



# Τεχνητή Νοημοσύνη

*8η διάλεξη (2024-25)*

Ίων Ανδρουτσόπουλος

<http://www.aueb.gr/users/ion/>

Οι διαφάνειες αυτής της διάλεξης βασίζονται στο βιβλίο *Artificial Intelligence – A Modern Approach* των S. Russel και P. Norvig, 2<sup>η</sup> και 4<sup>η</sup> έκδοση, Prentice Hall, 2003 και 2020. Τα περισσότερα σχήματα των διαφανειών προέρχονται από αντίστοιχες διαφάνειες του ίδιου βιβλίου.

# Τι θα ακούσετε σήμερα

- Περισσότερα για την **εξαγωγή συμπερασμάτων με προτασιακή λογική**:
  - **Εγκυρότητα και ικανοποιησιμότητα.**
  - Εξαγωγή συμπερασμάτων με **αναζήτηση απόδειξης.**
  - Εξαγωγή συμπερασμάτων με τον **κανόνα της ανάλυσης (resolution).**

# Ταυτολογική ισοδυναμία

- $\alpha \equiv \beta$  σημαίνει:  $\alpha \models \beta$  και  $\beta \models \alpha$ .
  - Οι προτάσεις  $\alpha$  και  $\beta$  αληθεύουν **στα ίδια ακριβώς μοντέλα**.

$$(\alpha \wedge \beta) \equiv (\beta \wedge \alpha) \quad \text{commutativity of } \wedge$$

$$(\alpha \vee \beta) \equiv (\beta \vee \alpha) \quad \text{commutativity of } \vee$$

$$((\alpha \wedge \beta) \wedge \gamma) \equiv (\alpha \wedge (\beta \wedge \gamma)) \quad \text{associativity of } \wedge$$

$$((\alpha \vee \beta) \vee \gamma) \equiv (\alpha \vee (\beta \vee \gamma)) \quad \text{associativity of } \vee$$

$$\neg(\neg\alpha) \equiv \alpha \quad \text{double-negation elimination}$$

$$(\alpha \Rightarrow \beta) \equiv (\neg\beta \Rightarrow \neg\alpha) \quad \text{contraposition}$$

$$(\alpha \Rightarrow \beta) \equiv (\neg\alpha \vee \beta) \quad \text{implication elimination}$$

$$(\alpha \Leftrightarrow \beta) \equiv ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha)) \quad \text{biconditional elimination}$$

$$\neg(\alpha \wedge \beta) \equiv (\neg\alpha \vee \neg\beta) \quad \text{de Morgan}$$

$$\neg(\alpha \vee \beta) \equiv (\neg\alpha \wedge \neg\beta) \quad \text{de Morgan}$$

$$(\alpha \wedge (\beta \vee \gamma)) \equiv ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \quad \text{distributivity of } \wedge \text{ over } \vee$$

$$(\alpha \vee (\beta \wedge \gamma)) \equiv ((\alpha \vee \beta) \wedge (\alpha \vee \gamma)) \quad \text{distributivity of } \vee \text{ over } \wedge$$

# Εγκυρότητα και ικανοποιησιμότητα

- «Η  $\alpha$  είναι **ταυτολογία**» ή «**έγκυρη**» σημαίνει:  $\alpha \equiv \text{True}$ .
  - Η πρόταση  $\alpha$  αληθεύει σε όλα τα μοντέλα.
- Θεώρημα:  $\alpha \models \beta$  ανν  $(\alpha \Rightarrow \beta)$  **έγκυρη (ταυτολογία)**.
  - Απόδειξη με ορισμό του  $\models$  και πίνακα αληθείας του  $\Rightarrow$ . (Άσκηση)
- «Η  $\alpha$  είναι **ικανοποιήσιμη**» σημαίνει υπάρχει μοντέλο στο οποίο η  $\alpha$  είναι αληθής.
  - Π.χ. η πρόταση  $(P \wedge \neg P)$  είναι **μη ικανοποιήσιμη**.
  - Επομένως  $\alpha$  **έγκυρη** ανν  $\neg\alpha$  **μη ικανοποιήσιμη**.
- $\alpha \not\models \beta$  ανν  $(\alpha \wedge \neg\beta)$  **μη ικανοποιήσιμη**
  - Η γνωστή μας «απαγωγή σε άτοπο».
  - $\alpha \models \beta$  ανν  $(\alpha \Rightarrow \beta)$  **έγκυρη** ανν  $\neg(\alpha \Rightarrow \beta)$  **μη ικανοποιήσιμη** ανν  $\neg(\neg\alpha \vee \beta)$  **μη ικανοποιήσιμη** ανν  $(\alpha \wedge \neg\beta)$  **μη ικανοποιήσιμη**.
  - Δυστυχώς ο έλεγχος ικανοποιησιμότητας στην προτασιακή λογική είναι **NP-πλήρες** πρόβλημα.

