Normal forms and linearity over nonflat domains

Basil A. Karádais

DRAFT OF 23 APR 2015, 1:15 P.M.

Abstract

We discuss normal forms of finite approximations of objects, in type systems which are modeled over nonflat base domains. We use our results to investigate the status of linearity: we show that we can work linearly in a systematic way within the nonlinear model, as well as restrict to a fully linear model whose ideals are in a bijective correspondence with the ones of the nonlinear.

1 Introduction

Thinking about computability in a practical way means, among other things, striving to reason as finitarily as it gets. In domain-theoretic denotational semantics, we understand a higher-order program through a collection of *approximations*, that is, partial descriptions of its input-output behavior, embodying consistent and complete information about it. Trying to get as finitary as possible, we base our model on truly *finite* approximations of programs, that is, finite sets of information tokens, and then work with an appropriate domain representation, as pioneered by Dana Scott in [24]: two finite approximations may give consistent information, and the information of one may entail the information of the other. The denotation of a program is then retrieved as a consistent and deductively closed set of tokens. Domain-theoretically, the deductive closures of finite approximations provide us with the compacts, and topologically, their upper cones provide us with a basis, hence their also being called *formal neighborhoods*.

A type system in this context is set up over inductively generated *Scott information systems* serving as interpretations for the base types, where the tokens—the atomic approximations that make up formal neighborhoods—are generated as a free algebra by constructors. One of the fundamental choices here concerns how exactly we want to let partiality enter the model. A fairly mainstream approach is to simply introduce *partiality as a pseudotoken*, which in particular does not participate in the formation of the other tokens, and so to end up with *flat domains* as base types. The semantics here is "strict": a constructor induces a non-injective mapping and different constructors have overlapping ranges. Nevertheless, the model is refined enough to allow addressing deep questions, like the issue of sequentiality, or the full abstraction for PCF, in an intelligent and informative way.

Another approach which has been drawing growing interest in the recent past—a fact reflected for example on the advent of realistic non-strict programming languages, like Haskell—is to introduce *partiality as a pseudoconstructor*, therefore allowing it to participate in the formation of the rest of the tokens, and ending up with *nonflat domains* as base types. The constructors regain their injectivity and their non-overlapping ranges, at the cost of making us work, already at base types, within a nontrivial, and in fact, pretty involved preordered set instead of a flat tree. Things get very combinatorial very quickly, and old answers may need novel tools to reestablish in the nonflat case, especially since we're interested in a constructive development.

In particular, a strand of this kind of research certainly tries to exploit the extra structure, hoping for results which wouldn't hold in the flat case. Martin Escardo examines such possibilities already in [6], where he shows that characteristic functions which fail to be computable in the flat domain of natural numbers, become computable when elevated to the corresponding nonflat domain. Another strand, advanced by the Munich logic group, involves adapting fundamental results, like computational adequacy, definabilility, and density, to the nonflat case [22, 19, 9, 11] (see also [23]), and also recasting previous approaches to important topics in the nonflat setting in an arguably more natural way—a recent example being work on exact real arithmetic by exploiting the inherent base-type non-strictness coinductively [14, 13]. It turns out that these two strands occasionally meet: Davide Rinaldi and the author have independently observed that the typical Berger-like argument for the Kleeny–Kreisel density theorem [1], can be recast in the nonflat setting in a way that it provides finite witnesses [10, 20]. These pages present some observations and techniques concerning formal neighborhoods, in type systems interpreted upon nonflat information systems, which have been largely developed to help attack certain general questions like the above.

We begin in section 2 with some basic facts regarding our chosen model. Since the first obstacle in our considerations appears in the potentially highly complicated guises of information, we investigate ways of finding *normal forms of neighborhoods*. In order to do this, we revisit in section 3 the notion of *neighborhood mapping* [11] and its appropriate particular notion of continuity. Then, in section 4, we use information-preserving neighborhood mappings, which, additionally, send equivalent arguments to the same value, thus providing us with normal forms. At base types, we discuss four distinct normal forms, namely, the straightforward supremum and deductive closure, together with a *path* and a *tree form*. Then we move on to higher types, where we introduce a streamlined version of *eigen-neighborhood* [11], and use it to establish a *higher-type normal form theorem*: normal forms at lower types induce normal forms at higher types.

We then turn to the issue of *linearity*, which we can also think of as *atomicity*: the entailment of a token by a neighborhood depends on only one token in the neighborhood. As we will recount in section 5, this property is tied quite naturally to the research on such topics as sequentiality and linear logic, since it is a statement of *linearity of entailment* brought down to the level of domain representations. We use our results on normal forms to extricate a notion of linearity from our general model, and show how to make this explicit by restricting it to appropriate subsystems.

We end in section 6 with a short discussion on future work.

2 Background

For the purposes of this paper it is enough to work within a fragment of the system of Schwichtenberg–Wainer [23, §6.1.4]. In particular, we will omit parametric and simultaneously defined base types, and we will focus on the finitary aspect of the fragment. Since by indulging ourselves in such simplifications we do not do justice to the theory, the interested reader should turn to [23] for a more complete exposition.

Types

Our types are built simultaneously by three rules, one for constructor types, one for base types, and one for higher types. Let ξ be a distinct type variable, to be used as a dummy variable.

- If $\vec{\sigma}_0, \dots, \vec{\sigma}_{n-1}$ are types for $n \geqslant 0$, then $(\vec{\sigma}_0 \to \xi) \to \cdots \to (\vec{\sigma}_{n-1} \to \xi) \to \xi$ is a *constructor type (of arity n)*. The type is called *finitary* if all σ_ν's are empty, and *infinitary* otherwise.

- If $\kappa_0, \dots, \kappa_{k-1}$ are constructor types for $k > 0$ and one of them nullary, then $\mu_\xi(\kappa_0, \dots, \kappa_{k-1})$ is a *type*.

- If ρ, σ are types then $\rho \to \sigma$ is a *type*.

By convention, we will use ι, η to denote arbitrary *base* types and ρ, σ to denote arbitrary types in general.

Examples of base types are the *unit type* $\mathbb{U} := \mu_\xi(\xi)$ with only one constructor, the *type of boolean values* $\mathbb{B} := \mu_\xi(\xi, \xi)$, with constructors $\mathtt{tt} : \mathbb{B}$ and $\mathtt{ff} : \mathbb{B}$, the *type of natural numbers* $\mathbb{N} := \mu_\xi(\xi, \xi \to \xi)$, with constructors $\mathtt{0} : \mathbb{N}$, $\mathtt{S} : \mathbb{N} \to \mathbb{N}$, the infinitary *type of ordinal numbers* $\mathbb{O} := \mu_\xi(\xi, \xi \to \xi, (\mathbb{N} \to \xi) \to \xi)$, with constructors $\mathtt{0} : \mathbb{O}$, $\mathtt{S} : \mathbb{O} \to \mathbb{O}$, and $\cup : (\mathbb{N} \to \mathbb{O}) \to \mathbb{O}$, and the *type of (extended) derivations* $\mathbb{D} := \mu_\xi(\xi, \xi, \xi \to \xi, \xi \to \xi \to \xi)$, with constructors $\mathtt{0} : \mathbb{D}$, $\mathtt{1} : \mathbb{D}$, $\mathtt{S} : \mathbb{D} \to \mathbb{D}$, and $\mathtt{B} : \mathbb{D} \to \mathbb{D} \to \mathbb{D}$.

Information systems

A *(Scott) information system* is a triple $(\mathrm{Tok}, \mathrm{Con}, \vdash)$, where Tok is a countable set of *tokens*, Con is a collection of finite sets of tokens which we call *consistent sets* or *(formal) neighborhoods*, and $\vdash$ is a subset of $\mathrm{Con} \times \mathrm{Tok}$, the *entailment*. These are subject to the axioms

$$
\begin{aligned}
&\{a\} \in \mathrm{Con}, \\
&U \subseteq V \wedge V \in \mathrm{Con} \to U \in \mathrm{Con}, \\
&U \in \mathrm{Con} \wedge a \in U \to U \vdash a, \\
&U \vdash V \wedge V \vdash c \to U \vdash c, \\
&U \in \mathrm{Con} \wedge U \vdash b \to U \cup \{b\} \in \mathrm{Con},
\end{aligned}
$$

where $U \vdash V$ stands for $U \vdash b$ for all $b \in V$. From the latter follows vacuously that $U \vdash \varnothing$ for all U, while $\varnothing \in \mathrm{Con}$ follows from the first two axioms. Occasionally we will use finite sets of tokens Γ which are not necessarily consistent, for which we write Fin, so $\mathrm{Con} \subseteq \mathrm{Fin}$. An information system is called *coherent* when in addition to the above it satisfies

$$\bigvee_{a,a' \in U} \{a, a'\} \in \mathrm{Con} \to U \in \mathrm{Con}. \tag{1}$$

By the coherence and the second axiom above, it follows that the consistency of a set of tokens is equivalent to the consistency of its pairs. Drawing on this property, we often write $a \asymp b$ for $\{a, b\} \in \mathrm{Con}$, and even $U \asymp V$ for $U \cup V \in \mathrm{Con}$.[1] In the following we restrict our attention to coherent systems.

[1] This is sometimes written as $U \uparrow V$.

Given two coherent information systems A and B, we form their *function space* $A \to B$: define its tokens by $\langle U, b\rangle \in \mathrm{Tok}$ if $U \in \mathrm{Con}_A$ and $b \in \mathrm{Tok}_B$, its consistency by $\langle U, b\rangle \asymp \langle U', b'\rangle$ if $U \asymp_A U'$ implies $b \asymp_B b'$, and its entailment by $W \vdash \langle U, b\rangle$ if $WU \vdash_B b$, where

$$b \in WU := \underset{U' \in \mathrm{Con}_A}{\exists} \left(\langle U', b\rangle \in W \wedge U \vdash_A U' \right).$$

The last operation is called *neighborhood application*; one can show that it is monotone in both arguments, that is, that $U \vdash U'$ implies $WU \vdash WU'$ and that $W \vdash W'$ implies $WU \vdash WU'$, for all appropriate U, U', W, W'.

Fact 2.1. *The function space of two coherent systems is itself a coherent information system.*

Information systems as representations of domains

An *ideal* (or *element*) of an information system ρ is a possibly infinite set of tokens $x \subseteq \mathrm{Tok}$, such that $U \in \mathrm{Con}$ for every $U \subseteq_f x$ (consistency), and $U \vdash b$ for some $U \subseteq_f x$ implies $b \in x$ (deductive closure). If x is an ideal of ρ, we write $x \in \mathrm{Ide}_\rho$ or $x : \rho$. Note that there is an empty ideal $\perp_\rho = \varnothing$ at every type ρ.

By a *(Scott–Ershov) domain* (with a countable basis) we mean here a directed complete partial order, which is additionally algebraic and bounded complete. It is furthermore *coherent* [18], if every set of compacts has a least upper bound exactly when each of its *pairs* has a least upper bound. The following fact is fundamental to our approach.

Fact 2.2 (Representation theorem). *Let $\rho = (\mathrm{Tok}_\rho, \mathrm{Con}_\rho, \vdash_\rho)$ be a coherent information system. Then $(\mathrm{Ide}_\rho, \subseteq, \varnothing)$ is a coherent domain with compacts given by $\left\{ \overline{U} \mid U \in \mathrm{Con}_\rho \right\}$. Conversely, every coherent domain can be represented by a coherent information system.*

An *approximable mapping* between two information systems ρ and σ is a relation $r \subseteq \mathrm{Con}_\rho \times \mathrm{Con}_\sigma$, that generalizes entailment in the following sense: $\langle \varnothing, \varnothing \rangle \in r$; if $\langle U, V_1\rangle, \langle U, V_2\rangle \in r$ then $\langle U, V_1 \cup V_2\rangle \in r$; and if $U \vdash_\rho U'$, $\langle U', V'\rangle \in r$, and $V' \vdash_\sigma V$, then $\langle U, V\rangle \in r$. One can show [24] that there is a bijective correspondence between the approximable mappings from ρ to σ and the ideals of the function space $\rho \to \sigma$, and moreover establish the categorical equivalence between domains with Scott continuous functions and information systems with approximable mappings. The equivalence is preserved if we restrict ourselves to the coherent case [12].

