Chapter 3

Composition Systems

3.1 Introduction

Compositionality is a mechanism to represent entities in a hierarchical way. Each entity is composed of several parts, which themselves are meaningful entities. Each entity is also re-usable in a near-infinite assortment of meaningful combinations to form other entities. Such hierarchical representation of meaningful entities is widely believed to be fundamental to language (Chomsky [2]) as well as to vision, or any other kind of cognition (Bienenstock [1]). On one hand, entities that convey information, such as sentences and scenes, decompose naturally into a hierarchy of meaningful and generic parts, with all the possible meanings of each part being examined. On the other hand, compositions of parts remove ambiguities, because interpretations of parts that do not fit the contextual constraints offered by the composition are removed from further consideration, making parts correctly interpreted at the top level of the hierarchy.

The fundamental importance of compositionality entails addressing the mechanism in a more principled way, and composition systems are devised for this purpose. A composition system includes four components: (1) a set of categories, or "labels", for the meaningful entities; (2) for each category, a set of parameters, or "attributes", that are used to describe entities falling into this category; (3) a set of constraints on compositions, or "composition rules"; and (4) a set of primitive entities, or "terminals", which can not be further decomposed, and which have definite interpretations and serve as the building blocks for other entities. Any entity that is built per the composition rules from the terminals is called an object generated by the composition system.

Even after a composition system is established, one still faces the following question: Why is it the case that the interpretation of a collection of objects as a single composite object, when possible, is generally favored over the interpretation of these same objects as independent entities? The answer is that the description length of a composite object is on average smaller than the total description length of its components. This answer clearly depends on how the objects are encoded, or, from the probability point of view, depends on the probability measure on objects. Any reasonable probability measure on objects generated by a composition system should of course address compositionality, which is the reason why it is called a "compositional probability measure". However, how the measure accommodates compositionality can be explained in many different ways which lead to different formalisms. The formalism that this chapter is devoted to is the one given by Geman, *et. al* [6].

The chapter proceeds as follows. In §3.2, we review the formalism of composition systems given in [6]. §§3.3-3.5 study probability distributions for discrete composition systems. For more general treatment, we refer the readers to [6].

3.2 Definitions and Notations

In this section, we collect the conventions and definitions for composition systems.

Convention. (***-Notation**) For a set S, we will use S^* to represent the set of finite *non-empty* strings of elements of S, i.e.,

$$S^* = \bigcup_{n=1}^{\infty} \{s_1 s_2 \cdots s_n : \; s_i \in S, i = 1, \ldots, n\}.$$

This is nonstandard — usually S^* includes the empty string. For any $\alpha^* \in S^*$, its length is defined as the total number of elements in the string and is denoted as $|\alpha^*|$,

If P is a measure on S, then P^* is a measure on S^*, such that for any (measurable) subset $C \subset S^*$,

$$P^*(C) = \sum_{n=1}^{\infty} P^n(C \cap S^n).$$

If f is a numerical function on S, then f^* is a numerical function on S^* such that for any $\alpha^* = \alpha_1 \cdots \alpha_n \in S^*$, $f^*(\alpha^*) = f(\alpha_1) \cdots f(\alpha_n)$. If g is a function on S which takes values in a general set V, then g^* is a function on S^* which takes values in the set V^*, such that $g^*(\alpha^*)$ is the string $g(\alpha_1) \cdots g(\alpha_n) \in V^*$. Without specification, a set is always assumed to be a general set, even if all its elements are numbers.

Definition 1. Given a *label set* N, *which is always assumed to be countable*, a *terminal set* T, the set of *labeled trees*, Θ, is the set of finite trees with nonterminal nodes labeled by elements of N and terminal (leaf) nodes labeled by elements of T.

Remark 2.

1. $T \subseteq \Theta$;
2. By the *label* of the tree $\omega \in \Theta$ we will mean the label of its root node. We use $\mathrm{L}(\omega)$ ($\mathrm{L} : \Theta \to T \cup N$) to represent the label of ω;
3. $\omega = l(\alpha^*)$, $\alpha^* = \alpha_1 \cdots \alpha_n$, means $\mathrm{L}(\omega) = l$ and the left-to-right daughter subtrees of ω are $\alpha_1, \ldots \alpha_n$;
4. The ordering of daughter nodes is distinguished. So, for example, $l(\alpha, \beta) \neq l(\beta, \alpha)$ unless $\alpha = \beta$;

5. For any $\omega \in \Theta$, define $|\omega|$ as the total number of nodes (including terminals) in ω and $h(\omega)$ as the height (including terminals) of ω;

6. The *yield* of any tree $\omega \in \Theta$, denoted $Y(\omega)$, is the left-to-right string of terminals of ω.

Definition 2. A *composition rule* for the label $l \in N$ is a pair $(B_l, \mathcal{S}_l)$ where B_l, the *binding function*, maps $\Theta^* = \cup_{n=1}^{\infty} \Theta^n$ into an arbitrary *range space*, $\mathcal{R}_l$:

$$B_l : \Theta^* \to \mathcal{R}_l,$$

and $\mathcal{S}_l$, the *binding support*, is a distinguished non-empty subset of $\mathcal{R}_l$, $\emptyset \neq \mathcal{S}_l \subseteq \mathcal{R}_l$. The triple

$$\mathcal{C} = (T, N, \{B_l, \mathcal{S}_l\}_{l \in N})$$

is called a *composition system*.

Remark 3.

1. The *attribute value* of any $\omega = l(\alpha^*) \in \Theta^*$ is the value of $B_l(\alpha^*)$ and is denoted as $A(\omega)$;

2. The *type* of any $\omega \in \Theta$, denoted $\mathrm{T}(\omega)$, is defined as as follows. If $\omega \in T$, then $\mathrm{T}(\omega)$ is ω itself. If $\omega \in \Theta_l$, then $\mathrm{T}(\omega)$ is the pair $(l, A(\omega))$.

3. For any type t, define Θ_t as the set $\{\omega \in \Theta : \ \mathrm{T}(\omega) = t\}$. If $t = (l, b)$, also write Θ_t as $\Theta_{l,b}$.

Definition 3. Given a composition system $\mathcal{C} = (T, N, \{B_l, \mathcal{S}_l\}_{l \in N}\})$, the set of *objects* Ω is the closure of T under $\{(B_l, \mathcal{S}_l)\}_{l \in N}$ in Θ. That is, $\omega \in \Theta$ is an object ($\omega \in \Omega$) if and only if either $\omega \in T$ or $\omega = l(\alpha^*)$, where $\alpha^* \in \Omega^*$ and $B_l(\alpha^*) \in \mathcal{S}_l$. The set of yields of all objects in Ω, i.e.,

$$\{Y(\omega) : \omega \in \Omega\},$$

is called the *language* generated by $\mathcal{C}$.

Remark 4.

1. $\mathcal{S}_l$ is required to be minimal. In other words, for any $b \in \mathcal{S}_l$, there is an $\omega \in \Omega$ such that $\mathrm{L}(\omega) = l$ and $A(\omega) = b$;

2. We use $\mathcal{T}$ to represent the set of all types of objects, i.e.

   $$\mathcal{T} = T \cup \{(l, b) : \ \exists\, \omega \text{ with } \mathrm{L}(\omega) = l \text{ and } A(\omega) = b\}.$$

   Because $\mathcal{S}_l$, $l \in N$, are minimal,

   $$\mathcal{T} = T \cup \{(l, b) : l \in N, \ b \in \mathcal{S}_l\}.$$

3. For any type t, define $\Omega_t = \Omega \cap \Theta_t$.

Definition 4. The *observable measures* are

1. Q, a probability measure on $T \cup N$ with its support being the whole $T \cup N$;
2. Q_l, a probability measure on $\mathcal{R}_l$ with support $\mathcal{S}_l$, for any $l \in N$.

