int.1 Craig's Interpolation Theorem

mod:int:prf: sec

mod:int:prf: thm:interpol **Theorem int.1** (Craig's Interpolation Theorem)**.** *If $\vDash \varphi \to \psi$, then there is a sentence χ such that $\vDash \varphi \to \chi$ and $\vDash \chi \to \psi$, and every constant symbol, function symbol, and predicate symbol (other than $=$) in χ occurs both in φ and ψ. The sentence χ is called an* interpolant *of φ and ψ.*

Proof. Suppose $\mathcal{L}_1$ is the language of φ and $\mathcal{L}_2$ is the language of ψ. Let $\mathcal{L}_0 = \mathcal{L}_1 \cap \mathcal{L}_2$. For each $i \in \{0, 1, 2\}$, let $\mathcal{L}'_i$ be obtained from $\mathcal{L}_i$ by adding the infinitely many new constant symbols $c_0, c_1, c_2, \dots$.

If φ is unsatisfiable, $\exists x\, x \neq x$ is an interpolant. If $\neg\psi$ is unsatisfiable (and hence ψ is valid), $\exists x\, x = x$ is an interpolant. So we may assume also that both φ and $\neg\psi$ are satisfiable.

In order to prove the contrapositive of the Interpolation Theorem, assume that there is no interpolant for φ and ψ. In other words, assume that $\{\varphi\}$ and $\{\neg\psi\}$ are inseparable in $\mathcal{L}_0$.

Our goal is to extend the pair $(\{\varphi\}, \{\neg\psi\})$ to a maximally inseparable pair (Γ^*, Δ^*). Let φ_0, φ_1, φ_2, …enumerate the sentences of $\mathcal{L}_1$, and ψ_0, ψ_1, ψ_2, …enumerate the sentences of $\mathcal{L}_2$. We define two increasing sequences of sets of sentences (Γ_n, Δ_n), for $n \geq 0$, as follows. Put $\Gamma_0 = \{\varphi\}$ and $\Delta_0 = \{\neg\psi\}$. Assuming (Γ_n, Δ_n) are already defined, define Γ_{n+1} and Δ_{n+1} by:

1. If $\Gamma_n \cup \{\varphi_n\}$ and Δ_n are inseparable in $\mathcal{L}'_0$, put φ_n in Γ_{n+1}. Moreover, if φ_n is an existential formula $\exists x\, \sigma$ then pick a new constant symbol c not occurring in Γ_n, Δ_n, φ_n or ψ_n, and put $\sigma[c/x]$ in Γ_{n+1}.

2. If Γ_{n+1} and $\Delta_n \cup \{\psi_n\}$ are inseparable in $\mathcal{L}'_0$, put ψ_n in Δ_{n+1}. Moreover, if ψ_n is an existential formula $\exists x\, \sigma$, then pick a new constant symbol c not occurring in Γ_{n+1}, Δ_n, φ_n or ψ_n, and put $\sigma[c/x]$ in Δ_{n+1}.

Finally, define:

$$\Gamma^* = \bigcup_{n \geq 0} \Gamma_n, \qquad \Delta^* = \bigcup_{n \geq 0} \Delta_n.$$

By simultaneous induction on n we can now prove:

mod:int:prf: part-a
1. Γ_n and Δ_n are inseparable in $\mathcal{L}'_0$;

mod:int:prf: part-b
2. Γ_{n+1} and Δ_n are inseparable in $\mathcal{L}'_0$.

The basis for (1) is given by **??**. For part (2), we need to distinguish three cases:

1. If $\Gamma_0 \cup \{\varphi_0\}$ and Δ_0 are separable, then $\Gamma_1 = \Gamma_0$ and (2) is just (1);

2. If $\Gamma_1 = \Gamma_0 \cup \{\varphi_0\}$, then Γ_1 and Δ_0 are inseparable by construction.

3. It remains to consider the case where φ_0 is existential, so that $\Gamma_1 = \Gamma_0 \cup \{\exists x\,\sigma, \sigma[c/x]\}$. By construction, $\Gamma_0 \cup \{\exists x\,\sigma\}$ and Δ_0 are inseparable, so that by **??** also $\Gamma_0 \cup \{\exists x\,\sigma, \sigma[c/x]\}$ and Δ_0 are inseparable.

This completes the basis of the induction for (1) and (2) above. Now for the inductive step. For (1), if $\Delta_{n+1} = \Delta_n \cup \{\psi_n\}$ then Γ_{n+1} and Δ_{n+1} are inseparable by construction (even when ψ_n is existential, by **??**); if $\Delta_{n+1} = \Delta_n$ (because Γ_{n+1} and $\Delta_n \cup \{\psi_n\}$ are separable), then we use the induction hypothesis on (2). For the inductive step for (2), if $\Gamma_{n+2} = \Gamma_{n+1} \cup \{\varphi_{n+1}\}$ then Γ_{n+2} and Δ_{n+1} are inseparable by construction (even when φ_{n+1} is existential, by **??**); and if $\Gamma_{n+2} = \Gamma_{n+1}$ then we use the inductive case for (1) just proved. This concludes the induction on (1) and (2).

It follows that Γ^* and Δ^* are inseparable; if not, by compactness, there is $n \geq 0$ that separates Γ_n and Δ_n, against (1). In particular, Γ^* and Δ^* are consistent: for if the former or the latter is inconsistent, then they are separated by $\exists x\, x \neq x$ or $\forall x\, x = x$, respectively.

