Dipartimento di Filosofia, Comunicazione e Spettacolo
Corso di Laurea Magistrale in Scienze Filosofiche

Tesi di Laurea Magistrale in Logica

Uno studio della normalizzazione in logica classica

Candidato
Fabio Massaioli

Relatore
Prof. Lorenzo Tortora de Falco

Correlatore
Prof. Matteo Morganti

Anno Accademico 2019/2020

Indice

Introduzione

Il presente lavoro è dedicato allo studio delle procedure di normalizzazione nei sistemi deduttivi per la logica classica. In particolare, oggetto di studio principale sarà il processo di riduzione non deterministico proprio della versione più generale del calcolo dei sequenti classico **LK**.

Il capitolo 1 è dedicato a una sommaria retrospettiva sullo sviluppo e i risultati più importanti della teoria della dimostrazione, accompagnata da alcune riflessioni di carattere filosofico. Lo studio formale del contenuto e della struttura delle dimostrazioni ha radici molto antiche e si può ricondurre direttamente ad Aristotele, considerato l'iniziatore in Occidente della logica formale come disciplina autonoma. Il filosofo è noto come padre del sillogismo, spesso inteso semplicemente come schema di inferenza e di argomentazione per la logica dei predicati. Questa di fatto è stata la visione trasmessa dalla logica medievale, ma non si deve commettere l'errore di attribuirla allo stesso Aristotele.

Come ha osservato Natali [Nat14], il suo approccio è invece caratterizzato da una peculiare distinzione, non del tutto esplicita, fra gli schemi dimostrativi di uso comune, dipendenti da assiomi logici generali quali non contraddizione e terzo escluso, e il sillogismo inteso come contenuto "profondo" della dimostrazione, presentato (specialmente nella forma più semplice) come un argomento la cui validità dipende esclusivamente dalla sua struttura, non da principi esterni o dal contenuto di premesse e conclusioni. Gli altri schemi di argomentazione non sono squalificati, ma la loro *ammissibilità* (per dirla con termini recenti) dipende dalla possibilità di tradurli in sillogismi; potremmo dire in altre parole che un ragionamento valido "contiene" dei sillogismi impliciti.

Le idee appena descritte sono state presto dimenticate (lo stesso nome di *Organon* conferito dai posteri ai trattati logici di Aristotele indica che la logica era concepita spesso come strumento piuttosto che come campo di indagine filosofica) e sono rimaste di fatto marginali fino alla loro riscoperta attraverso tre passaggi chiave: *primo*, la formalizzazione della logica classica ad opera di Frege, Hilbert e Tarski, che ha offerto un quadro di riferimento per uno studio rigoroso della logica e ha aperto la strada all'idea di *dimostrazione come oggetto simbolico manipolabile*; *secondo*, l'intuizionismo di Brouwer che ha contribuito a spostare l'attenzione dalla *dimostrabilità delle proposizioni* alla *forma delle dimostrazioni*; *terzo*, la scoperta da parte di Gentzen della dimostrazione come oggetto *dotato di una struttura interna* e delle *procedure di normalizzazione*: procedure, cioè, per trasformare dimostrazioni liberamente costruite in strutture, dette forme normali, accomunate da proprietà peculiari.

La logica intuizionistica, grazie alle sue caratteristiche, si è rivelata terreno fertile per le ricerche sulla struttura e sulla normalizzazione delle dimostrazioni: inizialmente con la scoperta di Gentzen (il cui significato è stato compreso più tardi, grazie alla logica lineare) che la differenza fra logica classica e intuizionistica può essere espressa come differenza molto semplice nella struttura interna delle dimostrazioni; in seguito con la scoperta ad opera di Curry e Howard della corrispondenza precisa fra dimostrazioni e programmi, con le procedure di normalizzazione che equivalgono all'esecuzione del programma corrispondente alla dimostrazione.

Questa scoperta ha permesso di formalizzare le intuizioni sul "contenuto" e il "significato" delle dimostrazioni: una dimostrazione corretta non è più, semplicemente, un ragionamento che non conclude il falso, quanto piuttosto un oggetto dotato di significato come *algoritmo per la costruzione del risultato* (l'idea, a grandi linee, è che se una proposizione afferma che dati due interi è possibile calcolarne la somma, la dimostrazione sarà un algoritmo per calcolare effettivamente la somma di due interi qualsiasi).

Si apre così la prospettiva di pensare e studiare gli schemi di ragionamento non più come semplici "inferenze valide" in base a principi esterni, ma come operazioni dotate di un significato proprio e interno. Se la logica intuizionistica è in un certo senso quella che usiamo quando ragioniamo su informazioni e processi conoscitivi, d'altra parte è quella classica che utilizziamo quando ragioniamo sulla realtà che ci circonda. È naturale allora chiedersi cosa caratterizzi, da questo nuovo punto di vista, gli schemi di ragionamento propri della logica classica.

È ben noto però che quest'ultima pone alcune grosse difficoltà e non si presta, almeno superficialmente, alle analisi appena menzionate. Dopo i primi risultati di Gentzen [Gen35] e le successive ricerche di Kreisel [Kre51] e Prawitz [Pra81], la teoria della dimostrazione per la logica classica ha dovuto attendere i primi anni '90 del XX secolo per muovere alcuni significativi passi avanti.

Un primo importante risultato è stato quello di Griffin [Gri90]. La scoperta che la legge classica $\neg\neg A \to A$ può essere utilizzata per *tipare* un operatore di controllo nel λ-calcolo con continuazioni ha aperto la strada a una serie di proposte tra cui il calcolo λ_c di Krivine e il λμ-calcolo di Parigot [Par92b] sono forse le più note e studiate.

Sul fronte del calcolo dei sequenti troviamo il sistema **FD** di Parigot [Par92a], che contiene il λμ-calcolo tipato come frammento e consiste in una forma di calcolo dei sequenti che condivide alcune importanti proprietà della deduzione naturale. Le ricerche in logica lineare hanno generato invece il calcolo **LC** di Girard [Gir91] che permette di fornire una semantica denotazionale per la logica classica utilizzando una forma particolare degli spazi coerenti.

Le proposte menzionate prevedono in alcuni casi l'estensione dei sistemi intuizionistici con regole di riduzione *ad hoc*, in altri casi una dissimmetrizzazione delle procedure di riduzione classiche che permetta di distinguere in ogni caso rilevante input (premesse) e output (conclusioni). Quest'ultimo approccio è stato studiato nella forma più generale da Danos, Joinet e Schellinx che hanno introdotto in [DJS97] il calcolo dei sequenti $\mathbf{LK}^{tq}$. Si tratta di un sistema deduttivo per la logica classica che è traducibile induttivamente in logica lineare, gode di una procedura di riduzione confluente e fortemente normalizzante e in cui è possibile ricostruire come frammenti molti dei sistemi menzionati sopra. Il prezzo (minimo) da pagare per queste proprietà è l'aggiunta di un'informazione ad ogni formula (il colore) che indica se la formula in questione si comporti dal punto di vista computazionale come un input o come un output.

Un merito di questo approccio è di identificare con precisione le cause del non determinismo della procedura originale di Gentzen. È legittimo però chiedersi se questa sia una caratteristica essenziale della logica classica, piuttosto che un difetto. Sembra interessante in particolare studiare il protocollo di riduzione per **LK** ottenuto da $\mathbf{LK}^{tq}$ "dimenticando" il colore delle formule o, equivalentemente, modificando liberamente il colore dopo ogni passo di riduzione. Tale protocollo è non deterministico, e tuttavia non soffre dei difetti insiti nella definizione originale di [Gen35]. Presentiamo e discutiamo brevemente le definizioni nel capitolo 2; il risultato principale di questa sezione è una variante del *lemma di stabilità* di [DJS97] valida per la versione non deterministica di **LK** (lemma 2.47).

Un approccio simile è quello scelto da Barbanera e Berardi nella definizione del λ-calcolo simmetrico in [BB96]. Si tratta di un sistema sviluppato con due obiettivi: primo, fornire una versione del λ-calcolo con tipi classici che sia semplice e "naturale" (nel senso che non si introducono riduzioni *ad hoc*); secondo, permettere l'estrazione di testimoni dalle dimostrazioni di formule Σ^0_1 nel linguaggio dell'aritmetica.

Nel capitolo 3 ci concentriamo sul primo aspetto e in particolare sul frammento proposizionale $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$. Presentiamo e analizziamo la tecnica dei *candidati simmetrici* utilizzata in [BB96] per dimostrare la normalizzazione forte della procedura di riduzione dei termini tipati.

Seguendo l'intuizione, accennata anche in [DJS97], che il λ-calcolo simmetrico equivalga a una versione "libera" di $\mathbf{LK}^{tq}$ in cui il colore delle formule di taglio può essere scelto liberamente a ogni passo, procediamo quindi a un confronto fra $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$ e la versione non deterministica di **LK** definita nel capitolo 2. Accenniamo a una serie di restrizioni su **LK** che permettono di tradurre le derivazioni in termini di $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$: queste considerazioni forniscono una guida per applicare la tecnica dei candidati simmetrici alla versione completa e non ristretta di **LK**.

Presentiamo infine nel capitolo 4 una dimostrazione di normalizzazione forte per il calcolo dei sequenti proposizionale $\mathbf{LK}^{\mathbf{0}}$ (teorema 4.23) e per quello proposizionale del secondo ordine $\mathbf{LK}^{\mathbf{2}}$ (teorema 4.33) utilizzando con una piccola correzione la nozione di riduzione definita nel capitolo 2. La nostra analisi delle procedure di riduzione ci permette di individuare una ulteriore fonte di non determinismo che discutiamo brevemente nel capitolo 5 (sezione 5.3). Otteniamo inoltre un risultato generale (proposizione 4.9) che insieme all'osservazione appena menzionata potrebbe essere alla base di una ulteriore analisi dei processi di riduzione, con l'idea di semplificare la dimostrazione dei teoremi di normalizzazione forte o identificare qualche genere di invariante.

Capitolo 1

La logica come studio delle dimostrazioni

Chiunque abbia avuto un'esposizione scolastica (o anche una basilare esposizione universitaria) alla logica è abituato a pensare ad essa come allo studio della validità delle proposizioni e della correttezza del ragionamento. Può capitare poi, a seconda del proprio ambito di studio, di entrare in contatto con il mondo dei linguaggi formali e la tetrade dei connettivi classici: congiunzione, disgiunzione, implicazione, negazione. In tal caso si potrà avere l'idea che la logica si occupi di determinare il significato di questi "componenti elementari" del discorso (le *costanti logiche*) e gli strumenti tecnici di ragionamento che li riguardano: tavole di verità, algebra di Boole, sistemi assiomatici, deduzione naturale, *tableaux* analitici *et cetera*.

Questa visione, salvo i dovuti raffinamenti, non è lontana dal vero. Da una parte la logica aristotelica prende le sue mosse proprio dall'analisi delle predicazioni, dell'affermazione e della negazione, del concetto di verità, come anche dall'analisi della correttezza e della giustificazione del ragionamento:[1] è l'inizio di una tradizione che attraversa i secoli fino ad oggi, e che peraltro molto presto ha identificato i connettivi più noti (specialmente l'implicazione e in generale le espressioni condizionali) come fulcro del dibattito, accanto agli schemi deduttivi.

D'altra parte anche la logica formale contemporanea, frutto delle faticose elaborazioni dell'800 e della prima metà del '900, si divide a grandi linee in due branche che ruotano rispettivamente intorno ai concetti di *soddisfacibilità* e *dimostrabilità*.

1.1 Soddisfacibilità e dimostrabilità

Lo studio della *soddisfacibilità* riguarda la relazione fra la proposizione e l'oggetto del discorso, e per la precisione:

a. a quali condizioni una proposizione sia "vera" rispetto a un dato oggetto o "stato di cose", e viceversa

b. cosa caratterizza gli "stati di cose" che soddisfano una certa proposizione.

Naturalmente questo comporta la precisazione di un concetto di "verità", di "oggetto del discorso" o "stato di cose": tutto ciò che passa solitamente sotto il

[1]Per un'introduzione si veda ad esempio [Nat14].

nome di *semantica* e che corrisponde all'idea intuitiva delle proposizioni come "descrizioni". L'approccio più noto e forse più diffuso (ma non certamente l'unico) è quello che fa capo alla tradizione semantica *tarskiana* e che trova la sua espressione più raffinata nella *teoria dei modelli*.

Gli sforzi per la formalizzazione dei concetti di proposizione, verità logica o soddisfacibilità, "stato di cose", seppur mantenendo l'impianto e i principi della logica classica di ispirazione aristotelica, hanno generato definizioni di volta in volta più astratte, conducendo la teoria della soddisfacibilità e dei modelli verso l'indipendenza da assunzioni di carattere *ontologico* sulle proprietà e la forma degli "oggetti del discorso". L'attenzione si è spostata sulla forma delle proposizioni, del discorso stesso se vogliamo, mettendo al centro le cosiddette *costanti logiche* o, nei termini di Popper, i *segni formativi*:[2] quei segni cioè che per convenzione rappresentano non il contenuto ma la forma di una proposizione.

La novità non sta nella limitazione dello studio logico ad un linguaggio rigido e artificiale: come sottolinea Popper [Pop47, p. 272], Aristotele è il primo ad adottare questa soluzione restringendo il campo d'indagine alle proposizioni categoriche, per le quali la copula sarà un *segno formativo*. Piuttosto, la scelta iniziale delle costanti logiche *e della loro interpretazione* mette in evidenza un minimo insieme di assunzioni (precisamente quelle necessarie per definire il significato dei segni) da cui si fa derivare ogni altro teorema: se pure qualcuno dovesse considerare le assunzioni discutibili, le conseguenze però sono certe e indiscutibili, universalmente valide *in relazione alle costanti scelte*.

In questo modo il dibattito si allontana dalla metafisica e si concentra sulla scelta delle costanti logiche; a seguire, e separatamente, ci si potrà chiedere se la scelta effettuata sia adeguata a "descrivere" la realtà, fino a che punto e in quali ambiti. Un risultato importante della formalizzazione è dunque la chiarezza raggiunta intorno all'oggetto proprio della ricerca logica.

Ricordiamo invece che Aristotele deduce i principi della sua logica da quelli che ritiene caratteri ontologici indubitabili della realtà,[3] e molti dibattiti sulla validità di questa o quella forma di ragionamento, di questa o quella dimostrazione, sono stati spesso *dibattiti sulla concezione della realtà* (o dell'oggetto del discorso, comunque inteso). Riportiamo un esempio intorno al principio di non contraddizione. Una nota definizione fornita da Aristotele (*Metaph.* Γ 4, 1005b19–21) è la seguente:

> *«È infatti impossibile che lo stesso attributo appartenga e non appartenga allo stesso soggetto nello stesso tempo e sotto lo stesso aspetto.»*

È evidente che si tratta di un principio ontologico e non puramente formale, dal quale si traggono due conseguenze immediate:

[2]A questo proposito è interessante e poco conosciuto [Pop47], in cui Popper affronta il problema in modo originale anticipando le proposte più tarde che discuteremo in seguito.

[3]Come mette in evidenza il già citato [Nat14] o anche l'apparato di note al libro Γ della Metafisica a cura di Enrico Berti in [Ari17].

a. è un errore affermare o negare uno stesso predicato di uno stesso soggetto: questa è la legge logica corrispondente al principio ontologico;

b. *«è impossibile pensare che la stessa cosa sia e non sia la stessa cosa»* (*Metaph.* Γ 4, 1005b29–30).

Potrebbe sorprendere quest'ultima affermazione e anche suonare come una petizione di principio, ma se ne coglie facilmente la coerenza osservando che "pensare che sia" e "pensare che non sia" sono per l'appunto *attributi contrari* della stessa sostanza che pensa! Penserò dunque cose diverse e contrarie in momenti diversi, ma non allo stesso tempo (purché io intenda effettivamente ciò che Aristotele intende per negazione: a questo serve la precisazione «sotto lo stesso aspetto»).

Diversamente, la semantica classica formale comincia con la distinzione di due possibili *denotazioni* di ogni proposizione: il *vero* e il *falso* (principio di bivalenza). Si continua poi con una interpretazione delle costanti logiche per la negazione e per la congiunzione: se A è una proposizione, *non-A* è vera quando A è falsa e viceversa; se A e B sono proposizioni, *A-e-B* è vera se e solo se A è vera e B è vera. Bastano questi passaggi a stabilire che *A-e-non-A* non può mai essere vera, dunque è falsa e vale il principio di non contraddizione: *non-(A-e-non-A)*. Evidentemente ogni disaccordo non può che concentrarsi sulle definizioni di denotazione, negazione, congiunzione: la domanda sulla loro adeguatezza alla realtà è diventata secondaria.

Lo stesso atteggiamento guida lo sviluppo degli studi sulla dimostrabilità. A partire da Frege (che pubblica nel 1893 uno dei testi più influenti in materia) e poi specialmente con Hilbert, il concetto di dimostrazione viene formalizzato attraverso i cosiddetti sistemi assiomatici, caratterizzati da un numero estremamente ridotto di *regole di inferenza* (tipicamente il solo *modus ponens*) e da una ampia scelta, più o meno libera, di *schemi di assiomi* a cui si affida il compito di fornire regole specifiche per i vari connettivi. Una dimostrazione diventa in senso formale una sequenza di formule puramente simboliche che possano essere ricondotte a qualche schema assiomatico o derivate da formule precedenti tramite le regole di inferenza, ed è possibile verificare meccanicamente e in un numero finito di passi se una data sequenza di formule sia o meno una dimostrazione in un certo sistema di assiomi. Diremo allora che una formula A è derivabile in un sistema S da un insieme di assunzioni Γ, e scriveremo

$$\Gamma \vdash_S A$$

quando esiste una sequenza di formule (*derivazione*) che termina con A e tale che ogni altro elemento della sequenza sia un elemento di Γ, o altrimenti sia riconducibile a un assioma di S o derivabile da altre formule della sequenza tramite qualche regola di inferenza di S.

Il concetto di *derivabilità* in un sistema assiomatico sostituisce il concetto più informale di dimostrabilità e si pone naturalmente il problema dei rapporti fra derivabilità e soddisfacibilità. Come nel caso della soddisfacibilità, la relazione di derivabilità in un sistema dato è fuori discussione e ogni disaccordo si concentra sulla scelta degli assiomi: quale ragione abbiamo per scegliere un sistema piuttosto che un altro? Quali insiemi di assiomi sono “interessanti”? La soluzione tradizionale consiste nell’accoppiare strettamente derivabilità e soddisfacibilità, chiedendo a ogni sistema di assiomi di rispettare la condizione di correttezza rispetto a una certa nozione di soddisfacibilità:

$$\Gamma \vdash A \implies \Gamma \models A$$

dove $\Gamma \models A$ significa che quando le premesse Γ sono soddisfatte, anche la conclusione A è soddisfatta. In particolare, se A è derivabile allora è infalsificabile (sempre soddisfatta). È questa la versione formale dell’intuizione secondo cui un ragionamento corretto *trasmette la verità* dalle premesse alla conclusione (e cosa altrettanto importante, come evidenziato da Popper nonché da Lakatos [Pop47; Lak79], *ritrasmette la falsità* dalle conclusioni alle premesse).[4]

Altrettanto interessante è la domanda inversa: tutto ciò che è soddisfacibile è anche derivabile? Vale cioè

$$\Gamma \models A \implies \Gamma \vdash A\,?$$

Questo è precisamente il contenuto del teorema di *completezza della logica*, uno dei primi importanti risultati *metalogici*, dimostrato da Gödel nel 1929 per la logica classica dei predicati del primo ordine: in parole povere, al primo ordine e per la semantica classica, la soddisfacibilità equivale alla derivabilità e pertanto è possibile utilizzare i sistemi deduttivi per studiare anche la soddisfacibilità.

1.2 Verso la teoria della dimostrazione

L’esperienza filosofica insegna che ogni sistema di ragionamento semplice, flessibile e potente porta con sé anche la possibilità di grandi banalizzazioni: quelle sottigliezze che si è scelto inizialmente di mettere da parte per semplificare il problema, vengono presto trattate come se non esistessero affatto. A maggior ragione questo accade con le formalizzazioni della logica e con la loro capacità di scomporre problemi complessi in pochi elementi semplici.

Pensiamo a due tipiche reazioni di fronte ai risultati della logica moderna: da una parte una visione puramente strumentale, che non riconosce alla logica la dignità di scienza indipendente né una profondità filosofica propria, per esempio

[4] “Trasmette la verità” nel senso che, quando le premesse sono vere, allora anche le conclusioni sono vere, mentre non vale necessariamente il viceversa. Popper e Lakatos osservano che è importante anche sottolineare il converso: quando le conclusioni sono false, allora qualche premessa deve essere falsa. Sebbene sia formalmente equivalente alla prima affermazione, questo secondo punto di vista è cruciale dal punto di vista euristico, quando si cerca di capire “cosa non va” in una dimostrazione

riducendola a uno studio meramente propedeutico alla matematica o alla filosofia.[5] Dall'altra parte, una fede smisurata nella logica come soluzione ad ogni problema teorico nella storia dell'umanità: in questo caso la logica finisce per diventare *l'unico* oggetto di studio della filosofia; osserviamo anche che i successi del formalismo hanno conferito spesso a questo atteggiamento la forma di una speranza nella meccanizzazione totale del sapere.[6]

Un tratto che accomuna i due atteggiamenti è quello di tendere ad un certo tipo di trivializzazione dei linguaggi e dei sistemi formali. Sappiamo che una volta scelte le costanti logiche e gli assiomi "tutto funziona"; l'idea allora suona più o meno così: scegliamo le costanti logiche (cioè i segni formativi, il linguaggio), le interpretazioni e gli assiomi che più ci convengono per rappresentare "la logica del nostro oggetto di indagine". Seguiranno uno dopo l'altro tanti utili teoremi, chiari ed irrefutabili, e scomparirà ogni dubbio sull'argomento. – Certo – potrà rispondere un avversario – nel tuo sistema! Ma *il mio sistema* è ben diverso. – E diverso ancora sarà il suo sistema, il loro e così via *ad libitum*: l'unico effetto di questo approccio è di trasformare dispute filosofiche in dispute sui formalismi, tutti banalmente *corretti e completi* rispetto a… se stessi![7] Così quel "distacco dalla realtà" (ma dovremmo dire: dalle *concezioni* della realtà) che è la forza delle astrazioni logiche formali, diventa spesso anche la loro debolezza.

L'idea che il compito della logica sia quello di annunciare verità astratte, sviscerare teoremi o generare certezze è leggermente ingenua e del tutto estranea a un serio concetto di logica come scienza o come ambito di indagine filosofica. Da una parte, proprio perché il formalismo si fonda sull'astrazione e sul distacco dalla realtà, acquisire una certezza nell'ambito di un sistema formale non significa necessariamente acquisire una certezza in senso assoluto. Potremmo dire che la logica è in un certo senso lo studio *di ciò che è relativo*: alle costanti logiche, alle interpretazioni, a determinate assunzioni e così via. Non sarebbe neanche sbagliato concepire la teoria dei modelli e della soddisfacibilità come uno studio delle relazioni (sebbene esistano sicuramente altri punti di vista). D'altra parte, l'acquisizione di verità logiche comporta un processo di ricerca e verifica delle dimostrazioni che è soggetto ad errori e non si può in alcun modo dare per scontato.[8] Certamente i formalismi sono stati elaborati per aumentare

[5] Si tratta peraltro di un atteggiamento antico e radicato, espresso già dal titolo *Organon* ("strumento") attribuito a posteriori all'opera logica di Aristotele che invece a questa visione era probabilmente estraneo, come sottolinea Natali in [Nat14, p. 39].

[6] Sebbene gli studi di Gödel sull'incompletezza abbiano smorzato non poco questi ottimismi, è facile constatare che questa forma di estremismo tende a ripresentarsi ad ogni progresso tecnico o tecnologico. D'altra parte anche questo è un difetto umano (e specialmente filosofico) che ha radici antichissime.

[7] Per una critica salace di questi atteggiamenti, forse a volte eccessivamente dura, si può indulgere nella lettura di [Rin90] e dei primi paragrafi di [Gir08]

[8] Bisogna anche ricordare a questo proposito che l'idea di "verificabilità meccanica" per un logico è puramente astratta: la macchina potrebbe impiegare un battito di ciglia come anche un millennio, l'importante è che sia possibile!

il rigore e la chiarezza, ma spesso la ricompensa di questi sforzi è la scoperta di errori e paradossi e semmai una riduzione delle certezze.[9]

Un contributo importante verso nuove direzioni di ricerca venne dall'ingresso di L.E.J. Brouwer e dell'*intuizionismo* nel dibattito sui fondamenti. Le proposte di Brouwer nascevano da considerazioni di carattere epistemologico e spesso ontologico che, sebbene abbiano lasciato un segno importante e durevole nella filosofia della matematica, sono risultate spesso controverse e ad oggi non sono ritenute parte dei fondamenti della scienza matematica. Tuttavia la ricerca matematica ispirata alle concezioni dell'intuizionismo ha prodotto risultati concreti e strumenti utili in svariati ambiti fra cui spiccano l'informatica e la logica. A questo proposito, si potrebbe paragonare la scoperta della logica intuizionistica a quella delle geometrie non-euclidee: per la prima volta qualcuno mostrò che la scienza logica non consisteva necessariamente nello studio dei principi "classici" e che invece era possibile – e fecondo – studiare sistemi differenti. Come nel caso della geometria, l'intuizionismo non ha prodotto frammentazione, ma ha incoraggiato la ricerca di più profondi principi comuni.

Muovendo da una particolare concezione degli oggetti e della scienza matematica, dunque proprio in base a considerazioni sulla natura degli "oggetti del discorso" e delle leggi logiche,[10] Brouwer riportò l'attenzione su un aspetto che era stato messo in secondo piano dallo studio della derivabilità e dei sistemi assiomatici: *la dimostrazione*, la sua struttura e i suoi contenuti.

1.2.1 La centralità della dimostrazione

Non sarebbe affatto sbagliato vedere questo cambio di prospettiva come un ritorno alle origini piuttosto che come una novità. Lo studio delle dinamiche del pensiero e dell'argomentazione è la più profonda vocazione della logica, riflessa anche nell'etimologia del nome oltre che nella pratica filosofica. Il cuore della stessa logica Aristotelica non è tanto lo studio della verità e del linguaggio quanto soprattutto lo studio degli schemi di ragionamento e specialmente del sillogismo. Quest'ultimo ha due tratti interessanti: il primo è che Aristotele, nella sua pratica argomentativa, non si limita certamente all'uso dei sillogismi; piuttosto sembra concepire il sillogismo come una struttura "profonda" dell'argo-

[9]Per lungo tempo le «rivoluzioni del rigore» logico hanno messo in crisi interi settori della matematica che si credevano solidamente acquisiti. Quelli che oggi sono ritenuti standard essenziali del rigore sono stati visti in passato, anche da grandi matematici, come atteggiamenti inutili o dannosi. Lakatos offre una bella ricostruzione di questi dibattiti nel già citato [Lak79].

[10]Citiamo da [Kle52] (trad. nostra): *«Brouwer, in un articolo intitolato 'L'inaffidabilità dei principi logici', mise in discussione la convinzione che le regole della logica classica, che in buona sostanza sono giunte fino a noi da Aristotele, abbiano una validità assoluta, indipendente dall'oggetto del discorso a cui sono applicate.»* È significativo il fatto che in realtà per Aristotele le leggi logiche avessero validità *precisamente* in forza della natura degli oggetti a cui si applicavano. È noto d'altra parte che Brouwer ritiene la logica classica perfettamente valida per insiemi finiti di oggetti, e come fa notare Berti (sempre nelle sue annotazioni al libro Γ in [Ari17]) questo era proprio l'ambito del discorso che Aristotele aveva in mente.

mentazione: le inferenze di uso più comune sono valide perché *potenzialmente traducibili* in sillogismi (ancora una volta seguiamo [Nat14] su questo punto). Il secondo tratto interessante è che il sillogismo è concepito come un'inferenza valida esclusivamente per l'esistenza di determinate relazioni fra i suoi termini e indipendentemente dal loro contenuto: dunque completare un sillogismo significa mettere in evidenza uno schema di relazioni fra premesse e conclusione che implica *necessariamente* (ma seguendo un'osservazione di Abrusci potremmo tradurre meglio *automaticamente*) la loro inseparabilità.[11] Sono questi due tratti che ritroveremo a breve nei sistemi deduttivi introdotti da Gentzen e nell'idea di una semantica basata sulle dimostrazioni.

D'altra parte dobbiamo osservare che la pratica matematica, da molto tempo, riconosce un ruolo centrale alla dimostrazione. Spesso chi non si occupa direttamente della scienza matematica è abituato a concepirla come un insieme di proposizioni (teoremi) indubitabilmente vere («la matematica non è un'opinione..!»). Nella visione comune la dimostrazione è un'attività secondaria collegata alla ricerca della certezza che deve caratterizzare le affermazioni della matematica. Se ipotizzassimo di poter meccanizzare completamente il processo dimostrativo (ed esiste un senso preciso in cui questo è possibile), la ricerca matematica si ridurrebbe – così si tende a pensare – all'attività di elencare le proposizioni dimostrabili; anche chi scrive non può negare di essere stato di questa opinione, prima di avvicinarsi allo studio della logica matematica.

Il problema è che la maggior parte di queste proposizioni (e dimostrazioni) saranno trivialità prive di importanza e spesso incomprensibili.[12] Nel lavoro del matematico c'è spesso un aspetto molto più fondamentale della banale verifica dei teoremi: si tratta di elaborare concetti, di cercare formulazioni e modi di presentare i problemi che favoriscano la comprensibilità della domanda, della risposta e della dimostrazione, e dove possibile anche la semplicità. Una dimostrazione corretta, ma inaccessibile alle capacità di lettura e comprensione di un essere umano, viene sempre accolta con sospetto: non per disprezzo gratuito, ma perché non ha nulla da insegnare! Leggendo un lavoro di matematica, non ci aspettiamo solamente che ci convinca della validità di ogni affermazione, ma che ci insegni a *pensare il problema*: questo è compito della dimostrazione.

Aspetti profondi della razionalità matematica sono legati al contenuto delle dimostrazioni. Rientra certamente fra questi, ad esempio, la tematica delle dimostrazioni cosiddette "costruttive": questo genere di dimostrazioni non si limita ad appurare in astratto la verità della conclusione, ma fornisce una *procedura* per raggiungere la conclusione ogni volta che le premesse siano soddisfatte. In particolare, una dimostrazione costruttiva di *esistenza* fornirà una procedura per esibire concretamente un membro (il cosiddetto *testimone*) della classe di

[11]*«Il sillogismo è un discorso nel quale, poste alcune cose, qualcosa di diverso da esse deriva necessariamente [per forza, automaticamente], per il fatto che esse sono come sono.»* (*A.pr.* A 1, 24b18–20)

[12]Come mettono in luce ad esempio i suggestivi capitoli introduttivi della tesi di dottorato di Paolo Pistone [Pis15].

oggetti di cui si è dimostrata l'esistenza: un esempio classico di questo stile sono le dimostrazioni con riga e compasso che forniscono una procedura per costruire (letteralmente) l'oggetto cercato una volta dati i punti di partenza. Per fornire un esempio ulteriore, consideriamo proprio le dimostrazioni del teorema di completezza: queste spesso si basano sulla presentazione di una procedura per ottenere un'interpretazione falsa delle proposizioni che non sono dimostrabili, procedura che gode di un interesse proprio e indipendente dal teorema di completezza.

Un secondo aspetto è quello delle tecniche dimostrative, che possono essere studiate in astratto e generalizzate a molte classi di problemi. Un terzo aspetto fondamentale è legato alla comprensione e all'approfondimento del contenuto delle proposizioni tramite l'analisi delle loro dimostrazioni. Una trattazione approfondita di questo aspetto è stata fornita da Lakatos nel suo [Lak79]: una delle sue osservazioni più interessanti è che anche la dimostrazione (incompleta) di una proposizione falsa può favorire la comprensione del problema e offrire spunti di ricerca fecondi. Chi scrive può testimoniare personalmente come questo accada continuamente nella pratica della ricerca.

1.2.2 La proposta intuizionista e l'opera di Gentzen

Torniamo a Brouwer, dunque. Come è noto, una delle caratteristiche salienti della sua filosofia era l'opposizione all'uso indiscriminato del principio classico del *terzo escluso* e in particolare alla sua validità generale, anche per universi del discorso infiniti. Le ragioni sono molteplici e non è possibile discuterle nel dettaglio: di seguito ci concentriamo su quelle più strettamente inerenti alle dimostrazioni.[13]

Una famosa formulazione del principio in questione si trova nel già citato libro Γ della Metafisica:

> *«Non può esserci alcuna via di mezzo della contraddizione, bensì è necessario, di uno stesso soggetto, o affermare o negare uno stesso predicato, qualunque esso sia.»* (1011a23–24)

Possiamo scomporre la definizione di Aristotele in due passaggi:

a. un primo passaggio ha carattere *negativo* (*«non può esserci alcuna via di mezzo della contraddizione»*): sotto questo punto di vista il principio esprime il fatto che non esistano alternative ai modi dell'affermazione e della negazione;

b. un secondo passaggio ha carattere *positivo* (*«è necessario affermare o negare»*): esprime l'idea, ben più controversa, che una delle due alternative sia necessariamente vera (ciò che è comunemente espresso nella logica moderna dalla validità della formula $A \vee \neg A$ *per ogni* A).

[13] Per un approfondimento si può consultare [MB93] o la voce della SEP *Intuitionism in the Philosophy of Mathematics* [Iem20]

È importante sottolineare che l'intuizionismo *non rifiuta* la formulazione negativa. Si afferma spesso che la logica intuizionista rinuncia al *principio di bivalenza*,[14] ma questa affermazione non deve essere fraintesa come se si ammettessero *più di due valori di verità*. Il punto è che la logica intuizionistica si concentra sulla *dimostrabilità*, ed esistono precisamente due modi della dimostrazione: prova e refutazione. Non è affatto detto però (neanche classicamente!) che per ogni proposizione io riesca a concludere una dimostrazione in uno dei due modi. Affermare e negare qui devono essere intesi come atti del giudizio e non come operazioni interne alla logica: il fatto che in alcuni casi sia impossibile emettere un giudizio non vale affatto come terza via.[15]

La seconda formulazione invece è precisamente quella rigettata da Brouwer e colleghi.[16] Per la precisione, l'intuizionismo rifiuta la forma positiva del *terzo escluso* come principio dimostrativo generale; il motivo di questa scelta deve essere cercato nella struttura delle dimostrazioni che ne fanno uso: quelle *per assurdo* o *non costruttive*.[17]

L'andamento tipico di un argomento per assurdo è il seguente: vogliamo verificare che un certo oggetto x esista; mostriamo che se non esistesse si avrebbe una contraddizione (*reductio ad absurdum*); concludiamo che l'oggetto deve esistere. C'è qualcosa di profondamente controintuitivo in questo schema di ragionamento: non abbiamo modo di indicare *quale sia* l'oggetto cercato, e tuttavia ci viene chiesto di credere alla sua esistenza. Ora questo è abbastanza naturale in domini del discorso finiti o enumerabili o più in generale quando siamo in grado di "immaginare" o intuire la forma dell'oggetto. Tipicamente questi saranno ambiti in cui la logica intuizionistica è equivalente a quella classica. Ma se questo non fosse il caso, non potremmo mai chiedere alla dimostrazione per assurdo di "darci un'idea" dell'oggetto. Che forma ha? Dove posso trovarlo? Niente di tutto questo: sappiamo solo che esiste. È precisamente quello che l'intuizionista rifiuta di accettare: in questa linea di pensiero un puro predicato di esistenza è privo di significato.[18] Per Brouwer dimostrare una proposizione significa fornire una

[14]Il principio per cui esistono al più due valori di verità, tipico della logica classica.

[15]Anche lo stesso Aristotele considerava questa possibilità, come risulta dalle sue famose riflessioni sulle verità future nel *Perì Hermeneias*; in altri passi però sembra escludere completamente questa possibilità rispetto alle realtà attuali: *«Tutto ciò che è pensabile o intelligibile, il pensiero o lo afferma o lo nega»* (1012a1). In senso negativo (non esistono alternative) saremo certamente d'accordo; la frase però ha decisamente un tono positivo e non è chiaro come debba essere interpretata.

[16]Curiosamente, anche Aristotele rigetta l'espressione *"A-o-non-A"* che per lui rappresenta proprio l'impossibile punto medio fra affermazione e negazione: *«L'ente non è detto 'non essere o essere'»* (1011a29). È chiaro però che accetta la disgiunzione dei contraddittori ("o qualcosa è, o non è") come principio della dimostrazione (si pensi agli assiomi di un sistema hilbertiano), e sappiamo che questo è sufficiente a giustificare la validità della formula $A \vee \neg A$ all'interno del sistema.

[17]Per essere precisi, è perfettamente possibile ragionare per assurdo utilizzando comunque un argomento costruttivo. Tuttavia (almeno nella logica classica) l'unico modo per argomentare *non costruttivamente*, è procedere per assurdo: l'intuizionista dunque ottiene la restrizione a dimostrazioni costruttive eliminando completamente il ragionamento per assurdo

[18]Non tanto "se non vedo non credo", perché spesso la costruzione dell'oggetto cercato sarà im-

procedura che permetta di "costruire" la conclusione a partire dalle premesse, e in particolare dimostrare un risultato di esistenza significa offrire una costruzione dell'oggetto cercato. Egli aveva indentificato nel principio generalizzato del terzo escluso il fondamento della validità dei ragionamenti per assurdo, e proponeva dunque di respingerne la validità (ma senza affermarne la falsità!) per restringere la pratica logica alla classe delle dimostrazioni *costruttive*.

La novità e l'interesse di questo approccio stanno precisamente nell'accettare o respingere i principi logici non in base a considerazioni "semantiche" od opportunistiche, ma con l'obiettivo di mettere sotto controllo la forma delle derivazioni. Fu certamente questo uno dei motivi che attrassero l'attenzione di studiosi (come Gödel e Gentzen o la nascente famiglia degli informatici) che pure non sottoscrivevano necessariamente la filosofia intuizionista. E fu appunto Gerhard Gentzen, della scuola di Hilbert, a compiere il passo più importante verso la futura teoria della dimostrazione.

La prima idea fondamentale di Gentzen fu quella di sostituire i sistemi assiomatici, difficili da trattare sia intuitivamente che formalmente, con un sistema «che si avvicinasse il più possibile al ragionamento reale» [Gen35]. Questo sistema, noto per l'appunto come "deduzione naturale", è composto da un numero ampio di regole di inferenza (almeno due per ogni connettivo) e da un numero di assiomi molto ristretto: un'approccio dunque completamente opposto a quello dei sistemi precedenti. Come conseguenza di questa scelta, Gentzen riesce a presentare le dimostrazioni sotto forma di alberi che possono essere composti e scomposti in vari modi e su cui è possibile ragionare per induzione.[19]

Il naturale passo successivo consiste nello studio della struttura generale di queste dimostrazioni (per cui si parla di *teoria strutturale della dimostrazione*). Due principi informano la scelta delle regole in deduzione naturale: la *separazione* delle costanti (ogni regola riguarda una singola costante logica) e la dualità fra *introduzioni ed eliminazioni* (ogni costante logica avrà una o più regole che la introducono nelle conclusioni e una o più regole che la eliminano dalle premesse).[20] In figura 1.1 (pagina seguente) mostriamo ad esempio le regole per la congiunzione e per l'implicazione. È facile osservare che la regola $\to E$ corrisponde al *modus ponens*; invece la regola $\to I$ si legge in questa maniera: se ipotizzando A si riesce, in un certo numero di passi, a derivare B, allora si può concludere $A \to B$ e *dimenticare l'ipotesi* A. Quest'ultimo passaggio è espresso dalle parentesi quadre aggiunte intorno all'ipotesi: diciamo che l'ipotesi è stata *scaricata* o che non è più *attiva*. Se esiste una derivazione in cui Γ è l'insieme del-

praticabile in atto. Il motto intuizionista suona piuttosto "credo solo se, almeno in linea di principio, potrei vedere"

[19] Per albero indendiamo qui un oggetto matematico con una struttura "ramificata": si parte da un elemento detto "radice" da cui si diramano vari percorsi (rami) che portano verso ulteriori elementi, da cui si diramano ulteriori percorsi e così via. Su una simile struttura si può "ragionare per induzione" nel senso del principio di induzione matematica, che permette di dimostrare proprietà generali di insiemi di oggetti ordinati adeguatamente. Nel caso dell'albero l'ordine è dato proprio dai "cammini" che dalla radice si allontanano seguendo i rami.

[20] Su questo punto seguiamo la presentazione di [Sch18].

$$\frac{A \qquad B}{A \wedge B}\wedge I \qquad \frac{A \wedge B}{A}\wedge E_1 \qquad \frac{A \wedge B}{B}\wedge E_2$$

$$\frac{\begin{array}{c}[A]\\ \vdots\\ B\end{array}}{A \to B}\to I \qquad \frac{A \to B \qquad A}{B}\to E$$

Figura 1.1: Regole di deduzione naturale per congiunzione e implicazione.

le ipotesi attive e A è la conclusione, scriviamo come prima $\Gamma \vdash A$ (A è derivabile da Γ). Una derivazione senza ipotesi attive (che chiameremo *derivazione chiusa*) dimostra la validità generale della sua conclusione; per esempio la derivazione seguente

$$\frac{[A]}{A \to A}\to I$$

dimostra la validità della tautologia $A \to A$ (e scriveremo $\vdash A \to A$).

Ora immaginiamo di voler dimostrare una proposizione, diciamo B, riguardo a un certo caso particolare. Per semplificare il lavoro abbiamo già mostrato un lemma: se la condizione A è soddisfatta, allora seguirà B ($A \to B$). Chiamiamo p la derivazione di questo lemma in deduzione naturale. Ora non ci resta che mostrare che nel nostro caso particolare vale A (sia q tale derivazione) e applicare la regola di *eliminazione dell'implicazione*:

$$\dfrac{\dfrac{\begin{array}{c}[A]\\ \vdots\, p\\ B\end{array}}{A \to B}\to I \qquad \begin{array}{c}\vdots\, q\\ A\end{array}}{B}\to E$$

Gentzen chiamava questo genere di strutture "digressioni":[21] un'introduzione seguita immediatamente da un'eliminazione dello stesso connettivo. La seconda idea fondamentale di Gentzen fu dunque di definire una procedura per eliminare da una derivazione tutte le "digressioni". In questo caso particolare, ad esempio, si procederebbe così:

$$\begin{array}{c}\vdots\, q\\ A\\ \vdots\, p\\ B\end{array}$$

[21] *Umwege*: la resa non è facile ma l'idea è quella di "fare un giro" invece di raggiungere direttamente la destinazione.

