Polimorficzne osłabianie bez aksjomatu rozdzielności

Jacek Chrząszcz

Streszczenie

Relacja osłabiania polimorficznego w rachunku λ drugiego rzędu (system **F**), wprowadzona przez J. C. Mitchella, zdefiniowana jest przez cztery aksjomaty: zwrotności, instancjacji, generalizacji i rozdzielności oraz reguły przechodniości i domknięcia przez kontekst. Wiadomo, że relacja ta jest nierozstrzygalna, jednak wszystkie znane dotychczas dowody tego faktu istotnie wykorzystywały aksjomat rozdzielności.

Praca niniejsza pokazuje nierozstrzygalność relacji osłabiania polimorficznego, gdy usuniemy z jej definicji ten aksjomat. Ponadto przedstawiona jest również pełna równościowa charakteryzacja relacji równoważności indukowanej przez nową relację osłabiania.

Praca magisterska z informatyki

Wydział Matematyki, Informatyki i Mechaniki
Uniwersystet Warszawski

Promotor: Paweł Urzyczyn

Warszawa, 10 września 1997

Podziękowania. Gorące podziękowania należą się mojemu promotorowi, Pawłowi Urzyczynowi, bez pomocy którego praca niniejsza nie mogłaby powstać. Jego słowa mobilizujące do pracy oraz cenne rady dochodziły do mnie nawet gdy dzieliły nas tysiące kilometrów. Pragnę również podziękować wszystkim członkom Zakładu Logiki Stosowanej za przyjacielską atmosferę i za to, że zawsze mogłem na nich liczyć. Wielką i bezcenną pomocą było również ciepło, którym otoczyli mnie moi najbliżsi. Dziękuję.

Spis rzeczy

1 Wprowadzenie **2**

2 Notacja **6**

3 Relacja osłabiania **7**

4 Wyprowadzanie nierówności przez przepisywanie **7**

5 Wyprowadzenia normalne **9**

6 Typy bikoercyjne **12**

7 Derywacje uporządkowane **17**

8 Wymuszanie nierówności **19**

9 Nierozstrzygalność relacji osłabiania **21**

10 Podsumowanie **24**

1 Wprowadzenie

Polimorficzne typy danych i polimorficzne procedury pojawiają się coraz częściej w dzisiejszej praktyce programowania. Zwykle jest to polimorfizm związany z programowaniem obiektowym, ale można znaleźć również konstrukcje do tworzenia typów i funkcji polimorficznych w sensie parametrycznym. Przykładem może być choćby znana z języka C++ konstrukcja `template`<*par*>*Typ*.

Oczywiście polimorfizm parametryczny występuje również w językach funkcyjnych, takich jak ML czy Haskell. Można znaleźć tam polimorficzne typy predefiniowane jak np. listy czy pary oraz konstrukcje do definiowania własnych algebraicznych typów z parametrami. Parametry w typach oznaczane są zmiennymi, a dla podkreślenia, że mogą one przyjmować dowolne wartości typowe, typy polimorficzne poprzedzane są jeszcze kwantyfikatorami. Wspomniany już polimorficzny typ listowy oznaczany jest $\forall\alpha\,(\alpha\,\texttt{list})$.

Jednak w systemie typów, używanym przez znane języki funkcyjne, mamy do czynienia z tzw. *płytkim polimorfizmem*. Oznacza to, że kwantyfikatory mogą występować w typach polimorficznych wyłącznie na najwyższym poziomie. Istnieją wprawdzie pewne eksperymentalne języki programowania, jak np. Quest autorstwa Giuseppe Longo i Luki Cardellego [CL91], które wyposażone są w polimorfizm wyższego rzędu, ale wszystkie popularne języki funkcyjne ograniczają się do płytkiej wersji polimorfizmu. Inaczej jest w odmianie rachunku λ, zwanej systemem **F** (zob. np. [Gir89]), która uważana jest za model języka programowania z polimorficznymi typami wyższego rzędu. Typy są tu dowolnymi formułami logicznymi zbudowanymi ze zmiennych, symboli $\to$ oraz kwantyfikatorów.

System **F** ma jeszcze jedną ważną własność. W systemie tym, pomimo braku stałych typowych, można wyrazić wszystkie typy danych, a czyste λ-termy (również bez dodatkowych stałych) wyrażają niezbędne konstruktory. Na przykład wspomniany już typ listowy można zakodować w następujący sposób:

$$\tau\,\texttt{list} \;=\; \forall\gamma\,(\gamma \to (\tau \to \gamma \to \gamma) \to \gamma)$$

Konstruktory typu listowego: lista pusta $\texttt{nil} : \tau\,\texttt{list}$ oraz głowa listy $\texttt{cons} : \tau \to \tau\,\texttt{list} \to \tau\,\texttt{list}$ reprezentowane są przez następujące termy:

$$\begin{array}{rcl} \texttt{nil} & = & \Lambda\gamma.\lambda n^{\gamma}.\lambda c^{\tau\to\gamma\to\gamma}.n \\ \texttt{cons}(h,t) & = & \Lambda\gamma.\lambda n^{\gamma}.\lambda c^{\tau\to\gamma\to\gamma}.c(h, t\gamma(n,c)) \end{array}$$

W tak rozbudowanym systemie narzuca się pytanie, czy termy danego typu σ nie mogą zastępować termow innego typu τ. Jeśli tak jest, mówimy, że typ σ jest *bardziej ogólny* od typu τ, lub że typ τ jest *słabszy* od typu σ. Oznaczamy to przez $\sigma \sqsubseteq \tau$. W klasycznym systemie **F** relację tę scharakteryzować można następującym aksjomatem:

$$\forall\alpha_1\ldots\alpha_n\,\sigma \;\sqsubseteq_{\mathbf{F}}\; \forall\beta_1\ldots\beta_m\,\sigma\{\alpha_1\mapsto\rho_1,\ldots,\alpha_n\mapsto\rho_n\} \tag{1}$$

gdzie $\beta_1,\ldots,\beta_m$ nie są wolne w $\forall\alpha_1\ldots\alpha_n\,\sigma$, a $\rho_1,\ldots,\rho_n$ są dowolnymi typami. Jednak relacji tej brakuje reguł kompozycjonalności. Nawet wiedząc, że $\sigma \sqsubseteq_{\mathbf{F}} \sigma'$, nie można udowodnić, że $\tau \to \sigma \sqsubseteq_{\mathbf{F}} \tau \to \sigma'$. John C. Mitchell w [Mit88] rozszerzył relację osłabiania do relacji $\sqsubseteq_{\mathrm{M}}$, której definicja, oprócz aksjomatów podobnych do (1) zawiera również aksjomat rozdzielności kwantyfikatora oraz reguły domykania przez kontekst, z których najważniejszą jest reguła **(→-kontekst)** antymonotoniczna ze względu na pierwszy argument i monotoniczna ze względu na drugi, tzn.

$$\sigma_1 \to \tau_1 \sqsubseteq_{\mathrm{M}} \sigma_2 \to \tau_2 \qquad \text{jeśli} \qquad \sigma_2 \sqsubseteq_{\mathrm{M}} \sigma_1 \quad \text{oraz} \quad \tau_1 \sqsubseteq_{\mathrm{M}} \tau_2$$

Praca Mitchella zawiera również dowód pełności relacji $\sqsubseteq_{\mathrm{M}}$ względem semantyki osłabiania. Nierówność $\sigma \sqsubseteq_{\mathrm{M}} \tau$ da się wyprowadzić wtedy i tylko wtedy, gdy we wszystkich *prostych modelach inferencji*[1] zachodzi $D_{[\![\sigma]\!]} \subseteq D_{[\![\tau]\!]}$, gdzie $D_{[\![\rho]\!]}$ jest zbiorem znaczeń wszystkich λ-termów typu ρ.

System **F**, wzbogacony o regułę osłabiania używającą nowej relacji $\sqsubseteq_{\mathrm{M}}$

$$\frac{E \vdash M : \sigma \quad \sigma \sqsubseteq_{\mathrm{M}} \tau}{E \vdash M : \tau}$$

jest właściwym rozszerzeniem klasycznego systemu **F**. Istnieją nawet termy typowalne w nowym systemie, które nie są typowalne w systemie klasycznym. To samo rozszerzenie, jak dowodzi Mitchell, uzyskać można dodając do klasycznego systemu **F** regułę wnioskowania

$$\textbf{(eta)} \qquad \frac{E \vdash M : \sigma \quad M \longrightarrow_{\eta} N}{E \vdash N : \sigma}$$

opartą o regułę redukcji $\lambda x.Mx \longrightarrow_{\eta} M$ jeśli x nie występuje wolno w M.

Równoważność tych dwóch rozszerzeń wynika z faktu, że jeśli $\sigma \sqsubseteq_{\mathrm{M}} \tau$, to istnieje funkcja zmiany typu $V : \sigma \to \tau$, zwana koercją, taka że V jest η-redukowalne do identyczności, po wytarciu informacji typowych.

Główną wadą rozszerzonego systemu $\mathbf{F}_{\eta}$, jak również klasycznego systemu **F** jest nierozstrzygalność większości ważnych problemów dotyczących tych systemów. I tak, dla systemu **F**, nierozstrzygalność problemów sprawdzania i rekonstrukcji typów została udowodniona przez Joe Wellsa w pracy [Wel94]. Pewną nadzieją dla systemu $\mathbf{F}_{\eta}$ była praca Trevora Jima [Jim95], który zbudował redukcję rekonstrukcji typów dla systemu $\mathbf{F}_{\eta}$ do nierówności $\sqsubseteq_{\mathrm{M}}$. Niestety praca Jerzego Tiuryna i Pawła Urzyczyna [TU95] pokazuje nierozstrzygalność relacji $\sqsubseteq_{\mathrm{M}}$, co implikuje również nierozstrzygalność sprawdzania typów. Na koniec Wells udowodnił w [Wel96], że rekonstrukcja typów również jest nierozstrzygalna.

Wspomniana praca Tiuryna i Urzyczyna [TU95] zawiera redukcję problemu stopu maszyny dwulicznikowej do nierówności $\sqsubseteq_{\mathrm{M}}$. Jako element pośredni, autorzy definiują *maszynę rejestrów stosowych*. Jest to deterministyczna maszyna z jednym stanem i skończoną liczbą rejestrów, która działając zdejmuje instrukcje z pierwszego rejestru stosowego i wykonuje je modyfikując odpowiednio zawartość wszystkich rejestrów. Zatrzymuje się po przeczytaniu wyróżnionej instrukcji końcowej. Problem stopu tej maszyny jest nierozstrzygalny, gdyż może ona symulować dowolną deterministyczną maszynę dwulicznikową.

[1] ang. *simple inference model*

Autorzy prezentują kodowanie chwilowej konfiguracji maszyny w postaci dwóch typów σ i τ w taki sposób, że nierówność $\sigma \sqsubseteq_{\mathrm{M}} \tau$ zachodzi wtedy i tylko wtedy, gdy konfiguracja ta prowadzi do zatrzymania maszyny. Dowód powyższego opiera się w ich pracy na pojęciu wagi nierówności zdefiniowanej na podstawie alternatywnego systemu wyprowadzania nierówności $\sqsubseteq_{\mathrm{M}}$, zaproponowanego przez Longo, Milsted i Sołowiewa w pracy [LMS95]. Dzięki udowodnionym przez nich pewnym własnościom relacji $\sqsubseteq_{\mathrm{M}}$ oraz wyprowadzeń w systemie Longo-Milsted-Sołowiew, autorzy pokazują następujący lemat:

> Niech σ_1 i τ_1 kodują daną chwilową konfigurację maszyny. Jeśli konfiguracja ta doprowadza do zatrzymania w jednym kroku, to trywialnie $\sigma_1 \sqsubseteq_{\mathrm{M}} \tau_1$. W przeciwnym wypadku łatwo ustalić parę typów σ_2 i τ_2, kodującą konfigurację po jednym kroku i to w taki sposób, że nierówność $\sigma_1 \sqsubseteq_{\mathrm{M}} \tau_1$ zachodzi wtedy i tylko wtedy, gdy $\sigma_2 \sqsubseteq_{\mathrm{M}} \tau_2$. Ponadto, jeśli obie nierówności zachodzą, waga pierwszej jest ostro większa od wagi drugiej.

Z lematu tego wynika, że jeśli nierówność kodująca bieżącą konfigurację chwilową maszyny zachodzi, to ponieważ za każdym krokiem maszyny waga nierówności kodującej maleje, więc maszyna nie może działać w nieskończoność. W drugą stronę, jeśli maszyna zatrzymuje się, to skoro nierówność kodująca jest prawdziwa dla konfiguracji końcowej, to również dla wszystkich ją poprzedzających.

Istota tego dowodu nierozstrzygalności polega na tym, iż nie da się ograniczyć stopnia *zagnieżdżenia* instancjacji niezbędnych do udowodnienia nierówności kodujących konfiguracje prowadzące do zatrzymania maszyny. System Longo-Milsted-Sołowiew zawiera regułę **(∀-lewy)**, która mówi, że aby udowodnić nierówność $\forall\alpha\,\sigma \sqsubseteq_{\mathrm{M}} \tau$, wystarczy najpierw pokazać $\sigma\{\alpha \mapsto \xi\} \sqsubseteq_{\mathrm{M}} \tau$ dla jakiegoś typu ξ.

Każdy krok maszyny, który intuicyjnie odpowiada pewnemu fragmentowi dowodu nierówności kodującej pierwotną konfigurację, powoduje konieczność użycia tej reguły. Ponadto przy kolejnym kroku maszyny, reguła ta będzie używana jak gdyby *wewnątrz* wstawionego typu ξ.

