PPA

Un asistente de demostración para lógica de primer orden con extracción de testigos usando la traducción de Friedman

Manuel Panichelli

Deparatamento de Computación, FCEyN, UBA

Diciembre 2024

Introducción

Asistentes de demostración

- Los **asistentes de demostración** son herramientas que facilitan la escritura y el chequeo de demostraciones por computadora.
- Usos usuales: formalización de teoremas matemáticos y verificación de programas.
- Ventajas:[1]
  - Facilitan la colaboración a gran escala (mediante la confianza en el asistente).
  - Habilitan generación automática de demostraciones con IA. Por ej. un *LLM* (como *ChatGPT*) suele devolver alucinaciones, que pueden ser filtradas automáticamente con un asistente.

[1] Terrence Tao - Machine Assisted Proof

Asistentes de demostración

Implementan distintas *teorías*. (TODO: No me gusta teoria, se usa para teorias de primer orden) Ejemplos:

- Mizar (lógica de primer orden)
- Coq (teoría de tipos)
- Agda (teoría de tipos)
- Isabelle (lógica de orden superior / teoría de conjuntos ZF)

PPA

- Diseñamos e implementamos **PPA** (*Pani's Proof Assistant*), un asistente de demostración para lógica **clásica** de primer orden.
- Permite hacer extracción de testigos: dada una demostración de $\exists x.p(x)$, encuentra t tal que $p(t)$.
- Aporte principal: implementación de extracción de testigos para lógica clásica de forma directa (vemos los detalles después).

Representación de demostraciones

Queremos escribir demostraciones en la computadora. ¿Cómo las representamos?. Veamos un ejemplo.

- Tenemos dos premisas
  1. Los alumnos que faltan a los exámenes, los reprueban.
  2. Si se reprueba un final, se recursa la materia.
- A partir de ellas, podríamos demostrar que si un alumno falta a un final, entonces recursa la materia.

Teorema

Si ((falta entonces reprueba) y (reprueba entonces recursa)) y falta, entonces recursa

Demostración

- Asumo que falta. Quiero ver que recursa.
- Sabemos que si falta, entonces reprueba. Por lo tanto reprobó.
- Sabemos que si reprueba, entonces recursa. Por lo tanto recursó. □

Sistemas deductivos

- La demostración anterior es poco precisa. No se puede representar rigurosamente.
- Necesitamos **sistemas deductivos**: sistemas lógicos formales usados para demostrar setencias. Pueden ser representados como un tipo abstracto de datos.
- Usamos **deducción natural**. Compuesto por,
  - **Lenguaje formal**: lógica de primer orden.
  - **Reglas de inferencia**: lista de reglas que se usan para probar teoremas a partir de axiomas y otros teoremas. Por ejemplo, *modus ponens* (si es cierto $A \rightarrow B$ y A, se puede concluir B) o *modus tollens* (si es cierto $A \rightarrow B$ y $\neg B$, se puede concluir $\neg A$)
  - **Axiomas**: fórmulas de L que se asumen válidas. Todos los teoremas se derivan de axiomas. Se usan para modelar *teorías* de primer orden (por ej. teoría de estudiantes en la facultad).

Lógica de primer orden

Definición (Términos)

Los términos están dados por la gramática:

$$
\begin{aligned}
t ::= {} & x && \text{(variables)} \\
& \mid f(t_1, \ldots, t_n) && \text{(funciones)}
\end{aligned}
$$

Definición (Fórmulas)

Las fórmulas están dadas por la gramática:

$$
\begin{aligned}
A, B ::= {} & p(t_1, \ldots, t_n) && \text{(predicados)} \\
& \mid \bot \mid \top && \text{(falso o } \textit{bottom} \text{ y verdadero o } \textit{top}\text{)} \\
& \mid A \wedge B \mid A \vee B && \text{(conjunción y disyunción)} \\
& \mid A \to B \mid \neg A && \text{(implicación y negación)} \\
& \mid \forall x.A \mid \exists x.A && \text{(cuantificador universal y existencial)}
\end{aligned}
$$

Los predicados son **fórmulas atómicas**. Los de aridad 0 además son llamados *variables proposicionales*.