Remark 5. Q and Q_l induce a probability measure on $\mathcal{T}$ which is identical to Q on T, and equals $Q(l)Q_l$ on $\mathcal{S}_l$ for each $l \in N$. The induced measure is still written as Q.

3.3 Compositional Probability Distribution and Its Existence

We only consider the case where T is countable. Because N is always countable, therefore Ω is also countable. Since by definition, $\mathcal{S}_l$ is minimal, then $\mathcal{S}_l$ must be countable. Because for any $l \in N$, the support of Q_l is $\mathcal{S}_l$, for each $b \in \mathcal{S}_l$, $Q_l(b) > 0$.

Definition 5. A *compositional probability measure P on Ω with observable probability measures Q and Q_l* is a probability measure such that

$$P(\omega) = \begin{cases} Q(\omega), & \text{for any } \omega \in T \\ Q(l)Q_l(b)\dfrac{P^*(\alpha^*)}{P^*\left(\{\beta^* \in \Omega^* : \ B_l(\beta^*) = b\}\right)}, & \text{for any } \omega = l(\alpha^*) \in \Omega_{l,b}. \end{cases} \tag{3.1}$$

For explanations of this formulation, see [6].

We now address the issue of existence of compositional probability distributions. Obviously, existence depends not only on the composition rules, but also on the observable measures Q and Q_l. However, we are more interested in results on existence which only depend on composition rules. Firstly, as the term "observable" suggests, Q and Q_l are determined by data and cannot be alternated artificially to accommodate the existence of solution for (3.1). Secondly, results only depending on composition rules are more informative about the structures of composition systems, hence offering more insight into the criteria for "good" composition systems.

Our basic result on existence is the following proposition.

Proposition 4. If for any $l \in N$ and any $b \in \mathcal{S}_l$,

$$\max\{h(\omega) : \ \omega \in \Omega_{l,b}\} < \infty \tag{3.2}$$

and

$$\max\{|\alpha^*| : \ l(\alpha^*) \in \Omega_{l,b}\} < \infty, \tag{3.3}$$

then for any observable probabilities Q and Q_l, there exists a compositional probability measure satisfying (3.1).

The proof of Proposition 4 is quite complicated. We put it in Appendix at the end of the chapter.

Suppose (3.1) has a solution P. For $l \in N$ and $b \in \mathcal{S}_l$, write

$$Z_{l,b} = \frac{Q(l)Q_l(b)}{P^*\Big(\{\beta^* \in \Omega^* : \ B_l(\beta^*) = b\}\Big)}. \tag{3.4}$$

Then for $t = (l, b) \in \mathcal{T}$ and $\omega = l(\alpha^*) \in \Omega_t$, (3.1) can be written as

$$P(\omega) = P^*(\alpha^*)Z_t,$$

Let $f(t;\omega)$ be the number of subtrees of ω with type t. By induction, it is easy to see that

$$P(\omega) = \prod_{t \in T} Q(t)^{f(t;\omega)} \prod_{t \in \mathcal{T} \setminus T} Z_t^{f(t;\omega)}.$$

Because $\prod_{t \in T} Q(t)^{f(t;\omega)} = Q^*(Y(\omega))$,

$$P(\omega) = Q^*(Y(\omega)) \prod_{t \in \mathcal{T} \setminus T} Z_t^{f(t;\omega)} = Q^*(Y(\omega))Z^{f(\omega)}, \tag{3.5}$$

where $Z = \{Z_t\}_{t \in \mathcal{T} \setminus T}$, $f(\omega) = \{f(t;\omega)\}_{t \in \mathcal{T} \setminus T}$ and $Z^{f(\omega)}$ is the product of all $Z_t^{f(t;\omega)}$. Because P is a compositional probability distribution, for any $t \in \mathcal{T} \setminus T$,

$$\sum_{\omega \in \Omega_t} Q^*(Y(\omega))Z^{f(\omega)} = \sum_{\omega \in \Omega_t} P(\omega) = Q(t).$$

Recall that for $t = (l, b)$, $Q(t) = Q(l)Q_l(t)$.

Therefore, we have proved that if (3.1) has a solution, then the equation system induced by the composition system with Z as the unknowns,

$$\sum_{\omega \in \Omega_t} Q^*(Y(\omega))Z^{f(\omega)} = Q(t), \quad \text{for all } t \in \mathcal{T} \setminus T, \tag{3.6}$$

has a solution given by (3.4). Conversely, if (3.6) has a solution Z, then P given by (3.5) satisfies (3.1). Therefore, the existence of solution for (3.1) is equivalent to the existence of solution for (3.6).

Based on Proposition 4, we can prove another result on existence without assuming (3.2).

Proposition 5. Assume the set $\mathcal{T} \setminus T$ is finite. Also assume for each $(l, b) \in \mathcal{T}$, (3.3) is satisfied. If for every $t \in \mathcal{T} \setminus T$, the domain of convergence of the series

$$\sum_{\omega \in \Omega_t} Q^*(Y(\omega))Z^{f(\omega)} \tag{3.7}$$

is open inside the region $Z > 0$, then there is a solution for (3.6).

Proof. Because of (3.3),

$$h(l,b) = |\{|\alpha^*| : \ l(\alpha^*) \in \Omega_{l,b} \ \}| < \infty.$$

For $n \in \mathbf{N}$, let $\Omega_n = \{\omega \in \Omega, \ h(\omega) \leq n\}$. Because $\mathcal{T}\backslash T = \{(l,b) : \Omega_{l,b} \neq \emptyset\}$ is finite, when n is large enough, Ω_n intersects with each Ω_t, $t \in \mathcal{T}$. For such Ω_n, both (3.2) and (3.3) are satisfied, i.e.,

$$\max\{h(\omega) : \ \omega \in \Omega_{l,b} \cap \Omega_n\} < \infty$$

and

$$\max\{|\alpha^*| : \ l(\alpha^*) \in \Omega_{l,b} \cap \Omega_n\} < \infty.$$

Then by Proposition 4, there is a compositional probability distribution P_n on Ω_n, such that for any $\omega = l(\alpha^*) \in \Omega_{l,b} \cap \Omega_n$,

$$P_n(\omega) = Q(l)Q_l(b)\frac{P_n^*(\alpha^*)}{P_n^*\Big(\{\beta^* \in \Omega_n^* : \ B_l(\beta^*) = b\}\Big)}.$$

Note that if $\omega = l(\alpha^*) \in \Omega_n$, then $\alpha^* \in \Omega_n^*$, and therefore $P_n^*(\alpha^*)$ in the above formula makes sense.

Define $Z_n = \{Z_{l,b,n}\}$ as in (3.4), i.e.,

$$Z_{l,b,n} = \frac{Q(l)Q_l(b)}{P_n^*\Big(\{\beta^* \in \Omega^* : \ B_l(\beta^*) = b\}\Big)}. \tag{3.8}$$

Then as in (3.6),

$$\sum_{\substack{\omega \in \Omega_{l,b} \\ h(\omega) \leq n}} Q^*(Y(\omega))Z_n^{f(\omega)} = Q(l)Q_l(b).$$

Since for each $(l,b) \in \mathcal{T}\backslash T$,

$$\frac{Q(l)Q_l(b)}{h(l,b)} \leq Z_{l,b,n} = \frac{Q(l)Q_l(b)}{\sum\limits_{\substack{B_l(\beta^*)=b \\ l(\beta^*) \in \Omega_n}} P_n^*(\beta^*)} \leq \frac{Q(l)Q_l(b)}{\sum\limits_{\substack{B_l(\beta^*)=b \\ l(\beta^*) \in \Omega_n}} D^*(\beta^*)},$$

Z_n are bounded. The definition of D is given by (A3.3) in Appendix.