Inny dowód nierozstrzygalności relacji $\sqsubseteq_{\mathrm{M}}$ zawiera praca Wellsa [Wel95]. Tutaj niemożność decyzji, czy dane dwa typy są w relacji $\sqsubseteq_{\mathrm{M}}$, wynika nie z faktu, że nie da się przewidzieć koniecznego stopnia zagnieżdżenia instancjacji, lecz z tego, że nie da się oszacować rozmiarów typów wstawianych (oznaczanych powyżej przez ξ). Wells używa problemu semi-unifikacji, który jest następujący:

> Mając dane dwie pary termów σ_1, τ_1 i σ_2, τ_2, zbudowanych ze zmiennych i symboli $\to$ rozstrzygnąć, czy istnieje podstawienie θ takie, że dla pewnych podstawień θ_1 i θ_2 zachodzą równości $\sigma_1\theta\theta_1 = \tau_1\theta$ oraz $\sigma_2\theta\theta_2 = \tau_2\theta$.

Nierozstrzygalność tego problemu również wynika z nieprzewidywalności rozmiaru pierwszego podstawienia θ.

Redukcja polega na uniwersalnej konstrukcji typu $\Pi(\delta_1, \zeta_1, \delta_2, \zeta_2)$ takiego, że jeśli $\vec{\gamma}$ są wszystkimi zmiennymi występującymi w $\sigma_1, \tau_1,\ \sigma_2, \tau_2$, to nierówność

$$\forall\beta\,(\Pi(\beta,\beta,\beta,\beta) \to \bot)\ \sqsubseteq_{\mathrm{M}}\ (\forall\vec{\gamma}\,\Pi(\sigma_1,\tau_1,\sigma_2,\tau_2)) \to \bot \tag{2}$$

zachodzi wtedy i tylko wtedy, gdy czwórka $\sigma_1, \tau_1,\ \sigma_2, \tau_2$ jest semi-unifikowalna.

Istota równoważności polega na tym, że wspólne podstawienie θ wymuszone jest przez wspólne kwantyfikowanie zmiennych wolnych $\vec{\gamma}$ we wszystkich czterech typach $\sigma_1, \tau_1,\ \sigma_2, \tau_2$. Następnie, dzięki aksjomatowi rozdzielności kwantyfkatora $\forall\alpha\,(\rho_1 \to \rho_2) \sqsubseteq_{\mathrm{M}} \forall\alpha\,\rho_1 \to \forall\alpha\,\rho_2$, podstawienia θ_1 i θ_2, odpowiednio w $\sigma_1\theta$ i w $\sigma_2\theta$, mogą być wykonane niezależnie. Dzięki odpowiedniemu ustawieniu typów $\sigma_1, \tau_1,\ \sigma_2, \tau_2$ w $\Pi(\sigma_1, \tau_1, \sigma_2, \tau_2)$ oraz dzięki nowemu systemowi wyprowadzania nierówności $\sqsubseteq_{\mathrm{M}}$, wyrażenie (2) może być wyprowadzone tylko na jeden sposób.

Pomimo nierozstrzygalności relacji $\sqsubseteq_{\mathrm{M}}$, w językach posiadających polimorficzne typy wyższego rzędu zachodzi potrzeba odpowiadania na pytanie, czy dany term U, posiadający typ σ może być użyty w miejscu, gdzie wymagany jest typ τ. Narzuca się więc problem znalezienia nietrywialnej rozstrzygalnej relacji osłabiania $\sqsubseteq$, zawartej w oryginalnej relacji osłabiania Mitchella. Pierwszą nieudaną próbą jest usunięcie z definicji $\sqsubseteq_{\mathrm{M}}$ najmniej intuicyjnie jasnego aksjomatu rozdzielności.

W kodowaniu semi-unifikacji Wellsa aksjomat ten jest kluczowy. Bez niego nie można, po wspólnym podstawieniu θ, dokonać dwóch niezależnych podstawień θ_1 i θ_2. W pracy Tiuryna i Urzyczyna natomiast, aksjomat rozdzielności nie wydaje się niezbędnym w kodowaniu obliczeń maszyny, ale ponieważ jest on ukryty w systemie Longo-Milsted-Sołowiew, bez tego aksjomatu, dowód odpowiednich własności nowej relacji $\sqsubseteq$ należy przeprowadzić inaczej.

Dlatego niniejsza praca, oparta o kodowanie problemu stopu maszyny rejestrów stosowych z pracy [TU95], definiuje pojęcie wagi nierówności w oparciu o pewną klasę wyprowadzeń w systemie przepisywania typów, łatwo wynikającym z oryginalnych aksjomatów Mitchella. Pozostałe własności potrzebne w kodowaniu obliczeń maszyny, udowodnione są przy pomocy jeszcze węższej podklasy wyprowadzeń.

Po niniejszym wstępie, rozdział 2 wprowadza formalnie podstawowe pojęcia i oznaczenia, z których będziemy korzystać. Kolejne dwa rozdziały prezentują oryginalny system Mitchella, który po usunięciu aksjomatu rozdzielności, definiuje rozpatrywaną relację osłabiania $\sqsubseteq$ oraz blisko z nim związany system przepisywania typów.

W rozdziale 5 określony jest warunek, jaki muszą spełniać *wyprowadzenia normalne* w tym systemie, a następnie lemat 5.4 pokazuje, że klasa ta wystarcza do dowodzenia nierówności typowych i dlatego w oparciu o nią może być zdefiniowane pojęcie wagi nierówności.

Rozdział 6 poświęcony jest relacji równoważności związanej z quasi-porządkiem $\sqsubseteq$. Dowiedzione są wszystkie potrzebne w dalszym ciągu pracy własności tej relacji oraz dodatkowo podana jest jej pełna równościowa charakteryzacja, na wzór podanej przez Tiuryna w [Tiu95] charakteryzacji relacji równoważności $\sim_{\mathrm{M}}$ związanej z oryginalną relacją Mitchella.

W rozdziale 7 ze wszystkich normalnych wyprowadzeń wydzielone są wyprowadzenia uporządkowane, podklasa potrzebna w rozdziale 8 do dowiedzenia własności relacji $\sqsubseteq$ bezpośrednio wykorzystywanych w kodowaniu maszyny rejestrów stosowych. Szkic tego kodowania, na podstawie pracy [TU95], podany jest w rozdziale 9, po którym następuje podsumowanie.

2 Notacja

Będziemy pracować z typami polimorficznego rachunku λ (system $\mathbf{F}$), zdefiniowanymi na podstawie przeliczalnego zbioru zmiennych V przez następującą gramatykę:

$$\mathcal{T} ::= V \mid \forall V\, \mathcal{T} \mid \mathcal{T} \to \mathcal{T}$$

Zmienne występujące w typach oznaczane będą literami z początku alfabetu greckiego α, β, γ, ciągi zmiennych kwantyfikowanych $\vec{\delta}$, $\vec{\zeta}$, $\vec{\vartheta}$, a typy innymi literami greckimi, jak σ, τ, ρ... Często typ $\forall\alpha\forall\beta\,\tau$ zapisywany będzie w skrócie jako $\forall\alpha\beta\,\tau$. Przyjmujemy, że kwantyfikatory wiążą silniej niż operator $\to$ oraz że symbol $\perp$ oznacza typ $\forall\alpha\,\alpha$.

Liczbę symboli $\to$ występującą w danym typie nazwiemy jego *wagą*. Definicja indukcyjna jest nastepująca:

$$\begin{aligned}
waga(\alpha) &= 0\\
waga(\forall\alpha\,\sigma) &= waga(\sigma)\\
waga(\sigma_1 \to \sigma_2) &= waga(\sigma_1) + waga(\sigma_2) + 1
\end{aligned}$$

Ze względu na obecność kwantyfikatorów, zmienne występujące w danym typie dzielimy na zmienne związane oraz zmienne wolne, czyli nie związane przez żaden kwantyfikator. Dla danego typu σ, zbiór zmiennych wolnych występujących w tym typie oznaczany będzie przez $FV(\sigma)$. Formalnie definiujemy:

$$\begin{aligned}
FV(\alpha) &= \{\alpha\}\\
FV(\sigma_1 \to \sigma_2) &= FV(\sigma_1) \cup FV(\sigma_2)\\
FV(\forall\beta\,\sigma) &= FV(\sigma) \setminus \{\beta\}
\end{aligned}$$

Typy różniące się jedynie nazwami zmiennych związanych będziemy uznawać za równe. Będziemy również przyjmować, że w każdym typie zmienne wolne mają inne nazwy, niż zmienne związane oraz że nazwy różnych zmiennych związanych nie pokrywają się.

Funkcję częściową o skończonej dziedzinie ze zbioru zmiennych V w zbiór typów $\mathcal{T}$ nazywamy *podstawieniem*. Dowolne podstawienie θ przedłuża się w naturalny sposób do funkcji ze zbioru $\mathcal{T}$ w zbiór $\mathcal{T}$. Wynikiem aplikacji podstawienia θ do typu σ, co będziemy oznaczać postfiksowo $\sigma\theta$ będzie typ σ zmodyfikowany w taki sposób, że zamiast każdego wystąpienia zmiennej wolnej α należącej do dziedziny θ wystąpi wartość θ na tej zmiennej. Pozostałe zmienne wolne w σ oraz zmienne związane pozostają bez zmian. Ponadto dokonując niezbędnych przemianowań zmiennych związanych, gwarantujemy, że zmienne wolne w typach $\alpha\theta$ pozostają wolne w typie $\sigma\theta$. Podstawienie θ o dziedzinie $\alpha_1, \ldots, \alpha_n$, takie że $\alpha_i\theta = \tau_i$ dla każdego i, będzie również oznaczane $\{\alpha_1 \mapsto \tau_1, \ldots, \alpha_n \mapsto \tau_n\}$. W zapisie operacja podstawienia wiązać będzie silniej niż kwantyfikatory i dlatego w typie $\forall\beta\,\sigma\{\alpha \mapsto \xi\}$ wolne wystąpienia zmiennej β w typie ξ znajdą się w zasięgu kwantyfikatora $\forall\beta$.

Typy polimorficzne możemy również utożsamiać z drzewami etykietowanymi. W liściach znajdują się zmienne, w węzłach wewnętrznych symbole $\to$, oraz każdy wierzchołek jest etykietowany skończonym ciągiem kwantyfikatorów.

Zbiór pozycji w takim drzewie możemy zatem utorzsamiać ze skończonym zbiorem skończonych łańcuchów nad alfabetem $\{0, 1\}$. Pozycje, które mają parzystą liczbę zer nazywane są pozytywnymi, a te które maja nieparzystą liczbę zer, to pozycje negatywne. Łańcuch pusty będziemy oznaczali przez ϵ, a mając dane dwa łańcuchy p_1 i p_2 ich konkatenacja oznaczana będzie $p_1 p_2$.

Pozycje będziemy porównywać między sobą przy pomocy dwóch różnych częściowych porządków. Porządek prefiksowy będzie oznaczany $\leq$, a zapis $p \not\gtrless r$ będzie oznaczać, że p i r są nieporównywalne. Drugim używanym porządkiem będzie porządek leksykograficzny, oznaczany $\preceq$, indukowany przez $0 \prec 1$.

Jeśli σ jest typem, a p pozycją w drzewie σ, przez $\sigma|_p$ oznaczamy poddrzewo na pozycji p. Jeśli σ i ϱ są typami oraz p jest pozycją w drzewie σ, to $\sigma[\varrho]_p$ oznacza typ σ z poddrzewem na pozycji p zastąpionym przez ϱ. W odróżnieniu od podstawienia, przy tej operacji niektóre zmienne wolne typu ϱ mogą znaleźć się w zasięgu kwantyfikatorów z σ.

3 Relacja osłabiania

Relacja osłabiania $\sqsubseteq$, której będziemy używać, zdefiniowana jest przez 3 aksjomaty i 3 reguły:

$$\begin{array}{rl} \textbf{(zwr)} & \sigma \sqsubseteq \sigma \\ \textbf{(inst)} & \forall\alpha\,\sigma \sqsubseteq \sigma\{\alpha \mapsto \rho\} \\ \textbf{(kwant)} & \sigma \sqsubseteq \forall\alpha\,\sigma \quad \text{jeśli } \alpha \notin FV(\sigma) \end{array}$$

Reguły:

$$\textbf{(}\rightarrow\textbf{-kontekst)} \quad \frac{\sigma_2 \sqsubseteq \sigma_1 \quad \tau_1 \sqsubseteq \tau_2}{\sigma_1 \to \tau_1 \sqsubseteq \sigma_2 \to \tau_2}$$

$$\textbf{(}\forall\textbf{-kontekst)} \quad \frac{\sigma \sqsubseteq \tau}{\forall\alpha\,\sigma \sqsubseteq \forall\alpha\,\tau}$$

$$\textbf{(przech)} \quad \frac{\sigma \sqsubseteq \rho \quad \rho \sqsubseteq \tau}{\sigma \sqsubseteq \tau}$$

W oryginalnej relacji osłabiania na typach polimorficznych, wprowadzonej przez Mitchella w [Mit88] był jeszcze aksjomat

$$\textbf{(rozdz)} \quad \forall\alpha(\sigma \to \tau) \sqsubseteq_{\mathrm{M}} \sigma \to \forall\alpha\,\tau \quad \text{jeśli } \alpha \notin FV(\sigma)$$

4 Wyprowadzanie nierówności przez przepisywanie

Z przyczyn technicznych zamiast wyprowadzeń w powyższym systemie wygodniej nam będzie używać przepisywania termów zgodnie z podanymi niżej regułami:

$$
\begin{array}{rcl@{\qquad}rcl}
C^+[\forall\alpha\,\sigma] & \Longrightarrow_i^+ & C^+[\sigma\{\alpha\mapsto\rho\}] & C^-[\sigma\{\alpha\mapsto\rho\}] & \Longrightarrow_i^- & C^-[\forall\alpha\,\sigma] \\
C^+[\sigma] & \Longrightarrow_q^+ & C^+[\forall\alpha\,\sigma] & C^-[\forall\alpha\,\sigma] & \Longrightarrow_q^- & C^-[\sigma] \\
 & & \text{jeśli } \alpha\notin FV(\sigma) & & & \text{jeśli } \alpha\notin FV(\sigma)
\end{array}
$$

gdzie $C^+[\,]$ i $C^-[\,]$ oznaczają dowolny kontekst z dziurą na pozycji dodatniej i, odpowiednio, ujemnej. Jeśli pozycja ta jest istotna, będziemy ją umieszczać nad strzałką, np. $\sigma \overset{p}{\Longrightarrow}_q^+ \tau$ i nazywać pozycją danego kroku. Oznaczmy teraz dowolny krok dodatni przez $\Longrightarrow^+$ oraz dowolny krok ujemny przez $\Longrightarrow^-$. Niech $\Longrightarrow$ oznacza dowolny krok przepisywania oraz $\Longrightarrow\!\!\!\!\!\!\!\Longrightarrow$ dowolną skończoną liczbę kroków. *Wagą* kroku $\sigma \Longrightarrow \tau$ nazwiemy wartość bezwzględną z różnicy wag σ i τ, czyli intuicyjnie liczbę symboli $\to$ wprowadzonych przez ten krok.

