Monotonicity Ground Decidability: A Proof Theoretic View

Daniel Li

University of California, Berkeley
li.daniel2001@berkeley.edu

Abstract. Monotonic functions from mathematics have been shown to be useful in modeling various phenomena in natural language semantics. The monotonicity calculus is a proof system that formalizes monotonicity reasoning in natural language inferences through order statements (inequalities) involving higher-order functions. In this paper, we prove that it is decidable over ground terms, i.e., terms that do not contain variables.

Keywords: Monotonicity · Decidability · Subterm property

1 Introduction

Monotonicity is a prevalent feature in natural language. Many words – notably determiners such as *every*, *some*, and *all* – induce monotonic (upward) or antitonic (downward) entailments. For instance, loosely speaking, *run* entails *move*, and *Every man runs* entails *Every man moves*. Meanwhile, *man* entails *person*, and *Every person runs* entails *Every man runs*. These can be formalized as follows.

$$\frac{run \leq move}{every(man)(\boldsymbol{run}) \leq every(man)(\boldsymbol{move})} \quad \frac{man \leq person}{every(\boldsymbol{person})(run) \leq every(\boldsymbol{man})(run)}$$

So we say that *every* is monotonic in its second argument and antitonic in its first argument.

This paper is based on the typed logical system presented in [2, 3] which formalizes reasoning about natural language inferences in the presence of monotonicity. Specifically, we show that its proof system, called the *monotonicity calculus*, is decidable over ground terms. Recently, we found that Fyodorov et al. have done similar work [1], but our methods are different, and we also extend their work by removing an assumption they make in the fragment of interest. Inspired by the proof of decidability for equational logic over ground terms in [5], our approach relies on exploiting the subterm property, to be defined below. Our methods are proof theoretic, and we believe one advantage of this is that it is more illustrative of the algorithmics. Amidst recent interest in finding decidable fragments, there has also been related work in the programming language theory literature on similar inequational logics (e.g. subtyping [4]).

2 Background

We present some relevant background information, focusing on syntactic aspects.

We begin with base types $b \in \mathcal{B}$. The set of types is defined inductively by

$$\tau \ ::= \ b \mid \tau \xrightarrow{+} \tau \mid \tau \xrightarrow{-} \tau$$

The **valence** of a type is defined inductively as follows: $\text{valence}(b) = 0$ for $b \in \mathcal{B}$, and $\text{valence}(\sigma \xrightarrow{p} \tau) = \text{valence}(\tau) + 1$ (where $p \in \{+, -\}$ is called the polarity marking).

Each type has an associated set of constants. In this paper, we will only work with **ground terms**, which are terms that do not contain any variables. All constants are ground terms, and for any ground terms F with type $\sigma \xrightarrow{p} \tau$ and A with type σ, $F(A)$ is a ground term with type τ.[1] If terms S and T have the same type disregarding their polarity markings, then we can write the **inequality** $S \leq T$.[2]

If F has type $\sigma \xrightarrow{+} \tau$ for some σ, τ, then we say F is **monotonic** (denoted $F : +$ or F^+). If F has type $\sigma \xrightarrow{-} \tau$ for some σ, τ, then we say F is **antitonic** (denoted $F : -$ or F^-).

We now define **subterms**. For a constant A, $\text{Subterm}(A) = \{A\}$, and for a term $F(A)$,

$$\text{Subterm}(F(A)) = \{F(A)\} \cup \text{Subterm}(F) \cup \text{Subterm}(A)$$

Given a set Γ of inequalities and terms S, T, the subterm set of (Γ, S, T) is

$$\text{Subterm}(S) \cup \text{Subterm}(T) \cup \bigcup_{(U \leq V) \in \Gamma} \text{Subterm}(U) \cup \text{Subterm}(V)$$

Instead of speaking of *subterms of* (Γ, S, T), we will often refer to them simply as *subterms* when it is understood what Γ, S, T are.

Definition 1. *The monotonicity calculus $\mathcal{M}$ (over ground terms) is the proof system with the following rules:*

$$(F : +)\ \frac{A \leq B}{F(A) \leq F(B)}\ (\text{Mono}) \qquad\qquad (G : -)\ \frac{A \leq B}{G(B) \leq G(A)}\ (\text{Anti})$$

$$\frac{F \leq G}{F(A) \leq G(A)}\ (\text{Point}) \qquad \frac{A \leq B \qquad B \leq C}{A \leq C}\ (\text{Trans}) \qquad \frac{}{A \leq A}\ (\text{Refl})$$

We will view formal derivations as Gentzen-style inverted trees. Let $\boldsymbol{D}$ be a derivation. If the conclusion of $\boldsymbol{D}$ is $S \leq T$, we say that $\boldsymbol{D}$ is a *derivation of*

[1] We use lowercase letters for actual terms and uppercase letters for metavariables over terms.

[2] We will generally not be explicit about types and just assume they are compatible.

$S \leq T$.[3] A *subderivation* within $\boldsymbol{D}$ is a subtree of $\boldsymbol{D}$. A *proper subderivation* within $\boldsymbol{D}$ is a subderivation that is not $\boldsymbol{D}$ itself. For an instantiation of an axiom or a derivation with (Refl), the *height* of the derivation is 1. The height of a derivation whose last step is (Mono), (Anti), or (Point) is $h + 1$ where h is the height of the subderivation of its premise. The height of a derivation whose last step is (Trans) is $\max(h_1, h_2) + 1$ where h_1, h_2 are the heights of the subderivations of its premises.

In this paper, we prove that given a finite set Γ of ground inequalities and ground terms S, T, it is decidable whether or not $\Gamma \vdash_{\mathcal{M}} S \leq T$. Incidentally, the fragment we investigate coincides with a fragment of the logic in [1]. Fyodorov et al. provide a decision algorithm for this fragment, but they impose the restriction that derivations cannot have any occurrences of

$$\frac{F_1^+ \leq F_2^- \qquad F_2^- \leq F_3^+}{F_1^+ \leq F_3^+}\,\text{(Trans)} \qquad\qquad \frac{F_1^- \leq F_2^+ \qquad F_2^+ \leq F_3^-}{F_1^- \leq F_3^-}\,\text{(Trans)}$$

which they term *mixed monotonicity*, conjecturing that such patterns do not occur in natural language. The main result in this paper extends the work of Fyodorov et al. by removing this restriction. The issue of whether it is linguistically innocuous to eliminate mixed monotonicity remains an open question, but regardless, monotonicity is ubiquitous in general mathematics, and formal properties of monotonicity reasoning are of theoretical interest.