Intuitivamente, abbiamo sviluppato il ragionamento fino a dedurre A per il nostro caso particolare; quindi abbiamo applicato immediatamente il ragionamento con cui si dimostra il lemma ($A \rightarrow B$), ma senza fare appello esplicitamente al lemma, arrivando così a concludere B. Osserviamo che quest'ultima derivazione è stata ottenuta scomponendo quella precedente e ricomponendo alcune delle sue parti: è possibile definire procedure simili anche per tutti gli altri connettivi (e più in generale per tutte le costanti logiche):

Possiamo ora presentare il risultato fondamentale di Gentzen, noto come *Hauptsatz* o *teorema di eliminazione del taglio*:

a. una derivazione è in *forma normale* quando è "priva di digressioni";[22]

b. ogni derivazione *chiusa* può essere ridotta in forma normale attraverso un numero finito di trasformazioni.

Non si deve sottostimare l'importanza e la fecondità di questo risultato, che di fatto fonda la teoria della dimostrazione aprendo la strada a un nuovo ramo della scienza logica. Nella prossima sezione discuteremo alcune conseguenze importanti delle scoperte di Gentzen; intanto ne presentiamo due che permettono di "chiudere" il discorso sull'intuizionismo che abbiamo aperto all'inizio di questo paragrafo.

Quando si adotta il sistema di deduzione naturale di tipo intuizionistico, è possibile dimostrare che ogni derivazione chiusa e in forma normale termina con una regola di introduzione. Come corollario del *Hauptsatz* abbiamo dunque che ogni derivazione chiusa può essere trasformata in una derivazione normale *che termina con una regola di introduzione che introduce la conclusione*. In questo modo si riesce a dare un senso formale molto preciso all'idea di "dimostrazione costruttiva": nel caso particolare di una dimostrazione di esistenza, riducendo la derivazione in forma normale si ottiene che l'ultima regola introduca un quantificatore esistenziale, e la premessa di questa regola conterrà un'esempio concreto del tipo di oggetto cercato![23]

Osserviamo che non è particolarmente importante ridurre *attualmente* le derivazioni in forma normale (cosa che in alcuni casi richiederebbe quantità impraticabili di tempo e spazio), quanto piuttosto *la possibilità* di farlo e le conseguenze di questa potenzialità.

---

[22] *«Er hat keine Umwege»* [Gen35].

[23] Consideriamo per esempio il teorema di Euclide sui numeri primi, che può essere espresso nella forma seguente: *se n è un numero primo, allora esiste un numero m primo e maggiore di n.* Il teorema può essere formalizzato nel linguaggio dell'aritmetica e dimostrato in deduzione naturale intuizionistica; possiamo inoltre dimostrare che $n = 3$ è un numero primo e applicare la regola $\rightarrow E$ per concludere che esiste un numero m primo e maggiore di 3. Riducendo la derivazione così ottenuta in forma normale, troveremo che l'ultima regola introduce il quantificatore esistenziale, e la sua premessa conterrà un'espressione aritmetica che denota effettivamente qualche numero primo e maggiore di 3.

1.2.3 Il calcolo dei sequenti

Concludiamo la presentazione ricordando che Gentzen non riuscì a dimostrare il teorema di eliminazione del taglio direttamente per la deduzione naturale (il risultato in questa forma è dovuto a [Pra65]), in particolare non per la deduzione naturale classica, che come vedremo sotto presenta delle difficoltà. Elaborò invece un calcolo logico alternativo, noto come "calcolo dei sequenti". Un sequente è un'espressione di questo tipo:

$$A, B, \dots \vdash C, D, \dots$$

dove le formule sulla sinistra sono interpretate, come ci si potrebbe aspettare, come "premesse" (o meglio antecedenti), e le formule a destra come "conclusioni" (o conseguenti). In particolare la sequenza di destra è interpretata come una disgiunzione: la dimostrabilità di *tutti* gli antecedenti implica la dimostrabilità di *almeno uno* dei conseguenti. Dualmente, la refutabilità di *tutti* i conseguenti implica la refutabilità di *almeno uno* degli antecedenti. Il sequente vuoto indica il falso logico o assurdità. Le idee di fondo del calcolo sono due:

1. se il sequente $\Gamma \vdash A$ è derivabile nel calcolo dei sequenti, allora esiste una deduzione naturale con premesse Γ e conclusione A;

2. le regole di eliminazione della deduzione naturale sono interpretate come regole di introduzione delle premesse; ad esempio la regola $\wedge E_1$ permette di trasformare una derivazione con premessa A in una derivazione con premessa $A \wedge B$:

$$\cfrac{A \wedge B}{A}\,\wedge E_1 \\ \vdots \\ C$$

La regola corrispondente nel calcolo dei sequenti sarà allora:

$$\vdots \\ \cfrac{A \vdash C}{A \wedge B \vdash C}\,(\wedge L_1)$$

Il calcolo, sebbene non sia privo di difetti, gode di alcune proprietà molto vantaggiose:

– la presenza di "digressioni" viene resa esplicita grazie alla regola (cut), o regola di taglio:[24]

$$\cfrac{\Gamma \vdash A \qquad A \vdash \Delta}{\Gamma \vdash \Delta}\,(cut)$$

[24]Ne presentiamo una versione semplificata per mettere in evidenza la relazione fra premesse e conclusioni; la stessa cosa faremo con le altre regole. Per una presentazione completa del calcolo si veda il capitolo 2.

che esprime la transitività della relazione di derivabilità e dunque la possibilità di comporre due derivazioni; il teorema di eliminazione del taglio fornisce allora una procedura per eliminare da qualsiasi dimostrazione tutte le occorrenze della regola (cut);

- si semplifica la definizione di forma normale: una derivazione è normale sse è *senza tagli*. Per il caso intuizionistico e per derivazioni di conclusione $\vdash A$ (senza antecedenti) continuano a valere le proprietà già enunciate riguardo all'ultima regola. Vale inoltre la *proprietà della sottoformula*: ogni occorrenza di formula in una derivazione normale è sottoformula di qualche occorrenza di formula nella conclusione;

- la forma del sequente permette di "tracciare" proprietà *globali* della struttura della derivazione, semplificando le definizioni delle regole. Consideriamo la regola $\to I$ e quella corrispondente nel calcolo dei sequenti:

$$\dfrac{\begin{matrix}[A]\\ \vdots\\ B\end{matrix}}{A \to B}\to I \qquad \dfrac{\begin{matrix}\vdots\\ A \vdash B\end{matrix}}{\vdash A \to B}\,(\to R)$$

  La prima cosa che salta all'occhio è la mancanza dell'assunzione scaricata: questa è rappresentata dalla scomparsa della formula A sulla sinistra del sequente. L'aspetto meno evidente è che la regola di deduzione naturale è *globale*: per verificarne la correttezza è necessario ispezionare l'intera derivazione. La regola $(\to R)$ invece è *locale*: la sua correttezza è garantita dal rispetto della struttura richiesta per i sequenti premessa e conclusione;

- come conseguenza del punto precedente, anche le regole di trasformazione (per l'eliminazione del taglio) sono *locali*, e cioè si limitano a modificare solamente le regole che precedono immediatamente il taglio che si vuole eliminare. Come vedremo nel prossimo capitolo, questa caratteristica crea alcune difficoltà, tuttavia permette di semplificare molto la definizione e lo studio delle trasformazioni.

Da ultimo presentiamo una delle caratteristiche più importanti del calcolo dei sequenti: le regole strutturali esplicite (così chiamate perché "modificano la struttura" della derivazione). Si tratta delle regole di contrazione, indebolimento (weakening) e scambio (exchange):

$$\dfrac{A, A, \Gamma \vdash \Delta}{A, \Gamma \vdash \Delta}\,(C) \qquad \dfrac{\Gamma \vdash \Delta}{A, \Gamma \vdash \Delta}\,(W) \qquad \dfrac{A, B, \Gamma \vdash \Delta}{B, A, \Gamma \vdash \Delta}\,(X)$$

Ovviamente possono essere applicate liberamente anche sulla destra del sequente. La seconda regola è abbastanza intuitiva: se Δ segue da Γ, allora segue anche da Γ con l'aggiunta di A; esprime la possibilità di ignorare le premesse e corrisponde al fatto che, quando si applica la regola $\to I$ in deduzione naturale, non è obbligatorio scaricare un'assunzione:

$$\dfrac{\begin{array}{c}\Gamma\\ \vdots\\ B\end{array}}{A\to B}\to I \qquad \dfrac{\dfrac{\begin{array}{c}\vdots\\ \Gamma\vdash B\end{array}}{\Gamma, A\vdash B}\,(W)}{\Gamma\vdash A\to B}\,(\to R)$$

La contrazione esprime la possibilità di usare una premessa più volte (due occorrenze di A vengono presentate come una sola) e corrisponde al fatto che quando si applica la regola $\to I$ è consentito scaricare anche più di un'assunzione:

$$\dfrac{\begin{array}{c}\Gamma\ [A]\ [A]\\ \vdots\\ B\end{array}}{A\to B}\to I \qquad \dfrac{\dfrac{\begin{array}{c}\vdots\\ \Gamma, A, A\vdash B\end{array}}{\Gamma, A\vdash B}\,(C)}{\Gamma\vdash A\to B}\,(\to R)$$

Durante l'eliminazione del taglio, la presenza di una contrazione comporta la duplicazione di una parte della derivazione e se usata in modo opportuno può far esplodere la taglia della forma normale: sono caratteristiche già presenti nella deduzione naturale, ma vengono alla luce chiaramente nel calcolo dei sequenti.

In generale queste tre regole hanno ricevuto per lungo tempo una scarsa considerazione: erano ritenute (specialmente la terza) una fastidiosa necessità tecnica dovuta alla complessità del calcolo dei sequenti e si preferiva quando possibile utilizzarle in forma implicita. Tuttavia fu già Gentzen nel suo [Gen35] a osservare che il calcolo dei sequenti intuizionistico **LJ** poteva essere ottenuto da quello classico **LK** *restringendo il lato destro del sequente a una singola conclusione*, il che, crucialmente, equivale a proibire l'uso della contrazione sulla destra del sequente: per la prima volta in modo formale ed esplicito la logica intuizionista viene caratterizzata *come una restrizione della forma delle dimostrazioni* e non come una scelta di principi.

Dobbiamo alla logica lineare (introdotta da Jean-Yves Girard in [Gir87]) la scoperta del fatto che questo fenomeno non è accidentale: le regole strutturali determinano buona parte del comportamento del calcolo e restringendone l'uso in vari modi si possono ottenere varie famiglie di sistemi. Basti pensare al fatto che la contrazione esprime l'idempotenza della congiunzione e della disgiunzione ($A \wedge A \sim A \vee A \sim A$), mentre la regola di scambio ne esprime la proprietà commutativa ($A \wedge B \sim B \wedge A$); si pensi anche al fatto che l'unico assioma del calcolo dei sequenti è l'identità $A \vdash A$ che è valida praticamente in ogni sistema logico, mentre la legge del terzo escluso diventa derivabile non appena si consenta l'uso della contrazione a destra del sequente. In questo senso la struttura della dimostrazione diventa primitiva rispetto alla scelta dei principi logici.

1.3 Dimostrazioni e semantica

Nei prossimi paragrafi vogliamo presentare brevemente alcuni dei risultati più interessanti della teoria della dimostrazione, con un occhio rivolto anche alle possibili letture filosofiche.

1.3.1 Consistenza "senza semantica"

Come corollario del teorema di eliminazione del taglio si può ottenere un risultato di correttezza puramente interno al calcolo logico, senza riferimento a una interpretazione esterna come la semantica classica tarskiana.

Chiamiamo un sistema di regole *consistente* quando per ogni formula A è impossibile derivare sia A che la sua refutazione. Supponiamo che esista una coppia di derivazioni (chiuse o pure, cioè senza assunzioni) di A e $\neg A$. La regola di eliminazione della negazione ($\neg E$) in deduzione naturale e quella di taglio nel calcolo dei sequenti permettono di derivare il *falso* (o il sequente vuoto):

$$\dfrac{\begin{matrix}\vdots \\ \neg A\end{matrix} \qquad \begin{matrix}\vdots \\ A\end{matrix}}{\bot}\,\neg E \qquad\qquad \dfrac{\begin{matrix}\vdots \\ \vdash \neg A\end{matrix} \qquad \dfrac{\begin{matrix}\vdots \\ \vdash A\end{matrix}}{\neg A \vdash}\,(\neg L)}{\vdash}\,(cut)$$

Ora dovrebbe essere possibile applicare lo *Hauptsatz* a entrambe le derivazioni e ottenere così una derivazione chiusa e normale del falso in deduzione naturale, o una derivazione pura e senza tagli del sequente vuoto nel calcolo dei sequenti. Sappiamo però che:

1. ogni derivazione chiusa e normale in deduzione naturale termina con una regola di introduzione, ma non esiste alcuna regola che introduca il falso;

2. per la *proprietà della sottoformula*, ogni occorrenza di formula in una derivazione senza tagli deve essere sottoformula di qualche occorrenza nella conclusione: ma la conclusione è il sequente vuoto, dunque la derivazione non può contenere alcuna formula, il che è assurdo;

da cui possiamo concludere che:

a. in entrambi i sistemi, non esiste una derivazione chiusa (o pura) e normale del falso (o del sequente vuoto) e di conseguenza

b. se una formula A è derivabile, $\neg A$ non è derivabile e viceversa: *sia la derivazione naturale che il calcolo dei sequenti sono consistenti*.

L'importanza del risultato non giace semplicemente nella possibilità di fare a meno di nozioni "esterne" sull'interpretazione del linguaggio logico; piuttosto tale indipendenza suggerisce che alcuni sistemi di regole logiche godano di proprietà interessanti che possono essere studiate indipendentemente dalla scelta di una semantica per le proposizioni.

Se poi ricordiamo che il requisito di correttezza (rispetto alla semantica delle proposizioni) era una strategia per selezionare insiemi di regole "interessanti" e non triviali, è naturale chiedersi se generalizzando il risultato si possa ottenere una strategia alternativa a questo scopo. Come vedremo nelle prossime righe, la risposta è positiva.

1.3.2 Semantica *con le dimostrazioni*[25]

L'ambito di ricerca denominato *proof-theoretic semantics* ruota intorno all'idea di studiare la semantica dei linguaggi logici a partire dai sistemi deduttivi. Sono due i punti di partenza storici di questo programma:

1. l'interpretazione BHK[26] della logica intuizionistica: avendo sostituito la verità con la dimostrabilità, l'idea è quella di interpretare la congiunzione $A \wedge B$ come la proposizione dimostrabile da una coppia di derivazioni rispettivamente di A e di B, l'implicazione $A \to B$ come la conclusione dimostrabile da una procedura costruttiva per trasformare ogni derivazione di A in una derivazione di B, e così via;

2. un'osservazione di Gentzen in [Gen35] che sottolinea come il senso dei connettivi logici sia intuitivamente "determinato" dalle loro regole di introduzione ed eliminazione; osserva inoltre che sembra esistere una certa relazione fra introduzioni ed eliminazioni, tale che specificando solamente le introduzioni si potrebbe giustificare ogni eliminazione ammissibile (dunque determinare il senso del connettivo, cfr. [Sch18]).

È facile vedere che esiste un'analogia fra le due idee: se pensiamo alla regola $\wedge I$ della deduzione naturale, esprime precisamente il fatto che i due congiunti siano derivabili allo stesso tempo e indipendentemente:

$$\dfrac{\begin{matrix}\vdots \\ A\end{matrix} \quad \begin{matrix}\vdots \\ B\end{matrix}}{A \wedge B}\ \wedge I$$

Allo stesso modo la regola per l'implicazione "racchiude" (nascondendo l'assunzione A) un frammento di derivazione che può essere composto con una derivazione

[25]Scegliamo di rendere così l'espressione *proof-theoretic semantics*, per distinguere questo ambito da quello della semantica *delle* dimostrazioni (*proof semantics*).

[26]Brouwer-Heyting-Kolmogorov. L'idea è ben più complessa di quanto descritto qui; la riprenderemo con più precisione nel paragrafo sulla semantica delle dimostrazioni.

qualsiasi di A per ottenere una derivazione di B:

$$\dfrac{\begin{matrix}[A] \\ \vdots \\ B\end{matrix}}{A \to B} \to I$$

Osserviamo che è possibile interpretare ogni formula come l'espressione di qualche proprietà comune a tutte le derivazioni che la dimostrano: dunque il significato della congiunzione e degli altri connettivi non andrebbe cercato in una 'funzione di verità' *à la Boole* o in una definizione *à la Tarski*, quanto piuttosto nella struttura dei ragionamenti di cui rappresentano le conclusioni.

Michael Dummett si è fatto interprete filosofico di questo approccio sotto due punti di vista. Primo, identificando nella scelta di una concezione della verità (classica o intuizionista) una presa di posizione metafisica (realista o antirealista) rispetto all'oggetto del discorso [Dum91]. Potremmo dire che la scelta classica mette al primo posto il modello della realtà, studiato con l'appoggio di una teoria logica, mentre la scelta intuizionista mette al primo posto il processo conoscitivo, rappresentato dalla struttura delle dimostrazioni, e contempla quindi la "terza posizione" dell'ignoranza (abbiamo visto però che definirla 'terza' è inesatto: si tratta più che altro di una "non posizione").

Secondo, osservando che comunque si scelga sul concetto di verità, l'asserzione e il giudizio condividono nei due casi una struttura di fondo che è precisamente quella della dimostrazione. Asserisco *A-e-B* quando sono pronto ad asserire sia A che B, viceversa se ascolto qualcuno asserire *A-e-B*, mi aspetto che sia in grado indifferentemente di asserire A o B.[27]

Torniamo dunque al punto di prima: il significato delle espressioni linguistiche per Dummett deve essere cercato nella struttura della loro *asseribilità*. La tradizione è chiaramente quella di matrice Wittgensteiniana del "significato come uso" (*meaning as use*) e potremmo vedere la posizione di Dummett come una precisazione dell'idea più generale. Un ulteriore passaggio è compiuto dal filosofo in *Truth* [Dum58] con il suggerimento che la verità sia in primo luogo una proprietà delle asserzioni o giudizi (intesi come atti) e solamente in secondo luogo delle proposizioni. Possiamo leggere questa posizione come una variante dell'approccio intuizionista e trarne una formulazione alternativa dell'approccio classico: la verità è in primo luogo una proprietà delle proposizioni. È interessante notare che entrambe le posizioni non sarebbero estranee al pensiero aristotelico: se è vero che le proposizioni categoriche possono essere considerate vere o false in base alla loro corrispondenza o meno con cio che è, d'altra parte è anche vero che, secondo il Filosofo, la verità consiste nel «dire che l'ente è e il non

[27] L'uso della disgiunzione "o" in quest'ultimo passaggio non è accidentale: dipende dal fatto, meno visibile in deduzione naturale ma evidente nel calcolo dei sequenti, che esiste una dualità fra congiunzione e disgiunzione espressa anche nella profonda simmetria fra introduzioni ed eliminazioni.

ente non è» (*Metaph.* Γ 7, 1011a23–24), che suona come un riferimento all'atto del pronunciarsi. Anche nell'interpretazione medievale di Tommaso d'Aquino, la verità è predicata in primo luogo dell'intelletto nell'atto del giudicare, e in secondo luogo delle proposizioni in quanto potenziali oggetti di giudizio; bisogna dire che Dummett è più radicale: propone l'idea dal sapore pragmatista che una affermazione sia falsa quando una persona che ne conosce l'uso ritiene che affermarla costituirebbe una menzogna [Dum58].

La precisazione tecnica di queste idee risale agli anni '60 e '70 del XX secolo, specialmente ad opera di Dag Prawitz. Non possiamo entrare in dettagli, ma vogliamo soffermarci su un'obiezione fondamentale a questo approccio e sulla risposta a tale obiezione.

Se le regole di introduzione ed eliminazione sono sufficienti a definire un connettivo, cosa mi impedisce di selezionare un insieme qualsiasi di regole e definire connettivi a mio piacimento? In particolare, cosa mi impedisce di definire connettivi triviali, che permettono di derivare il falso o qualsiasi proposizione?[28]

La risposta giace nel cosiddetto principio di *armonia logica*, o *principio di inversione*, enunciato in [Pra65]. Non qualsiasi insieme di introduzioni ed eliminazioni definisce un connettivo valido, ma solamente quelli che soddisfano la condizione seguente:

> Se componendo una regola di introduzione α e una di eliminazione β (dello stesso connettivo) è possibile derivare la conclusione *B*, allora componendo le sole derivazioni delle premesse è possibile derivare *B* *senza aggiungere* α *e* β.

In buona sostanza, ricordando la terminologia di Gentzen, la condizione chiede che sia possibile eliminare la "digressione" formata da introduzione ed eliminazione dello stesso connettivo con una procedura simile a quella descritta da Gentzen: in altre parole, il principio di armonia chiede che sia possibile *l'eliminazione del taglio*! Come abbiamo visto questo produce immediatamente un sistema consistente.

Un'obiezione più forte che si può muovere a questo approccio è invece la seguente: sembra che non si faccia altro che riprendere le definizioni di Tarksi, sostituendo il termine 'vero' con 'derivabile'. Avremo dunque definizioni del tipo *A-e-B* è derivabile sse *A* è derivabile e *B* è derivabile. Qual è il vantaggio?

L'obiezione ci sembra leggittima; vogliamo tentare una risposta seguendo un brillante articolo di Popper (il già citato *Logic without assumptions* [Pop47]), in cui il filosofo propone e analizza un sistema inferenziale molto simile al calcolo dei sequenti di Gentzen (è sorprendente il fatto, osservato in [Sch18], che Popper fosse all'epoca quasi certamente ignaro degli studi di Gentzen).

Il problema della semantica tarskiana, secondo Popper riconosciuto dallo stesso Tarski, è la dipendenza da una nozione di *segno formativo*, in altre parole dalla necessità di scegliere le costanti logiche: questa scelta sarà in generale

[28] È famoso il connettivo *tonk* proposto in chiave polemica in [Pri60].

arbitraria. La soluzione proposta consiste nella ricerca di regole di inferenza che prescindano completamente dalla struttura delle proposizioni[29] e più in generale dalla scelta del linguaggio, e che siano valide in modo assoluto (bisogna tenere a mente che per Popper un linguaggio è sempre accompagnato da una semantica propria, diversamente dall'approccio formale comune che separa i due concetti).

Si comincia con due regole basilari:

$$A \vdash A;$$
$$\textit{se } A \vdash B \textit{ e } B \vdash C \textit{ allora } A \vdash C.$$

Adottiamo la notazione dei sequenti in quanto sostanzialmente equivalente a quella introdotta nell'articolo. Evidentemente le due regole corrispondono all'assioma identità e alla regola di taglio, e chiaramente la loro correttezza non dipende in alcun modo dal contenuto delle proposizioni, ma esclusivamente dalla riflessività e transitività della relazione di derivabilità ($\vdash$).

Popper procede quindi a definire una serie di regole per i vari connettivi; consideriamo come sempre l'esempio della congiunzione:

> *A esprime la congiunzione di B e C sse vale* $A \vdash B$, $A \vdash C$, $B, C \vdash A$.

Fin qui niente di strano, non è difficile riconoscere nella regola una formulazione alternativa delle regole di deduzione naturale o del calcolo dei sequenti: la differenza è che si evita di fare riferimento al simbolo $\wedge$ e alla struttura delle proposizioni. È invece estremamente interessante il passaggio successivo:

> *Il linguaggio $\mathcal{L}$ contiene un* segno formativo per la congiunzione *se per qualche segno $\diamond$ vale sempre che $B \diamond C$* esprime la congiunzione di B e C.

dove con $B \diamond C$ intendiamo una struttura sintattica qualsiasi che utilizza il segno "$\diamond$" (siamo liberi di sbizzarrirci: $\diamond BC$, $B!\diamond!C$, *etc.*; nelle intenzioni di Popper la definizione si potrebbe anche applicare a una lingua naturale come l'Italiano).

A nostro avviso questo approccio spiega con eleganza il senso delle definizioni inferenziali dei connettivi logici. Per essere precisi, non è il *segno o connettivo linguistico* ad essere definito dalle regole, ma *il concetto stesso di congiunzione*; diremo quindi che la logica classica e la logica intuizionistica *contengono una congiunzione*. Non si vuole qui favorire la derivabilità contro la verità classica e tarskiana: sono concetti indipendenti ed entrambi fecondi in modo diverso e ci sembra giusto accogliere l'obiezione sotto questo aspetto. Si tratta piuttosto di stabilire una *priorità delle strutture del ragionamento* nel determinare il significato dei *concetti* logici: vogliamo *concepire* la congiunzione come una particolare modalità del ragionamento piuttosto che come una funzione fra valori di verità.

Il collegamento delle forme di inferenza con la scelta di una semantica delle proposizioni può naturalmente avvenire in un secondo momento: non si tratta

[29] Fatte le dovute precisazioni, l'approccio di Popper avrebbe in questa scelta degli interessanti punti di contatto con quello aristotelico.

di un'operazione di poco conto, come ha sottolineato Dummett. Questo rimane possibile grazie ai già menzionati risultati di correttezza e completezza, che a questo punto non hanno più il compito di *giustificare* le forme del ragionamento quanto piuttosto quello di valutarne l'adeguatezza alla scelta semantica.

Sempre a Dummett è dovuta l'importante osservazione che i connettivi possono essere definiti anche a partire dalle regole di eliminazione invece che dalle introduzioni, che possono occupare un ruolo secondario. Dal punto di vista tecnico i due approcci sono equivalenti (si deve attendere l'introduzione della logica lineare perché queste differenze possano essere espresse formalmente). Esiste tuttavia una differenza di carattere "morale": il primo approccio può essre considerato *verificazionista*, il secondo invece *pragmatista*.[30] Per comprendere il senso di questa affermazione esaminiamo ancora una volta il caso della congiunzione:

– la regola di introduzione esprime il fatto che siamo in possesso di dimostrazioni per entrambe le premesse: abbiamo dunque "verificato" la validità della congiunzione;

– la regola di eliminazione invece esprime il fatto che una derivazione di $A \wedge B$ può essere utilizzata indifferentemente come se fosse una derivazione di A o una derivazione di B.

Il primo approccio dunque è *verificazionista* nel senso che indica come costruire correttamente una dimostrazione della congiunzione e considera questo aspetto prioritario nella definizione, mentre il secondo è *pragmatista* nel senso che indica come *utilizzare* correttamente una dimostrazione della congiunzione (e di conseguenza inverte le priorità). La distinzione sarà resa ancora più chiara dall'analogia fra dimostrazioni e programmi che stiamo per presentare.

1.3.3 Semantica *delle dimostrazioni*

Abbiamo già accennato all'interpretazione BHK della logica intuizionistica. Sebbene comprenda anche l'aspetto dell'interpretazione delle proposizioni, questo non è certamente primario: il punto cruciale è invece l'interpretazione delle *dimostrazioni*. Che cos'è, o cosa significa, una dimostrazione? Che cosa distingue due dimostrazioni diverse, e in che senso possono dirsi "diverse"? Cosa accomuna due dimostrazioni equivalenti?

La proposta di Heyting è la seguente (omettiamo per semplicità i casi della disgiunzione e del quantificatore esistenziale):

– una dimostrazione di $A \wedge B$ è una coppia (a, b) dove a è una dimostrazione di A e b è una dimostrazione di B;

– una dimostrazione di $A \to B$ è una *funzione* f che associa a ogni dimostrazione di A una dimostrazione di B e che sia calcolabile effettivamente;

[30] Il riferimento è sempre [Dum91]; per un'interessante analisi tecnica e formale di questa idea si veda invece [Zei09].

– una dimostrazione di $\forall x.A(x)$ è una funzione f che associa a ogni oggetto x del dominio una dimostrazione di $A(x)$.

Due dimostrazioni della stessa formula saranno allora equivalenti se sono la stessa funzione o la stessa coppia. L'originalità della proposta sta nella capacità di spiegare l'essenza di una dimostrazione attraverso un oggetto matematico più primitivo, come una funzione; la formulazione incappa però in due difficoltà: il dominio di definizione delle funzioni menzionate non è affatto chiaro, così come non è chiaro in che senso una dimostrazione chiusa come la seguente, che si presenta come una coppia di dimostrazioni di un'implicazione e di A, sia invece una coppia di dimostrazioni di B e di C:

$$\dfrac{\begin{array}{c}\vdots\\ A \to B \wedge C\end{array} \qquad \begin{array}{c}\vdots\\ A\end{array}}{B \wedge C} \to E$$

Anticipando quanto diremo a breve, la risposta si trova nel processo di normalizzazione e nelle proprietà della forma normale: ricordiamo che normalizzando la derivazione appena esibita ne otterremo una che termina con una regola di introduzione per $B \wedge C$, le cui premesse saranno precisamente una dimostrazione di B e una di C. In questo modo ci allontaniamo dallo studio "statico" della struttura delle dimostrazioni e portiamo al centro dell'attenzione la dinamica delle procedure di riduzione e dell'interazione fra dimostrazioni.

Gli strumenti per chiarire e concretizzare l'approccio BHK sono stati forniti dalla scoperta dell'analogia fra dimostrazioni e programmi, nota dal punto di vista formale come *isomorfismo Curry-Howard*. L'idea si sviluppa sullo sfondo del λ-calcolo, un formalismo per la rappresentazione e lo studio di programmi introdotto da Alonzo Church negli anni '30 del secolo scorso: ne presentiamo rapidamente gli elementi essenziali seguendo a grandi linee [GLT89].

Il λ-calcolo è composto da una serie di operatori per costruire *termini* e da un insieme di passi di riduzione applicabili a termini di forma particolare (*redex*). Supponiamo che t, u siano termini, allora:

– x è un termine, dove la lettera x indica un qualsiasi nome di variabile;

– $\lambda x.t$ è l'operatore di *astrazione*, che costruisce una funzione nella variabile x;

– (tu) è l'operatore di *applicazione*, che per l'appunto applica il parametro u a una funzione t.

Il passo di β-riduzione è definito come segue:

$$(\lambda x.t)u \rightsquigarrow_\beta t[u/x]$$

dove la notazione $t[u/x]$ indica che abbiamo sostituito con u ogni occorrenza della variabile x in t. Il passo si applica a qualsiasi termine o sottotermine che abbia la forma indicata a sinistra della freccia, che chiameremo *redex*. Un termine che non contiene redex sarà *in forma normale*. Osserviamo che questa regola di riscrittura intende riprodurre l'operazione di calcolo di una funzione. Immaginiamo che il termine $\lambda x.t$ rappresenti una funzione così definita:

$$t(x) = x + 2$$

supponiamo inoltre che il termine u sia uguale a 3. La regola di β-riduzione allora esprime intuitivamente il fatto che

$$t(u) = (x + 2)[u/x] = u + 2 = 3 + 2.$$

Con questo vogliamo solamente offrire un'intuizione sul funzionamento del formalismo: le notazioni dell'aritmetica non fanno parte del λ-calcolo puro (sebbene possano essere aggiunte con opportune estensioni). È notevole tuttavia il fatto che i soli operatori riportati sopra siano sufficienti, con la regola di β-riduzione, a definire un sistema di calcolo equivalente alla macchina di Turing.

È naturale porsi alcune domande sui termini del calcolo: ogni termine ha una forma normale? Se esiste una forma normale, è unica? Ogni strategia di riduzione raggiunge una forma normale?

Solamente la seconda domanda ha risposta positiva: deriva dalla proprietà conosciuta come *confluenza*. È possibile allora caratterizzare il sottoinsieme dei termini fortemente normalizzanti (quelli per cui la risposta alla terza domanda è positiva)? Sappiamo che questo non è possibile in generale perché la domanda equivale al problema dell'arresto;[31] Church tuttavia riuscì a caratterizzare un sottoinsieme dei termini fortemente normalizzanti utilizzando la nozione di *tipo*:

- la variabile x può avere un tipo qualsiasi α;
- l'astrazione $\lambda x.t$ ha tipo $\alpha \to \beta$ dove α è il tipo di x e β è il tipo di t;
- l'applicazione (tu) ha tipo β se t è di tipo $\alpha \to \beta$ e u è di tipo α.

L'osservazione di Church è che, se costruiamo un termine qualsiasi rispettando queste regole (in particolare rispettando le condizioni imposte sull'applicazione), otteniamo certamente un termine fortemente normalizzante. Parliamo in tal caso di termini *tipati*.

Fu poi H. Curry nel 1958 [CF58] ad osservare che, se interpretiamo i tipi come formule logiche (in particolare $\alpha \to \beta$ come un'implicazione), allora esiste qualche termine t di tipo α se e solo se α è una tautologia valida intuizionisticamente: questo suggerisce che un programma di tipo α equivalga a qualche dimostrazione intuizionistica della proposizione α. Infine W. A. Howard nel 1969

[31] Cioè il problema di decidere, dato un programma e un input qualsiasi, se l'esecuzione del programma avrà termine (se cioè si arresterà). L'insolubilità di questo problema nel caso generale è un risultato fondamentale nella teoria della calcolabilità.

osservò che i termini del λ-calcolo semplicemente tipato potevano essere messi in corrispondenza diretta con le derivazioni in deduzione naturale intuizionistica.[32]

La corrispondenza si ottiene trattando le variabili come assunzioni, le astrazioni come regole $\to I$ e le applicazioni come regole $\to E$. È facile verificare che la struttura coincide:

$$\dfrac{\begin{array}{c}[x:A]\\ \vdots\\ t:B\end{array}}{\lambda x.t : A\to B}\to I \qquad \dfrac{t:A\to B \qquad u:A}{(tu):B}\to E$$

la corrispondenza può essere estesa agli altri connettivi aggiungendo al λ-calcolo operatori e riduzioni opportune. Ciò che è più sorprendente – ed è il motivo per cui parliamo di *isomorfismo* – è che anche la regola di β-riduzione coincide esattamente con le riduzioni di Gentzen:

$$\dfrac{\dfrac{\begin{array}{c}[x:A]\\ \vdots\\ t:B\end{array}}{\lambda x.t:A\to B}\to I \qquad \begin{array}{c}\vdots\\ u:A\end{array}}{(\lambda x.t)u:B}\to E \qquad \rightsquigarrow \qquad \begin{array}{c}\vdots\\ u/x:A\\ \vdots\\ t[u/x]:B\end{array}$$

in particolare, l'operazione di sostituzione del λ-calcolo corrisponde alla composizione di derivazioni operata da Gentzen e i *redex* del λ-calcolo corrispondono alle "divagazioni": introduzioni seguite da eleminazioni. Una conseguenza immediata dell'isomorfismo è che il processo di riduzione di Gentzen-Prawitz è confluente e fortemente normalizzante.

La scoperta dell'isomorfismo Curry-Howard permette di concretizzare l'idea di BHK risolvendo le due difficoltà menzionate prima: il dominio della funzione dalle dimostrazioni di A a dimostrazioni di B sarà quello dei λ-termini di tipo A; inoltre, come avevamo anticipato, una dimostrazione di $A\to B$ è interpretata come una funzione (o una dimostrazione di $A \wedge B$ come una coppia) con riferimento alla forma normale che, grazie all'isomorfismo, è nota essere esistente e unica. Possiamo dunque conferire un significato nuovo alla consistenza di una dimostrazione: ricordando il risultato "interno" di consistenza ottenuto grazie all'eliminazione del taglio, potremo dire che la caratteristica di una dimostrazione corretta è quella di rappresentare una certa funzione o algoritmo; in altre parole, una dimostrazione corretta "ha un contenuto".

Quest'ultima osservazione può essere raffinata fino allo sviluppo di una *semantica denotazionale* per le dimostrazioni. Si tratta di un concetto ben noto e ampiamente studiato nell'ambito dell'informatica teorica e non è possibile presentarlo in dettaglio: cercheremo di abbozzare una spiegazione tenendo a mente che la definizione formale dovrà essere molto più sofisticata.

[32][How80], l'articolo fu pubblicato molto più tardi.

L'idea è di associare a ogni formula atomica A un *dominio* $[\![A]\!]$ di "oggetti di tipo A". $[\![A \wedge B]\!]$ denoterà allora il dominio delle coppie $\langle a, b \rangle$ dove $a \in [\![A]\!]$ e $b \in [\![B]\!]$; $[\![A \to B]\!]$ denoterà invece il dominio delle *funzioni* da $[\![A]\!]$ in $[\![B]\!]$, dove il termine *funzione* è inteso nel senso più comune, senza riferimento al λ-calcolo.

Come ultimo passo definiamo una interpretazione *delle derivazioni*, cioè una funzione che associa ad ogni derivazione t di conclusione A un oggetto nel dominio $[\![A]\!]$. Ancora una volta ritroviamo l'idea di associare una coppia a ogni derivazione della congiunzione e una funzione a ogni derivazione dell'implicazione; il passaggio cruciale tuttavia consiste nel chiedere che l'interpretazione di una derivazione sia *invariante sotto riduzione*: se $t \rightsquigarrow t'$ allora $[\![t]\!] = [\![t']\!]$. Diremo che la derivazione t *denota* un oggetto nel dominio A.

Una costruzione di questo genere è naturalmente candidata a rispondere alle domande poste in apertura del paragrafo: cosa rende due dimostrazioni equivalenti? Rispondiamo: t_1 e t_2 sono equivalenti se *denotano lo stesso oggetto* e cioè se vale $[\![t_1]\!] = [\![t_2]\!]$, sono diverse invece se denotano oggetti diversi. Ci sono tre differenze fra questo approccio approccio e quello meno raffinato basato sulle forme normali: *primo*, poiché abbiamo ottenuto un invariante della riduzione, siamo in grado di studiare i processi di riduzione in maniera più raffinata; *secondo*, è possibile stabilire l'equivalenza di forme normali superficialmente diverse: è noto che esistono molti algoritmi diversi che calcolano *la stessa funzione* da A in B e ognuno di essi corrisponderà in generale a una diversa dimostrazione normale di $A \to B$, ma la loro *denotazione* sarà la stessa; *terzo*, le dimostrazioni possono essere trattate come elementi di un dominio che (nella definizione più rigorosa) avrà la struttura di uno *spazio* e potrà essere studiato con i metodi dell'algebra e della topologia.

Come promesso, possiamo ora presentare in forma più evidente la distinzione di Dummett fra approccio *verificazionista* e approccio *pragmatista*. Il primo è in relazione con lo stile di Church, in cui le regole di *typing* sono viste come una guida alla costruzione dei termini: se costruisco un termine rispettando determinate regole, allora godrà di determinate proprietà e denoterà un oggetto nel dominio del suo tipo. Il secondo invece corrisponde allo stile di Curry: se posso *utilizzare* un termine qualsiasi secondo determinate regole, allora può essere tipato dalla formula che esprime questa possibilità. Per esempio, se è possibile interpretare un termine qualsiasi come *tipato* dalla formula $A \to B$, allora denoterà effettivamente una funzione da $[\![A]\!]$ in $[\![B]\!]$ e *si comporterà correttamente* se applicato ad un secondo termine di tipo A (se cioè operiamo un'eliminazione dell'implicazione), dove "correttamente" significa: si comporterà come un termine fortemente normalizzante di tipo B.

È anche possibile arricchire l'idea della definizione *inferenziale* dei connettivi. Ricordiamo che la proposta era di estrarre il senso di un connettivo dalle sue regole di introduzione ed eliminazione. Per quanto l'approccio potesse essere interessante, si obiettava che lo sforzo non sembra giustificabile, visto che l'unico risultato sembra quello di guadagnare l'indipendenza dalle definizioni tarskiane senza ulteriori approfondimenti. A nostro avviso è proprio il discorso appena concluso che giustifica lo sforzo inferenzialista.

L'isomorfismo Curry-Howard e la semantica denotazionale forniscono un modo differente e più concreto (più che intuitivo o puramente formale) di pensare il contenuto delle proposizioni logiche. Dal punto di vista verificazionista esse denotano una classe di atti di pensiero che è necessario compiere per dimostrarle; viste come conclusioni o *risposte* indicano che un particolare atto è stato compiuto: nel caso della congiunzione, ad esempio, l'avvenuta fusione di due linee di ragionamento indipendenti. Dal punto di vista pragmatista invece denotano una classe di atti di pensiero *possibili* o *disponibili* una volta che siano dimostrate; viste come premesse o *domande* indicano che la disponibilità di un particolare atto è richiesta per completare un ragionamento: nel caso della congiunzione la possibilità di recuperare indipendentemente una linea di ragionamento per ogni congiunto.

I due approcci sono chiaramente complementari e i processi di normalizzazione offrono una cornice formale per lo studio dell'interazione fra domanda e risposta, produzione e consumo, progettazione e verifica. Vogliamo suggerire qui l'idea che la logica, intesa come scienza delle dimostrazioni, fornisca gli strumenti per studiare in maniera astratta ed estremamente generale alcuni fenomeni che si presentano in ogni ambito delle attività umane, e non solo nel ragionamento.

Chiudiamo il discorso osservando che l'approccio appena discusso mette al centro del discorso logico i processi epistemici, l'intensionalità e in un certo senso il "soggetto" del discorso,[33] laddove l'approccio della teoria dei modelli mette al centro la descrizione di uno stato di cose, l'estensionalità e l'oggetto del discorso. Ancora una volta possiamo constatare che le due tradizioni non si escludono a vicenda e sembra giusto anzi considerarle complementari; d'altra parte questo permette di comprendere meglio lo sforzo di separare i sistemi deduttivi dalle semantiche estensionali.

1.4 La situazione della logica classica

Il discorso appena concluso si applica senza alcuna difficoltà alla logica intuizionistica, che in un certo senso era stata concepita "apposta per questo". Non appena si voglia passare agli schemi di ragionamento classico, però, si incappa in numerose difficoltà. Cos'è che non funziona?