Znaczenie wprowadzonego systemu przepisywania wyraża formalnie następujący prosty lemat:

Lemat 4.1 *Dla dowolnych dwóch typów σ i τ zachodzi:*

$$\sigma \sqsubseteq \tau \quad \textit{wtedy i tylko wtedy, gdy} \quad \sigma \Longrightarrow\!\!\!\!\!\!\!\Longrightarrow \tau$$

Dowód. Załóżmy, że $\sigma \sqsubseteq \tau$. Udowodnimy przez indukcję ze względu na budowę dowodu tego faktu, że $\sigma \Longrightarrow\!\!\!\!\!\!\!\Longrightarrow \tau$. Przypadek bazowy, to zastosowanie aksjomatu. Jeśli był to aksjomat **(zwr)**, to $\sigma = \tau$, w przypadku pozostałych aksjomatów $\sigma \overset{\epsilon}{\Longrightarrow}^+ \tau$.

Jeśli ostatnio zastosowaną regułą było **(przech)**, to znaczy z założenia indukcyjnego, że istnieje ρ takie, że $\sigma \Longrightarrow\!\!\!\!\!\!\!\Longrightarrow \rho$ oraz $\rho \Longrightarrow\!\!\!\!\!\!\!\Longrightarrow \tau$. Wynika stąd natychmiast, że $\sigma \Longrightarrow\!\!\!\!\!\!\!\Longrightarrow \tau$.

Jeśli zastosowano regułę **($\forall$-kontekst)**, to znaczy, że $\sigma = \forall\alpha\,\sigma'$, $\tau = \forall\alpha\,\tau'$ oraz z indukcji, że $\sigma' \Longrightarrow\!\!\!\!\!\!\!\Longrightarrow \tau'$. Wynika z tego oczywiście, że $\forall\alpha\,\sigma' \Longrightarrow\!\!\!\!\!\!\!\Longrightarrow \forall\alpha\,\tau'$.

Zastosowanie reguły **($\to$-kontekst)** wymaga nieco więcej wyjaśnień. Wiemy, że $\sigma = \sigma_1 \to \sigma_2$ i $\tau = \tau_1 \to \tau_2$. Z założenia indukcyjnego wiemy również, że istnieją wyprowadzenia

$$\tau_1 \overset{p_1}{\Longrightarrow} \cdots \overset{p_n}{\Longrightarrow} \sigma_1 \qquad\qquad \text{oraz} \qquad\qquad \sigma_2 \overset{s_1}{\Longrightarrow} \cdots \overset{s_m}{\Longrightarrow} \tau_2$$

Na podstawie tych dwóch wyprowadzeń, można skonstruować następujące:

$$\sigma_1 \to \sigma_2 \;\overset{0p_n}{\Longrightarrow}\; \cdots \;\overset{0p_1}{\Longrightarrow}\; \tau_1 \to \sigma_2 \;\overset{1s_1}{\Longrightarrow}\; \cdots \;\overset{1s_m}{\Longrightarrow}\; \tau_1 \to \tau_2$$

Pierwszy ciąg, to kroki z wyprowadzenia $\tau_1 \overset{p_1}{\Longrightarrow} \cdots \overset{p_n}{\Longrightarrow} \sigma_1$ w których na skutek dodania ścieżkom początkowego symbolu 0 zmieniono znaki z + na − i odwrotnie. Kroki te ustawione są w odwrotnej kolejności i w odwrotnym kierunku.

Dla dowodu implikacji w drugą stronę, wystarczy pokazać, że jeśli $\sigma \Longrightarrow \tau$, to $\sigma \sqsubseteq \tau$. Przypuśćmy, że $\sigma = C[\sigma'] \overset{p}{\Longrightarrow} C[\tau'] = \tau$. Zastosujemy indukcję ze względu na budowę kontekstu $C[\,]$. Jeśli jest on pusty, to $\sigma \sqsubseteq \tau$ dzięki aksjomatowi **(kwant)** lub **(inst)**.

W przeciwnym przypadku są dwie możliwości. Jeżeli $C[\,] = \forall\alpha\; C'[\,]$, to z założenia indukcyjnego $C'[\sigma'] \sqsubseteq C'[\tau']$ i wówczas tezę dostajemy dzięki regule

($\forall$-kontekst). Jeśli natomiast pierwszym symbolem w $C[\,]$ jest $\to$, to należy rozpatrzyć dwa przypadki. Jeśli $p = 0p'$ dla jakiegoś p', to $C[\,] = C_0[\,] \to \rho$ i wówczas $C_0[\tau'] \overset{p'}{\Longrightarrow} C_0[\sigma']$, ponieważ znak p' jest przeciwny do znaku p. Z założenia indukcyjnego $C_0[\tau'] \sqsubseteq C_0[\sigma']$ i, jako że $\rho \sqsubseteq \rho$, na mocy reguły **($\to$-kontekst)**,

$$C_0[\sigma'] \to \rho \sqsubseteq C_0[\tau'] \to \rho$$

Podobnie, jeśli $p = 1p'$, wówczas $C[\,] = \rho \to C_1[\,]$ i z indukcji $C_1[\sigma'] \sqsubseteq C_1[\tau']$. Po zastosowaniu aksjomatu **(zwr)** do typu ρ i reguły **($\to$-kontekst)**, dostajemy

$$\rho \to C_1[\sigma'] \sqsubseteq \rho \to C_1[\tau']$$

■

5 Wyprowadzenia normalne

Celem niniejszego rozdziału jest pokazanie, że każdą nierówność da się udowodnić za pomocą wyprowadzenia *normalnego*, tj. sekwencji kroków dodatnich (które zwiększają rozmiar typu), po której następuje sekwencja kroków ujemnych.

Definicja 5.1 `(wyprowadzenia normalne)`
Wyprowadzenie $\sigma \Longrightarrow\!\!\!\!\!\!\Longrightarrow \tau$ nazywamy wyprowadzeniem normalnym, *jeśli jest ono postaci $\sigma \Longrightarrow\!\!\!\!\!\!\Longrightarrow^{+} \rho \Longrightarrow\!\!\!\!\!\!\Longrightarrow^{-} \tau$ dla pewnego typu ρ.*

Wagą wyprowadzenia normalnego $\sigma \Longrightarrow\!\!\!\!\!\!\Longrightarrow^{+} \rho \Longrightarrow\!\!\!\!\!\!\Longrightarrow^{-} \tau$ nazwiemy sumę wag poszczególnych kroków. Łatwo pokazać, że jest ona równa $2 \times \mathit{waga}(\rho) - \mathit{waga}(\sigma) - \mathit{waga}(\tau)$.

Zanim pokażemy, jak doprowadzić dowolne wyprowadzenie do postaci normalnej, zdefiniujemy najpierw pojęcie bezpiecznych kroków.

Definicja 5.2 `(bezpieczne kroki)`
Krok $\sigma \Longrightarrow \tau$ jest krokiem bezpiecznym, *jeśli spełniony jest jeden z następujących warunków*

1. *jest to krok $\Longrightarrow_q$;*
2. *jest to krok $\Longrightarrow_i$, w którym typem wstawianym jest $\bot$;*
3. *jest to krok $\overset{p}{\Longrightarrow}_i$, w którym typem wstawianym jest zmienna β, związana na pozycji $p' \leq p$, jeśli znak p' jest ten sam, co znak p.*

Do drugiego rodzaju kroków bezpiecznych zaliczymy w szczególności wszystkie instancjacje pustych kwantyfikatorów, to znaczy kroki $C^+[\forall\alpha\,\sigma] \Longrightarrow_i^+ C^+[\sigma]$ oraz odpowiednio $C^-[\sigma] \Longrightarrow_i^- C^-[\forall\alpha\,\sigma]$ dla $\alpha \notin FV(\sigma)$.

W ostatnim punkcie definicji bezpiecznych kroków, chodzi o sytuacje następujące:

$$C^+[\forall\beta\, C^+[\forall\alpha\,\sigma]] \quad \Longrightarrow_i^+ \quad C^+[\forall\beta\, C^+[\sigma\{\alpha \mapsto \beta\}]]$$

$$C^-[\forall\beta\, C^+[\sigma\{\alpha \mapsto \beta\}]] \quad \Longrightarrow_i^- \quad C^-[\forall\beta\, C^+[\forall\alpha\,\sigma]]$$

Wszystkie kroki bezpieczne mają wagę 0. Ich nazwa wynika z faktu, że wsortowanie ich w normalne wyprowadzenie nie zmienia jego wagi. Formalne sformułowanie tej własności, jak również przykład kroku o wadze 0, który nie jest bezpieczny, zostaną podane na końcu niniejszego rozdziału.

Poniższy lemat pokazuje jak zamieniać miejscami kroki ujemne z dodatnimi,[2] co jest niezbędne do przekształcania wyprowadzeń do postaci normalnej.

Lemat 5.3 `(zamiana kroków ujemnych i dodatnich)`
Niech σ i τ będą typami takimi, że $\sigma \Longrightarrow^- \rho \Longrightarrow^+ \tau$. Wtedy istnieje typ ρ' taki, że $\sigma \Longrightarrow^+ \rho' \Longrightarrow^- \tau$. Ponadto jeśli krok $\sigma \Longrightarrow^- \rho$ był bezpieczny, to po zamianie krok $\rho' \Longrightarrow^- \tau$ też będzie bezpieczny i podobnie jeśli krok $\rho \Longrightarrow^+ \tau$ był bezpieczny, to krok $\sigma \Longrightarrow^+ \rho'$ również taki będzie.

Dowód. Niech p i r będą odpowiednio ścieżkami pierwszego i drugiego kroku. W zależności od wzajemnego położenia tych pozycji oraz od rodzaju kroków wyróżniamy następujące przypadki :

- $p \not\gtrless r$ — wtedy zamiana kroków nie stanowi problemu, kroki bezpieczne zostają bezpiecznymi;

- $p < r$, $r = p \cdot s$. Jeśli obydwa kroki są rodzaju $\Longrightarrow_q$, zmiana kolejności nie stanowi problemu. Jeśli pierwszy krok jest rodzaju $\Longrightarrow_i$, a drugi $\Longrightarrow_q$, to mamy następującą sytuację:

$$\begin{aligned} \sigma = \sigma[\forall\vec{\delta}\,\varrho[\forall\vec{\zeta}\,\upsilon]_s\{\alpha \mapsto \xi\}]_p & \overset{p}{\Longrightarrow}{}_i^- & \sigma[\forall\vec{\delta}\alpha\,\varrho[\forall\vec{\zeta}\,\upsilon]_s]_p \\ & \overset{r}{\Longrightarrow}{}_q^+ & \sigma[\forall\vec{\delta}\alpha\,\varrho[\forall\vec{\zeta}\beta\,\upsilon]_s]_p = \tau \end{aligned}$$

  Ponieważ β jest nową zmienną, możemy przyjąć, że β nie występuje w $FV(\xi)$, a więc i w $FV(\upsilon\{\alpha \mapsto \xi\})$. Mamy zatem

$$\begin{aligned} \sigma = \sigma[\forall\vec{\delta}\,\varrho[\forall\vec{\zeta}\,\upsilon]_s\{\alpha \mapsto \xi\}]_p & \overset{r}{\Longrightarrow}{}_q^+ & \sigma[\forall\vec{\delta}\,\varrho[\forall\vec{\zeta}\beta\,\upsilon]_s\{\alpha \mapsto \xi\}]_p \\ & \overset{p}{\Longrightarrow}{}_i^- & \sigma[\forall\vec{\delta}\alpha\,\varrho[\forall\vec{\zeta}\beta\,\upsilon]_s]_p = \tau \end{aligned}$$

  Jeśli zachodzi sytuacja odwrotna, tzn. pierwszy krok jest rodzaju $\Longrightarrow_q$, a drugi $\Longrightarrow_i$, mamy

$$\begin{aligned} \sigma = \sigma[\forall\vec{\delta}\beta\,\varrho[\forall\vec{\zeta}\alpha\,\upsilon]_s]_p & \overset{p}{\Longrightarrow}{}_q^- & \sigma[\forall\vec{\delta}\,\varrho[\forall\vec{\zeta}\alpha\,\upsilon]_s]_p \\ & \overset{r}{\Longrightarrow}{}_i^+ & \sigma[\forall\vec{\delta}\,\varrho[\forall\vec{\zeta}\,\upsilon\{\alpha \mapsto \xi\}]_s]_p = \tau \end{aligned}$$

  Ponieważ β nie występuje w ϱ, możemy przyjąć, iż nie występuje również w ξ, a zatem $\beta \notin FV(\varrho[\forall\vec{\zeta}\,\upsilon\{\alpha \mapsto \xi\}]_s)$. Mamy więc

$$\begin{aligned} \sigma = \sigma[\forall\vec{\delta}\beta\,\varrho[\forall\vec{\zeta}\alpha\,\upsilon]_s]_p & \overset{r}{\Longrightarrow}{}_i^+ & \sigma[\forall\vec{\delta}\beta\,\varrho[\forall\vec{\zeta}\,\upsilon\{\alpha \mapsto \xi\}]_s]_p \\ & \overset{p}{\Longrightarrow}{}_q^- & \sigma[\forall\vec{\delta}\,\varrho[\forall\vec{\zeta}\,\upsilon\{\alpha \mapsto \xi\}]_s]_p = \tau \end{aligned}$$

  W obydwu powyższych sytuacjach typ podstawiany (oznaczany ξ) nie zmieniał się, zatem kroki bezpieczne pozostawały bezpiecznymi.