Notación

Usamos

- $x, y, z, \ldots$ como **variables**.
- $f, g, h, \ldots$ como **símbolos de función**.
- $p, q, r, \ldots$ como **símbolos de predicado**.
- $t, u, \ldots$ para referirnos a **términos**.
- $a, b, c, \ldots, A, B, C, \ldots$ y $\varphi, \psi, \ldots$ para referirnos a **fórmulas**.

Deducción natural

Deducción natural

Definiciones

- Γ es un **contexto de demostración**, conjunto de fórmulas que se asumen válidas
- Notación: $\Gamma, \varphi = \Gamma \cup \{\varphi\}$
- $\vdash$ es la **relación de derivabilidad** definida a partir de las *reglas de inferencia*. Permite escribir juicios $\Gamma \vdash \varphi$.
- Intuición: *"φ es una consecuencia de las suposiciones de Γ"*
- El juicio es cierto si en una cantidad finita de pasos podemos concluir φ a partir de las fórmulas de Γ, los axiomas y las reglas de inferencia.
- Decimos que φ es *derivable* a partir de Γ.

Reglas de inferencia

Definición (Reglas de inferencia)

$$\frac{}{\Gamma, A \vdash A}\ \mathrm{Ax} \qquad \frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash A \wedge B}\ \mathrm{I}\wedge$$

$$\frac{\Gamma, A \vdash B}{\Gamma \vdash A \to B}\ \mathrm{I}{\to} \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash A}\ \mathrm{E}\wedge_1$$

$$\frac{\Gamma \vdash A \to B \quad \Gamma \vdash A}{\Gamma \vdash B}\ \mathrm{E}{\to} \qquad \frac{\Gamma \vdash A \wedge B}{\Gamma \vdash B}\ \mathrm{E}\wedge_2$$

Dos tipos para cada conectivo y cuantificador, dada una fórmula formada con un conectivo:

- **Introducción**: ¿Cómo la demuestro?
- **Eliminación**: ¿Cómo la uso para demostrar otra?

Ejemplo

Vamos a demostrar el ejemplo informal en deducción natural. Lo modelamos para un alumno y materia particulares. Notamos:

- $X \equiv \mathrm{reprueba}(\mathit{juan}, \mathrm{final}(\mathit{logica}))$
- $R \equiv \mathrm{recursa}(\mathit{juan}, \mathit{logica})$
- $F \equiv \mathrm{falta}(\mathit{juan}, \mathrm{final}(\mathit{logica}))$

Queremos probar entonces

$$\Big((F \to X) \land (X \to R)\Big) \to (F \to R)$$

Ejemplo

$$
\dfrac{\dfrac{\dfrac{\dfrac{}{\Gamma \vdash (F \to X) \wedge (X \to R)}\,\mathrm{Ax}}{\Gamma \vdash X \to R}\,\mathrm{E}\wedge_1 \qquad \begin{array}{c}\Pi \\ \Gamma \vdash X\end{array}}{\Gamma = (F \to X) \wedge (X \to R),\; F \vdash R}\,\mathrm{E}\!\to}{\dfrac{(F \to X) \wedge (X \to R) \vdash F \to R}{\vdash \Big((F \to X) \wedge (X \to R)\Big) \to (F \to R)}\,\mathrm{I}\!\to}\,\mathrm{I}\!\to
$$

donde

$$
\Pi = \dfrac{\dfrac{\dfrac{}{\Gamma \vdash (F \to X) \wedge (X \to R)}\,\mathrm{Ax}}{\Gamma \vdash F \to X}\,\mathrm{E}\wedge_2 \qquad \dfrac{}{\Gamma \vdash F}\,\mathrm{Ax}}{\Gamma \vdash X}\,\mathrm{E}\!\to
$$