3 A Special Case

To motivate the solution to the full problem, we first consider a restricted system $\mathcal{M}_+$ without antitonic functions and without the proof rule (Anti). Many of the ideas from this section will carry over to the general case, and we develop the intuition and details necessary for the full solution so that we can just highlight the differences instead of repeating these details later.

Definition 2. $\mathcal{M}_+$ *has the proof rules* (Mono), (Point), (Trans), and (Refl) from $\mathcal{M}$.

The main approach in this paper is to exploit the subterm property. This essentially makes the search space finite for proof search, and decidability follows closely after.

Definition 3. *Let Γ be a set of inequalities, and S, T be terms. A **subterm derivation** of $S \leq T$ from axioms in Γ is a derivation that contains only terms from the subterm set of (Γ, S, T). A proof system $\mathcal{P}$ has the **subterm property** if $\Gamma \vdash_{\mathcal{P}} S \leq T$ implies $\Gamma \vdash_{\mathcal{P}} S \leq T$ via a subterm derivation.*

First note that $\mathcal{M}_+$ does not have the subterm property. As a very simple counterexample,

$$\{f \leq g, a \leq b\} \vdash f(a) \leq g(b)$$

[3] We will sometimes interchangeably refer to a derivation by its conclusion.

but it is not possible to obtain a subterm derivation. The simplest derivations we can find are

$$\dfrac{\dfrac{a \leq b}{f(a) \leq f(b)}\,(\text{Mono}) \qquad \dfrac{f \leq g}{f(b) \leq g(b)}\,(\text{Point})}{f(a) \leq g(b)}\,(\text{Trans}) \qquad \dfrac{\dfrac{f \leq g}{f(a) \leq g(a)}\,(\text{Point}) \qquad \dfrac{a \leq b}{g(a) \leq g(b)}\,(\text{Mono})}{f(a) \leq g(b)}\,(\text{Trans})$$

which contains non-subterms $f(b)$ and $g(a)$ respectively.

To that end, consider a new proof system

Definition 4. $\mathcal{M}_+'$ *has the proof rules:*[4]

$$\dfrac{F \leq G \qquad A \leq B}{F(A) \leq G(B)}\,(\text{App}) \qquad \dfrac{A \leq B \qquad B \leq C}{A \leq C}\,(\text{Trans}) \qquad \dfrac{}{A \leq A}\,(\text{Refl})$$

Essentially, this system resolves the issue above. We claim that $\mathcal{M}_+'$ is equivalent to $\mathcal{M}_+$.

Proposition 1. *$\Gamma \vdash_{\mathcal{M}_+} S \leq T$ if and only if $\Gamma \vdash_{\mathcal{M}_+'} S \leq T$*

Proof. For both proof systems, we show that each of the proof rules can be derived in the other system. For the forward direction, (Mono) and (Point) can be derived, respectively, through

$$\dfrac{\dfrac{}{F \leq F}\,(\text{Refl}) \qquad A \leq B}{F(A) \leq F(B)}\,(\text{App}) \qquad\qquad \dfrac{F \leq G \qquad \dfrac{}{A \leq A}\,(\text{Refl})}{F(A) \leq G(A)}\,(\text{App})$$

For the backwards direction, (App) can be derived essentially by the example above.

We now introduce a few technical notions in order to show that $\mathcal{M}_+'$ has the subterm property.

Definition 5. *Given a derivation of $S \leq T$ whose last step is (Trans), its **bottommost chain** is the maximal sequence $(S = S_1, S_2, \ldots, S_n = T)$ such that $S_1 \leq S_2; \ldots; S_{n-1} \leq S_n$ all appear as premises of (Trans). That is, for each $k = 1, \ldots, n-1$, the derivation of $S_k \leq S_{k+1}$ does not have (Trans) as its last step. The **bottommost subterm chain** is the maximal subsequence of the bottommost chain containing only subterms of (Γ, S, T).*

Definition 6. *A **subterm segment** is a subsequence $(S_1', S_2', \ldots, S_{n-1}', S_n')$ of the bottommost chain such that S_1', S_n' are subterms and $S_2', \ldots, S_{n-1}'$ are not subterms. In particular, if $n = 2$, then (S_1', S_2') is a subterm segment if S_1', S_2' are subterms.*

[4] Height for (App) is defined in the same way as (Trans), since it also has two premises.

Definition 7. *Let* $(S^{(1)}, S^{(2)}, \ldots, S^{(i)})$ *be the bottommost subterm chain of a derivation. The bottommost chain of the derivation is* ***subterm-separated*** *if* $S^{(1)} \leq S^{(2)}; \ldots; S^{(i-1)} \leq S^{(i)}$ *all appear in the derivation. A derivation is subterm-separated if the bottommost chain of each of its subderivations is subterm-separated. In this case, the bottommost chain of each* $S^{(j)} \leq S^{(j+1)}$ *is a subterm segment.*

Lemma 1. *If* $\Gamma \vdash_{\mathcal{M}'_+} S \leq T$*, then* $\Gamma \vdash_{\mathcal{M}'_+} S \leq T$ *via a subterm-separated derivation.*

Proof. By induction on the size of the bottommost chain for any arbitrary subderivation. Just rearrange the order in which (Trans) is applied in.

We now discuss the main result of this section.

Theorem 1. *If* $\Gamma \vdash_{\mathcal{M}'_+} S \leq T$ *via a subterm-separated derivation, then* $\Gamma \vdash_{\mathcal{M}'_+} S \leq T$ *via a subterm derivation.*

Proof. Refer to the appendix.

We provide an illustrative example and an informal proof sketch below.