[2] W drugą stronę zamiana nie zawsze jest możliwa. Nie da się np. odwrócić kolejności kroków $\forall\alpha\,\alpha \overset{\epsilon}{\Longrightarrow}{}_i^+ (\beta \to \beta) \to \beta \overset{0}{\Longrightarrow}{}_i^- (\forall\alpha\,\alpha) \to \beta$

Jeśli obydwa kroki są rodzaju $\Longrightarrow_i$ mamy

$$\begin{array}{rcl} \sigma = \sigma[\forall\vec{\delta}\,\varrho[\forall\vec{\zeta}\beta\,\upsilon]_s\{\alpha\mapsto\xi_\alpha\}]_p & \stackrel{p}{\Longrightarrow}{}_i^- & \sigma[\forall\vec{\delta}\alpha\,\varrho[\forall\vec{\zeta}\beta\,\upsilon]_s]_p \\ & \stackrel{r}{\Longrightarrow}{}_i^+ & \sigma[\forall\vec{\delta}\alpha\,\varrho[\forall\vec{\zeta}\,\upsilon\{\beta\mapsto\xi_\beta\}]_s]_p = \tau \end{array}$$

Wiemy, że $\beta \notin FV(\xi_\alpha)$, ale α może wystąpić w ξ_β. Przyjmując $\bar{\varrho} = \varrho\{\alpha \mapsto \xi_\alpha\}$ oraz $\bar{\upsilon} = \upsilon\{\alpha \mapsto \xi_\alpha\}$, możemy zamienić te kroki w następujący sposób:

$$\begin{array}{rcl} \sigma = \sigma[\forall\vec{\delta}\,\bar{\varrho}[\forall\vec{\zeta}\beta\,\bar{\upsilon}]_s]_p & \stackrel{r}{\Longrightarrow}{}_i^+ & \sigma[\forall\vec{\delta}\,\bar{\varrho}[\forall\vec{\zeta}\,\bar{\upsilon}\{\beta\mapsto\xi_\beta\{\alpha\mapsto\xi_\alpha\}\}]_s]_p \\ & = & \sigma[\forall\vec{\delta}\,\varrho[\forall\vec{\zeta}\,\upsilon\{\beta\mapsto\xi_\beta\}]_s\{\alpha\mapsto\xi_\alpha\}]_p \\ & \stackrel{p}{\Longrightarrow}{}_i^- & \sigma[\forall\vec{\delta}\alpha\,\varrho[\forall\vec{\zeta}\,\upsilon\{\beta\mapsto\xi_\beta\}]_s]_p = \tau \end{array}$$

gdzie środkowa równość jest prostą własnością podstawienia. Jeśli krok ujemny był bezpieczny, to i takim pozostanie, ponieważ podstawiany typ (oznaczany ξ_α) nie zmienił się. Jeśli krok dodatni był bezpieczny, to $\xi_\beta = \bot$ lub $\xi_\beta = \gamma$ dla jakiejś zmiennej γ związanej na pozycji dodatniej; w obu przypadkach $\alpha \notin FV(\xi_\beta)$, a więc $\xi_\beta\{\alpha \mapsto \xi_\alpha\} = \xi_\beta$;

- $p > r$. W tej sytuacji patrzymy na dowód dualny, tzn. jeśli

$$\sigma \stackrel{p}{\Longrightarrow}{}^- \rho \stackrel{r}{\Longrightarrow}{}^+ \tau$$

to również dualnie

$$\tau \to \bot \stackrel{0r}{\Longrightarrow}{}^- \rho \to \bot \stackrel{0p}{\Longrightarrow}{}^+ \sigma \to \bot$$

gdzie oczywiście $0r < 0p$, więc możemy skorzystać z poprzedniego przypadku. ■

Lemat 5.4 *Jeśli dane są dwa typy σ i τ, takie że $\sigma \sqsubseteq \tau$, to istnieje normalne wyprowadzenie $\sigma \Longrightarrow\!\!\!\!\!\!\!\!\gg^+ \Longrightarrow\!\!\!\!\!\!\!\!\gg^- \tau$.*

Dowód. Ponieważ dla danych typów σ i τ istnieje jakieś wyprowadzenie, wystarczy użyć odpowiednią liczbę razy Lematu 5.3, otrzymując wyprowadzenie normalne. ■

Po dowiedzeniu powyższego bardzo istotnego faktu możemy zdefiniować *wagę* nierówności

Definicja 5.5 `(waga nierówności)`
Waga nierówności $\sigma \sqsubseteq \tau$ to minimalna waga jej wyprowadzenia normalnego $\sigma \Longrightarrow\!\!\!\!\!\!\!\!\gg^+ \rho \Longrightarrow\!\!\!\!\!\!\!\!\gg^- \tau$.

Po podaniu powyższej definicji możemy formalnie wyjaśnić znaczenie nazwy kroków bezpiecznych

Stwierdzenie 5.6 *Niech $\sigma_1 \sqsubseteq \tau$ i niech $\sigma_0 \Longrightarrow \sigma_1$ będzie bezpiecznym krokiem. Wówczas waga nierówności $\sigma_0 \sqsubseteq \tau$ jest nie większa od wagi nierówności $\sigma_1 \sqsubseteq \tau$.*

Dowód. Stwierdzenie to wynika bezpośrednio z Lematu 5.3. ■

Podamy teraz przykład kroku o wadze 0, który nie jest bezpieczny. Rozpatrzmy następujące jednokrokowe wyprowadzenie o wadze $waga(\xi)$:

$$\forall\alpha\,(\forall\beta\,(\bot\to\beta)\to\alpha)\ \overset{\epsilon}{\Longrightarrow}{}_i^{+}\ \forall\beta\,(\bot\to\beta)\to\xi$$

Jeśli poprzedzimy je krokiem o zerowej wadze

$$\forall\alpha\,((\bot\to\alpha)\to\alpha)=\forall\alpha\,((\bot\to\beta)\{\beta\mapsto\alpha\}\to\alpha)\ \overset{0}{\Longrightarrow}{}_i^{-}\ \forall\alpha\,(\forall\beta\,(\bot\to\beta)\to\alpha)$$

to dostaniemy wyprowadzenie, w którym krok ujemny poprzedza krok dodatni. Pierwszy krok nie jest bezpieczny, gdyż zmienna α jest związana na pozycji dodatniej ϵ, a nie ujemnej, tak jak zmienna β. Po zamianie tych kroków zgodnie z Lematem 5.3, otrzymujemy wyprowadzenie

$$\forall\alpha\,((\bot\to\alpha)\to\alpha)\ \overset{\epsilon}{\Longrightarrow}{}_i^{+}\ ((\bot\to\xi)\to\xi)\ \overset{0}{\Longrightarrow}{}_i^{-}\ \forall\beta\,(\bot\to\beta)\to\xi$$

którego waga jest równa $3\times waga(\xi)$, czyli więcej niż waga pierwotnego kroku dodatniego, o ile tylko $waga(\xi)>0$.

6 Typy bikoercyjne

Typy σ i τ określamy jako *bikoercyjne*, jeśli $\sigma\sqsubseteq\tau$ i $\tau\sqsubseteq\sigma$. Pokażemy w niniejszym rozdziale, że relacja bikoercyjności może być utożsamiana z najmniejszą kongruencją zawierającą aksjomaty

(B1) $\forall\alpha\forall\beta\,\sigma\sim\forall\beta\forall\alpha\,\sigma$

(B2) $\forall\alpha\,\sigma\sim\sigma\{\alpha\mapsto\bot\}$ (jeśli wszystkie wystąpienia α w σ są pozytywne)

Relacja równoważności związana z oryginalnym systemem Mitchella została scharakteryzowana przez Tiuryna w pracy [Tiu95]. Definiowana jest ona przez aksjomaty **(B1)**, **(B2)** oraz aksjomat

(B3) $\forall\alpha\,(\sigma_1\to\sigma_2)\sim_{\mathrm{M}}\sigma_1\to\forall\alpha\,\sigma_2$ (jeśli $\alpha\notin FV(\sigma_1)$)

który w naszym systemie nie jest prawdziwy.[3] Dalsza część niniejszego rozdziału poświęcona jest dowodowi faktu, że relacja $\sim$ definiowana jako najmniejsza kongruencja zawierająca aksjomaty **(B1)** i **(B2)**, jest relacją równoważności indukowaną przez quasi-porządek $\sqsubseteq$. Aby to udowodnić, wprowadzimy pojęcie kroków odwracalnych.

Definicja 6.1 `(kroki odwracalne)`
Pojedynczy krok redukcji $\sigma\Longrightarrow\tau$ nazwiemy odwracalnym, *jeśli $\tau\sqsubseteq\sigma$.*

Lemat 6.2 *Niech $\sigma\Longrightarrow_i^{+}\tau$ i niech p i r będą takie, że $\sigma|_{pr}=\forall\vec{\delta}\alpha$, gdzie α jest zmienną związaną na pozycji dodatniej p. Jeśli w τ na pozycji pr nie występuje zmienna związana na pozycji dodatniej, to krok $\sigma\Longrightarrow_i^{+}\tau$ nie jest odwracalny.*

[3] Nie jest prawdą, że $\forall\alpha\,(\beta\to(\alpha\to\beta))\sim\beta\to\forall\alpha\,(\alpha\to\beta)$

Dowód. Załóżmy, że krok ten jest odwracalny, czyli że $\tau \sqsubseteq \sigma$. Istnieje zatem wyprowadzenie normalne, $\tau \Longrightarrow\!\!\!\!\!\!\Rightarrow^{+} \rho \Longrightarrow\!\!\!\!\!\!\Rightarrow^{-} \sigma$.

Jeśli $\tau|_{pr} = \forall\vec{\delta}\vec{\zeta}\,\beta$, gdzie β jest zmienną wolną lub związaną na pozycji ujemnej, to również $\rho|_{pr} = \forall\vec{\delta}\vec{\zeta}\,\beta$, ponieważ w fazie dodatniej zmienne związane na pozycjach ujemnych nie są instancjowane, a zmienne wolne w ogóle nie mogą być instancjowane.

Jeśli natomiast $\tau|_{pr} = \forall\vec{\delta}\vec{\zeta}\,(\xi_1 \to \xi_2)$, dla jakichś ξ_1, ξ_2, to podobnie $\rho|_{pr} = \forall\vec{\delta}\vec{\zeta}\,(\xi_1' \to \xi_2')$ (dla jakichś ξ_1', ξ_2'), ponieważ w fazie dodatniej waga typu nie maleje, czyli symbol $\to$ nie może zostać zastąpiony zmienną.

Zatem $\rho|_{pr}$ nie jest zmienną związaną na pozycji dodatniej. Wynika z tego również, że $\sigma|_{pr}$ nie jest zmienną związaną na pozycji dodatniej, ponieważ kroki $\Longrightarrow_q^{-}$ nie mają wpływu na wiązanie zmiennych, a kroki $\Longrightarrow_i^{-}$ nie tworzą nowych zmiennych związanych na pozycji dodatniej. Sprzeczność. ∎

W kolejnym lemacie skorzystamy z następującego oznaczenia. Niech σ będzie typem, w którym na pozycji s znajduje się zmienna α, związana przez kwantyfikator na pozycji p (oczywiście $p \leq s$). Wówczas niech $\Pi_\sigma(s)$ oznacza zbiór pozycji wszystkich wystąpień zmiennej α (związanej przez kwantyfikator $\forall\alpha$ z pozycji p) w typie σ. Jeśli z kontekstu będzie wynikało o jaki typ chodzi, indeks σ zostanie opuszczony.

Lemat 6.3 *Niech $\sigma \Longrightarrow\!\!\!\!\!\!\Rightarrow^{+} \rho \Longrightarrow\!\!\!\!\!\!\Rightarrow^{-} \sigma$ będzie dowolnym normalnym wyprowadzeniem nierówności $\sigma \sqsubseteq \sigma$. Niech p będzie pozycją dodatnią, a r takie, że pr jest pozycją ujemną i $\sigma|_{pr} = \forall\vec{\vartheta}\,\alpha$, gdzie α jest zmienną związaną na pozycji dodatniej p. Wówczas w każdym typie w wyprowadzeniu $\sigma \Longrightarrow\!\!\!\!\!\!\Rightarrow^{+} \rho \Longrightarrow\!\!\!\!\!\!\Rightarrow^{-} \sigma$ na pozycji pr występuje zmienna związana przez kwantyfikator z pozycji p, a zbiór $\Pi(pr)$ nie ulega zmianie.*

Dowód. Wszystkie kroki w rozpatrywanym wyprowadzeniu

$$\sigma = \sigma_0 \Longrightarrow^{+} \sigma_1 \Longrightarrow^{+} \cdots \Longrightarrow^{+} \sigma_n = \rho \Longrightarrow\!\!\!\!\!\!\Rightarrow^{-} \sigma$$

są oczywiście odwracalne, ponieważ wszystkie te typy należą do jednej klasy abstrakcji relacji równoważności indukowanej przez $\sqsubseteq$. Zgodnie z Lematem 6.2, dodatnie kroki odwracalne nie mogą zmieniać kształtu drzewa typu, ani znaku pozycji wiązania zmiennych. Zatem kształt typu oraz znaki pozycji wiązania zmiennych w typie ρ są takie same, jak w typie σ. Kroki ujemne $\Longrightarrow_i^{-}$ mogą jedynie zmniejszać wagę typu oraz zamieniać zmienne wolne lub związane na pozycjach dodatnich w związane na pozycjach ujemnych. Wynika z tego, że w fazie ujemnej kształt drzewa typu, ani znaki pozycji wiązań zmiennych również nie mogą ulec zmianie.