Reglas de inferencia

Definición (Reglas de inferencia)

$$\frac{}{\Gamma \vdash A \vee \neg A}\ \mathrm{LEM}$$

$$\frac{\Gamma \vdash \bot}{\Gamma \vdash A}\ \mathrm{E}\bot \qquad \frac{\Gamma, A \vdash \bot}{\Gamma \vdash \neg A}\ \mathrm{I}\neg$$

$$\frac{}{\Gamma \vdash \top}\ \mathrm{I}\top \qquad \frac{\Gamma \vdash \neg A \qquad \Gamma \vdash A}{\Gamma \vdash \bot}\ \mathrm{E}\neg$$

$$\frac{\Gamma \vdash A}{\Gamma \vdash A \vee B}\ \mathrm{I}\vee_1 \qquad \frac{\Gamma \vdash B}{\Gamma \vdash A \vee B}\ \mathrm{I}\vee_2$$

$$\frac{\Gamma \vdash A \vee B \qquad \Gamma, A \vdash C \qquad \Gamma, B \vdash C}{\Gamma \vdash C}\ \mathrm{E}\vee$$

Reglas de inferencia

Definición (Sustitución)

Notamos como $A\{x := t\}$ a la sustitución de todas las ocurrencias libres de la variable x por el término t en la fórmula A.

Definición (Reglas de cuantificadores)

$$\frac{\Gamma \vdash A \qquad x \notin fv(\Gamma)}{\Gamma \vdash \forall x.A}\ \mathrm{I}\forall \qquad\qquad \frac{\Gamma \vdash \forall x.A}{\Gamma \vdash A\{x := t\}}\ \mathrm{E}\forall$$

$$\frac{\Gamma \vdash A\{x := t\}}{\Gamma \vdash \exists x.A}\ \mathrm{I}\exists$$

$$\frac{\Gamma \vdash \exists x.A \qquad \Gamma, A \vdash B \qquad x \notin fv(\Gamma, B)}{\Gamma \vdash B}\ \mathrm{E}\exists$$

Reglas admisibles

- Mencionamos *modus tollens* pero no aparece en las reglas de inferencia.
- Queremos un sistema lógico **minimal**: no agregamos como regla de inferencia lo que podemos derivar a partir de las existentes, las reglas **admisibles**.
- Se implementan como *macros*: cada uso de la regla admisible se reemplaza por su demostración.

Lema (Modus tollens)

$$
\dfrac{
  \dfrac{\dfrac{}{\Gamma \vdash (A \to B) \wedge \neg B}\,Ax}{\Gamma \vdash \neg B}\,E\wedge_2
  \qquad
  \dfrac{
    \dfrac{\dfrac{}{\Gamma \vdash (A \to B) \wedge \neg B}\,Ax}{\Gamma \vdash A \to B}\,E\wedge_1
    \quad
    \dfrac{}{\Gamma \vdash A}\,Ax
  }{\Gamma \vdash B}\,E\to
}{
  \dfrac{
    \dfrac{\Gamma = (A \to B) \wedge \neg B, A \vdash \bot}{(A \to B) \wedge \neg B \vdash \neg A}\,I\neg
  }{\vdash (A \to B \wedge \neg B) \to \neg A}\,I\to
}\,E\neg
$$

Sustitución sin capturas

Para la sustitución $A\{x := t\}$ queremos evitar la **captura de variables**, por ejemplo

$$(\forall y.p(x))\{x := y\} \stackrel{?}{=} \forall y.p(y)$$

sustituyendo sin más, capturamos a la variable x que ahora está ligada. Lo evitamos **automáticamente**: cuando se encuentra con una captura, se renombra la variable ligada de forma que no ocurra

$$(\forall y.p(x))\{x := y\} = \forall z.p(y)$$

Alfa equivalencia

- Si tenemos una hipótesis $\exists x.p(x)$ queremos poder usarla para demostrar $\exists y.p(y)$.
- No son iguales, pero son α-**equivalentes**: si renombramos variables ligadas de forma apropiada, son iguales.
- Algoritmo naíf: cuadrático en la estructura de la fórmula, renombrando recursivamente.
- Algoritmo cuasilineal: manteniendo dos sustituciones, una por fórmula.