Because $\mathcal{T}\backslash T$ is finite, there is a subsequence Z_{n_i} of Z_n which is uniformly convergent to, say, $\xi = \{\xi_{l,b}\}$. Given any $\epsilon = \{\epsilon_{l,b}\}$, with $0 < \epsilon_{l,b} < \xi_{l,b}$, for large enough i, $Z_{n_i} > \xi - \epsilon$, that is, for each (l,b), $Z_{l,b,n_i} > \xi_{l,b} - \epsilon_{l,b}$ Therefore

$$\sum_{\substack{\omega \in \Omega_{l,b} \\ h(\omega) \leq n_i}} Q^*(Y(\omega))(\xi - \epsilon)^{f(\omega)} \leq Q(l)Q_l(b).$$

Letting $i \to \infty$ and then $\epsilon \to 0$, we get

$$\sum_{\omega \in \Omega_{l,b}} Q^*(Y(\omega))\xi^{f(\omega)} \le Q(l)Q_l(b).$$

By the assumption that the domain of convergence of the series of (3.7) is open for each $t \in \mathcal{T} \backslash T$, for any $\beta = \{\beta_{l,b}\}$ with $\beta_{l,b} > 0$ being small enough,

$$\sum_{\omega \in \Omega_{l,b}} Q^*(Y(\omega))(\xi + \beta)^{f(\omega)} < \infty.$$

When i is large enough, $Z_{n_i} \le \xi + \beta$. Therefore,

$$\sum_{\omega \in \Omega_{l,b}} Q^*(Y(\omega))(\xi + \beta)^{f(\omega)} \ge Q(l)Q_l(b).$$

Letting $\beta \to 0$, we get

$$\sum_{\omega \in \Omega_{l,b}} Q^*(Y(\omega))\xi^{f(\omega)} \ge Q(l)Q_l(b).$$

Therefore, ξ is a solution of (3.6). □

Example 1. We consider the following composition system (also see §4.3, [6],). Let $T = \{t\}$, and $N = \{S\}$. If

$$B_S(\alpha^*) = \begin{cases} 1 & \text{when } \alpha^* = (\beta_1, \beta_2), |Y(\beta_1)| = |Y(\beta_2)| \\ 0 & \text{otherwise} \end{cases}$$

and $\mathcal{S}_S = \{1\}$, then Ω is the set of balanced binary trees. The associated language is the set of strings of t of length 2^n, $n \ge 0$. Let $Q(S) = p$ and $Q(t) = q = 1 - p$, with $p \in (0, 1)$. Then the corresponding equation system is

$$\sum_{n=1}^{\infty} q^{2^n} Z^{2^n - 1} = p.$$

The convergence interval of the series on the left hand side of the equation is $(-1/q, 1/q)$, which is open. Therefore, there is a solution of the equation on $\{Z > 0\}$.

If in the above system, we change the binding function B_S to

$$B_S(\alpha^*) = \begin{cases} 1 & \text{when } \alpha^* = (\beta_1, \beta_2), |Y(\beta_1)| = |Y(\beta_2)| \text{ or } |Y(\beta_2)| + 1 \\ 0 & \text{otherwise} \end{cases}$$

while keeping everything else unchanged, then the generated language is the set of strings t^n, $t \ge 1$. The corresponding equation is

$$\sum_{n=2}^{\infty} q^n Z^{n-1} = p.$$

Again, the convergence interval of the series on the left hand side is $(-1/q, 1/q)$, which implies there is a solution for the equation on $\{Z > 0\}$. □

The following example shows the optimality of Proposition 5.

Example 2. Take $T = \{t\}$, $N = \{S\}$, and

$$B_S(\alpha^*) = \begin{cases} 1 & \text{if } \alpha^* = t \\ 2 & \text{if } \alpha^* = (\alpha, \beta) \text{ and } \mathrm{L}(\alpha) = \mathrm{L}(\beta) = S \\ 0 & \text{otherwise.} \end{cases}$$

Then $\mathcal{S}_S = \{1, 2\}$ corresponds to the context-free grammar

$$S \to SS, \ S \to t.$$

Take $Q(t) = u$ and $Q(S) = v = 1 - u$. The probabilities

$$Q_S(b) = \begin{cases} p & \text{if } b = 2 \\ q = 1 - p & \text{if } b = 1 \end{cases}$$

correspond to the production probabilities $P(S \to SS) = p$ and $P(S \to t) = q$. The string that the only tree in $\Omega_{S,1}$ generates is t, and the set of strings that trees in $\Omega_{S,2}$ generate is $\{t^n\}_{n \geq 2}$. For each $n \geq 2$, there are $\Gamma(2n-1)/\Gamma(n)\Gamma(n+1)$ trees with the same yield t^n. For each such tree ω, $f(S, 1; \omega) = n$, and $f(S, 2; \omega) = n - 1$. Hence the corresponding equation system is

$$\begin{cases} uZ_{S,1} = vq \Rightarrow Z_{S,1} = \dfrac{vq}{u} \\ \displaystyle\sum_{n=2}^{\infty} \frac{(2n-2)!}{(n-1)!n!} u^n Z_{S,1}^n Z^{n-1} = vp. \end{cases}$$

Substitute $Z_{S,1} = vqu^{-1}$ into the second equation. The convergence domain of the resulting power series

$$F(Z) = \sum_{n=2}^{\infty} \frac{(2n-2)!}{(n-1)!n!} (vq)^n Z^{n-1}$$

is the closed interval $[-1/4vq, 1/4vq]$. We know that if $p > 1/2$, then there is no compositional probability distribution for the grammar (see §4.3, [6]). When $Z = 1/4vq$, the value of $F(Z)$ is vq. In order that there is a solution, it is necessary and sufficient that $F(1/4vq) \geq vp$, i.e., $q \geq p$, or $p \leq 1/2$. □

Example 3. The composition systems in Example 1 share the following properties.

1. The set $\mathcal{T} = T \cup \{(l, b) : l \in N, b \in \mathcal{S}_l\}$ is finite;
2. The arity of each B_l is 2;
3. For each $t \in \mathcal{T} \setminus T$,

$$\limsup_{n \to \infty} |\{Y(\omega) : \ \omega \in \Omega_t, \ |Y(\omega)| = n\}|^{1/n} = 1.$$

and

$$\limsup_{n \to \infty} \max_{\substack{Y \in T^* \\ |Y| = n}} |\{\omega \in \Omega_t : \ |Y(\omega)| = Y\}|^{1/n} = 1.$$

4. For each $t \in \mathcal{T}\backslash\mathcal{T}$, there are constants $0 \le \beta_{s,t} \le 1$, and $K_{s,t} \ge 0$ for all $s \in \mathcal{T}$, and $\omega_n \in \Omega_t$ with $|\omega_n| \to \infty$, such that

$$\beta_{s;t}|\omega_n| - K_{s;t} \le f(s;\omega_n) \le \beta_{s;t}|\omega_n| + K_{s;t}.$$

We show that if a composition system satisfies the above conditions, then for each $t \in \mathcal{T}$, the domain of convergence of the induced series

$$F_t(Z) = \sum_{\omega\in\Omega_t} Q^*(Y(\omega))Z^{f(\omega)}$$

is open inside the region $Z > 0$ for each $t \in \mathcal{T}\backslash T$, and hence compositional probability measures always exist.