Zbadajmy teraz, jaki wpływ na pozycję wiązania zmiennej α i zbiór $\Pi(pr)$ może mieć pojedynczy krok dodatni $\overset{s}{\Longrightarrow}_i^{+}$, dopuszczalny w myśl Lematu 6.2

- jeśli $s \not\gtrless p$, to krok ten nie ma wpływu na zmienną α;
- jeśli $s \leq p$ i kwantyfikator $\forall\alpha$ na pozycji p jest w zasięgu instancjowanego kwantyfikatora, krok ten nie ma wpływu na pozycję wiązania zmiennej α, ani na zbiór $\Pi(pr)$;

- jeśli $s = p$ i instancjowanym kwantyfikatorem jest $\forall\alpha$, to pozycja wiązania może zmienić się z p na inną pozycję dodatnią p', gdzie $p = p'r'$ dla jakiegoś r'

$$\sigma_i[\forall\vec{\delta}\beta\,\varrho[\forall\vec{\zeta}\alpha\,\upsilon]_{r'}]_{p'} \overset{p}{\Longrightarrow}{}_i^+ \sigma_i[\forall\vec{\delta}\beta\,\varrho[\forall\vec{\zeta}\upsilon\{\alpha \mapsto \forall\vec{\vartheta}\,\beta\}]_{r'}]_{p'} = \sigma_{i+1}$$

- jeśli $s \geq p$ i instancjowany kwantyfikator jest w zasięgu kwantyfikatora $\forall\alpha$, to może zwiększyć się zbiór $\Pi(pr)$ (przyjmujemy, że $s = pr''$ dla jakiegoś r'')

$$\sigma_i[\forall\vec{\zeta}\alpha\,\upsilon[\forall\vec{\delta}\gamma\,\chi]_{r''}]_p \overset{s}{\Longrightarrow}{}_i^+ \sigma_i[\forall\vec{\zeta}\alpha\,\upsilon[\forall\vec{\delta}\,\chi\{\gamma \mapsto \forall\vec{\vartheta}\,\alpha\}]_{r''}]_p = \sigma_{i+1}$$

Nie można zatem zmniejszyć zbioru $\Pi(pr)$ ani zmienić pozycji wiązania na większą w porządku $<$. W przypadku, gdy instancjowany jest kwantyfikator $\forall\alpha$, zmiana jest odwracalna w dodatniej fazie derywacji tylko w sytuacji trywialnej, gdy $p' = p$ i $\beta \notin FV(\varrho)$. Druga zmiana, kiedy instancjowany kwantyfikator jest w zasięgu $\forall\alpha$, jest odwracalna również wyłącznie w przypadku trywialnym, kiedy $\gamma \notin FV(\chi)$.

Ponieważ ujemna faza derywacji nie może zmieniać dodatnich miejsc wiązania zmiennych, więc zmiany pozycji wiązania zmiennych mogą być tylko trywialne. ∎

Lemat 6.4 *Jeśli $\sigma \Longrightarrow \tau$ jest krokiem odwracalnym, to wyrażenie $\sigma \sim \tau$ da się wyprowadzić za pomocą aksjomatów* **(B1)** *i* **(B2)**.

Dowód. Jeśli krok $\sigma \Longrightarrow \tau$ jest rodzaju $\Longrightarrow_q$, wystarczy użyć równości **(B2)**. W przeciwnym przypadku jest to krok rodzaju $\Longrightarrow_i$. Załóżmy, że jest to krok dodatni

$$\sigma = \sigma[\forall\vec{\delta}\alpha\,\varrho]_p \overset{p}{\Longrightarrow}{}_i^+ \sigma[\forall\vec{\delta}\varrho\{\alpha \mapsto \xi\}]_p = \tau$$

Jeśli $\alpha \notin FV(\varrho)$, możemy przyjąć, że $\xi = \bot$ i $\sigma \sim \tau$ na mocy aksjomatu **(B2)**.

Załóżmy zatem, że $\alpha \in FV(\varrho)$. Z Lematu 6.2 wynika, że $\xi = \forall\alpha_1 \ldots \alpha_n\,\beta$, gdzie β jest albo jedną ze zmiennych $\alpha_1, \ldots, \alpha_n$, albo inną zmienną związaną na pozycji dodatniej.

Jeśli $\beta = \alpha_i$ dla jakiegoś i, to z Lematu 6.2 wynika, iż zmienna α może występować w ϱ tylko na pozycjach dodatnich. W tym przypadku, wyrażenie $\sigma \sim \tau$ wyprowadzamy używając aksjomatu **(B2)** na pozycji p, zastępując $\forall\alpha\,\varrho$ przez $\varrho\{\alpha \mapsto \bot\}$, a następnie w każdym wystąpieniu zmiennej α w ϱ dostawiamy brakujące kwantyfikatory, zamieniając wszędzie $\bot = \forall\alpha_i\,\alpha_i$ na $\forall\alpha_1 \ldots \alpha_n\,\alpha_i$ za pomocą $(n-1)$-krotnego użycia aksjomatu **(B2)**.

Jeśli β jest inną zmienną związaną na pozycji dodatniej p', gdzie $p = p'r'$ dla pewnego r', to na mocy Lematu 6.3 mamy następujące możliwości:

- możliwość trywialna — $p' = p$ oraz $\beta \notin FV(\varrho)$;
- wszystkie wystąpienia zmiennych α i β są pozytywne.

W pierwszym przypadku mamy do czynienia z następującą sytuacją:

$$\sigma|_p = \forall\delta_1, \ldots, \delta_{i-1}\,\beta\,\delta_{i+1}, \ldots, \delta_n\,\alpha\,\varrho$$

gdzie $\forall\beta$ jest pustym kwantyfikatorem. Wystarczy odpowiednią liczbę razy użyć aksjomatu **(B1)**, aby doprowadzić do sytuacji

$$\forall\delta_1, \ldots, \delta_{i-1}\,\beta\,\alpha\,\delta_{i+1}, \ldots, \delta_n\,\varrho$$

a następnie raz aksjomatu **(B2)**, aby wyeliminować pusty kwantyfikator $\forall\beta$.

W drugim przypadku zamieniamy najpierw wszystkie wystąpienia zmiennych α i β na $\perp$ za pomocą aksjomatu **(B2)**, a następnie zamieniamy wszystkie wystąpienia $\perp$ na pozycjach zajmowanych oryginalnie przez α i β na nową zmienną γ dostawiając kwantyfikator $\forall\gamma$ w miejscu oryginalnego kwantyfikatora $\forall\beta$

$$\begin{array}{lcl}
 & \sigma & = \sigma[\forall\vec{\delta}\beta\,\upsilon[\forall\vec{\zeta}\alpha\,\varrho]_{r'}]_{p'} \\
\overrightarrow{\textbf{(B2)}} & & \sim \sigma[\forall\vec{\delta}\beta\,\upsilon[\forall\vec{\zeta}\,\varrho\{\alpha\mapsto\perp\}]_{r'}]_{p'} \\
\overrightarrow{\textbf{(B2)}} & & \sim \sigma[\forall\vec{\delta}\,\upsilon[\forall\vec{\zeta}\,\varrho\{\alpha\mapsto\perp\}]_{r'}\{\beta\mapsto\perp\}]_{p'} \\
\overleftarrow{\textbf{(B2)}} & & \sim \sigma[\forall\vec{\delta}\gamma\,\upsilon[\forall\vec{\zeta}\,\varrho\{\alpha\mapsto\gamma\}]_{r'}\{\beta\mapsto\gamma\}]_{p'} \\
 & & \equiv \sigma[\forall\vec{\delta}\beta\,\upsilon[\forall\vec{\zeta}\,\varrho\{\alpha\mapsto\beta\}]_{r'}]_{p'}
\end{array}$$

W obu przypadkach jeszcze trzeba odpowiednią liczbą zastosowań aksjomatu **(B2)** dostawić we wszystkich koniecznych miejscach brakujące puste kwantyfikatory $\forall\alpha_1 \ldots \alpha_n$.

Jeśli rozpatrywany krok $\tau \Longrightarrow \sigma$ był krokiem $\xRightarrow{p}_i^-$, to sytuacja jest dualna. Wystarczy w podobny sposób przeanalizować krok $\tau \to \gamma \xRightarrow{0p}_i^+ \sigma \to \gamma$ (gdzie γ jest nową zmienną), który również jest odwracalny, ponieważ $\sigma \to \gamma \sqsubseteq \tau \to \gamma$.

Zatem dla każdego kroku odwracalnego $\sigma \Longrightarrow \tau$, wyrażenie $\sigma \sim \tau$ da się wyprowadzić używając wyłącznie aksjomatów **(B1)** i **(B2)**. ■

Lemat 6.5 *Jeśli wyrażenie $\sigma \sim \tau$ może być udowodnione jednokrotnym użyciem aksjomatu* **(B1)** *lub* **(B2)**, *to $\sigma \Longrightarrow\!\!\!\!\Rightarrow \tau$ oraz $\tau \Longrightarrow\!\!\!\!\Rightarrow \sigma$. Ponadto w wyprowadzeniach tych używamy samych bezpiecznych redukcji.*

Dowód. Musimy rozpatrzeć cztery możliwości, ze względu na to, który aksjomat został użyty i na jakiej pozycji

- jeśli do dowodu $\sigma \sim \tau$ użyto aksjomatu **(B1)** na pozycji dodatniej p, tzn.

  $$\sigma = \sigma[\forall\vec{\delta}\alpha\beta\,\varrho]_p \sim \sigma[\forall\vec{\delta}\beta\alpha\,\varrho]_p = \tau$$

  to używamy następujących kroków:

  $$\sigma[\forall\vec{\delta}\alpha\beta\,\varrho]_p \xRightarrow{p}_q^+ \sigma[\forall\vec{\delta}\gamma\alpha\beta\,\varrho]_p \xRightarrow{p}_i^+ \sigma[\forall\vec{\delta}\gamma\alpha\,\varrho\{\beta\mapsto\gamma\}]_p \equiv \tau$$

  gdzie γ jest nową zmienną. Krok $\xRightarrow{p}_i^+$ jest bezpieczny, ponieważ term podstawiany jest zmienną związaną na pozycji o identycznym znaku. Wyprowadzenie nierówności w drugą stronę jest analogiczne;

- jeśli użyto aksjomatu **(B1)** na pozycji ujemnej, to korzystając z faktu, że

$$\forall\alpha\beta\,\varrho \equiv \forall\gamma\beta\,\varrho\{\alpha \mapsto \gamma\}$$

gdzie γ jest nową zmienną, możemy użyć kroków następujących

$$\sigma[\forall\vec{\delta}\alpha\beta\,\varrho]_p \equiv \sigma[\forall\vec{\delta}\gamma\beta\,\varrho\{\alpha \mapsto \gamma\}]_p \overset{p}{\Longrightarrow}{}_i^- \sigma[\forall\vec{\delta}\gamma\beta\alpha\,\varrho]_p \overset{p}{\Longrightarrow}{}_q^- \sigma[\forall\vec{\delta}\beta\alpha\,\varrho]_p$$

Krok $\overset{p}{\Longrightarrow}{}_i^-$ jest bezpieczny, jak poprzednio. Podobnie, wyprowadzenie drugiej nierówności jest analogiczne;

- jeśli użyty został aksjomat **(B2)** na pozycji dodatniej p, to mamy do czynienia z następującą sytuacją:

$$\sigma = \sigma[\forall\vec{\delta}\alpha\,\varrho]_p \sim \sigma[\forall\vec{\delta}\,\varrho\{\alpha \mapsto \bot\}]_p = \tau$$

gdzie wystąpienia zmiennej α w typie ϱ są wyłącznie pozytywne. Aby pokazać nierówność w jedną stronę, używamy bezpiecznego kroku

$$\sigma[\forall\vec{\delta}\alpha\,\varrho]_p \overset{p}{\Longrightarrow}{}_i^+ \sigma[\forall\vec{\delta}\,\varrho\{\alpha \mapsto \bot\}]_p$$