Information systems as interpretations of types

We assign an information system to each type. Every higher type is naturally assigned a function space, so it suffices to discuss the information systems for *finitary* algebras. Let ι be a finitary algebra given by certain constructors. Each constructor C comes with an arity $r \geqslant 0$ (and at least one of them equal to zero). To the given constructors we add an extra nullary pseudoconstructor $*$ to denote partiality. Write $U \sim V$ for $U \vdash V \wedge V \vdash U$. We define the following.

- If C is an r-ary constructor and $a_1, \ldots, a_r \in \mathrm{Tok}_\iota$ then $Ca_1 \cdots a_r \in \mathrm{Tok}_\iota$. For its *head* write $\mathrm{hd}(Ca_1 \cdots a_r) = C$; for its i-th component write $a(i)$, that is, $(Ca_1 \cdots a_r)(i) = a_i$ for $i = 1, \ldots, r$.

- It is $a \asymp_\iota *$ and $* \asymp_\iota a$ for all $a \in \mathrm{Tok}_\iota$. Furthermore, if C is an r-ary constructor and $a_1 \asymp_\iota b_1, \dots, a_r \asymp_\iota b_r$ then $Ca_1 \cdots a_r \asymp_\iota Cb_1 \cdots b_r$. Finally, it is $U \in \mathrm{Con}_\iota$ if $a \asymp_\iota a'$ for all $a, a' \in U$.
- It is $U \vdash_\iota *$ for all $U \in \mathrm{Con}_\iota$. Furthermore, if C is an r-ary constructor, $U_1, \dots, U_r \in \mathrm{Con}_\iota$ are inhabited and $U_1 \vdash_\iota b_1, \dots, U_r \vdash_\iota b_r$, then $U \vdash_\iota Cb_1 \cdots b_r$ for all $U \in \mathrm{Con}_\iota$ which are *sufficient for C on $U_1, \dots, U_r$*, in the sense that for each $i = 1, \dots, r$ and each $a_i \in U_i$ there exists an $a \in U$ such that $\mathrm{hd}(a) = C$ and $a(i) = a_i$. Finally, if $U \vdash_\iota b$, then also $U \cup \{*\} \vdash_\iota b$.

Note that the definition of Con_ι incorporates (1), so it follows that $\varnothing \vdash_\iota \{*\}$. Concerning the notion of sufficiency, note that (a) it is $U \sim_\iota CU_1 \cdots U_r$, whenever U is sufficient for C on $U_1, \dots, U_r$, where

$$CU_1 \cdots U_r := \left\{ Ca_1 \cdots a_r \mid a_1 \in U_1, \dots, a_r \in U_r \right\},$$

and (b) in case C is a proper constructor, U is sufficient for C on $U_1, \dots, U_r$ if and only if $U \cup \{*\}$ is, if and only if $U \backslash \{*\}$ is. More generally, every neighborhood U which is *nontrivial* (meaning $U \not\sim_\iota \{*\}$) is equivalent to one of the form $CU_1 \cdots U_r$: if

$$U \backslash \{*\} = \left\{ Ca_{11} \cdots a_{r1}, \dots, Ca_{1m} \cdots a_{rm} \right\},$$

we let $U_i := \left\{ a_{i1}, \dots, a_{im} \right\}$ for every $i = 1, \dots, r$. Finally, the finite set $CU_1 \cdots U_r$ is consistent if every U_i is consistent.

It is straightforward, but tedious, to check that all these make sense.

Fact 2.3. *Let ι be an algebra given by constructors. The triple $(\mathrm{Tok}_\iota, \mathrm{Con}_\iota, \vdash_\iota)$ is a coherent information system.*

Finally, note that at finitary base types if U is a consistent finite set then also its *deductive closure* $\overline{U}$ is a consistent finite set, where $b \in \overline{U}$ if and only if $U \vdash b$.

By straightforward induction on the formation of tokens we can show that finitary base types are antisymmetric *on tokens* (write $a \vdash_\rho b$ for $\{a\} \vdash_\rho b$).

Lemma 2.4 (Antisymmetry)**.** *Let ι be a finitary base type. For all tokens $a, b \in \mathrm{Tok}_\iota$, if $a \sim_\iota b$ then $a = b$.*

On the other hand, it is easy to see that base types are not antisymmetric *on neighborhoods*, since for example $\{\mathtt{S}*\} \sim_{\mathbb{N}} \{\mathtt{S}*, *\}$ and $\{\mathtt{B0}*, \mathtt{B}*\mathtt{1}\} \sim_{\mathbb{D}} \{\mathtt{B01}\}$. It follows that antisymmetry does not carry over to higher types, either for tokens or for neighborhoods. This already motivates our pursue of normal forms in section 4.

Measuring tokens

For the study of base-type normal forms we need to make some further elementary observations. Define the *height* $|a|$ of $a \in \mathrm{Tok}_\iota$ by

$$\begin{aligned} |*_\iota| &:= 0, \\ |Ca_1 \cdots a_r| &:= 1 + \max\left\{ |a_1|, \dots, |a_r| \right\}, \end{aligned}$$

and the *size* $\|a\|$ of $a \in \mathrm{Tok}_\iota$ to be the number of its proper constructors:

$$\begin{aligned} \|*\| &:= 0, \\ \|Ca_1 \cdots a_r\| &:= 1 + \|a_1\| + \cdots + \|a_r\|. \end{aligned}$$

Proposition 2.5. *Let ι be a finitary algebra and $a, b, a_1, \ldots, a_r \in \mathrm{Tok}_\iota$. The following hold.*

1. *If $a \vdash_\iota b$ then $|a| \geqslant |b|$ and $\|a\| \geqslant \|b\|$.*
2. *It is $|a| \leqslant \|a\|$. Moreover, it is $|a| = \|a\|$ if and only if*

$$\underset{b \in \mathrm{Tok}_\iota}{\forall} \left(|b| = |a| \to \|b\| \geqslant \|a\| \right). \tag{2}$$

3. *Let $m = \max\left\{|a_1|, \ldots, |a_r|\right\}$. If $m = \|a_1\| + \cdots + \|a_r\|$, then $|a_i| = \|a_i\|$ for all i among $1, \ldots, r$. Moreover, if $m > 0$, then there exists a unique i among $1, \ldots, r$, such that*

$$|a_i| = \|a_i\| = m \ \wedge \ \underset{j \neq i}{\forall} |a_j| = \|a_j\| = 0.$$

Proof. The formulas in 1 are shown by straightforward induction, as well as that $|a| \leqslant \|a\|$ in 2. We show that $|a| = \|a\|$ if and only if (2) holds. From left to right, let a be a token with height and size equal, and let b be such that $|b| = |a|$. It is

$$\|b\| \geqslant |b| = |a| = \|a\|.$$

For the other way around, if $a = *$, then it's immediate, while for $a = Ca_1 \cdots a_r$ with $|a| = n$, consider the token

$$b := \underbrace{C \cdots C}_{n} \vec{*};$$

it is $|b| = n$, so $\|a\| \leqslant \|b\|$ by (2); by the construction of b it is $\|a\| \leqslant |a|$; by the assumption we get $\|a\| = |a|$.

We show 3 by cases on m. If $m = 0$, then for all $i = 1, \ldots, r$ it is $|a_i| = 0$, that is, $a_i = *$, so also $\|a_i\| = 0$. If $m > 0$, then there are $i = 1, \ldots, r$, for which $|a_i| = m$; assume there are k such a_i's ($k > 0$), and let l be the sum of the heights of the rest, that is, of all a_j's with $|a_j| \neq m$; by (2) it is

$$\|a_1\| + \cdots + \|a_r\| \geqslant k \cdot m + l;$$

by the assumption we get $m \geqslant k \cdot m + l$, from which we obtain $k = 1$ and $l = 0$; which is exactly what we wanted. ☐

3 Neighborhood mappings

By way of heuristics, we'd rather avoid working with the whole class of approximable maps between two information systems. The reason is that we would like to spare ourselves having to check after the fact if the maps that we used were "finitary" enough. Instead, we concentrate on mappings that operate explicitly on formal neighborhoods, and so seem to fit our setting more naturally.

A *neighborhood mapping* from type ρ to type σ is a mapping that sends neighborhoods from Con_ρ to neighborhoods in Con_σ. Such a mapping f is *compatible (with equientailment)* if $f(U_1) \sim_\sigma f(U_2)$, whenever $U_1 \sim_\rho U_2$. It is *monotone* if $U_1 \vdash_\rho U_2$ implies $f(U_1) \vdash_\sigma f(U_2)$, and *consistent* if $U_1 \asymp_\rho U_2$ implies $f(U_1) \asymp_\sigma f(U_2)$.

All three of the above notions are fundamental to our development. Compatibility with equientailment is arguably a sine qua non, but, as should be expected, it is too weak to ensure either monotonicity or consistency; for example the mapping from $\mathrm{Con}_{\mathbb{B}}$ to $\mathrm{Con}_{\mathbb{B}}$ defined by

$$\begin{aligned} \varnothing, \{*_{\mathbb{B}}\} &\mapsto \{\mathtt{tt}\}, \\ \{*_{\mathbb{B}}, \mathtt{tt}\}, \{\mathtt{tt}\} &\mapsto \{\mathtt{ff}\}, \\ \{*_{\mathbb{B}}, \mathtt{ff}\}, \{\mathtt{ff}\} &\mapsto \{\mathtt{tt}\}, \end{aligned}$$

is compatible but neither monotone nor consistent. Furthermore, there are consistent mappings that are not monotone, like the mapping from $\mathrm{Con}_{\mathbb{B}}$ to $\mathrm{Con}_{\mathbb{B}}$ defined by

$$\begin{aligned} \varnothing, \{*_{\mathbb{B}}\} &\mapsto \{\mathtt{tt}\}, \\ \{*_{\mathbb{B}}, \mathtt{tt}\}, \{\mathtt{tt}\} &\mapsto \{*_{\mathbb{B}}\}, \\ \{*_{\mathbb{B}}, \mathtt{ff}\}, \{\mathtt{ff}\} &\mapsto \{*_{\mathbb{B}}\}, \end{aligned}$$

and, moreover, there are consistent mappings that are not even compatible (see example below). Not surprisingly, monotone mappings are the safest ones to work with.

Lemma 3.1. *Let* $f : \mathrm{Con}_\rho \to \mathrm{Con}_\sigma$ *be a neighborhood mapping.*

1. *It is monotone if and only if it is compatible with equientailment and* $f(U_1 \cup U_2) \vdash_\sigma f(U_1) \cup f(U_2)$ *for every* $U_1, U_2 \in \mathrm{Con}_\rho$ *with* $U_1 \asymp_\rho U_2$.
2. *If it is monotone, then it is also consistent.*

Proof. For 1, assume that f is compatible and satisfies the above condition; let $U_1, U_2 \in \mathrm{Con}_\rho$ with $U_1 \vdash_\rho U_2$; then $U_1 \sim_\rho U_1 \cup U_2$, and by compatibility $f(U_1) \sim_\sigma f(U_1 \cup U_2)$, so the assumption yields $f(U_1) \vdash_\sigma f(U_2)$. Conversely, assume that f is monotone; then compatibility is immediate, and letting $U_1, U_2 \in \mathrm{Con}_\rho$ with $U_1 \asymp_\rho U_2$, we have $U_1 \cup U_2 \vdash_\rho U_i$ for $i = 1, 2$, so $f(U_1 \cup U_2) \vdash_\rho f(U_i)$, by monotonicity. The statement 2 follows immediately. □

Example. At each type ρ, easy examples of neighborhood mappings are the *identity mapping* given by $U \mapsto U$, and the *constant mappings* given by $U \mapsto U_0$ for any fixed $U_0 \in \mathrm{Con}_\rho$; all of these are monotone.