Ejemplo

$$
\begin{array}{rlr}
(\exists x.f(x)) \stackrel{\alpha}{=} (\exists y.f(y)) & & \{\}, \{\} \\
\iff & f(x) \stackrel{\alpha}{=} f(y) & \{x \mapsto z\}, \{y \mapsto z\} \\
\iff & x \stackrel{\alpha}{=} y & \{x \mapsto z\}, \{y \mapsto z\} \\
\iff & z = z. &
\end{array}
$$

PPA

Mathematical vernacular

Aparentemente hay una forma canónica de presentar demostraciones matemáticas[2]. Descubierta e implementada independientemente en Mizar, Isar (Isabelle), etc. Combinación de ideas:

- **Deducción natural en estilo de *Fitch***. Notación equivalente en la cual las demostraciones son representadas como listas de fórmulas en lugar de árboles. Las que aparecen antes justifican las que aparecen después.
- **Reglas de inferencia *declarativas***: una forma de afirmar que $A_1, \ldots, A_n \vdash A$ es válida, sin tener que demostrarlo a mano.
- **Sintaxis similar a un lenguaje de programación** en lugar del lenguaje natural usado para demostraciones.

[2] *Mathematical Vernacular* de Freek Wiedijk

PPA

Diseñamos e implementamos el *lenguaje* PPA, inspirado en el *mathematical vernacular*. Veamos la **interfaz** completa y luego la implementación.

Ejemplo demostración

```
axiom falta_reprueba: forall A . forall E .
   falta(A, E) -> reprueba(A, E)
axiom reprueba_recursa: forall A . forall M .
   reprueba(A, final(M)) -> recursa(A, M)

theorem falta_entonces_recursa: forall A . forall M .
   falta(A, final(M)) -> recursa(A, M)
proof
   let A
   let M
   suppose falta: falta(A, final(M))
   have reprueba: reprueba(A, final(M)) by falta, falta_reprueba
   thus recursa(A, M) by reprueba, reprueba_recursa
end
```

Programas

Un **programa** de PPA consiste en una lista de **declaraciones**, que pueden ser

- **Axiomas**: fórmulas que se asumen válidas

```
axiom <name> : <form>
```

- **Teoremas**: fórmulas junto con sus demostraciones.

```
theorem <name> : <form>
proof
  <steps>
end
```

Identificadores

- **Variables** (<var>)

```
(\_|[A-Z])[a-zA-Z0-9\_\-]*(\')*
```

- **Identificadores** (<id>)

```
[a-zA-Z0-9\_\-\?!#\$\%\*\+\<\>\=\?\@\^]+(\')*
```

- **Nombres** (<name>)
  Pueden ser identificadores o strings arbitrarios encerrados por comillas dobles.

```
<id> | \"[^\"]*\"
```

Fórmulas y términos

Términos:

- Variables: `<var>`
- Funciones: `<id>(<term>, ..., <term>)`

Funciones:

- Predicados: `<id>(<term>, ..., <term>)`
- `<form> & <form>`
- `<form> | <form>`
- `<form> -> <form>`
- `<form> <-> <form>`
- `~ <form>`
- **`exists`** `<var> . <form>`
- **`forall`** `<var> . <form>`
- **`true`**, **`false`**
- `(<form>)`

Demostraciones

- Lista de comandos que reducen sucesivamente la *tesis* (fórmula a demostrar) hasta agotarla por completo.
- Corresponden aproximadamente a reglas de inferencia de deducción natural (vistas como una demostración en el estilo de Fitch).
- Tienen disponible un **contexto** con todas las hipótesis asumidas (como axiomas) o demostradas (teoremas y comandos que demuestran hipótesis auxiliares).

by - El mecanismo principal de demostración

- `<form>` **by** `<h1>, ..., <hn>` afirma que la fórmula es una consecuencia lógica de las fórmulas que corresponden a las hipótesis provistas.
- Los nombres de las hipótesis son del tipo `<name>` (o bien identificadores o *strings* arbitrarios).
- Por debajo usa un *solver* completo para lógica proposicional pero heurístico para primer órden.
- Se usa para eliminar implicaciones y universales.
- Usado por dos comandos principales: **`thus`** y **`have`**.

Thus

```
thus <form> by <h1>, ..., <hn>
```

Si `<form>` es *parte* de la tesis, y el *solver* puede demostrar la implicación, lo demuestra automáticamente y lo descarga de la tesis.