Aby pokazać nierówność w drugą stronę, oznaczmy przez $r_1, \ldots, r_n$ pozycje wszystkich wystąpień zmiennej α w typie ϱ. Dla każdego i mamy $\varrho\{\alpha \mapsto \bot\}|_{r_i} = \forall\vec{\zeta_i}\bot = \forall\vec{\zeta_i}\beta\,\beta$, zatem

$$\begin{array}{rcl}
\sigma[\forall\vec{\delta}\,\varrho\{\alpha \mapsto \bot\}]_p & \overset{p}{\Longrightarrow}{}_q^+ & \sigma[\forall\vec{\delta}\gamma\,\varrho\{\alpha \mapsto \bot\}]_p \\
& \overset{pr_1}{\Longrightarrow}{}_i^+ & \sigma[\forall\vec{\delta}\gamma\,\varrho\{\alpha \mapsto \bot\}[\forall\vec{\zeta_1}\,\gamma]_{r_1}]_p \\
& \vdots & \\
& \overset{pr_n}{\Longrightarrow}{}_i^+ & \sigma[\forall\vec{\delta}\gamma\,\varrho\{\alpha \mapsto \bot\}[\forall\vec{\zeta_1}\,\gamma]_{r_1} \cdots [\forall\vec{\zeta_n}\,\gamma]_{r_n}]_p \\
& = & \sigma[\forall\vec{\delta}\gamma\,\varrho\{\alpha \mapsto \gamma\}]_p \\
& \equiv & \sigma[\forall\vec{\delta}\alpha\,\varrho]_p
\end{array}$$

gdzie γ jest nową zmienną. Kroki $\overset{pr_i}{\Longrightarrow}{}_i^+$ są poprawne i bezpieczne, ponieważ wszystkie r_i, a co za tym idzie również wszystkie pr_i są pozytywne, a wstawiana w miejsce kolejnych $\forall\beta\,\beta$ zmienna γ jest związana na pozycji dodatniej p;

- jeżeli aksjomat **(B2)** został użyty na pozycji ujemnej, sytuacja jest dualna. Nierówność $\tau \sqsubseteq \sigma$ dowodzimy łatwo

$$\tau = \sigma[\forall\vec{\delta}\,\varrho\{\alpha \mapsto \bot\}]_p \overset{p}{\Longrightarrow}{}_i^- \sigma[\forall\vec{\delta}\alpha\,\varrho]_p = \sigma$$

a drugą nierówność dowodzimy tak, jak poprzednio, oznaczając $r_1, \ldots, r_n$ pozycje wszystkich wystąpień zmiennej α w typie ϱ (dla każdego i mamy $\varrho|_{r_i} = \forall\vec{\zeta_i}\,\alpha$)

$$\begin{array}{rcl}
\sigma[\forall\vec{\delta}\alpha\,\varrho]_p & \overset{pr_1}{\Longrightarrow}{}_i^- & \sigma[\forall\vec{\delta}\alpha\,\varrho[\forall\vec{\zeta_1}\,\bot]_{r_1}]_p \\
& \vdots & \\
& \overset{pr_n}{\Longrightarrow}{}_i^- & \sigma[\forall\vec{\delta}\alpha\,\varrho[\forall\vec{\zeta_1}\,\bot]_{r_1} \cdots [\forall\vec{\zeta_n}\,\bot]_{r_n}]_p \\
& = & \sigma[\forall\vec{\delta}\alpha\,\varrho\{\alpha \mapsto \bot\}]_p \\
& \overset{p}{\Longrightarrow}{}_q^- & \sigma[\forall\vec{\delta}\,\varrho\{\alpha \mapsto \bot\}]_p
\end{array}$$

■

Twierdzenie 1 (charakteryzacja bikoercyjności)
Nierówności $\sigma \sqsubseteq \tau$ oraz $\tau \sqsubseteq \sigma$ zachodzą równocześnie wtedy i tylko wtedy, gdy $\sigma \sim \tau$.

Dowód. Jeżeli zachodzą obie nierówności $\sigma \sqsubseteq \tau$ oraz $\tau \sqsubseteq \sigma$, to istnieje wyprowadzenie normalne $\sigma \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{+} \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{-} \tau$ złożone z samych kroków odwracalnych. Z Lematu 6.4 wynika, że $\sigma \sim \tau$.

W drugą stronę, jeżeli $\sigma \sim \tau$, to z Lematu 6.5 i z przechodniości relacji $\sqsubseteq$ mamy $\sigma \sqsubseteq \tau$ oraz $\tau \sqsubseteq \sigma$. ∎

Własność typów bikoercyjnych, której będziemy używać w dalszej części pracy wyraża następujące stwierdzenie:

Stwierdzenie 6.6 *Jeśli $\sigma_0 \sim \sigma_1$ i $\sigma_1 \sqsubseteq \tau$ to zachodzi również nierówność $\sigma_0 \sqsubseteq \tau$ oraz waga tej ostatniej nierówności równa jest wadze nierówności $\sigma_1 \sqsubseteq \tau$.* ∎

Dowód. Stwierdzenie to wynika łatwo z Lematu 6.5 oraz ze Stwierdzenia 5.6. ∎

7 Derywacje uporządkowane

Aby pokazać własności relacji $\sqsubseteq$, będziemy używać uporządkowanej wersji normalnych derywacji. Intuicyjnie, pozytywne kroki derywacji będą wykonywać się w kolejności wyznaczonej przez porządek leksykograficzny pozycji, indukowany przez $0 \prec 1$. W ramach tej samej pozycji najpierw wykonywać się będą wszystkie kroki dostawiające kwantyfikatory, a następnie kroki instancjujące, ułożone od zewnętrznych do wewnętrznych. Kroki negatywne uporządkowane będą dualnie. Przepisywanie typów będzie się zatem odbywać najpierw na pozycjach pozytywnych od góry do dołu, a następnie na pozycjach negatywnych od dołu do góry.

Definicja 7.1 (derywacje uporządkowane)
Ciąg kroków pozytywnych $\sigma_0 \overset{p_1}{\Longrightarrow}^{+} \cdots \overset{p_n}{\Longrightarrow}^{+} \sigma_n$ jest uporządkowany, jeśli spełnione są następujące warunki:

1. *dla każdego i, j jeśli $i < j$, to $p_i \preceq p_j$;*
2. *dla ustalonego p kroki $\overset{p}{\Longrightarrow}^{+}_{q}$ poprzedzają kroki $\overset{p}{\Longrightarrow}^{+}_{i}$;*
3. *nie występują pary kroków*

$$\begin{aligned} \sigma_j[\forall\vec{\delta}\,\alpha\,\vec{\zeta}\beta\,\varrho]_p &\overset{p}{\Longrightarrow}^{+}_{i} \sigma_j[\forall\vec{\delta}\,\alpha\,\vec{\zeta}\,\varrho\{\beta \mapsto \xi_\beta\}]_p \\ &\overset{p}{\Longrightarrow}^{+}_{i} \sigma_j[\forall\vec{\delta}\,\vec{\zeta}\,\varrho\{\beta \mapsto \xi_\beta\}\{\alpha \mapsto \xi_\alpha\}]_p \end{aligned}$$

Ciąg kroków negatywnych $\tau_0 \Longrightarrow^{-} \cdots \Longrightarrow^{-} \tau_n$ jest uporządkowany, jeśli dualny ciąg kroków pozytywnych $\tau_n \to \gamma \Longrightarrow^{+} \cdots \Longrightarrow^{+} \tau_0 \to \gamma$ (gdzie γ jest nową zmienną) jest uporządkowany.

Normalną derywację $\sigma \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{+} \rho \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{-} \tau$ nazywamy uporządkowaną *jeśli ciągi $\sigma \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{+} \rho$ oraz $\rho \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{-} \tau$ są uporządkowane.*

Udowodnimy teraz, że każdą derywację da się uporządkować. Potrzebny nam będzie następujący kluczowy lemat.

Lemat 7.2 *Jeśli ciąg kroków* $\sigma \stackrel{p}{\Longrightarrow}^{+} \rho \stackrel{p'}{\Longrightarrow}^{+} \tau$ *nie jest uporządkowany, to istnieje typ* ρ' *taki, że derywacja* $\sigma \Longrightarrow^{+} \rho' \Longrightarrow^{+} \tau$ *jest uporządkowana.*

Dowód. Jeśli obydwa kroki są krokami $\Longrightarrow_q$, to zamiana ich miejscami nie stanowi żadnego problemu.

Jeśli pierwszy jest krok $\Longrightarrow_q$, a drugi $\Longrightarrow_i$ i derywacja nie jest uporządkowana, to znaczy, że $p' \prec p$. W każdym przypadku, nowo dostawiony kwantyfikator nie wiąże żadnej zmiennej w typach ρ i τ, zatem można go wstawić również po instancjacji.

Jeśli pierwszy krok jest krokiem $\Longrightarrow_i$, a drugi $\Longrightarrow_q$, to zamiana również jest oczywista.

Jeśli obydwa kroki wyprowadzenia są typu $\Longrightarrow_i$, to znaczy, że $p' \preceq p$. Jeśli p' nie jest prefiksem p, to zamiana jest trywialna. W przeciwnym przypadku, dla r takiego że $p = p'r$ (być może $r = \epsilon$), mamy następującą sytuację

$$\begin{array}{rcl} \sigma \ = \ \sigma[\forall\vec{\delta}\alpha\, \varrho[\forall\vec{\zeta}\beta\, \upsilon]_r]_{p'} & \stackrel{p}{\Longrightarrow}{}_i^{+} & \sigma[\forall\vec{\delta}\alpha\, \varrho[\forall\vec{\zeta}\, \upsilon\{\beta \mapsto \xi_\beta\}]_r]_{p'} \\ & \stackrel{p'}{\Longrightarrow}{}_i^{+} & \sigma[\forall\vec{\delta}\, \varrho[\forall\vec{\zeta}\, \upsilon\{\beta \mapsto \xi_\beta\}]_r\{\alpha \mapsto \xi_\alpha\}]_{p'} \ = \ \tau \end{array}$$

Stosując fakt, że $\upsilon\{\beta \mapsto \xi_\beta\}\{\alpha \mapsto \xi_\alpha\} = \upsilon\{\alpha \mapsto \xi_\alpha\}\{\beta \mapsto \xi_\beta\{\alpha \mapsto \xi_\alpha\}\}$ i oznaczając $\overline{\varrho} = \varrho\{\alpha \mapsto \xi_\alpha\}$ oraz $\overline{\upsilon} = \upsilon\{\alpha \mapsto \xi_\alpha\}$ możemy dokonać następującej zamiany

$$\begin{array}{rcl} \sigma \ = \ \sigma[\forall\vec{\delta}\alpha\, \varrho[\forall\vec{\zeta}\beta\, \upsilon]_r]_{p'} & \stackrel{p}{\Longrightarrow}{}_i^{+} & \sigma[\forall\vec{\delta}\, \varrho[\forall\vec{\zeta}\beta\, \upsilon]_r\{\alpha \mapsto \xi_\alpha\}]_{p'} \\ & = & \sigma[\forall\vec{\delta}\, \overline{\varrho}[\forall\vec{\zeta}\beta\, \overline{\upsilon}]_r]_{p'} \\ & \stackrel{p'}{\Longrightarrow}{}_i^{+} & \sigma[\forall\vec{\delta}\, \overline{\varrho}[\forall\vec{\zeta}\, \overline{\upsilon}\{\beta \mapsto \xi_\beta\{\alpha \mapsto \xi_\alpha\}\}]_r]_{p'} \\ & = & \sigma[\forall\vec{\delta}\, \varrho[\forall\vec{\zeta}\, \upsilon\{\beta \mapsto \xi_\beta\}]_r\{\alpha \mapsto \xi_\alpha\}]_{p'} \ = \ \tau \end{array}$$

W każdym przypadku można więc zamieniać kolejność źle uporządkowanych par kroków. ∎

Lemat 7.3 `(porządkowanie derywacji)`
Jeśli $\sigma \sqsubseteq \tau$*, to istnieje derywacja uporządkowana* $\sigma \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{+} \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{-} \tau$*, której waga jest równa wadze nierówności.*

Dowód. Na podstawie Lematu 7.2 wnioskujemy, że dowolne ciągi kroków dodatnich mogą być sortowane. Stosując metodę dualną sortować można również kroki ujemne.

Niech $\sigma \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{+} \rho \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{-} \tau$ będzie wyprowadzeniem nierówności $\sigma \sqsubseteq \tau$ o minimalnej wadze. Sortowanie kroków pozytywnych i negatywnych nie zmienia oczywiście typu ρ, więc waga wyprowadzenia $\sigma \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{+} \rho \Longrightarrow\!\!\!\!\!\!\!\!\Longrightarrow^{-} \tau$ równa $2 \times \mathit{waga}(\rho) - \mathit{waga}(\sigma) - \mathit{waga}(\tau)$ nie ulega zmianie. ∎

8 Wymuszanie nierówności

Ten rozdział zawiera dowód własności relacji $\sqsubseteq$ używanych bezpośrednio w dowodzie nierozstrzygalności.

Lemat 8.1 *Niech $\sigma \sqsubseteq \tau$ oraz $\sigma|_p = \varrho_\sigma$ i $\tau|_p = \varrho_\tau$ są podtermami, których zmienne wolne nie są związane odpowiednio w σ i τ. Jeżeli p jest ścieżką pozytywną, to $\varrho_\sigma \sqsubseteq \varrho_\tau$, a jeśli jest ścieżką negatywną, to $\varrho_\tau \sqsubseteq \varrho_\sigma$.*

Dowód. Niech $\sigma = \sigma_0 \stackrel{p_1}{\Longrightarrow}^+ \cdots \stackrel{p_n}{\Longrightarrow}^+ \sigma_n = \tau_m \stackrel{q_m}{\Longrightarrow}^- \cdots \stackrel{q_1}{\Longrightarrow}^- \tau_0 = \tau$ będzie uporządkowanym wyprowadzeniem nierówności $\sigma \sqsubseteq \tau$. Ponieważ ciągi $p_1, \ldots, p_n$ oraz $q_1, \ldots, q_m$ są posortowane leksykograficznie, więc pozycje, których prefiksem jest p występują jako kolejne. Niech $p_k, \ldots, p_{k'}, q_l, \ldots, q_{l'}$ będą wszystkimi takimi pozycjami i niech $r_k, \ldots, r_{k'}, s_l, \ldots, s_{l'}$ będą takie, że $p_i = pr_i$ oraz $q_j = ps_j$.