Esistono vari modi per estendere il calcolo di deduzione naturale alla logica classica. Uno di questi è l'aggiunta di una regola *ad hoc* per le doppie negazioni:

$$\frac{\neg\neg A}{A}\,\mathrm{DNE}$$

[33] Alcune linee di riflessione filosofica su questi temi sono state sviluppate in vari scritti di J.-Y. Girard.

Il difetto di questa regola è che viola proprio quei principi di *armonia* che garantiscono l'eliminabilità delle "divagazioni" (o redex): l'eliminazione del taglio diventa estremamente difficile (la definizione si complica e richiede passaggi *ad hoc*) e per lungo tempo si è creduto che non fosse possibile associare a questa regola un operatore in qualche forma del λ-calcolo. Esiste certamente una connessione fra questa difficoltà e quelle che spinsero Gentzen a introdurre il calcolo dei sequenti.[34]

La difficoltà diventa più evidente considerando il contenuto della regola: essa afferma che è possibile considerare una *refutazione* di $\neg A$ come una *dimostrazione* di A e questo fa saltare l'idea che una dimostrazione di A denoti un oggetto di tipo A. In generale la *refutazione* di $\neg A$ denoterà una funzione che trasforma ogni dimostrazione di $\neg A$ in una dimostrazione del *falso* (da cui deduciamo che *non può esistere una dimostrazione di* $\neg A$ e dunque vale $\neg\neg A$): sarà dunque una funzione da $[\![\neg A]\!]$ in $[\![\bot]\!]$, che evidentemente non appartiene al dominio $[\![A]\!]$.

Un primo passo verso una soluzione del problema è stato compiuto da T. G. Griffin in [Gri90]. Una derivazione che contiene la regola DNE può essere trasformata come segue:

$$\begin{array}{c}[\neg A]\\ \vdots\,\mathrm{p}\\ \dfrac{\dfrac{\bot}{\neg\neg A}\,\neg I}{A}\,\mathrm{DNE}\\ \vdots\,\mathrm{q}\\ B\end{array} \quad\rightsquigarrow\quad \begin{array}{c} \dfrac{\dfrac{[\neg B]\qquad \begin{array}{c}[A]\\ \vdots\,\mathrm{q}\\ B\end{array}}{\bot}\,\neg E}{\neg A}\,\neg I \\ \vdots\,\mathrm{p}\\ \dfrac{\dfrac{\bot}{\neg\neg B}\,\neg I}{B}\,\mathrm{DNE}\end{array}$$

In buona sostanza abbiamo trasformato una derivazione *classica* di B in una derivazione *intuizionista* di $\neg\neg B$, quindi abbiamo applicato la regola DNE per ottenere B. Due sono le osservazioni di Griffin:

1. selezionando particolari strategie di riduzione, e restringendo in certi modi la forma delle derivazioni, è possibile associare alla regola DNE l'operatore `call/cc` di Scheme (un linguaggio di programmazione simile al λ-calcolo): si tratta di un *operatore di controllo*, cioè un operatore che modifica il flusso di esecuzione del programma;

2. applicando una serie di trasformazioni è possibile trasformare ogni derivazione classica in una derivazione completamente intuizionistica seguita dall'applicazione di una singola regola DNE.

Osserviamo in particolare che le regole per la negazione classica permettono di scambiare arbitrariamente conclusioni e premesse (è questo il contenuto della

[34] L'approccio di Gentzen era leggermente diverso e consisteva nell'aggiungere il principio $A \vee \neg A$ come assioma alla deduzione naturale: si presentano comunque difficoltà simili

negazione *lineare*, che classicamente è equivalente a quella intuizionistica). Dal punto di vista algoritmico, questo significa che esiste una completa simmetria fra *input* e *output* di un programma. Seguendo il secondo punto, possiamo trattare ogni derivazione classica come una derivazione intuizionistica del falso in cui la conclusione classica figura negata fra le premesse: in ogni momento sarà possibile estrarre la conclusione classica utilizzando la regola DNE.

L'approccio di Griffin non manca di difetti e conserva in buona sostanza un carattere *ad hoc*. Ciò nonostante le sue osservazioni hanno ispirato tutti gli approcci successivi, come il calcolo λ_c di Krivine o il $\lambda\mu$-calcolo di Parigot [Par92b]. In entrambi i casi si aggiungono operatori di controllo; l'approccio di Parigot consiste inoltre nel distinguere nettamente premesse e conclusioni (rompendo la simmetria input-output) e identificare una conclusione classica "principale" che può essere scambiata con le premesse utilizzando gli operatori per la negazione.

Diverso, ma simile in spirito, è l'approccio utilizzato dal sistema **LC** di Girard [Gir91] che, imponendo restrizioni sulla forma dei sequenti e delle formule, riesce a costruire derivazioni classiche in cui in ogni momento è possibile identificare una conclusione "intuizionistica" (per essere precisi sarà una conclusione *lineare*) e su questa base tradurre le derivazioni nel calcolo intuizionista (o lineare) preservando i processi di riduzione.

L'impressione però è che non si facciano grandi passi nella comprensione delle caratteristiche specifiche del ragionamento classico. Un approccio diverso può essere tentato partendo dal calcolo dei sequenti. Ricordiamo che la caratteristica fondamentale del calcolo classico **LK** è quella di utilizzare sequenti multi-conclusione, dove le varie conclusioni si intendono disgiunte. Questo fa sì che la disgiunzione classica acquisisca un contenuto differente da quella intuizionista.

$$\frac{\vdash A}{\vdash A \vee B}\,(\vee_a^1) \qquad \frac{\vdash B}{\vdash A \vee B}\,(\vee_a^2) \qquad \frac{\vdash A, B}{\vdash A \vee B}\,(\vee_m)$$

La disgiunzione intuizionista, caratterizzata dalle prime due regole, denota una scelta effettuata fra una dimostrazione di A e una di B. Nel caso classico si aggiunge una terza possibilità ed è evidente che nella terza regola non avviene alcuna scelta.

Piuttosto, possiamo interpretare una derivazione di conclusione $\vdash A, B$ come una procedura "sospesa" in attesa di ulteriori informazioni: per raggiungere una conclusione "definitiva", è necessario eliminare uno dei due rami della disgiunzione. L'informazione mancante è allora una refutazione di A o una refutazione di B e la disgiunzione classica può essere interpretata come una sorta di "transazione": posso offrire una prova di B a chi offre una refutazione di A; d'altra parte input e output sono perfettamente simmetrici, perciò esistono altri tre modi di interpretarla, ad esempio "posso offrire una prova di A a chiunque offra una refutazione di B".

Questa analisi diventa molto significativa nel caso di una derivazione strettamente non intuizionistica. Ricordando che la differenza sta nel permettere o meno la contrazione sulla destra del sequente, tali derivazioni avranno in generale la forma seguente:

$$\frac{\vdash A, A}{\vdash A}\,(C)$$

In che senso questa sarebbe una dimostrazione “per contraddizione”? Applichiamo l’interpretazione “transazionale”: l’informazione mancante per concludere A è…una refutazione di A! La contrazione sulla destra allora esprime il fatto *classico* che, se una refutazione di A permette di concludere contraddittoriamente A, tale refutazione è impossibile e possiamo ritenerci autorizzati a concludere A, sapendo che nessuno potrà chiamarci a rendere conto della nostra conclusione.[35]

Le difficoltà però non spariscono magicamente. La simmetria fra premesse e conclusioni, tipica del calcolo dei sequenti classico, fa sì che la procedura di normalizzazione diventi non deterministica: in molti passaggi è possibile compiere scelte che portano a forme normali molto differenti, e non c’è motivo di preferire una scelta all’altra.[36] Sebbene si conservi l’esistenza di una forma normale, vengono a mancare le proprietà di confluenza e unicità della forma normale. Inoltre, i passi di riduzione *locali* proposti da Gentzen permettono in alcuni casi di procedere all’infinito: si perde così la proprietà di normalizzazione forte e ogni speranza di una semantica denotazionale. Seguendo [DJS97], un primo passo per rimediare consiste nel *delocalizzare* alcuni passi di riduzione, rendendoli più simili a quelli della deduzione naturale. A questo punto si aprono due strade.

[35]Vogliamo sottolineare un interessante punto di contatto fra questa osservazione e un passaggio di *Metaph.* Γ 4 (1006a11–25) in cui Aristotele considera la possibilità di dimostrare il principio di *non contraddizione* “in modo confutatorio”. Di per sé, egli dice, il principio non è dimostrabile e non deve essere dimostrato. Supponiamo però di trovarci di fronte un avversario che lo neghi: il Filosofo ci avverte che è impossibile procedere a una *refutazione*, o riduzione all’assurdo, perché questo non costituirebbe un problema per qualcuno che accetta le contraddizioni. Piuttosto, dobbiamo attendere che l’avversario apra bocca: appena parlerà, se davvero intende dire qualcosa di sensato, utilizzerà di fatto il principio di non contraddizione, ammettendolo implicitamente. A noi non resta che rendere esplicita questa ammissione. Berti [Ari17] sottolinea nelle note che qui non si fa uso del terzo escluso, tuttavia la strategia dimostrativa “attendista” è del tutto analoga: aspettare la mossa dell’avversario, quindi “condurlo” ad affermare precisamente ciò che intende negare.

[36]Dal punto di vista di un essere umano che riduce manualmente si tratterà solamente di scegliere liberamente fra varie alternative: in questo senso si può dire che la procedura di riduzione è *sottodeterminata* dalle regole. Utilizziamo l’espressione “non determinismo” nel senso tecnico dell’informatica teorica, con riferimento a una macchina ipotetica (reale o teorica) che opera le riduzioni, mentre l’essere umano assume il punto di vista di un osservatore. Se le informazioni a disposizione dell’osservatore sono limitate allo stato iniziale e alle regole di riduzione, è impossibile prevedere lo stato finale della macchina, anche se è possibile (quando vale la normalizzazione forte) determinare l’insieme degli stati finali possibili. In particolare, se il processo di scelta della macchina fosse non deterministico in senso ontologico, allora indipendentemente dalla quantità di informazione posseduta non si potrebbe mai prevedere in anticipo lo stato finale, e tuttavia le regole sarebbero rispettate. In questo senso si dice che la riduzione non è deterministica.

Continuando a seguire [DJS97], possiamo osservare che non è necessario introdurre una rigida divisione fra premesse e conclusioni (input e output, come nella soluzione di Parigot) o addirittura identificare un singolo output, come nelle soluzioni di Girard. È sufficiente aggiungere un elemento d'informazione ad ogni formula e sottoformula (detto *colore*) che indica se la formula debba essere trattata come input o come output.

Consideriamo la formula $(A^q \wedge_m B^q)^t$, dove t e q sono i due colori possibili. Quando questa formula compare fra le conclusioni a destra di un sequente, sarà considerata come un *output* di tipo $A^q \wedge_m B^q$. I congiunti A^q e B^q invece saranno considerati come *input*. In generale, il colore t rispetta la posizione della formula nel sequente, mentre il colore q la inverte: le premesse colorate q diventano output e le conclusioni input. Rispettando le informazioni fornite dai colori, si ottiene un protocollo di riduzione confluente e fortemente normalizzante e si può costruire una semantica denotazionale traducendo le derivazioni in logica lineare classica. Un punto a favore di questo approccio è che le derivazioni "colorate" non sono immediatamente traducibili in derivazioni intuizionistiche, segno che non stiamo "trascurando" le dinamiche classiche.

Ci sarebbe però una seconda strada. È possibile che il non-determinismo dei processi di riduzione sia una caratteristica essenziale, e non accidentale, della logica classica? È possibile mettere in qualche modo "sotto controllo" il non-determinismo e studiare il sistema di Gentzen non ristretto con strumenti semantici? Questo è l'approccio scelto da Barbanera e Berardi in [BB96] con la proposta del *λ-calcolo simmetrico*: un sistema con tipi classici, non deterministico (quindi non confluente) ma fortemente normalizzante.

Il cuore di questa tesi è costituito da uno studio del sistema non deterministico ottenuto combinando la libertà del calcolo dei sequenti classico **LK** di Gentzen con i passi di riduzione *non locali* di [DJS97]. Si applica a questo sistema la tecnica dimostrativa di [BB96] con l'obiettivo di capire se la riduzione in **LK** sia fortemente normalizzante. In tal caso, sebbene le forme normali non siano uniche, saremo sempre certi di raggiungerne una, qualunque sia il cammino di riduzione scelto: questo ci permetterà di parlare del*le forme normali* di una derivazione, che saranno in numero finito, dunque del "contenuto" di una derivazione, e potrebbe rappresentare un primo passo per "domare" il non determinismo e tentare un'analisi più approfondita delle dinamiche classiche.

Capitolo 2

Calcolo dei sequenti e normalizzazione

Abbiamo già introdotto l'idea del *sequente* come strumento per "tracciare" le proprietà globali di una derivazione e trasformare le regole globali in regole locali. Nelle prossime pagine presenteremo nel dettaglio e rigorosamente il sistema di regole per la logica classica: il calcolo dei sequenti **LK**. Discuteremo quindi le relative procedure di normalizzazione, il teorema di eliminazione del taglio e le difficoltà peculiari del sistema. Seguiamo principalmente [AT14] e [DJS97], lasciando quando possibile sottointesi i dettagli più noti.

2.1 Linguaggio, notazioni e nozioni preliminari

Definizione 2.1 (Termini). Fissiamo:

- un insieme infinito $\mathcal{V} = \{x, y, z, \dots\}$ di *variabili vincolabili individuali*;
- un insieme $\mathcal{C} = \{c, c_1, c_2, \dots\}$ di simboli per *costanti individuali*;
- per ogni $n \geq 1$, un insieme $\mathcal{C}_n$ di simboli per *funzioni di arietà n*;

e definiamo attraverso una grammatica l'insieme $\mathcal{T}$ dei termini:

$$t_1, t_2, \dots \quad ::= \quad x \quad | \quad c \quad | \quad f_n(t_1, \dots, t_n)$$

dove $f_n \in \mathcal{C}_n$ è un simbolo per funzioni di arietà n qualsiasi.

Definizione 2.2 (Linguaggio del secondo ordine). Per ogni $n \geq 0$ fissiamo un insieme infinito $\mathcal{V}_n = \{X, Y, Z, \dots\}$ di *variabili per predicati di arietà n* e siano $t_1, t_2, \dots \in \mathcal{T}$ termini.

L'insieme $\mathcal{F}$ delle formule della logica dei predicati del secondo ordine è definito dalla grammatica seguente:

$$\begin{array}{ccccccccccc} & & & & \mathcal{F}, A, B & ::= & \mathcal{P} & | & \mathcal{N} & & \\ \mathcal{P} & ::= & X_n(t_1, \dots, t_n) & | & A \wedge_m B & | & A \vee_a B & | & \exists x.A & | & \exists X.A \\ \mathcal{N} & ::= & \neg X_n(t_1, \dots, t_n) & | & A \vee_m B & | & A \wedge_a B & | & \forall x.A & | & \forall X.A \end{array}$$

dove $X_n \in \mathcal{V}_n$ è una variabile per predicati di arietà $n \geq 0$ qualsiasi. La negazione sarà una funzione $\neg(\cdot) : \mathcal{F} \to \mathcal{F}$ definita come segue:

$$\begin{aligned}
\neg(X_n(t_1,\dots,t_n)) &= \neg X_n(t_1,\dots,t_n) & \neg(\neg X_n(t_1,\dots,t_n)) &= X_n(t_1,\dots,t_n)\\
\neg(A \wedge_m B) &= \neg(A) \vee_m \neg(B) & \neg(A \vee_m B) &= \neg(A) \wedge_m \neg(B)\\
\neg(A \vee_a B) &= \neg(A) \wedge_a \neg(B) & \neg(A \wedge_a B) &= \neg(A) \vee_a \neg(B)\\
\neg(\exists x.A) &= \forall x.\neg A & \neg(\forall x.A) &= \exists x.\neg A\\
\neg(\exists X.A) &= \forall X.\neg A & \neg(\forall X.A) &= \exists X.\neg A
\end{aligned}$$

Osserviamo che la funzione $\neg(\cdot)$ è un'involuzione: vale $\neg\neg(A) = A$. Chiamiamo *positive* tutte le formule appartenenti all'insieme $\mathcal{P}$, *negative* quelle appartenenti all'insieme $\mathcal{N}$.

Osservazione 2.3. La distinzione fra connettivi *moltiplicativi* e *additivi* ($\wedge_m/\wedge_a$, $\vee_m/\vee_a$) proviene dalla logica lineare. In logica classica parliamo di *stile* del connettivo e della regola corrispondente: come è noto le due congiunzioni e le due disgiunzioni sono classicamente equivalenti fra loro e le rispettive regole sono interderivabili utilizzando le regole strutturali. Sebbene la distinzione dei diversi stili a livello linguistico sia artificiale (dal punto di vista della semantica classica sono indistinguibili), abbiamo voluto introdurla seguendo [DJS97], sia per mettere in evidenza le diverse dinamiche di riduzione, sia per ridurre il numero di casi da trattare.

L'implicazione è classicamente equivalente alla disgiunzione e può essere trattata come un connettivo definito; anche in questo caso però dovremo distinguere due stili:

$$\begin{aligned}
A \overset{m}{\to} B &= \neg A \vee_m B\\
A \overset{a}{\to} B &= \neg A \vee_a B
\end{aligned}$$

Definizione 2.4 (Frammenti). L'insieme $\mathcal{F}^0$ delle formule proposizionali è il più grande sottoinsieme di $\mathcal{F}$ che contiene solamente formule prive di quantificatori e variabili per predicati *non 0-ari* (appartenenti a $\mathcal{V}_n$ per qualche $n > 0$).

L'insieme $\mathcal{F}^1$ delle formule del primo ordine è il più grande sottoinsieme di $\mathcal{F}$ che contiene solamente formule prive di quantificatori del secondo ordine.

L'insieme $\mathcal{F}^2$ delle formule proposizionali del secondo ordine è il più grande sottoinsieme di $\mathcal{F}$ che contiene solamente formule prive di quantificatori del primo ordine e variabili per predicati *non 0-ari* (appartenenti a $\mathcal{V}_n$ per qualche $n > 0$).

Come abbiamo anticipato già nelle definizioni, utilizziamo le lettere $A, B, C, \dots$ per denotare formule qualsiasi (dipenderà dal contesto se si tratti di formule proposizionali, del primo o del secondo ordine) e le lettere $X, Y, Z, \dots$ per denotare (variabili per) predicati. Utilizziamo le lettere minuscole $t, u, v, \dots$ per termini qualsiasi, $x, y, z, \dots$ per variabili vincolabili del primo ordine.

Chiamiamo *formula atomica* qualsiasi formula che abbia la forma $X(t_1,\dots,t_n)$ oppure $\neg X(t_1,\dots,t_n)$ per qualche $n \geq 0$.

Diamo per scontata la familiarità con le interpretazioni *Tarskiane* dei linguaggi per la logica classica: di seguito useremo l'espressione *semantica classica*.

2.1.1 Sostituzioni al primo ordine

Denotiamo con l'espressione $t[u/x]$ il termine ottenuto rimpiazzando *simultaneamente*[1] ogni occorrenza di x in t con un'occorrenza di u.

Definizione 2.5 (Occorrenze libere, vincolate). Chiamiamo le occorrenze di una variabile individuale x in una formula *libere* o *vincolate* in base alla seguente definizione. Sia A una formula e x una variabile che occorre in A:

- se A è atomica, allora tutte le occorrenze di x in A sono *libere*;
- se $A = B \wedge_{m,a} C$, $A = B \vee_{m,a} C$, $A = \exists X.B$ oppure $A = \forall X.B$, diremo che x occorre *libera* (risp. *vincolata*) in A sse occorre *libera* (risp. *vincolata*) in B o in C;
- se $A = \exists y.B$ oppure $A = \forall y.B$ con $y \neq x$, diremo che x occorre *libera* (risp. *vincolata*) in A sse occorre *libera* (risp. *vincolata*) in B;
- se $A = \exists x.B$ oppure $A = \forall x.B$, allora tutte le occorrenze di x in A sono vincolate.

Scriveremo $A(x_1, \dots, x_n)$ per denotare una formula A in cui potrebbero occorrere libere alcune fra le variabili $x_1, \dots, x_n$; denotiamo con $FV^1(A)$ l'insieme delle variabili del primo ordine libere in A.

Denotiamo con $A(t/x)$ la formula ottenuta rimpiazzando *simultaneamente* ogni occorrenza *libera* di x in A con il termine t. Come è noto, questa è un'operazione delicata se t contiene variabili che sono vincolate da qualche quantificatore. Come di consueto, consideriamo identiche due formule che differiscono solamente per i nomi delle variabili vincolate: operando una sostituzione sceglieremo implicitamente una presentazione opportuna della formula che eviti conflitti di questo genere; non ci soffermiamo però a fornire la giustificazione rigorosa di questo approccio.

2.1.2 Sostituzioni al secondo ordine

Definizione 2.6. Analogamente alla definizione 2.5, diremo che una variabile per predicati X occorre *libera* in una formula A sse $A = X(t_1, \dots, t_n)$ o $A = \neg X(t_1, \dots, t_n)$ oppure X occorre libera in qualche sottoformula di A. Diremo che occorre *vincolata* sse $A = \exists X.B$ o $A = \forall X.B$ oppure X occorre vincolata in qualche sottoformula di A. Denotiamo con $FV^2(A)$ l'insieme delle variabili proposizionali libere in A.

Denotiamo con $A[B/Xx_1 \dots x_n]$ la formula ottenuta rimpiazzando *simultaneamente* ogni sottoformula atomica 'libera' di A del tipo $X(t_1, \dots, t_n)$ (risp. $\neg X(t_1, \dots, t_n)$) con la formula $B(t_1/x_1, \dots, t_n/x_n)$ (risp. $\neg B(t_1/x_1, \dots, t_n/x_n)$).

[1] Il lettore esperto conoscerà già l'importanza della precisazione: se u contiene qualche occorrenza di x che viene introdotta in $t[u/x]$ questa non deve essere a sua volta sostituita.

Anche in questo caso evitiamo la cattura delle variabili di B durante la sostituzione utilizzando l'identità tra formule a meno di differenze nei nomi delle variabili vincolate (del secondo ma anche del primo ordine: dobbiamo evitare che eventuali variabili del primo ordine libere in B siano catturate!).

2.1.3 Sequenti

Definizione 2.7. Un *multinsieme* su X è una funzione $\mu : X \to \mathbb{N}$. L'insieme X è detto anche *supporto di* μ e useremo la notazione $|\mu| = X$.

Intuitivamente un multinsieme è equivalente a una successione di elementi di X trascurando l'ordine degli elementi: se $X = \{A, B, C\}$ e $\mu = [A, A, B]$ allora $\mu(A) = 2$, $\mu(B) = 1$, $\mu(C) = 0$.

Un multinsieme μ su X è *finito* quando l'insieme $\{x : \mu(x) > 0\}$ è finito. Denotiamo con $\mathcal{M}_{fin}(X)$ l'insieme dei multinsiemi finiti su X.

Definizione 2.8 (Sequente). Un *sequente* è un multinsieme finito su $\mathcal{F}$. Un sequente è interpretato dalla semantica classica come la disgiunzione delle sue occorrenze di formule; il sequente vuoto è interpretato dal falso.

Denotiamo con le lettere greche $\Gamma, \Delta, \ldots$ le successioni di formule (proposizionali, del primo ordine, del secondo ordine). Se Γ è una successione di tutti gli elementi di un sequente S, allora $\vdash \Gamma$ è una *presentazione* di S. Il sequente con una singola formula A e il sequente vuoto hanno entrambi una sola presentazione: $\vdash A$ e $\vdash$.

Di quando in quando sarà necessario distinguere e mettere in evidenza una particolare occorrenza di formula in un sequente. Lo faremo aggiungendo un indice in apice alla formula: $\vdash \Gamma, A^x$.

Osservazione 2.9. La definizione di sequente adottata è adeguata allo studio della *derivabilità* classica; d'altra parte le procedure di normalizzazione sono definite, e operano, rispetto a occorrenze *distinte* di formule nei sequenti: sarebbe dunque più appropriato definire il sequente come una successione di occorrenze o come un *insieme di nomi o indici* accompagnato da una funzione che associa una formula a ogni nome. Per semplicità scegliamo la presentazione tradizionale, tenendo a mente che le occorrenze sono da considerarsi distinte e mettendole in evidenza quando necessario.

Osserviamo che la scelta di definire il sequente come un multinsieme anziché come una successione ha l'effetto di rendere implicita la regola strutturale (*exchange*) del calcolo dei sequenti, che permette di riordinare le occorrenze di formule ed esprime la commutatività dei connettivi.

Osservazione 2.10. Il lettore attento avrà notato la presenza di una negazione involutiva ($\neg\neg A = A$) che sfrutta la dualità di De Morgan e la presentazione monolaterale dei sequenti (tutte le occorrenze di formule si trovano a destra nella zona "disgiuntiva", ricordando il discorso del capitolo precedente) e potrà chiedersi se questo non equivalga a presupporre la *classicità* della logica già nella sintassi del linguaggio, magari nascondendo con questo "stratagemma" qualche difetto o difficoltà delle derivazioni classiche.

Seguendo Girard in [Gir91] vogliamo sottolineare che non è questo il caso: imponendo opportune restrizioni sulla forma e il numero delle occorrenze in un sequente monolaterale (che si traducono inevitabilmente in restrizioni sulla forma delle derivazioni) e aggiungendo al linguaggio le unità logiche vero e falso, si può ottenere un sistema intuizionista. La classicità del sistema sarà data dall'assenza di simili restrizioni sui sequenti e sulle derivazioni; d'altra parte la scelta di questo formato nel caso classico ha l'effetto di semplificare le definizioni e riduce molto il numero dei casi da trattare.

En passant osserviamo che abbiamo omesso dalla nostra definizione del linguaggio le unità logiche: abbiamo fatto questa scelta seguendo l'impostazione di [DJS97], ancora una volta per ridurre il numero di regole e semplificare il lavoro, ma tutti i risultati dovrebbero potersi applicare senza troppe difficoltà al sistema esteso.

2.2 Il calcolo dei sequenti LK

Il calcolo dei sequenti per la logica classica è composto dai seguenti gruppi di regole:

Gruppo identità: assioma e taglio

$$\frac{}{\vdash \neg A, A}\,(ax) \qquad\qquad \frac{\vdash \Gamma, A \qquad \vdash \Delta, \neg A}{\vdash \Gamma, \Delta}\,(cut)$$

Regole strutturali: contrazione e indebolimento

$$\frac{\vdash \Gamma, A, A}{\vdash \Gamma, A}\,(C) \qquad\qquad \frac{\vdash \Gamma}{\vdash \Gamma, A}\,(W)$$

Regole logiche

$$\frac{\vdash \Gamma, A \qquad \vdash \Delta, B}{\vdash \Gamma, \Delta, A \wedge_m B}\,(\wedge_m) \qquad\qquad \frac{\vdash \Gamma, A, B}{\vdash \Gamma, A \vee_m B}\,(\vee_m)$$

$$\frac{\vdash \Gamma, A_i}{\vdash \Gamma, A_1 \vee_a A_2}\,(\vee_a^i) \qquad\qquad \frac{\vdash \Gamma, A \qquad \vdash \Gamma, B}{\vdash \Gamma, A \wedge_a B}\,(\wedge_a)$$

dove $i \in \{1, 2\}$. Chiamiamo *moltiplicative* le regole $(\wedge_m)$ e $(\vee_m)$, *additive* le regole $(\vee_a^1)$, $(\vee_a^2)$ e $(\wedge_a)$.

Quantificatori del primo ordine

$$\frac{\vdash \Gamma, A(t/x)}{\vdash \Gamma, \exists x.A(x)}\ (\exists) \qquad\qquad \frac{\vdash \Delta, A(x)}{\vdash \Delta, \forall x.A}\ (\forall)$$

dove t è un termine qualsiasi e la variabile x, detta *variabile propria* della regola $(\forall)$, non occorre libera in Δ (scriveremo anche $x \notin FV^1(\Delta)$).

Quantificatori del secondo ordine

$$\frac{\vdash \Gamma, A[B/Xx_1 \dots x_n]}{\vdash \Gamma, \exists X.A}\ (\exists^2) \qquad\qquad \frac{\vdash \Delta, A}{\vdash \Delta, \forall X.A}\ (\forall^2)$$

dove B è una formula qualsiasi, X una variabile per predicati di arietà n che non occorre libera in Δ ($X \notin FV^2(\Delta)$).

Diremo 0-arie, unarie, binarie le regole che hanno nessuna, una o due premesse. Denoteremo con **LK** il calcolo dei predicati del secondo ordine composto da tutte le regole elencate sopra. Denoteremo con $\mathbf{LK^0}$, $\mathbf{LK^1}$, $\mathbf{LK^2}$ i frammenti del calcolo ottenuti per restrizione rispettivamente all'insieme delle formule proposizionali $\mathcal{F}^0$, delle formule del primo ordine $\mathcal{F}^1$, delle formule proposizionali del secondo ordine $\mathcal{F}^2$. Le definizioni e i risultati che nelle prossime pagine saranno enunciati per **LK** si applicheranno senza difficoltà ad ognuno dei frammenti appena elencati.

Definizione 2.11 (Formula attiva, principale). Ricordiamo che le occorrenze di formule nei sequenti sono distinte: le regole di **LK** appena definite operano dunque su occorrenze particolari dei loro sequenti premessa e conclusione, ad esempio:

$$\frac{\vdash \Gamma, A^x \qquad \vdash \Delta, B^y}{\vdash \Gamma, \Delta, (A \wedge_m B)^z}\ (\wedge_m)$$

dove $(A \wedge_m B)^z$ è un'occorrenza "nuova", non presente nelle premesse. Chiamiamo conclusione *principale* di una regola logica l'occorrenza di formula introdotta dalla regola (nel nostro esempio z), mentre diciamo *attive* le occorrenze di formule utilizzate nelle premesse (nell'esempio x, y). Tutte le altre occorrenze saranno *passive* e saranno chiamate anche *contesto*.

La regola *cut* ha due formule attive, dette anche *formule di taglio*, e nessuna conclusione principale. L'assioma invece ha due conclusioni principali.

Nel caso delle regole strutturali non parleremo di formule attive o principali, ma di premesse o conclusioni *contratte* o *indebolite*.

2.2.1 Derivazioni, forma normale, correttezza e completezza

Richiamiamo rapidamente la nozione di albero e passiamo a definire il concetto di *derivazione* in **LK**. Seguendo [AT14] adottiamo la definizione di *albero* come insieme non vuoto e ordinato di *nodi*, tale che ogni sottoinsieme non vuoto abbia un estremo superiore e l'insieme dei maggioranti di ogni nodo sia finito e totalmente ordinato. Si dimostra facilmente che in questo ordine esiste un solo elemento massimale che chiamiamo *radice*, mentre chiamiamo *foglie* gli elementi minimali. Ogni nodo è *padre* dei suoi predecessori immediati nell'ordine arborescente e figlio del suo successore immediato. Diamo per scontate le nozioni di *cammino*, *ramo*, *ramificazione finita* e *infinita*, *altezza* dell'albero (sarà nulla quando l'albero contenga la sola radice).

Definizione 2.12 (Derivazione, derivabilità). Una *derivazione* π *in* **LK** è un albero *finito* i cui nodi sono occorrenze di regole di **LK** con l'aggiunta dell'*assioma proprio*: una regola 0-aria con conclusione qualsiasi.

Ad ogni nodo (regola) r è associato un sequente σ_r che rappresenta la conclusione di r; le conclusioni dei figli di r sono le premesse di r: questo significa che una regola binaria avrà esattamente due figli, una regola unaria avrà un figlio, una regola 0-aria sarà una foglia.

Chiamiamo *conclusione* di π il sequente $\vdash \Gamma$ associato alla radice di π, *premesse* di π i sequenti associati agli eventuali assiomi propri. Se M è l'insieme delle premesse di π, diremo che π è una derivazione di $\vdash \Gamma$ da M. Se $M = \emptyset$, diremo che π è una derivazione di $\vdash \Gamma$ e scriveremo $\pi \vdash \Gamma$.

Diremo che un sequente $\vdash \Gamma$ è *derivabile* in **LK** sse esiste una derivazione π tale che $\pi \vdash \Gamma$.

Osservazione 2.13. Gli assiomi propri indicano la "posizione" in cui due derivazioni possono essere composte. Sia $\pi_1 \vdash A$ e π_2 una derivazione di $\vdash \Gamma$ da $\vdash A$:

$$
\begin{array}{c} \vdots\ \pi_1 \\ \vdash A \end{array}
\qquad\qquad
\begin{array}{c} \vdash A \\ \vdots\ \pi_2 \\ \vdash \Gamma \end{array}
\qquad\qquad
\begin{array}{c} \vdots\ \pi_1 \\ \vdash A \\ \vdots\ \pi_2 \\ \vdash \Gamma \end{array}
$$

π_1 e π_2 possono essere composte in modo da ottenere una derivazione di $\vdash \Gamma$. Osserviamo che la regola di taglio permette di fare a meno degli assiomi propri: sostituendo in π_2 l'assioma proprio con una regola (ax) di conclusione $\vdash \neg A, A$, otteniamo una derivazione $\pi_2' \vdash \neg A, \Gamma$ che possiamo tagliare con π_1 ottenendo una derivazione di Γ:

$$
\dfrac{\begin{array}{c} \vdots\ \pi_1 \\ \vdash A \end{array} \qquad \begin{array}{c} \dfrac{}{\vdash \neg A, A}\,(ax) \\ \vdots\ \pi_2' \\ \vdash \neg A, \Gamma \end{array}}{\vdash \Gamma}\,(cut)
$$

Per questo motivo, da ora in poi ci limiteremo a considerare derivazioni *pure*, cioè senza assiomi propri.

Osservazione 2.14. Solitamente le derivazioni sono definite come alberi di sequenti in cui padri e figli sono messi in relazione da qualche regola di **LK**. Poiché nello studio dei processi di normalizzazione è spesso utile riferirsi alle occorrenze di regole in una derivazione, abbiamo scelto una definizione alternativa che non modifica la sostanza: ogni derivazione nel senso della definizione 2.12 induce una derivazione nel senso tradizionale sostituendo ogni nodo con il sequente associato.

Teorema 2.15 (Correttezza). *Il calcolo dei sequenti* **LK** *(risp.* $\mathbf{LK^0}$, $\mathbf{LK^1}$, $\mathbf{LK^2}$*) è corretto rispetto alla semantica classica per la logica dei predicati del secondo ordine (risp. proposizionale, del primo ordine, proposizionale del secondo ordine).*

Dimostrazione. Per induzione sull'altezza delle derivazioni, verificando che ogni regola "preserva la verità" (ogni modello delle premesse è anche un modello delle conclusioni). □

Definizione 2.16. Una derivazione π di Γ è *normale* se non contiene regole (cut) (diremo anche *cut-free* o *senza tagli*). È *analitica* se ogni occorrenza di formula in π è una *sottoformula*[2] di qualche occorrenza di formula in Γ.

Proprietà 2.17 (Proprietà della sottoformula). *Ogni derivazione normale è analitica.*

Dimostrazione. Per induzione sull'altezza della derivazione, osservando che la regola (cut) è l'unica ad avere occorrenze nelle premesse che non siano sottoformule della conclusione. □

Teorema 2.18 (Eliminabilità del taglio). *La regola (cut) è ridondante: ogni sequente derivabile in* **LK** *(risp. in* $\mathbf{LK^0}$, $\mathbf{LK^1}$, $\mathbf{LK^2}$*), è derivabile analiticamente.*

Dimostreremo più avanti (e in una forma più raffinata) questo risultato che è complesso in generale e particolarmente al secondo ordine. Abbiamo voluto menzionarlo in quanto è stato il punto di partenza per la tecnica dimostrativa del seguente teorema.

Teorema 2.19 (Completezza). *Il calcolo dei predicati* $\mathbf{LK^1}$ *è completo rispetto alla semantica classica per la logica dei predicati del primo ordine (ogni formula valida è derivabile).*

Osserviamo che il risultato implica immediatamente la completezza del frammento proposizionale $\mathbf{LK^0}$. La dimostrazione si basa su una tecnica di *proof-search*, per i dettagli si veda ad esempio [AT14, pp. 81–103]. Come è noto non esistono risultati di completezza al secondo ordine.

[2] Utilizziamo qui implicitamente il concetto di *sottoformula estesa*, per cui $A(t/x)$ è considerata sottoformula di $\exists x.A$ e similmente $A[B/Xx_1 \dots x_n]$ sarà sottoformula di $\exists X.A$.

2.3 Passi di riduzione e normalizzazione debole

Come abbiamo anticipato, il teorema 2.18 si dimostra in una versione raffinata che fornisce una procedura effettiva per trasformare ogni derivazione con tagli in una normale. L'importanza di questa procedura non può essere sottovalutata: in primo luogo perché fornisce una dimostrazione puramente interna della consistenza di **LK**, senza fare appello a un concetto esterno di semantica; in secondo luogo per la sua efficacia nel dimostrare risultati di derivabilità o inderivabilità; in terzo luogo per la sua natura computazionale, già menzionata in precedenza, e per la possibilità di sviluppare una *semantica delle dimostrazioni*. Proprio questo terzo aspetto diventa estremamente complesso nel caso della logica classica ed è l'ambito di indagine principale di questa tesi.

Cominciamo col definire una serie di trasformazioni basilari sulle derivazioni con tagli, quindi dimostreremo il teorema componendole in passi più complessi.

Passi logici moltiplicativi

$$\dfrac{\dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma_1, A} \quad \overset{\vdots\,\pi_2}{\vdash \Gamma_2, B}}{\vdash \Gamma_1, \Gamma_2, A \wedge_m B}(\wedge_m) \qquad \dfrac{\overset{\vdots\,\pi_3}{\vdash \Delta, \neg A, \neg B}}{\vdash \Delta, \neg A \vee_m \neg B}(\vee_m)}{\vdash \Gamma_1, \Gamma_2, \Delta}(cut)$$

$$\rightsquigarrow_{(\wedge_m/\vee_m^1)} \dfrac{\dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma_1, A} \quad \overset{\vdots\,\pi_3}{\vdash \Delta, \neg A, \neg B}}{\vdash \Gamma_1, \Delta, \neg B}(cut) \qquad \overset{\vdots\,\pi_2}{\vdash \Gamma_2, B}}{\vdash \Gamma_1, \Gamma_2, \Delta}(cut)$$

$$\dfrac{\dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma_1, A} \quad \overset{\vdots\,\pi_2}{\vdash \Gamma_2, B}}{\vdash \Gamma_1, \Gamma_2, A \wedge_m B}(\wedge_m) \qquad \dfrac{\overset{\vdots\,\pi_3}{\vdash \Delta, \neg A, \neg B}}{\vdash \Delta, \neg A \vee_m \neg B}(\vee_m)}{\vdash \Gamma_1, \Gamma_2, \Delta}(cut)$$

$$\rightsquigarrow_{(\wedge_m/\vee_m^2)} \dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma_1, A} \qquad \dfrac{\overset{\vdots\,\pi_3}{\vdash \Delta, \neg A, \neg B} \quad \overset{\vdots\,\pi_2}{\vdash \Gamma_2, B}}{\vdash \Gamma_2, \Delta, \neg B}(cut)}{\vdash \Gamma_1, \Gamma_2, \Delta}(cut)$$

Passi logici additivi

$$\dfrac{\dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma, A} \qquad \overset{\vdots\,\pi_2}{\vdash \Gamma, B}}{\vdash \Gamma, A \wedge_a B}\,(\wedge_a) \qquad \dfrac{\overset{\vdots\,\pi_3}{\vdash \Delta, \neg A}}{\vdash \Delta, \neg A \vee_a \neg B}\,(\vee_a^1)}{\vdash \Gamma, \Delta}\,(cut)$$

$$\rightsquigarrow_{(\wedge_a/\vee_a^1)} \quad \dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma, A} \qquad \overset{\vdots\,\pi_3}{\vdash \Delta, \neg A}}{\vdash \Gamma, \Delta}\,(cut)$$

$$\dfrac{\dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma, A} \qquad \overset{\vdots\,\pi_2}{\vdash \Gamma, B}}{\vdash \Gamma, A \wedge_a B}\,(\wedge_a) \qquad \dfrac{\overset{\vdots\,\pi_3}{\vdash \Delta, \neg B}}{\vdash \Delta, \neg A \vee_a \neg B}\,(\vee_a^2)}{\vdash \Gamma, \Delta}\,(cut)$$

$$\rightsquigarrow_{(\wedge_a/\vee_a^2)} \quad \dfrac{\overset{\vdots\,\pi_2}{\vdash \Gamma, B} \qquad \overset{\vdots\,\pi_3}{\vdash \Delta, \neg B}}{\vdash \Gamma, \Delta}\,(cut)$$

Passi logici per i quantificatori

$$\dfrac{\dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma, A(x)}}{\vdash \Gamma, \forall x. A}\,(\forall) \qquad \dfrac{\overset{\vdots\,\pi_2}{\vdash \Delta, \neg A(t/x)}}{\vdash \Delta, \exists x. \neg A}\,(\exists)}{\vdash \Gamma, \Delta}\,(cut)$$

$$\rightsquigarrow_{(\forall/\exists)} \quad \dfrac{\overset{\vdots\,\pi_1[t/x]}{\vdash \Gamma, A(t/x)} \qquad \overset{\vdots\,\pi_2}{\vdash \Delta, \neg A(t/x)}}{\vdash \Gamma, \Delta}\,(cut)$$

$$\dfrac{\dfrac{\overset{\vdots\,\pi_1}{\vdash \Gamma, A}}{\vdash \Gamma, \forall X. A}\,(\forall^2) \qquad \dfrac{\overset{\vdots\,\pi_2}{\vdash \Delta, \neg A[B/Xx_1 \dots x_n]}}{\vdash \Delta, \exists X. \neg A}\,(\exists^2)}{\vdash \Gamma, \Delta}\,(cut)$$

$$\rightsquigarrow_{(\forall^2/\exists^2)} \quad \dfrac{\overset{\vdots\,\pi_1[B/Xx_1 \dots x_n]}{\vdash \Gamma, A[B/Xx_1 \dots x_n]} \qquad \overset{\vdots\,\pi_2}{\vdash \Delta, \neg A[B/Xx_1 \dots x_n]}}{\vdash \Gamma, \Delta}\,(cut)$$

L'espressione $\pi_1[t/x]$ (risp. $\pi_1[B/Xx_1 \dots x_n]$) indica che abbiamo sostituito con t (risp. con $B(x_1,\dots,x_n)$) tutte le occorrenze della variabile individuale x nella derivazione π_1 (risp. della variabile per predicati X) che sono “riconducibili” ad occorrenze libere nella conclusione di π_1. L'operazione non è affatto banale: oltre ad evitare la cattura non voluta da parte di quantificatori, dovremo evitare di sostituire le variabili proprie di qualche regola $(\forall)$ o $(\forall^2)$. La definizione rigorosa di questa operazione è molto complessa, rimandiamo per esempio a [AT14, pp. 117–121].