Example 1. Let $\Gamma = \{f_1 \leq f_2, f_2 \leq f_3, f_3 \leq f_4, a_1 \leq a_2, a_2 \leq a_3, a_3 \leq a_4\}$. Assume we begin with the following derivation.

$$\dfrac{\dfrac{\dfrac{f_1 \leq f_2 \quad a_1 \leq a_2}{f_1(a_1) \leq f_2(a_2)}\,(\text{App}) \quad \dfrac{f_2 \leq f_3 \quad a_2 \leq a_3}{f_2(a_2) \leq f_3(a_3)}\,(\text{App})}{f_1(a_1) \leq f_3(a_3)}\,(\text{Trans}) \quad \dfrac{f_3 \leq f_4 \quad a_3 \leq a_4}{f_3(a_3) \leq f_4(a_4)}\,(\text{App})}{f_1(a_1) \leq f_4(a_4)}\,(\text{Trans})$$

Note that the bottommost chain of the derivation $(f_1(a_1), f_2(a_2), f_3(a_3), f_4(a_4))$ is a single subterm segment, so the derivation is subterm-separated. The non-subterms here are $f_2(a_2)$ and $f_3(a_3)$. We apply the following transformations to eliminate these non-subterms.

$$\dfrac{\dfrac{\dfrac{\boldsymbol{f_1 \leq f_2} \quad \boldsymbol{f_2 \leq f_3}}{\boldsymbol{f_1 \leq f_3}}\,(\text{Trans}) \quad \dfrac{\boldsymbol{a_1 \leq a_2} \quad \boldsymbol{a_2 \leq a_3}}{\boldsymbol{a_1 \leq a_3}}\,(\text{Trans})}{\boldsymbol{f_1(a_1) \leq f_3(a_3)}}\,(\text{App}) \quad \dfrac{f_3 \leq f_4 \quad a_3 \leq a_4}{f_3(a_3) \leq f_4(a_4)}\,(\text{App})}{f_1(a_1) \leq f_4(a_4)}\,(\text{Trans})$$

$$\dfrac{\dfrac{\dfrac{f_1 \leq f_2 \quad f_2 \leq f_3}{\boldsymbol{f_1 \leq f_3}}\,(\text{Trans}) \quad \boldsymbol{f_3 \leq f_4}}{\boldsymbol{f_1 \leq f_4}}\,(\text{Trans}) \quad \dfrac{\dfrac{a_1 \leq a_2 \quad a_2 \leq a_3}{\boldsymbol{a_1 \leq a_3}}\,(\text{Trans}) \quad \boldsymbol{a_3 \leq a_4}}{\boldsymbol{a_1 \leq a_4}}\,(\text{Trans})}{\boldsymbol{f_1(a_1) \leq f_4(a_4)}}\,(\text{App})$$

The idea is to apply (Trans) as early as possible.

In general, the terms involved may be more complex than those displayed in the example above. For instance, f_i might be replaced with $f_i(a_i)$, and a_i with $g_i(b_i)$. So consider the derivation above but where each f_i is replaced with some complex term F_i and each a_i with some complex term A_i. Moreover, assume that each $F_i \leq F_{i+1}$ and each $A_i \leq A_{i+1}$ is derived rather than being an axiom. We now need to use induction on derivation height.[5] Note that the height of $F_1 \leq F_4$ in the final derivation is less than the that of $F_1(A_1) \leq F_4(A_4)$ in the original derivation, so we can apply the inductive hypothesis to $F_1 \leq F_4$, and similarly, to $A_1 \leq A_4$, and this gives us a subterm derivation because F_1, F_4, A_1, A_4 are all subterms.

We can further generalize and view $(F_1(A_1), F_2(A_2), F_3(A_3), F_4(A_4))$ as a subterm segment in a larger subterm-separated derivation. Then we can iteratively apply the same argument to each subterm segment.

Corollary 1. $\{(\Gamma, S, T) \colon \Gamma \vdash_{\mathcal{M}_+} S \leq T\}$ *is decidable.*

Proof. From Proposition 1, Lemma 1, and Theorem 1, it follows that $\Gamma \vdash_{\mathcal{M}_+} S \leq T$ if and only if $\Gamma \vdash_{\mathcal{M}'_+} S \leq T$ via a subterm derivation. Also, note that for any derivation $\boldsymbol{D}$ of $S \leq T$, if the conclusion $S \leq T$ appears in a proper subderivation $\boldsymbol{D''}$ within $\boldsymbol{D}$, then $\boldsymbol{D''}$ is also a derivation of $S \leq T$, so the remaining portion of $\boldsymbol{D}$ was unnecessary. Similarly, for any subderivation $\boldsymbol{D'}$ of $S' \leq T'$ within the derivation $\boldsymbol{D}$ of $S \leq T$, it would be redundant if $S' \leq T'$ appeared in a proper subderivation $\boldsymbol{D''}$ within $\boldsymbol{D'}$. Let $\boldsymbol{D'} \subseteq \boldsymbol{D}$ denote that $\boldsymbol{D'}$ is a subderivation within $\boldsymbol{D}$, and let $\boldsymbol{D'} \subsetneq \boldsymbol{D}$ denote that $\boldsymbol{D'}$ is a proper subderivation within $\boldsymbol{D}$. Then, $\Gamma \vdash_{\mathcal{M}'_+} S \leq T$ via a subterm derivation if and only if there is a derivation of $S \leq T$ in the set $\boldsymbol{\mathcal{S}}$ of subterm derivations $\boldsymbol{D}$ such that there is no $\boldsymbol{D'} \subseteq \boldsymbol{D}$ for which there is $\boldsymbol{D''} \subsetneq \boldsymbol{D'}$ that contains the conclusion of $\boldsymbol{D'}$. In the language of trees, this is the set of all trees in which no branch has any inequality appearing more than once in it. Let N be the size of the subterm set of (Γ, S, T). Then there are at most N^2 subterm inequalities, so the height of a derivation in $\boldsymbol{\mathcal{S}}$ is at most N^2. Also, there is a finite set of proof rules, so $\boldsymbol{\mathcal{S}}$ is finite. Thus, we can enumerate all of its elements.

4 The General Case

We now return to discussing the monotonicity calculus $\mathcal{M}$. The goal is still to define a new proof system that is equivalent to $\mathcal{M}$ in some sense and prove that it has the subterm property. We begin by pointing out a key difference from the previous section and providing some intuition. In $\mathcal{M}$, we may have derivations of the form

$$\dfrac{\dfrac{\dfrac{\dfrac{}{A \leq B}}{F(A) \leq F(B)}\,(\text{Mono}) \qquad \dfrac{\dfrac{}{F^+ \leq G^-}}{F(B) \leq G(B)}\,(\text{Point})}{F(A) \leq G(B)}\,(\text{Trans}) \qquad \dfrac{\dfrac{}{C \leq B}}{G(B) \leq G(C)}\,(\text{Anti})}{F(A) \leq G(C)}\,(\text{Trans})$$

where B is not a subterm. Then $F(B)$ is also not a subterm, assuming it is not in an axiom. In the previous section, if we had the conclusion $F(A) \leq G(C)$

[5] In the full proof in the appendix, we will have to define a more nuanced notion of height in order for the induction to work.

then $F \le G$ and $A \le C$ must have both appeared in the derivation, so we could eliminate $F(B)$ from the derivation and then eliminate B by induction. But we see that this is not the case here. We want to be able to go directly from $F^+ \le G^-$ and $A \le B \ge C$ (where $A \le B \ge C$ is shorthand for $A \le B; C \le B$) to $F(A) \le G(C)$, but it appears that we do not have an immediate analogue for the step in the proof of Theorem 1 that allowed us to eliminate non-subterms.