The constructors of an algebra provide further natural examples of monotone neighborhood mappings. The constructor $\mathtt{B}$ in $\mathbb{D}$ for instance induces a mapping $\mathrm{Con}_{\mathbb{D}} \times \mathrm{Con}_{\mathbb{D}} \to \mathrm{Con}_{\mathbb{D}}$, by $(U_1, U_2) \mapsto \mathtt{B}U_1U_2$.[2] Another example of a rather useful monotone mapping is the *partial length* mapping for (finitary) base types. Consider the *token* mapping $\mathrm{pl} : \mathrm{Tok}_{\mathbb{D}} \to \mathrm{Tok}_{\mathbb{D}}$ given (intuitively) by $\mathrm{pl}(a) := \mathtt{S}^{|a|}*_{\mathbb{D}}$, and extend it to neighborhoods by letting

$$\begin{aligned} \varnothing &\mapsto \{*_{\mathbb{D}}\}, \\ \{U\} &\mapsto \{\mathrm{pl}(a) \mid a \in U\}. \end{aligned}$$

An example of a consistent mapping which is not compatible would be the rather crude detotalizing mapping $d(U) := U \backslash \{a \in U \mid a \text{ total}\}$ (a is *total* if it does not involve $*$); by using such a mapping we don't harm consistency, but we rather unwarrantedly lose information, since $\{\mathtt{S}*_{\mathbb{D}}, \mathtt{S0}\} \sim_{\mathbb{D}} \{\mathtt{S0}\}$, but $d(\{\mathtt{S}*_{\mathbb{D}}, \mathtt{S0}\}) = \{\mathtt{S}*_{\mathbb{D}}\} \not\sim_{\mathbb{D}} \varnothing = d(\{\mathtt{S0}\})$.

[2]This is a bit of a cheat here though, see section 6.

Finally, an example of a neighborhood mapping that is neither consistent nor compatible would be the following: if m_U is the maximum size to be found among the tokens of $U \in \mathrm{Con}_{\mathbb{D}}$, set $m(U) := \{\mathtt{S}^{m_U}\mathtt{0}\}$. Then $\{\mathtt{B0}*, \mathtt{B}*\mathtt{0}\} \sim_{\mathbb{D}} \{\mathtt{B00}\}$ but $m(\{\mathtt{B0}*, \mathtt{B}*\mathtt{0}\}) = \{\mathtt{SS0}\}$ and $m(\{\mathtt{B00}\}) = \{\mathtt{SSS0}\}$, which are neither equivalent nor even consistent. $\square$

Ideals from neighborhood mappings

Despite the merits of monotonicity, it turns out that the weaker property of consistency suffices for a neighborhood mapping, because it is exactly what we need to naturally capture the notion of continuity for ideals.

Recall that an ideal is a possible infinite set of tokens which is (a) consistent, that is, every two of its tokens are consistent, and (b) deductively closed, that is, if some finite part of it entails a token, then this token must also belong to it. The idea of a neighborhood mapping f is obviously to achieve these two requirements by appropriately working on the level of neighborhoods: intuitively, the (right-flattened) graph of f should correspond to an ideal. To ensure that (a) holds, it is fitting that we require consistency from f, but what about (b); should we require something more? We show that we don't.

Define the *idealization* $\hat{f}$ of a neighborhood mapping $f : \mathrm{Con}_\rho \to \mathrm{Con}_\sigma$ to be the token set

$$\hat{f} := \Big\{\langle U, b\rangle \in \mathrm{Tok}_{\rho\to\sigma} \mid \underset{U_1,\dots,U_m \in \mathrm{Con}_\rho}{\exists} \Big(U \vdash_\rho \bigcup_{j=1}^{m} U_j \wedge \bigcup_{j=1}^{m} f(U_j) \vdash_\sigma b\Big)\Big\}.$$

Note that the term $\bigcup_{j=1}^m f(U_j)$ in the definition (which accounts for the essential nonlinearity of our setting) is implicitly required to be consistent—otherwise it wouldn't be allowed to appear on the left of an entailment.

Proposition 3.2. *Let ρ, σ be types, and f be a neighborhood mapping at type $\rho \to \sigma$. Then $\hat{f}$ is an ideal if and only if f is consistent.*

Proof. Assume that $\hat{f}$ is an ideal, and let $U_1, U_2 \in \mathrm{Con}_\rho$, with $U_1 \asymp_\rho U_2$. Since $U_i \vdash_\rho U_i$ and $f(U_i) \vdash_\sigma f(U_i)$ for each $i = 1, 2$, it is $\langle U_i, f(U_i)\rangle \subseteq \hat{f}$, by the definition of idealization, so the consistency of $\hat{f}$ yields $f(U_1) \asymp_\sigma f(U_2)$, and f is consistent.

Now assume that f is consistent. For the consistency of $\hat{f}$, let $\langle U_i, b_i\rangle \in \hat{f}$, with $U_1 \asymp_\rho U_2$. By the definition of idealization there exist $U^1_1, \dots U^1_{m_1}, U^2_1, \dots U^2_{m_2} \in \mathrm{Con}_\rho$, such that

$$U_i \vdash_\rho \bigcup_{j_i=1}^{m_i} U^i_{j_i} \wedge \bigcup_{j_i=1}^{m_i} f(U^i_{j_i}) \vdash_\sigma b_i, \tag{$\star$}$$

for each $i = 1, 2$. Since U_1 and U_2 are consistent, the propagation of consistency at type ρ gives us $\bigcup_{j_1=1}^{m_1} U^1_{j_1} \asymp_\rho \bigcup_{j_2=1}^{m_2} U^2_{j_2}$, which in turn yields $\bigcup_{j_1=1}^{m_1} f(U^1_{j_1}) \asymp_\sigma \bigcup_{j_2=1}^{m_2} f(U^2_{j_2})$, due to the consistency of f; by the propagation at type σ we get $b_1 \asymp_\sigma b_2$.

For the deductive closure of $\hat{f}$, let $W \subseteq \hat{f}$ and $\langle U, b\rangle \in \mathrm{Tok}_{\rho\to\sigma}$ be such that $W \vdash_{\rho\to\sigma} \langle U, b\rangle$. By the definition of entailment, there are $\langle U_i, b_i\rangle \in W$, $i = 1, \dots, n$, such that $U \vdash_\rho \bigcup_{i=1}^n U_i$ and $\{b_1, \dots, b_n\} \vdash_\sigma b$. Now each $\langle U_i, b_i\rangle$ is in $\hat{f}$, so there exist neighborhoods $U^i_1, \dots, U^i_{m_i}$, such that ($\star$) holds, but now for $i = 1, \dots, n$. By propagation at ρ, *all* $U^i_{j_i}$'s are consistent; moreover, the consistency of f ensures that *all*

$f(U^i_{j_i})$'s are consistent; then, by the transitivity of entailment, it is

$$U \vdash_\rho \bigcup_{i=1}^{n} \bigcup_{j_i=1}^{m_i} U^i_{j_i} \wedge \bigcup_{i=1}^{n} \bigcup_{j_i=1}^{m_i} f(U^i_{j_i}) \vdash_\sigma b,$$

so $\langle U, b\rangle \in \hat{f}$, by the definition. □

Let us again note that not all ideals can be given by neighborhood mappings by way of Proposition 3.2: a counterexample at type $\mathbb{N} \to \mathbb{N}$ would be the ideal given by $\{\langle 0, \mathtt{S}^n *\rangle \mid n = 0, 1, \ldots\}$. Having stressed that, this result justifies us to say *continuous* for a *neighborhood* mapping, exactly when it is consistent; for reasons of clarity though, we will refrain from using the term.

4 Normal forms

We mentioned in the introduction that the complexity of nonflat base types can become unwieldy very early. To illustrate the point, consider the base type $\mathbb{N}$, where two neighborhoods are equivalent exactly when their highest tokens coincide, for example $\{\mathtt{SS0}, \mathtt{S}*\} \sim_{\mathbb{N}} \{\mathtt{SS0}, \mathtt{SS}*, *\}$. By an elementary combinatorial argument we have that the number of equivalent neighborhoods whose higher token is some known a, is the number of all subsets of the set $\overline{a} \setminus \{a\}$; this means, for example, that merely for the natural number 9 (that is, for the numeral SSSSSSSSS0), there are already a thousand and twenty four equivalent neighborhood representations in the model. It is clear that we could use canonical ways to spot neighborhoods with the desired information, and work exclusively with them—as expected, one of these canonical forms will indeed be the singleton form $\{a\}$.

In our context we look at normal forms not so much as irreducible elements in a rewriting system as in, say, [3], but rather as values of special continuous neighborhood *endo*mappings. Let ρ be a type; a neighborhood-mapping $f : \mathrm{Con}_\rho \to \mathrm{Con}_\rho$ is a *normal form mapping (at type ρ)* if $f(U) \sim_\rho U$ (preservation of information) and $U_1 \sim_\rho U_2$ implies $f(U_1) = f(U_2)$ (uniqueness) for all $U, U_1, U_2 \in \mathrm{Con}_\rho$. By the first requirement, it is clear that every normal form mapping is monotone, so also compatible and consistent, by Lemma 3.1.

We first investigate normal form mappings at arbitrary base types. Then we give a method of inducing a normal form mapping at a higher type, provided we have normal form mappings at the lower types at our disposal.

4.1 Normal forms at base types

Closures and suprema

Perhaps the normal forms which are easiest to recognize at (finitary) base types are given by the deductive closure and the supremum.

Define the *supremum* $\sup(a, b)$ of two consistent base-type tokens inductively over their structure by

$$\begin{aligned} &\sup(a, *) = \sup(*, a) = a, \\ &\sup(Ca_1 \cdots a_r, Cb_1 \cdots b_r) = C \sup(a_1, b_1) \cdots \sup(a_r, b_r), \end{aligned}$$

for every constructor C of arity r. Further, for a base-type neighborhood U define its supremum (or *eigentoken*) $\sup(U) \in \mathrm{Tok}$ by

$$\begin{aligned}&\sup(\varnothing) := *,\\ &\sup(\{a_1,\dots,a_m\}) := \sup(\cdots\sup(a_1,a_2)\cdots,a_m).\end{aligned}$$

Proposition 4.1 (Closure and supremum normal form)**.** *Let ι be a finitary base type. The mappings $U \mapsto \overline{U}$ and $U \mapsto \{\sup(U)\}$ are normal form mappings at type ι.*

Proof. For the deductive closure it is clear. We show that the supremum yields a normal form. We concentrate first on the case of two tokens $a,b \in \mathrm{Tok}_\iota$, where $a \asymp_\iota b$. If one of them is trivial, say $b = *$, then $\{\sup(a,*)\} = \{a\} \sim_\iota \{a,*\}$. If not, then $a = Ca_1\cdots a_r$ and $b = Cb_1\cdots b_r$, with $a_i \asymp_\iota b_i$ for all i's; the induction hypothesis is that $\{\sup(a_i,b_i)\} \sim_\iota \{a_i,b_i\}$ for all i's, so it is

$$\begin{aligned}\{a,b\} &= \{Ca_1\cdots a_r, Cb_1\cdots b_r\}\\ &\sim_\iota C\{a_1,b_1\}\cdots\{a_r,b_r\}\\ &\sim_\iota C\{\sup(a_1,b_1)\}\cdots\{\sup(a_r,b_r)\}\\ &= \{C\sup(a_1,b_1)\cdots\sup(a_r,b_r)\}\\ &= \{\sup(a,b)\}.\end{aligned}$$

Now let $U \in \mathrm{Con}_\iota$. If $U = \varnothing$ then $\sup(U) = *$. If $U = \{a_1,\dots,a_m\}$ then

$$\begin{aligned}U &\sim_\iota \{\sup(a_1,a_2),a_3,\dots,a_m\}\\ &\sim_\iota \cdots\\ &\sim_\iota \{\sup(\cdots\sup(a_2,a_1)\cdots,a_m)\}\\ &= \{\sup(U)\},\end{aligned}$$

based on the previous.