Passi strutturali[3]

$$\dfrac{\begin{matrix}\vdots\ \pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \dfrac{}{\vdash A, \neg A}\,(ax)}{\vdash \Gamma, A}\,(cut) \quad \rightsquigarrow_{(ax)} \quad \begin{matrix}\vdots\ \pi_1\\ \vdash \Gamma, A\end{matrix}$$

$$\dfrac{\begin{matrix}\vdots\ \pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \dfrac{\begin{matrix}\vdots\ \pi_2\\ \vdash \Delta\end{matrix}}{\vdash \Delta, \neg A}\,(W)}{\vdash \Gamma, \Delta}\,(cut) \quad \rightsquigarrow_{(W)} \quad \dfrac{\begin{matrix}\vdots\ \pi_2\\ \vdash \Delta\end{matrix}}{\vdash \Gamma, \Delta}\,(W)$$

$$\dfrac{\begin{matrix}\vdots\ \pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \dfrac{\begin{matrix}\vdots\ \pi_2\\ \vdash \Delta, \neg A, \neg A\end{matrix}}{\vdash \Delta, \neg A}\,(C)}{\vdash \Gamma, \Delta}\,(cut)$$

$$\rightsquigarrow_{(C)} \quad \dfrac{\dfrac{\begin{matrix}\vdots\ \pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \dfrac{\begin{matrix}\vdots\ \pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \begin{matrix}\vdots\ \pi_2\\ \vdash \Delta, \neg A, \neg A\end{matrix}}{\vdash \Delta, \neg A}\,(cut)}{\vdash \Gamma, \Gamma, \Delta}\,(cut)}{\vdash \Gamma, \Delta}\,(C)$$

[3] Le doppie righe indicano che abbiamo effettuato il numero di indebolimenti o contrazioni necessarie ad ottenere la conclusione.

Passi commutativi

$$\dfrac{\begin{matrix}\vdots\;\pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \dfrac{\begin{matrix}\vdots\;\pi_2\\ \vdash \Delta, \neg A^x\end{matrix}}{\vdash \Delta', \neg A^x}\,(R)}{\vdash \Gamma, \Delta'}\,(cut) \quad \rightsquigarrow_{(cc)} \quad \dfrac{\dfrac{\begin{matrix}\vdots\;\pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \begin{matrix}\vdots\;\pi_2\\ \vdash \Delta, \neg A^x\end{matrix}}{\vdash \Gamma, \Delta}\,(cut)}{\vdash \Gamma, \Delta'}\,(R)$$

dove R è una regola unaria e $\neg A^x$ non è conclusione principale di R.

$$\dfrac{\begin{matrix}\vdots\;\pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \dfrac{\begin{matrix}\vdots\;\pi_2\\ \vdash \Delta_1, \neg A, B\end{matrix} \qquad \begin{matrix}\vdots\;\pi_3\\ \vdash \Delta_2, C\end{matrix}}{\vdash \Delta_1, \Delta_2, B \wedge_m C, \neg A}\,(\wedge_m)}{\vdash \Gamma, \Delta_1, \Delta_2, B \wedge_m C}\,(cut)$$

$$\rightsquigarrow_{(cc)} \quad \dfrac{\dfrac{\begin{matrix}\vdots\;\pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \begin{matrix}\vdots\;\pi_2\\ \vdash \Delta_1, B, \neg A\end{matrix}}{\vdash \Gamma, \Delta_1, B}\,(cut) \qquad \begin{matrix}\vdots\;\pi_3\\ \vdash \Delta_2, C\end{matrix}}{\vdash \Gamma, \Delta_1, \Delta_2, B \wedge_m C}\,(\wedge_m)$$

$$\dfrac{\begin{matrix}\vdots\;\pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \dfrac{\begin{matrix}\vdots\;\pi_2\\ \vdash \Delta, \neg A, B\end{matrix} \qquad \begin{matrix}\vdots\;\pi_3\\ \vdash \Delta, \neg A, C\end{matrix}}{\vdash \Delta, B \wedge_a C, \neg A}\,(\wedge_a)}{\vdash \Gamma, \Delta, B \wedge_a C}\,(cut)$$

$$\rightsquigarrow_{(ccadd)} \quad \dfrac{\dfrac{\begin{matrix}\vdots\;\pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \begin{matrix}\vdots\;\pi_2\\ \vdash \Delta, B, \neg A\end{matrix}}{\vdash \Gamma, \Delta, B}\,(cut) \qquad \dfrac{\begin{matrix}\vdots\;\pi_1\\ \vdash \Gamma, A\end{matrix} \qquad \begin{matrix}\vdots\;\pi_3\\ \vdash \Delta, C, \neg A\end{matrix}}{\vdash \Gamma, \Delta, C}\,(cut)}{\vdash \Gamma, \Delta, B \wedge_a C}\,(\wedge_a)$$

I passi commutativi si applicano quando non sia possibile effettuare altri passi e hanno lo scopo di "avvicinare" il taglio alle regole che introducono le formule di taglio, finché non sia possibile applicare al taglio un passo logico o strutturale.

Osserviamo che sia nel caso dei passi strutturali, sia nel caso dei passi commutativi, può capitare che sia necessario compiere una scelta: quando *entrambe* le formule di taglio siano indebolite, contratte o non principali sarà possibile ridurre il taglio in modi diversi. Come vedremo più avanti, questo è uno dei motivi per cui la procedura di riduzione in **LK** è non deterministica.

2.3.1 Il teorema di eliminazione del taglio (normalizzazione debole)

Definizione 2.20 (Passo logico). Il passo di riduzione logico $\rightsquigarrow_L$ è l'unione dei passi $(\wedge_m/\vee_m^1)$, $(\wedge_m/\vee_m^2)$, $(\wedge_a/\vee_a^1)$, $(\wedge_a/\vee_a^2)$, $(\wedge_a/\vee_m^1)$, $(\wedge_a/\vee_m^2)$, $(\wedge_m/\vee_a^1)$, $(\wedge_m/\vee_a^2)$, $(\forall/\exists)$, $(\forall^2/\exists^2)$. Chiamiamo *taglio logico* (o *taglio di tipo* L) ogni occorrenza della regola (cut) a cui si applichi un passo $\rightsquigarrow_L$. Equivalentemente, un taglio è logico (di tipo L) sse *entrambe* le formule di taglio sono *principali in una regola logica*.

Vogliamo ora definire un generico passo di riduzione strutturale che raccolga tutti i passi strutturali e logici. Non è possibile però utilizzare, come nel caso logico la, strategia di unire i passi appena definiti (le ragioni saranno chiarite più avanti) e per questo abbiamo bisogno di alcune nozioni preliminari.

Definizione 2.21. Sia A^x una occorrenza di formula non principale nella conclusione di qualche (occorrenza di) regola r. Diremo che una occorrenza di formula A^y in qualche premessa di r è *ascendente immediata* di A^x se:

a. r è la regola $(\wedge_a)$, A^x occorre nel contesto della conclusione e A^y è una delle due occorrenze corrispondenti ad A^x nei due contesti delle premesse.

b. r è una contrazione di conclusione A^x e A^y è una delle due premesse contratte;

c. A^x occorre nel contesto della conclusione e A^y è l'occorrenza corrispondente ad A^x nel contesto di qualche premessa.

Sia $\pi \vdash \Gamma, A^x$. Una occorrenza qualsiasi A^y in π è *ascendente* di A^x sse esiste in una sequenza $A^{x_1}, \ldots, A^{x_n}$ di occorrenze in π tale che $A^{x_1} = A^x$, $A^{x_n} = A^y$ e per ogni $1 \leq i < n$, $A^{x_{i+1}}$ è ascendente immediata di A^{x_i}. Diremo anche che A^x è *discendente* di A^y.

$$\frac{\vdash \Gamma, A^y, B \qquad \vdash \Delta, C}{\vdash \Gamma, A^x, \Delta, B \wedge_m C}\,(\wedge_m) \qquad\qquad \frac{\vdash \Gamma, A^u, A^v}{\vdash \Gamma, A^z}\,(C)$$

Figura 2.1: A^y è ascendente (immediata) di A^x, A^u e A^v di A^z.

Osservazione 2.22. Secondo la definizione 2.21, ogni occorrenza A^x è ascendente di se stessa: la sequenza A^x soddisfa in maniera vacua i requisiti necessari.

Osserviamo che la definizione si appoggia in modo intuitivo sulla "corrispondenza" fra i contesti di premesse e conclusioni. Ricordando però che i contesti delle premesse sono copiati nella conclusione e che supponiamo di poter distinguere diverse occorrenze della stessa formula in un sequente, è chiaro che ogni occorrenza nel contesto di una conclusione sarà ottenuta "copiando" una particolare occorrenza nel contesto di qualche premessa. La precisazione di questa idea richiede (come dicevamo nell'osservazione 2.9) di definire il sequente come una successione o un insieme di formule con indice.

Spesso per mettere in evidenza le occorrenze ascendenti nelle premesse utilizzeremo lo stesso indice applicato alle formule nella conclusione:

$$\frac{\vdash \Gamma, A^x, B}{\vdash \Gamma, A^x, B \vee_a C}\ (\vee_a^1)$$

Definizione 2.23 (Albero strutturale). Sia π una derivazione che contiene un'occorrenza di formula A^x: l'*albero strutturale di A^x in π* è il sottoinsieme ordinato di π composto da ogni occorrenza di regola la cui conclusione contenga un'ascendente di A^x (l'ordine dei nodi sarà lo stesso di π).

Intuitivamente, la costruzione dell'albero strutturale di A^x in π consiste nel "risalire" la derivazione partendo dalla regola r in cui occorre A^x e "tracciando la storia" delle occorrenze da cui A^x discende: la radice dell'albero strutturale sarà r, le foglie saranno le regole logiche, assiomi o indebolimenti che introducono A^x in π.

Definizione 2.24 (Passo strutturale). Il passo di riduzione strutturale $\rightsquigarrow_S$ si applica ad ogni *taglio strutturale*, caratterizzato dalla forma seguente:

$$\pi = \cfrac{\begin{array}{c}\vdots\ \pi_1 \\ \vdash \Gamma, A^x\end{array} \qquad \begin{array}{c}\vdots\ \pi_2 \\ \vdash \Delta, \neg A^y\end{array}}{\vdash \Gamma, \Delta}\ (cut)$$

dove almeno una delle due formule di taglio *non è* conclusione principale di una regola logica. Evidentemente nessun taglio logico è strutturale e viceversa.

Supponiamo senza perdita di generalità che $\neg A^y$ non sia conclusione principale di una regola logica, allora $\pi \rightsquigarrow_S \pi'$ dove π' è costruita come segue:

1. si costruisce l'albero strutturale di $\neg A^y$ in π_2 e si considera separatamente ogni foglia dell'albero;

2. ogni foglia (W) di conclusione A viene sostituita in π_2 da un numero di regole (W) sufficiente a concludere Γ;

3. ogni foglia (ax) viene sostituita in π_2 da π_1: la conclusione A^x di π_1 rimpiazza la conclusione A dell'assioma;

4. per ogni foglia logica con conclusione principale $\neg A$ si introduce in π_2 un taglio[4] fra tale foglia π_1;

[4] Per essere precisi, sia r una regola logica che è foglia dell'albero strutturale di $\neg A^y$ in π_2 e sia δ il sottoalbero di π_2 radicato in r, cioè la sottoderivazione di π_2 che termina con r. Il passo $\rightsquigarrow_S$ sostituisce δ in π_2 con un taglio fra π_1 e δ, usando come formule di taglio A^x e la conclusione principale $\neg A$ di r.

5. si sostituisce in π_2 con Γ ogni occorrenza di $\neg A$ che (i) sia ascendente di $\neg A^y$ e (ii) non occorra in una foglia dell'albero strutturale di $\neg A^y$; si sostituiscono in π_2 le contrazioni che operano su ascendenti di $\neg A^y$ con il numero di regole (C) necessarie a contrarre Γ, Γ in Γ.

Diciamo che π_1 è la *sottoderivazione trasportata* e che il passo $\rightsquigarrow_S$ *trasporta* π_1.[5] Diciamo inoltre che i nuovi tagli eventualmente introdotti sono *residui* del taglio ridotto. Si verifica facilmente che π' può essere ottenuta applicando un certo numero di volte i passi $\rightsquigarrow_{(ax)}$, $\rightsquigarrow_{(W)}$, $\rightsquigarrow_{(C)}$ e i passi commutativi, di volta in volta spostando, eliminando o duplicando sempre π_1.

Osservazione 2.25. Sia π una derivazione che termina con un taglio strutturale:

$$\pi = \cfrac{\begin{array}{c}\vdots\ \pi_1 \\ \vdash \Gamma, A^x\end{array} \qquad \begin{array}{c}\vdots\ \pi_2 \\ \vdash \Delta, \neg A^y\end{array}}{\vdash \Gamma, \Delta}\ (cut)$$

Se π_2 è un assioma con conclusione $\vdash A, \neg A^y$, allora $\pi \rightsquigarrow_S \pi_1$. Se tutte le foglie dell'albero strutturale di $\neg A^y$ in π_2 sono indebolimenti, il passo $\rightsquigarrow_S$ *cancella* π_1. In particolare, se l'ultima regola di π_2 è un indebolimento che introduce $\neg A^y$, allora π si riduce a π_2 con l'indebolimento finale sostituito da un certo numero di indebolimenti su Γ.

Si osservi infine che (come nel caso dei vari passi strutturali e commutativi) quando sia A^x che $\neg A^y$ non sono principali, è possibile applicare il passo $\rightsquigarrow_S$ in due modi diversi: trasportando π_1 o trasportando π_2.

Definizione 2.26. Denotiamo con $\rightsquigarrow_1$ (riduzione in un passo) l'unione delle relazioni $\rightsquigarrow_L$ e $\rightsquigarrow_S$; denotiamo con $\rightsquigarrow$ (riduzione) la chiusura riflessiva e transitiva di $\rightsquigarrow_1$.

Teorema 2.27 (Eliminazione del taglio, Hauptsatz). *Esiste una procedura che trasforma ogni derivazione π in $\mathbf{LK^1}$ di un sequente qualsiasi $\vdash \Gamma$ in una derivazione senza tagli dello stesso sequente.*

Seguendo [AT14] presentiamo una variante dell'argomento proposto da Danos, Joinet e Schellinx in [DJS97].

Lemma 2.28. *Sia $\pi \vdash \Gamma, \Delta$ qualsiasi derivazione con la forma seguente:*

$$\pi = \cfrac{\begin{array}{c}\vdots\ \pi_1 \\ \vdash \Gamma, A^x\end{array} \qquad \begin{array}{c}\vdots\ \pi_2 \\ \vdash \Delta, \neg A^y\end{array}}{\vdash \Gamma, \Delta}\ (cut)$$

dove π_1 e π_2 sono senza tagli: diremo che π è quasi senza tagli. *Esiste una procedura che trasforma π in una derivazione* senza tagli *con la stessa conclusione.*

[5] Intuitivamente π_1 viene "trasportata verso l'alto" seguendo i rami dell'albero strutturale di $\neg A^y$ in π_2.

Dimostrazione. Sia t il taglio finale di π. Definiamo l'*energia* di t come la coppia $(d(A), s(t))$ dove $d(A)$ è il numero di simboli in A, detto anche *grado* di A, e $s(t)$ è il tipo di t: L se t è logico, $S2$ se è strutturale e una delle due formule di taglio è conclusione principale di una regola logica, $S1$ se entrambe le formule di taglio sono non principali. Le energie sono ordinate lessicograficamente con $S1 > S2 > L$. Procediamo per induzione sull'energia di t:

- se $s(t) = L$, allora applicando un passo $\rightsquigarrow_L$ otteniamo uno o due tagli su sottoformule di A e $\neg A$ che avranno grado strettamente minore di A. Applicando l'ipotesi induttiva una o due volte possiamo concludere.

- Se $s(t) = S2$, allora applicando un passo $\rightsquigarrow_S$ e trasportando la derivazione che contiene la formula di taglio principale otteniamo una derivazione con zero o più tagli residui di tipo L (per definizione del passo $\rightsquigarrow_S$: i tagli vengono introdotti solamente su foglie logiche dell'albero strutturale). I nuovi tagli avranno dunque energia strettamente minore di t e potremo applicare l'ipotesi induttiva a ciascun residuo separatamente.

- Se $s(t) = S1$, allora scegliamo quale sottoderivazione trasportare fra π_1 e π_2 e applichiamo un passo $\rightsquigarrow_S$: tutti i residui saranno di tipo $S2$ (ancora una volta per la definizione del passo $\rightsquigarrow_S$) e potremo applicare a ciascuno l'ipotesi induttiva.

Il fatto che la conclusione sia la stessa si verifica facilmente osservando che ognuno dei passi definiti in precedenza preserva le conclusioni. □

Dimostrazione del teorema 2.27. Per induzione sul numero di tagli in π. Osserviamo che grazie all'ordine arborescente sulle regole di π è sempre possibile trovare una sottoderivazione δ di π quasi senza tagli: il lemma 2.28 permette di trasformare δ in una derivazione normale δ' con la stessa conclusione. Sostituendo δ con δ' in π otteniamo una derivazione π' con conclusione $\vdash \Gamma$ e numero di tagli strettamente minore di π. □

Lo stesso risultato, come vedremo a breve, vale anche per **LK** non ristretto al primo ordine, ma come è noto la dimostrazione al secondo ordine è molto più difficile (il passo logico per i quantificatori del secondo ordine può aumentare arbitrariamente il grado delle formule di taglio). Per ora ci limitiamo a esporre il risultato su $\mathbf{LK^1}$ che è dimostrabile in modo più semplice.

Dal punto di vista computazionale, il teorema 2.27 equivale a un risultato di *normalizzazione debole* o esistenza della forma normale e può essere riformulato in questi termini: esiste almeno una *strategia di riduzione* che raggiunge sempre una forma normale.

Definizione 2.29. Denotiamo con WN $\subseteq$ **LK** l'insieme di derivazioni tale che $\pi \in$ WN sse esiste π' normale tale che $\pi \rightsquigarrow \pi'$.

Definizione 2.30. Sia $\pi \in \mathbf{LK}$. Denotiamo con $\nu(\pi)$ l'altezza (eventualmente infinita) dell'albero di riduzione di π. SN $\subseteq \mathbf{LK}$ è l'insieme di derivazioni tale che $\pi \in$ SN sse $\nu(\pi) \in \mathbb{N}$.

Teorema 2.31 (Normalizzazione debole). $\mathbf{LK^1} \subseteq$ WN.

Definizione 2.32. Diciamo che $F \subseteq \mathbf{LK}$ è un *frammento computazionale* di **LK** se per ogni $\pi \in F$ e per ogni π' tale che $\pi \rightsquigarrow \pi'$ vale $\pi' \in F$ (F è chiuso sotto riduzione).

Proposizione 2.33. $\mathbf{LK^0}$, $\mathbf{LK^1}$, $\mathbf{LK^2}$ *sono frammenti computazionali di* **LK**.

Dimostrazione. Nessun passo di riduzione introduce regole sui quantificatori. □

Come caso particolare della proposizione appena enunciata, $\mathbf{LK^0}$ è un frammento computazionale di $\mathbf{LK^1}$ dunque il risultato di normalizzazione debole si applica anche a $\mathbf{LK^0}$.

Corollario 2.34. $\mathbf{LK^0} \subseteq$ WN.

2.4 Non confluenza di LK

È naturale chiedersi se per le riduzioni su $\mathbf{LK^1}$ (o su **LK**) valgano anche le proprietà di *normalizzazione forte* (ogni strategia raggiunge una forma normale) e di *confluenza* (che implica l'unicità della forma normale). Certamente nel secondo caso si può dare una risposta negativa. Abbiamo già accennato ai motivi: analizziamo ora nel dettaglio le difficoltà che in [DJS97] sono state battezzate "dilemma strutturale" e "dilemma logico".

2.4.1 Dilemma strutturale

Una prima difficoltà in vista della normalizzazione forte è data dal comportamento dei passi strutturali e commutativi. Consideriamo una derivazione con tagli del sequente $\vdash \neg A \wedge_m \neg A, A \wedge_m A$:

$$
\dfrac{
  \dfrac{
    \dfrac{
      \dfrac{}{\vdash \neg A, A}\,(ax) \quad \dfrac{}{\vdash \neg A, A}\,(ax)
    }{\vdash \neg A \wedge_m \neg A, A, A}\,(\wedge_m)
  }{\vdash \neg A \wedge_m \neg A, A}\,(C)
  \quad
  \dfrac{
    \dfrac{
      \dfrac{}{\vdash A, \neg A}\,(ax) \quad \dfrac{}{\vdash A, \neg A}\,(ax)
    }{\vdash A \wedge_m A, \neg A, \neg A}\,(\wedge_m)
  }{\vdash A \wedge_m A, \neg A}\,(C)
}{\vdash \neg A \wedge_m \neg A, A \wedge_m A}\,(cut)
$$

l'applicazione di un singolo passo $\rightsquigarrow_{(C)}$ permette di ottenere il ridotto seguente:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\vdash \neg A, A} \quad \dfrac{}{\vdash \neg A, A}}{\vdash \neg A \wedge_m \neg A, A, A}}{\vdash \neg A \wedge_m \neg A, A} \qquad \dfrac{\dfrac{\dfrac{\dfrac{}{\vdash \neg A, A} \quad \dfrac{}{\vdash \neg A, A}}{\vdash \neg A \wedge_m \neg A, A, A}}{\vdash \neg A \wedge_m \neg A, A} \quad \dfrac{\dfrac{}{\vdash A, \neg A} \quad \dfrac{}{\vdash A, \neg A}}{\vdash A \wedge_m A, \neg A, \neg A}\,(cut)}{\vdash \neg A \wedge_m \neg A, A \wedge_m A, \neg A}\,(cut)}{\vdash \neg A \wedge_m \neg A, \neg A \wedge_m \neg A, A \wedge_m A}\,(C)}{\vdash \neg A \wedge_m \neg A, A \wedge_m A}$$

Ora si danno due alternative. Continuare a trasportare verso l'alto la sottoderivazione sinistra fino a ottenere la seguente forma normale:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{\vdash \neg A, A}\,(ax) \quad \dfrac{}{\vdash \neg A, A}\,(ax)}{\vdash \neg A \wedge_m \neg A, A, A}\,(\wedge_m)}{\vdash \neg A \wedge_m \neg A, A}\,(C) \qquad \dfrac{\dfrac{\dfrac{}{\vdash \neg A, A}\,(ax) \quad \dfrac{}{\vdash \neg A, A}\,(ax)}{\vdash \neg A \wedge_m \neg A, A, A}\,(\wedge_m)}{\vdash \neg A \wedge_m \neg A, A}\,(C)}{\vdash \neg A \wedge_m \neg A, \neg A \wedge_m \neg A, A \wedge_m A}\,(\wedge_m)}{\vdash \neg A \wedge_m \neg A, A \wedge_m A}\,(C)$$

oppure cambiare strategia e trasportare la sottoderivazione destra: si osserva facilmente che applicando in modo opportuno il passo $\rightsquigarrow_C$ si può procedere all'infinito senza mai raggiungere una forma normale (in particolare duplicando sempre la sottoderivazione che termina con il taglio superiore).

Si tratta di una difficoltà caratteristica della sintassi del calcolo dei sequenti: come è noto, in deduzione naturale non sono necessari passi commutativi e i passi "strutturali" operano sostituendo le ipotesi di una dimostrazione con una dimostrazione delle ipotesi. L'introduzione del passo strutturale generico $\rightsquigarrow_S$ ha l'obiettivo di ovviare a questo inconveniente: una volta scelta la sottoderivazione da trasportare, si persiste nella scelta fino a raggiungere gli assiomi, gli indebolimenti o le regole logiche che introducono la formula di taglio. La "risalita" fino agli assiomi rispecchia il processo di riduzione della deduzione naturale.

Quanto appena detto però lascia aperta la questione della scelta: in ogni taglio di tipo $S1$ sarà possibile applicare due riduzioni diverse (trasportando la sottoderivazione destra o sinistra) e i due ridotti saranno nel caso generale molto differenti. È chiaro quindi che per ottenere un protocollo di riduzione confluente sarà necessario risolvere questa ambiguità.

2.4.2 Dilemma logico

Esiste però una seconda fonte di ambiguità che finora abbiamo passato sotto silenzio. Consideriamo un taglio logico moltiplicativo:

$$\dfrac{\dfrac{\begin{matrix}\vdots\,\pi_1\\ \vdash\Gamma_1,A\end{matrix}\qquad\begin{matrix}\vdots\,\pi_2\\ \vdash\Gamma_2,B\end{matrix}}{\vdash\Gamma_1,\Gamma_2,A\wedge_m B}\,(\wedge_m)\qquad\dfrac{\begin{matrix}\vdots\,\pi_3\\ \vdash\Delta,\neg A,\neg B\end{matrix}}{\vdash\Delta,\neg A\vee_m\neg B}\,(\vee_m)}{\vdash\Gamma_1,\Gamma_2,\Delta}\,(cut)$$

Possiamo applicare due passi di riduzione, $\rightsquigarrow_{(\wedge_m/\vee_m^1)}$ e $\rightsquigarrow_{(\wedge_m/\vee_m^2)}$, ottenendo le due derivazioni seguenti:

$$\dfrac{\dfrac{\begin{matrix}\vdots\,\pi_1\\ \vdash\Gamma_1,A\end{matrix}\qquad\begin{matrix}\vdots\,\pi_3\\ \vdash\Delta,\neg A,\neg B\end{matrix}}{\vdash\Gamma_1,\Delta,\neg B}\,(cut)\qquad\begin{matrix}\vdots\,\pi_2\\ \vdash\Gamma_2,B\end{matrix}}{\vdash\Gamma_1,\Gamma_2,\Delta}\,(cut)$$

$$\dfrac{\begin{matrix}\vdots\,\pi_1\\ \vdash\Gamma_1,A\end{matrix}\qquad\dfrac{\begin{matrix}\vdots\,\pi_3\\ \vdash\Delta,\neg A,\neg B\end{matrix}\qquad\begin{matrix}\vdots\,\pi_2\\ \vdash\Gamma_2,B\end{matrix}}{\vdash\Gamma_2,\Delta,\neg B}\,(cut)}{\vdash\Gamma_1,\Gamma_2,\Delta}\,(cut)$$

In molti casi le due forme sono equivalenti. Tuttavia, quando entrambi i tagli sono di tipo strutturale con A e B conclusioni *non principali* di π_1 e π_2, hanno dei ridotti non comuni. Consideriamo la prima delle due e supponiamo di ridurre il taglio superiore trasportando π_3 all'interno di π_1, quindi riduciamo il taglio finale facendo salire la sottoderivazione così ottenuta lungo l'albero strutturale di B in π_2. Si verifica facilmente che la derivazione risultante non può essere ottenuta per riduzione dalla seconda forma: comunque si proceda, una riduzione sul taglio finale non potrà fare altro che trasportare anche π_2 all'interno di π_1, portando a forme normali in generale molto differenti. Questo problema è stato identificato con chiarezza da Girard nel suo [Gir91], ma è stato osservato, per esempio, anche in [Par92a].

2.5 Il protocollo di riduzione tq

Sono state proposte in letteratura varie soluzioni ai problemi appena discussi. Il dilemma strutturale viene risolto “orientando” i tagli, cioè aggiungendo alle derivazioni (o estraendone) qualche informazione che indichi quale sottoderivazione trasportare nel caso $S1$: sono esempi di questo approccio il sistema **LC** di Girard (introdotto in [Gir91]) o il sistema **FD** (*free deduction*) e il λμ-calcolo di Parigot (risp. in [Par92a; Par92b]). Quest'ultimo permette tra l'altro di associare a ogni derivazione classica un termine di una versione modificata del λ-calcolo.

Queste proposte differiscono per il modo in cui risolvono la difficoltà sui passi commutativi e il dilemma logico. In **LC** troviamo una soluzione simile al passo $\rightsquigarrow_S$ di [DJS97] e l'uso della regola $(\wedge_m)$ è ristretto nei casi rilevanti per il dilemma logico, con il risultato che l'ordine in cui effettuare i due tagli sarà sempre determinato. Invece **FD** e il suo frammento λμ-calcolo adottano una sintassi più vicina a quella della deduzione naturale (il passo strutturale si limita a sostituire alcuni assiomi); inoltre in **FD** si omette completamente la regola $(\vee_m)$, mentre nel λμ-calcolo l'orientamento dei tagli strutturali è fissato in maniera tale che il dilemma logico non sia rilevante (cfr. [DJS97, p. 795]).

Nei prossimi paragrafi presentiamo la soluzione proposta da Danos, Joinet e Schellinx in [DJS97]: il sistema $\mathbf{LK}^{tq}$ e il protocollo di riduzione tq. Questo approccio ha il merito di essere il più generale: come osservato dagli autori, tutti gli altri sistemi possono essere ricostruiti come frammenti computazionali di $\mathbf{LK}^{tq}$ e le loro riduzioni possono essere simulate.

Definizione 2.35. Una *formula colorata* è una formula $A \in \mathcal{F}$ accompagnata da una funzione che associa a ogni sottoformula di A un colore $\epsilon \in \{t, q\}$. Utilizziamo la notazione A^ϵ per denotare una formula A la cui "sottoformula principale"[6] ha colore ϵ. Diremo che ϵ è il *colore di A*; parleremo invece di *colorazione di A* per riferirci ai colori di tutte le sottoformule.

Definiamo la funzione involutiva $\bar{\epsilon} : \{t, q\} \to \{t, q\}$ che scambia i due colori:

$$\begin{aligned} \bar{t} &= q \\ \bar{q} &= t. \end{aligned}$$

Evidentemente $\bar{\bar{\epsilon}} = \epsilon$. Definiamo infine la *negazione* di una formula colorata $\neg(A^\epsilon)$ per induzione sull'altezza di A come per le formule non colorate, ma la negazione invertirà anche il colore di ogni sottoformula. Ad esempio avremo:

$$\neg(X^t \wedge_m Y^q)^t = ((\neg X)^q \vee_m (\neg Y)^t)^q.$$

In generale una coppia di formule duali sarà colorata dualmente: una delle due avrà colore t e l'altra q e la stessa cosa vale per le rispettive sottoformule. Due formule colorate sono considerate uguali quando differiscono solamente per il nome delle variabili vincolate (in particolare, le *colorazioni* saranno identiche).

2.5.1 Regole e derivazioni in LKtq

Le regole di $\mathbf{LK}^{tq}$ sono quelle di **LK** ma con formule colorate e con l'aggiunta di condizioni sulla colorazione delle formule. In particolare, *la colorazione dei contesti è sempre preservata*. Inoltre la regola $(\vee_m)$ è sostituita dalla coppia di regole $(\vee_m^1)$ e $(\vee_m^2)$, che saranno utilizzate per risolvere il dilemma logico.

[6] Ci riferiamo alla formula A che è sempre sottoformula (impropria) di A ed è inoltre l'unica sottoformula di A uguale ad A.

Gruppo identità

$$\frac{}{\vdash \neg A^{\bar{\epsilon}}, A^{\epsilon}}\ (ax) \qquad \frac{\vdash \Gamma, A^{\epsilon} \qquad \vdash \Delta, \neg A^{\bar{\epsilon}}}{\vdash \Gamma, \Delta}\ (cut)$$

A e $\neg A$ si intendono formule *colorate* duali, dunque duali anche rispetto alla colorazione. Le conclusioni dell'assioma possono essere colorate liberamente purché si rispetti questa condizione.

Regole strutturali

$$\frac{\vdash \Gamma, A^{\epsilon}, A^{\epsilon}}{\vdash \Gamma, A^{\epsilon}}\ (C) \qquad \frac{\vdash \Gamma}{\vdash \Gamma, A^{\epsilon}}\ (W)$$

La conclusione indebolita introdotta dalla regola (W) può essere colorata liberamente. Nella regola (C) invece si richiede che le due premesse contratte abbiano colorazione identica alla conclusione.

Regole logiche

$$\frac{\vdash \Gamma, A^{\epsilon_1} \qquad \vdash \Delta, B^{\epsilon_2}}{\vdash \Gamma, \Delta, (A^{\epsilon_1} \wedge_m B^{\epsilon_2})^{\epsilon}}\ (\wedge_m) \qquad \frac{\vdash \Gamma, A^{\epsilon_1}, B^{\epsilon_2}}{\vdash \Gamma, (A^{\epsilon_1} \vee_m B^{\epsilon_2})^{\epsilon}}\ (\vee_m^i)$$

$$\frac{\vdash \Gamma, A_i^{\epsilon_i}}{\vdash \Gamma, (A_1^{\epsilon_1} \vee_a A_2^{\epsilon_2})^{\epsilon}}\ (\vee_a^i) \qquad \frac{\vdash \Gamma, A^{\epsilon_1} \qquad \vdash \Gamma, B^{\epsilon_2}}{\vdash \Gamma, (A^{\epsilon_1} \wedge_a B^{\epsilon_2})^{\epsilon}}\ (\wedge_a)$$

dove $i \in \{1,2\}$. Il colore ϵ della conclusione può essere scelto liberamente, mentre la colorazione delle sottoformule deve essere identica a quella delle premesse. Nel caso delle regole $(\vee_a^i)$, una delle due sottoformule non compare nelle premesse e può essere colorata liberamente.

Quantificatori del primo ordine

$$\frac{\vdash \Gamma, A(t/x)^{\epsilon_1}}{\vdash \Gamma, (\exists x.A(x)^{\epsilon_1})^{\epsilon}}\ (\exists) \qquad \frac{\vdash \Delta, A(x)^{\epsilon_1}}{\vdash \Delta, (\forall x.A^{\epsilon_1})^{\epsilon}}\ (\forall)$$

dove t è un termine qualsiasi e la variabile x, detta *variabile propria* della regola $(\forall)$, non occorre libera in Δ (scriveremo anche $x \notin FV^1(\Delta)$). Il colore ϵ della conclusione principale può essere scelto liberamente, mentre la colorazione di A deve essere preservata.

Quantificatori del secondo ordine

$$\frac{\vdash \Gamma, A[B/Xx_1 \dots x_n]^{\epsilon_1}}{\vdash \Gamma, (\exists X.A^{\epsilon_1})^{\epsilon}} (\exists^2) \qquad\qquad \frac{\vdash \Delta, A^{\epsilon_1}}{\vdash \Delta, (\forall X.A^{\epsilon_1})^{\epsilon}} (\forall^2)$$

dove B è una formula qualsiasi, X una variabile per predicati di arietà n che non occorre libera in Δ ($X \notin FV^2(\Delta)$). Si noti che il colore di B è irrilevante nella sostituzione: la nuova sottoformula prenderà caso per caso il colore dell'occorrenza di X sostituita. Come sempre, il colore ϵ della conclusione principale può essere scelto liberamente.

Definizione 2.36 (Derivazione in LKtq). Una derivazione in $\mathbf{LK}^{tq}$ è un albero di occorrenze di regole scelte fra quelle appena introdotte. Come nella definizione 2.12, a ogni occorrenza di regola r è associato un sequente conclusione σ_r di formule colorate. Le nozioni di *conclusione*, *ascendente* (e discendente), *albero strutturale* sono definite analogamente a **LK**.

Proprietà 2.37. *Il calcolo* $\mathbf{LK}^{tq}$ *è corretto rispetto alla semantica classica per la logica dei predicati del secondo ordine e completo al primo ordine.*

Dimostrazione. Ogni dimostrazione in $\mathbf{LK}^{tq}$ può essere tradotta in **LK** dimenticando i colori delle formule e sostituendo le regole $(\vee^1_m)$ e $(\vee^2_m)$ con la regola $(\vee_m)$.

Viceversa, ogni derivazione in **LK** può essere tradotta in $\mathbf{LK}^{tq}$ sostituendo liberamente la regola $(\vee_m)$ e colorando le formule in maniera adeguata. Almeno una colorazione è sempre possibile, per esempio assegnando il colore t a ogni formula e sottoformula *positiva* e q a ogni formula e sottoformula *negativa*. È facile verificare che formule duali avranno colorazione duale e formule uguali avranno colorazione identica (dunque sarà sempre possibile contrarre). $\square$

Osservazione 2.38. La dimostrazione appena conclusa riposa sul fatto che la distinzione delle formule in *positive* e *negative* rappresenta una partizione dell'insieme delle formule $\mathcal{F}$ che gode della seguente proprietà:

$$A \in \mathcal{P} \iff \neg A \in \mathcal{N}.$$

Questo fatto ci sarà utile in più occasioni.

2.5.2 Protocollo di riduzione e normalizzazione

Il protocollo di riduzione tq è definito in maniera simile alla riduzione in **LK**, con la differenza che si utilizzano i colori e le regole speciali per la disgiunzione moltiplicativa per eliminare il non determinismo dei passi strutturali e moltiplicativi.

Definizione 2.39 (Passo di riduzione tq). Il passo di riduzione $\rightsquigarrow_{Ltq}$ per derivazioni di $\mathbf{LK}^{tq}$ è definito come il passo $\rightsquigarrow_L$, con la differenza che il passo $(\wedge_m/\vee_m^1)$ (risp. $(\wedge_m/\vee_m^2)$) si applica solamente alla regola $(\vee_m^1)$ (risp. alla regola $(\vee_m^2)$).

Il passo di riduzione $\rightsquigarrow_{Stq}$ è definito come il passo $\rightsquigarrow_S$, con la differenza che, nel caso di un taglio di tipo $S1$, si trasporta sempre la sottoderivazione la cui formula di taglio ha colore t. Per questo motivo diremo che una formula di colore q è *attrattiva*.

Denotiamo con $\rightsquigarrow_{tq}$ l'unione dei passi $\rightsquigarrow_{Ltq}$ e $\rightsquigarrow_{Stq}$, con $\rightsquigarrow_{tq}^*$ la chiusura riflessiva e transitiva di $\rightsquigarrow_{tq}$.

Teorema 2.40. *La riduzione $\rightsquigarrow_{tq}^*$ è confluente e fortemente normalizzante ($\mathbf{LK}^{tq} \subseteq \mathrm{SN}_{tq}$). Le forme normali sono uniche a meno di indebolimenti contratti e permutazioni di regole strutturali.*

La dimostrazione del teorema si fonda sulla possibilità di immergere $\mathbf{LK}^{tq}$ in un frammento confluente e fortemente normalizzante della logica lineare. Per i dettagli rimandiamo a [DJS97]; osserviamo però che tutte le fonti di non determinismo identificate nella sezione 2.4 sono state eliminate: ogni scelta è completamente predeterminata.

Abbiamo già osservato che è possibile tradurre le derivazioni da **LK** in $\mathbf{LK}^{tq}$ e viceversa preservandone la struttura. Osserviamo inoltre che la definizione dei passi di riduzione è la stessa, con l'unica eccezione del processo di scelta nei casi dubbi. In altre parole, la riduzione in $\mathbf{LK}^{tq}$ corrisponde a quella in **LK** purché si facciano determinate scelte dove è necessario: come conseguenza otteniamo una procedura debolmente normalizzante per **LK** al secondo ordine.

Corollario 2.41. $\mathbf{LK} \subseteq \mathrm{WN}$.

Domanda. Il risultato di normalizzazione forte per $\mathbf{LK}^{tq}$ implica la normalizzazione forte per **LK**?

La risposta è con tutta probabilità negativa. Sia π una derivazione in **LK** e $Col(\pi)$ l'insieme di tutte le sue possibili traduzioni[7] in $\mathbf{LK}^{tq}$: a ogni traduzione corrisponde una forma normale in $\mathbf{LK}^{tq}$ che sarà anche una forma normale di π in **LK**.

Tuttavia, come osservato in [BBS97], esistono forme normali di π raggiungibili in **LK** (e nel λ-calcolo simmetrico utilizzato nell'articolo) che non sono raggiungibili in $\mathbf{LK}^{tq}$. La ragione è sottile ed è stata discussa brevemente in [Lai01]. Supponiamo che π contenga il taglio seguente:

$$\dfrac{\begin{array}{c}\vdots\ \pi_1\\ \vdash \Gamma, A\end{array} \qquad \begin{array}{c}\vdots\ \pi_2\\ \vdash \Delta, \neg A\end{array}}{\vdash \Gamma, \Delta}\ (cut)$$

[7] Ogni traduzione consiste in una colorazione delle formule che occorrono in π più una scelta fra $(\vee_m^1)$ e $(\vee_m^2)$ per ogni occorrenza della regola $(\vee_m)$.

e supponiamo che il taglio sia strutturale di tipo $S1$ (entrambe le formule di taglio non principali). Supponiamo inoltre che π si riduca in π' con un passo $\rightsquigarrow_S$ che trasporta π_1: π' conterrà un certo numero di copie di π_1 (supponiamo che siano più di una).

Ora scegliamo una traduzione $\pi_c \in Col(\pi)$ di π in $\mathbf{LK}^{tq}$, tale che la formula di taglio $\neg A$ sia attrattiva (di colore q) e A non attrattiva (t). Riducendo lo stesso taglio con un passo $\rightsquigarrow_{Stq}$ otterremo $\pi'_c \in Col(\pi')$ che conterrà alcune copie della traduzione di π_1. Il problema è che ognuna di queste copie sarà colorata allo stesso modo e pertanto si ridurrà allo stesso modo (facendo le stesse scelte), mentre in **LK** niente proibisce di fare scelte diverse per ogni copia: in buona sostanza **LK** rappresenta una versione di $\mathbf{LK}^{tq}$ in cui è possibile cambiare completamente colorazione dopo ogni passo di riduzione (c'è un'osservazione simile in [DJS97] a proposito del λ-calcolo simmetrico di [BB96]).