Suppose F_t converges at some $Z > 0$. We want to show Z is an inner point in the domain of convergence of F_t. The series in x

$$F_t(Zx) = \sum_{\omega\in\Omega_t} Q^*(Y(\omega))Z^{f(\omega)}x^{|f(\omega)|}$$

is a univariate power series, where $|f(\omega)| = \sum f(t;\omega)$. By condition 2,

$$|f(\omega)| = \frac{|\omega| - 1}{2}.$$

Define power series

$$f_t(x) = \sum_{\omega\in\Omega_t} Q^*(Y(\omega))Z^{f(\omega)}x^{|\omega|},$$

Let ρ be the radius of convergence of f_t. We will show $\rho > 1$. Once this is proved, it is easy to see every $Z' < \rho Z$ is in the domain of convergence of F_t, implying Z is an inner point of the domain of convergence of F_t.

By conditions 2 and 3,

$$\limsup_{n\to\infty} |\{\omega \in \Omega_t : \ |\omega| = n\}|^{1/n} = 1.$$

Therefore, by condition 4,

$$\frac{1}{\rho} = \limsup_{\substack{|\omega|\to\infty \\ \omega\in\Omega_t}} \left|Q^*(Y(\omega))Z^{f(\omega)}\right|^{1/|\omega|} = \prod_{s\in T} Q(s)^{\beta_{s;t}} \prod_{s\in\mathcal{T}\backslash T} Z_s^{\beta_{s;t}}.$$

There are infinitely many $\omega \in \Omega_t$, such that

$$\begin{aligned}
& Q^*(Y(\omega))Z^{f(\omega)} \\
= \ & \prod_{s\in T} Q(s)^{f(s;\omega)} \prod_{s\in\mathcal{T}\backslash T} Z_s^{f(s;\omega)} \\
\ge \ & \prod_{s\in T} Q(s)^{\beta_{s;t}|\omega|+K_{s;t}} \prod_{\substack{s\in\mathcal{T}\backslash T \\ Z_s\ge 1}} Z_s^{\beta_{s;t}|\omega|-K_{s;t}} \prod_{\substack{s\in\mathcal{T}\backslash T \\ Z_s\le 1}} Z_s^{\beta_{s;t}|\omega|+K_{s;t}} \\
\ge \ & \rho^{-|\omega|} \prod_{s\in T} Q(s)^{K_{s;t}} \prod_{\substack{s\in\mathcal{T}\backslash T \\ Z_s\ge 1}} Z_s^{-K_{s;t}} \prod_{\substack{s\in\mathcal{T}\backslash T \\ Z_s\le 1}} Z_s^{K_{s;t}}.
\end{aligned}$$

Because $F_{l,b}(Z)$ converges, $1/\rho < 1$. Thus $\rho > 1$. □

3.4 Subsystems

Suppose we have a composition system $\mathcal{C}' = (T', N', \{B_l, \mathcal{S}_l\}_{l \in N'}$ with Ω' as the set of trees. We can build a new composition system in the following way. First, take Ω' as part of a new terminal set T''. Suppose $T'' = \Omega' \cup A$, where $A \cap \Omega' = \emptyset$. Then we define a label set N'' disjoint from N', and for each label $l \in N''$, a composition rule $(B_l, \mathcal{S}_l)$. The new composition system $\mathcal{C}'' = (T'', N'', \{B_l, \mathcal{S}_l\}_{l \in N''})$ is not a super-system of $\mathcal{C}'$, because $\mathcal{C}'$ and $\mathcal{C}''$ have disjoint label sets and composition rules and their terminal sets are different. On the other hand, the composition system $\mathcal{C}$ with terminal set $T' \cup A = T' \cup (T'' \backslash \Omega')$, label set $N \cup N'$, and composition rules $\{B_l, \mathcal{S}_l\}_{l \in N \cup N'}$ is a super-sytem of $\mathcal{C}'$.

The above construction can be formulated into the following definition.

Definition 6. Suppose $\mathcal{C} = (N, \{B_l, \mathcal{S}_l\}_{l \in N}, T)$ is a composition system with Ω being the set of objects. Suppose T' and N' are non-empty subsets of T and N, respectively. For each $l \in N'$, assume $\mathcal{S}'_l$ is a non-empty subset of $\mathcal{S}_l$. Let $\mathcal{C}'$ be the composition system formed by T', N', and $\{B_l, \mathcal{S}'_l\}_{l \in N'}$. Let Ω' be the set of objects generated by $\mathcal{C}'$. Since $T' \neq \emptyset$, Ω' is not empty.

$\mathcal{C}'$ is said to be a subsystem of $\mathcal{C}$, denoted as $\mathcal{C}' \subset \mathcal{C}$, if Ω' contains all $\omega \in \Omega$ with $\mathrm{L}(\omega) = l \in N'$ and $A(\omega) \in \mathcal{S}'_l$. The composition system with terminal set $T \cup \Omega'$, label set $N_1 \cup N_2$, where $N_1 = N \backslash N'$, and $N_2 = \{l \in N' : \ \mathcal{S}_l \backslash \mathcal{S}'_l \neq \emptyset\}$, and composition rules $\{B_l, \mathcal{S}_l\}_{l \in N_1} \cup \{B_l, \mathcal{S}_l \backslash \mathcal{S}'_l\}_{l \in N_2}$, is called the quotient system of $\mathcal{C}$ over $\mathcal{C}'$ and is denoted as $\mathcal{C}/\mathcal{C}'$. The set of objects generated by $\mathcal{C}/\mathcal{C}'$ is denoted as Ω/Ω'. □

Intuitively speaking, $\mathcal{C}/\mathcal{C}'$ is an abstraction of $\mathcal{C}$. It takes objects in $\mathcal{C}'$ as terminals, which, by definition, are not decomposable, thus losing the details about them. On the other hand, $\mathcal{C}$ can be thought of as being more detailed than $\mathcal{C}/\mathcal{C}'$. The information about $\mathcal{C}$ is determined by that about both $\mathcal{C}/\mathcal{C}'$ and $\mathcal{C}'$.

Subsystems can be used to construct of compositional probability distributions. For example, if both $\mathcal{C}'$ and $\mathcal{C}/\mathcal{C}'$ satisfy the conditions of Proposition 4, then for any Q on $T \cup N$ and Q_l on $\mathcal{S}_l$, $l \in N$, both

$$P_1(\omega) = \begin{cases} Q(\omega), & \text{if } \omega \in T', \\ Q(l) Q_l(B_l(\alpha^*)) \dfrac{P_1^*(\alpha^*)}{P_1^*\Big(\{\beta^* \in \Omega^* : B_l(\beta^*) = B_l(\alpha^*)\}\Big)}, & \text{if } \omega = l(\alpha^*) \in \Omega', \end{cases}$$

and

$$P_2(\omega) = \begin{cases} Q(\omega), & \text{if } \omega \in T \backslash T', \\ P_1(\omega), & \text{if } \omega \in \Omega', \\ Q(l) Q_l(B_l(\alpha^*)) \dfrac{P_2^*(\alpha^*)}{P_2^*\Big(\{\beta^* \in \Omega^* : B_l(\beta^*) = B_l(\alpha^*)\}\Big)}, & \text{if } \omega = l(\alpha^*) \in \Omega/\Omega', \end{cases}$$

have solutions. Note that neither P_1 nor P_2 is a probability distribution, because each of the sums of P_1 and P_2 is less than 1. The existence of the solutions is guaranteed by Proposition

6 in Appendix. On the other hand, the measure

$$P(\omega) = \begin{cases} P_1(\omega), & \text{if } \omega \in \Omega' \\ P_2(\omega), & \text{if } \omega \in \Omega/\Omega' \end{cases}$$

is a compositional probability distribution on Ω.