To address this issue, we will introduce a new relation symbol $\lesseqgtr$ called a *zigzag*. Intuitively, $S_1 \lesseqgtr S_n$ represents $S_1 \le S_2 \ge \cdots \le S_n$ for some terms $S_2, \ldots, S_{n-1}$ which are abstracted away. Then, after introducing new (Trans) and (App) rules, we can get a derivation like

$$\dfrac{F^+ \le G^- \qquad \dfrac{A \le B \qquad B \ge C}{A \lesseqgtr C}\,(\text{Trans})}{F(A) \le G(C)}\,(\text{App})$$

because antitonic G reverses the direction of the second inequality. But this is just a simplification. In the general case, there are many components we have to consider, as discussed here below.

Call the left premise of (App) *function position* and the right premise *argument position*. Observe that in the example above, if G were monotonic, then we would be not be able to derive $F(A) \le G(C)$, and likewise, if we had $A \lesseqgtr C$ representing $A \le B_1 \ge B_2 \le C$ in argument position, we would also not be able to. So in argument position, we have to encode the number of inequality direction flips, and in function positon, we have to encode the number of polarity changes. Intuitively, more polarity changes and fewer inequality direction flips will yield stronger conclusions. We also have to encode the number of inequality direction flips in function position though because the conclusion of (App) is in general a zigzag, rather than an inequality, and the number of flips in function position affects the number of flips in the conclusion. If this conclusion occurs in argument position in a larger derivation, then it makes a difference.

We will replace all inequalities with zigzags and encode all the information discussed above in *information tables* to be defined below. Every zigzag $S_1 \lesseqgtr S_n$ is associated with an information table χ, and the idea is that information tables record intermediate computation (that would have appeared in $\mathcal{M}$) using finite resources – they record the polarity markings of the intermediate terms $S_1 \le S_2 \ge \cdots \le S_{n-1} \ge S_n$, rather than the terms themselves. Note that for a term with a higher-order type, the information table encodes all polarity markings in its type. Also, for uniformity, polarity changes are also encoded for zigzags that occur in argument position but can be ignored intuitively.

The definitions below formalize the details discussed above. They may seem slightly obscure and complicated at first, so we hope the intuition provided above is helpful.

Definition 8. *An* ***information table*** χ *is an* $m \times n$ *matrix* ($m \le L$)[6] *consisting of polarity markings* $p \in \{+, -\}$ *and inequality symbols* $* \in \{\le, \ge\}$ *such that for each column* j,

[6] L is maximum valence among all terms in the subterm set of (Γ, S, T).

– *if j is odd, the column corresponds to the polarity markings of some term S: for each $i \leq m$, $\chi_{i,j} = p_i$ where $\sigma_1 \xrightarrow{p_1} \dots \xrightarrow{p_{i-1}} \sigma_i \xrightarrow{p_i} \dots \xrightarrow{p_m} \sigma_{m+1}$ is the type of S; let $\boldsymbol{p}_S$ denote such a column*
– *if j is even, all entries in the column are the same inequality symbol: for each $i \leq m$, $\chi_{i,j} = *_j \in \{\leq, \geq\}$*

The entire information table can then be represented as $\chi = [\boldsymbol{p}_0 *_1 \boldsymbol{p}_1 *_2 \cdots *_n \boldsymbol{p}_n]$.

Definition 9. *In an information table χ, an **inequality flip** occurs at j if j is even and $*_j \neq *_{j-2}$. A **polarity change** occurs at (i,j) if j is odd and $\chi_{i,j} \neq \chi_{i,j-2}$.*

Definition 10. *A sequence of inequalities $S_0 *_1 S_1 *_2 \cdots *_n S_n$[7] in $\mathcal{M}$ **respects** χ if $\chi = [\boldsymbol{p}_{S_0} *_1 \boldsymbol{p}_{S_1} *_2 \cdots *_n \boldsymbol{p}_{S_n}]$.*

Definition 11. *$\mathcal{M}_\infty$ is the proof system with zigzags of the form $S \lesseqgtr_\chi T$ where χ is an information table. The proof rules are*

$$\frac{F \lesseqgtr_\phi G \qquad A \lesseqgtr_\psi B}{F(A) \lesseqgtr_{\phi(\psi)} G(B)} \text{(App)} \quad \frac{A \lesseqgtr_\phi B \qquad B \lesseqgtr_\psi C}{A \lesseqgtr_{\phi+\psi} C} \text{(Trans)} \quad \frac{}{A \lesseqgtr_{[\chi_A \leq \chi_A]} A} \text{(Refl)}$$

where $\phi + \psi$ is the concatenated matrix $[\phi\ \psi']$ where ψ' is ψ with its first column deleted. $\phi(\psi)$ is a set, which is defined below, so (App) represents a family of proof rules. For each proof rule, we also allow replacing the conclusion $S \lesseqgtr_\chi T$ with $S \lesseqgtr_{\chi'} T$ where χ' deletes any occurrences of $[\cdots *_i \boldsymbol{p}_i \dots]$ from χ for which $*_i = *_{i-1}$.