Questa difficoltà provoca il fallimento delle strategie più semplici ma anche di un approccio più raffinato. In generale si può mostrare che se $\pi \rightsquigarrow_1 \pi'$ allora esistono $\pi_c \in Col(\pi)$ e $\pi'_c \in Col(\pi')$ tali che $\pi_c \rightsquigarrow_{tq} \pi'_c$. L'idea allora è di considerare l'albero di riduzione di π e ad ogni nodo ρ associare l'insieme delle sue traduzioni $Col(\rho)$. Si vorrebbe definire una misura su questo insieme e sfruttare il fatto appena enunciato per mostrare che la misura decresce lungo ogni cammino dell'albero di riduzione, da cui si avrebbe che la lunghezza di ogni ramo è finita e $\pi \in \mathrm{SN}$.

Purtroppo nel caso generale esisterà sempre $\delta \in Col(\pi')$ tale che per nessun $\gamma \in Col(\pi)$ valga $\gamma \rightsquigarrow^*_{tq} \delta$: in altre parole esistono delle colorazioni di π' che sono irraggiungibili per tq-riduzione dalle colorazioni di π. Per questo motivo sembra impossibile mettere in relazione una eventuale misura non triviale su $Col(\pi)$ con la stessa misura su $Col(\pi')$, se non in casi particolari (come derivazioni lineari o riduzioni di tagli $S2$).

2.6 Residui, interspazi e lemma di stabilità

Concludiamo il capitolo presentando alcuni strumenti tecnici per l'analisi delle derivazioni e dei processi di riduzione.

Definizione 2.42. Sia π una derivazione in **LK** e c un'occorrenza della regola di taglio in π. Chiamiamo *passo di riduzione di c* e denotiamo con $\rightsquigarrow_{[c]}$ il sottoinsieme di $\rightsquigarrow_1$ tale che $\pi \rightsquigarrow_{[c]} \pi'$ sse $\pi \rightsquigarrow_1 \pi'$ *riducendo il taglio c*.

Osservazione 2.43. In un protocollo di riduzione deterministico come quello di $\mathbf{LK}^{tq}$ (o del λ-calcolo puro) ogni taglio può essere ridotto in un solo modo senza ambiguità: data una derivazione π con un'occorrenza di taglio c esiste un'unica derivazione π' tale che $\pi \rightsquigarrow_{[c]} \pi'$. In **LK**, invece, la riduzione di un taglio strutturale $S1$ o di un taglio logico moltiplicativo è ambigua e in questi casi esisteranno *due* derivazioni π'_1 e π'_2 tali che $\pi \rightsquigarrow_{[c]} \pi'_1$ e $\pi \rightsquigarrow_{[c]} \pi'_2$.

Consideriamo una coppia di derivazioni π, π' in **LK** o $\mathbf{LK}^{tq}$ tali che $\pi \rightsquigarrow_1 \pi'$ (in **LK**) o $\pi \rightsquigarrow_{tq} \pi'$ (in $\mathbf{LK}^{tq}$). Osserviamo che, necessariamente, ogni occorrenza di una regola *logica o assioma* r in π' *proviene da un'unica occorrenza* r_* della stessa regola in π.[8] Analogamente, ogni occorrenza di un taglio c in π' proviene da un'unica occorrenza c_* in π con le stesse formule di taglio,[9] a meno che c non sia uno dei nuovi tagli introdotti durante un passo di riduzione logico.

Definizione 2.44 (Lift, residui). Per qualsiasi coppia di derivazioni π, π' in **LK** (o in $\mathbf{LK}^{tq}$) tale che $\pi \rightsquigarrow_1 \pi'$ ($\pi \rightsquigarrow_{tq} \pi'$) e per ogni occorrenza di regola logica o assioma r in π' definiamo il *lift* r_* *di* r *in* π come l'unica occorrenza della stessa regola in π da cui r proviene (nel senso appena discusso). Inoltre per ogni occorrenza di formula A^x *attiva in* r definiamo il lift A^x_* in π come l'occorrenza corrispondente (a meno di sostituzioni) attiva in r_*.

Per ogni taglio c in π' che non sia stato creato da un passo di riduzione logico $\pi \rightsquigarrow_L \pi'$ ($\pi \rightsquigarrow_{Ltq} \pi'$) definiamo il lift c_* di c in π come l'unica occorrenza di taglio in π da cui c proviene. Anche in questo caso ogni formula di taglio A^x di c corrisponderà a una formula di taglio di c_* identica a meno di sostituzioni che denoteremo con A^x_*.

Per ogni occorrenza s in π di regola logica o taglio, definiamo l'insieme dei *residui di* s *in* π' come l'insieme delle occorrenze di regole r in π' tali che $r_* = s$.

Diremo che le occorrenze di taglio *create* in un passo logico che riduce *pi* in π' *non hanno lift* in π, e il taglio logico ridotto *non ha residui* in π'.

Definizione 2.45 (Interspazio principale-attiva). Data una derivazione π in **LK** (o in $\mathbf{LK}^{tq}$) e un'occorrenza di formula A^x che sia conclusione principale o indebolita di una occorrenza di regola r,[10] possiamo definire il sottoinsieme ordinato B_{A^x} di π che contiene tutte le occorrenze di regola s la cui conclusione σ_s contenga un'occorrenza A^y discendente di A^x.[11]

Evidentemente l'insieme ordinato B_{A^x} è un cammino in π: la prima occorrenza di regola[12] in B_{A^x} sarà r, denotiamo l'ultima con r' e con A^z l'occorrenza che è discendente di A^x nella conclusione $\sigma_{r'}$ di r'. Avremo che r' è la radice di π oppure A^z è attiva in qualche regola logica o taglio in π e in ogni caso non ha discendenti: B_{A^x} è un ramo dell'albero strutturale di A^z in π e r è una delle foglie.

Chiamiamo B_{A^x} l'*interspazio principale-attiva di* A^x (o *interspazio di* A^x per brevità). Ad ogni occorrenza di formula A^z in π attiva in una regola logica o

[8] Questo può accadere in due modi: r_* appartiene alla porzione di π che non viene modificata dal passo di riduzione e rimane identica in π' a meno di sostituzioni di sottoformule o termini, oppure r_* appartiene alla sottoderivazione di π che è stata trasportata da un passo $\rightsquigarrow_S$ ($\rightsquigarrow_{Stq}$ nel caso di $\mathbf{LK}^{tq}$) e r occorre in una delle copie di tale sottoderivazione.

[9] In tal caso, oltre alle possibilità discusse nella nota precedente, si aggiungerà la possibilità che il taglio c sia stato introdotto da un passo $\rightsquigarrow_S$ ($\rightsquigarrow_{Stq}$) che riduceva c_*.

[10] Dunque r sarà una regola logica, un assioma o un indebolimento.

[11] Un'occorrenza di cui A^x sia ascendente (definizione 2.21).

[12] L'estremo inferiore di B_{A^x} nell'ordine arborescente, la più lontana dalla radice di π.

strutturale, possiamo associare l'insieme finito $\{B_{A^{x_1}}, \dots, B_{A^{x_n}}\}$ degli interspazi che terminano con A^z, che coincide con l'insieme dei rami dell'albero strutturale di A^z in π dove ogni A^{x_i} occorre in una delle foglie.

Intuitivamente, l'interspazio B_{A^x} di A^x "traccia il percorso" dal punto in cui A^x viene introdotta nella derivazione fino al punto in cui viene utilizzata. Quando B_{A^x}, oltre alla regola che introduce A^x, contiene solamente regole strutturali che *non operano* su discendenti di A^x, diremo che A^x ha un interspazio *lucido*: in tal caso, se A^y è l'ultima occorrenza discendente di A^x (quella attiva o conclusiva), allora B_{A^x} è *l'unico ramo* dell'albero strutturale di A^y e l'unico elemento dell'insieme degli interspazi. Come caso particolare, quando B_{A^x} contiene solamente la regola che introduce A^x (cioè A^x è utilizzata immediatamente) diremo che A^x ha un interspazio *piatto*.

$$
\cfrac{\cfrac{\cfrac{\cfrac{}{\vdash \Gamma, A^x}\,m/w}{\vdots}\,s_1}{\vdash \Gamma', A^x}\,s_n}{\vdots}\,a
\qquad\qquad
\cfrac{\cfrac{}{\vdash \Gamma, A^x}\,m/w}{\vdots}\,a
$$

Figura 2.2: Esempi di interspazio lucido e piatto. Le regole $s_1, \dots, s_n$ sono strutturali su occorrenze che *non discendono* da A^x. a è la regola logica o taglio che utilizza A^x e non appartiene all'interspazio.

È utile chiedersi cosa accade gli interspazi durante il processo di riduzione. Abbiamo detto in particolare che per una certa occorrenza A^y attiva in una regola logica o in un taglio in qualche derivazione π, se uno degli interspazi B_{A^x} associati ad A^y è lucido o piatto, allora è anche *l'unico* interspazio associato ad A^y. Supponiamo dunque che $\pi \rightsquigarrow_1 \pi'$ ($\pi \rightsquigarrow_{tq} \pi'$ in $\mathbf{LK}^{tq}$): per la definizione 2.44 possiamo parlare dei *residui* di A^y in π'. Tali residui saranno occorrenze attive di A (a meno di sostituzioni) alle quali saranno associati vari interspazi che per estensione chiameremo *residui di* B_{A^x}: questi ultimi godono in $\mathbf{LK}^{tq}$ di una proprietà molto utile detta *lemma di stabilità*.

Lemma 2.46 (Lemma di stabilità per LKtq). *Sia π una derivazione in $\mathbf{LK}^{tq}$, π' tale che $\pi \rightsquigarrow_{tq} \pi'$, e sia B_{A^x} l'interspazio di una occorrenza di formula attrattiva*[13] *A^x. Se B_{A^x} è lucido (risp. piatto), allora tutti i residui di B_{A^x} in π' sono lucidi (risp. piatti).*

Per la dimostrazione rimandiamo a [DJS97, pp. 776–777], l'argomento è simile a quello che presenteremo a breve per **LK**. Ricordiamo che l'ipotesi 'attrattiva'

[13] Cioè di colore q, nella nostra versione di $\mathbf{LK}^{tq}$.

su A^x è necessaria solo nel caso in cui A^x sia la conclusione di un assioma: sfruttiamo questo fatto per adattare il risultato a **LK** (dove ovviamente l'ipotesi non può essere formulata allo stesso modo).

Lemma 2.47 (Lemma di stabilità per LK). *Sia π una derivazione in* **LK***, π' tale che $\pi \rightsquigarrow_1 \pi'$, e sia B_{A^x} l'interspazio di una occorrenza di formula A^x indebolita, principale in una regola logica, o principale in un assioma che ha dei residui in π'. Se B_{A^x} è lucido (risp. piatto), allora tutti i residui di B_{A^x} in π' sono lucidi (risp. piatti).*

Dimostrazione. Sia c il taglio ridotto nel passaggio da π a π' (equivalentemente, sia c tale che $\pi \rightsquigarrow_{[c]} \pi'$). Se c è di tipo L, allora dobbiamo considerare due possibilità:

a. A^x è la formula di taglio: in tal caso, ricordando che per definizione un taglio logico ridotto non ha residui, né A^x né a maggior ragione B_{A^x} hanno residui in π' e la conclusione è soddisfatta in modo vacuo;

b. A^x non è la formula di taglio: per lucidità di B_{A^x}, c non occorre in B_{A^x} e l'interspazio non è modificato dalla riduzione (a meno di sostituzioni di sottoformule o termini).

Altrimenti c è un taglio strutturale e possiamo identificare in π (i) una sottoderivazione trasportata δ e (ii) l'albero strutturale della formula di taglio di c che ha *comportamento attrattivo*.[14] Sia dunque A^y l'ultima occorrenza che discendente da A^x (cioè quella attiva che termina l'interspazio); ricordiamo che i residui di B_{A^x} sono definiti utilizzando i residui di A^y, che a loro volta sono definiti utilizzando i residui della regola in cui A^y è attiva (definizione 2.44).

Distinguiamo due casi. Primo, A^y è una delle due formule di taglio (dunque ogni residuo di A^y sarà attivo in un residuo di c) e si danno due sottocasi:

a. A^y occorre nella conclusione di δ (la sottoderivazione trasportata). In tal caso B_{A^x} è interamente contenuto in δ e ogni residuo di B_{A^x} in π è contenuto in una copia di δ che avrà fra le sue conclusioni un'occorrenza A^z residua di A^y e attiva in un residuo di c, dunque ogni residuo di B_{A^x} è una copia di B_{A^x} ed è lucido (risp. piatto);

b. A^y è la formula di taglio con comportamento attrattivo. Ricordiamo che B_{A^x} è l'unico ramo dell'albero strutturale di A^y in π. Se A^x è conclusione di un assioma o di un indebolimento, allora B_{A^x} non ha residui in π', altrimenti A^x è conclusione di una regola logica e l'unico residuo di B_{A^x} è piatto.

[14]Cioè quella delle due che non occorre nella conclusione di δ. Ovviamente in **LK** le formule non sono colorate, perciò non possiamo parlare di *formula attrattiva*; tuttavia, quando riduciamo un taglio strutturale, dobbiamo scegliere una sottoderivazione da trasportare e potremo dire che l'altra formula di taglio si comporta "come se fosse attrattiva in $\mathbf{LK}^{tq}$".

Secondo e ultimo caso, A^y non è una delle formule di taglio di c. Distinguiamo ancora una volta tre sottocasi:

a. B_{A^x} fa parte di δ: in tal caso, per lucidità, B_{A^x} è completamente contenuto in δ, ogni eventuale residuo è una copia di B_{A^x} ed è pertanto lucido (risp. piatto);

b. B_{A^x} non fa parte di δ e *non contiene* foglie dell'albero strutturale della formula di taglio con comportamento attrattivo: in tal caso l'unico residuo di B_{A^x} in π' è uguale a B_{A^x} a meno di un "cambio di contesto": eventuali contrazioni nell'albero strutturale della formula di taglio sono rimpiazzate da contrazioni sul contesto di δ. Pertanto il residuo potrà essere lucido o piatto, e in particolare sarà piatto se lo è anche B_{A^x};

c. B_{A^x} non fa parte di δ e *contiene* una o più foglie dell'albero strutturale della formula di taglio. Se una di queste foglie è un assioma, sarà la prima regola di B_{A^x}, cioè lo stesso assioma che introduce A^x: in tal caso l'assioma non ha residui in π' (è sostituito dalla sottoderivazione trasportata) e l'ipotesi non è soddisfatta. Altrimenti B_{A^x} è lucido e le foglie sono indebolimenti che occorrono in B_{A^x} (e non possono essere la prima regola di B_{A^x}): in tal caso il residuo di B_{A^x} in π' potrà contenere alcuni indebolimenti ed eventualmente contrazioni in più sul contesto di δ e sarà ancora lucido.

Con questo tutte le possibilità sono esaurite e possiamo concludere la dimostrazione. $\square$

Capitolo 3

Il λ-calcolo simmetrico

Questo capitolo è dedicato alla presentazione del λ-calcolo simmetrico introdotto da F. Barbanera e S. Berardi in [BB96]. Come annunciato nell'introduzione, ci concentreremo sul frammento proposizionale $\lambda^{\text{Sym}}_{\text{Prop}}$: analizzeremo la tecnica dei *candidati simmetrici* utilizzata per dimostrare il teorema di normalizzazione forte e presenteremo un confronto fra la struttura delle derivazioni in **LK** e quella dei termini di $\lambda^{\text{Sym}}_{\text{Prop}}$che fornirà una guida per l'applicazione della stessa tecnica dimostrativa alla riduzione in **LK**.

3.1 Tipi e termini del frammento proposizionale

Definizione 3.1. Fissato l'insieme $\mathcal{A} = \{a, b, \dots\}$ dei *tipi atomici* e l'insieme $\mathcal{A}^{\perp} = \{a^{\perp}, b^{\perp}, \dots\}$ dei *tipi atomici negati*, definiamo attraverso le seguenti grammatiche

(i) l'insieme degli *m-tipi* di $\lambda^{\text{Sym}}_{\text{Prop}}$:

$$A, B ::= \alpha \mid \alpha^{\perp} \mid A \wedge B \mid A \vee B$$

dove $\alpha, \alpha^{\perp}$ sono elementi qualsiasi rispettivamente di $\mathcal{A}$ e $\mathcal{A}^{\perp}$.

(ii) l'insieme dei tipi di $\lambda^{\text{Sym}}_{\text{Prop}}$:

$$C ::= A \mid \perp$$

La negazione involutiva $(\cdot)^{\perp}$ è definita sugli *m-tipi* come di consueto utilizzando le dualità di De Morgan:

$$\begin{aligned}
(\alpha)^{\perp} &= \alpha^{\perp} \\
(\alpha^{\perp})^{\perp} &= \alpha \\
(A \wedge B)^{\perp} &= (A)^{\perp} \vee (B)^{\perp} \\
(A \vee B)^{\perp} &= (A)^{\perp} \wedge (B)^{\perp}
\end{aligned}$$

Si verifica per induzione su A che $A^{\perp\perp} = A$.

Gli m-tipi sono definiti separatamente dai tipi per impedire che la formula per l'assurdità $\perp$ sia utilizzata come sottoformula propria di un tipo qualsiasi.

Come osservato in [BB96], questa non è una limitazione perché la formula $A \wedge \perp$ è classicamente equivalente a $\perp$ e la formula $A \vee \perp$ è equivalente ad A. Questa scelta è giustificata da ragioni tecniche: sarà utile nel seguito poter assumere che i termini di tipo $\perp$ abbiano una forma particolare.

Vedremo attraverso il confronto con **LK** che il tipo $\perp$ non denota l'unità logica del falso (l'elemento neutro della disgiunzione), ma l'assenza di una conclusione "in evidenza" (dal punto di vista della deduzione naturale, la presenza di un'assurdità nella dimostrazione): come osservato da Parigot in [Par00], questo è un passaggio chiave per ottenere un sistema tipato classico consistente (nel senso che i tipi sono preservati sotto riduzione).

Nel seguito di questo capitolo denoteremo gli m-tipi con $A, B, A_1, A_2, \ldots$, mentre i tipi saranno denotati da $C, D, C_1, C_2, \ldots$.

Definizione 3.2. L'insieme dei termini di $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$ è definito dalla grammatica

$$t,u \ ::= \ x \mid \langle t,u \rangle \mid \sigma_i(t) \mid t \star u \mid \lambda x.t$$

dove x è una variabile, $i \in \{1,2\}$. I termini sono tipati dalle seguenti regole:

$$x : A \quad (var) \qquad\qquad \begin{array}{c} [x:A] \\ \vdots \\ t : \perp \\ \hline \lambda x.t : A^{\perp} \end{array} (\lambda)$$

$$\frac{t : A \qquad u : A^{\perp}}{t \star u : \perp}(\star) \qquad\qquad \frac{t : A \qquad u : B}{\langle t,u \rangle : A \wedge B}(\wedge) \qquad \frac{t : A_i}{\sigma_i(t) : A_1 \vee A_2}(\vee_i)$$

Come d'abitudine, consideriamo identici i termini che differiscono solamente per il nome delle variabili vincolate. Nel seguito il tipo di un termine sarà spesso indicato in apice: x^A indica che la variabile x è di tipo A.

Proposizione 3.3. *Il sistema di tipi di $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$ è corretto e completo rispetto alla derivabilità classica proposizionale.*

Dimostrazione. Si verifica facilmente che le regole sono corrette; l'implicazione si può trattare come un connettivo definito; le regole mancanti rispetto alla deduzione naturale tradizionale sono derivabili. Mostriamo di seguito la derivazione della regola di eliminazione della congiunzione (per $i \in \{1,2\}$):

$$\dfrac{\dfrac{A_1 \wedge A_2 \qquad \dfrac{[A_i^{\perp}]}{A_1^{\perp} \vee A_2^{\perp}}(\vee_i)}{\perp}(\star)}{A_i}(\lambda)$$

□

L'operatore $\star$ è detto *applicazione simmetrica* perché, dati i termini t^A e $u^{A^\perp}$, sia $t \star u$ che $u \star t$ sono termini corretti di $\lambda^{\text{Sym}}_{\text{Prop}}$. Questa simmetria è riflessa nelle regole di riduzione che presenteremo a breve in coppie duali.

3.2 Regole di riduzione e non confluenza

Definizione 3.4. Definiamo le regole di riduzione sui termini di $\lambda^{\text{Sym}}_{\text{Prop}}$come segue:

$$\begin{cases} (\beta) & (\lambda x.t) \star u \rightsquigarrow_\beta \ t[u/x] \\ (\beta^\perp) & u \star (\lambda x.t) \rightsquigarrow_{\beta^\perp} t[u/x] \end{cases}$$

$$\begin{cases} (\eta) & \lambda x.t \star x \rightsquigarrow_\eta \ t \quad x \notin FV(t) \\ (\eta^\perp) & \lambda x.x \star t \rightsquigarrow_{\eta^\perp} t \quad x \notin FV(t) \end{cases}$$

$$\begin{cases} (\pi) & \langle t_1, t_2 \rangle \star \sigma_i(u) \rightsquigarrow_\pi \ t_i \star u \quad i \in \{1,2\} \\ (\pi^\perp) & \sigma_i(u) \star \langle t_1, t_2 \rangle \rightsquigarrow_{\pi^\perp} u \star t_i \quad i \in \{1,2\} \end{cases}$$

Denotiamo con $\rightsquigarrow_{Sym}$ l'unione delle relazioni appena definite, con $\rightsquigarrow^*_{Sym}$ la chiusura riflessiva e transitiva di $\rightsquigarrow_{Sym}$.

La regola di riduzione (π) è ispirata alla regola di riduzione $(\wedge_a/\vee^i_a)$ del calcolo dei sequenti:

$$\dfrac{\dfrac{t_1 : A_1 \qquad t_2 : A_2}{\langle t_1, t_2 \rangle : A_1 \wedge A_2}(\wedge) \qquad \dfrac{u : A_i^\perp}{\sigma_i(u) : A_1^\perp \vee A_2^\perp}(\vee_i)}{\langle t_1, t_2 \rangle \star \sigma_i(u) : \perp}(\star) \quad \rightsquigarrow_\pi \quad \dfrac{t_i : A_i \qquad u : A_i^\perp}{t_i \star u : \perp}(\star)$$

La regola di riduzione (β) invece ha uno statuto particolare. Siano $t^\perp, u^\perp$ una coppia di termini di tipo $\perp$ e consideriamo il termine seguente:

$$\lambda x^A.t \star \lambda y^{A^\perp}.u.$$

Ciascuno dei due sottotermini $\lambda x.t$ e $\lambda y.u$ può essere trattato come una funzione o come un argomento:

$$\lambda x.t \star \lambda y.u \rightsquigarrow_\beta \ t[\lambda y.u/x],$$
$$\lambda x.t \star \lambda y.u \rightsquigarrow_{\beta^\perp} u[\lambda x.t/y].$$

Si può vedere che intuitivamente la λ-astrazione seguita da applicazione simmetrica in $\lambda^{\text{Sym}}_{\text{Prop}}$ corrisponde al taglio strutturale in **LK** e i passi (β) e $(\beta^\perp)$ corrispondono ai due modi di ridurre un taglio di tipo $S1$: l'analogia sarà resa più precisa in seguito. In generale i due ridotti potranno essere molto diversi fra di loro e si può mostrare che in alcuni casi avranno forme normali diverse (si veda p. es. [BB96, p. 105] o l'analisi approfondita in [BBS97]). La riduzione $\rightsquigarrow^*_{Sym}$ è dunque non confluente; nonostante questo si può mostrare che $\lambda^{\text{Sym}}_{\text{Prop}}$ gode di normalizzazione forte.

3.3 Normalizzazione forte

La normalizzazione forte per $\lambda^{\text{Sym}}_{\text{Prop}}$ è dimostrata in [BB96] attraverso una variante non triviale del metodo della riducibilità di Tait e Girard (si vedano ad esempio [GLT89; Gir87]), che chiameremo "tecnica dei *candidati simmetrici*". Abbiamo scelto nella presentazione di favorire la leggibilità rispetto alla precisione: per questo utilizzeremo una notazione più leggera rispetto a quella di [BB96] lasciando spesso sottointeso il tipo degli insiemi di termini.

Definizione 3.5. Sia t un termine in $\lambda^{\text{Sym}}_{\text{Prop}}$. Denotiamo con $\nu(t)$ l'altezza (eventualmente infinita) dell'albero di riduzione di t. Denotiamo con $\text{SN}_{Sym} \subseteq \lambda^{\text{Sym}}_{\text{Prop}}$ l'insieme di termini tale che $t \in \text{SN}_{Sym}$ sse $\nu(t) \in \mathbb{N}$.

Teorema 3.6 (Normalizzazione forte per $\boldsymbol{\lambda^{\text{Sym}}_{\text{Prop}}}$). $\lambda^{\text{Sym}}_{\text{Prop}} \subseteq \text{SN}_{Sym}$.

Useremo la notazione $\mathcal{V}_C$ per l'insieme delle *variabili* di tipo C, $\mathcal{T}_C$ per l'insieme dei *termini* di tipo C, SN_C per l'insieme dei termini *fortemente normalizzanti* di tipo C ($\text{SN}_{Sym} \cap \mathcal{T}_C$).

3.3.1 Candidati e riducibilità

Si vuole associare ad ogni tipo C un insieme $\text{RED}(C)$ [1] di *termini riducibili* di tipo C e mostrare che $\text{RED}(C) \subseteq \text{SN}_{Sym}$, dopodiché mostreremo per induzione sull'altezza dei termini che ogni termine di tipo C è riducibile.

Per poter completare quest'ultima dimostrazione, è necessario fornire delle condizioni sufficienti "abbastanza deboli" perché un termine di tipo C con una certa forma appartenga all'insieme $\text{RED}(C)$:

(i) $x^A \in \text{RED}(A)$;

(ii) $\langle t, u\rangle \in \text{RED}(A \wedge B) \iff t \in \text{RED}(A)$ e $u \in \text{RED}(B)$;

(iii) $\sigma_i(t) \in \text{RED}(A_1 \vee A_2) \iff t \in \text{RED}(A_i)$ con $i \in \{1,2\}$;

(iv) $\lambda x.t \in \text{RED}(A) \iff \forall u \in \text{RED}(A^\perp).\, t[u/x] \in \text{RED}(\perp)$;

(v) $t \star u \in \text{RED}(\perp) \iff t \star u \in \text{SN}_{Sym}$.

La prima e l'ultima condizione possono anche essere espresse così: $\mathcal{V}_A \subseteq \text{RED}(A)$, $\text{RED}(\perp) = \text{SN}_\perp$. Ricordiamo che i termini di tipo $\perp$ avranno sempre la forma di un'applicazione simmetrica.

Osserviamo con [BB96] che queste condizioni non sono sufficienti a fornire una definizione induttiva dell'insieme $\text{RED}(A)$. Il motivo è il riferimento all'insieme

[1] Abbiamo preferito la notazione *à la Girard* $\text{RED}(C)$ a quella di [BB96] $[\![C]\!]$ considerando il fatto che nell'ambito della logica lineare la notazione $[\![-]\!]$ è spesso utilizzata per la semantica denotazionale.

$\mathrm{RED}(A^{\perp})$ nella quarta condizione: questo a sua volta sarebbe definito con un riferimento a $\mathrm{RED}(A^{\perp\perp}) = \mathrm{RED}(A)$ e le due definizioni sarebbero circolari. I due insiemi dunque devono essere definiti *in coppia*: per questo motivo Barbanera e Berardi parlano di *candidati simmetrici*.

La tecnica di [BB96] somiglia in parte a quella proposta da Girard in [Gir87] per dimostrare la normalizzazione forte della logica lineare al secondo ordine: si definisce una nozione di ortogonalità fra candidati e si pone $\mathrm{RED}(A^{\perp}) = \mathrm{RED}(A)^{\perp}$. Per ottenere questo però la strategia è diversa da quella di Girard: si utilizza il teorema del punto fisso di Tarski e si definisce $\mathrm{RED}(A)$ come il più piccolo punto fisso del biortogonale $(\cdot)^{\perp\perp}$ e $\mathrm{RED}(A^{\perp})$ come $\mathrm{RED}(A)^{\perp}$.

Cominciamo introducendo un piccolo abuso di notazione: vogliamo applicare gli operatori del λ-calcolo simmetrico a insiemi di termini. Siano dunque $X \subseteq \mathcal{T}_A$, $X_1 \subseteq \mathcal{T}_{A_1}, X_2 \subseteq \mathcal{T}_{A_2}, i \in \{1,2\}$:

$$\begin{aligned}
\langle X_1, X_2 \rangle &= \{\langle t, u \rangle : t \in X_1 \wedge u \in X_2\}, \\
\sigma_i(X_i) &= \{\sigma_i(t) : t \in X_i\}, \\
\lambda(X) &= \{\lambda x.t : x \in \mathcal{V}_A \wedge t \in \mathcal{T}_{\perp} \wedge \forall u \in X.\, t[u/x] \in \mathrm{SN}_{Sym}\}.
\end{aligned}$$

Osserviamo che gli elementi di $\langle X_1, X_2 \rangle$ saranno di tipo $A_1 \wedge A_2$, gli elementi di $\sigma_i(X_i)$ saranno di tipo $A_1 \vee A_2$ e quelli di $\lambda(X)$ saranno di tipo $A^{\perp}$: è necessario prestare attenzione al fatto che l'operatore $\lambda(X)$ nega il tipo dell'insieme X.

Per formulare correttamente la definizione di ortogonalità è necessario disporre di una *partizione* dell'insieme degli m-tipi in due sottoinsiemi (diciamo $\mathcal{F}_1$ e $\mathcal{F}_2$) che soddisfi la seguente condizione:

$$A \in \mathcal{F}_1 \iff A^{\perp} \in \mathcal{F}_2.$$

Questo è possibile grazie alla suddivisione dei tipi atomici nei due insiemi $\mathcal{A}$ e $\mathcal{A}^{\perp}$ (definizione 3.1).

Definizione 3.7. Sia A un m-tipo di forma α o $A_1 \wedge A_2$ con A_1 e A_2 m-tipi qualsiasi. Siano inoltre $X \subseteq \mathcal{T}_A$ e $Y \subseteq \mathcal{T}_{A^{\perp}}$ insiemi qualsiasi di termini di tipo A e $A^{\perp}$ e definiamo per induzione sull'altezza di A gli insiemi $X^{\perp} \subseteq \mathcal{T}_{A^{\perp}}$ e $Y^{\perp} \subseteq \mathcal{T}_A$. Definiamo contemporaneamente ad ogni passo induttivo l'insieme $\mathrm{RED}(A)$ come il più piccolo insieme tale che $\mathrm{RED}(A) = \mathrm{RED}(A)^{\perp\perp}$ e poniamo $\mathrm{RED}(A^{\perp}) = \mathrm{RED}(A)^{\perp}$.

– $A = \alpha, A^{\perp} = \alpha^{\perp}$:

$$\begin{aligned}
X^{\perp} &= \mathcal{V}_{A^{\perp}} \cup \lambda(X), \\
Y^{\perp} &= \mathcal{V}_A \cup \lambda(Y).
\end{aligned}$$

– $A = A_1 \wedge A_2, A^{\perp} = A_1^{\perp} \vee A_2^{\perp}$:

$$\begin{aligned}
X^{\perp} &= \mathcal{V}_{A^{\perp}} \cup \sigma_1(\mathrm{RED}(A_1^{\perp})) \cup \sigma_2(\mathrm{RED}(A_2^{\perp})) \cup \lambda(X), \\
Y^{\perp} &= \mathcal{V}_A \cup \langle \mathrm{RED}(A_1), \mathrm{RED}(A_2) \rangle \cup \lambda(Y).
\end{aligned}$$

Osserviamo che $\mathrm{RED}(A) = \mathrm{RED}(A)^{\perp\perp} = \mathrm{RED}(A^{\perp})^{\perp}$: per qualsiasi m-tipo B vale dunque in generale $\mathrm{RED}(B^{\perp}) = \mathrm{RED}(B)^{\perp}$. Poniamo infine $\mathrm{RED}(\perp) = \mathrm{SN}_{\perp}$.

Osservazione 3.8. La definizione 3.7 è ben posta. Considerando gli insiemi $\mathcal{P}(\mathcal{T}_A)$ e $\mathcal{P}(\mathcal{T}_{A^\perp})$ ordinati per inclusione, si verifica facilmente che l'operatore $\lambda(X) : \mathcal{P}(\mathcal{T}_A) \to \mathcal{P}(\mathcal{T}_{A^\perp})$ è decrescente, cioè per $X, Y \subseteq \mathcal{T}_A$ vale

$$X \subseteq Y \implies \lambda(Y) \subseteq \lambda(X).$$

Di conseguenza, anche l'operatore $X^\perp$ è decrescente e l'operatore $X^{\perp\perp}$ ottenuto per composizione sarà crescente:

$$X \subseteq Y \implies X^{\perp\perp} \subseteq Y^{\perp\perp}.$$

Allora, per il teorema del punto fisso di Tarski, esiste un minimo punto fisso del biortogonale e la definizione degli insiemi $\mathrm{RED}(A)$ è ben posta. Osserviamo che gli insiemi $\mathrm{RED}(A)$ devono essere definiti per induzione simultaneamente agli ortogonali $X^\perp, Y^\perp$ perché la definizione di questi ultimi si appoggia sull'esistenza di $\mathrm{RED}(B)$ per qualche m-tipo B di altezza strettamente minore di A.

Osserviamo infine che la definizione di $\mathrm{RED}(A)$ soddisfa i cinque requisiti che avevamo elencato. La verifica è banale per l'ultimo punto relativo a $RED(\perp)$; per gli altri osserviamo che $RED(A) = RED(A^\perp)^\perp$ e i vari punti seguiranno dalla definizione di ortogonale (per esempio, se $t \in \mathrm{RED}(A)$ e $u \in \mathrm{RED}(B)$ allora evidentemente $\langle t, u\rangle \in \langle \mathrm{RED}(A), \mathrm{RED}(B)\rangle \subseteq RED(A^\perp \vee B^\perp)^\perp = \mathrm{RED}(A \wedge B)$).

Proposizione 3.9. $\mathrm{RED}(C) \subseteq \mathrm{SN}_{Sym}$.

Dimostrazione. Per induzione su C considerando la struttura di $t \in \mathrm{RED}(C)$:

- $t = x^C$: la conclusione è immediata.

- $t = \langle u, v\rangle$ con $C = A \wedge B$: per la condizione (ii) abbiamo $u \in \mathrm{RED}(A)$ e $v \in \mathrm{RED}(B)$ e per ipotesi induttiva $u, v \in \mathrm{SN}_{Sym}$. Dato che ogni riduzione su t è una riduzione su u o su v, possiamo concludere che $t \in \mathrm{SN}_{Sym}$.

- $t = \sigma_i(u)$ con $C = A \vee B$, $i \in \{1, 2\}$: il ragionamento è analogo al caso precedente sfruttando la condizione (iii).

- $t = \lambda x^{C^\perp}.u$: per la condizione (i) sappiamo che $x \in \mathrm{RED}(C^\perp)$ e per la condizione (iv) $u = u[x/x] \in \mathrm{SN}_{Sym}$. Come nei casi precedenti, ogni riduzione su t è una riduzione su u e possiamo concludere.

- $t = u \star v$ con $C = \perp$: per definizione $\mathrm{RED}(\perp) \subseteq \mathrm{SN}_{Sym}$. □

Proposizione 3.10. *Sia A un m-tipo e $t \in \mathcal{T}_A$. Se per ogni $u \in \mathrm{RED}(A^\perp)$ vale*

$$t \star u \in \mathrm{SN}_{Sym} \text{ oppure } u \star t \in \mathrm{SN}_{Sym}$$

allora $t \in \mathrm{RED}(A)$.

Dimostrazione. Per simmetria dell'applicazione, basta provare che la conclusione vale quando $t \star u \in \mathrm{SN}_{Sym}$; l'argomento sarà simmetrico nell'altro caso. Procediamo per induzione sull'altezza di A distinguendo vari casi in base alla struttura di t:

- $t = x^A$: allora $t \in \mathrm{RED}(A)$ per la condizione (i).
- $t = \langle t_1, t_2 \rangle$ con $A = A_1 \wedge A_2$: vogliamo mostrare che $t_i \in \mathrm{RED}(A_i)$ per ogni $i \in \{1,2\}$. Per ipotesi induttiva, è sufficiente mostrare che per ogni $u \in \mathrm{RED}(A_i^\perp)$ vale $t_i \star u \in \mathrm{SN}_{Sym}$. Scegliamo allora $u \in \mathrm{RED}(A_i^\perp)$ qualsiasi: per la condizione (iii) avremo $\sigma_i(u) \in \mathrm{RED}(A_1^\perp \vee A_2^\perp)$ e per ipotesi $t \star \sigma_i(u) \in \mathrm{SN}_{Sym}$. Possiamo concludere osservando che $t \star \sigma_i(u) \rightsquigarrow_\pi t_i \star u$.
- $t = \sigma_i(u)$ con $A = A_1 \vee A_2$, $i \in \{1,2\}$: questo caso è analogo al precedente.
- $t = \lambda x.u$: per qualsiasi $v \in \mathrm{RED}(A^\perp)$ abbiamo che $t \star v \rightsquigarrow_\beta u[v/x]$ e $v \star t \rightsquigarrow_{\beta^\perp} u[v/x]$. Per ipotesi possiamo affermare in entrambi i casi che $u[v/x] \in \mathrm{SN}_{Sym}$, da cui segue la conclusione per la condizione (iv).
- $t = u \star v$: questo caso non può verificarsi perché $\perp$ non è un m-tipo (dunque $A \neq \perp$). □

Proposizione 3.11. *Se* $t \in \mathrm{RED}(C)$ *e* $t \rightsquigarrow_{Sym} t'$*, allora* $t' \in \mathrm{RED}(C)$.

Dimostrazione. Per induzione su C, considerando ancora una volta le diverse forme di t:

- $t = x^C$: questo caso non può verificarsi perché x è normale.
- $t = \langle u, v \rangle$ con $C = A \wedge B$: in tal caso $u \rightsquigarrow_{Sym} u'$ e $t' = \langle u', v \rangle$ oppure $v \rightsquigarrow_{Sym} v'$ e $t' = \langle u, v' \rangle$. Sappiamo che $u \in \mathrm{RED}(A)$ e $v \in \mathrm{RED}(B)$ e per ipotesi induttiva $u' \in \mathrm{RED}(B)$ o $v' \in \mathrm{RED}(B)$. La condizione (ii) permette di concludere.
- $t = \sigma_i(u)$ con $C = A \vee B$: si procede come nel caso precedente.
- $t = \lambda x.u$: sappiamo per la condizione (iv) che per ogni $v \in \mathrm{RED}(C^\perp)$ vale $u[v/x] \in \mathrm{SN}_{Sym}$. Si danno due possibilità: $t \rightsquigarrow_{\eta/\eta^\perp} t'$, oppure $u \rightsquigarrow_{Sym} u'$ con $t' = \lambda x.u'$.

  Nel primo caso avremo che $u = w \star x$ oppure $u = x \star w$ e $t' = w$ per qualche w di tipo C: dal fatto che $u[v/x] \in \mathrm{SN}_{Sym}$ possiamo concludere che $w \star v \in \mathrm{SN}_{Sym}$ o $v \star w \in \mathrm{SN}_{Sym}$ e per la proposizione 3.10 avremo $t' = w \in \mathrm{RED}(C)$.

  Nel secondo caso avremo che $u[v/x] \rightsquigarrow_{Sym} u'[v/x]$ da cui possiamo concludere che $u'[v/x] \in \mathrm{SN}_{Sym}$ e per la condizione (iv) $t' = \lambda x.u' \in \mathrm{RED}(C)$.
- $t = u \star v$: sappiamo che $t \in \mathrm{RED}(\perp) = \mathrm{SN}_\perp$ e se $t \rightsquigarrow_{Sym} t'$ allora anche $t' \in \mathrm{SN}_\perp$. □

Proposizione 3.12. *Per qualsiasi m-tipo A, $t \in \mathrm{RED}(A)$, $u \in \mathrm{RED}(A^\perp)$ vale*

$$t \star u \in \mathrm{SN}_{Sym}.$$

Dimostrazione. Per la proposizione 3.9 e la definizione 3.5 abbiamo $\nu(t), \nu(u) \in \mathbb{N}$. Procediamo per induzione lessicografica sulla coppia $(h(A), \nu(t) + \nu(u))$. L'enunciato è equivalente alla seguente proposizione:

$$t \star u \rightsquigarrow_{Sym} s \implies s \in SN_{Sym}.$$

Consideriamo tre coppie di casi simmetrici:

1. $t \rightsquigarrow_{Sym} t'$ con $s = t' \star u$ oppure $u \rightsquigarrow_{Sym} u'$ con $s = t \star u'$: per la proposizione 3.11 $t' \in \mathrm{RED}(A)$ e $u' \in \mathrm{RED}(A^\perp)$. Avremo anche $\nu(t') < \nu(t)$ e $\nu(u') < \nu(u)$: in entrambi i casi la somma delle due misure è strettamente minore di $\nu(t) + \nu(u)$ e possiamo applicare l'ipotesi induttiva.

2. $t = \langle t_1, t_2 \rangle$, $u = \sigma_i(v)$, $s = t_i \star v$ per qualche $i \in \{1,2\}$: in tal caso $A = A_1 \wedge A_2$ e $A^\perp = A_1^\perp \vee A_2^\perp$; per le condizioni (ii) e (iii) possiamo affermare che $t_i \in \mathrm{RED}(A_i)$ e $v \in \mathrm{RED}(A_i^\perp)$ e applicare l'ipotesi induttiva a $s = t_i \star v$. Il caso simmetrico si ottiene per $t = \sigma_i(v)$, $u = \langle u_1, u_2 \rangle$ e $s = v \star u_i$.

3. $t = \lambda x.v$ con $s = v[u/x]$ oppure $u = \lambda x.v$ con $s = v[t/x]$. In entrambi i casi possiamo concludere osservando che $v[u/x] \in \mathrm{SN}_{Sym}$ o $v[t/x] \in \mathrm{SN}_{Sym}$ per la condizione (iv). □

3.3.2 Dimostrazione del teorema

Come nella tecnica di Tait e Girard [GLT89], l'enunciato del teorema segue da un lemma che mostra come ottenere termini riducibili partendo da termini qualsiasi.