There is another application of subsystems. Assume we have a compositional probability distribution P_1 on a system $\mathcal{C}' = (T', N', \{B'_l, \mathcal{S}'_l\}_{l \in N'})$, and the observable measures are $Q'(l)$ and $Q'_l(b)$. Suppose $\mathcal{C}'$ is expanded into a larger system $\mathcal{C}$. Assume the expansion does not change $\mathcal{S}_l$ for any $l \in N'$. It just adds more terminals to T', and more labels to N', and sets up rules for the new labels.

$Q'(l)$ now becomes the conditional probability measure on N'. Thus in $\mathcal{C}$, the probability of each $l \in N'$ is changed to $\lambda Q'(l)$, for some constant λ. However, for any $l \in N'$ and any $b \in \mathcal{S}'_l$, $Q'_l(b)$ is not changed. If all the binding functions B_l, $l \in N'$, have the same arity, then the probability of $\omega \in \Omega'$ is simply changed to $\lambda P_1(\omega)$ when ω is considered as an element in Ω. This makes enlarging a system and adjusting the probability distribution easy.

3.5 The Gibbs Form of Compositional Probability Distributions

We now discuss the Gibbs form of compositional probability distributions. Suppose P is a compositional probability distribution on Ω. Then P can be formulated as in (3.5). Extend $Z = \{Z_t\}_{t \in \mathcal{T} \setminus T}$ to $\{Z_t\}_{t \in \mathcal{T}}$, where for $t \in T$, $Z_t = Q(t)$. Also extend $f(\omega) = \{f(t;\omega)\}_{t \in \mathcal{T} \setminus T}$ to $\{f(t;\omega)\}_{t \in \mathcal{T}}$. Finally, let $\lambda = \{\log Z_t\}_{t \in \mathcal{T}}$. Then $P(\omega)$ takes the form of Gibbs distribution,

$$P(\omega) = P_\lambda(\omega) = \exp\Big(\lambda \cdot f(\omega)\Big). \tag{3.9}$$

A special property of the Gibbs distribution (3.9) is that its normalization constant is 1.

For an arbitrary λ, P_λ is a positive measure on Ω, but not necessarily a probability measure. Among all the λ's which make P_λ a probability measure on Ω, $\lambda = \{\log Z_t\}_{t \in \mathcal{T}}$ has the following minimization property,

$$\lambda = \underset{\substack{\lambda':\ P_{\lambda'} \text{ is} \\ \text{a prob.}}}{\arg\min} \sum_{t \in \mathcal{T}} Q(t) \log \frac{Q(t)}{P_{\lambda'}(\Omega_t)}. \tag{3.10}$$

Indeed, the sum on the right hand side of (3.10) is always non-negative. If an compositional distribution exists, then the sum achieves 0 at $\lambda = \{\log Z_t\}_{t \in \mathcal{T}}$. Therefore λ is a minimizer.

Bibliography

[1] E. Bienenstock. Composition. A. Aertsen and V. Braitenberg (editors). *Brain Theory: Biological Basis and Computational Theory of Vision*. In preparation.

[2] N. Chomsky. *Syntactic Structures*. Mouton. 1976.

[3] S. Geman, D. Potter, and Z. Chi. *Compositional Systems*. Technical Report, Division of Applied Mathematics, Brown University. February, 1998.

[4] D. F. Potter. *Compositional Pattern Recognition*. PhD Thesis, Division of Applied Mathematics, Brown University. 1998. In preparation.

Appendix

In this Appendix we will prove Proposition 4. First, we need to introduce some notations.

Definition 7.

1. The mapping $l : \alpha^* \mapsto l(\alpha^*)$ can be thought of as a function from Θ^* to Θ_l, which is one-to-one and onto. We write its inverse as l^{-1};

2. The *graph* of any tree $\omega \in \Theta$ is a tree with the same topology as ω but with all nodes being unlabeled (Figure 3.1);

3. That a tree ω is *compatible* with a tree graph g, denoted as $\omega \sim g$ means the following. If g is a tree with a single node, then $\omega \sim g$. If g is a tree with daughter subtrees $g_1, \ldots, g_n$, then $\omega = l(\alpha_1, \ldots, \alpha_m)$ is compatible with g if and only if $m = n$ and each α_i, $1 \le i \le m$, is compatible with g_i. If $\omega \sim g$, then for each node $v \in g$, let $\omega(v)$ represent the subtree of ω with v as the root;

4. The *arrangement* of any $\omega \in \Omega$, denoted $E(\omega)$, is a tree with the same topology as ω but with each node being annotated by its type (Figure 3.1);

5. For any $\omega \in \Theta$, the depth of a subtree ω' is the depth of the root of ω' in the tree ω and is denoted as $d(\omega', \omega)$. By this definition, $d(\omega, \omega) = 1$. □

Proposition 4 can be expressed in a little more general form, where Q is a finite positive measure instead of a probability measure on $\mathcal{T}$.

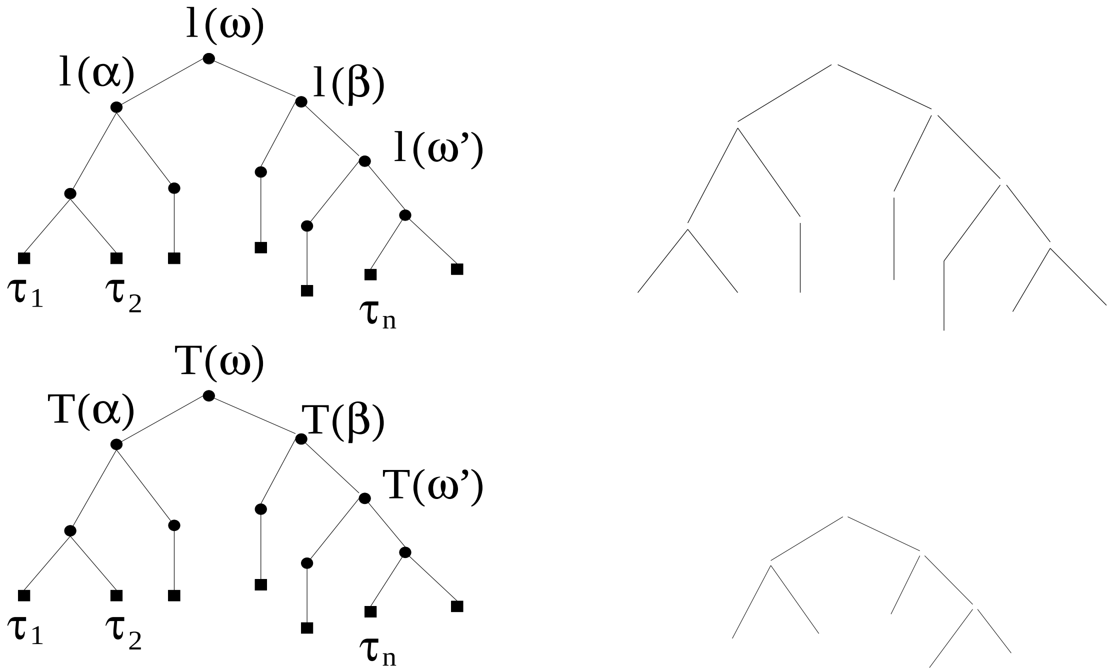

Figure 3.1: A tree $\omega = l(\alpha, \beta)$ (upper left), its graph (upper right), its arrangement (lower left) and a compatible graph. Circles are nonterminals and squares are terminals