Eliminación de implicación

```
axiom ax1: a -> b
axiom ax2: b -> c

theorem t1: a -> c
proof
   suppose a: a

   // La tesis ahora es c
   thus c by a, ax1, ax2
end
```

Eliminación de universal

```
axiom ax: forall X . f(X)

theorem t: f(n)
proof
   thus f(n) by ax
end
```

Have

```
have <name>: <form> by <h1>, ..., <hn>
```

Análogo a **`thus`**, pero introduce una afirmación *auxiliar* sin reducir la tesis, agregándola al contexto.

Eliminación de implicación en dos pasos

```
axiom ax1: a -> b
axiom ax2: b -> c

theorem t1: a -> c
proof
    suppose a: a
    have b: b by a, ax1
    thus c by b, ax2
end
```

Hipótesis anterior

Ambas pueden referirse a la hipótesis anterior con guión medio (-), y pueden hacerlo implícitamente usando **hence** y **have**.

Comando	Alternativo	¿Reduce la tesis?
thus	**hence**	Sí
have	**then**	No

Eliminación en dos pasos

```
axiom ax1: a -> b
axiom ax2: b -> c

theorem t1: a -> c
proof
   suppose a: a
   have b: b by a, ax1
   thus c by b, ax2
end
```

Alternativas equivalentes

```
proof
   suppose a: a
   have b: b by -, ax1
   thus c by -, ax2
end
proof
   suppose -: a
   then -: b by ax1
   hence c by ax2
end
```

By opcional

- El **by** es opcional
- Si se omite, la fórmula debe ser demostrable por el *solver* sin partir de ninguna hipótesis
- Vale para todas las tautologías proposicionales.

Tautología proposicional

```
theorem "distributiva de negación sobre disyunción":
    ~(a | b) <-> ~a & ~b
proof
    thus ~(a | b) <-> ~a & ~b
end
```

Comandos y reglas de inferencia

Regla	Comando
LEM	**`cases`**
Ax	**`by`**
$I\exists$	**`take`**
$E\exists$	**`consider`**
$I\forall$	**`let`**
$E\forall$	**`by`**
$I\vee_1$	**`by`**
$I\vee_2$	**`by`**
$E\vee$	**`cases`**

Regla	Comando
$I\wedge$	**`by`**
$E\wedge_1$	**`by`**
$E\wedge_2$	**`by`**
$I\rightarrow$	**`suppose`**
$E\rightarrow$	**`by`**
$I\neg$	**`suppose`**
$E\neg$	**`by`**
$I\top$	**`by`**
$E\bot$	**`by`**

Suppose (I→/ I¬)

`suppose <name>: <form>` (I→/ I¬)

- Si la tesis es una implicación $A \to B$, agrega el antecedente A como hipótesis con el nombre dado y reduce la tesis al consecuente B
- Viendo la negación como una implicación $\neg A \equiv A \to \bot$, permite introducir negaciones, tomando $B = \bot$.

Introducción de implicación

```
theorem "suppose":
    a -> (a -> b) -> b
proof
    suppose h1: a
    suppose h2: a -> b
    thus b by h1, h2
end
```

Introducción de negación

```
theorem "not intro":
    ~b & (a -> b) -> ~a
proof
    suppose h: ~b & (a -> b)
    suppose a: a
    hence false by h, a
end
```

Cases (E$\vee$)

cases by <h1>, ..., <hn> $(\mathrm{I}\to / \ \mathrm{I}\neg)$

- Permite razonar por casos a partir de una disyunción. Para cada uno, se debe demostrar la tesis en su totalidad.
- Si los casos son <f1> a <fn>, tiene que valer <f1> | ... | <fn> **by** <h1>, ..., <hn>.
- Se puede omitir el **by** para razonar mediante LEM (casos φ y $\neg\varphi$).

Cases

```
theorem "cases":
    (a & b) | (c & a) -> a
proof
    suppose h: (a & b) | (c & a)
    cases by h
        case a & b
            hence a
        case right: a & c
            thus a by right
    end
end
```

Take (I∃)

`take <var> := <term>`   (I∃)

- Introduce un existencial instanciando su variable y reemplazándola por un término.
- Si la tesis es `exists X . p(X)`, luego de `take X := a`, se reduce a `p(a)`.