Lemma 3.13. *Sia C un tipo e $t \in \mathcal{T}_C$ un termine di tipo C, con le variabili libere di t contenute nell'insieme $\{x_1^{A_1}, \dots, x_n^{A_n}\}$ per qualche $n \geq 0$. Siano inoltre $t_1 \in \mathrm{RED}(A_1), \dots, t_n \in \mathrm{RED}(A_n)$. Allora*

$$t[t_1/x_1, \dots, t_n/x_n] \in \mathrm{RED}(C).$$

Dimostrazione. Per induzione sull'altezza di t:

– $t = x_i$ per qualche $1 \leq i \leq n$: in tal caso $C = A_i$ e per ipotesi $t[t_1/x_1, \dots, t_n/x_n] = t_i \in \mathrm{RED}(A_i) = \mathrm{RED}(C)$.

– $t = \langle u, v \rangle$, $t = \sigma_{1/2}(u)$, $t = u \star v$: in tutti questi casi la conclusione segue per ipotesi induttiva su u, v e sfruttando rispettivamente le condizioni (ii) e (iii) e la proposizione 3.12.

- $t = \lambda x.u$: poiché possiamo modificare il nome delle variabili vincolate, siamo liberi di assumere che $x \notin \{x_1, \dots, x_n\}$. Vogliamo applicare la condizione (iv) mostrando che, per ogni $v \in \mathrm{RED}(C^\perp)$, vale $u[t_1/x_1, \dots, t_n/x_n, v/x] \in \mathrm{SN}_{Sym}$. Dato che u è di tipo $\perp$, per ipotesi induttiva avremo che $u[t_1/x_1, \dots, t_n/x_n, v/x] \in \mathrm{RED}(\perp) \subseteq \mathrm{SN}_{Sym}$. □

Dimostrazione del teorema 3.6. Sia $t \in \lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$ un termine qualsiasi di tipo C e sia $FV(t) = \{x_1^{A_1}, \dots, x_n^{A_n}\}$ l'insieme delle sue variabili libere (per qualche $n \geq 0$). Per la condizione (i) sappiamo che $x_i \in \mathrm{RED}(A_i)$ per ogni $1 \leq i \leq n$ e allora per il lemma 3.13 avremo $t = t[x_1/x_1, \dots, x_n/x_n] \in \mathrm{RED}(C) \subseteq \mathrm{SN}_{Sym}$. □

3.4 Analisi e confronto con LK

Vogliamo mettere in luce alcuni aspetti fondamentali della tecnica dimostrativa appena utilizzata per poi discutere la possibilità di applicarla alla riduzione in **LK**.

3.4.1 Condizioni di riducibilità

Abbiamo già sottolineato la necessità di definire gli insiemi di termini riducibili in coppie per tipi duali. Questo punto diventa essenziale nel momento in cui si voglia utilizzare la tecnica dei candidati di [GLT89] per estendere il teorema di normalizzazione a un calcolo con tipi del secondo ordine.

Com'è noto, la tecnica di Girard consiste nell'identificare alcune condizioni (*condizioni di riducibilità*) che caratterizzino in astratto un insieme di termini riducibili (detto *candidato di riducibilità*). Le condizioni devono fornire un'ipotesi induttiva sufficiente a completare la dimostrazione di una proposizione nello stile del lemma 3.13, in assenza di qualsiasi altra informazione sull'insieme di termini in questione.

Si procede quindi a definire una nozione di *riducibilità parametrica*: l'idea è che se $\mathcal{R}$ è un candidato per il tipo A, allora $\mathrm{RED}(\alpha)[\mathcal{R}/\alpha] = \mathcal{R}$ è un candidato per il tipo $\alpha[A/\alpha] = A$. La difficoltà è che in un sistema di tipi classico con negazione involutiva è inevitabile fare riferimento a un qualche candidato per il tipo $A^\perp$: quale valore assegnare all'espressione $\mathrm{RED}(\alpha^\perp)[\mathcal{R}/\alpha]$?

Girard affronta un problema simile per la logica lineare in [Gir87]. La sua soluzione consiste in una definizione di ortogonalità fra candidati che sia abbastanza generale da permettere di scrivere $RED(\alpha^\perp)[\mathcal{R}/\alpha] = \mathcal{R}^\perp$. Nel nostro caso però, come abbiamo già osservato, la nozione di ortogonalità per il tipo A fa riferimento alla riducibilità $\mathrm{RED}(B)[\mathcal{R}/\alpha]$ per tipi B di altezza strettamente minore di A, nozione che a sua volta conterrà un riferimento al candidato $\mathcal{R}$ e al suo ortogonale $\mathcal{R}^\perp$: avremmo dunque una circolarità.

La nostra soluzione consiste nel chiedere che i candidati *siano sempre dati in coppie* per tipi duali: non avremo semplicemente un insieme $\mathcal{R}$, ma una coppia $\mathcal{R}, \mathcal{R}^\neg$ di *candidati simmetrici per* $A, A^\perp$ e potremo porre $\mathrm{RED}(\alpha^\perp)[\mathcal{R}/\alpha] = \mathcal{R}^\neg$.

Riteniamo dunque che la presenza di coppie di candidati sia un elemento essenziale della tecnica e non un mero accidente legato alle difficoltà nella definizione. Volendo, si potrebbe anche dire che la tecnica di [Gir87] contenga lo stesso approccio, che rimane però implicito grazie alle buone proprietà della definizione di ortogonalità.

Un semplice confronto fra la dimostrazione di [GLT89] e quella di [BB96] appena presentata suggerisce che le *condizioni di riducibilità* si possano riconoscere nelle proposizioni 3.9, 3.10, 3.11 e 3.12: definiremo allora una coppia di candidati simmetrici come una coppia di insiemi che soddisfano queste quattro proprietà. Vedremo nel prossimo capitolo che in realtà la definizione può essere semplificata: le sole proprietà 3.10 e 3.12 implicano le altre due.

3.4.2 Riducibilità per le astrazioni

Un secondo aspetto fondamentale della tecnica va identificato nella definizione di riducibilità per termini che hanno la forma di astrazioni:

$$\lambda x.t \in \mathrm{RED}(A) \iff \forall u \in \mathrm{RED}(A^{\perp}).\, t[u/x] \in \mathrm{SN}_{Sym}.$$

La definizione di riducibilità utilizzata in [GLT89] per tipi funzionali è invece:

$$t \in \mathrm{RED}(A \to B) \iff \forall u \in \mathrm{RED}(A).\, (tu) \in \mathrm{RED}(B).$$

In particolare, se $t = \lambda x.t'$ si chiede che valga $(\lambda x.t')u \in \mathrm{RED}(B)$. Una prima differenza evidente è l'assenza di tipi per funzioni in $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$: l'operatore di astrazione esprime una negazione. Vogliamo concentrarci però sulla definizione. Se traducessimo la seconda in $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$ otterremmo la proposizione seguente:

$$\lambda x.t \in \mathrm{RED}(A) \iff \forall u \in \mathrm{RED}(A^{\perp}).\, (\lambda x.t) \star u \in \mathrm{RED}(\perp).$$

Ricordando che $\mathrm{RED}(\perp) = \mathrm{SN}_{\perp}$, supponiamo di voler verificare che un termine $\lambda x.t$ di tipo A appartenga a $\mathrm{RED}(A)$ (come accade nella dimostrazione del lemma 3.13): dovremo verificare che per ogni $u \in \mathrm{RED}(A^{\perp})$ valga $(\lambda x.t) \star u \in \mathrm{SN}_{Sym}$. Questa però, a causa della simmetria dell'applicazione, è una proposizione troppo forte. In generale anche u potrà avere la forma di un'astrazione $\lambda y.v$: se pure sapessimo (ad esempio per ipotesi induttiva) che $t[u/x] \in \mathrm{SN}_{Sym}$, dovremmo dimostrare anche che $v[t/x] \in \mathrm{SN}_{Sym}$ e non avremmo informazioni sufficienti per concludere, senza parlare del fatto che questo andrebbe mostrato per ogni ridotto di t.

Pertanto, la definizione di [BB96] ha un ruolo cruciale nel gestire il non determinismo della riduzione in $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$ e naturalmente vorremmo adattare questa strategia ai tagli strutturali non deterministici di **LK**. A questo proposito, dobbiamo mettere in luce un altro punto critico del ragionamento di [BB96]. Nella dimostrazione della proposizione 3.11, secondo caso del quarto punto, si sfrutta il fatto che

$$\textit{se } t[u/x] \in \mathrm{SN}_{Sym} \textit{ e } t \rightsquigarrow_{Sym} t', \textit{ allora } t'[u/x] \in \mathrm{SN}_{Sym}.$$

Questo passaggio rimane in sordina nella presentazione originale dell'argomento, perché si appoggia su una proprietà della riduzione che viene data per scontata:

$$t \rightsquigarrow_{Sym} t' \implies t[u/x] \rightsquigarrow_{Sym} t'[u/x].$$

Tuttavia, osserviamo che $t[u/x]$ può essere visto come il risultato di un passo di β-riduzione applicato al termine $(\lambda x.t) \star u$: la proposizione enunciata sopra rappresenta dunque una proprietà "parziale" di confluenza:

$$\begin{tikzcd}
(\lambda x.t) \star u \arrow[r, "Sym"] \arrow[d, "\beta"] & (\lambda x.t') \star u \arrow[d, "\beta"] \\
t[u/x] \arrow[r, "Sym"] & t'[u/x]
\end{tikzcd}$$

vedremo nel prossimo capitolo che la dimostrazione della stessa proprietà in **LK** è molto delicata.

3.4.3 Riducibilità per derivazioni in LK

In [BBS97] gli autori osservano che le differenze fra $\lambda^{\text{Sym}}_{\text{Prop}}$ e **LK** sono troppo forti per tentare un confronto preciso; nonostante questo si ha anche l'intuizione che le somiglianze siano altrettanto forti e che meritino un approfondimento.

Un tentativo simile si trova in [Lai01], che propone un'immersione del frammento additivo di **LK** in $\lambda^{\text{Sym}}_{\text{Prop}}$. La traduzione è costruita in maniera indiretta, utilizzando sequenti monolaterali *sinistri* (sequenti di forma $\Gamma \vdash$), e non è una biiezione: esistono termini e specialmente forme normali in $\lambda^{\text{Sym}}_{\text{Prop}}$ che non sono (nel senso di [Lai01]) traduzioni di derivazioni nel frammento di **LK** scelto. In particolare, ogni traduzione avrà la forma $t \star u$ e ogni derivazione in **LK** corrisponde a una derivazione in $\lambda^{\text{Sym}}_{\text{Prop}}$ con conclusione $\perp$; il vantaggio immediato di questo approccio è che la riduzione del λ-calcolo simmetrico simula quella in **LK** e permette di affermare che quest'ultima (per il frammento additivo) è fortemente normalizzante.

Si può fare di meglio? Ricordiamo (seguendo [Gir91] e [DJS97]) che i sistemi di deduzione naturale possono essere messi in relazione con il calcolo dei sequenti con *stoup*, cioè con una formula in evidenza. È possibile associare ogni termine di $\lambda^{\text{Sym}}_{\text{Prop}}$ a una derivazione nel frammento additivo di **LK** in cui i sequenti hanno la forma $\vdash \Gamma; A$ (dove A può essere assente) e si impongono tre restrizioni:

(i) solamente la formula nello *stoup* può essere attiva;

(ii) la conclusione principale è sempre nello *stoup*;

(iii) non è possibile applicare regole strutturali alle formule nello stoup.

Le formule fuori dallo *stoup* denotano le variabili libere, quella in evidenza invece denota la conclusione della derivazione in $\lambda^{\text{Sym}}_{\text{Prop}}$. La regola di taglio assume la forma seguente:

$$\frac{\begin{array}{c}\vdots\\ \vdash \Gamma; t : A\end{array} \qquad \begin{array}{c}\vdots\\ \vdash \Delta; u : A^{\perp}\end{array}}{\vdash \Gamma, \Delta; t \star u}\ (cut)$$

Possiamo dunque dare un senso preciso all'affermazione fatta in precedenza che il tipo $\perp$ non sia una vera e propria formula, ma rappresenti l'assenza di una conclusione in evidenza: è per l'appunto il risultato di un taglio.

Come è possibile riportare una conclusione in evidenza? Una possibilità è quella di aggiungere una regola *ad hoc* che rappresenti la λ-astrazione:

$$\frac{\begin{array}{c}\vdots\\ \vdash \Gamma, x^{\perp} : A; t \star u\end{array}}{\vdash \Gamma; \lambda x.t \star u : A}\ (\lambda)$$

dove la notazione $x^{\perp}$ indica il fatto che la variabile x è di tipo $A^{\perp}$ (in generale un'occorrenza di A fuori dallo *stoup* corrisponde a una variabile libera di tipo $A^{\perp}$). La β-riduzione di $\lambda^{\text{Sym}}_{\text{Prop}}$ corrisponderà allora con precisione al passo $\rightsquigarrow_S$ di **LK**. È possibile continuare su questa strada e ottenere una corrispondenza completa; questo approccio però non è privo di difetti (vedremo nella sezione 5.2 che l'assioma di **LK** ha un comportamento diverso dalla variabile di $\lambda^{\text{Sym}}_{\text{Prop}}$) ed è necessario aggiungere alcune restrizioni *ad hoc* per ottenere una traduzione biiettiva *e* la simulazione reciproca dei processi di riduzione.

Ci interessa invece la possibilità di tradurre la tecnica dei candidati simmetrici nel frammento completo di **LK** introdotto nel capitolo precedente; ci chiediamo in particolare come formulare il predicato $\text{RED}(A)$ di riducibilità per il tipo A. Seguendo sia l'intuizione dello *stoup*, sia l'approccio di Girard in [Gir87], l'idea è di considerare derivazioni in **LK** con una occorrenza di formula in evidenza nella conclusione.

Quando l'occorrenza è principale in una regola logica, tradurre le condizioni (i)–(v) sarà banale. Una derivazione con la forma seguente:

$$\frac{\begin{array}{c}\vdots\ \pi_1\\ \vdash \Gamma, A^{x}\end{array} \qquad \begin{array}{c}\vdots\ \pi_2\\ \vdash \Gamma, B^{y}\end{array}}{\vdash \Gamma, A \wedge_a B}\ (\wedge_a)$$

apparterrà all'insieme $\text{RED}(A \wedge_a B)$ se $\pi_1 \in \text{RED}(A)$ e $\pi_2 \in \text{RED}(B)$ rispetto alle occorrenze in evidenza A^x, B^y. Allo stesso modo possiamo formulare le condizioni per le regole $(\vee_a)$ e $(\wedge_m)$.

Più complesso è il caso della disgiunzione moltiplicativa $(\vee_m)$, che non ha corrispondenti immediati in $\lambda^{\text{Sym}}_{\text{Prop}}$. L'idea è di trovare delle condizioni che siano abbastanza deboli ma sufficienti a provare che un taglio moltiplicativo è fortemente normalizzante; in particolare vogliamo mostrare che la derivazione seguente è fortemente normalizzante:

$$\dfrac{\dfrac{\begin{matrix}\vdots\,\pi\\ \vdash \Gamma, A, B\end{matrix} \qquad \begin{matrix}\vdots\,\pi_1\\ \vdash \Delta_1, \neg A\end{matrix}}{\vdash \Gamma, \Delta_1, B}\,(cut) \qquad \begin{matrix}\vdots\,\pi_2\\ \vdash \Delta_2, \neg B\end{matrix}}{\vdash \Gamma, \Delta_1, \Delta_2}\,(cut)$$

sapendo che π soddisfa le condizioni che stiamo cercando, $\pi_1 \in \mathrm{RED}(\neg A)$ e $\pi_2 \in \mathrm{RED}(\neg B)$. Sappiamo inoltre che un taglio fra due derivazioni è fortemente normalizzante se appartengono a una coppia di candidati simmetrici per tipi duali; basterà chiedere allora che il taglio superiore sia in $\mathrm{RED}(B)$, e cioè che:

$$\forall \pi_1 \in \mathrm{RED}(\neg A).\, cut(\pi, \pi_1) \in \mathrm{RED}(B);$$

per coprire l'altro caso relativo al dilemma logico (l'altra riduzione possibile per un taglio logico) dovremo chiedere anche che

$$\forall \pi_2 \in \mathrm{RED}(\neg B).\, cut(\pi, \pi_2) \in \mathrm{RED}(A).$$

Cosa dire infine di una derivazione in cui la conclusione in evidenza non sia principale in una regola logica? L'idea è che potremo trattarla come una conclusione principale applicando implicitamente una regola (λ). Per una derivazione $\pi \vdash \Gamma, A^x$ con A^x non principale diremo dunque che

$$\pi \in \mathrm{RED}(A) \iff \forall \delta \in \mathrm{RED}(\neg A).\, \pi[\delta / \vdash A^x] \in \mathrm{SN},$$

dove l'espressione $\pi[\delta / \vdash A^x]$ dovrà denotare la riduzione di un passo strutturale in cui A^x si comporta attrattivamente.

Capitolo 4

Normalizzazione forte per LK

Come anticipato nel capitolo precedente, il lavoro da svolgere si articola in quattro parti. Dobbiamo innanzitutto definire con precisione la nozione di 'sostituzione' per derivazioni in **LK** e mostrare che gode delle proprietà necessarie. In secondo luogo, sfruttando l'analisi svolta nel paragrafo 3.4.3, dobbiamo definire i candidati simmetrici e il predicato di riducibilità $\mathrm{RED}(A)$. Procederemo quindi alla dimostrazione del lemma 4.24 sulla riducibilità di derivazioni qualsiasi, da cui seguirà l'enunciato del teorema 4.23 di normalizzazione forte per $\mathbf{LK^0}$. Infine, seguendo [GLT89] e [Gir87], definiremo la nozione di riducibilità parametrica ed estenderemo la dimostrazione al frammento proposizionale del secondo ordine $\mathbf{LK^2}$ (teorema 4.33).

Richiamando le definizioni del capitolo 2, ricordiamo che la notazione

$$\pi \vdash \Gamma$$

indica che π è una derivazione la cui conclusione è il sequente $\vdash \Gamma$; ricordiamo inoltre il significato della notazione $\pi \rightsquigarrow_{[c]} \pi'$: π si trasforma in π' riducendo il taglio c. Per semplificare il testo useremo spesso una notazione alternativa per l'ultima regola delle derivazioni:

$$R(\pi_1, \dots, \pi_n) = \frac{\pi_1 \vdash \Gamma_1 \quad \cdots \quad \pi_n \vdash \Gamma_n}{\vdash \Gamma} \ (R)$$

dove (R) è una regola qualsiasi; ad esempio, $cut(\pi_1, \pi_2)$ denota la derivazione ottenuta tagliando π_1 con π_2 (le formule attive saranno chiare dal contesto). Come caso particolare,

$$ax(A) = \frac{}{\vdash \neg A, A} \ (ax).$$

4.1 Sostituzione per derivazioni in LK

Vogliamo definire una nozione di 'sostituzione' per **LK** analoga a quella del λ-calcolo. Il candidato naturale per rappresentare questa operazione è il passo di riduzione strutturale: abbiamo visto che, traducendo **LK** in $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$, il taglio strutturale corrisponde ad un'applicazione e il passo S del protocollo tq corrisponde a un passo di β-riduzione, cioè per l'appunto una sostituzione.

Definizione 4.1 (Sostituzione, prima definizione). Sia $\pi_1 \vdash \Gamma, A^x$ e $\pi_2 \vdash \Delta, \neg A^y$ e poniamo $\pi = cut(\pi_1, \pi_2)$ con $A^x, \neg A^y$ formule di taglio. L'espressione $\pi_1[\pi_2/\vdash A^x]$ denota:

- π_2, se $\pi_1 = ax(A)$;
- π, se A^x è principale;
- altrimenti denota la derivazione ottenuta da π riducendo il taglio finale con un passo $\rightsquigarrow_S$ che trasporta π_2.

L'idea è che l'espressione di sostituzione denoti il risultato della "risalita" di π_2 attraverso π_1, cioè la derivazione ottenuta tagliando o sostituendo con π_2 ogni foglia dell'albero strutturale di A^x in π_1:

$$\dfrac{\begin{array}{c}\vdots\ \pi_1 \\ \vdash \Gamma, A^x\end{array} \qquad \begin{array}{c}\vdots\ \pi_2 \\ \vdash \Delta, \neg A^y\end{array}}{\vdash \Gamma, \Delta}\ (cut)$$

Come abbiamo osservato nel capitolo precedente, per utilizzare la tecnica di [BB96] è fondamentale dimostrare la seguente proposizione:

Proposizione. *Sia $\pi_1 \vdash \Gamma, A^x$, $\pi_1 \rightsquigarrow_1 \pi_1'$ e $\pi_2 \vdash \Delta, \neg A^y$. Allora*

$$\pi_1[\pi_2/\vdash A^x] \in \mathrm{SN} \implies \pi_1'[\pi_2/\vdash A^x] \in \mathrm{SN}.$$

Abbiamo già spiegato che, nel λ-calcolo simmetrico, la proposizione appena enunciata è una banale conseguenza del fatto che, per t, t', u termini qualsiasi,

$$t \rightsquigarrow t' \implies t[u/x] \rightsquigarrow t'[u/x].$$

Purtroppo la nozione di sostituzione appena definita non permette di dimostrare questo fatto. L'assioma di **LK**, come accennato prima, si comporta diversamente dalla variabile di $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$: è possibile che sia coinvolto in due tagli e, se sono entrambi strutturali, ridurli in ordine diverso porta a ridotti molto diversi; riportiamo un'analisi dettagliata del problema nella sezione 5.2 del prossimo capitolo, in cui giustifichiamo la soluzione che stiamo per presentare: definiamo una versione modificata del passo S, che simula parzialmente quanto avviene in $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$ e gode delle proprietà di cui abbiamo bisogno.

Definizione 4.2 (Passo Sd). Siano $\pi_1 \vdash \Gamma, A^x$, $\pi_2 \vdash \Delta, \neg A^y$ con almeno una delle due formule in evidenza non principale. Il taglio strutturale fra π_1 e π_2

$$\dfrac{\pi_1 \vdash \Gamma, A^x \qquad \pi_2 \vdash \Delta, \neg A^y}{\vdash \Gamma, \Delta}\ (cut)$$

si riduce con un passo Sd (dove d sta per *delayed*) in base al suo tipo:

(S2) si adotta il passo S consueto trasportando la derivazione in cui la formula di taglio è principale in una regola logica;

(S1) nessuna delle formule di taglio è principale in una regola logica, scegliamo la derivazione da trasportare (supponiamo che sia π_2) ed esaminiamo le foglie dell'albero strutturale di A^x in π_1:

- per ogni regola logica con conclusione A^x introduciamo un taglio residuo con π_2;
- ogni indebolimento con conclusione A^x è sostituito da indebolimenti sul contesto Δ;
- per ogni assioma s distinguiamo quattro possibilità esaminando la derivazione trasportata e le due conclusioni di s:
  a. se la formula di taglio $\neg A^y$ è conclusione principale di π_2, sostituiamo l'assioma s con π_2;[1]
  b. se la conclusione di *s non coinvolta nel taglio* (cioè $\neg A$) ha un *interspazio piatto*, sostituiamo l'assioma s con π_2;
  c. se la conclusione di *s coinvolta nel taglio*[2] ha un *interspazio piatto*, sostituiamo l'assioma s con π_2;
  d. altrimenti *introduciamo un taglio residuo fra π_2 e l'assioma s*.

Denotiamo con $\rightsquigarrow_d$ l'unione dei passi $\rightsquigarrow_L$ e $\rightsquigarrow_{Sd}$, con $\rightsquigarrow_d^*$ la chiusura transitiva e riflessiva di $\rightsquigarrow_d$, con $\nu_d(\pi)$ l'altezza dell'albero di riduzione di π in $\rightsquigarrow_d^*$, infine denotiamo con $\mathrm{SN}_d \subseteq \mathbf{LK}$ l'insieme tale che $\pi \in \mathrm{SN}_d$ sse $\nu_d(\pi) \in \mathbb{N}$. Le nozioni di *lift* e *residuo* possono essere definite come di consueto (definizione 2.44). Denotiamo con $\rightsquigarrow_{d[c]}$ l'applicazione di un passo $\rightsquigarrow_d$ al taglio c.

Osservazione 4.3. Possiamo enunciare i casi b e c della definizione 4.2 anche come un'unico caso: se *almeno una* delle due conclusioni dell'assioma ha un *interspazio piatto*, allora la derivazione trasportata rimpiazza l'assioma.

Osservazione 4.4. Il lemma di stabilità (2.47) come formulato nel capitolo 2 non si adatta immediatamente alla d-riduzione appena definita. Consideriamo un'occorrenza di assioma s le cui conclusioni abbiano entrambe un interspazio lucido. È possibile che un passo di riduzione Sd introduca un taglio con s: in tal caso l'assioma avrà un residuo nel ridotto e una delle due conclusioni avrà un interspazio piatto (quella attiva nel taglio), ma l'altra non avrà più un interspazio lucido! La via d'uscita più semplice consiste nel limitare l'enunciato del lemma

[1] Questo può accadere solamente quando π_2 è un assioma, altrimenti il taglio sarebbe di tipo $S2$.

[2] Sarà dunque la conclusione A ascendente di A^x. Il caso che ci interessa realmente è quello precedente; la precisazione è necessaria solo perché altrimenti non si potrebbe più eliminare direttamente un taglio con assioma: la conclusione *non attiva* nel taglio infatti non avrebbe *mai* un interspazio piatto. Osserviamo anche che questa è *l'unica* situazione in cui il caso è rilevante.

agli interspazi piatti. Sarebbe possibile recuperare un risultato sugli interspazi lucidi rafforzando ulteriormente le ipotesi, ma nella nostra situazione non è necessario.

Lemma 4.5 (Stabilità sotto d-riduzione). *Sia π una derivazione in* **LK***, π' tale che $\pi \rightsquigarrow_d \pi'$, e sia B_{A^x} l'interspazio di una occorrenza di formula A^x indebolita, principale in una regola logica, o principale in un assioma che ha dei residui in π'. Se B_{A^x} è piatto, allora tutti i residui di B_{A^x} in π' sono piatti.*

Osservazione 4.6. Il motivo per cui il passo S del protocollo tq sostituisce immediatamente gli assiomi invece di introdurre tagli residui è che, se si introducesse sempre un taglio residuo con assioma, si potrebbero avere cammini di riduzione ciclici (cfr. [DJS97, p. 764]):

$$\dfrac{\begin{array}{c}\vdots\;\gamma\\ \vdash \Gamma, A\end{array} \qquad \dfrac{\dfrac{}{\neg A^x, A^y}\,(ax) \qquad \begin{array}{c}\vdots\;\delta\\ \vdash \Delta, \neg A\end{array}}{\vdash \neg A^x, \Delta}\,(cut)}{\vdash \Gamma, \Delta}\,(cut)$$

Trasportando γ nella derivazione qui sopra, otterremmo un taglio fra γ ed assioma seguito da un taglio con δ. Trasportando poi δ, otterremmo nuovamente la derivazione di partenza.

È importante dunque assicurarsi che questa possibilità sia completamente esclusa. Questo si verifica facilmente per assiomi in cui l'interspazio di qualche conclusione è piatto, che saranno sempre sostituiti (in particolare il controesempio esibito sopra non darebbe difficoltà); gli altri invece potrebbero essere sostituiti in alcuni casi da un taglio con assioma. A questo punto l'interspazio della formula di taglio è piatto e avremo solamente due possibilità: la riduzione di un taglio che coinvolga come foglia qualche conclusione dell'assioma lo sostituirà; ogni altra riduzione, per il lemma 4.5, preserva l'interspazio.

Proposizione 4.7. $\mathrm{SN}_d \subseteq \mathrm{SN}$.

Dimostrazione. Sia c un taglio in π con $\pi \rightsquigarrow_{[c]} \pi_1$. Se c è di tipo L allora evidentemente $\pi \rightsquigarrow_{d[c]} \pi_1$. Altrimenti c è strutturale ed esiste π_2 tale che $\pi \rightsquigarrow_{d[c]} \pi_2$ (riducendo nella stessa direzione) e $\pi_2 \rightsquigarrow^*_{Sd} \pi_1$ riducendo gli eventuali tagli con assioma residui di c introdotti dal passo Sd.[3] Viceversa se $\pi \rightsquigarrow_{d[c]} \pi_2$ allora $\pi \rightsquigarrow_{[c]} \pi_2$ oppure non esiste alcun cammino di riduzione tale che $\pi \rightsquigarrow \pi_2$. Possiamo concludere che ogni cammino di riduzione in **LK** può essere simulato da un cammino di *d-riduzione* di lunghezza maggiore o uguale e vale dunque $\nu(\pi) \leq \nu_d(\pi)$, da cui segue l'enunciato. □

[3]Questa osservazione fornisce immediatamente anche una dimostrazione di normalizzazione debole per la riduzione modificata.

Grazie al risultato appena dimostrato, possiamo applicare la tecnica dei candidati simmetrici alla d-riduzione e ottenere un teorema di normalizzazione forte per la riduzione "classica". Modifichiamo allora la nozione di sostituzione utilizzando il passo *Sd* e verifichiamo che goda delle proprietà desiderate.

Definizione 4.8 (Sostituzione, seconda definizione). Sia $\pi_1 \vdash \Gamma, A^x$, $\pi_2 \vdash \Delta, \neg A^y$ e poniamo $\pi = cut(\pi_1, \pi_2)$ con $A^x, \neg A^y$ formule di taglio. L'espressione $\pi_1[\pi_2/\vdash A^x]$ denota:

- π_2, se $\pi_1 = ax(A)$;
- π, se A^x è principale;
- altrimenti denota la derivazione ottenuta da π riducendo il taglio finale *con un passo Sd* che trasporta π_2.

Proposizione 4.9. *Sia $\pi_1 \vdash \Gamma, A^x$, c un taglio in π_1 tale che $\pi_1 \rightsquigarrow_{d[c]} \pi_1'$ e $\pi_2 \vdash \Delta, \neg A^y$. Allora, utilizzando la definizione 4.8,*

$$\pi_1[\pi_2/\vdash A^x] \rightsquigarrow^*_{d[c']} \pi_1'[\pi_2/\vdash A^x]$$

dove c' è l'unico residuo di c in $\pi_1[\pi_2/\vdash A^x]$ e $\rightsquigarrow^*_{d[c']}$ denota la chiusura riflessiva e transitiva del passo che d-riduce c'.[4] *Equivalentemente, il diagramma seguente commuta:*

$$\begin{tikzcd}
\pi_1 \arrow[r, "d[c]"] \arrow[d, "{[\pi_2/\vdash A^x]}"] & \pi_1' \arrow[d, "{[\pi_2/\vdash A^x]}"] \\
\pi_1[\pi_2/\vdash A^x] \arrow[r, dashed, two heads, "{d[c']^*}"] & \pi_1'[\pi_2/\vdash A^x]
\end{tikzcd}$$

Dimostrazione. Osserviamo che π_1 non può essere normale, e in particolare non può essere un assioma: in tal caso l'espressione di sostituzione denota la derivazione seguente o un suo ridotto:

$$\dfrac{\begin{array}{c}\vdots\ \pi_1 \\ \vdash \Gamma, A^x\end{array} \qquad \begin{array}{c}\vdots\ \pi_2 \\ \vdash \Delta, \neg A^y\end{array}}{\vdash \Gamma, \Delta}\ (cut) = k$$

dove indichiamo con k il taglio finale.

Se A^x è conclusione principale di π_1, allora $\pi_1[\pi_2/\vdash A^x]$ denota la derivazione appena esibita. Osservando che A^x ha un interspazio piatto e che questo è preservato sotto d-riduzione (lemma 4.5), concludiamo che $\pi_1'[\pi_2/\vdash A^x]$ denota la derivazione $cut(\pi_1', \pi_2)$, $c' = c$ ed evidentemente $cut(\pi_1, \pi_2) \rightsquigarrow_{d[c']} cut(\pi_1', \pi_2)$.

[4] In altre parole, non è sempre necessario ridurre c'.

Altrimenti, ricordando che k denota il taglio fra π_1 e π_2 rappresentato sopra, l'enunciato della proposizione equivale ad affermare che il seguente diagramma commuta:

$$\begin{tikzcd}
cut(\pi_1,\pi_2) \arrow[r, "d[c]"] \arrow[d, "Sd[k]"] & cut(\pi_1',\pi_2) \arrow[d, "Sd[k]"] \\
\pi_1[\pi_2/\vdash A^x] \arrow[r, dashed, two heads, "d[c']^*"] & \pi_1'[\pi_2/\vdash A^x]
\end{tikzcd}$$

consideriamo cioè l'operazione di sostituzione come un passo Sd che riduce k trasportando π_2.

Mostriamo innanzitutto che esiste un unico residuo c' di c in $\pi_1[\pi_2/\vdash A^x]$ (in altre parole, ricordando la definizione 2.44, esiste un unico taglio c' in $\pi_1[\pi_2/\vdash A^x]$ di cui $c = c'_*$ è il *lift* in π_1): basta osservare che c non appartiene alla sottoderivazione trasportata dal passo $Sd[k]$ (che è π_2) ed essendo un taglio non può in alcun modo appartenere all'albero strutturale di A^x in π_1. Pertanto, ha un residuo c' in $\pi_1[\pi_2/\vdash A^x]$ e tale residuo è unico.

Se c è l'ultima regola di π_1 e riducendolo A^x diventa conclusione principale di π_1', allora $\pi_1'[\pi_2/\vdash A^x] = cut(\pi_1',\pi_2)$ e dobbiamo distinguere due casi:

a. c è un taglio con assioma, A^x discende dalla conclusione dell'assioma e A^z è conclusione principale di π_1':

$$\dfrac{\dfrac{\begin{array}{c}\vdots\ \pi_1' \\ \vdash \Gamma, A^z\end{array} \quad \nearrow d[c] \qquad \dfrac{}{\vdash \neg A, A^x}\,(ax)}{\vdash \Gamma, A^x}\,c \qquad Sd[k] \nwarrow \begin{array}{c}\vdots\ \pi_2 \\ \vdash \Delta, \neg A^y\end{array}}{\vdash \Gamma, \Delta}\,k$$

In tal caso operando la sostituzione, cioè applicando il passo $Sd[k]$ che trasporta π_2, l'assioma viene rimpiazzato da π_2 (perché la conclusione $\neg A$ ha interspazio piatto); abbiamo dunque $\pi_1[\pi_2/\vdash A^x] = cut(\pi_1',\pi_2)$ e cioè $\pi_1[\pi_2/\vdash A^x] = \pi_1'[\pi_2/\vdash A^x]$ e possiamo concludere senza ridurre c'.

b. c è un taglio strutturale e A^x è conclusione principale di una delle sottoderivazioni premessa di c. Supponiamo senza perdita di generalità che A^x sia conclusione principale di γ nella derivazione seguente:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\ \gamma \\ \vdash \Gamma_1, A^x, B\end{array} \qquad \begin{array}{c}\vdots\ \delta \\ \vdash \Gamma_2, \neg B\end{array}}{\vdash \Gamma, A^x}\,c \qquad Sd[k] \nwarrow \begin{array}{c}\vdots\ \pi_2 \\ \vdash \Delta, \neg A^y\end{array}}{\vdash \Gamma, \Delta}\,k$$

dove $\pi_1 = cut(\gamma, \delta)$. Riducendo k per primo otteniamo $\pi_1[\pi_2/\vdash A^x]$ che avrà la forma seguente:

$$\dfrac{\dfrac{\begin{matrix}\vdots\,\gamma\\ \vdash \Gamma_1, A^x, B\end{matrix} \qquad \begin{matrix}\vdots\,\pi_2\\ \vdash \Delta, \neg A^y\end{matrix}}{\vdash \Gamma_1, \Delta, B}\,k \qquad \begin{matrix}\vdots\,\delta\\ \vdash \Gamma_2, \neg B\end{matrix}}{\vdash \Gamma, \Delta}\,c'$$

Se δ è un assioma e il passo $d[c]$ trasporta γ, allora $\pi_1' = \gamma$, avremo $\pi_1'[\pi_2/\vdash A^x] = cut(\gamma, \pi_2)$ e la conclusione segue applicando il passo $d[c']$ che sostituirà δ con $cut(\gamma, \pi_2) = \pi_1'[\pi_2/\vdash A^x]$.

Altrimenti il passo $d[c]$ trasporta δ risultando in $\pi_1' = \gamma[\delta/\vdash B]$, avremo dunque $\pi_1'[\pi_2/\vdash A^x] = cut(\gamma[\delta/\vdash B], \pi_2)$ e potremo concludere applicando a $\pi_1[\pi_2/\vdash A^x]$ il passo $d[c']$ e trasportando δ.

Altrimenti consideriamo le trasformazioni dell'albero strutturale di A^x in π_1 durante la riduzione. Se c è un taglio logico dobbiamo distinguere due casi: un passo moltiplicativo non può modificare l'albero strutturale di A^x e riducendo c' in $\pi_1[\pi_2/\vdash A^x]$ otteniamo $\pi_1'[\pi_2/\vdash A^x]$; un passo additivo cancella una delle premesse della regola $(\wedge_a)$ e tutte le foglie dell'albero strutturale di A^x contenute in tale premessa: riducendo c' in $\pi_1[\pi_2/\vdash A^x]$ cancelliamo tutte le occorrenze di π_2 introdotte dal passo $Sd[k]$ nella sottoderivazione cancellata e otteniamo così $\pi_1'[\pi_2/\vdash A^x]$.

Se invece c è un taglio strutturale, una sottoderivazione δ di π_1 sarà trasportata e dobbiamo considerare quattro possibilità:

a. c è un taglio fra δ e un assioma che appartiene anche all'albero strutturale di A^x, la riduzione $d[c]$ sostituisce l'assioma con δ:

$$\dfrac{\begin{matrix}\dfrac{\begin{matrix}\vdots\,\delta\\ \vdash \Sigma, A^z\end{matrix} \;\nearrow d[c] \qquad \dfrac{}{\vdash A^x, \neg A}\,(ax)}{\vdash \Sigma, A^x}\,c\\ \vdots\\ \vdash \Gamma, A^x\end{matrix} \qquad Sd[k] \nwarrow\; \begin{matrix}\vdots\,\pi_2\\ \vdash \Delta, \neg A^y\end{matrix}}{\vdash \Gamma, \Delta}\,k$$

Osserviamo che almeno una conclusione dell'assioma ha un interspazio piatto (è utilizzata immediatamente nel taglio c con δ): dunque la riduzione $Sd[k]$ rimpiazza l'assioma con π_2, e $\pi_1[\pi_2/\vdash A^x]$ avrà la forma seguente:

$$\dfrac{\begin{array}{c}\vdots\ \delta\\ \vdash \Sigma, A^z\end{array} \quad \overset{d[c']^*}{\nwarrow} \quad \begin{array}{c}\vdots\ \pi_2\\ \vdash \Delta, \neg A^y\end{array}}{\begin{array}{c}\vdash \Sigma, \Delta\\ \vdots\\ \vdash \Gamma, \Delta\end{array}}\, c'$$

Se la formula di taglio A^z è conclusione principale di δ, allora $\pi_1'[\pi_2/\vdash A^x]$ avrà la stessa forma perché applicando il passo $Sd[k]$ a $cut(\pi_1', \pi_2)$ (colonna destra del diagramma) la riduzione introdurrà un taglio residuo fra δ e π_2: avremo dunque $\pi_1[\pi_2/\vdash A^x] = \pi_1'[\pi_2/\vdash A^x]$.