As for uniqueness, if $U_1 \sim_\iota U_2$, then $\sup(U_1) \sim_\iota \sup(U_2)$ by transitivity, and by Lemma 2.4 we get $\sup(U_1) = \sup(U_2)$. ☐

We may write nf^c for $U \mapsto \overline{U}$ and nf^s for $U \mapsto \{\sup(U)\}$. In terms of cardinality, these would be the biggest and the smallest normal forms respectively. In the following we will see some interesting intermediate ones.

Paths

The common idea behind the following two normal forms is to consider the entailment diagram of the closure of a given neighborhood, and then eliminate the circles that appear in it. What we have to show is that in doing so we won't lose any information.

Call $a \in \mathrm{Tok}_\iota$ a *path*, and write $a \in \mathrm{Tok}_\iota^p$, if it is built inductively by the following clauses:

- $* \in \mathrm{Tok}_\iota^p$;
- if C is a constructor and $a \in \mathrm{Tok}_\iota^p$, then $C\vec{*}a\vec{*} \in \mathrm{Tok}_\iota^p$ (where the vectors $\vec{*}$ may be empty).

For example, `B*S0` is a path, whereas `B0S*` isn't. The choice of the name stems from the fact that a path's deductive closure contains no circles (see Proposition 4.2.1 below). By convention, we write $C\overrightarrow{*}a^i\overrightarrow{*}$, if we want to indicate that a^i possesses the i-th position in the arity of C, that is, that $a^i = (C\overrightarrow{*}a^i\overrightarrow{*})(i)$ using the notation of section 2.

Proposition 4.2. *Let ι be a base type.*

1. *Let $a \in \mathrm{Tok}_\iota^p$ and $b_1, b_2 \in \mathrm{Tok}_\iota$. The following* path comparability *property holds: if $a \vdash_\iota b_1$ and $a \vdash_\iota b_2$, then $b_1 \vdash_\iota b_2$ or $b_2 \vdash_\iota b_1$.*
2. *Let $a \in \mathrm{Tok}_\iota^p$ and $b \in \mathrm{Tok}_\iota$. If $a \vdash_\iota b$, then $b \in \mathrm{Tok}_\iota^p$.*
3. *Let $U \in \mathrm{Con}_\iota \backslash \varnothing$, and $b \in \mathrm{Tok}_\iota^p$. The following* path linearity *property holds: if $U \vdash_\iota b$ then there exists an $a \in U$, such that $\{a\} \vdash_\iota b$.*

Proof. For 1. By induction on a. If $a = *$ then $b_m = *$ for both $m = 1, 2$. If $a = C\overrightarrow{*}a^i\overrightarrow{*}$, with $a^i \in \mathrm{Tok}_\iota^p$, then for each $m = 1, 2$ it is $b_m = C\overrightarrow{*}b_m^i\overrightarrow{*}$, with $a^i \vdash_\iota b_m^i$. The induction hypothesis yields $b_1^i \vdash_\iota b_2^i$ or $b_2^i \vdash_\iota b_1^i$, and the definition of entailment does the rest.

For 2. By induction on the path a. If $a = *$ then also $b = *$. Let $a = C\overrightarrow{*}a^i\overrightarrow{*}$, with $a^i \in \mathrm{Tok}_\iota^p$ (for some i within the arity of C). If $b = *$ then it's again trivial, otherwise there exists a $b^i \in \mathrm{Tok}_\iota$ such that $b = C\overrightarrow{*}b^i\overrightarrow{*}$ and $a^i \vdash_\iota b^i$; the induction hypothesis yields $b^i \in \mathrm{Tok}_\iota^p$, so $b \in \mathrm{Tok}_\iota^p$ as well, by definition.

For 3. By induction on b. If $b = *$ then any element of U will do (there is at least one element since U is inhabited). If $b = C\overrightarrow{*}b^i\overrightarrow{*}$, with $b^i \in \mathrm{Tok}_\iota^p$, then $U \backslash \{*\}$ has the form $\{C\overrightarrow{a_1}, \ldots, C\overrightarrow{a_m}\}$, where $\{a_{1i}, \ldots, a_{mi}\} \vdash_\iota b^i$, by the definition of entailment. By the induction hypothesis there exists a $j = 1, \ldots, m$, such that $\{a_{ji}\} \vdash_\iota b^i$; it follows that $\{C\overrightarrow{a_j}\} \vdash_\iota b$. ☐

A nice characterization of paths comes from the minimality of their size.

Proposition 4.3. *A token $a \in \mathrm{Tok}_\iota$ is a path if and only if it has minimal size, that is, if $|a| = \|a\|$.*

Proof. From left to right, let $a \in \mathrm{Tok}_\iota^p$. By induction on the information of a. If $a = *$, then both its height and its size are zero by definition. If $a = C\overrightarrow{*}b\overrightarrow{*}$ for some constructor C and $b \in \mathrm{Tok}_\iota^p$, then the induction hypothesis gives that $|b| = \|b\| = m$ for some $m \geqslant 0$; by the definition of height and size we obtain $|a| = 1 + m = \|a\|$.

For the other way around, let a be such that $|a| = \|a\| = m$ for some $m \geqslant 0$. We perform induction on m. For $m = 0$, it is $a = *$, which is a path by definition. For $m + 1 \geqslant 0$, we have $a = Ca_1 \cdots a_r$ with

$$m = \max\{|a_1|, \ldots, |a_r|\} = \|a_1\| + \cdots + \|a_r\|;$$

by Proposition 2.5.3, it is $|a_i| = \|a_i\| =: m_i$ for all $i = 1, \ldots, r$, and either $m = m_i = 0$, so $a = C\overrightarrow{*}$, or else there is exactly one i such that $m_i > 0$, so $a = C\overrightarrow{*}a_i\overrightarrow{*}$, with $a_i \in \mathrm{Tok}_\iota^p$ by the induction hypothesis; in both cases it is $a \in \mathrm{Tok}_\iota^p$ by definition. ☐

It turns out that paths facilitate a natural notion of a "maximal normal form" for neighborhoods. If $U \in \mathrm{Con}_\rho$, its *maximal elements* $\mathrm{mxl}(U)$ are those tokens $a \in U$, such that if $a' \in U$ is some other token with $a' \vdash_\rho a$, then $a' \sim_\rho a$. Say that a neighborhood U is *path reduced*, and write $U \in \mathrm{Con}_\iota^p$, if every token in it is a path and is maximal in U. For example, $\{$`B00`$\}$ and $\{$`B0*,B*0,B**`$\}$ are not path reduced, but $\{$`B0*,B*0`$\}$ is.

Proposition 4.4. *If $U \in \mathrm{Con}_\iota^p$, then the following* maximality property *holds:*

$$U_0 \vdash_\iota a \to a \in U_0,$$

for all inhabited $U_0 \subseteq U$ and $a \in U$.

Proof. Let $U \in \mathrm{Con}_\iota^p$, $U_0 \subseteq U$ an inhabited subneighborhood, and $a \in U$. Since U consists of paths, Proposition 4.2.3 yields a single $a' \in U_0$ such that $a' \vdash_\iota a$. But U is reduced, so $a' = a$. □

A *path form* of a neighborhood is an equivalent neighborhood which is path reduced; for example, the finite set $\{\mathtt{B0*}, \mathtt{B*0}\}$ is a path form of both $\{\mathtt{B00}\}$ and $\{\mathtt{B0*}, \mathtt{B*0}, \mathtt{B**}\}$.

Theorem 4.5 (Path normal form)**.** *Every neighborhood at a finitary base type has a unique path form.*

Proof. We consider first path forms for tokens $a \in \mathrm{Tok}_\iota$ (and thus cover the singleton finite sets). Let $\mathrm{nf}^p : \mathrm{Tok}_\iota \to \mathrm{Fin}_\iota$ be the mapping defined recursively by the clauses

$$\mathrm{nf}^p(C) = \{C\} \text{ for } C \text{ nullary},$$
$$\mathrm{nf}^p(Ca_1 \cdots a_r) = \bigcup_{i=1}^{r} C\overrightarrow{\{*\}}\mathrm{nf}_*^p(a_i)\overrightarrow{\{*\}},$$

where $\mathrm{nf}_*^p(a) := \mathrm{nf}^p(a) \backslash \{*\}$ (it is always $\mathrm{nf}_*^p(a) = \mathrm{nf}^p(a)$, when $a \neq *$). Note the use of our notational convention of page 5, and that in the first clause the pseudoconstructor $*$ is counted in.

It is direct to see that if $b \in \mathrm{nf}^p(a)$ then $b \in \mathrm{Tok}_\iota^p$. Such a b must also be maximal: if $a = *$, then also $b = *$; otherwise $a = Ca_1 \cdots a_r$ and $b = C\vec{*}a^i\vec{*}$ for some $i = 1, \dots, r$ and $a^i \in \mathrm{nf}^p(a_i)$ with $a^i \neq *$; assuming that there is a $b' \in \mathrm{nf}^p(a)$, such that $b' \vdash_\iota b$, it follows that $b' = Cb_1 \cdots b_i \cdots b_r$, with $b_i \vdash_\iota a^i$; by the construction of nf^p and the induction hypothesis for $\mathrm{nf}^p(a_i)$, it must be $b_j = *$ for all $j \neq i$, and $b_i = a^i$, so $b' = b$. We've shown then that $\mathrm{nf}^p(a)$ is path reduced for every $a \in \mathrm{Tok}_\iota$. The preservation of information follows from the induction hypotheses $\mathrm{nf}^p(a_i) \sim_\iota \{a_i\}$ for each $i = 1, \dots, r$, and the definition of entailment. As for uniqueness, it follows immediately from Lemma 2.4, since $a \sim_\iota b$ implies $a = b$, so $\mathrm{nf}^p(a) = \mathrm{nf}^p(b)$.

Moving on to neighborhoods $U \in \mathrm{Con}_\iota$, we may set $\mathrm{nf}^p(U) := \mathrm{nf}^p(\sup(U))$; this is a normal form mapping by the previous and Proposition 4.1. □

Remark. In previous approaches to maximal normal forms, as in [21] and [11], the restriction to nonsuperunary constructors and binary entailment made it possible to avoid paths and obtain a normal form directly from the mapping $U \mapsto \mathrm{mxl}(U)$, which was both "linear" (see section 5) and "maximal" in the sense of Proposition 4.4. This doesn't work in the general case of an algebra with superunary constructors and a full entailment predicate. Take for example the neighborhood $U = \{\mathtt{BB00*}, \mathtt{BB0*0}, \mathtt{BB*00}\}$, for which it already holds that $U = \mathrm{mxl}(U)$; this is neither linear (we've agreed that it entails $\mathtt{BB000}$, but it cannot do so with one token) nor even maximal (for the subset $U_0 := \{\mathtt{BB00*}, \mathtt{BB0*0}\}$ it is $U_0 \vdash_{\mathbb{D}} \mathtt{BB*00}$ but $\mathtt{BB*00} \notin U_0$). Moreover, even if we restricted entailment to its binary version and we tried to find *subsets* of U which do satisfy Proposition 4.4 and are themselves linear and maximal, we would actually find three: every pair of tokens in U forms such a neighborhood; but there would be no natural reason to prefer one over the other as a normal form in order to have uniqueness.[3] □

Trees

From the path form of a neighborhood we can easily obtain its "tree form" by taking atomic closures. Call a neighborhood $U \in \mathrm{Con}_\iota$ a *(full) tree*, and write $U \in \mathrm{Con}_\iota^t$, if for every $a \in U$ it is $a \in \mathrm{Tok}_\iota^p$ and $\overline{a} \subseteq U$ (the name is justified by Proposition 4.2.1). For example, the neighborhood $\{\mathtt{B0*}, \mathtt{B*0}, \mathtt{B00}\}$ is no tree, but $\{\mathtt{B0*}, \mathtt{B*0}, \mathtt{B**}, *\}$ is, consisting of the closures of the paths $\mathtt{B0*}$ and $\mathtt{B*0}$. So *the root* of the trees we consider here is always $*$ (that's why we think of them as "full"), while their *leaves* are simply their tokens of maximal information. A *tree form* of a neighborhood is an equivalent neighborhood which is a tree; for example, a tree-form of $\{\mathtt{B0*}, \mathtt{B*0}, \mathtt{B00}\}$ is $\{\mathtt{B0*}, \mathtt{B*0}, \mathtt{B**}, *\}$—actually, the only one.