Consider (E∃)

`consider <var> st <name>: <form> by <h1>, ..., <hn>` (E∃)

- Si se puede justificar **exists** `X. p(X)`, permite razonar sobre tal `X`.
- Agrega `<form>` como hipótesis al contexto, con nombre `<name>`. No reduce la tesis.
- Debe valer `exists <var> . <form> by <h1>, ..., <hn>`
- Permite α-equivalencias: Si podemos justificar **exists** `X. p(X)`, podemos usarlo como `consider Y st h: p(Y) by ...`.

Let (I∀)

`let` `<var>` (I∀)

- Permite demostrar un cuantificador universal.
- Si la tesis es **`forall`** `X . p(X)`, luego de **`let`** `X`, la tesis se reduce a `p(X)`.
- Permite renombrar la variable, por ejemplo luego de **`let`** `Y` la tesis se reduce a `p(Y)`.

Descarga de conjunciones

Si la tesis es una conjunción, se puede probar un subconjunto de ella y se reduce el resto.

Descarga simple

```
theorem "and discharge":
    a -> b -> (a & b)
proof
    suppose "a" : a
    suppose "b" : b
    // La tesis es a & b
    hence b by "b"

    // La tesis es a
    thus a by "a"
end
```

Descarga compleja

```
axiom "a": a
axiom "b": b
axiom "c": c
axiom "d": d
axiom "e": e
theorem "and discharge":
    (a & b) & ((c & d) & e)
proof
    thus a & e by "a", "e"
    thus d by "d"
    thus b & c by "b", "c"
end
```

Equivalently

```
equivalently <form>
```

- Permite reducir la tesis a una fórmula equivalente
- Se puede usar por ejemplo para descarga de conjunciones, o para razonar por el absurdo mediante la eliminación de la doble negación.

Descarga de conjunción

```
axiom a1: ~a
axiom a2: ~b

theorem "ejemplo" : ~(a | b)
proof
   equivalently ~a & ~b
   thus ~a by a1
   thus ~b by a2
end
```

Razonamiento por el absurdo

```
theorem t: <form>
proof
   equivalently ~~<form>
   suppose <name>: ~<form>
   // Demostración de <form>
   // por el absurdo,
   // asumiendo ~<form>
   // y llegando a una
   // contradicción (false).
end
```

Claim

```
claim <name>: <form>
```

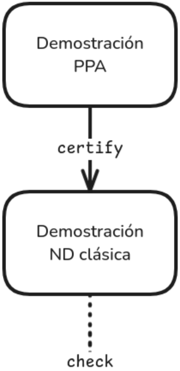
Permite demostrar una afirmación auxiliar. Útil para ordenar las demostraciones sin tener que definir otro teorema.

```
theorem t: <form1>
proof
   claim <name>: <form2>
   proof
      // Demostración de <form2>.
   end
   // Demostración de <form1> refiriéndose a <name>.
end
```

Certificador

Certificados

- Las demostraciones de PPA se *certifican* generando una demostración de deducción natural.
- No deberían generarse demostraciones erróneas, pero son chequeadas independientemente como mecanismo de *fallback*.



Criterio de de Bruijn

Un asistente de demostración cumple con el criterio de de Bruijn si satisface que sus demostraciones puedan ser chequeadas por un programa independiente, pequeño y confiable.

Contexto global

Se generan N demostraciones de deducción natural para cada programa, y se guardan en el *contexto global*. El chequeo se extiende a contextos.

```
axiom ax1: q
axiom ax2: q -> p
axiom ax3: p -> r

theorem t1: p
proof
    thus p by ax1, ax2
end

theorem t2: r
proof
    thus r by t1, ax3
end
```

