into a leaf labeled by t', and
- decrementing all remaining internal node labels,

where $m = \mu t$ is the smallest internal node label of t.

Lemma 4.5 (Friedman). If $\mu t_1 = \mu t_2$ and t_1^* gap⁺-embeds into t_2^*, then t_1 gap⁺-embeds into t_2.



(This has turned out to be the toughest part of the proof to formalize in Agda! Part of the reason is that Friedman only gives the one-line proof "Straightforward".)

Minimal Bad Sequence Argument

The proof of Lemma 4.7 is based on Nash-Williams's *minimal bad sequence argument.*
Let size t be the number of nodes of t.

MBS. If there is a bad sequence $f : \mathbb{N} \to$ Tree A n, then there is a bad sequence $g : \mathbb{N} \to$ Tree A n that is minimal with respect to the lexicographic ordering of $\mathbb{N} \to$ Tree A n based on size.

I need to reformulate this lemma to the following:

Lemma. Suppose $f : \mathbb{N} \to$ Tree A n is a bad sequence. Then there is a bad sequence that is minimal among all sequences $g : \mathbb{N} \to$ Tree A n such that for some subsequence $\hat{f}$ of f, $g\,i$ is a subtree of $\hat{f}\,i$ (for every $i \in \mathbb{N}$).

In other words, the minimal bad sequence is obtained from the given bad sequence f by optionally skipping $f\,i$ and/or replacing $f\,i$ by some subtree of $f\,i$ for each i.

Agda vs. Coq

Agda is very similar to Haskell (in fact it's written in Haskell), and Agda code can be made almost as human readable as Haskell code.
Agda has control sequences that play similar roles to some tactics of Coq; you just don't record those key strokes in Agda!

Agda	Coq
C-c C-c	intros
C-c C-c x	destruct x
C-c C-c x	induction x
C-c C-r	split

金沢　誠／法政大学理工学部　創生科学科
Web: makotokanazawa.ws.hosei.ac.jp　Email: kanazawa@hosei.ac.jp