Proposition 6. Suppose for any $l \in N$ and any $b \in \mathcal{S}_l$,

$$\max\{h(\omega) : \ \omega \in \Omega_{l,b}\} < \infty \tag{A3.1}$$

and

$$\max\{|l^{-1}(\omega)| : \ \omega \in \Omega_{l,b}\} < \infty. \tag{A3.2}$$

Assume Q is a positive measure on $\mathcal{T}$, with $Q(t) > 0$ for each $t \in \mathcal{T}$. If $Q(\mathcal{T}) < \infty$, then there exists a compositional probability measure satisfying (3.1).

Our proof of Proposition 6 is based on the following fixed point theorem, which is due to Schauder.

Theorem Suppose X is a Banach space, $C \subset X$ is closed and convex. If $F : C \to C$ is continuous and $F(C)$ is sequentially compact, then F has a fixed point in C. □

Proof of Proposition 6: For any type $t = (l, b)$, let

$$\begin{aligned} h(t) &= \max_{\omega \in \Omega_t} h(\omega), \\ m(t) &= |\{l^{-1}(\omega) : \ \omega \in \Omega_t\}|, \\ n(t) &= \max_{\omega \in \Omega_t} |l^{-1}(\omega)|. \end{aligned}$$

Then $h(t)$, $m(t)$, and $n(t)$ are finite. Also write $h(l, b)$, $m(l, b)$ and $n(l, b)$ for $h(t)$, $m(t)$, and $n(t)$, respectively. For consistency, define, for $\tau \in T$, $h(\tau) = 1$, $m(\tau) = 1$ and $n(\tau) = 0$.

Let X be the l^1 space on Ω, i.e.

$$X = \left\{ x : \Omega \to \mathbf{R} : \ x(\tau) = Q(\tau), \ \forall\, \tau \in T, \ \sum_{\omega \in \Omega} |x(\omega)| < \infty \right\}.$$

Let $M = \max\{Q(\mathcal{T}), 1\}$. Define a positive measure D on Ω inductively as follows. For $\tau \in T$, $D(\tau) = Q(\tau) > 0$. For any $t = (l, b) \in \mathcal{T} \backslash T$ and $\omega = l(\alpha^*) \in \Omega_t$,

$$D(\omega) = \frac{Q(t)}{m(t) M^{n(t)}} D^*(\alpha^*). \tag{A3.3}$$

For consistency, we define $D^*(\emptyset) = 1$. Then $D(\tau)$ can also be written in the form of (A3.3).

For ease of typing, we introduce a new notation. If $\beta^* \in \Omega^*$ satisfies $B_l(\beta^*) = b \in \mathcal{S}_l$, then we say β^* is compatible with type $t = (l, b)$ and use $\beta^* \sim t$ to represent this. For consistency, we define $\emptyset$ to be the only string that is compatible with a type t if $t \in T$.

Lemma 4. D has the following properties,

$$D(\Omega) \le M \tag{A3.4}$$

$$0 < \sum_{\beta^* \sim t} D^*(\beta^*) \le m(t) M^{n(t)}. \tag{A3.5}$$

Proof. We will get (A3.4) by showing for all $n \ge 1$,

$$\sum_{h(\omega) \le n} D(\omega) \le M, \tag{A3.6}$$

When $n = 1$, the sum equals $\sum_T Q(\tau) \le M$. Assume (A3.6) is true for $n \le k$. Then

$$\sum_{h(\omega) \le k+1} D(\omega) = \sum_{t \in \mathcal{T}} \frac{Q(t)}{m(t) M^{n(t)}} \sum_{\substack{\alpha^* \sim t \\ h(\alpha^*) \le k}} D^*(\alpha^*),$$

where $h(\alpha^*) = \max_{\alpha \in \alpha^*} h(\alpha)$. By induction hypothesis,

$$\sum_{\substack{\alpha^* \sim t \\ h(\alpha^*) \le k}} D^*(\alpha^*) = \sum_{j=1}^{\infty} \sum_{\substack{\alpha^* \sim t \\ h(\alpha^*) \le k \\ |\alpha^*| = j}} D^*(\alpha^*) \le \sum_{j=1}^{\infty} \mathbf{1}_{\{\exists \omega \in \Omega_t, \ |l^{-1}(\omega)| = j\}} M^{n(t)} = m(t) M^{n(t)}, \tag{A3.7}$$

which, together with last equation, implies (A3.6). Letting $n \to \infty$ in (A3.6), we then prove (A3.4). Letting $k \to \infty$ in (A3.7), we get (A3.5). □

Lemma 5. Let g be a tree graph. To each node $v \in g$, assign a type $t(v)$, such that $t(v) \in \mathcal{T} \backslash T$ unless v is a leaf of g. Then

$$\sum_{\substack{\omega \sim g, \forall v \in g \\ \mathrm{T}(\omega(v)) = t(v)}} D(\omega) \le E(g, t), \tag{A3.8}$$

where

$$E(g,t) = \prod_{v\in g}{}' \frac{Q(t(v))}{\sum\limits_{\beta^*\sim t(v)} D^*(\beta^*)} \cdot \prod_{v\in g}{}'' Q(t(v)), \tag{A3.9}$$

where the production $\prod'$ runs over all non-terminal nodes of g and $\prod''$ over all terminal nodes of g.

Proof. When $h(g) = 1$, the right hand side of (A3.8) is $Q(t)$, where t is the type assigned to the only node in g. The left hand side of (A3.8) is the sum of

$$\sum_{\omega\in\Omega_t} D(\omega) = \sum_{\alpha^*\sim t} \frac{Q(t)}{m(t)M^{n(t)}} D^*(\alpha^*).$$

By (A3.5), the sum is less than $Q(t)$.

Suppose (A3.8) is true for all finite graphs g with $h(g) \le k$. Given a tree graph g with height $k+1$ and daughter subtrees $g_1, \ldots, g_n$, by (A3.3) and (A3.5), for any $\omega \sim g$ with $\mathrm{T}(\omega) = t(v_0)$, where v_0 is the root of g,

$$D(\omega) \le \frac{Q(t(v_0))}{\sum\limits_{\beta^*\sim t(v_0)} D^*(\beta^*)} D^*(\alpha^*),$$

which leads to

$$\begin{aligned}
\sum_{\substack{\omega\sim g, \forall v\in g\\ \mathrm{T}(\omega(v))=t(v)}} D(\omega) &\le \frac{Q(t(v_0))}{\sum\limits_{\beta^*\sim t(v_0)} D^*(\beta^*)} \sum_{\substack{\alpha^*\sim t(v_0)\\ \text{for } i=1,\ldots n, \alpha_i\sim g_i,\\ \forall v\in g_i, \mathrm{T}(\alpha_i(v))=t(v)}} D^*(\alpha^*) \\
&\le \frac{Q(t(v_0))}{\sum\limits_{\beta^*\sim t(v_0)} D^*(\beta^*)} \prod_{i=1}^{n} \sum_{\substack{\alpha_i\sim g_i, \forall v\in g_i\\ \mathrm{T}(\alpha_i(v))=t(v)}} D(\alpha_i).
\end{aligned}$$