Załóżmy najpierw, że p jest pozytywne. Wówczas kroki

$$\sigma_{k-1}|_p \stackrel{r_k}{\Longrightarrow}^+ \cdots \stackrel{r_{k'}}{\Longrightarrow}^+ \sigma_{k'}|_p = \tau_{l'}|_p \stackrel{s_{l'}}{\Longrightarrow}^- \cdots \stackrel{s_l}{\Longrightarrow}^- \tau_{l-1}|_p$$

stanowią poprawne wyprowadzeniem nierówności $\varrho_\sigma \sqsubseteq \varrho_\tau$.

Istotnie, spośród pozostałych kroków, których pozycjami są $p_1, \ldots, p_{k-1}$, $p_{k'+1}, \ldots, p_n$ oraz $q_1, \ldots, q_{l-1}, q_{l'+1}, \ldots, q_m$, na podterm na pozycji p mogłyby wpływać wyłącznie te, w których p_i (bądź q_j) jest prefiksem p. Ponieważ wyprowadzenie jest uporządkowane, więc takie pozytywne kroki wystąpią tylko przed krokami $p_k, \ldots, p_{k'}$, a negatywne po krokach $q_l, \ldots, q_{l'}$. Zgodnie z założeniem żadne zmienne w ϱ_σ nie są związane w σ, ani zmienne w ϱ_τ nie są związane w τ. Z tego względu kroki na pozycjach, które są prefiksami p, również nie mogą zmieniać podtypu na pozycji p i zachodzi $\sigma_{k-1}|_p = \sigma|_p = \varrho_\sigma$ oraz $\tau_{l'+1}|_p = \tau_p = \varrho_\tau$.

Niektóre kroki $\Longrightarrow_i^+$ spośród $p_k, \ldots, p_{k'}$ mogą związać pewne zmienne z podtypu na pozycji p przez kwantyfikatory znajdujące się powyżej p

$$\sigma_{i-1}[\forall\vec{\delta}\beta\,\rho[\varrho[\forall\vec{\zeta}\alpha\,\upsilon]_{r_i}]_u]_t \stackrel{p_i}{\Longrightarrow}_i^+ \sigma_{i-1}[\forall\vec{\delta}\beta\,\rho[\varrho[\forall\vec{\zeta}\,\upsilon\{\alpha \mapsto \xi_\alpha\}]_{r_i}]_u]_t = \sigma_i$$

(gdzie $p = tu$, $p_i = pr_i$ jak poprzednio), w przypadku gdy $\beta \in FV(\xi_\alpha)$. Jednak jako że wyprowadzenie jest uporządkowane, te ostatnie kwantyfikatory nie mogą być dalej instancjonowane. Można więc zamiast zmiennych związanych przez te kwantyfikatory, wprowadzać zmienne wolne, ponieważ i tak w fazie ujemnej zostaną one wszystkie zlikwidowane (przez kroki $\Longrightarrow_i^-$), tak jak zostałyby zlikwidowane wszystkie zmienne związane przez kwantyfikatory występujące powyżej pozycji p. Odpowiednikiem tego kroku w nowym wyprowadzeniu, będzie

$$\varrho[\forall\vec{\zeta}\alpha\,\upsilon]_{r_i} \stackrel{r_i}{\Longrightarrow}_i^+ \varrho[\forall\vec{\zeta}\,\upsilon\{\alpha \mapsto \xi_\alpha\}]_{r_i}$$

gdzie $\beta \in FV(\xi_\alpha)$ jest zmienną wolną.

Jeśli p jest negatywne, podobna analiza dowodzi, że nierówność $\varrho_\tau \sqsubseteq \varrho_\sigma$ może być udowodniona przez ciąg kroków

$$\tau_{l-1}|_p \stackrel{s_l}{\Longrightarrow}^+ \cdots \stackrel{s_{l'}}{\Longrightarrow}^+ \tau_{l'}|_p = \sigma_{k'}|_p \stackrel{r_{k'}}{\Longrightarrow}^- \cdots \stackrel{r_k}{\Longrightarrow}^- \sigma_{k-1}|_p$$

■

Lemat 8.2 *Nierówność $\sigma' \to \sigma'' \sqsubseteq \tau' \to \tau''$ zachodzi wtedy i tylko wtedy, gdy zachodzą nierówności $\sigma' \sqsupseteq \tau'$ i $\sigma'' \sqsubseteq \tau''$. Ponadto, jeśli nierówności te są prawdziwe, to waga pierwszej jest równa sumie wag pozostałych.*

Dowód. Implikacja od prawej strony do lewej jest prostą konsekwencją reguły **($\to$-kontekst)**, a od lewej do prawej — poprzedniego lematu.

Zależność wag tych nierówności jest również łatwa do ustalenia. Niech

$$\tau' = \tau'_0 \stackrel{t_1}{\Longrightarrow}^+ \cdots \stackrel{t_k}{\Longrightarrow}^+ \tau'_k = \sigma'_l \stackrel{u_l}{\Longrightarrow}^- \cdots \stackrel{u_1}{\Longrightarrow}^- \sigma'_0 = \sigma' \tag{3}$$

będzie uporządkowanym wyprowadzeniem nierówności $\sigma' \sqsupseteq \tau'$ o minimalnej wadze. Podobnie, niech

$$\sigma'' = \sigma''_0 \stackrel{p_1}{\Longrightarrow}^+ \cdots \stackrel{p_n}{\Longrightarrow}^+ \sigma''_n = \tau''_m \stackrel{q_m}{\Longrightarrow}^- \cdots \stackrel{q_1}{\Longrightarrow}^- \tau''_0 = \tau'' \tag{4}$$

będzie minimalnym wyprowadzeniem nierówności $\sigma'' \sqsubseteq \tau''$. Wówczas wyprowadzenie

$$\begin{aligned} & \sigma'_0 \to \sigma''_0 \stackrel{0u_1}{\Longrightarrow}^+ \cdots \stackrel{0u_l}{\Longrightarrow}^+ \sigma'_l \to \sigma''_0 \stackrel{1p_1}{\Longrightarrow}^+ \cdots \stackrel{1p_n}{\Longrightarrow}^+ \sigma'_l \to \sigma''_n = \\ = \; & \tau'_k \to \tau''_m \stackrel{1q_m}{\Longrightarrow}^- \cdots \stackrel{1q_1}{\Longrightarrow}^- \tau'_k \to \tau''_0 \stackrel{0t_k}{\Longrightarrow}^- \cdots \stackrel{0t_k}{\Longrightarrow}^- \tau'_0 \to \tau''_0 \end{aligned} \tag{5}$$

jest uporządkowanym wyprowadzeniem normalnym nierówności $\sigma' \to \sigma'' \sqsubseteq \tau' \to \tau''$. Zatem waga tej ostatniej jest nie większa od sumy wag pozostałych.

W drugą stronę, każde uporządkowane wyprowadzenie normalne nierówności $\sigma' \to \sigma'' \sqsubseteq \tau' \to \tau''$ zaczyna się od kroków $\stackrel{\epsilon}{\Longrightarrow}^+_q$, poprzedzających kroki $\stackrel{\epsilon}{\Longrightarrow}^+_i$. Mamy zatem następującą sytuację:

$$\sigma' \to \sigma'' \stackrel{\epsilon}{\Longrightarrow}^+_q \cdots \stackrel{\epsilon}{\Longrightarrow}^+_q \forall\alpha_1 \ldots \alpha_n\, \sigma' \to \sigma''$$

gdzie oczywiście $\alpha_1, \ldots, \alpha_n \notin FV(\sigma' \to \sigma'')$. Zatem kolejne kroki $\stackrel{\epsilon}{\Longrightarrow}^+_i$ są pustymi instancjacjami

$$\forall\alpha_1 \ldots \alpha_n\, \sigma' \to \sigma'' \stackrel{\epsilon}{\Longrightarrow}^+_i \cdots \stackrel{\epsilon}{\Longrightarrow}^+_i \forall\alpha_{i_1} \ldots \alpha_{i_k}\, \sigma' \to \sigma''$$

Ponieważ do końca wyprowadzenia nie ma już więcej kroków $\stackrel{\epsilon}{\Longrightarrow}$, a typ końcowy jest postaci $\tau' \to \tau''$, więc ciąg $\alpha_{i_1}, \ldots, \alpha_{i_k}$ musi być ciągiem pustym. Zatem całe wyprowadzenie jest równoważne wyprowadzeniu bez kroków $\stackrel{\epsilon}{\Longrightarrow}^+$. Pozostałe kroki dzielą się dokładnie tak, jak w (5), na wyprowadzenia nierówności $\sigma' \sqsupseteq \tau'$ (3) oraz $\sigma'' \sqsubseteq \tau''$ (4). Zatem suma wag tych dwóch nierówności jest nie większa niż waga $\sigma' \to \sigma'' \sqsubseteq \tau' \to \tau''$. ■

Lemat 8.3 *Niech σ i τ będą takimi typami, że istnieją dwie ścieżki: pozytywna p i negatywna s, takie że $\sigma|_p = \sigma|_s = \alpha$ i $\tau|_p = \tau|_s = \alpha$ dla pewnej zmiennej wolnej α. Ponadto typ τ nie jest postaci $\forall\gamma\, \tau'$. Wówczas*

1. *$\forall\alpha\, \sigma \sqsubseteq \tau\{\alpha \mapsto \xi\}$ wtedy i tylko wtedy, gdy $\sigma\{\alpha \mapsto \xi\} \sqsubseteq \tau\{\alpha \mapsto \xi\}$;*

2. *jeśli obie nierówności zachodzą, to waga pierwszej jest większa lub równa wadze drugiej;*

3. ponadto jeśli $waga(\xi) > 0$ to waga pierwszej nierówności jest ostro większa od wagi drugiej.

Dowód. Jeśli zachodzi nierówność $\sigma\{\alpha \mapsto \xi\} \sqsubseteq \tau\{\alpha \mapsto \xi\}$ dla typów σ i τ jak w treści lematu, to ponieważ $\forall\alpha\, \sigma \overset{\epsilon}{\Longrightarrow}{}^{i}_{+} \sigma\{\alpha \mapsto \xi\}$, więc prawdziwa jest również nierówność

$$\forall\alpha\, \sigma \sqsubseteq \tau\{\alpha \mapsto \xi\} \tag{6}$$

Aby pokazać wynikanie w drugą stronę, załóżmy, że zachodzi nierówność (6). Rozpatrzmy wyprowadzenie uporządkowane tej nierówności o minimalnej wadze. Zaczyna się ono od pewnej liczby (być może zerowej) kroków $\overset{\epsilon}{\Longrightarrow}{}^{+}$ o zerowej wadze

$$\begin{aligned} \forall\alpha\, \sigma \quad & \overset{\epsilon}{\Longrightarrow}{}^{+}_{q} \cdots \overset{\epsilon}{\Longrightarrow}{}^{+}_{q} \quad \forall\alpha_1 \ldots \alpha_n \alpha\beta_1 \ldots \beta_m\, \sigma \\ & \overset{\epsilon}{\Longrightarrow}{}^{+}_{i} \cdots \overset{\epsilon}{\Longrightarrow}{}^{+}_{i} \quad \forall\alpha_{i_1} \ldots \alpha_{i_k} \alpha\beta_1 \ldots \beta_m\, \sigma \end{aligned}$$

Gdyby ciąg $\alpha_{i_1} \ldots \alpha_{i_k}$ nie był pusty, kwantyfikatory te zostałyby na pozycji ϵ aż do końca derywacji, co jest niemożliwe. Z tego samego powodu kolejnym krokiem derywacji musi być krok instancjujący kwantyfikator $\forall\alpha$

$$\forall\alpha\, \sigma' \overset{\epsilon}{\Longrightarrow}{}^{+}_{i} \sigma'\{\alpha \mapsto \chi\} \tag{7}$$

gdzie $\sigma' = \forall\beta_1 \ldots \beta_m\, \sigma$. Oczywiście $\beta_1, \ldots, \beta_m \notin FV(\sigma'\{\alpha \mapsto \chi\})$. Z Lematu 8.1 wynika, że $\xi \sim \chi$, a zatem również $\sigma'\{\alpha \mapsto \xi\} \sim \sigma'\{\alpha \mapsto \chi\}$. Ponadto kwantyfikatory $\forall\beta_1 \ldots \beta_m$ są puste, więc $\sigma\{\alpha \mapsto \xi\} \sim \sigma'\{\alpha \mapsto \chi\}$ i co za tym idzie

$$\sigma\{\alpha \mapsto \xi\} \sqsubseteq \tau\{\alpha \mapsto \xi\}$$

Waga tej ostatniej nierówności, na mocy Stwierdzenia 6.6, jest nie większa od wagi wyprowadzenia $\sigma'\{\alpha \mapsto \chi\} \Longrightarrow\!\!\!\!\!\Rightarrow \tau\{\alpha \mapsto \xi\}$ zatem nie większa niż waga nierówności (6) odjąć waga kroku (7). Waga tego kroku jest większa od zera wtedy i tylko wtedy, gdy $waga(\xi) > 0$, ponieważ wagi ξ i χ są równe. ∎

9 Nierozstrzygalność relacji osłabiania

Niniejszy rozdział zamknie dowód nierozstrzygalności relacji $\sqsubseteq$. Po udowodnieniu Lematów 8.2 i 8.3, wiemy już, że relacja $\sqsubseteq$ ma wszystkie własności relacji $\sqsubseteq_{\mathrm{M}}$ używane w pracy [TU95] do kodowania problemu stopu maszyny rejestrów stosowych. Pokażemy w pewnym skrócie na czym to kodowanie polega.