Proposition 4.6 (Tree normal form)**.** *Every neighborhood at a finitary base type has a unique tree form.*

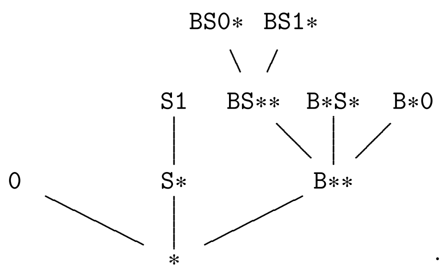
Proof. Let $U \in \mathrm{Con}_\iota$. By Theorem 4.5 we can assume that U is path reduced. Set $\mathrm{nf}^t(U) := \bigcup_{a \in U} \overline{a}$. It is clear by the construction that this is a tree. The preservation of information and the uniqueness are both straightforward. □

Remark. The tree form of a neighborhood can of course be generated without appeal to its path form. For tokens we can first set

$$\mathrm{nf}^t(Ca_1 \cdots a_r) := \{*\} \cup \bigcup_{i=1}^{r} \{C \overrightarrow{*} a_{0i} \overrightarrow{*} \mid a_{0i} \in \mathrm{nf}^t(a_i)\},$$

and then $\mathrm{nf}^t(U) := \bigcup_{a \in U} \mathrm{nf}^t(a)$. □

Example. At type $\mathbb{D}$, the singleton $\{\mathtt{0}\}$ has tree form $\{*, \mathtt{0}\}$, and the singleton $\{\mathtt{S1}\}$ has tree form $\{*, \mathtt{S*}, \mathtt{S1}\}$. The singleton $\{\mathtt{BS0S*}\}$, which involves a binary constructor, has tree form $\{*, \mathtt{B**}, \mathtt{BS**}, \mathtt{BS0*}, \mathtt{B*S*}\}$, and similarly the singleton $\mathtt{BS10}$ has tree form $\{*, \mathtt{B**}, \mathtt{BS**}, \mathtt{BS1*}, \mathtt{B*0}\}$. The union of these tree forms yields the following picture.



.

□

[3] See also the discussion of "eigen-maximality" below, where it is shown how a natural notion of maximality is achieved at higher types—still failing to be a normal form.

4.2 Normal forms at higher types

Once we've found normal forms for base types, we would like to systematically induce normal forms from lower to higher types. Our approach will make crucial use of "eigen-neighborhoods".

If $\Theta \in \mathrm{Fin}_{\rho\to\sigma}$ is a finite set with $\Theta = \left\{\langle U_j, b_j\rangle \mid j = 1,\dots,l\right\}$, write $L(\Theta)$ for $\bigcup_j U_j$ (notice that this is a *flattening*), and $R(\Theta)$ for $\bigcup_j \left\{b_j\right\}$. Furthermore, if $U \in \mathrm{Con}_\rho$ and $\Delta \in \mathrm{Fin}_\sigma$, write $\langle U, \Delta\rangle$ for $\left\{\langle U, b\rangle \mid b \in \Delta\right\}$ (note that $\langle U, \varnothing\rangle = \varnothing$).

Eigen-neighborhoods

Consider a neighborhood $W = \left\{\langle U_j, b_j\rangle \mid j = 1,\dots l\right\}$ at some higher type $\rho \to \sigma$, which we apply to some information, say U, of type ρ. If this U is above U_j and U_k for some $j,k = 1,\dots l$, then (a) U_j and U_k must be consistent, and (b) both b_j and b_k will belong to the value WU. Furthermore, if U is above U_j, which in turn is above some other U_k, then (a) U must be above U_k as well, and (b) both b_j and b_k will again belong to the value WU. These two basic facts regarding application motivate the definition of the *eigen-neighborhoods* of W, Eig_W.

At a base type ι, by convention, we say that E is an eigen-neighborhood of $U \in \mathrm{Con}_\iota$ if $E = \varnothing$ or $E = U$. At a higher type $\rho \to \sigma$, a neighborhood H is an eigen-neighborhood of $W \in \mathrm{Con}_{\rho\to\sigma}$ if it is of the form $H = \langle U, V\rangle$ and it features the following properties:

- *consistency*: $U \in \mathrm{Con}_{L(W)}$,
- *left closure*: $U = \overline{U} \cap L(W)$,
- *right closure*: $V = \overline{WU} \cap R(W)$;

by the first requirement, it is clear that Eig_W is a finite set of neighborhoods for every W, and by the first and third, that every eigen-neighborhood is indeed consistent. Notice that the concept of eigen-neighborhoods is *not* given inductively over types; in fact, in the following we concentrate on eigen-neighborhoods at higher types, as it's there where they prove essential.

Given $W \in \mathrm{Con}_{\rho\to\sigma}$ as above, every $U \in \mathrm{Con}_\rho$ induces a characteristic eigen-neighborhood $W\|\!\!\upharpoonright_U$, the *eigen-restriction of W to U*, in the following natural way:

$$W\|\!\!\upharpoonright_U := \left\langle \overline{U} \cap L(W), \overline{WU} \cap R(W)\right\rangle;$$

the eigen-neighborhood $W\|\!\!\upharpoonright_U$ is basically the "support" of W with respect to U, that is, the part of W that answers to the input U.[4] One can directly check the following.

Lemma 4.7. *Let ρ, σ be types.*

1. *For all $W \in \mathrm{Con}_{\rho\to\sigma}$ and $U \in \mathrm{Con}_\rho$, it is $WU \vdash_\sigma b$ if and only if $R(W\|\!\!\upharpoonright_U) \vdash_\sigma b$.*
2. *For all $W \in \mathrm{Con}_{\rho\to\sigma}$, if $H \in \mathrm{Eig}_W$ then $H = W\|\!\!\upharpoonright_{L(H)}$.*
3. *For all $W \in \mathrm{Con}_{\rho\to\sigma}$ and $U, U' \in \mathrm{Con}_\rho$, if $U \vdash_\rho U'$ then $P(W\|\!\!\upharpoonright_{U'}) \subseteq P(W\|\!\!\upharpoonright_U)$, where P stands for L and R.*

[4]Note that this is quite different than the more modest *restriction of W to U*, which would be just the subneighborhood $W\!\upharpoonright_U := \left\{\langle U', b\rangle \in W \mid U \vdash_\rho U'\right\}$.

4. *The only nonempty eigen-neighborhood of a neighborhood of the form $\langle U,V\rangle$ is itself, up to equientailment.*

The following establishes our intuition of eigen-neighborhoods as “generalized tokens”. Write eig for the mapping $U \mapsto \bigcup \mathrm{Eig}_U$; it is trivially $U = \mathrm{eig}(U)$ for all base-type U's, whereas at higher types it is

$$\mathrm{eig}(W) := \bigcup_{U \in \mathrm{Con}_{L(W)}} \left\langle \overline{U} \cap L(W), \overline{WU} \cap R(W) \right\rangle.$$

Proposition 4.8 (Consistent eigenform)**.** *Let ρ and σ be types, and $W, W_1, W_2 \in \mathrm{Con}_{\rho\to\sigma}$.*

1. *It is $W \sim_{\rho\to\sigma} \mathrm{eig}(W)$. Moreover, $\mathrm{Eig}_{\mathrm{eig}(W)} = \mathrm{Eig}_W$, therefore the mapping* eig *is idempotent, that is,* $\mathrm{eig}(\mathrm{eig}(W)) = \mathrm{eig}(W)$.

2. *It is $W_1 \vdash_{\rho\to\sigma} W_2$ if and only if for every $H_2 \in \mathrm{Eig}_{W_2}$ there exists an $H_1 \in \mathrm{Eig}_{W_1}$ such that $H_1 \vdash_{\rho\to\sigma} H_2$. Similarly, it is $W_1 \asymp_{\rho\to\sigma} W_2$ if and only if for all $H_1 \in \mathrm{Eig}_{W_1}$, $H_2 \in \mathrm{Eig}_{W_2}$, it is $H_1 \asymp_{\rho\to\sigma} H_2$.*

Proof. For 1: Let W be a neighborhood at type $\rho \to \sigma$. From left to right, let $\langle U,V\rangle$ be one of its eigen-neighborhoods. By the definition we have $V = \overline{WU} \cap R(W)$, from which we get that $WU \vdash_\sigma V$, that is, that $W \vdash_{\rho\to\sigma} \langle U,V\rangle$. For the other way around, let $\langle U,b\rangle \in W$. For the induced eigen-neighborhood $W\|\!\upharpoonright_U$ we have $W\|\!\upharpoonright_U U = \overline{WU} \cap R(W)$, and $b \in WU \cap R(W)$, so $W\|\!\upharpoonright_U \vdash_{\rho\to\sigma} \langle U,b\rangle$.

Before showing the idempotence, we first notice that (i) $L(W) = L(\mathrm{eig}(W))$ and $R(W) = R(\mathrm{eig}(W))$. Indeed, for the left part, if $a \in L(W)$, then there is a $\langle U,b\rangle \in W$, with $a \in U$; then $a \in L(W\|\!\upharpoonright_U)$, where $W\|\!\upharpoonright_U \in \mathrm{Eig}_W$, so $a \in L(\mathrm{eig}(W))$; conversely, if $a \in L(\mathrm{eig}(W))$, then there is an $H \in \mathrm{Eig}_W$, such that $a \in L(H)$, which means that $a \in L(W)$ immediately. For the right part, if $b \in R(W)$, then there is a $\langle U,b\rangle \in W$; then $b \in R(W\|\!\upharpoonright_U)$, where $W\|\!\upharpoonright_U \in \mathrm{Eig}_W$, so $b \in R(\mathrm{eig}(W))$; conversely, if $b \in R(\mathrm{eig}(W))$, then there's an $H \in \mathrm{Eig}_W$, such that $b \in R(H)$, which means that $b \in R(W)$ by the definition of eigen-neighborhoods. Then we also notice the easy fact that (ii) at a type ρ, if $U \sim_\rho U'$ then $\overline{U} = \overline{U'}$ for all $U \in \mathrm{Con}_\rho$.

For idempotence, it suffices to show that both $\mathrm{eig}(W)$ and W share the same eigen-neighborhoods. Let $H = \langle U,V\rangle$; by the definition of eigen-neighborhoods, it is $H \in \mathrm{Eig}_{\mathrm{eig}(W)}$ if and only if

$$U \in \mathrm{Con}_{L(\mathrm{eig}(W))} \;\wedge\; U = \overline{U} \cap L(\mathrm{eig}(W)) \;\wedge\; V = \overline{\mathrm{eig}(W)U} \cap R(\mathrm{eig}(W)),$$

which, by (i) and (ii) above (as well as the fact that $W \sim_{\rho\to\sigma} W'$ implies $WU \sim_\sigma W'U$ for all W, W', U), holds if and only if

$$U \in \mathrm{Con}_{L(W)} \;\wedge\; U = \overline{U} \cap L(W) \;\wedge\; V = \overline{WU} \cap R(W);$$

this conjunction yields by definition $H \in \mathrm{Eig}_W$.