Given a set Γ of axioms in $\mathcal{M}$, the set of axioms $\Gamma_{\lesseqgtr}$ in $\mathcal{M}_\infty$ is

$$\{(S \lesseqgtr_{[\chi_S \leq \chi_T]} T) \colon (S \leq T) \in \Gamma\} \cup \{(T \lesseqgtr_{[\chi_T \geq \chi_S]} S) \colon (S \leq T) \in \Gamma\}$$

Definition 12. *Let $F * \cdots * G$ be a sequence of k (hypothetical) axioms in $\mathcal{M}$ that respects ϕ. Let $A * \cdots * B$ be a sequence of k' (hypothetical) axioms that respects ψ. $\boldsymbol{\chi \in \phi(\psi)}$ iff $\Gamma \vdash_{\mathcal{M}} F(A) * \cdots * G(B)$ for some sequence of $k + k'$ inequalities that respects χ. In particular, information tables in $\phi(\psi)$ have one less row than ϕ.* [8]

Example 2. Here is a procedure for visualizing Definition 12. Assume we have

$$\phi = \begin{bmatrix} + \leq_1 - \geq_2 + \\ + \leq_1 + \geq_2 - \end{bmatrix}, \ \psi = \begin{bmatrix} + \leq_3 + \geq_4 + \leq_5 + \geq_6 + \end{bmatrix}$$

We have labeled the inequality symbols for explanatory purposes – the indices do not have any meaning. Begin by writing out the sequence of inequality symbols from ϕ:

$$\boldsymbol{\leq_1 \ \geq_2}$$

[7] This is shorthand for $S_0 *_1 S_1; S_1 *_2 S_2; \dots ; S_{n-1} *_n S_n$.
[8] Note that since we are only encoding polarity markings rather than specific terms, the set $\phi(\psi)$ is finite.

Then insert the alternating inequality symbols from ψ into the sequence such that $\leq_3$ is before $\geq_4$ which is before $\leq_5$ which is before $\geq_6$:

$$\leq_3 \ \boldsymbol{\leq_1} \ \geq_4 \ \leq_5 \ \boldsymbol{\geq_2} \ \geq_6 \tag{1}$$

Since '$-$' appears between $\leq_1$ and $\geq_2$ in the first row of ϕ, reverse the inequalities between $\leq_1$ and $\geq_2$ in the sequence:

$$\leq_3 \ \boldsymbol{\leq_1} \ \leq_4 \ \geq_5 \ \boldsymbol{\geq_2} \ \geq_6$$

The polarity markings of the resulting information table χ are determined by the second row of ϕ:

$$\chi = \left[+ \leq_3 + \boldsymbol{\leq_1} + \leq_4 + \geq_5 + \boldsymbol{\geq_2} - \geq_6 -\right]$$

By Definition 11, we can reduce this to

$$\chi = \left[+ \leq + \geq + \geq -\right]$$

Note that we have reduced 3 flips in ψ to 1 flip in χ. To get the entire set $\phi(\psi)$, we consider all possible insertion orders of $\leq_3, \ldots, \geq_6$ in (1).

With regard to the extra implicit proof rule in Definition 11, it is straightforward to verify that deleting columns with the same inequality direction in information tables does not introduce additional derivable statements.

Lemma 2. *Given $\Gamma_{\lessgtr}$, let $\Gamma'_{\lessgtr} = \{(S \lessgtr_{\chi'} T) \colon (S \lessgtr_{\chi} T) \in \Gamma_{\lessgtr}\}$, where χ' is as described in Definition 11. If $\Gamma'_{\lessgtr} \vdash_{\mathcal{M}_\infty} U \lessgtr_{\phi} V$, then $\Gamma_{\lessgtr} \vdash_{\mathcal{M}_\infty} U \lessgtr_{\phi} V$.*

Proof. By induction on valence.

Another consequence of Definitions 11 and 12 is

Lemma 3. *Any $\chi \in \phi(\psi)$ has at least as many inequality flips as ϕ.*

This is because in $\mathcal{M}$, (Point) does not have an analogue like (Mono) does with (Anti).

The following shorthand will be useful: $S \lessgtr T$ means $S \lessgtr_{\chi} T$ for some χ. $S \lessgtr_k T$ (where $k \in \mathbb{N}$) means $S \lessgtr_{\chi} T$ for some χ that has k flips. $S \lessgtr_X T$ (where X is a set of information tables) means $S \lessgtr_{\chi} T$ for some $\chi \in X$.

What we have just defined is an infinite proof system. There are many proof rules that turn out to be unnecessary, and we will systematically eliminate them. We choose to use $\mathcal{M}_\infty$ as an intermediate system because it provides for a more natural flow in reasoning. But eventually, we will end with a finite proof system $\mathcal{M}_\Gamma$, to be defined below, in order to demonstrate decidability.

An outline of the proof of decidability is

1. $\Gamma \vdash_{\mathcal{M}} S \leq T$ if and only if $\Gamma_{\lessgtr} \vdash_{\mathcal{M}_\infty} S \lessgtr_0 T$
2. $\Gamma_{\lessgtr} \vdash_{\mathcal{M}_\infty} S \lessgtr_0 T$ if and only if $\Gamma_{\lessgtr} \vdash_{\mathcal{M}_\infty} S \lessgtr_0 T$ via a subterm derivation

3. $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_0 T$ via a subterm derivation if and only if $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\Gamma} S \lesseqgtr_0 T$ via a subterm derivation

with each roughly corresponding to one of the next three subsections.

Remark 1. The definitions of **subterm derivation**, **subterm property**, **bottommost chain**, **subterm-separated**, and **subterm segment** are analogous to the previous section. Then, we also have an analogue for Lemma 1.

4.1 Part One

We first show that $\mathcal{M}$ embeds into $\mathcal{M}_\infty$. The following result is analogous to the forward direction of Proposition 1.

Proposition 2. *If $\Gamma \vdash_{\mathcal{M}} S \leq T$, then $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_0 T$*

Corollary 2. *If $\Gamma \vdash_{\mathcal{M}} S \leq S_1 \geq \cdots \leq S_n \geq T$, then $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_n T$.*

Proof. By induction on n. Apply (Trans).

Proposition 3. *If $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_\chi T$, then $\Gamma \vdash_{\mathcal{M}} S \leq \cdots \geq T$ via a sequence of derivations that respects χ.*

Proof. By structural induction on the proof rules. Due to Lemma 2, this follows directly from Definitions 11 and 12.

4.2 Part Two

Theorem 2. *If $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_\chi T$ via a subterm-separated derivation, then $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_\chi T$ via a subterm derivation.*

Proof. The proof is almost identical to the proof of Theorem 1, but there is an extra consideration regarding information tables. Previously, we had a single inequality relation, but now we have a family of zigzag relations.