Maszyna rejestrów stosowych $\mathcal{M}$ posiada dwa rejestry główne V_1 i V_2 oraz skończoną liczbę rejestrów pomocniczych $v_1, \ldots, v_n$. Rejesty te zawierają niepuste słowa nad skończonym alfabetem Σ, złożonym z etykiet instrukcji, w tym wyróżnionej etykiety końcowej e. Krok maszyny polega na zdjęciu etykiety instrukcji z pierwszego rejestru głównego, wykonaniu jej na rejestrach pomocniczych oraz zamianie miejscami rejestrów głównych. Jeśli przeczytana etykieta była ostatnią w rejestrze V_1, to na jego miejsce kopiujemy zawartość pierwszego rejestru pomocniczego v_1 sprzed wykonania instrukcji. Instrukcje maszyny są równoległymi przypisaniami zawartości jednych rejestrów na inne, z ewentualnym dopisaniem na wierzchołku dodatkowych etykiet. Dla danej etykiety $\ell \neq e$, zapisujemy to

$$\ell(r_1, \dots, r_n) = (w_1 r_{i_1}, \dots, w_n r_{i_n}) \qquad \text{gdzie } w_i \in \Sigma^*$$

Mając daną chwilową konfigurację maszyny — krotkę $(V_1, V_2, v_1, \dots, v_n)$ złożoną z $n+2$ słów — następny krok spowoduje następującą zmianę, w zależności od zawartości pierwszego rejestru głównego:

- jeśli $V_1 = eV_1'$, to maszyna się zatrzymuje;
- jeśli $V_1 = \ell V_1'$, gdzie $\ell \neq e$ oraz $V_1 \neq \epsilon$, to następną konfiguracją będzie $(V_2, V_1', v_1', \dots, v_n')$, jeśli $\ell(v_1, \dots, v_n) = (v_1', \dots, v_n')$;
- jeśli $V_1 = \ell$, dla $\ell \neq e$, to następną konfiguracją będzie $(V_2, v_1, v_1', \dots, v_n')$, dla $v_1', \dots, v_n'$ jak poprzednio.

Konfiguracja chwilowa maszyny $(V_1, V_2, v_1, \dots, v_n)$ kodowana jest przez $n+2$ typy $\sigma_{V_1}, \sigma_{V_2}, \sigma_{v_1}, \dots, \sigma_{v_n}$, zawierające pewną ustaloną zmienną wolną γ, łatwo konstruowalne na podstawie zawartości odpowiednich rejestrów. Kluczem do dowodu nierozstrzygalności jest następujące stwierdzenie:

Stwierdzenie 9.1 *Dla danej maszyny rejestrów stosowych $\mathcal{M}$, jej konfiguracja chwilowa $(V_1, V_2, v_1, \dots, v_n)$ prowadzi do zatrzymania wtedy i tylko wtedy, kiedy zachodzi nierówność kodująca*

$$\sigma_{V_1} \sqsubseteq (\sigma_{v_1} \to \sigma_{v_1}) \to \cdots (\sigma_{v_n} \to \sigma_{v_n}) \to \sigma_{V_2} \to \gamma \tag{8}$$

Szkic dowodu:[4] Równoważność ta opiera się na następujących własnościach kodowania i relacji $\sqsubseteq$

- jeśli $(V_1, V_2, v_1, \dots, v_n)$ jest konfiguracją końcową, tj. $V_1 = eV_1'$, to $\sigma_{V_1} = \bot$;
- jeśli $V_1 = \ell V_1'$ dla $V_1' \neq \epsilon$ oraz $\ell(r_1, \dots, r_n) = (w_1 r_{i_1}, \dots, w_n r_{i_n})$, mamy

$$\begin{aligned} \sigma_{V_1} \quad = \quad & \forall \alpha_1 \dots \alpha_n \Big((\alpha_1 \to \alpha_1) \to \dots \to (\alpha_n \to \alpha_n) \to \\ & \Big((\sigma_{w_1}[\alpha_{i_1}] \to \sigma_{w_1}[\alpha_{i_1}]) \to \cdots \to (\sigma_{w_n}[\alpha_{i_n}] \to \sigma_{w_n}[\alpha_{i_n}]) \to \\ & \sigma_{V_1'} \to \gamma \Big) \to \gamma \Big) \end{aligned}$$

  Jeśli zachodzi nierówność kodująca (8), to na mocy n-krotnego użycia Lematu 8.3, wiemy, że jest ona równoważna nierówności

$$\begin{aligned} & (\sigma_{v_1} \to \sigma_{v_1}) \to \dots \to (\sigma_{v_n} \to \sigma_{v_n}) \to \\ & \Big((\sigma_{w_1}[\sigma_{v_{i_1}}] \to \sigma_{w_1}[\sigma_{v_{i_1}}]) \to \dots \to (\sigma_{w_n}[\sigma_{v_{i_n}}] \to \sigma_{w_n}[\sigma_{v_{i_n}}]) \to \sigma_{V_1'} \to \gamma \Big) \to \gamma \\ \sqsubseteq \quad & (\sigma_{v_1} \to \sigma_{v_1}) \to \dots \to (\sigma_{v_n} \to \sigma_{v_n}) \to \sigma_{V_2} \to \gamma \end{aligned}$$

[4] Szczegóły kodowania oraz niniejszego dowodu mogą być znalezione w [TU95]

Po $(n+1)$-krotnym zastosowaniu Lematu 8.2 mamy z kolei równoważność tej nierówności z nierównością kodującą konfigurację po jednym kroku

$$\sigma_{V_2} \sqsubseteq (\sigma_{w_1}[\sigma_{v_{i_1}}] \to \sigma_{w_1}[\sigma_{v_{i_1}}]) \to \ldots \to (\sigma_{w_n}[\sigma_{v_{i_n}}] \to \sigma_{w_n}[\sigma_{v_{i_n}}]) \to \sigma_{V_1'} \to \gamma$$

Waga ostatniej nierówności jest nie większa od wagi nierówności (8), a różnica wag staje się ściśle dodatnią, jeśli choć jeden z typów $\sigma_{v_1}, \ldots, \sigma_{v_n}$ miał niezerową wagę;

- podobnie, jeśli $V_1 = \ell$, dla ℓ jak wyżej, mamy

$$\begin{aligned} \sigma_{V_1} \quad = \quad & \forall \alpha_1 \ldots \alpha_n \Big((\alpha_1 \to \alpha_1) \to \ldots \to (\alpha_n \to \alpha_n) \to \\ & \Big((\sigma_{w_1}[\alpha_{i_1}] \to \sigma_{w_1}[\alpha_{i_1}]) \to \cdots \to (\sigma_{w_n}[\alpha_{i_n}] \to \sigma_{w_n}[\alpha_{i_n}]) \to \\ & \alpha_1 \to \gamma \Big) \to \gamma \Big) \end{aligned}$$

  i stąd, rozumując jak poprzednio, otrzymujemy równoważność nierówności kodującej (8) z nierównością

$$\sigma_{V_2} \sqsubseteq (\sigma_{w_1}[\sigma_{v_{i_1}}] \to \sigma_{w_1}[\sigma_{v_{i_1}}]) \to \ldots \to (\sigma_{w_n}[\sigma_{v_{i_n}}] \to \sigma_{w_n}[\sigma_{v_{i_n}}]) \to \sigma_{v_1} \to \gamma$$

Łatwo zatem pokazać, że jeśli konfiguracja $(V_1, V_2, v_1, \ldots, v_n)$ jest końcowa, to nierówność kodująca zachodzi. W przeciwnym przypadku mamy równoważność

$$\begin{aligned} & \sigma_{V_1} \sqsubseteq (\sigma_{v_1} \to \sigma_{v_1}) \to \cdots (\sigma_{v_n} \to \sigma_{v_n}) \to \sigma_{V_2} \to \gamma \\ \iff \; & \sigma_{U_1} \sqsubseteq (\sigma_{u_1} \to \sigma_{u_1}) \to \cdots (\sigma_{u_n} \to \sigma_{u_n}) \to \sigma_{U_2} \to \gamma \end{aligned}$$

gdzie $(U_1, U_2, u_1, \ldots, u_n)$ jest konfiguracją maszyny po wykonaniu jednego kroku. Ponadto waga tej drugiej nierówności jest nie większa od wagi pierwszej, a jeśli nie wszystkie $v_1, \ldots, v_n$ zaczynają się od etykiety końcowej e, to różnica wag jest ściśle dodatnia.

Dowód faktu, że jeśli dla danej konfiguracji maszyna zatrzymuje się, to nierówność kodująca zachodzi, jest łatwy i był już opisany we wstępie.

W drugą stronę, jeśli zachodzi nierówność kodująca konfigurację chwilową maszyny, to nie może ona zrobić nieskończonej liczby kroków, gdyż przy każdym kroku waga nierówności kodującej zmniejsza się lub pozostaje bez zmian przy jednocześnie malejącej liczbie etykiet w rejestrach głównych. W ostatnim przypadku, jeśli opróżni się któryś z rejestrów głównych, to jego następna zawartość będzie zaczynać się od etykiety końcowej, co w kolejnym kroku doprowadzi do zatrzymania maszyny. ■

Twierdzenie 2
Relacja osłabiania polimorficznego bez aksjomatu rozdzielności $\sqsubseteq$ *jest nierozstrzygalna.*

Dowód. Twierdzenie wynika łatwo z właśnie udowodnionego Stwierdzenia 9.1 oraz z Lematu 16 z pracy Tiuryna i Urzyczyna [TU95], redukującego nierozstrzygalny problem stopu deterministycznych maszyn dwulicznikowych do problemu stopu maszyn rejestrów stosowych. ■

10 Podsumowanie

Praca zawiera dowód nierozstrzygalności relacji osłabiania polimorficznego $\sqsubseteq$ powstałej z relacji osłabiania Mitchella z pracy [Mit88] przez usunięcie aksjomatu rozdzielności kwantyfikatora. Dowód przebiega poprzez redukcję problemu stopu maszyny rejestrów stosowych z pracy Tiuryna i Urzyczyna [TU95]. Niezbędne własności relacji $\sqsubseteq$ dowiedzione są przy użyciu systemu przepisywania typów opartego na aksjomatach Mitchella. Dodatkowo rozdział 6 prezentuje równościową charakteryzację relacji równoważności indukowanej przez quasi-porządek $\sqsubseteq$.

Niemal identyczny dowód nierozstrzygalności mógłby zostać przeprowadzony dla relacji powstałej z $\sqsubseteq$ przez usunięcie kolejnego aksjomatu — aksjomatu dostawiania pustego kwantyfikatora **(kwant)**. Poważniejsze różnice wystąpiłyby jedynie w rozdziale 6, ale najważniejsze dla dalszego dowodu Stwierdzenie 6.6 pozostaje prawdziwe.

Pożądaną kontynuacją pracy byłoby znalezienie jakiejś rozstrzygalnej relacji osłabiania zawartej w relacji $\sqsubseteq_{\mathrm{M}}$ Mitchella. Pewnym pomysłem na rozstrzgalność jest stratyfikacja typów polimorficznych, opisana w pracy Daniela Leivanta [Lei91]. Mając daną rozłączną sumę zbiorów zmiennych $V = \cup_{i \geq 0} V_i$ określamy rząd typu jako maksymalny rząd zmiennej (wolnej lub związanej), która w nim występuje. Następnie kroki **(inst)** mogą być wykonywane wyłacznie wtedy, gdy typ wstawiany ma rząd ściśle mniejszy niż rząd instancjowanej zmiennej.

Przy takim systemie dowód nierozstrzygalności z pracy [TU95], oparty na kodowaniu obliczeń maszyny, nie mógłby być zastosowany, gdyż jak powiedzieliśmy we wstępie, opiera się on na nieograniczonym zagnieżdżeniu instancjacji. Za to prawdopodobnie dowód nierozstrzygalności Wellsa (por. [Wel95]) mógłby zostać przeprowadzony bez problemów. Jednak, ponieważ opiera się on bardzo silnie na aksjomacie rozdzielności, to system bez tego aksjomatu, z ograniczoną przez stratyfikcję instancjacją mógłby okazać się rozstrzygalny.

Bibliografia

[CL91] L. Cardelli and G. Longo. "A semantic basis for Quest", *Journal of Functional Programming*, 1:417-458, October 1991.

[Gir89] J.-Y. Girard, Y. Lafont, and P. Taylor. *Proofs and Types.* Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1989.

[Jim95] T. Jim, "System F plus subsumption reduces to Mitchell's subtyping relation", Manuscript (1995).

[Lei91] D. Leivant, "Finitely stratified polymorphism", *Information and Computation*, **93** (1), 1991, 93–113.

[LMS95] G. Longo, K. Milsted, and S. Soloviev, "A logic of subtyping", *Proc. 10-th IEEE Symp. Logic in Computer Science*, San Diego (1995), pp. 292-299.

[Mit88] J.C. Mitchell, "Polymorphic type inference and containment", *Information and Computation*, **76** (2-3), 1988, pp. 211–249.

[Tiu95] J. Tiuryn, "Equational Axiomatization of Bicoercibility for Polymorphic Types", *Proc. Conf. Foundations of Software Technology and Teoretical Computer Science'95*, Bangalore, India; 18-20 Dec 1995. Lecture Notes in Computer Science 1026, pp. 166-179.

[TU95] J. Tiuryn and P. Urzyczyn, "The Subtyping Problem For Second-Order Types in Undecidable" *Proceedings, 11th Annual IEEE Symposium on Logic in Computer Science*, pp. 74-85, 27-30 July 1996.

[Wel94] J.B. Wells, "Typability and type checking in the second-order λ-calculus calculus are equivalent and undecidable", *Proc. 9th LICS*, IEEE, 1994, pp. 176–185.

[Wel95] J.B. Wells, "The Undecidabilility of Mitchell's Subtyping Relationship", Technical Report, Computer Science Department, Boston University, Number 95-019, December 10 1995.

[Wel96] J.B. Wells, "Typability is Undecidable for F+Eta" Technical Report, Computer Science Department, Boston University, Number 96-022, March 9 1996.