Every $h(g_i) \le k$. Then by induction, we prove (A3.8). □

Now define C as the set of all $x \in X$ which satisfy the following conditions,

- C1. For any $\tau \in T$, $x(\tau) = Q(\tau)$;
- C2. For any $\omega \in \Omega$, $x(\omega) \ge D(\omega)$;
- C3. For any tree graph g, any assignment $t : \{v \in g\} \to \mathcal{T}$ with $t(v) \notin T$ unless v is a terminal of g,

$$\sum_{\substack{\omega\sim g, \forall v\in g\\ \mathrm{T}(\omega(v))=t(v)}} x(\omega) \le E(g,t). \tag{A3.10}$$

C is not empty, because $D \in C$. We want to use Schauder's fixed point theorem to prove there is a solution for (3.1) in C. To this end, define a mapping $F : C \to R^{\Omega}$, such that

$$\begin{cases} (Fx)(\tau) = x(\tau), & \forall\ \tau \in T \\ (Fx)(\omega) = Q(l)Q_l(b)\dfrac{x^*(\alpha^*)}{\sum\limits_{\substack{\beta^*\in\Omega^* \\ B_l(\beta)=b}} x^*(\beta^*)}, & \forall\ \omega = l(\alpha^*) \in \Omega, B_l(\alpha^*) = b \end{cases} \tag{A3.11}$$

The definition (A3.11) makes sense because

$$0 < \sum_{\beta^*\sim t} x^*(\beta^*) \leq m(t)M^{n(t)}. \tag{A3.12}$$

The second half of (A3.12) can be proved in the same way as (A3.5).

It is clear that C is convex and closed. In order to show that $F(C)$ is sequentially compact, it is enough to show that $F(C) \subset C$ and C is tight. First we shall show that C is tight.

Lemma 6. For any $\epsilon > 0$, $n \geq 2$, and finite $I \subset \mathcal{T}$, there is a finite $J \subset \mathcal{T}$ with $J \supset I$, such that

$$\sum_{(g,t)\in G} E(g,t) < \epsilon, \tag{A3.13}$$

where $G = G_n(I, J)$ is the set of pairs (g, t) satisfying the following conditions,

- G1 For each $(g,t) \in G$, $h(g) = n$, and $t : \{v \in g\} \to \mathcal{T}$ is a mapping such that $t(v) \notin T$ unless v is a terminal of g;
- G2 For any $v \in g$ with $d(v, g) \leq n - 1$, $t(v) \in I$;
- G3 There is a $v \in g$ with $d(v, g) = n$ such that $t(v) \notin J$;
- G4 The set $\{\omega \in \Omega : \omega \sim g$, and for every $v \in g, \mathrm{T}(\omega(v)) = t(v)\}$ is not empty;
- G5 Every $(g, t) \in G$ is maximum. That is, there are no (g, t) and (g', t'), such that $g \subset g'$ and for any $v \in g$, $t(v) = t'(v)$.

Proof. Let $N = \max_{t\in I} n(t)$. Here t represents an element in $\mathcal{T}$ instead of a mapping to $\mathcal{T}$. Then N is the maximum number of daughter subtrees a tree ω whose type is in I can have. By (A3.2), N is finite. Fix $J \supset I$ and let $G = G_n(I, J)$. If g is a tree graph with $(g, t) \in G$ for some mapping $t : \{v \in g\} \to \mathcal{T}$, then by condition G4, $h(g)$ has to be n. For any $v \in g$ with $d(v, g) \leq n - 1$, since $t(v) \in I$, the number of daughter subtrees of v must be less or equal to N, otherwise there would not be an $\omega \in \Omega$ with $\omega \sim g$ and $\mathrm{T}(\omega(v)) \in I$, contradicting to G4. Therefore, the set of all g with $(g, t) \in G$ for some t is finite. In addition, this set is independent of the selection of $J \supset I$.

Given $(g, t) \in G$,

$$E(g,t) \leq \prod_{\substack{v\ \text{non-} \\ \text{terminal}}} \frac{Q(t(v))}{\sum\limits_{\beta^*\sim t(v)} D^*(\beta^*)} \prod_{\substack{v\ \text{terminal} \\ d(v,g)<n}} Q(t(v)) \prod_{\substack{v\ \text{terminal} \\ d(v,g)=n}} Q(t(v)).$$

Let

$$R_0 = \max_{t \in I} \frac{Q(t)}{\sum_{\beta^* \sim t} D^*(\beta^*)},$$

and $R = \max\{R_0, M\}$ (recall $M = \max\{Q(\mathcal{T}), 1\}$). Since a non-terminal of g necessarily has depth less than n, then

$$E(g,t) \le R^{|g|} \prod_{\substack{v \text{ terminal} \\ d(v,g)=n}} Q(t(v)).$$

Because there are only finite number of g with $(g,t) \in G$ for some t, $|g|$ is bounded by a constant, say, A. So we get

$$\sum_{(g,t)\in G} E(g,t) \le R^A \sum_{(g,t)\in G} \prod_{\substack{v \text{ terminal} \\ d(v,g)=n}} Q(t(v)).$$

Notice that A is independent of the selection of J.

G is the union of disjoint sets G_α which have the following two properties,

1. For any (g,t), and $(g',t') \in G_\alpha$, $g = g'$, and for any $v \in g$ with $d(v,g) < n$, $t(v) = t'(v)$;
2. If $\alpha \ne \beta$, then for $(g,t) \in G_\alpha$ and $(g',t') \in G_\beta$, either $g \ne g'$ or there is a $v \in g$ with $d(v,g) < n$, such that $t(v) \ne t'(v)$.

It is easy to check that the number of G_α's is finite. In addition, the number is independent of the selection of J. Let the number be K_0. For any G_α, consider

$$\sum_{(g,t)\in G_\alpha} \prod_{\substack{v \text{ terminal} \\ d(v,g)=n}} Q(t(v)).$$

Since at least one of the $t(v)$ is not in J, then the sum is bounded $M^a - Q(J)^a \le M^{|g|} - Q(J)^{|g|} \le M^A - Q(J)^A$, where a is the number of $v \in g$ with $d(v,g) = n$. We then get

$$\sum_{(g,t)\in G} E(g,t) \le K_0 R^A (M^A - Q(J)^A).$$

Again, the bound is independent of the selection of $J \supset I$. Therefore, we can choose $J \supset I$ large enough to make the right hand side less than ϵ. This proves the lemma. □

Lemma 7. C is tight.

Proof. Fix $\epsilon > 0$. Then there is a finite set $I_1 \subset \mathcal{T}$ such that

$$\sum_{t \in I_1} Q(t) < \frac{\epsilon}{2}.$$

Define

$$H = \max_{t \in I_1} h(t).$$

By Lemma 6, there is a nested sequence of finite sets $I_2 \subset I_3 \ldots \subset I_H$ with $I_2 \supset I_1$, such that

$$\sum_{(g,t) \in G_k(I_{k-1}, I_k)} E(g,t) < \frac{\epsilon}{2H}, \qquad \text{for } 2 \leq k \leq H,$$

where $G_k(I_{k-1}, I_k)$ are defined as in Lemma 6.