We now show that Γ^* is maximally consistent in $\mathcal{L}_1'$ and likewise Δ^* in $\mathcal{L}_2'$. For the former, suppose that $\varphi_n \notin \Gamma^*$ and $\neg\varphi_n \notin \Gamma^*$, for some $n \geq 0$. If $\varphi_n \notin \Gamma^*$ then $\Gamma_n \cup \{\varphi_n\}$ is separable from Δ_n, and so there is $\chi \in \mathcal{L}_0'$ such that both:

$$\Gamma^* \vDash \varphi_n \rightarrow \chi, \qquad\qquad \Delta^* \vDash \neg\chi.$$

Likewise, if $\neg\varphi_n \notin \Gamma^*$, there is $\chi' \in \mathcal{L}_0'$ such that both:

$$\Gamma^* \vDash \neg\varphi_n \rightarrow \chi', \qquad\qquad \Delta^* \vDash \neg\chi'.$$

By propositional logic, $\Gamma^* \vDash \chi \vee \chi'$ and $\Delta^* \vDash \neg(\chi \vee \chi')$, so $\chi \vee \chi'$ separates $\Gamma^*$and Δ^*. A similar argument establishes that Δ^* is maximal.

Finally, we show that $\Gamma^* \cap \Delta^*$ is maximally consistent in $\mathcal{L}_0'$. It is obviously consistent, since it is the intersection of consistent sets. To show maximality, let $\sigma \in \mathcal{L}_0'$. Now, Γ^* is maximal in $\mathcal{L}_1' \supseteq \mathcal{L}_0'$, and similarly Δ^* is maximal in $\mathcal{L}_2' \supseteq \mathcal{L}_0'$. It follows that either $\sigma \in \Gamma^*$ or $\neg\sigma \in \Gamma^*$, and either $\sigma \in \Delta^*$ or $\neg\sigma \in \Delta^*$. If $\sigma \in \Gamma^*$ and $\neg\sigma \in \Delta^*$ then σ would separate Γ^* and Δ^*; and if $\neg\sigma \in \Gamma^*$ and $\sigma \in \Delta^*$ then Γ^* and Δ^* would be separated by $\neg\sigma$. Hence, either $\sigma \in \Gamma^* \cap \Delta^*$ or $\neg\sigma \in \Gamma^* \cap \Delta^*$, and $\Gamma^* \cap \Delta^*$ is maximal.

Since Γ^* is maximally consistent, it has a model $\mathfrak{M}_1'$ whose domain $|\mathfrak{M}_1'|$ comprises all and only the elements $c^{\mathfrak{M}_1'}$ interpreting the constant symbols—just like in the proof of the completeness theorem (**??**). Similarly, Δ^* has a model $\mathfrak{M}_2'$ whose domain $|\mathfrak{M}_2'|$ is given by the interpretations $c^{\mathfrak{M}_2'}$ of the constant symbols.

Let $\mathfrak{M}_1$ be obtained from $\mathfrak{M}_1'$ by dropping interpretations for constant symbols, function symbols, and predicate symbols in $\mathcal{L}_1' \setminus \mathcal{L}_0'$, and similarly for $\mathfrak{M}_2$. Then the map $h\colon M_1 \rightarrow M_2$ defined by $h(c^{\mathfrak{M}_1'}) = c^{\mathfrak{M}_2'}$ is an isomorphism in $\mathcal{L}_0'$, because $\Gamma^* \cap \Delta^*$ is maximally consistent in $\mathcal{L}_0'$, as shown. This follows because any $\mathcal{L}_0'$-sentence either belongs to both Γ^* and Δ^*, or to neither: so $c^{\mathfrak{M}_1'} \in P^{\mathfrak{M}_1'}$ if and only if $P(c) \in \Gamma^*$ if and only if $P(c) \in \Delta^*$ if and only if

$c^{\mathfrak{M}'_2} \in P^{\mathfrak{M}'_2}$. The other conditions satisfied by isomorphisms can be established similarly.

Let us now define a model $\mathfrak{M}$ for the language $\mathcal{L}_1 \cup \mathcal{L}_2$ as follows:

1. The domain $|\mathfrak{M}|$ is just $|\mathfrak{M}_2|$, i.e., the set of all elements $c^{\mathfrak{M}'_2}$;
2. If a predicate symbol P is in $\mathcal{L}_2 \setminus \mathcal{L}_1$ then $P^{\mathfrak{M}} = P^{\mathfrak{M}'_2}$;
3. If a predicate P is in $\mathcal{L}_1 \setminus \mathcal{L}_2$ then $P^{\mathfrak{M}} = h(P^{\mathfrak{M}'_2})$, i.e., $\langle c_1^{\mathfrak{M}'_2}, \dots, c_n^{\mathfrak{M}'_2} \rangle \in P^{\mathfrak{M}}$ if and only if $\langle c_1^{\mathfrak{M}'_1}, \dots, c_n^{\mathfrak{M}'_1} \rangle \in P^{\mathfrak{M}'_1}$.
4. If a predicate symbol P is in $\mathcal{L}_0$ then $P^{\mathfrak{M}} = P^{\mathfrak{M}'_2} = h(P^{\mathfrak{M}'_1})$.
5. Function symbols of $\mathcal{L}_1 \cup \mathcal{L}_2$, including constant symbols, are handled similarly.

Finally, one shows by induction on formulas that $\mathfrak{M}$ agrees with $\mathfrak{M}'_1$ on all formulas of $\mathcal{L}'_1$ and with $\mathfrak{M}'_2$ on all formulas of $\mathcal{L}'_2$. In particular, $\mathfrak{M} \vDash \Gamma^* \cup \Delta^*$, whence $\mathfrak{M} \vDash \varphi$ and $\mathfrak{M} \vDash \neg\psi$, and $\nvDash \varphi \to \psi$. This concludes the proof of Craig's Interpolation Theorem. $\square$

Photo Credits

Bibliography