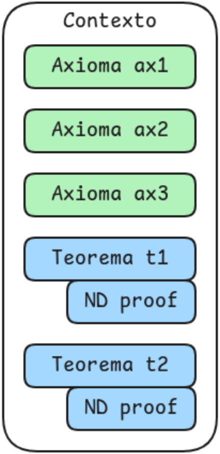


Figura: Contexto resultante de certificar un programa

Certificado de demostraciones

El certificado de una demostración es recursivo: se certifica cada comando, generando una demostración en deducción natural cuyas premisas son el certificado del resto de la demostración en PPA.

```
theorem t:
    p(v) -> exists X . p(X)
proof
    suppose h: p(v)
    take X := v
    thus p(v) by h
end
```

$$\dfrac{\dfrac{\dfrac{}{h : p(v) \vdash p(v)}\,\mathsf{Ax}_h}{h : p(v) \vdash \exists x.p(X)}\,\mathsf{I}\exists}{\vdash p(v) \to \exists x.p(X)}\,\mathsf{I}{\to}_h$$

Figura: Ejemplo de certificado generado para un programa

Contexto local

Cada demostración tiene un contexto local a ella con las hipótesis agregadas por ciertos comandos (**suppose**, **consider**, **have**, **claim**, etc.). Necesaria para obtener las fórmulas asociadas a las hipótesis en el **by**.

```
axiom ax1: p -> q
theorem t: (q -> r) -> p -> r
proof
    suppose h1: (q -> r)
    suppose h2: p
    then tq: q by ax1
    hence r by h1
end
```

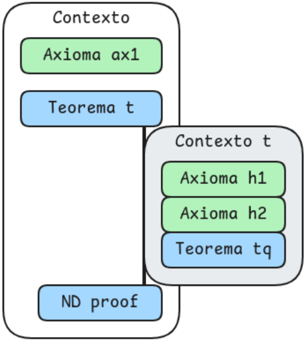


Figura: Ejemplo de contexto local

Certificado del by

Teniendo $\Gamma = \{h_1 : B_1, \ldots, h_n : B_n\}$, para certificar

`thus A by h1, ..., hn:`

1. Buscamos las hipótesis en el contexto. Queremos demostrar
$$B_1 \wedge \ldots \wedge B_n \rightarrow A$$
2. **Razonamos por el absurdo**: Asumiendo la negación buscamos una contradicción
$$\begin{aligned} \neg(B_1 \wedge \ldots \wedge B_n \rightarrow A) &\equiv \neg(\neg(B_1 \wedge \ldots \wedge B_n) \vee A) \\ &\equiv B_1 \wedge \ldots \wedge B_n \wedge \neg A \end{aligned}$$
3. Convertimos la negación a forma normal disyuntiva (**DNF**)
$$(a_1 \wedge \ldots \wedge a_n) \vee \ldots \vee (b_1 \wedge \ldots \wedge b_m)$$
4. Buscamos una **contradicción** refutando cada cláusula individualmente. Será refutable si
    - Contiene $\perp$ o dos fórmulas opuestas $(a, \neg a)$,
    - Eliminando existenciales consecutivos y reiniciando el proceso, se consigue una refutación $(\neg p(k), \forall x.p(x))$

Ejemplo sin cuantificadores (1/2)

Tenemos el siguiente programa

```
axiom ax1: a -> b
axiom ax2: a
theorem t: b
proof
   thus b by ax1, ax2
end
```

1. Para certificar **thus** b **by** ax1, ax2 hay que generar una demostración para la implicación

$$\big((a \to b) \wedge a\big) \to b$$

2. Negamos la fórmula

$$\neg\big[\big((a \to b) \wedge a\big) \to b\big]$$

Ejemplo sin cuantificadores (2/2)

3. La convertimos a DNF

$$
\begin{aligned}
&\neg[((a \to b) \wedge a) \to b] \\
&\equiv \neg[\neg((a \to b) \wedge a) \vee b] && (A \to B \equiv \neg A \vee B) \\
&\equiv \neg\neg((a \to b) \wedge a) \wedge \neg b && (\neg(A \vee B) \equiv \neg A \wedge \neg B) \\
&\equiv ((a \to b) \wedge a) \wedge \neg b && (\neg\neg A \equiv A) \\
&\equiv (\neg a \vee b) \wedge a \wedge \neg b && (A \to B \equiv \neg A \vee B) \\
&\equiv (\neg a \vee b) \wedge a \wedge \neg b && ((A \vee B) \wedge C \equiv (A \wedge C) \vee (B \wedge C)) \\
&\equiv (\neg a \wedge a \wedge \neg b) \vee \\
&\quad (b \wedge a \wedge \neg b)
\end{aligned}
$$