Altrimenti π_2 sarà trasportata dal passo $Sd[k]$ all'interno di δ e $\pi_1'[\pi_2/\vdash A^x]$ conterrà la sottoderivazione $\delta[\pi_2/\vdash A^z]$. Potremo allora applicare a $\pi_1[\pi_2/\vdash A^x]$ il passo $d[c']$, ma con attenzione: dovremo trasportare π_2 e non δ. Il passo $d[c']$ sul residuo dovrà cioè essere effettuato in direzione contraria a quella scelta nel passo $d[c]$.[5]

b. c è un taglio di tipo $S1$ con assioma, δ è l'assioma che viene trasportato durante la riduzione e δ appartiene all'albero strutturale di A^x in π_1:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\ \gamma\\ \vdash \Sigma, A\end{array} \quad \overset{d[c]}{\nwarrow} \quad \dfrac{}{\vdash A^x, \neg A}\,(ax) = \delta}{\begin{array}{c}\vdash \Sigma, A^x\\ \vdots\\ \vdash \Gamma, A^x\end{array}}\, c \qquad \overset{Sd[k]}{\nwarrow} \quad \begin{array}{c}\vdots\ \pi_2\\ \vdash \Delta, \neg A^y\end{array}}{\vdash \Gamma, \Delta}\, k$$

Anche in questo caso il passo $Sd[k]$ rimpiazza l'assioma δ con π_2 e avremo $\pi_1[\pi_2/\vdash A^x] =$

$$\dfrac{\begin{array}{c}\vdots\ \delta\\ \vdash \Sigma, A^z\end{array} \quad \overset{d[c']}{\nwarrow} \quad \begin{array}{c}\vdots\ \pi_2\\ \vdash \Delta, \neg A^y\end{array}}{\begin{array}{c}\vdash \Sigma, \Delta\\ \vdots\\ \vdash \Gamma, \Delta\end{array}}\, c'$$

A questo punto si verifica facilmente che $\pi_1[\pi_2/\vdash A^x] \rightsquigarrow_{d[c']} \pi_1'[\pi_2/\vdash A^x]$ riducendo c' e trasportando π_2. Se δ sostituisse in π_1' qualche assioma che non ha interspazi piatti, la sostituzione $\pi_1'[\pi_2/\vdash A^x]$ potrebbe contenere

[5] Si tratta di un caso critico della riduzione non deterministica in cui, cambiando l'ordine dei passi di riduzione, cambia anche l'insieme delle forme normali accessibili. In particolare, applicando prima il passo $d[c]$, si perderanno le forme normali in cui δ viene trasportata all'interno di π_2. Discutiamo la questione più ampiamente nella sezione 5.3 del prossimo capitolo.

tagli fra tali assiomi e π_2: ma in tal caso la riduzione di c' introdurrà tagli fra assioma e π_2 nelle stesse posizioni.

c. δ non è un assioma e contiene alcune foglie dell'albero strutturale di A^x in π_1. In particolare, δ avrà conclusione $\vdash \Sigma, A^{x_1}, \dots, A^{x_n}$ per qualche $n > 0$, dove le occorrenze $A^{x_1}, \dots, A^{x_n}$ sono ascendenti di A^x. Durante la riduzione $d[c]$, δ sarà spostata, duplicata ed eventualmente cancellata, introducendo dove necessario contrazioni e indebolimenti sulle occorrenze contestuali $A^{x_1}, \dots, A^{x_n}$, che faranno parte dell'albero strutturale di A^x in π_1'. Osserviamo facilmente che il passo $Sd[k]$ costruirà $\pi_1[\pi_2/\vdash A^x]$ rimpiazzando δ con $\delta' = \delta[\pi_2/\vdash A^{x_1}, \dots, \pi_2/\vdash A^{x_n}]$ di conclusione $\vdash \Sigma, \Delta^{x_1}, \dots, \Delta^{x_n}$: riducendo quindi c' trasporteremo δ' e introdurremo indebolimenti e contrazioni su $\Delta^{x_1}, \dots, \Delta^{x_n}$ proprio dove li avremmo introdotti su $A^{x_1}, \dots, A^{x_n}$.

d. δ sostituisce, o viene tagliata con, qualche assioma che è *una foglia dell'albero strutturale di A^x in π_1*. Si tratta di un caso molto delicato e precisamente quello che giustifica l'introduzione della d-riduzione. Supponiamo che s sia uno qualsiasi degli assiomi in questione: necessariamente nessuna delle conclusioni di s avrà un interspazio piatto.[6] Dobbiamo distinguere una serie di sotto-casi e verificare che il diagramma commuti:
   - la formula di taglio A è conclusione principale di δ e $\neg A^y$ è conclusione principale di π_2: riducendo c si rimpiazza s con δ, dopodiché riducendo k introduciamo un taglio fra π_2 e δ nella posizione occupata in precedenza da s; viceversa riducendo k per primo si rimpiazza s con π_2, quindi riducendo c' si introduce nella stessa posizione un taglio fra π_2 e δ.
   - A *non è* conclusione principale di δ mentre $\neg A^y$ è conclusione principale di π_2: riducendo c si ottiene un taglio residuo fra l'assioma s e δ, che riducendo k diventa un taglio fra π_2 e δ; viceversa riducendo k si rimpiazza s con π_2 e poiché $\neg A^y$ è principale, riducendo c' si ottiene un taglio residuo fra π_2 e δ.
   - A è conclusione principale di δ ma $\neg A^y$ *non è* conclusione principale di π_2: come nel caso precedente invertendo i ruoli.
   - Entrambe le conclusioni non principali: la prima ad essere trasportata fra δ e π_2 lascia un taglio residuo con l'assioma s, la seconda sostituisce s in quanto seguito immediatamente da un taglio.

Mentre i casi a e b escludono gli altri, i casi c e d non si escludono a vicenda; si tratta però di problemi indipendenti e la congiunzione dei due ragionamenti permette di concludere anche nel caso generale. Se nessuno di questi casi si verifica, le foglie dell'albero strutturale di A^x in π_1 rimangono inalterate in π_1' dalla riduzione di c e la conclusione è immediata riducendo c' alla stessa maniera. □

[6] Tutti i casi in cui s potrebbe avere qualche conclusione con interspazio piatto sono stati discussi e risolti in precedenza.

Corollario 4.10. *Sia* $\pi_1 \vdash \Gamma, A^x$, $\pi_1 \leadsto_d \pi_1'$ *e* $\pi_2 \vdash \Delta, \neg A^y$. *Allora*

$$\pi_1[\pi_2/\vdash A^x] \in \mathrm{SN}_d \implies \pi_1'[\pi_2/\vdash A^x] \in \mathrm{SN}_d.$$

Osserviamo che per π_1, π_2 qualsiasi, la proposizione $\pi_1[\pi_2/\vdash A^x] \in \mathrm{SN}_d$ intesa secondo la definizione 4.8 è (a volte strettamente) più forte della proposizione $\pi_1[\pi_2/\vdash A^x] \in \mathrm{SN}_d$ intesa secondo la definizione 4.1, per le stesse ragioni esposte nella dimostrazione della proposizione 4.7. Da ora in avanti dunque faremo sempre riferimento esclusivamente alla nozione di sostituzione della definizione 4.8 (che usa il passo $\leadsto_{Sd}$).

4.2 Candidati simmetrici e riducibilità

Nelle prossime pagine ci occuperemo di adattare la definizione dei candidati simmetrici e dei predicati di riducibilità alle derivazioni di $\mathbf{LK^0}$ (ricordiamo dunque che tutte le formule saranno proposizionali). Utilizziamo le stesse notazioni presentte nel capitolo precedente, riadattate al caso del calcolo dei sequenti.

Definizione 4.11. Una *derivazione di tipo* A è una derivazione $\pi \vdash \Gamma, A^x$ con una occorrenza di formula A^x in evidenza nella conclusione. Denotiamo con $\mathcal{D}_A$ l'insieme di tutte le derivazioni di tipo A.

A seguire ometteremo di quando in quando l'indice sull'occorrenza di formula in evidenza, qualora sia ovvia dal contesto. Commetteremo più volte l'abuso di applicare gli operatori logici di **LK** a insiemi di derivazioni. Ad esempio, sia X (rispettivamente Y) un insieme di derivazioni di tipo A (risp. B). Scriveremo $\wedge_m(X,Y)$ per denotare l'insieme seguente:

$$\wedge_m(X,Y) = \left\{ \frac{\pi_1 \vdash \Gamma, A^x \qquad \pi_2 \vdash \Delta, B^y}{\vdash \Gamma, \Delta, A \wedge_m B^z}\,(\wedge_m) \;:\, \pi_1 \in X, \pi_2 \in Y \right\}$$

e analogamente per gli altri operatori. Un caso particolare è rappresentato dall'operatore $(\wedge_a)$:

$$\wedge_a(X,Y) = \left\{ \frac{\pi_1 \vdash \Gamma, A^x \qquad \pi_2 \vdash \Delta, B^y}{\vdash \Gamma, A \wedge_a B^z}\,(\wedge_a) \;:\, \pi_1 \in X, \pi_2 \in Y, \Gamma = \Delta \right\}$$

Definizione 4.12 (Candidati simmetrici). Sia A una formula. Chiamiamo *candidati simmetrici per* $A, \neg A$ ogni *coppia* di insiemi $\mathcal{R} \subseteq \mathcal{D}_A$, $\mathcal{R}^\neg \subseteq \mathcal{D}_{\neg A}$ *non vuoti* e tali che:

(i) per qualsiasi $\pi \in \mathcal{D}_A$, se $\forall \delta \in \mathcal{R}^\neg.\ \pi[\delta/\vdash A^x] \in \mathrm{SN}_d$ allora $\pi \in \mathcal{R}$;

(ii) per qualsiasi $\pi \in \mathcal{D}_{\neg A}$, se $\forall \delta \in \mathcal{R}.\ \pi[\delta/\vdash \neg A^x] \in \mathrm{SN}_d$ allora $\pi \in \mathcal{R}^\neg$;

(iii) per ogni $\gamma \in \mathcal{R}$, $\delta \in \mathcal{R}^\neg$ vale $cut(\gamma, \delta) \in \mathrm{SN}_d$.

Osservazione 4.13. I punti (i) e (ii) della definizione esprimono una condizione sufficiente perché una derivazione π di tipo A appartenga a un candidato $\mathcal{R}$ per A: se, per ogni elemento δ del candidato "avversario" $\mathcal{R}^{\neg}$, l'espressione $\pi[\delta/\vdash A^x]$ denota una derivazione fortemente normalizzante, allora $\pi \in \mathcal{R}$; la stessa cosa vale, *mutatis mutandis*, per $\mathcal{R}^{\neg}$. In altre parole, se una derivazione π di tipo A *non appartiene* ad un candidato $\mathcal{R}$ per A, allora deve esistere in $\mathcal{R}^{\neg}$ almeno una derivazione δ di tipo $\neg A$ tale che $\pi[\delta/\vdash A^x]$ *non sia* fortemente normalizzante.

Osservazione 4.14. Dalla definizione 4.12 segue immediatamente $\mathcal{R}, \mathcal{R}^{\neg} \subseteq \mathrm{SN}_d$ (per la condizione (iii)) e di conseguenza $ax(A) \in \mathcal{R}$ e $ax(\neg A) \in \mathcal{R}^{\neg}$ (per le condizioni (i) e (ii)).

Proprietà 4.15. *Sia $\mathcal{R}, \mathcal{R}^{\neg}$ una coppia di candidati simmetrici per $A, \neg A$: per ogni $\gamma \in \mathcal{R}$ (risp. $\gamma \in \mathcal{R}^{\neg}$), se $\gamma \rightsquigarrow_d \gamma'$ allora $\gamma' \in \mathcal{R}$ (risp. $\gamma \in \mathcal{R}^{\neg}$).*

Dimostrazione. Per $\gamma \in \mathcal{R}$, scegliamo qualsiasi δ in $\mathcal{R}^{\neg}$: per la condizione (iii) vale $cut(\gamma, \delta) \in \mathrm{SN}_d$, da cui $cut(\gamma', \delta) \in \mathrm{SN}_d$ e di conseguenza $\gamma'[\delta/\vdash A^x] \in \mathrm{SN}_d$. La condizione (i) permette di concludere; l'argomento è analogo per $\mathcal{R}^{\neg}$ sfruttando la condizione (ii). □

Vogliamo ora definire, per induzione sul tipo A, gli insiemi di *derivazioni riducibili* di tipo A; fatto questo verificheremo che gli insiemi di derivazioni riducibili di tipo A e $\neg A$ costituiscano una coppia di candidati simmetrici per $A, \neg A$. In particolare, procederemo per induzione sulle formule *positive* e definiremo la coppia di insiemi $\mathrm{RED}(A), \mathrm{RED}(\neg A)$ ($\neg A$ sarà una formula negativa).

Definizione 4.16 (Riducibilità). Sia A una formula *positiva*, $P \subseteq \mathcal{D}_A$ e $N \subseteq \mathcal{D}_{\neg A}$ insiemi di derivazioni di tipo A (risp. di tipo $\neg A$). Definiamo innanzitutto gli insiemi

$$\mathrm{Subst}(P) = \{\pi : \pi \in \mathcal{D}_{\neg A} \wedge \forall \delta \in P.\ \pi[\delta/\vdash \neg A^x] \in \mathrm{SN}_d\},$$
$$\mathrm{Subst}(N) = \{\pi : \pi \in \mathcal{D}_{A} \wedge \forall \delta \in N.\ \pi[\delta/\vdash A^x] \in \mathrm{SN}_d\}.$$

Osserviamo che $\mathrm{Subst}(P) \subseteq \mathcal{D}_{\neg A}$ e $\mathrm{Subst}(N) \subseteq \mathcal{D}_A$: l'operatore Subst "nega" il tipo degli insiemi di derivazioni a cui viene applicato.

Definiamo quindi per induzione su A gli insiemi $P^{\perp} \subseteq \mathcal{D}_{\neg A}$, $N^{\perp} \subseteq \mathcal{D}_A$. Definiamo inoltre *contemporaneamente*, ad ogni passo induttivo, l'insieme $\mathrm{RED}(A) \subseteq \mathcal{D}_A$ come il più piccolo insieme tale che $\mathrm{RED}(A) = \mathrm{RED}(A)^{\perp\perp}$ e poniamo $\mathrm{RED}(\neg A) = \mathrm{RED}(A)^{\perp}$.

– $A = X, \neg A = \neg X$ formule atomiche:

$$P^{\perp} = \{ax(\neg A)\} \cup \mathrm{Subst}(P),$$
$$N^{\perp} = \{ax(A)\} \cup \mathrm{Subst}(N).$$

– $A = B \wedge_m C$, $\neg A = \neg B \vee_m \neg C$:

$$P^\perp = \{ax(\neg A)\} \cup \vee_m(\mathrm{RED}(\neg B, \neg C)) \cup \mathrm{Subst}(P),$$
$$N^\perp = \{ax(A)\} \cup \wedge_m(\mathrm{RED}(B), \mathrm{RED}(C)) \cup \mathrm{Subst}(N)$$

dove $\mathrm{RED}(\neg B, \neg C)$ è un insieme di derivazioni $\pi \vdash \Gamma, \neg B^x, \neg C^y$ (con Γ qualsiasi) definito come segue:

$$\mathrm{RED}(\neg B, \neg C) = \left\{ \pi : \begin{array}{l} \forall \gamma \in \mathrm{RED}(B).\ cut(\pi, \gamma) \in \mathrm{RED}(\neg C) \\ \forall \delta \in \mathrm{RED}(C).\ cut(\pi, \delta) \in \mathrm{RED}(\neg B) \end{array} \right\}$$

dove i tagli si intendono effettuati sulle conclusioni in evidenza $\neg B^x$ e $\neg C^y$.

– $A = B \vee_a C$, $\neg A = \neg B \wedge_a \neg C$:

$$P^\perp = \{ax(\neg A)\} \cup \wedge_a(\mathrm{RED}(\neg B), \mathrm{RED}(\neg C)) \cup \mathrm{Subst}(P),$$
$$N^\perp = \{ax(A)\} \cup \vee_a^1(\mathrm{RED}(B)) \cup \vee_a^2(\mathrm{RED}(C)) \cup \mathrm{Subst}(N).$$

Osservazione 4.17. La definizione 4.16 è ben posta. Per ogni formula A consideriamo gli insiemi $\mathcal{P}(\mathcal{D}_A)$, $\mathcal{P}(\mathcal{D}_{\neg A})$ ordinati per inclusione: è facile verificare che l'operatore $\mathrm{Subst} : \mathcal{P}(\mathcal{D}_A) \to \mathcal{P}(\mathcal{D}_{\neg A})$ è decrescente, cioè vale

$$X \subseteq Y \implies \mathrm{Subst}(Y) \subseteq \mathrm{Subst}(X).$$

Di conseguenza anche l'ortogonale $(\cdot)^\perp : \mathcal{P}(\mathcal{D}_A) \to \mathcal{P}(\mathcal{D}_{\neg A})$ è decrescente e il biortogonale $(\cdot)^{\perp\perp} : \mathcal{P}(\mathcal{D}_A) \to \mathcal{P}(\mathcal{D}_A)$ sarà un operatore crescente (vale cioè $X^{\perp\perp} \subseteq Y^{\perp\perp}$), che per il teorema del punto fisso di Tarski possiede un minimo punto fisso.

Osservazione 4.18. Le condizioni (i) e (ii) della definizione 4.12 possono essere espresse anche chiedendo che valga $\mathrm{Subst}(\mathcal{R}^\neg) \subseteq \mathcal{R}$ e $\mathrm{Subst}(\mathcal{R}) \subseteq \mathcal{R}^\neg$. Si vede chiaramente allora che la coppia di insiemi $\mathrm{RED}(A)$, $\mathrm{RED}(\neg A)$ soddisfa le suddette condizioni.

4.2.1 Adeguatezza della definizione di riducibilità

Proposizione 4.19. $\mathrm{RED}(A) \subseteq \mathrm{SN}_d$.

Dimostrazione. È facile verificare che $\mathrm{RED}(A) = \mathrm{RED}(\neg A)^\perp$: per A negativa è immediato per definizione di $\mathrm{RED}(A)$, per A positiva basta osservare che $\mathrm{RED}(A) = \mathrm{RED}(A)^{\perp\perp} = \mathrm{RED}(\neg A)^\perp$. Dimostriamo la conclusione per induzione su A e considerando le varie forme di $\gamma \in \mathrm{RED}(A)$:

– $\gamma = ax(A)$: banale.

– $A = B \wedge_m C$, $\gamma = \wedge_m(\gamma_1, \gamma_2)$ con $\gamma_1 \in \mathrm{RED}(B)$ e $\gamma_2 \in \mathrm{RED}(C)$: ogni passo di riduzione su γ è un passo di riduzione su γ_1 o γ_2 e per ipotesi induttiva $\mathrm{RED}(B), \mathrm{RED}(C) \subseteq \mathrm{SN}_d$.

- $A = B \vee_m C$, $\gamma = \vee_m(\gamma_1)$ con $\gamma_1 \in \mathrm{RED}(B,C)$: per ipotesi induttiva $\mathrm{RED}(C) \subseteq \mathrm{SN}_d$; abbiamo inoltre per definizione $\gamma_1 \in \mathrm{RED}(B,C)$ e sappiamo che $ax(\neg B) \in \mathrm{RED}(\neg B)$. Il fatto che $cut(\gamma_1, ax(\neg B)) \in \mathrm{RED}(C)$ permette di concludere $\gamma_1 \in \mathrm{SN}_d$.

- $A = B_1 \vee_a B_2$, $\gamma = \vee_a^i(\gamma_1)$ con $\gamma_1 \in \mathrm{RED}(B_i)$, $i \in \{1,2\}$: applichiamo l'ipotesi induttiva a $\mathrm{RED}(B_i)$.

- $A = B \wedge_a C$, $\gamma = \wedge_a(\gamma_1, \gamma_2)$ con $\gamma_1 \in \mathrm{RED}(B)$ e $\gamma_2 \in \mathrm{RED}(C)$: applichiamo l'ipotesi induttiva a $\mathrm{RED}(B)$ e $\mathrm{RED}(C)$.

- $\gamma \in \mathrm{Subst}(\mathrm{RED}(\neg A))$: ricordando che $ax(\neg A) \in \mathrm{RED}(\neg A)$, per definizione $\gamma[ax(\neg A)/\vdash A^x] \in \mathrm{SN}_d$ e vale $\gamma[ax(\neg A)/\vdash A^x] \rightsquigarrow_{Sd}^* \gamma$ (riducendo gli eventuali tagli residui con assioma). □

Proposizione 4.20. *Sia* $\gamma \in \mathrm{RED}(A)$, $\gamma \rightsquigarrow_d \gamma'$. *Allora* $\gamma' \in \mathrm{RED}(A)$.

Dimostrazione. Per induzione su A considerando le forme di γ (possiamo ovviamente assumere che γ non sia normale):

- $A = B \wedge_m C$, $\gamma = \wedge_m(\gamma_1, \gamma_2)$ con $\gamma_1 \in \mathrm{RED}(B)$ e $\gamma_2 \in \mathrm{RED}(C)$: in tal caso $\gamma_1 \rightsquigarrow_d \gamma_1'$ (oppure risp. $\gamma_2 \rightsquigarrow_d \gamma_2'$) e $\gamma' = \wedge_m(\gamma_1', \gamma_2)$ (risp. $\gamma' = \wedge_m(\gamma_1, \gamma_2')$). Per ipotesi induttiva $\gamma_1' \in \mathrm{RED}(B)$ (risp. $\gamma_2' \in \mathrm{RED}(C)$). In entrambi i casi per definizione $\gamma' \in \mathrm{RED}(A)$.

- $A = B \vee_m C$, $\gamma = \vee_m(\gamma_1)$ con $\gamma_1 \in \mathrm{RED}(B,C)$: avremo che $\gamma_1 \rightsquigarrow_d \gamma_1'$ con $\gamma' = \vee_m(\gamma_1')$. Dobbiamo mostrare che $\gamma_1' \in \mathrm{RED}(B,C)$ e cioè che:

$$\forall \delta \in \mathrm{RED}(\neg B).\, cut(\gamma_1', \delta) \in \mathrm{RED}(C),$$
$$\forall \delta \in \mathrm{RED}(\neg C).\, cut(\gamma_1', \delta) \in \mathrm{RED}(B).$$

  Evidentemente $cut(\gamma_1, \delta) \rightsquigarrow_d cut(\gamma_1', \delta)$ per qualsiasi δ: entrambi i punti seguono per ipotesi induttiva su $\mathrm{RED}(B)$ e $\mathrm{RED}(C)$.

- I casi della disgiunzione e congiunzione additiva sono semplici varianti dell'argomento per la congiunzione moltiplicativa.

- $\gamma \in \mathrm{Subst}(\mathrm{RED}(\neg A))$: per ogni $\delta \in \mathrm{RED}(\neg A)$ vale $\gamma[\delta/\vdash A^x] \in \mathrm{SN}_d$; per il corollario 4.10 avremo anche $\gamma'[\delta/\vdash A^x] \in \mathrm{SN}_d$ da cui $\gamma' \in \mathrm{RED}(A)$. □

Proposizione 4.21. *Per qualsiasi* $\gamma \in \mathrm{RED}(A)$, $\delta \in \mathrm{RED}(\neg A)$ *vale*

$$\pi = \frac{\gamma \vdash \Gamma, A^x \qquad \delta \vdash \Delta, \neg A^y}{\vdash \Gamma, \Delta}\,(cut) \in \mathrm{SN}_d.$$

Dimostrazione. Sappiamo già che $\gamma, \delta \in \mathrm{SN}_d$ (proposizione 4.19), per cui $\nu_d(\gamma), \nu_d(\delta) \in \mathbb{N}$. Procediamo per induzione lessicografica sulla coppia ordinata $(h(A), \nu_d(\gamma) + \nu_d(\delta))$: vogliamo mostrare che, per ogni derivazione π',

$$\pi \rightsquigarrow_d \pi' \implies \pi' \in \mathrm{SN}_d.$$

È facile verificare che quest'ultima proposizione è equivalente all'enunciato della proprietà. Esaminiamo separatamente ogni caso:

- $\gamma \rightsquigarrow_d \gamma'$, $\pi' = cut(\gamma', \delta)$: in tal caso, per la proprietà 4.20, $\gamma' \in \mathrm{RED}(A)$. Vale inoltre $\nu_d(\gamma') < \nu_d(\gamma)$ e possiamo applicare l'ipotesi induttiva.
- $\delta \rightsquigarrow_d \delta'$, $\pi' = cut(\gamma, \delta')$: caso identico al precedente.
- $\gamma = ax(A)$, $\pi' = \delta$ oppure $\delta = ax(\neg A)$, $\pi' = \gamma$: immediato.
- $\gamma \in \mathrm{Subst}(\mathrm{RED}(\neg A))$: se A^x è principale, $\pi = \gamma[\delta/\vdash A^x]$ e la conclusione è immediata; altrimenti $\pi \rightsquigarrow_{Sd} \pi' = \gamma[\delta/\vdash A^x]$ e per definizione di Subst possiamo concludere $\pi' \in \mathrm{SN}_d$.
- $\delta \in \mathrm{Subst}(\mathrm{RED}(A))$: caso identico al precedente.
- $A = B \wedge_m C$, $\neg A = \neg B \vee_m \neg C$, $\gamma = \wedge_m(\gamma_1, \gamma_2)$, $\delta = \vee_m(\delta_1)$ dove $\gamma_1 \in \mathrm{RED}(B)$, $\gamma_2 \in \mathrm{RED}(C)$, $\delta_1 \in \mathrm{RED}(\neg B, \neg C)$. Sono possibili due riduzioni:

$$\pi' = \dfrac{\dfrac{\gamma_1 \vdash \Gamma, B \qquad \delta_1 \vdash \Sigma, \neg B, \neg C}{\vdash \Gamma, \Sigma, \neg C}\,(cut) \qquad \gamma_2 \vdash \Delta, C}{\vdash \Gamma, \Sigma, \Delta}\,(cut)$$

  oppure

$$\pi' = \dfrac{\gamma_1 \vdash \Gamma, B \qquad \dfrac{\delta_1 \vdash \Sigma, \neg B, \neg C \qquad \gamma_2 \vdash \Delta, C}{\vdash \Sigma, \Delta, \neg B}\,(cut)}{\vdash \Gamma, \Sigma, \Delta}\,(cut)$$

  Nel primo caso, dal fatto che $\delta_1 \in \mathrm{RED}(\neg B, \neg C)$ e $\gamma_1 \in \mathrm{RED}(B)$ concludiamo che $cut(\gamma_1, \delta_1) \in \mathrm{RED}(\neg C)$. Dato che $\gamma_2 \in \mathrm{RED}(C)$, per ipotesi induttiva $\pi' = cut(cut(\gamma_1, \delta_3), \gamma_2) \in \mathrm{SN}_d$. L'argomento è simile nel secondo caso; inoltre per $A = \neg B \vee_m \neg C$ e scambiando γ con δ si ottiene un caso completamente simmetrico.

- $A = B_1 \vee_a B_2$, $\neg A = \neg B_1 \wedge_a \neg B_2$, $\gamma = \vee_a^i(\gamma_1)$, $\delta = \wedge_a(\delta_1, \delta_2)$ dove $\gamma_1 \in \mathrm{RED}(B_i)$, $\delta_1 \in \mathrm{RED}(\neg B_1)$, $\delta_2 \in \mathrm{RED}(\neg B_2)$, $i \in \{1, 2\}$. La riduzione è

$$\pi' = \dfrac{\gamma_1 \vdash \Gamma, B_i \qquad \delta_i \vdash \Delta, \neg B_i}{\vdash \Gamma, \Delta}\,(cut)$$

  Per ipotesi induttiva $\pi' \in \mathrm{SN}_d$; per $A = \neg B_1 \wedge_a \neg B_2$ e scambiando γ con δ si ottiene di nuovo un caso simmetrico. □

Proposizione 4.22. *Per ogni formula A, gli insiemi* $\mathrm{RED}(A)$ *e* $\mathrm{RED}(\neg A)$ *sono una coppia di candidati simmetrici per* $A, \neg A$.

Dimostrazione. I due insiemi sono evidentemente non vuoti. Abbiamo già osservato (4.18) che le condizioni (i) e (ii) della definizione 4.12 sono soddisfatte per il fatto che $\mathrm{Subst}(\mathrm{RED}(\neg A)) \subseteq \mathrm{RED}(A)$ e $\mathrm{Subst}(\mathrm{RED}(A)) \subseteq \mathrm{RED}(\neg A)$. La proposizione 4.21 mostra che la coppia $\mathrm{RED}(A)$, $\mathrm{RED}(\neg A)$ soddisfa anche la condizione (iii). □

4.3 Dimostrazione del teorema

Teorema 4.23 (Normalizzazione forte). $\mathbf{LK^0} \subseteq \mathrm{SN}$.

Come abbiamo già fatto per $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$, proviamo il teorema come corollario di un lemma più generale che ha l'obiettivo di fornire un'ipotesi induttiva adeguata:

Lemma 4.24. *Sia $n > 0$, $\pi \vdash A_1, \dots, A_n$ e $\delta_i \vdash \Gamma_i, \neg A_i \in \mathrm{RED}(\neg A_i)$ per ogni $1 \le i \le n$. Allora $\pi' = \pi[\delta_1/\vdash A_1, \dots, \delta_n/\vdash A_n] \in \mathrm{SN}_d$.*

Dimostrazione. Si procede per induzione sull'altezza di π. Per alleggerire la notazione commetteremo l'abuso di scrivere $\pi[\delta_i]$ invece che $\pi[\delta_i/\vdash A_i]$. È anche utile, per semplificare il seguito, mettere subito in evidenza una conseguenza dell'ipotesi induttiva. Sia $m \geq 0$, $\rho \vdash B_1, \dots, B_m, B^x$ qualsiasi di altezza strettamente minore di π e per ogni $1 \le i \le m$ scegliamo $\gamma_i \in \mathrm{RED}(\neg B_i)$. Ora, per ogni $\gamma \in \mathrm{RED}(\neg B)$, per ipotesi induttiva vale

$$\rho[\gamma_1, \dots, \gamma_m, \gamma/\vdash B^x] \in \mathrm{SN}_d$$

e di conseguenza $\rho[\gamma_1, \dots, \gamma_m] \in \mathrm{Subst}(\mathrm{RED}(\neg B)) \subseteq \mathrm{RED}(B)$. In altre parole, sostituendo tutte le conclusioni di ρ tranne B^x ottengo una derivazione in $\mathrm{RED}(B)$.

Passiamo dunque all'analisi dei casi. Se l'ultima regola di π non è un taglio, poiché l'ordine delle conclusioni è irrilevante, supponiamo che A_n sia la conclusione principale, contratta o indebolita dell'ultima regola di π:

- $\pi = ax(A_n)$: in tal caso la conclusione di π è $\vdash \neg A_2, A_2$ e abbiamo $A_1 = \neg A_2$, $\delta_1 \in \mathrm{RED}(A_2)$ e $\delta_2 \in \mathrm{RED}(\neg A_2)$. Per definizione di sostituzione abbiamo che $\pi[\delta_1] = \delta_1$, dunque $\pi' = \pi[\delta_1, \delta_2] = \delta_1[\delta_2/\vdash A_2]$. Per la proposizione 4.21 vale $cut(\delta_1, \delta_2) \in \mathrm{SN}_d$ da cui segue $\pi' \in \mathrm{SN}_d$.

- $A_n = B \wedge_m C$, $\pi = \wedge_m(\pi_1, \pi_2)$: in questo caso, per qualche $0 \le k < n$,

$$\pi = \frac{\pi_1 \vdash A_1, \dots, A_k, B \qquad \pi_2 \vdash A_{k+1}, \dots, A_{n-1}, C}{\vdash A_1, \dots, \neg A_{n-1}, A_n = B \wedge_m C} \, (\wedge_m).$$

  Per ipotesi induttiva, $\pi_1' = \pi_1[\delta_1, \dots, \delta_k] \in \mathrm{RED}(B)$ e allo stesso modo vale $\pi_2' = \pi_2[\delta_{k+1}, \dots, \delta_{n-1}] \in \mathrm{RED}(C)$. Allora, per definizione,

$\pi[\delta_1, \dots, \delta_{n-1}] = \wedge_m(\pi'_1, \pi'_2) \in \mathrm{RED}(B \wedge_m C)$ e per la proposizione 4.21 $\pi' = cut(\wedge_m(\pi'_1, \pi'_2), \delta_n) \in \mathrm{SN}_d$.

- $A_n = B \vee_m C$, $\pi = \vee_m(\pi_1)$: in tal caso $\pi_1 \vdash A_1, \dots, A_{n-1}, B^x, C^y$ e vogliamo mostrare che $\pi'_1 = \pi_1[\delta_1, \dots, \delta_{n-1}] \in \mathrm{RED}(B, C)$. Ne seguirà che $\pi[\delta_1, \dots, \delta_{n-1}] \in \mathrm{RED}(B \vee_m C)$ e potremo concludere $\pi' \in \mathrm{SN}_d$ argomentando come nel caso precedente. Scegliamo qualsiasi $\gamma_1 \in \mathrm{RED}(\neg B)$ e $\gamma_2 \in \mathrm{RED}(\neg C)$ e consideriamo le due derivazioni seguenti:

$$\rho_1 = \frac{\pi'_1 \vdash \Gamma, B^x, C^y \qquad \gamma_1 \vdash \Delta_1, \neg B}{\vdash \Gamma, C^y, \Delta_1}\ (cut)$$

$$\rho_2 = \frac{\pi'_1 \vdash \Gamma, B^x, C^y \qquad \gamma_2 \vdash \Delta_2, \neg C}{\vdash \Gamma, B^x, \Delta_2}\ (cut)$$

  Verifichiamo che:

  a. $\rho_1 \in \mathrm{RED}(C)$: è evidente che $\rho_1[\gamma_2/\vdash C^y] = cut(\pi'_1[\gamma_2/\vdash C^y], \gamma_1)$. Per ipotesi induttiva, $\pi'_1[\gamma_2/\vdash C^y] = \pi_1[\delta_1, \dots, \delta_{n-1}, \gamma_2] \in \mathrm{RED}(B)$: allora per la proposizione 4.21 abbiamo $\rho_1[\gamma_2/\vdash C^y] \in \mathrm{SN}_d$ da cui (ricordando che γ_2 è scelta liberamente) $\rho_1 \in \mathrm{Subst}(\mathrm{RED}(\neg C)) \subseteq \mathrm{RED}(C)$.
  b. $\rho_2 \in \mathrm{RED}(B)$: il ragionamento è analogo.

- I casi delle regole logiche additive fanno semplicemente uso dell'ipotesi induttiva e della proprietà 4.21, come nel caso della congiunzione moltiplicativa. Si presterà attenzione al fatto che, nel caso della regola $(\wedge_a)$, le derivazioni $\delta_1, \dots, \delta_{n-1}$ sono duplicate anziché distribuite fra le premesse.

- $\pi = W(\pi_1)$ dove $\pi_1 \vdash A_1, \dots, A_{n-1}$:

$$\frac{\pi_1 \vdash A_1, \dots, A_{n-1}}{\vdash A_1, \dots, A_{n-1}, A_n}\ (W).$$

  Per ipotesi induttiva, $\pi'_1 = \pi_1[\delta_1, \dots, \delta_{n-1}] \in \mathrm{SN}_d$ e π' avrà la forma seguente:

$$\frac{\pi'_1 \vdash \Gamma_1, \dots, \Gamma_{n-1}}{\vdash \Gamma_1, \dots, \Gamma_{n-1}, \Gamma_n}\ (W).$$

  Evidentemente se $\pi'_1 \in \mathrm{SN}_d$ allora anche $\pi' \in \mathrm{SN}_d$.

- $\pi = C(\pi_1)$ dove $\pi_1 \vdash A_1, \dots, A_{n-1}, A_n^x, A_n^y$: poiché questo caso è molto simile al precedente, omettiamo le illustrazioni. Per ipotesi induttiva,

$$\pi'_1 = \pi_1[\delta_1, \dots, \delta_{n-1}, \delta_n/\vdash A_n^x, \delta_n/\vdash A_n^y] \in \mathrm{SN}_d.$$

  Avremo inoltre $\pi' = C(\dots C(\pi'_1)\dots)$, dove il segno di ellissi indica che abbiamo effettuato il numero di contrazioni necessario per le due copie del contesto Γ_n di δ_n: come prima, da $\pi'_1 \in \mathrm{SN}_d$ segue $\pi' \in \mathrm{SN}_d$.

Resta da verificare il caso in cui l'ultima regola di π sia un taglio:

$$\pi = \frac{\pi_1 \vdash A_1, \dots, A_k, B \qquad \pi_2 \vdash A_{k+1}, \dots, A_n, \neg B}{\vdash A_1, \dots, A_n} \; (cut)$$

dove $0 \leq k \leq n$ e B è una formula qualsiasi. Per ipotesi induttiva, $\pi'_1 = \pi_1[\delta_1, \dots, \delta_k] \in \mathrm{RED}(B)$ e $\pi'_2 = \pi_2[\delta_{k+1}, \dots, \delta_n] \in \mathrm{RED}(\neg B)$. Allora per la proposizione 4.21 possiamo concludere che $\pi' = cut(\pi'_1, \pi'_2) \in \mathrm{SN}_d$. □

Dimostrazione del teorema 4.23. Sia $\pi \vdash A_1, \dots, A_n$ una derivazione qualsiasi con $n > 0$. Per ogni $1 \leq i \leq n$ scegliamo $\delta_i = ax(\neg A_i) \in \mathrm{RED}(\neg A_i)$. Per il lemma 4.24 vale

$$\pi' = \pi[\delta_1 / \vdash \neg A_1, \dots, \delta_n / \vdash \neg A_n] \in \mathrm{SN}_d$$

e sappiamo inoltre che $\pi' \rightsquigarrow^*_{Sd} \pi$ riducendo tutti i tagli residui con assioma introdotti in π' dalla sostituzione. Abbiamo già mostrato che $\mathrm{SN}_d \subseteq \mathrm{SN}$ (proposizione 4.7). □

4.4 Estensione al secondo ordine

Vogliamo estendere il risultato della sezione precedente a $\mathbf{LK^2}$ (dunque lavoreremo su formule proposizionali del secondo ordine). Dobbiamo innanzitutto estendere il nostro abuso degli operatori logici ai quantificatori del secondo ordine. Sia $X_1 \subseteq \mathcal{D}_A$ tale che per ogni $\pi \in X_1$, $\pi \vdash \Gamma, A^x$ con Y non libera in Γ e sia $X_2 \subseteq \bigcup_{V \in \mathcal{F}^2} \mathcal{D}_{A[V/Y]}$: poniamo

$$\forall^2_Y(X_1) = \left\{ \frac{\pi \vdash \Gamma, A^x}{\vdash \Gamma, \forall Y.A} \, (\forall^2) : \pi \in X_1 \right\},$$

$$\exists^2_Y(X_2) = \left\{ \frac{\pi \vdash \Gamma, A[V/Y]^x}{\vdash \Gamma, \exists Y.A} \, (\exists^2) : \pi \in X_2 \right\}.$$

Definizione 4.25 (Riducibilità parametrica). Sia A una formula positiva: fissiamo una sequenza $\underline{X} = X_1, \dots, X_n$ di variabili proposizionali con $FV(A) \subseteq \underline{X}$, una sequenza $\underline{U} = U_1, \dots, U_n$ di formule, una sequenza $\underline{\mathcal{R}} = \mathcal{R}_1, \dots, \mathcal{R}_n$ di insiemi tali che $\mathcal{R}_i, \mathcal{R}_i^{\neg}$ siano una coppia di candidati simmetrici per $U_i, \neg U_i$, per ogni $1 \leq i \leq n$. Definiamo per induzione su A la coppia di insiemi $\mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}] \subseteq \mathcal{D}_{A[\underline{U}/\underline{X}]}$, $\mathrm{RED}(\neg A)[\underline{\mathcal{R}}/\underline{X}] \subseteq \mathcal{D}_{\neg A[\underline{U}/\underline{X}]}$:

- $A = X_i$, $\neg A = \neg X_i$ per qualche i: poniamo $\mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}] = \mathcal{R}_i$ e $\mathrm{RED}(\neg A)[\underline{\mathcal{R}}/\underline{X}] = \mathcal{R}_i^{\neg}$. Poniamo inoltre:

$$\mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}]^{\perp} = \mathrm{RED}(\neg A)[\underline{\mathcal{R}}/\underline{X}],$$
$$\mathrm{RED}(\neg A)[\underline{\mathcal{R}}/\underline{X}]^{\perp} = \mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}].$$

– A non atomica: definiamo come prima gli insiemi $P^\perp, N^\perp$ per ogni coppia di insiemi $P \subseteq \mathcal{D}_{A[\underline{U}/\underline{X}]}$, $N \subseteq \mathcal{D}_{\neg A[\underline{U}/\underline{X}]}$; definiamo quindi $\mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}] \subseteq \mathcal{D}_{A[\underline{U}/\underline{X}]}$ come il più piccolo insieme tale che

$$\mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}] = \mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}]^{\perp\perp}$$

e poniamo

$$\mathrm{RED}(\neg A)[\underline{\mathcal{R}}/\underline{X}] = \mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}]^{\perp}.$$

I vari casi sono identici a quelli della definizione 4.16 avendo cura di propagare tutte le sostituzioni su tipi e variabili; enunciamo solamente il nuovo caso per $A = \exists Y.B$, $\neg A = \forall Y.\neg B$:

$$P^\perp = \{ax(\neg A[\underline{U}/\underline{X}])\} \cup \forall^2_Y(\mathrm{RED}_{\forall Y}(\neg B)[\underline{\mathcal{R}}/\underline{X}]) \cup \mathrm{Subst}(P),$$
$$N^\perp = \{ax(A[\underline{U}/\underline{X}])\} \cup \exists^2_Y(\mathrm{RED}_{\exists Y}(B)[\underline{\mathcal{R}}/\underline{X}]) \cup \mathrm{Subst}(N)$$

dove $\mathrm{RED}_{\forall Y}(\neg B)[\underline{\mathcal{R}}/\underline{X}] \subseteq \mathcal{D}_{\neg B[\underline{\mathcal{R}}/\underline{X}]}$ è definito come l'insieme tale che $\pi \in \mathrm{RED}_{\forall Y}(\neg B)[\underline{\mathcal{R}}/\underline{X}]$ sse

(i) $\pi \vdash \Gamma, \neg B^x$ con Y non libera in Γ;

(ii) per ogni formula V e per ogni coppia $\mathcal{S}, \mathcal{S}^\neg$ di candidati simmetrici per $V, \neg V$ vale $\pi[V/Y] \in \mathrm{RED}(\neg B)[\underline{\mathcal{R}}/\underline{X}, \mathcal{S}/Y]$.

Dualmente $\mathrm{RED}_{\exists Y}(B)[\underline{\mathcal{R}}/\underline{X}]$ è definito come l'insieme di derivazioni tale che $\pi \in \mathrm{RED}_{\exists Y}(B)[\underline{\mathcal{R}}/\underline{X}]$ sse

(i) $\pi \vdash \Gamma, B[V/Y]^x$ per qualche formula V;

(ii) $\pi \in \mathrm{RED}(B)[\underline{\mathcal{R}}/\underline{X}, \mathcal{S}/Y]$ per qualche coppia $\mathcal{S}, \mathcal{S}^\neg$ di candidati simmetrici per $V, \neg V$.

Come prima la definizione di $\mathrm{RED}(A)$ è ben posta grazie al teorema del punto fisso di Tarksi.

Lemma 4.26 (Sostituzione).

$$\mathrm{RED}(A[V/Y])[\underline{\mathcal{R}}/\underline{X}] = \mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}, \mathrm{RED}(V)[\underline{\mathcal{R}}/\underline{X}]/Y].$$

Dimostrazione. L'enunciato si dimostra molto facilmente per induzione su A. Seguendo Girard in [Gir87] e [GLT89] ricordiamo che in questo passaggio si fa implicitamente uso dello schema di comprensione al secondo ordine per utilizzare il predicato $\mathrm{RED}(V)[\underline{\mathcal{R}}/\underline{X}]$ come un insieme. □

Dobbiamo verificare di nuovo che gli insiemi $\mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}]$, $\mathrm{RED}(\neg A)[\underline{\mathcal{R}}/\underline{X}]$ siano una coppia di candidati simmetrici per $A[\underline{U}/\underline{X}], \neg A[\underline{U}/\underline{X}]$. Per $A, \neg A$ atomiche le condizioni (i) e (ii) della definizione 4.12 sono soddisfatte per ipotesi su $\underline{\mathcal{R}}$; per formule non atomiche basta osservare che

$$\mathrm{Subst}(\mathrm{RED}(\neg A)[\underline{\mathcal{R}}/\underline{X}]) \subseteq \mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}].$$

Le dimostrazioni saranno molto simili a quelle già presentate, con l'aggiunta delle sostituzioni che rendono tutto meno leggibile: per questo verifichiamo solamente i nuovi casi e quelli in cui le sostituzioni hanno un ruolo attivo.