For 2: Let $W_1, W_2 \in \mathrm{Con}_{\rho\to\sigma}$. Concerning entailment, assume first that $W_1 \vdash_{\rho\to\sigma} W_2$, and let $H_2 \in \mathrm{Eig}_{W_2}$; by 1, it is clear that $W_1 \vdash_{\rho\to\sigma} H_2$, and setting $H_1 := W_1\|\!\upharpoonright_{L(H_2)}$ we do get that $H_1 \vdash_{\rho\to\sigma} H_2$. For the other way around, assuming that every eigen-neighborhood of W_2 is entailed by some neighborhood of W_1, it follows that a part of $\bigcup \mathrm{Eig}_{W_1}$ suffices to entail all of $\bigcup \mathrm{Eig}_{W_2}$, so 1 yields what we want.

Concerning consistency, it is $W_1 \asymp_{\rho\to\sigma} W_2$ if and only if, because of 1, $\bigcup \mathrm{Eig}_{W_1} \asymp_{\rho\to\sigma} \bigcup \mathrm{Eig}_{W_2}$, which holds exactly when $H_1 \asymp_{\rho\to\sigma} H_2$ for all $H_1 \in \mathrm{Eig}_{W_1}$ and $H_2 \in \mathrm{Eig}_{W_2}$. ☐

We say that $\operatorname{eig}(U)$ is the *eigenform* of U, and if $U = \operatorname{eig}(U)$, we say that U *is in eigenform*; every base-type neighborhood is in eigenform. Note that although eig deserves the definite article (due to its idempotence), it still *does not* yield a normal form mapping: two equivalent eigenforms are not necessarily equal, as one sees already at (nontrivial) base types. Its utility is rather that it rearranges and tidies up the information *as it is in the given neighborhood*, and brings it to a more manageable form. For example, we can use the eigenform of a neighborhood W to easily obtain *conservative extensions*: if $H \in \mathrm{Eig}_W$, then $W \sim_\rho W \cup W^H$ for every $W^H \in \mathrm{Con}_\rho$ with $H \vdash_\rho W^H$, a simple technique that often helps at higher types.

Furthermore, Proposition 4.8.2 suggests that eigen-neighborhoods bring into light a *linear* behavior of higher-type entailment; it is indeed a statement which we can think of as *implicit linearity*. This is a phenomenon that we will revisit in section 5.1.

Eigen-maximality

Based on the intuition of eigen-neighborhoods as generalized tokens, we may establish a form of a neighborhood—but still no *normal* form!—where no eigen-neighborhood is informationally redundant.

First we observe that entailment between eigen-neighborhoods reduces (contravariantly) to componentwise *inclusion*.

Lemma 4.9. *Let ρ, σ be types, $W \in \mathrm{Con}_{\rho\to\sigma}$, and $H_1, H_2 \in \mathrm{Eig}_W$ be such that $H_1 \vdash_{\rho\to\sigma} H_2$. Then $L(H_1) \subseteq L(H_2)$ and $R(H_2) \subseteq R(H_1)$, thus also $R(H_2) = R(H_1)$.*

Proof. Since $H_1 \vdash_{\rho\to\sigma} H_2$, it is $L(H_2) \vdash_\rho L(H_1)$ and $R(H_1) \vdash_\sigma R(H_2)$. The corresponding inclusions follow from the left and right closure properties that define eigen-neighborhoods (see also Lemma 4.7.3). The last claim follows from the monotonicity of application: since $L(H_2) \vdash_\rho L(H_1)$, it is $WL(H_2) \vdash_\sigma WL(H_1)$; then $\overline{WL(H_1)} \cap R(W) \subseteq \overline{WL(H_2)} \cap R(W)$, so $R(H_1) \subseteq R(H_2)$, and we're done. □

Write Eig_W^0 for the nonempty eigen-neighborhoods of W. Call $W \in \mathrm{Con}_{\rho\to\sigma}$ *eigen-maximal* if it is in eigenform and each $H \in \mathrm{Eig}_W$ is either empty or maximal (that is, if $H \in \mathrm{Eig}_W^0$, then for all $H' \in \mathrm{Eig}_W$ with $H' \vdash_{\rho\to\sigma} H$, it is $H' \sim_{\rho\to\sigma} H$); write $\mathrm{Eig}_W^{\max}$ for the set of maximal eigen-neighborhoods of W. An eigen-maximal neighborhood is "flat", in the sense that the inclusion diagram of its eigen-neighborhoods forms a flat tree. Note that by Lemma 4.9 it follows that in an eigen-maximal neighborhood, if $H' \vdash_{\rho\to\sigma} H$ and H is nonempty, then $H' = H$.

Lemma 4.10. *Let ρ, σ be types. For every $W \in \mathrm{Con}_{\rho\to\sigma}$ there exists an eigen-maximal $W' \in \mathrm{Con}_{\rho\to\sigma}$, such that $W \sim_{\rho\to\sigma} W'$.*

Proof. Let $W \in \mathrm{Con}_{\rho\to\sigma}$. It is $W \sim_{\rho\to\sigma} \bigcup \mathrm{Eig}_W$, by Proposition 4.8.1. Then $W' := \bigcup \mathrm{Eig}_W^{\max}$ is an equivalent eigen-maximal neighborhood. □

Obviously, there are in general several witnesses for the above fact, but since we will need to refer to at least one of them later, it helps to single out the one that we actually used in the proof. If $U \in \mathrm{Con}_\rho$, write emxl for the mapping $\mathrm{Con}_\rho \to \mathrm{Con}_\rho$ given by $\operatorname{emxl}(U) := \bigcup \mathrm{Eig}_U^{\max}$; by Lemma 4.10, this is clearly a monotone endomapping.

The following provides the crucial stepping stone towards our goal.

Proposition 4.11 (Eigencorrespondence)**.** *Let $W_1, W_2 \in \mathrm{Con}_{\rho\to\sigma}$, and W_1 be eigen-maximal. It is $W_1 \sim_{\rho\to\sigma} W_2$ if and only if for each $H_1 \in \mathrm{Eig}_{W_1}$ there is exactly one $H_2 \in \mathrm{Eig}_{W_2}$, up to equientailment, such that $H_2 \sim_{\rho\to\sigma} H_1$.*

Proof. Assume that $W_1 \sim_{\rho\to\sigma} W_2$, and let $H_1 \in \mathrm{Eig}_{W_1}$. By Proposition 4.8.2 there is an $H_2 \in \mathrm{Eig}_{W_2}$, such that $H_2 \vdash_{\rho\to\sigma} H_1$ for which, in turn, there is an $H_1' \in \mathrm{Eig}_{W_1}$, such that $H_1' \vdash_{\rho\to\sigma} H_2$. It follows that $H_1' \vdash_{\rho\to\sigma} H_1$, but W_1 is eigen-maximal, so $H_1' = H_1$, and consequently $H_2 \sim_{\rho\to\sigma} H_1$. The uniqueness of H_2 up to equientailment is clear. As for the converse, it follows immediately from Proposition 4.8.2. □

It follows that if both W_1 and W_2 are eigen-maximal, then their eigen-neighborhoods are in a one to one correspondence. Another nice property that follows from eigencorrespondence is the next one.

Corollary 4.12. *Let $W_1, W_2 \in \mathrm{Con}_{\rho\to\sigma}$ be eigen-maximal. If $W_1 \sim_{\rho\to\sigma} W_2$ then $W_1\|\!\upharpoonright_U \sim_{\rho\to\sigma} W_2\|\!\upharpoonright_U$ for every $U \in \mathrm{Con}_\rho$.*

Proof. Let $U \in \mathrm{Con}_\rho$. By Proposition 4.11, there is exactly one $H \in \mathrm{Eig}_{W_1}$, such that $H \sim_{\rho\to\sigma} W_2\|\!\upharpoonright_U$, that is, such that $L(H) \sim_\rho L(W_2\|\!\upharpoonright_U)$ and $R(H) \sim_\sigma R(W_2\|\!\upharpoonright_U)$. Now, on the one hand, since $U \vdash_\rho L(W_2\|\!\upharpoonright_U)$, it is $U \vdash_\rho L(H)$ by transitivity, so Lemma 4.7 (items 2 and 3) yields $L(H) \subseteq L(W_1\|\!\upharpoonright_U)$; on the other hand, since $W_1 \sim_{\rho\to\sigma} W_2$, it is $W_1 U \sim_\sigma W_2 U$, so $R(W_1\|\!\upharpoonright_U) \sim_\sigma R(W_2\|\!\upharpoonright_U)$, by Lemma 4.7.1. It follows that $H \vdash_{\rho\to\sigma} W_1\|\!\upharpoonright_U$, but W_1 is eigen-maximal, so $H = W_1\|\!\upharpoonright_U$. This means that $W_1\|\!\upharpoonright_U \sim_{\rho\to\sigma} W_2\|\!\upharpoonright_U$, so we're done. □

Eigenproducts of endomappings

Given two mappings $f : \mathrm{Con}_\rho \to \mathrm{Con}_\rho$ and $g : \mathrm{Con}_\sigma \to \mathrm{Con}_\sigma$, define their *eigenproduct* $\langle f, g\rangle : \mathrm{Con}_{\rho\to\sigma} \to \mathrm{Fin}_{\rho\to\sigma}$ by

$$\langle f, g\rangle(W) := \bigcup_{H \in \mathrm{Eig}^0_W} \langle f(L(H)), g(R(H))\rangle .$$

There is a simple criterion for the consistency of an eigenproduct. A neighborhood-mapping $f : \mathrm{Con}_\rho \to \mathrm{Con}_\rho$ is called *inflationary* (or *expansive*) when $f(U) \vdash_\rho U$ and *deflationary* (or *contractive*) when $U \vdash_\rho f(U)$ for $U \in \mathrm{Con}_\rho$. It is direct to see that a deflationary mapping is also consistent. Furthermore, we have the following.

Lemma 4.13. *Let f and g be neighborhood-mappings at types ρ and σ respectively. If f is inflationary and g is deflationary, then $\langle f, g\rangle$ is a deflationary neighborhood-mapping at type $\rho\to\sigma$ (and, a fortiori, consistently defined).*

Proof. Let $W \in \mathrm{Con}_{\rho\to\sigma}$. For every $H \in \mathrm{Eig}_W$ it is $f(L(H)) \vdash_\rho L(H)$ and $R(H) \vdash_\sigma g(R(H))$, so $H \vdash_{\rho\to\sigma} \langle f, g\rangle(H)$. It follows from Proposition 4.8.1 that $W \vdash_{\rho\to\sigma} \langle f, g\rangle(W)$, so we're done. □

Since a normal form mapping is trivially inflationary and deflationary, it follows that the eigenproduct of two normal form mappings is deflationary, so also consistently defined. The obvious question is when it is also inflationary, and so itself a normal form mapping.

Proposition 4.14. *Let f and g be normal form mappings at types ρ and σ respectively. Then their eigenproduct is a normal form mapping at type $\rho\to\sigma$, when restricted to eigen-maximal neighborhoods.*

Proof. Let $W, W_1, W_2 \in \mathrm{Con}_{\rho\to\sigma}$ be eigen-maximal. We want to show (i) that $\langle f, g\rangle(W) \sim_{\rho\to\sigma} W$, and (ii) that $W_1 \sim_{\rho\to\sigma} W_2$ implies $\langle f, g\rangle(W_1) = \langle f, g\rangle(W_2)$.

For (i), it is

$$\begin{aligned}\langle f,g\rangle(W) &= \bigcup_{H\in \mathrm{Eig}^0_W} \langle f(L(H)), g(R(H))\rangle \\ &\overset{(\star)}{\sim}_{\rho\to\sigma} \bigcup_{H\in \mathrm{Eig}^0_W} \langle L(H), R(H)\rangle \\ &\overset{(\star\star)}{\sim}_{\rho\to\sigma} W,\end{aligned}$$

where $(\star)$ holds by the induction hypotheses at types ρ and σ, and $(\star\star)$ by Proposition 4.8.1.