4.3 Part Three

Lemma 4. *If $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_\chi T$ via a subterm derivation, then $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_\psi T$ via a subterm derivation where ψ has at most 2^{N^2} inequality flips, where N is the size of the subterm set of $(\Gamma_{\lesseqgtr}, S, T)$.*

Proof. Reasoning in a similar way as in Corollary 1, if some subderivation of $S' \lesseqgtr_{\chi'} T'$ within the derivation of $S \lesseqgtr_\chi T$ contains $S' \lesseqgtr_{\chi''} T'$ (for some χ'' which is not necessarily χ') somewhere other than the conclusion, then there is a smaller subderivation of $S' \lesseqgtr T'$, and thus a smaller derivation of $S \lesseqgtr T$. So consider the set $\boldsymbol{S}^{\lesseqgtr}$ of subterm derivations with the property that any subderivation of $S' \lesseqgtr_{\chi'} T'$ does not contain $S' \lesseqgtr_{\chi''} T'$ (for any χ'') anywhere other than the conclusion in the subderivation. If $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_\chi T$ via a subterm derivation, then there must be a derivation of some $S \lesseqgtr T$ in this set. Now we just have to count the maximum number of flips for any derivation in $\boldsymbol{\mathcal{S}}^{\lesseqgtr}$. For any step, if the premises have k_1, k_2 flips, then the conclusion has at most $k_1 + k_2 + 1$ flips by Definition 11 and 12. Thus, the maximum number of flips for the conclusion of the entire derivation is at most the total number of instances of axioms in the derivation.[9] There are N^2 pairs of subterms, so the maximum number of terminal nodes (axioms) of a tree in $\boldsymbol{\mathcal{S}}^{\lesseqgtr}$ is 2^{N^2}.

We have just established that we only need to encode at most $K = 2^{N^2}$ inequality flips. In the following discussion, we bound the number of polarity columns we need to encode. Here is the essential idea: By Definition 12, $F \lesseqgtr_\chi G$, where χ has K polarity changes with the same inequality direction in its first row, is already sufficient to derive $F(A) \lesseqgtr_0 G(B)$ from $A \lesseqgtr_K B$ using (App). Now considering all $m \leq L$ rows of χ, each column has at most 2^L different possibilities, so it suffices to encode $2^L K$ polarity columns between each pair of inequality flips.

Lemma 5. *Let $K > 0$. Let χ' be the information table obtained by deleting all occurrences of $[\cdots *_k\, p_k \ldots]$ from χ for which there exist $n < k$ and a sequence $n = n_1 < \cdots < n_{K+1} = k$ such that*

1. *$*_n = *_{n+1} = \cdots = *_k$ and*
2. *$p_{n_1} = p_{n_2} = \cdots = p_{n_{K+1}}$*

If ψ has at most K inequality flips and

$$\{F \lesseqgtr_\phi G, A \lesseqgtr_\psi B\} \vdash F(A) \lesseqgtr_\chi G(B)$$

via (App), then

$$\{F \lesseqgtr_{\phi'} G, A \lesseqgtr_\psi B\} \vdash F(A) \lesseqgtr_{\chi'} G(B)$$

via (App).

Definition 13. *$\mathcal{M}_\Gamma$ is the restriction of $\mathcal{M}_\infty$ to zigzags whose information tables have at most $K = 2^{N^2}$ inequality flips and at most $2^L K^2$ polarity columns, where N is the size of the subterm set of $(\Gamma_{\lesseqgtr}, S, T)$.*

Proposition 4. *Let χ be an information table with at most K inequality flips. If $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\infty} S \lesseqgtr_\chi T$ via a subterm derivation, then $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\Gamma} S \lesseqgtr_{\chi'} T$ via a subterm derivation, where χ' is as defined in Lemma 5.*

Proof. The proof is by induction on derivation height. First consider the case where (Trans) is the last step in the derivation. Then we have

$$\dfrac{\dfrac{}{S \lesseqgtr_\phi U} \qquad \dfrac{}{U \lesseqgtr_\psi T}}{S \lesseqgtr_\chi T}\ \text{(Trans)}$$

[9] Note that a zigzag may occur multiple times in a derivation as long as none of its instances are part of the same branch. In certain cases, we have to instantiate the same axiom multiple times.

It follows by Definition 11 that ϕ, ψ have at most K inequality flips, so the result follows by the inductive hypothesis. In the other case, (App) is the last step in the derivation:

$$\dfrac{\overline{F \lesseqgtr_\phi G} \qquad \overline{A \lesseqgtr_\psi B}}{S = F(A) \lesseqgtr_\chi G(B) = T}\ \text{(App)}$$

It follows by Lemma 3 that ϕ has at most K flips. Additionally, we can assume that ψ has at most K flips by Lemma 4. By the inductive hypothesis, $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\Gamma} F \lesseqgtr_{\phi'} G$ via a subterm derivation. Then $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\Gamma} F(A) \lesseqgtr_{\chi'} G(B)$ by Lemma 5.

4.4 Putting It Together

Finally, here is the result we want.

Theorem 3. *$\Gamma \vdash_{\mathcal{M}} S \leq T$ if and only if $\Gamma_{\lesseqgtr} \vdash_{\mathcal{M}_\Gamma} S \lesseqgtr_0 T$ via a subterm derivation.*

Proof. Refer back to the outline at the beginning of the section. For the forwards direction, (1) is by Proposition 2, (2) is by Theorem 2 and the analogue of Lemma 1 mentioned in Remark 1, and (3) is by Proposition 4. For the backwards direction, (3) and (2) are trivial, and (1) is by Proposition 3.

Corollary 3. *$\{(\Gamma, S, T) : \Gamma \vdash_{\mathcal{M}} S \leq T\}$ is decidable.*

Proof. Let M be the number of information tables in $\mathcal{M}_\Gamma$, which is finite by Definition 13. Then there are at most MN^2 zigzags $S \lesseqgtr T$ where S, T are subterms. Since there is a finite number of proof rules in $\mathcal{M}_\Gamma$, we can replicate the argument in the proof of Corollary 1 by replacing N^2 with MN^2. Then the result follows by Theorem 3.