4. Refutamos cada cláusula

$$(\neg a \wedge a \wedge \neg b) \vee (b \wedge a \wedge \neg b)$$

Ejemplo con cuantificadores (1/3)

Tenemos el siguiente programa

```
axiom ax1: forall X . p(X) -> q(X)
axiom ax2: p(a)
theorem t: q(a)
proof
   thus q(a) by ax1, ax2
end
```

1. Para certificar **`thus`** `q(a)` **`by`** `ax1, ax2` hay que generar una demostración para la implicación

$$\Big((\forall x.(p(x) \rightarrow q(x))) \land p(a)\Big) \rightarrow q(a)$$

2. Negamos la fórmula

$$\neg\left[\Big((\forall x.(p(x) \rightarrow q(x))) \land p(a)\Big) \rightarrow q(a)\right]$$

Ejemplo con cuantificadores (2/3)

3. La convertimos a DNF

$$\begin{aligned}
&\neg \Big[\Big((\forall x.(p(x) \to q(x))) \land p(a)\Big) \to q(a)\Big] \\
&\equiv \neg \Big[\neg\Big((\forall x.(p(x) \to q(x))) \land p(a)\Big) \lor q(a)\Big] \\
&\equiv \neg\neg\Big((\forall x.(p(x) \to q(x))) \land p(a)\Big) \land \neg q(a) \\
&\equiv (\forall x.(p(x) \to q(x))) \land p(a) \land \neg q(a)
\end{aligned}$$

como a los ojos de DNF un $\forall$ es opaco, a pesar de que dentro tenga una implicación, la fórmula ya está en forma normal.

4. Buscamos una contradicción refutando cada cláusula. No hay forma encontrando literales opuestos o $\perp$, por ej. la cláusula $p(a)$ no es refutable.

Ejemplo con cuantificadores (3/3)

5. Probamos eliminando $\forall x.(p(x) \to q(x))$. Reemplazamos x por una meta-variable fresca u.

$$(p(\mathrm{u}) \to q(\mathrm{u})) \land p(a) \land \neg q(a)$$

6. Convertimos a DNF

$$\begin{aligned}
&(p(\mathrm{u}) \to q(\mathrm{u})) \land p(a) \land \neg q(a) \\
&\equiv (\neg p(\mathrm{u}) \lor q(\mathrm{u})) \land p(a) \land \neg q(a) \\
&\equiv ((\neg p(\mathrm{u}) \land p(a)) \lor (q(\mathrm{u}) \land p(a))) \land \neg q(a) \\
&\equiv (\neg p(\mathrm{u}) \land p(a) \land \neg q(a)) \lor \\
&\quad\ (q(\mathrm{u}) \land p(a) \land \neg q(a))
\end{aligned}$$

7. Buscamos una contradicción refutando cada cláusula. Los literales opuestos tienen que *unificar* en lugar de ser iguales.
   - $\neg p(\mathrm{u}) \land p(a) \land \neg q(a)$ tenemos $p(\mathrm{u}) \doteq p(a)$ con $\{\mathrm{u} := a\}$
   - $q(\mathrm{u}) \land p(a) \land \neg q(a)$ tenemos $q(\mathrm{u}) \doteq q(a)$ con $\{\mathrm{u} := a\}$

Desafío

¡Hay que generar una demostración en deducción natural!

Pasos

- **Razonamiento por el absurdo**: mediante las *reglas admisibles* cut y eliminación de la doble negación (E$\neg\neg$).
- **Conversión a DNF**: mediante la implementación de un *sistema de reescritura*.
- **Contradicciones**: mediante la *regla admisible* E$\wedge_\varphi$ + E$\vee$ + I$\neg$.
- **Eliminación de cuantificadores universales**: mediante unificación y E$\forall$.

Razonamiento por el absurdo

Conversión a DNF

Contradicciones

Alcance y limitaciones

- Completo para lógica proposicional y heurístico para primer orden. Esto es aceptable, la validez de LPO es indecidible (Teorema de Church).

Descarga de conjunciones

Extracción de testigos