Proposizione 4.27. $\mathrm{RED}(A)\,[\underline{\mathcal{R}}/\underline{X}] \subseteq \mathrm{SN}_d$.

Dimostrazione. Per induzione su A. Per A atomica la conclusione è immediata dal fatto che per ogni $\mathcal{R}_i$ in $\underline{\mathcal{R}}$, vale $\mathcal{R}_i, \mathcal{R}_i^\neg \subseteq \mathrm{SN}_d$ (osservazione 4.14). Mancano i casi dei due quantificatori (sempre considerando la forma di $\gamma \in \mathrm{RED}(A)\,[\underline{\mathcal{R}}/\underline{X}]$):

- $A = \exists Y.B$ e $\gamma = \exists^2_Y(\gamma')$ con $\gamma' \in \mathrm{RED}(B)\,[\underline{\mathcal{R}}/\underline{X}, \mathcal{S}/Y]$ per qualche formula V e coppia $\mathcal{S}, \mathcal{S}^\neg$ di candidati simmetrici per $V, \neg V$: ogni riduzione su γ è una riduzione su γ' e possiamo applicare l'ipotesi induttiva per concludere che $\gamma' \in \mathrm{SN}_d$.
- $A = \forall Y.B$ e $\gamma = \forall^2_Y(\gamma')$ con $\gamma'[V/Y] \in \mathrm{RED}(B)\,[\underline{\mathcal{R}}/\underline{X}, \mathcal{S}/Y]$ per ogni formula V e per ogni coppia $\mathcal{S}, \mathcal{S}^\neg$ di candidati simmetrici per $V, \neg V$: poniamo $V = Y$ e scegliamo $\mathcal{S}, \mathcal{S}^\neg$ qualsiasi; avremo $\gamma' = \gamma'[V/Y]$ e l'ipotesi induttiva permette di concludere.

Come prima, per A non atomica e $\gamma \in \mathrm{Subst}(\mathrm{RED}(\neg A)\,[\underline{\mathcal{R}}/\underline{X}])$ la conclusione è immediata dal fatto che $ax(\neg A[\underline{U}/\underline{X}]) \in \mathrm{RED}(\neg A)\,[\underline{\mathcal{R}}/\underline{X}]$ e $\gamma[ax(\neg A)/\vdash A^x] \rightsquigarrow^*_{Sd} \gamma$. □

Osservazione 4.28. Non sfuggirà all'occhio attento che la dimostrazione appena conclusa riposa sulla possibilità di scegliere (cioè sull'esistenza di) una coppia qualsiasi $\mathcal{S}, \mathcal{S}^\neg$ di candidati simmetrici per $Y, \neg Y$ formule atomiche. Non è possibile nel nostro caso seguire [GLT89] e scegliere l'insieme $\mathcal{D}_Y \cap \mathrm{SN}_d$ (le derivazioni fortemente normalizzanti di tipo Y), ma abbiamo già mostrato che la coppia di insiemi $\mathrm{RED}(Y), \mathrm{RED}(\neg Y)$ (senza sostituzioni, secondo la definizione 4.16) soddisfa le condizioni di riducibilità della definizione 4.12.

Proposizione 4.29. *Sia* $\gamma \in \mathrm{RED}(A)\,[\underline{\mathcal{R}}/\underline{X}]$, $\gamma \rightsquigarrow_d \gamma'$. *Allora* $\gamma' \in \mathrm{RED}(A)\,[\underline{\mathcal{R}}/\underline{X}]$.

Dimostrazione. Per induzione su A. Per A atomica segue dalla proprietà 4.15 applicata a $\underline{\mathcal{R}}$. Per A non atomica e $\gamma \in \mathrm{Subst}(\mathrm{RED}(\neg A)\,[\underline{\mathcal{R}}/\underline{X}])$ segue come prima dal corollario 4.10. Per i casi dei quantificatori la conclusione si dimostra facilmente applicando l'ipotesi induttiva alla sottoderivazione premessa: il ragionamento è analogo a quello svolto in precedenza per gli altri connettivi. □

Proposizione 4.30. *Per qualsiasi* $\gamma \in \mathrm{RED}(A)\,[\underline{\mathcal{R}}/\underline{X}]$, $\delta \in \mathrm{RED}(\neg A)\,[\underline{\mathcal{R}}/\underline{X}]$ *vale*

$$\pi = \frac{\gamma \vdash \Gamma, A[\underline{U}/\underline{X}]^x \qquad \delta \vdash \Delta, \neg A[\underline{U}/\underline{X}]^y}{\vdash \Gamma, \Delta}\,(cut) \in \mathrm{SN}_d.$$

Dimostrazione. L'argomento è impostato come nella dimostrazione della proposizione 4.21. Per $A, \neg A$ atomiche la conclusione segue come sempre dalle ipotesi su $\underline{\mathcal{R}}$. Tutti i casi trattati nella dimostrazione della proposizione 4.21 si ripetono con l'aggiunta delle sostituzioni sui candidati. Verifichiamo che la proposizione valga anche nel caso delicato di un taglio logico fra quantificatori. Se abbiamo che:

(i) $A = \exists Y.B$, $\neg A = \forall Y.\neg B$;

(ii) $\gamma = \exists^2_Y(\gamma_1)$, $\delta = \forall^2_Y(\delta_1)$;

(iii) $\gamma_1 \vdash \Gamma, B[\underline{U}/\underline{X}, V/Y]$ per qualche V, con $\gamma_1 \in \mathrm{RED}(B)[\underline{\mathcal{R}}/\underline{X}, \mathcal{S}/Y]$ per qualche coppia $\mathcal{S}, \mathcal{S}^\neg$ di candidati simmetrici per $V, \neg V$;

(iv) $\delta_1 \vdash \Delta, \neg B[\underline{U}/\underline{X}]$ con Y non libera in Δ e per ogni formula W e coppia $\mathcal{W}, \mathcal{W}^\neg$ di candidati per $W, \neg W$ vale $\delta_1[W/Y] \in \mathrm{RED}(\neg B)[\underline{\mathcal{R}}/\underline{X}, \mathcal{W}/Y]$;

allora π si riduce con un passo per quantificatori alla derivazione seguente:

$$\pi' = \frac{\gamma_1 \vdash \Gamma, B[\underline{U}/\underline{X}, V/Y] \qquad \delta_1[V/Y] \vdash \Delta, \neg B[\underline{U}/\underline{X}, V/Y]}{\vdash \Gamma, \Delta}\,(cut)$$

dove per l'ipotesi (iii) abbiamo $\gamma_1 \in \mathrm{RED}(B)[\underline{\mathcal{R}}/\underline{X}, \mathcal{S}/Y]$ e per l'ipotesi (iv) abbiamo $\delta_1 \in \mathrm{RED}(\neg B)[\underline{\mathcal{R}}/\underline{X}, \mathcal{S}/Y]$. Per ipotesi induttiva possiamo concludere $\pi' \in \mathrm{SN}_d$. □

Proposizione 4.31. *Per ogni formula A, $\underline{X} = X_1, \dots, X_n$ sequenza di variabili proposizionali con $FV(A) \subseteq \underline{X}$, fissata $\underline{U} = U_1, \dots, U_n$ sequenza qualsiasi di formule e $\underline{\mathcal{R}} = \mathcal{R}_1, \dots, \mathcal{R}_n$ sequenza di insiemi tali che, per ogni $1 \le i \le n$, $\mathcal{R}_i, \mathcal{R}_i^\neg$ siano una coppia di candidati simmetrici per $U_i, \neg U_i$, gli insiemi*

$$\mathrm{RED}(A)[\underline{\mathcal{R}}/\underline{X}],\ \mathrm{RED}(\neg A)[\underline{\mathcal{R}}/\underline{X}]$$

sono una coppia di candidati simmetrici per $A[\underline{U}/\underline{X}], \neg A[\underline{U}/\underline{X}]$.

4.4.1 Normalizzazione forte per $\mathbf{LK^2}$

Lemma 4.32. *Sia $n > 0$, $\pi \vdash A_1, \dots, A_n$ e $X = X_1, \dots, X_m$ la lista di tutte le variabili proposizionali libere nella conclusione di π. Per ogni lista di formule $\underline{U} = U_1, \dots, U_m$ a cui associamo una lista di candidati simmetrici $\underline{\mathcal{R}} = \mathcal{R}_1, \dots, \mathcal{R}_m$ (come sempre $\mathcal{R}_i, \mathcal{R}_i^\neg$ saranno candidati simmetrici per $U_i, \neg U_i$) e scelte infine n derivazioni $\delta_1, \dots, \delta_n$ tali che, per ogni $1 \le i \le n$, $\delta_i \in \mathrm{RED}(\neg A_i)[\underline{\mathcal{R}}/\underline{X}]$, vale*

$$\pi' = \pi[\underline{U}/\underline{X}][\delta_1/\vdash A_1[\underline{U}/\underline{X}], \dots, \delta_n/\vdash A_n[\underline{U}/\underline{X}]] \in \mathrm{SN}_d.$$

Dimostrazione. Per induzione sull'altezza di π. Verifichiamo solamente i casi che devono essere gestiti diversamente da prima, per tutti gli altri è sufficiente aggiungere le sostituzioni su formule e candidati nella dimostrazione del lemma 4.24:

- $A_n = \exists Y.B$, $\pi = \exists^2_Y(\pi_1)$: in tal caso $\pi_1 \vdash A_1, \dots, A_{n-1}, B[V/Y]$ per qualche formula V. Per ipotesi induttiva,

  $$\pi'_1 = \pi_1[\underline{U}/\underline{X}][\delta_1, \dots, \delta_{n-1}] \in \mathrm{RED}(B[V/Y])[\underline{\mathcal{R}}/\underline{X}, \underline{\mathcal{S}}/\underline{Z}]$$

  dove $\underline{Z}$ sono le variabili libere in V ma *non presenti in* $\underline{X}$ e $\underline{\mathcal{S}}$ è una sequenza di candidati simmetrici di tipo $\underline{Z}, \neg\underline{Z}$.[7] Il lemma 4.26 ci dice che

  $$\mathrm{RED}(B[V/Y])[\underline{\mathcal{R}}/\underline{X}, \underline{\mathcal{S}}/\underline{Z}] = \mathrm{RED}(B)[\underline{\mathcal{R}}/\underline{X}, \mathrm{RED}(V)[\underline{\mathcal{R}}/\underline{X}, \underline{\mathcal{S}}/\underline{Z}]/Y]:$$

  avremo dunque $\pi'_1 \in \mathrm{RED}_{\exists Y}(B)[\underline{\mathcal{R}}/\underline{X}]$ e $\exists^2_Y(\pi'_1) \in \mathrm{RED}(A_n)[\underline{\mathcal{R}}/\underline{X}]$. Possiamo concludere sfruttando la proposizione 4.30 e osservando che $\pi' = cut(\exists^2_Y(\pi'_1), \delta_n)$.

- $A_n = \forall Y.B$, $\pi = \forall^2_Y(\pi_1)$: in questo caso $\pi_1 \vdash A_1, \dots, A_{n-1}, B$ con Y libera *al più* in B. Sia $\pi'_1 = \pi_1[\underline{U}/\underline{X}][\delta_1, \dots, \delta_{n-1}]$: per ipotesi induttiva, scelta qualsiasi formula V e coppia $\mathcal{S}, \mathcal{S}^\neg$ di candidati simmetrici per $V, \neg V$, avremo $\pi'_1[V/Y] \in \mathrm{RED}(B)[\underline{\mathcal{R}}/\underline{X}, \mathcal{S}/Y]$. Dunque $\pi'_1 \in \mathrm{RED}_{\forall Y}(B)[\underline{\mathcal{R}}/\underline{X}]$ e $\forall^2_Y(\pi'_1) \in \mathrm{RED}(B)[\underline{\mathcal{R}}/\underline{X}]$. Ancora una volta osserviamo che $\pi' = cut(\forall^2_Y(\pi'_1), \delta_n)$ e sfruttiamo la proposizione 4.30.

- $\pi = cut(\pi_1, \pi_2)$ dove $\pi_1 \vdash A_1, \dots, A_k, B$ e $\pi_2 \vdash A_{k+1}, \dots, A_n, \neg B$ per qualche formula di taglio B e $0 \le k \le n$: il caso sarà gestito come nella dimostrazione del lemma 4.24, ma dovremo prestare attenzione alle variabili libere in $B, \neg B$ e non presenti in $\underline{X}$. Sia $\underline{Z}$ la lista di tali variabili: come nel caso esistenziale sceglieremo una lista $\underline{\mathcal{S}}$ (stessa lunghezza di $\underline{Z}$) di candidati simmetrici di tipo $\underline{Z}, \neg\underline{Z}$ e potremo concludere il ragionamento usando $\mathrm{RED}(B)[\underline{\mathcal{R}}/\underline{X}, \underline{\mathcal{S}}/\underline{Z}]$ e $\mathrm{RED}(\neg B)[\underline{\mathcal{R}}/\underline{X}, \underline{\mathcal{S}}/\underline{Z}]$.

È importante osservare che nel corso della dimostrazione si usano implicitamente più volte i punti (i) e (ii) della definizione 4.12 applicata a $\underline{\mathcal{R}}$: per la precisione in tutti i casi in cui si abbia A_n non principale e $A_n = X_i$ per qualche $1 \le i \le m$. □

Teorema 4.33 (Normalizzazione forte per $\mathbf{LK^2}$). $\mathbf{LK^2} \subseteq \mathrm{SN}$.

Dimostrazione. Applichiamo il lemma 4.32 a qualsiasi $\pi \vdash A_1, \dots, A_n$ con $\underline{U} = \underline{X}$, $\underline{\mathcal{R}}$ qualsiasi e $\delta_i = ax(\neg A_i)$ per ogni $1 \le i \le n$. □

[7] Per ogni Z_i in $\underline{Z}$, $\mathcal{S}_i, \mathcal{S}_i^\neg$ è una coppia di candidati simmetrici di tipo $Z_i, \neg Z_i$, che possiamo scegliere liberamente per l'osservazione 4.28.

Capitolo 5

Difficoltà e questioni aperte

Raccogliamo in quest'ultimo capitolo alcune questioni aperte e trattiamo alcune difficoltà che, per favorire la leggibilità, avevamo omesso di discutere nel capitolo precedente.

5.1 Normalizzazione forte per LK con stili misti

La distinzione fra connettivi additivi e moltiplicativi in logica classica è utile ai fini dello studio delle derivazioni, ma è chiaramente artificiale: in presenza delle regole strutturali i due stili sono completamente equivalenti dal punto di vista della derivabilità, mentre dal punto di vista della semantica classica non è neanche possibile tracciare una distinzione.

La definizione "naturale" del calcolo dei sequenti **LK** dunque non distingue formule di stile diverso e permette di contrarre e tagliare liberamente le conclusioni di regole di stile diverso (parliamo in tal caso di **LK** *hetero-style*, distinguendolo dalla versione *homo-style* da noi adottata). Le procedure di normalizzazione sono le stesse: possono essere definite come quelle di $\mathbf{LK}^{tq}$ dimenticando i colori e con l'aggiunta di alcuni passi per i tagli di stile misto [AT14]. È noto che, almeno in parte, tali passi non possono essere simulati da quelli *homo-style* [JST96].

È dunque naturale chiedersi se la tecnica dimostrativa esibita nel capitolo precedente possa adattarsi al sistema più generale possibile per la logica classica. Congetturiamo che la risposta sia positiva: il nostro argomento è stato formulato inizialmente in un sistema ristretto che ammetteva tagli misti (quale è peraltro $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$) e in seguito esteso alla versione completa, ma *homo-style*, di **LK**.

Sarà necessario riadattare la definizione di ortogonalità e i predicati di riducibilità. Ovviamente non è possibile sfruttare la partizione delle formule in positive e negative; si potrà utilizzare invece la partizione $\wedge/\vee$ già adottata nel capitolo 3 per $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$. Inoltre gli operatori moltiplicativi e additivi per lo stesso connettivo dovranno essere accoppiati nelle definizioni; ad esempio per $X \subseteq \mathcal{D}_{\neg A \vee \neg B}$ definiremo $X^{\perp}$ come segue:

$$\begin{aligned} X^{\perp} = \{ax(A \wedge B)\} &\cup \wedge_m(\mathrm{RED}(A), \mathrm{RED}(B)) \\ &\cup \wedge_a(\mathrm{RED}(A), \mathrm{RED}(B)) \cup \mathrm{Subst}(X) \end{aligned}$$

e dovremo fare la stessa cosa per la disgiunzione. Non si intravedono particolari difficoltà, ma non abbiamo analizzato la questione nel dettaglio.

5.2 Perché il passo Sd

Abbiamo già accennato al fatto che l'introduzione del passo $\rightsquigarrow_{Sd}$ è giustificata da una difficoltà relativa al comportamento dell'assioma di **LK** sotto riduzione, che è differente dal comportamento di una variabile del λ-calcolo simmetrico. Vediamo ora la questione nel dettaglio.

Avendo due conclusioni, l'assioma di **LK** può essere foglia di due alberi strutturali diversi nella stessa derivazione. La difficoltà nasce quando *entrambi* gli alberi sono coinvolti in un passo di riduzione strutturale: in tal caso, effettuare i due passi in ordine diverso può portare a forme normali molto diverse. Consideriamo la derivazione seguente:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\,\gamma\\ \vdash \Gamma, A\end{array} \qquad \begin{array}{c}\dfrac{}{\vdash \neg A^x, A^y}\,(ax)\\ \vdots\,\sigma\\ \vdash \Sigma, \neg A^x, A^y\end{array}}{\vdash \Gamma, \Sigma, A^y}\,(cut) \qquad \begin{array}{c}\vdots\,\delta\\ \vdash \Delta, \neg A\end{array}}{\vdash \Gamma, \Sigma, \Delta}\,(cut)$$

In questo caso particolare, le conclusioni dell'assioma non sono mai utilizzate all'interno di σ, ma questo non è rilevante: il problema si manifesta per qualsiasi forma degli interspazi di $\neg A^x$ e A^y.

Supponiamo che le formule di taglio di γ e δ abbiano entrambe interspazio non piatto: è perfettamente possibile nel caso generale. Osserviamo che riducendo prima il taglio superiore e poi quello inferiore – cioè trasportando prima γ e poi δ – trasporteremo δ all'interno di γ, mentre riducendo prima il taglio inferiore avremo la situazione opposta (γ all'interno di δ):

$$\dfrac{\begin{array}{c}\vdots\,\gamma\\ \vdash \Gamma, A^y\\ \vdots\,\sigma\\ \vdash \Gamma, \Sigma, A^y\end{array} \qquad \begin{array}{c}\vdots\,\delta\\ \vdash \Delta, \neg A\end{array}}{\vdash \Gamma, \Sigma, \Delta}\,(cut)$$

$$\dfrac{\begin{array}{c}\vdots\,\gamma\\ \vdash \Gamma, A\end{array} \qquad \begin{array}{c}\vdots\,\delta\\ \vdash \Delta, \neg A^x\\ \vdots\,\sigma\\ \vdash \Sigma, \Delta, \neg A^x\end{array}}{\vdash \Gamma, \Sigma, \Delta}\,(cut)$$

In buona sostanza, l'assioma rappresenta un taglio "futuro" fra γ e δ, "nascosto" dalla struttura arborescente della derivazione, e scegliere quale ridurre per primo equivale a scegliere come ridurre il taglio nascosto. Poniamo allora $\rho = cut(\gamma, \sigma)$ (il taglio superiore): evidentemente $\rho \rightsquigarrow_S \rho' = \sigma[\gamma / \vdash \neg A^x]$ riducendo il taglio superiore, mentre $\rho[\delta / \vdash A^y]$ è ottenuta riducendo il taglio inferiore, quindi in generale

$$\rho[\delta / \vdash A^y] \not\rightsquigarrow \rho'[\delta / \vdash A^y].$$

Abbiamo dunque per il passo S un controesempio alla proposizione che intendiamo provare (4.9): non sempre l'operatore di sostituzione commuta con un passo di riduzione nella derivazione su cui opera.

Il motivo per cui questa difficoltà non si presenta nel λ-calcolo simmetrico[1] è l'asimmetria fra premesse e conclusioni: una variabile non rappresenta "due conclusioni" come l'assioma, ma una premessa e una conclusione che possono essere scambiate usando gli operatori per la negazione. Pertanto, non è possibile che *due* sottotermini sostituiscano la stessa variabile durante la riduzione.

Volendo tradurre l'assioma di **LK** in $\lambda^{\text{Sym}}_{\text{Prop}}$, la resa migliore è quella proposta da Laird in [Lai01], cioè un taglio fra due variabili: $x \star y$. Supponendo di aver tradotto la derivazione σ di prima, avremo un termine con la forma seguente:

$$\sigma[x \star y].$$

Poiché abbiamo detto che le formule di taglio di γ e δ non sono principali, ricordando che la λ-astrazione permette di portare in evidenza conclusioni "fuori dallo stoup", le traduzioni delle due derivazioni avranno la forma seguente:

$$\lambda z.\gamma \qquad \lambda w.\delta$$

e potremo scrivere i due tagli come segue:

$$(\lambda y.((\lambda z.\gamma) \star \lambda x.\sigma[x \star y])) \star \lambda w.\delta.$$

La sostituzione dell'assioma $x \star y$ con δ allora non avviene in un passo di riduzione, ma in *due* passi:

$$\begin{aligned} & (\lambda y.((\lambda z.\gamma) \star \lambda x.\sigma[x \star y])) \star \lambda w.\delta \\ \rightsquigarrow_\beta \; & (\lambda z.\gamma) \star \lambda x.\sigma[x \star \lambda w.\delta] \\ \rightsquigarrow_{\beta^\perp} \; & (\lambda z.\gamma) \star \lambda x.\sigma[\delta[x/w]]; \end{aligned}$$

è questo "ritardo" che permette alle due riduzioni di commutare:

$$\begin{aligned} & (\lambda y.((\lambda z.\gamma) \star \lambda x.\sigma[x \star y])) \star \lambda w.\delta \\ \rightsquigarrow_\beta \; & (\lambda z.\gamma) \star \lambda x.\sigma[x \star \lambda w.\delta] \\ \rightsquigarrow_{\beta^\perp} \; & \sigma[(\lambda z.\gamma) \star \lambda w.\delta] \\ \rightsquigarrow_\beta \; & \sigma[\gamma[\lambda w.\delta/z]]. \end{aligned}$$

[1] Né tantomeno in $\mathbf{LK}^{tq}$, a causa dell'orientamento dell'assioma: solo una delle conclusioni sarà attrattiva, per cui sarà determinato fin dall'inizio quale fra γ e δ deve "salire dentro l'altra".

Da qui nasce l'idea del passo Sd, cioè *d-elayed*, ritardato, che lascia un taglio con assioma "in sospeso" dove necessario. Va osservato che la nostra definizione introdurrà tagli residui con assioma in alcuni casi in cui non sarebbe strettamente necessario, ma per aumentare l'accuratezza si dovrebbero ispezionare sottoderivazioni di taglia arbitrariamente grande, mentre la piattezza degli interspazi è una condizione verificabile in tempo costante.

5.3 L'ordine delle riduzioni è rilevante

Vogliamo discutere qui brevemente un fenomeno a cui abbiamo già accennato nella dimostrazione della proposizione 4.9: l'ordine in cui si effettuano passi di riduzione apparentemente indipendenti può influenzare l'insieme di forme normali raggiungibili (restringendolo in modi diversi), ed è pertanto una ulteriore fonte di non determinismo dei processi di riduzione in **LK** che si aggiunge al dilemma strutturale e a quello logico.

Un esempio evidente di questo fenomeno è il caso patologico discusso all'inizio di questa sezione: riducendo i due tagli in ordine diverso, si perdeva l'accesso a diversi insiemi di ridotti e dunque di forme normali.

La d-riduzione mitiga questo fenomeno, ma non lo elimina completamente. Il caso rilevante è proprio quello sottolineato in precedenza (dimostrazione della proposizione 4.9):

$$\begin{array}{c}
\dfrac{\begin{array}{c}\vdots\ \gamma\\ \vdash \Gamma, A^x\end{array} \ \overset{Sd}{\dashrightarrow}\quad \dfrac{}{\vdash A^y, \neg A^z}\,(ax)}{\vdash \Gamma, A^y}\,(cut)\\
\vdots\\
\vdash \Sigma
\end{array}$$

Supponendo sempre che A^x *non sia* la conclusione principale di γ, avremo due scelte per ridurre il taglio in figura: trasportare l'assioma o sostituirlo con γ (come indicato dalla freccia). Queste due scelte sono differenti solo in apparenza: in entrambi i casi, una riduzione strutturale che interagisse con la conclusione A^y dell'assioma trasporterebbe qualche altra sottoderivazione (diciamo δ) all'interno di γ. Supponiamo invece di effettuare prima quest'ultima operazione (sarà la riduzione di qualche altro taglio nella derivazione); ci troveremmo nella situazione seguente:

$$\begin{array}{c}
\dfrac{\begin{array}{c}\vdots\ \gamma\\ \vdash \Gamma, A^x\end{array} \ \overset{Sd}{\dashrightarrow}\quad \begin{array}{c}\vdots\ \delta\\ \vdash \Delta, \neg A^w\end{array}}{\vdash \Gamma, \Delta}\,(cut)\\
\vdots\\
\vdash \Sigma
\end{array}$$

Adesso, riducendo come indicato in figura, potremmo trasportare γ all'interno di δ, mentre riducendo prima il taglio superiore era impossibile! Questa sembra essere una caratteristica ineliminabile della versione non deterministica di **LK** che stiamo studiando.

Non è chiaro se questo aspetto sia essenziale o possa invece essere considerato un difetto. Se volessimo correggerlo, ancora una volta il λ-calcolo simmetrico verrebbe in nostro aiuto. Ricordiamo il ruolo della regola (λ) nel sistema con *stoup* a cui abbiamo accennato nel paragrafo 3.4.3 del capitolo 3:

$$\frac{\vdash \Gamma, A;}{\vdash \Gamma; A}\,(\lambda)$$

Dal punto di vista della derivabilità, questa regola permette di riportare una formula nello *stoup* (inizialmente vuoto), trasformando una occorrenza *non principale* in una conclusione *principale*. Dal punto di vista della dinamica delle riduzioni, questa regola permette di arrestare temporaneamente la "risalita" di un passo strutturale (dato che la sua conclusione è principale); quando la riduzione riprende a trasportare, la regola sarà eliminata (come accade alla λ-astrazione in $\lambda^{\mathrm{Sym}}_{\mathrm{Prop}}$). Nulla impedisce di dimenticare lo stoup, imponendo che la regola sia applicata a una occorrenza *non principale* trasformandola in una occorrenza *principale*:

$$\frac{\vdash \Gamma, A}{\vdash \Gamma, A}\,(\lambda)$$

chiederemo poi che ogni formula attiva sia anche principale (dunque che abbia interspazio piatto): potremo inserire occorrenze della regola (λ) (il nome è certamente discutibile) dove necessario. A questo punto il caso presentato prima prenderà la forma seguente:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\ \gamma\\ \vdash \Gamma, A^x\end{array}}{\vdash \Gamma, A^x}\,(\lambda) \qquad \dfrac{}{\vdash A^y, \neg A^z}\,(ax)}{\begin{array}{c}\vdash \Gamma, A^y\\ \vdots\\ \vdash \Sigma\end{array}}\,(cut)$$

e le due scelte possibili nella riduzione del taglio avranno effetti molto diversi:

$$\begin{array}{c}\vdots\ \gamma[ax(\neg A)/\vdash A^x]\\ \vdash \Gamma, A^y\\ \vdots\\ \vdash \Sigma\end{array} \qquad\qquad \begin{array}{c}\dfrac{\begin{array}{c}\vdots\ \gamma\\ \vdash \Gamma, A^x\end{array}}{\vdash \Gamma, A^x}\,(\lambda)\\ \vdots\\ \vdash \Sigma\end{array}$$

nel primo caso abbiamo trasportato l'assioma all'interno di γ eliminando la regola (λ), nel secondo caso invece abbiamo sostituito l'assioma *e preservato la regola* (λ); una riduzione successiva si arresterà all'altezza di quest'ultima regola permettendo di raggiungere più ridotti:

$$\dfrac{\dfrac{\dfrac{\vdots\ \gamma}{\vdash \Gamma, A^x}}{\vdash \Gamma, A^x}\,(\lambda) \qquad \dfrac{\vdots\ \delta}{\vdash \Delta, \neg A^w}}{\dfrac{\vdash \Gamma, A^y}{\dfrac{\vdots}{\vdash \Sigma}}}\,(cut)$$

L'effetto della regola, dunque, è di rendere evidente nella struttura della derivazione (e nella dinamica della riduzione) il compimento di una scelta che altrimenti sarebbe rimasta implicita.

Abbiamo estratto questa regola direttamente dal λ-calcolo simmetrico; non è chiaro se si tratti di una soluzione *ad hoc* o se goda di qualche significato più profondo come regola strutturale. Sicuramente non è una soluzione priva di difetti (ad esempio non è chiaro come debba essere applicata nel caso della disgiunzione moltiplicativa). D'altra parte l'utilità di una regola che permetta di "controllare" più finemente i processi di riduzione è comprovata (pensiamo ad esempio agli esponenziali in logica lineare). Osserviamo inoltre che la regola per sua natura indica le posizioni in cui è necessario aggiungere scatole (cioè sequenzializzare) in un *proof-net* per **LK** sul tipo di quelli ipotizzati da Girard in appendice al suo [Gir91].

5.4 Alternative ai candidati simmetrici

Concludiamo il capitolo discutendo una possibile alternativa alla definizione di riducibilità adottata nel capitolo 4. È possibile avvicinarsi alla definizione di Girard in [Gir87]? Parliamo dell'idea di candidato come insieme di derivazioni chiuso per biortogonale, sfruttando una definizione molto generale di ortogonalità fra insiemi di derivazioni.

A tale proposito, sarebbe naturale definire l'operatore ortogonale utilizzando esclusivamente l'operatore Subst:

$$X^{\perp} = \{ax(\neg A)\} \cup \mathrm{Subst}(X)$$

dove $X \subseteq \mathcal{D}_A$ (la presenza dell'assioma è necessaria per garantire che $X^{\perp\perp}$ sia incluso in SN_d). Si potrebbe pensare quindi di seguire l'approccio di Girard in [Gir87] e definire per esempio:

$$\mathrm{RED}(\neg A \vee_m \neg B) = (\{ax(A \wedge_m B)\} \cup \wedge_m(\mathrm{RED}(A), \mathrm{RED}(B)))^{\perp},$$
$$\mathrm{RED}(A \wedge_m B) = \mathrm{RED}(\neg A \vee_m \neg B)^{\perp}.$$

Purtroppo con questo approccio non si riesce a soddisfare la definizione 4.12 di *candidato simmetrico*. Una proprietà chiave della definizione di ortogonale data da Girard è che in generale vale $X \subseteq X^{\perp\perp}$: qui si riesce (forse) a dimostrare che

$$\wedge_m(\mathrm{RED}(A), \mathrm{RED}(B)) \subseteq \mathrm{RED}(A \wedge_m B),$$

e allora avremo

$$\mathrm{RED}(A \wedge_m B)^{\perp} \subseteq \mathrm{RED}(\neg A \vee_m \neg B)$$

soddisfacendo così i punti (i) e (ii) della definizione 4.12. Ma non è affatto chiaro se si possa mostrare il viceversa:

$$\mathrm{RED}(\neg A \vee_m \neg B) \subseteq \mathrm{RED}(A \wedge_m B)^{\perp}.$$

Questo però è un passaggio critico nella dimostrazione del punto (iii). Consideriamo il taglio seguente:

$$\frac{\pi_1 \vdash \Gamma, A \wedge_m B \qquad \pi_2 \vdash \Delta, \neg A \vee_m \neg B}{\vdash \Gamma, \Delta}\,(cut)$$

dove $\pi_1 \in \mathrm{RED}(A \wedge_m B)$ ma $\pi_1 \notin \wedge_m(\mathrm{RED}(A), \mathrm{RED}(B))$, e d'altra parte $\pi_2 \in \mathrm{RED}(A \vee_m B)$ ma $\pi_2 \notin \mathrm{RED}(A \wedge_m B)^{\perp}$. In assenza di una dimostrazione dell'inclusione menzionata sopra, dobbiamo ipotizzare che questa situazione sia possibile: vogliamo mostrare che il taglio è fortemente normalizzante, ma quando π_2 non termina con una regola $(\vee_m)$ che introduce $\neg A \vee_m \neg B$, uno dei ridotti possibili sarà $\pi_2[\pi_1/\vdash \neg A \vee_m \neg B]$ e non avremo elementi per concludere che quest'ultima derivazione sia in SN_d.

5.4.1 Utilizzando il punto fisso?

Una possibile alternativa sarebbe definire $\mathrm{RED}(A \wedge_m B)$ come un qualsiasi punto fisso del biortogonale che includa il sottoinsieme $\wedge_m(\mathrm{RED}(A), \mathrm{RED}(B))$. Questo è certamente possibile: il biortogonale per derivazioni di tipo $A \wedge_m B$ è un operatore crescente sull'insieme $\mathcal{P}(\mathcal{D}_{A\wedge_m B})$ ordinato per inclusione; inoltre il sottoinsieme di $\mathcal{P}(\mathcal{D}_{A\wedge_m B})$ costituito da tutti gli insiemi che includono $\wedge_m(\mathrm{RED}(A), \mathrm{RED}(B))$ è chiaramente chiuso sotto le operazioni di unione e intersezione, che forniscono gli estremi superiori e inferiori. Possiamo dunque applicare il teorema di Tarski per ottenere il punto fisso cercato; porremo quindi $\mathrm{RED}(\neg A \vee_m \neg B) = \mathrm{RED}(A \wedge_m B)^{\perp}$. Il problema in tal caso è dimostrare:

1. che $\vee_m(\mathrm{RED}(A,B)) \subseteq \mathrm{RED}(A \vee_m B) = \mathrm{RED}(\neg A \wedge_m \neg B)^{\perp}$: in altre parole, che i candidati definiti come ortogonali dei punti fissi soddisfino le condizioni sufficienti necessarie per i loro connettivi;

2. che la definizione così data (X è un candidato se $X = X^{\perp\perp}$) sia forte almeno quanto la definizione 4.12: in altre parole, che le condizioni di riducibilità

siano soddisfatte. I punti (i) e (ii) non creano difficoltà (seguono dall'inclusione di $\mathrm{Subst}(X^{\perp})$ in X). Invece il punto (iii) è più delicato. Bisogna mostrare che, per qualsiasi $\gamma \in X$ e $\delta \in X^{\perp}$, vale

$$\frac{\gamma \vdash \Gamma, A \qquad \delta \vdash \Delta, \neg A}{\vdash \Gamma, \Delta}\,(cut) \in \mathrm{SN}_d.$$

La conclusione è immediata quando qualche formula di taglio è principale (grazie al fatto che $\gamma \in \mathrm{Subst}(X^{\perp})$ e $\delta \in \mathrm{Subst}(X)$). Non è chiaro invece se si possa concludere nel caso di un taglio $S1$: quasi certamente si dovrà utilizzare la proposizione 4.9, ma come impostare l'induzione?

Se l'operazione riuscisse, si potrebbero eliminare tutte le proposizioni che seguono la definizione dei candidati e, come in [Gir87], passare subito al lemma 4.24 di riducibilità. In assenza di queste possibilità, dobbiamo seguire [BB96] e includere i sottoinsiemi di nostro interesse direttamente nella definizione di ortogonale.

Bibliografia

[Ari17] Aristotele. *Metafisica*. Cur. e trad. da Enrico Berti. Biblioteca Filosofica Laterza. GLF Editori Laterza, 2017.

[AT14] Vito Michele Abrusci e Lorenzo Tortora de Falco. *Logica: Volume 1 - Dimostrazioni e modelli al primo ordine*. UNITEXT. Springer Milan, 2014.

[AT18] Vito Michele Abrusci e Lorenzo Tortora de Falco. *Logica: Volume 2 - Incompletezza, teoria assiomatica degli insiemi*. UNITEXT. Springer Milan, 2018.

[BB96] Franco Barbanera e Stefano Berardi. «A Symmetric Lambda-Calculus for Classical Program Extraction». In: *Information and Computation* 125.2 (1996), pp. 103–117.

[BBS97] Franco Barbanera, Stefano Berardi e Michele Schivalocchi. «"Classical" programming-with-proofs in λ_{PA}^{Sym}:an analysis of non-confluence». In: *Proceedings of TACS'97*. Vol. 1281. Lecture Notes in Computer Science. Springer-Verlag, 1997.

[CF58] Haskell B. Curry e Robert Feys. *Combinatory Logic*. Combinatory Logic v. 1. North-Holland Publishing Company, 1958.

[DJS97] Vincent Danos, Jean-Baptiste Joinet e Harold Schellinx. «A New Deconstructive Logic: Linear Logic». In: *The Journal of Symbolic Logic* 62.3 (1997), pp. 755–807.

[Dum58] Michael Dummett. «Truth». In: *Proceedings of the Aristotelian Society* 59 (1958), pp. 141–162.

[Dum91] Michael Dummett. *The Logical Basis of Metaphysics*. The William James lectures delivered at Harvard University. Harvard University Press, 1991.

[Gen35] Gerhard Gentzen. «Untersuchungen über das logische Schließen. I». In: *Mathematische Zeitschrift* 39.1 (dic. 1935), pp. 176–210.

[Gir08] Jean-Yves Girard. «The phantom of transparency». Testo di un intervento, disponibile online: `https://girard.perso.math.cnrs.fr/longo1.pdf`. Feb. 2008.

[Gir87] Jean-Yves Girard. «Linear logic». In: *Theoretical Computer Science* 50.1 (1987), pp. 1–101.

[Gir91] Jean-Yves Girard. «A new constructive logic: classical logic». In: *Mathematical Structures in Computer Science* 1.3 (1991), pp. 255–296.

[GLT89] Jean-Yves Girard, Yves Lafont e Paul Taylor. *Proofs and Types*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1989.

[Gri90] Timothy G. Griffin. «A Formulae-as-Types Notion of Control». In: *Proceedings of the 17th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*. POPL '90. San Francisco, California, USA: Association for Computing Machinery, 1990, pp. 47–58.

[How80] William A Howard. «The formulae-as-types notion of construction». In: *To HB Curry: Essays on combinatory logic, lambda calculus and formalism*. A cura di J. Robert Hindley e J. P. Seldin. Academic Press, 1980.

[Iem20] Rosalie Iemhoff. «Intuitionism in the Philosophy of Mathematics». In: *The Stanford Encyclopedia of Philosophy*. A cura di Edward N. Zalta. Summer 2020. Metaphysics Research Lab, Stanford University, 2020.

[JST96] Jean-Baptiste Joinet, Harold Schellinx e Lorenzo Tortora de Falco. «Strong normalization for all-style LKtq». In: *Theorem Proving with Analytic Tableaux and Related Methods*. A cura di P. Miglioli et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 226–243.

[Kle52] Stephen Cole Kleene. *Introduction to Metamathematics*. Bibliotheca Mathematica. Wolters-Noordhoff, 1952.

[Kre51] G. Kreisel. «On the Interpretation of Non-Finitist Proofs–Part I». In: *The Journal of Symbolic Logic* 16.4 (1951), pp. 241–267.

[Lai01] James Laird. «A Deconstruction of Non-deterministic Classical Cut Elimination». In: *Typed Lambda Calculi and Applications*. A cura di Samson Abramsky. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 268–282.

[Lak79] Imre Lakatos. *Dimostrazioni e confutazioni. La logica della scoperta matematica*. A cura di Giulio Giorello. Trad. da Daniela Benelli. Filosofia della scienza. Feltrinelli, 1979.

[MB93] Corrado Mangione e Silvio Bozzi. *Storia della logica: da Boole ai nostri giorni*. Saggi blu. Garzanti, 1993.

[Nat14] Carlo Natali. *Aristotele*. Pensatori (Carocci). Carocci editore, 2014.

[Par00] Michel Parigot. «On the Computational Interpretation of Negation». In: *Computer Science Logic*. A cura di Peter G. Clote e Helmut Schwichtenberg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 472–484.

[Par92a] Michel Parigot. «Free deduction: An analysis of "Computations" in classical logic». In: *Logic Programming*. A cura di Andrei Voronkov. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, pp. 361–380.

[Par92b] Michel Parigot. «λμ-Calculus: An algorithmic interpretation of classical natural deduction». In: *Logic Programming and Automated Reasoning*. A cura di Andrei Voronkov. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, pp. 190–201.

[Pis15] Paolo Pistone. «On proofs and types in second order logic». Tesi di dott. Università degli studi Roma Tre, 2015.

[Pop47] K. R. Popper. «Logic without Assumptions». In: *Proceedings of the Aristotelian Society* 47 (1947), pp. 251–292.

[Pra65] Dag Prawitz. *Natural Deduction: A Proof-theoretical Study*. Acta Universitatis Stockholmiensis. Almqvist & Wiksell, 1965.

[Pra81] Dag Prawitz. «Validity and normalizability of proofs in first and second order classical and intuitionistic logic». In: *Atti del Congresso nazionale di logica: Montecatini Terme, 1-5 ottobre 1979*. A cura di Sergio Bernini. Bibliopolis Napoli, 1981, pp. 11–36.

[Pri60] A. N. Prior. «The Runabout Inference-Ticket». In: *Analysis* 21.2 (dic. 1960), pp. 38–39. eprint: `https://academic.oup.com/analysis/article-pdf/21/2/38/360684/21-2-38.pdf`.

[Rin90] Yann-Joachim Ringard. «Mustard Watches: an integrated approach to time and food». Pseudonimo di Jean-Yves Girard, articolo disponibile online: `https://girard.perso.math.cnrs.fr/mustard/article.html`. Ott. 1990.

[Sch18] Peter Schroeder-Heister. «Proof-Theoretic Semantics». In: *The Stanford Encyclopedia of Philosophy*. A cura di Edward N. Zalta. Spring 2018. Metaphysics Research Lab, Stanford University, 2018.

[Zei09] Noam Zeilberger. «The Logical Basis of Evaluation Order and Pattern-Matching». Tesi di dott. USA: Carnegie Mellon University, 2009.