For (ii), it is

$$\begin{aligned}\langle f,g\rangle(W_1) &= \bigcup_{H_1\in \mathrm{Eig}^0_{W_1}} \langle f(L(H_1)), g(R(H_1))\rangle \\ &\overset{(\star)}{\sim}_{\rho\to\sigma} \bigcup_{H_2\in \mathrm{Eig}^0_{W_2}} \langle f(L(H_2)), g(R(H_2))\rangle \\ &= \langle f,g\rangle(W_2),\end{aligned}$$

where $(\star)$ holds by Proposition 4.11. □

From Proposition 4.14 and Lemma 4.10 we immediately get the following general result.

Theorem 4.15 (Inductive normal forms)**.** *Let f and g be normal form mappings at types ρ and σ respectively. Then the mapping $\langle f, g\rangle \circ \mathrm{emxl}$ is a normal form mapping at type $\rho \to \sigma$.*

Note that the choice of emxl can be replaced by any choice of an endomapping that sends a neighborhood to an equivalent eigen-maximal neighborhood.

5 Linearity

In a similar way that we can reduce the consistency to a binary predicate by coherence (1), we can reduce entailment to a binary predicate by an appropriate property. Call an information system, as well as its corresponding type, *linear* (or *atomic*) if

$$U \vdash b \leftrightarrow \underset{a\in U}{\exists} \{a\} \vdash b; \tag{3}$$

in this case, write $U \vdash^A b$ (we implicitly mean that U is inhabited). In general, call a neighborhood *linear* if it satisfies (3) for all b's; for example every singleton forms a linear neighborhood. From a computational point of view, the property assures us that in order to decide $U \vdash b$, we don't have to check $U_0 \vdash b$ for *all* $U_0 \subseteq U$; it suffices to just check it for the singleton ones.

There are obvious technical reasons to want to work with linear systems, but there are also good conceptual reasons. Since Gordon Plotkin elaborated in [17] on the role of inherently nonsequential functionals like the "parallel or" in Scott's model, a lot of work focused on finding restricted models where this problem would not arise—this is the well known quest for a "fully abstract" model for Plotkin's PCF (see [5] for a succinct overview). Gerard Berry noticed very early that if a functional is to be called "sequential", it should at least be *stable* [2], that is, apart from being Scott continuous it should preserve consistent infima. The appropriate domains for stability are the so called "dI-domains", and as Guo-Qiang Zhang showed in [26] (see also [27, 28]), in order to represent stable domains by information systems, we have to require linearity. Furthermore, since Jean-Yves Girard based his linear logic on Berry's stability notion among other things (see [7]), linearity is a property where we arrive quite often when pursuing models of linear logic (see for example [4]).

It is easy to see that *flat* information systems induced by algebras are linear. In our *nonflat* setting, the base type $\mathbb{N}$ of natural numbers is linear: for example, the entailment $\{\mathtt{S}*,\mathtt{SS0}\} \vdash_{\mathbb{N}} \mathtt{SS}*$ reduces to $\{\mathtt{SS0}\} \vdash_{\mathbb{N}} \mathtt{SS}*$. Similarly for the type $\mathbb{B}$ of booleans. Moreover, like with coherence, linearity is a property that function spaces preserve (see Proposition 5.3), so one could naturally argue that the choice of *linear* nonflat information systems should be perfectly legitimate. Indeed, one can establish several fundamental results in the nonflat setting based on linear systems alone: Helmut Schwichtenberg dealt with density, preservation of values, and adequacy in [22], and we have also shown definability for the type system based on $\mathbb{B}$ and $\mathbb{N}$ in [11, 9]; also, in the same spirit, but independently, Fritz Müller has obtained similar results [16].

Notwithstanding its importance and facility, linearity seems to be natural, or at least explicit, only for those systems that are built by constructors with at most unary arity, like $\mathbb{B}$ and $\mathbb{N}$, whereas in $\mathbb{D}$ we have the paradox

$$\{\mathtt{B0}*,\mathtt{B}*\mathtt{0}\} \vdash \mathtt{B00} \;\wedge\; \{\mathtt{B0}*,\mathtt{B}*\mathtt{0}\} \nvdash^{A} \mathtt{B00}. \tag{4}$$

Indeed, linearity is one of the several things that get tricky when one decides to go nonflat—this echoes the situation with sequentiality, see for example Glynn Winskel's discussion in [25, pp. 340–341]. It would seem then that linear systems cannot be used for a general theory of higher-type computability (unless of course we restrict ourselves to base types with non-superunary constructors), and that one would be forced instead to deal with the intricate, not necessarily linear entailment, in its full generality. We see in this section that this is actually not the case. On the one hand, nonlinear systems turn out to be "implicitly linear" (as was already foreshadowed in Proposition 4.8.2): every neighborhood is equivalent to a linear one. More importantly, we can make this hidden linearity explicit by restricting our models to linear subsystems, without losing in expressivity at all.

5.1 Implicit linearity

Call a type *implicitly linear*, when every neighborhood has an equivalent one which is linear. It is not hard to see that all of our (finitary) base types are implicitly linear, since there are normal forms for every neighborhood which are linear, like nf^c or nf^s (see section 4.1). We show that we can elevate this fact to higher types.

We need one more easy observation concerning eigenproducts.

Lemma 5.1. *Let $W \in \mathrm{Con}_{\rho\to\sigma}$ be any neighborhood and $f : \mathrm{Con}_\rho \to \mathrm{Con}_\rho$, $g : \mathrm{Con}_\sigma \to \mathrm{Con}_\sigma$ be information preserving. Then $W \sim_{\rho\to\sigma} \langle f,g\rangle(W)$.*

Proof. From left to right, let $\langle U,b\rangle \in \langle f,g\rangle(W)$. By the definition, there is a (nonempty) eigen-neighborhood H of W, such that $U = L(H)$ and $b \in R(H)$. By the preservation of information of f and g, it is $W \vdash_{\rho\to\sigma} \langle f(L(H)), g(R(H))\rangle$, so it follows that $W \vdash_{\rho\to\sigma} \langle U, b\rangle$ as well. For the other way around, let H be an eigen-neighborhood of W. Then there is a subneighborhood of $\langle f, g\rangle(W)$, namely $\langle f(L(H)), g(R(H))\rangle$, which entails H, so, by Proposition 4.8.1, we're done. $\square$

Theorem 5.2 (Implicit linearity)**.** *Let ρ be an arbitrary type. There exists a neighborhood mapping* $\mathrm{at}_\rho : \mathrm{Con}_\rho \to \mathrm{Con}_\rho$*, such that* $\mathrm{at}_\rho(U)$ *is linear and equivalent to U for all $U \in \mathrm{Con}_\rho$.*

Proof by induction over types. At a base type ι for a neighborhood U set $\mathrm{at}_\iota(U) := \mathrm{nf}(U)$, where nf is a normal form mapping that sends to linear equivalent neighborhoods. At a higher type $\rho \to \sigma$, let $W \in \mathrm{Con}_{\rho\to\sigma}$; by Lemma 4.10, we may safely assume that it is eigen-maximal. Set

$$\mathrm{at}_{\rho\to\sigma}(W) := \langle \mathrm{id}, \mathrm{at}_\sigma\rangle(W),$$

where $\mathrm{id} : \mathrm{Con}_\rho \to \mathrm{Con}_\rho$ is the identity neighborhood-mapping, and at_σ is provided by the induction hypothesis at σ.

This is obviously a finite set, which is equivalent to W, by the induction hypothesis at σ and Lemma 5.1. It remains to show the implicit linearity property. Let $\langle U, b\rangle \in \mathrm{Tok}_{\rho\to\sigma}$ be such that $\mathrm{at}_{\rho\to\sigma}(W) \vdash_{\rho\to\sigma} \langle U, b\rangle$. Then also $W \vdash_{\rho\to\sigma} \langle U, b\rangle$, which means that $W\|\!\!\upharpoonright_U \vdash_{\rho\to\sigma} \langle U, b\rangle$, or, equivalently, $U \vdash_\rho L(W\|\!\!\upharpoonright_U)$ and $R(W\|\!\!\upharpoonright_U) \vdash_\sigma b$; now, at_σ maps to equivalent neighborhoods, so $\mathrm{at}_\sigma(R(W\|\!\!\upharpoonright_U)) \vdash_\sigma b$; by the induction hypothesis at σ, there exists a single token $b_0 \in \mathrm{at}_\sigma(R(W\|\!\!\upharpoonright_U))$ such that $\{b_0\} \vdash_\sigma b$, so for the token $c := \langle L(W\|\!\!\upharpoonright_U), b_0\rangle$, which belongs to $\mathrm{at}_{\rho\to\sigma}(W)$ by definition, we have $\{c\} \vdash_{\rho\to\sigma} \langle U, b\rangle$, and we're done. $\square$

Example. At type $\mathbb{D} \to \mathbb{D}$ consider the neighborhood W given by

$$\big\{\big\langle\{\mathtt{0}\},\mathtt{S1}\big\rangle,\big\langle\{\mathtt{1}\},\mathtt{S0}\big\rangle,\big\langle\{\mathtt{B0}*\},\mathtt{SBS}**\big\rangle,\big\langle\{\mathtt{B01}\},\mathtt{SB}*\mathtt{S}*\big\rangle\big\}.$$

Using the algorithm of the proof of Theorem 5.2 with $\mathrm{at}_{\mathbb{D}} := \sup$, we get the equivalent neighborhood $\mathrm{at}_{\mathbb{D}\to\mathbb{D}}(W)$ as

$$\big\{\big\langle\{\mathtt{0}\},\mathtt{S1}\big\rangle,\big\langle\{\mathtt{1}\},\mathtt{S0}\big\rangle,\big\langle\{\mathtt{B0}*\},\mathtt{SBS}**\big\rangle,\big\langle\{\mathtt{B}*\mathtt{1}\},\mathtt{SB}*\mathtt{S}*\big\rangle,\big\langle\{\mathtt{B0}*,\mathtt{B01}\},\mathtt{SBS}*\mathtt{S}*\big\rangle\big\},$$

which is linear: for example, the neighborhood W needs the subneighborhood $\big\{\langle\mathtt{B0}*,\mathtt{SBS}**\rangle,\langle\mathtt{B01},\mathtt{SB}*\mathtt{S}*\rangle\big\}$ to entail the information $\big\langle\{\mathtt{B0}*,\mathtt{B}*\mathtt{1}\},\mathtt{SBS}*\mathtt{S}*\big\rangle$, whereas $\mathrm{at}_{\mathbb{D}\to\mathbb{D}}(W)$ does this with the singleton $\big\{\big\langle\{\mathtt{B0}*,\mathtt{B01}\},\mathtt{SBS}*\mathtt{S}*\big\rangle\big\}$. Notice though that the latter is linear at type $\mathbb{D} \to \mathbb{D}$ while it presupposes the nonlinear entailment $\{\mathtt{B0}*,\mathtt{B}*\mathtt{1}\} \vdash_{\mathbb{D}} \{\mathtt{B0}*,\mathtt{B01}\}$ on the left—naturally, this nuance can be lifted if we first substitute all left neighborhoods by equivalent linear ones. $\square$

5.2 Explicit linearity

The witness we provided in the proof of Theorem 5.2 fails to be a normal form mapping because of the identity, but the argument would work just fine with any other mapping instead of id (which we used for simplicity), as long as it sends to equivalent neighborhoods; so there are plenty of witnesses of implicit linearity which are actually normal form mappings. In particular, there are witnesses which ensure that not only the entailments at $\rho \to \sigma$, but also all entailments of lower type will be linear (see our last example). One may naturally wonder: can we not just restrict ourselves to appropriate normal forms at every type, and work exclusively in a linear setting?

Indeed we can, just not in a downright naive way. First of all we observe that if we achieve linearity at base types then we're done, since all function spaces will also be linear, by the following known result (see [22]).