References

1. Fyodorov, Y., Winter, Y., Francez, N.: Order-based Inference in Natural Logic. Logic Journal of the IGPL **11**(4), 385–416 (2003)
2. Icard, T., Moss, L.: Recent Progress on Monotonicity. Linguistic Issues in Language Technology **9**, 167–194 (2014)
3. Icard, T., Moss, L., Tune, W.: A Monotonicity Calculus and Its Completeness. In: Proceedings of the 15th Meeting on the Mathematics of Language, pp. 75–87. Association for Computational Linguistics, London (2017). https://doi.org/10.18653/v1/W17-3408
4. Kennedy, A., Pierce, B.: On Decidability of Nominal Subtyping with Variance. In: International Workshop on Foundations and Developments of Object-Oriented Languages (2007)
5. Kozen, D.: Complexity of Finitely Presented Algebras. In: Proc. 9th Symposium on Theory of Computing, pp. 165–177. Association for Computing Machinery (1977)

A Proof of Main Theorem

Definition 14. *The **flattened height** of a subterm-separated derivation of $S_1 \leq S_n$, denoted $\boldsymbol{f}(S_1 \leq S_n)$, is defined as follows. If the last step of the derivation is (Trans), let $(S_1, \ldots, S_n)$ be the bottommost chain. If $(S_1, \ldots, S_n)$ is a subterm segment, then*

$$\boldsymbol{f}(S_1 \leq S_n) = \max\{\boldsymbol{f}(S_k \leq S_{k+1})\colon k = 1, \ldots, n-1\} + 1$$

Otherwise, let $(S_1 = S^{(1)}, \ldots, S^{(n')} = S_n)$ be the bottommost subterm chain, and then

$$\boldsymbol{f}(S_1 \leq S_n) = \max\{\boldsymbol{f}(S^{(k)} \leq S^{(k+1)})\colon k = 1, \ldots, n'-1\} + 1.^{10}$$

If the last step of the derivation is not (Trans), the flattened height is defined in the same way as the height.

Essentially, we want to mimic a proof system that allows (Trans) to have any number of premises, without actually defining an infinite set of proof rules. Then we can think of each subterm-separated bottommost chain as having the following form. This definition is for technical reasons in our proof by induction below.

$$\dfrac{\dfrac{\dfrac{}{S_1^{(1)} \leq S_2^{(1)}}\,(\text{App}) \quad \cdots \quad \dfrac{}{S_{n_1-1}^{(1)} \leq S_{n_1}^{(1)}}\,(\text{App})}{S_1^{(1)} \leq S_{n_1}^{(1)}}\,(\text{Trans}) \quad \cdots \quad \dfrac{\dfrac{}{S_1^{(i)} \leq S_2^{(i)}}\,(\text{App}) \quad \cdots \quad \dfrac{}{S_{n_i-1}^{(i)} \leq S_{n_i}^{(i)}}\,(\text{App})}{S_1^{(i)} \leq S_{n_i}^{(i)}}\,(\text{Trans})}{S_1^{(1)} \leq S_{n_i}^{(i)}}\,(\text{Trans})$$

Theorem 4. *If $\Gamma \vdash_{\mathcal{M}'_+} S \leq T$ via a subterm-separated derivation, then $\Gamma \vdash_{\mathcal{M}'_+} S \leq T$ via a subterm derivation.*

Proof. The proof is by nested induction. We will induct on h, the flattened height of the derivation, and i, the size of the bottommost subterm chain. We will also induct on n with respect to another claim we discuss below. Here is an outline of the proof (where the bolded step contains the main contents).

1. Base case $h = 1$
2. Induction on h
   (a) Base case $i = 2$
      i. Base case $n = 2$
      ii. **Induction on n**
   (b) Induction on i

We now begin the proof. The base case $h = 1$ is trivial, since a derivation that has (flattened) height 1 is either an axiom or follows by a use of (Refl).

Now assume the statement is true for any derivation with flattened height $h' < h$. Consider a derivation of $S \leq T$ with flattened height h. Consider the base case $i = 2$. In this case, the bottommost chain is a single subterm segment.

[10] Equivalently, $\boldsymbol{f}(S_1 \leq S_n) = \max\{\boldsymbol{f}(S_k \leq S_{k+1})\colon k = 1, \ldots, n-1\} + 2$.

Claim. For any $n \geq 2$, a subterm-separated derivation whose bottommost chain $(S_1, \ldots, S_n)$ is a single subterm segment of length n, has a subterm-separated derivation whose last step is (App) and moreover has at most the same flattened height (excluding the case where $n = 2$ and $S_1 \leq S_2$ is an axiom).

Proof. By induction on n. For the base case, consider $S_1 \leq S_2$ and assume it is not an axiom. Then, the last step of the derivation is not (Trans) by Definition 5, so it is (App).

For the induction, fix $n > 2$ and assume the claim for any $n' < n$. Consider a derivation of $S_1 \leq S_n$ with bottommost chain $(S_1, \ldots, S_n)$. The last step of the derivation is (Trans) since $n > 2$. Let $S_1 \leq S_k$ and $S_k \leq S_n$ (where $k \in \{2, \ldots, n-1\}$ and S_k is not a subterm) be the premises of this step. By the inductive hypothesis on n, we can obtain subterm-separated derivations of $S_1 \leq S_k$ and $S_k \leq S_n$ where the last step is (App), without increasing their flattened heights. So $S_1 = F_1(A_1), S_k = F_k(A_k), S_n = F_n(A_n)$ for some $F_1, A_1, F_k, A_k, F_n, A_n$:

$$\dfrac{\dfrac{\overline{F_1 \leq F_k} \quad \overline{A_1 \leq A_k}}{F_1(A_1) \leq F_k(A_k)}\,(\text{App}) \qquad \dfrac{\overline{F_k \leq F_n} \quad \overline{A_k \leq A_n}}{F_k(A_k) \leq F_n(A_n)}\,(\text{App})}{F_1(A_1) \leq F_n(A_n)}\,(\text{Trans}) \tag{2}$$

Since the derivations of $F_1(A_1) \leq F_k(A_k)$ and $F_k(A_k) \leq F_n(A_n)$ are subterm-separated, so are the derivations of $F_1 \leq F_k$, $A_1 \leq A_k$, $F_k \leq F_n$, and $A_k \leq A_n$ shown above. Let

$$(F_1 = F^{(1)}, F^{(2)}, \ldots, F^{(k'-1)}, F^{(k')} = F_k) \tag{3}$$

be the bottommost subterm chain of $F_1 \leq F_k$, and let

$$(F_k = F^{(k')}, F^{(k'+1)}, \ldots, F^{(n'-1)}, F^{(n')} = F_n) \tag{4}$$

be the bottommost subterm chain of $F_k \leq F_n$.

Let h be the flattened height of $F_1(A_1) \leq F_n(A_n)$ in (2). Since

$$(F_1(A_1), F_k(A_k), F_n(A_n))$$

is the bottommost chain, the subderivation of $F_1(A_1) \leq F_k(A_k)$ has flattened height at most $h - 1$, so then the subderivation of $F_1 \leq F_k$ has flattened height at most $h - 2$. Similary, the subderivation of $F_k \leq F_n$ has flattened height at most $h - 2$. Now we proceed by cases.