Define $\tilde{S}_1 = \{\omega : \ \mathrm{T}(\omega) \notin I_1\}$. For $2 \leq k \leq H$, define $\tilde{S}_k$ as the set of ω which satisfy the following conditions

1. For any i, $1 \leq i < k$, for any $\omega' \subset \omega$ with $d(\omega', \omega) = i$, $\mathrm{T}(\omega) \in I_i$;
2. There is an $\omega' \subset \omega$ with $d(\omega', \omega) = k$ such that $\mathrm{T}(\omega') \notin I_k$.

Then $\tilde{S}_i$ are disjoint and

$$\bigcup_{i=1}^{H} \tilde{S}_i = \{\omega : \ \ \text{there is an } i, 1 \leq i \leq k, \text{ and } \omega' \subset \omega \text{ with } d(\omega', \omega) = i, \ \mathrm{T}(\omega') \notin I_i\}$$

Because I_k are increasing, for k, $2 \leq k \leq H$, $\tilde{S}_k \subset S_k$, where S_k is the set of ω satisfying

1. For any $\omega' \subset \omega$ with $d(\omega', \omega) < k$, $\mathrm{T}(\omega) \in I_{k-1}$;
2. There is an $\omega' \subset \omega$ with $d(\omega', \omega) = k$ such that $\mathrm{T}(\omega') \notin I_k$.

It is easy to see that for k, $2 \leq k \leq H$,

$$S_k = \bigcup_{(g,t) \in G_k(I_{k-1}, I_k)} \{\omega : \ \omega \sim g, \ \mathrm{T}(\omega(v)) = t(v), \text{ for any } v \in g\}.$$

Therefore, by (A3.10), for k, $2 \leq k \leq H$,

$$\sum_{\omega \in \tilde{S}_k} x(\omega) \leq \sum_{\omega \in S_k} x(\omega) \leq \sum_{(g,t) \in G_k} E(g,t) \leq \frac{\epsilon}{2H}.$$

We also have

$$\sum_{\omega \in \tilde{S}_1} x(\omega) = \sum_{\mathrm{T}(\omega) \notin I_1} x(\omega) \leq \frac{\epsilon}{2}.$$

Thus we get

$$x\left(\bigcup_{i=1}^{H} \tilde{S}_i\right) \leq \epsilon.$$

Because every ω with $\mathrm{T}(\omega) \in I_1$ has height less or equal to H, therefore if $\omega \in A$, where

$$A = \left(\bigcup_{i=1}^{H} \tilde{S}_i \right)^c = \{\omega : \text{ for any } i, 1 \leq i \leq H, \text{ and } \omega' \subset \omega, \text{ with } d(\omega', \omega) = i, \mathrm{T}(\omega') \in I_i\},$$

then $h(\omega) \leq H$, and for each $\omega' \subset \omega$, $\mathrm{T}(\omega') \in I_i \subset I_H$. Therefore, each label of $E(\omega)$ is in I_H. Since the correspondence between objects and their arrangements is one-to-one, then A is a finite set. Thus we get $x(A^c) < \epsilon$. This completes the proof that C is tight. □

Now we prove $F(C) \subset C$. For any $x \in C$, condition C1 is clearly satisfied. By (A3.3), (A3.11), and (A3.12), for any type $t = (l, b) \in \mathcal{T} \backslash T$, for any $\omega = l(\alpha^*) \in \Omega_t$,

$$(Fx)(\omega) = Q(t) \frac{x^*(\alpha^*)}{\sum_{\beta^* \sim t} x^*(\beta^*)} \geq Q(t) \frac{D^*(\alpha^*)}{m(t) M^{n(t)}} = D(\omega).$$

As for C3, if a tree graph g is of height 1, then for any t assigned to the single node in g,

$$\sum_{\substack{\omega \sim g, \forall v \in g \\ \mathrm{T}(\omega(v)) = t(v)}} (Fx)(\omega) = Q(t) = E(g, t).$$

The case where $h(g) \geq 2$ can then be proved following the proof of (A3.8).

The only thing that remains to show is the continuity of F. For this purpose, we shall use the following version of dominance convergence theorem without giving its proof.

Lemma 8. Let ν be a positive measure on a measurable space X. Suppose $\{f_n\}$, $\{g_n\}$ are sequences of measurable functions on X such that $|f_n| \leq g_n$, $\forall\ n \geq 1$, $f_n \to f$, ν-a.s. and $g_n \to g$, ν-a.s. If

$$\lim_{n \to \infty} \int g_n \, d\nu = \int g \, d\nu < \infty,$$

then

$$\lim_{n \to \infty} \int f_n \, d\nu = \int f \, d\nu.$$

□

Continuing the proof, suppose $x_n \to x$ in C, i.e., $\sum_{\omega \in \Omega} \|x_n(\omega) - x(\omega)\| \to 0$. Let $y_n = F(x_n)$ and $y = F(x)$. We want to show $\|y_n - y\| = \sum_{\omega \in \Omega} \|y_n(\omega) - y(\omega)\| \to 0$. The sum is dominated by $\sum_{\omega \in \Omega} g_n(\omega)$, where $g_n = y_n + y$.

Our plan is to show that for each ω, $y_n(\omega) \to y(\omega)$. Then $g_n(\omega) \to 2y(\omega)$. Since $\sum_{\omega \in \Omega} g_n(\omega) \equiv 2 \sum_{\omega \in \Omega} y(\omega) = 2M$, then by the above dominance convergence result, $\sum_{\omega \in \Omega} \|y_n(\omega) - y(\omega)\| \to 0$.

Now we show $y_n(\omega) \to y(\omega)$. Given $t = (l, b) \in \mathcal{T} \backslash T$, for any $\omega = l(\alpha^*) \in \Omega_t$,

$$\left| \sum_{B_l(\alpha^*) = b} x_n^*(\alpha^*) - \sum_{B_l(\alpha^*) = b} x^*(\alpha^*) \right| \leq \sum_{k=1}^{h(t)} \left| \sum_{\substack{B_l(\alpha^*) = b \\ |\alpha^*| = k}} x_n^*(\alpha^*) - \sum_{\substack{B_l(\alpha^*) = b \\ |\alpha^*| = k}} x^*(\alpha^*) \right|$$

For each k, $1 \le k \le h(t)$,

$$\begin{aligned} & \sum_{\substack{B_l(\alpha^*)=b \\ |\alpha^*|=k}} |x_n^*(\alpha^*) - x^*(\alpha^*)| \\ \le \quad & \sum_{\substack{B_l(\alpha^*)=b \\ |\alpha^*|=k}} \sum_{i=1}^{k} x^*(\alpha_1 \cdots \alpha_{i-1}) \, |x_n(\alpha_i) - x(\alpha_i)| \, x_n^*(\alpha_{i+1} \cdots \alpha_k) \\ \le \quad & kM^{k-1} \|x_n - x\| \to 0, \end{aligned}$$

leading to

$$\sum_{B_l(\alpha^*)=b} x_n^*(\alpha^*) \to \sum_{B_l(\alpha^*)=b} x^*(\alpha^*) > 0.$$

Therefore,

$$Q(t) \frac{x_n^*(\alpha^*)}{\displaystyle\sum_{B_{\mathrm{L}(\omega)}(\beta^*)=b} x_n(\beta^*)} \to Q(t) \frac{x^*(\alpha^*)}{\displaystyle\sum_{B_{\mathrm{L}(\omega)}(\beta^*)=b} x(\beta^*),}$$

i.e., $y_n(\omega) \to y(\omega)$, completing the proof. □