Proposition 5.3. *Let ρ, σ be coherent information systems, with σ being linear. Then their function space $\rho \to \sigma$ is a linear coherent information system.*

Proof. Let $W \in \mathrm{Con}_{\rho\to\sigma}$ and $\langle U,b\rangle \in \mathrm{Tok}_{\rho\to\sigma}$ be such that $W \vdash_{\rho\to\sigma} \langle U,b\rangle$. This means that $WU \vdash_\sigma b$; since η is linear, there exists a single $b_0 \in WU$, such that $\{b_0\} \vdash_\sigma b$, so there exists a pair $\langle U_0,b_0\rangle \in W$, such that $\{U_0,b_0\} \vdash_{\rho\to\sigma} \langle U,b\rangle$. □

Now in order to obtain linear information systems for our base types, we naturally turn to the linear normal forms which we encountered in section 4.1. The naive way to go about the problem is to consider a class Con^{nf} of neighborhoods in some normal form which we know is linear, and check if the triple $(\mathrm{Tok},\mathrm{Con}^{nf},\vdash)$ will do. The only two choices that we have from our previous discussion are Con^{sup} for the suprema and Con^{cl} for the deductive closures; it turns out that both are bad choices. If we restrict Con to the neighborhoods that can serve as suprema, that is, to *singletons*, we lose the propagation of consistency by entailment, and if we restrict it to deductively closed neighborhoods, then we lose the consistency of singletons and the closure of consistency to subsets. Trying to mend these shortcomings by tweaking, for example, the definition of entailment, only seems to lead further from intuition.

A less naive idea is to capitalize on our results on *path normal forms*, namely, by restricting the *carrier* set to paths, and then adapting consistency and entailment accordingly.

Write $\rho \cong \sigma$, if the ideals of ρ and the ideals of σ are in a bijective correspondence. Moreover, if $x,y:\rho$, write $x \vdash_\rho y$ to mean that for every $V \subseteq_f y$ there is a $U \subseteq_f x$ with $U \vdash_\rho V$.

Proposition 5.4. *Let ι be a finitary base type. There exists a linear coherent information system η, such that $\eta \cong \iota$.*

Proofsketch. Given a finitary base type ι, define ι^p by letting Tok_{ι^p} be Tok_ι^p, Con_{ι^p} be $\mathrm{Con}_\iota \cap \mathscr{P}_f(\mathrm{Tok}_{\iota^p})$, and $\vdash_{\iota^p}$ be $\vdash_\iota \cap (\mathrm{Con}_{\iota^p} \times \mathrm{Tok}_{\iota^p})$ (notice that we allow every neighborhood of paths to be consistent in ι^p, and not just the path reduced ones). It is straightforward to check that ι^p is indeed a coherent information system; we call it the *path subsystem* of ι.

To see that it is linear, let $U \in \mathrm{Con}_{\iota^p}$ and $b \in \mathrm{Tok}_{\iota^p}$ be such that $U \vdash_{\iota^p} b$. Since b is a path, by Proposition 4.2.3 there is an $a \in U$ with $\{a\} \vdash_\iota b$. But a is itself a path, so $\{a\} \vdash_{\iota^p} b$.

For the equivalence of the ideals, consider the mappings $F : \iota \to \iota^p$ and $G : \iota^p \to \iota$, defined by

$$F(x) := \bigcup_{U \subseteq_f x} \{a \in \mathrm{Tok}_{\iota^p} \mid U \vdash_\iota a\},$$

$$G(x) := \bigcup_{U \subseteq_f x} \{a \in \mathrm{Tok}_\iota \mid U \vdash_\iota a\}.$$

Using Theorem 4.5, it is tediously straightforward to show that F and G are indeed well defined, injective (note that $F(x) \sim_\iota x$ and $G(x) \sim_\iota x$), and mutually inverse, thus $\iota \cong \iota^p$. □

It follows that if we are willing to dispose of succinct representations like B00, and restrict to their path-neighborhood representations like $\{\mathtt{B0}*, \mathtt{B}*\mathtt{0}\}$, we obtain a model of information systems, which are not only coherent, but also linear in an explicit way. More precisely, from Propositions 5.3 and 5.4 we obtain the following.

Theorem 5.5 (Linearity)**.** *Let ρ be a type. There exists a linear coherent information system ρ^A, such that $\rho^A \cong \rho$.*

6 Discussion

We showed how we can circumvent the inherent combinatorial complexity of nonflat information systems by working with canonical and even normal forms, and we used these results to show that we can fully recover linearity in a setting with a full, nonlinear entailment predicate. There are several things that suggest themselves as next steps.

On the technical side, we made heavy use of neighborhood mappings between two information systems, prefering them to the whole class of approximable maps. It could be fruitful to look into the *theory of neighborhood mappings* from a wider scope. For instance, a typical neighborhood mapping that is used by default constantly in the relevant literature—indeed, in the previous pages as well—is the neighborhood application, and so are the mappings induced by superunary constructors—we let one such sneak in as an example in section 3. These mappings are obviously well-behaved, since they are monotone in all arguments, but we cannot really talk about them officially with the scanty tools we have given here. It should be direct to generalize the notions of compatible, consistent and monotone mappings to the case of several arguments, and study their exact relationship to the corresponding ideals, and in doing so we may acquire some extra insight into such everyday operations between neighborhoods.

In the case of normal forms, as well, there is a prospect of generalization, if one is willing to take certain questions about the totality of all normal forms seriously. For instance, how many normal forms are there to choose from, anyway? It is easy to see that the class of all normal forms at a given type can be arranged as a semigroup in two natural ways, namely both under composition, and under consistent union $U \mapsto f(U) \cup g(U)$, whenever $f(U) \asymp g(U)$ for every U; both of these structures lack their natural neutral element, namely the identity map and the constant $U \mapsto \varnothing$, respectively. Moreover, both of these semigroups are actually “bands” (every element is idempotent), the former is particularly “rectangular” (the equation $f \circ g \circ h = f \circ h$ holds for all f, g, h) and even “left-zero” (all elements are left-neutral), while the latter is commutative. We list this array of terms to merely indicate that the *theory of semigroups* is rich enough to have thematized such concepts (see for example [8]). This suggests that the deceivingly meager structure of a semigroup might still give good answers on general questions regarding our navigation in the highly complex nonflat domains of choice.

Last, but most relevant for our interest in higher-type computability, the main message of the paper is that we can work linearly even with a nonlinear (that is, not necessarily binary) entailment, and therefore widen the scope of previous work like [15, 22, 16, 11] in a substantial way. There are two further points to stress. Firstly, the localization of linearity, provided in the last section, points the way to the *internalization of the study of sequentiality and linear semantics in the general nonflat case*. There is an enormous amount of work invested in these matters by several people already (see section 5 for some references), and to even adapt the basic ideas to our setting should be rewarding. Secondly, adapting such work to the nonflat case may even lead to surprisingly more powerful results, as we mentioned already in the introduction. In pursuit of such aims, it seems likely that the tools we have used in this work, like neighborhood mappings and their normal forms, may very well prove to be indispensable.

Dues

For their lending an ear, their feedback, their suggestions, thanks to Helmut Schwichtenberg, Davide Rinaldi, Matthias Hofer, and Rhea Marstaller. Matthias Hofer has also provided a proof for a special case of Theorem 4.15, using an earlier variant of eigenneighborhood.

References

[1] Ulrich Berger. Total sets and objects in domain theory. *Ann. Pure Appl. Logic*, 60(2):91–117, 1993.

[2] Gerard Berry. Stable models of typed λ-calculi. Automata, languages and programming,, 5th Colloq., Udine 1978, Lect. Notes Comput. Sci. 62, 72-89 (1978)., 1978.

[3] M. Bezem, J.W. Klop, R. de Vrijer, and Terese. *Term Rewriting Systems*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2003.

[4] Antonio Bucciarelli, Alberto Carraro, Thomas Ehrhard, and Antonino Salibra. On linear information systems. In *LINEARITY*, pages 38–48, 2009.

[5] Pierre-Louis Curien. Definability and full abstraction. In *Computation, meaning, and logic. Articles dedicated to Gordon Plotkin.*, pages 301–310. Amsterdam: Elsevier, 2007.

[6] Martín H. Escardó. On lazy natural numbers with applications to computability theory and functional programming. *SIGACT News*, 24(1):61–67, January 1993.

[7] Jean-Yves Girard, Paul Taylor, and Yves Lafont. *Proofs and types*, volume 7 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, 1989.

[8] John M. Howie. *Fundamentals of semigroup theory.* Oxford: Clarendon Press, 1995.

[9] Simon Huber, Basil A. Karádais, and Helmut Schwichtenberg. Towards a formal theory of computability. In *Ways of Proof Theory (Pohler’s Festschrift)*, pages 257–282. Ontos Verlag, Frankfurt, 2010.

[10] Basil A. Karádais. Implicit atomicity and finite density for non-flat domains. Available at http://www.math.unipd.it/ basilio/d.pdf, 2013.

[11] Basil A. Karádais. *Towards an Arithmetic with Approximations*. PhD thesis, Mathematisches Institut, LMU, 2013. Available at http://www.math.lmu.de/ karadais/t.pdf.

[12] Basil A. Karádais. Atomicity, coherence of information, and point-free structures. Submitted, 2015.

[13] Kenji Miyamoto. *Program Extraction from Coinductive Proofs and its Application to Exact Real Arithmetic*. PhD thesis, Mathematisches Institut, LMU, 2013.

[14] Kenji Miyamoto and Helmut Schwichtenberg. Program extraction in exact real arithmetic. *Mathematical Structures in Computer Science*, FirstView:1–13, 12 2014.

[15] Fritz Müller. Full abstraction for a recursively typed lambda calculus with parallel conditional. Technical Report SFB 124, FB 14, Informatik, Universität des Saarlandes, Saarbrücken, December 1993.

[16] Fritz Müller. Full abstraction for a recursively typed lambda calculus with parallel conditional. *Computing research repository*, abs/0806.1827, 2008. Revised version of [15].

[17] Gordon D. Plotkin. LCF considered as a programming language. *Theoret. Comput. Sci.*, 5(3):223–255, 1977/78.

[18] Gordon D. Plotkin. T^ω as a universal domain. *J. Comput. System Sci.*, 17(2):209–236, 1978.

[19] Florian Ranzi. Recursion and pcf-definability over the partial continuous functionals. Master’s thesis, Mathematisches Institut, LMU, 2008.

[20] Davide Rinaldi. *Formal methods in the theories of rings and domains*. PhD thesis, Mathematisches Institut, LMU, 2014.

[21] Helmut Schwichtenberg. Eine Normalform für endliche Approximationen von partiellen stetigen Funktionalen. In J. Diller, editor, *Logik und Grundlagenforschung, Festkolloquium zum 100. Geburtstag von Heinrich Scholz*, pages 89–95, 1986.

[22] Helmut Schwichtenberg. Recursion on the partial continuous functionals. In C. Dimitracopoulos, L. Newelski, D. Normann, and J. Steel, editors, *Logic Colloquium '05*, volume 28 of *Lecture Notes in Logic*, pages 173–201. Association for Symbolic Logic, 2006.

[23] Helmut Schwichtenberg and Stanley S. Wainer. *Proofs and computations*. Perspectives in Logic. Cambridge University Press, Cambridge, 2012.

[24] Dana S. Scott. Domains for denotational semantics. In *Automata, languages and programming (Aarhus, 1982)*, volume 140 of *Lecture Notes in Comput. Sci.*, pages 577–613. Springer, Berlin, 1982.

[25] Glynn Winskel. Event structures. In *Advances in Petri nets 1986, Part II (Bad Honnef, 1986)*, volume 255 of *Lecture Notes in Comput. Sci.*, pages 325–392. Springer, Berlin, 1987.

[26] Guo-Qiang Zhang. DI-Domains as information systems. Technical Report DAIMI-PB-282, Department of Computer Science, Aarhus University, Aarhus, Denmark, August 1989.

[27] Guo-Qiang Zhang. *Logic of domains.* Basel: Birkhäuser, 1991.

[28] Guo-Qiang Zhang. *dI*–domains as prime information systems. *Inf. Comput.*, 100(2):151–177, 1992.