<u>Case 1</u>: (F_k is a subterm of (Γ, F_1, F_n)) Then we can obtain a derivation

$$\dfrac{\overline{F_1 \leq F_k} \quad \overline{F_k \leq F_n}}{F_1 \leq F_n}\,(\text{Trans})$$

which is subterm-separated, since the subderivations of $F_1 \leq F_k$ and $F_k \leq F_n$ were already subterm-separated in (2) and F_k is a subterm. We also already know that these two subderivations have flattened height at most $h-2$. We now have three subcases: 1) The bottommost chains of both of these subderivations have multiple subterm segments; 2) Both of them have a single subterm segment; and 3) Exactly one of them has a single subterm segment. In the first subcase,

$$\max\{\boldsymbol{f}(F^{(j)} \leq F^{(j+1)})\colon j = 1, \ldots, k'-1\} = \boldsymbol{f}(F_1 \leq F_k) - 1 \leq h-3$$

$$\max\{\boldsymbol{f}(F^{(j)} \leq F^{(j+1)})\colon j = k', \ldots, n'-1\} = \boldsymbol{f}(F_k \leq F_n) - 1 \leq h-3$$

Since $(F^{(1)}, \ldots, F^{(k')}, \ldots, F^{(n')})$ is the bottommost subterm chain of $F_1 \leq F_n$, it follows that the derivation has flattened height $\max\{\boldsymbol{f}(F^{(j)} \leq F^{(j+1)})\colon j = 1, \ldots, n'-1\} + 1 \leq h-2$. We can also verify that the derivation has flattened height at most $h-1$ in the other subcases by similar calculations.

Case 2: (F_k is not a subterm of (Γ, F_1, F_n)) Again, consider the subcase where the bottommost chain of $F_1 \leq F_k$ and $F_k \leq F_n$ both consist of multiple subterm segments. The derivation in (2) contains subderivations of $F^{(k'-1)} \leq F_k$ and $F_k \leq F^{(k'+1)}$, from (3) and (4), and the bottommost chain of both of these subderivations are subterm segments. Then we can obtain a derivation

$$\dfrac{\dfrac{\overline{F_1 \leq F^{(k'-1)}} \quad \dfrac{\overline{F^{(k'-1)} \leq F_k} \quad \overline{F_k \leq F^{(k'+1)}}}{F^{(k'-1)} \leq F^{(k'+1)}}\,(\text{Trans})}{F_1 \leq F^{(k'+1)}}\,(\text{Trans}) \quad \overline{F^{(k'+1)} \leq F_n}}{F_1 \leq F_n}\,(\text{Trans})$$

where the bottommost subterm chain of $F_1 \leq F^{(k'-1)}$ is the subsequence

$$(F^{(1)}, \ldots, F^{(k'-1)})$$

of (3) and the bottommost subterm chain of $F^{(k'+1)} \leq F_n$ is is the subsequence $(F^{(k'+1)}, \ldots, F_n)$ of (4). Note that the bottommmost subterm chain of $F_1 \leq F_n$ is $(F^{(1)}, \ldots, F^{(k'-1)}, F^{(k'+1)}, \ldots, F^{(n')})$ since F_k is not a subterm, so this derivation is subterm-separated. As in Case 1,

$$\boldsymbol{f}(F^{(j)} \leq F^{(j+1)}) \leq h-3 \text{ for any } j = 1, \ldots, n'-1 \tag{5}$$

In particular, $\boldsymbol{f}(F^{(k'-1)} \leq F_k) \leq h-3$ and $\boldsymbol{f}(F_k \leq F^{(k'+1)}) \leq h-3$. Since F_k is not a subterm, observe that the bottommost chain of $F^{(k'-1)} \leq F^{(k'+1)}$, which is the concatenation of the bottommost chains of $F^{(k'-1)} \leq F_k$ and $F_k \leq F^{(k'+1)}$, is a single subterm segment, so then

$$\boldsymbol{f}(F^{(k'-1)} \leq F^{(k'+1)}) \leq h-3 \tag{6}$$

Finally, the derivation of $F_1 \leq F_n$ has flattened height

$$\max\{\boldsymbol{f}(F^{(k'-1)} \leq F^{(k'+1)}); \boldsymbol{f}(F^{(j)} \leq F^{(j+1)})\colon j = 1, \ldots, k'-2, k'+1, \ldots, n'-1\}+1$$

which is at most $h-2$ by (5) and (6).[11] We can verify that the other subcases, as in Case 1, also yield derivations with flattened height at most $h-1$.

[11] Note that we may have potentially modified the *height* of the derivation but not the *flattened height*. This is precisely why we introduced the notion of flattened height.

Thus, in all cases, we have found a subterm-separated derivation of $F_1 \leq F_n$ with flattened height at most $h-1$. By similar reasoning, we can find a subterm-separated derivation of $A_1 \leq A_n$ with flattened height at most $h-1$. It follows that

$$\dfrac{\dfrac{}{F_1 \leq F_n} \qquad \dfrac{}{A_1 \leq A_n}}{F_1(A_1) \leq F_n(A_n)}\ \text{(App)}$$

has flattened height at most h. □

Now, going back to the induction on h and base case on i, we can obtain a derivation of $S = F_1(A_1) \leq F_n(A_n) = T$ with flattened height h. Then we can apply the inductive hypothesis on h to $F_1 \leq F_n$ and $A_1 \leq A_n$, since they are subterm-separated, giving us a subterm derivation of $F_1(A_1) \leq F_n(A_n)$, as F_1, F_n, A_1, A_n are all subterms of $F_1(A_1)$ or $F_n(A_n)$.

For the induction on i, fix $i > 2$ and assume any subterm-separated derivation with $i-1$ subterms in its bottommost subterm chain has a subterm derivation. Consider a subterm-separated derivation with bottommost subterm chain $(S_1, \ldots, S_i)$. Without loss of generality, assume this derivation contains a subderivation of $S_1 \leq S_{i-1}$. By the inductive hypothesis, there is a subterm derivation of $S_1 \leq S_{i-1}$. By the base case, there is a subterm derivation of $S_{i-1} \leq S_i$. Then, we can obtain a subterm derivation of $S_1 \leq S_i$ by applying (Trans). ■