# Κανόνες εξαγωγής συμπερασμάτων

- Παράγουν συμπεράσματα από τύπους που υπάρχουν ήδη στη ΒΓ.
  - « $\alpha, \beta \vdash \gamma$ » σημαίνει ότι από τους τύπους  $\alpha$  και  $\beta$  ο κανόνας παράγει τον τύπο  $\gamma$ .
- $\{(\alpha \Rightarrow \beta), \alpha\} \vdash \beta$  (Modus Ponens)
  - Απόδειξη ορθότητας με πίνακα αληθείας. (Άσκηση)
  - Αποδεικνύουμε δηλαδή ότι:  $((\alpha \Rightarrow \beta) \wedge \alpha) \vDash \beta$ .
- $\{(\alpha \Rightarrow \beta), \neg\beta\} \vdash \neg\alpha$  (Modus Tollens)
  - Απόδειξη με πίνακα αληθείας. (Άσκηση)
- $(\alpha \wedge \beta) \vdash \alpha$  (απαλοιφή σύζευξης)
- $(\alpha \Leftrightarrow \beta) \vdash ((\alpha \Rightarrow \beta) \wedge (\beta \Rightarrow \alpha))$  (απαλοιφή  $\Leftrightarrow$ )
- Μπορούν να χρησιμοποιηθούν και όλες οι ταυτολογικές ισοδυναμίες, προς οποιαδήποτε κατεύθυνση.

# Εξαγωγή συμπεράσματος με αναζήτηση απόδειξης

- **Αρχική κατάσταση:** η ΒΓ στην αρχική της μορφή.
- **Τελική κατάσταση:** η ΒΓ σε μορφή που να περιέχει τον προς απόδειξη τύπο.
- **Τελεστές μετάβασης:** κανόνες εξαγωγής συμπερασμάτων
  - Προσθέτουν τύπους στη ΒΓ.
- **Απόδειξη:** μονοπάτι από αρχική σε τελική κατάσταση.
- Στην πράξη η αναζήτηση απόδειξης μπορεί να είναι **πιο αποδοτική** από τον εξαντλητικό έλεγχο μοντέλων.
  - Αν θέλουμε να δείξουμε ότι  $\alpha \not\models \beta$ , μπορούμε να ελέγξουμε **εξαντλητικά** όλα τα μοντέλα με τον TT-Entails?.
  - Ή να κατασκευάσουμε μια **απόδειξη** (μονοπάτι) από αρχική ΒΓ που περιέχει μόνο το  $\alpha$  σε τελική ΒΓ που περιέχει και το  $\beta$ .
- Αλλά πάλι στη χειρότερη περίπτωση, **εκθετικός χρόνος**.
  - Διαφορετικά θα είχαμε τρόπο να λύσουμε το **NP-πλήρες** πρόβλημα της ικανοποιησιμότητας σε πολυωνυμικό χρόνο!

# Παράδειγμα στον κόσμο του Wumpus

- Έστω ότι η ΒΓ περιέχει αρχικά:
  - $R_1: \neg P_{1,1}$        $R_4: \neg B_{1,1}$        $R_5: B_{2,1}$
  - $R_2: (B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}))$
  - $R_3: (B_{2,1} \Leftrightarrow (P_{1,1} \vee P_{2,2} \vee P_{3,1}))$
- Θέλουμε να αποδείξουμε  $(\neg P_{1,2} \wedge \neg P_{2,1})$ .
- Εφαρμογή απαλοιφής  $\Leftrightarrow$  στο  $R_2$ :
  - $R_6: (B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$
- Εφαρμογή απαλοιφής σύζευξης στο  $R_6$ :
  - $R_7: ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1})$
- Εφαρμογή αντιθετοαντιστροφής στο  $R_7$ :
  - $R_8: (\neg B_{1,1} \Rightarrow \neg (P_{1,2} \vee P_{2,1}))$
- Modus Ponens με  $R_8$  και  $R_4$ :
  - $R_9: \neg (P_{1,2} \vee P_{2,1})$
- Εφαρμογή κανόνα De Morgan στο  $R_9$  δίνει το ζητούμενο.

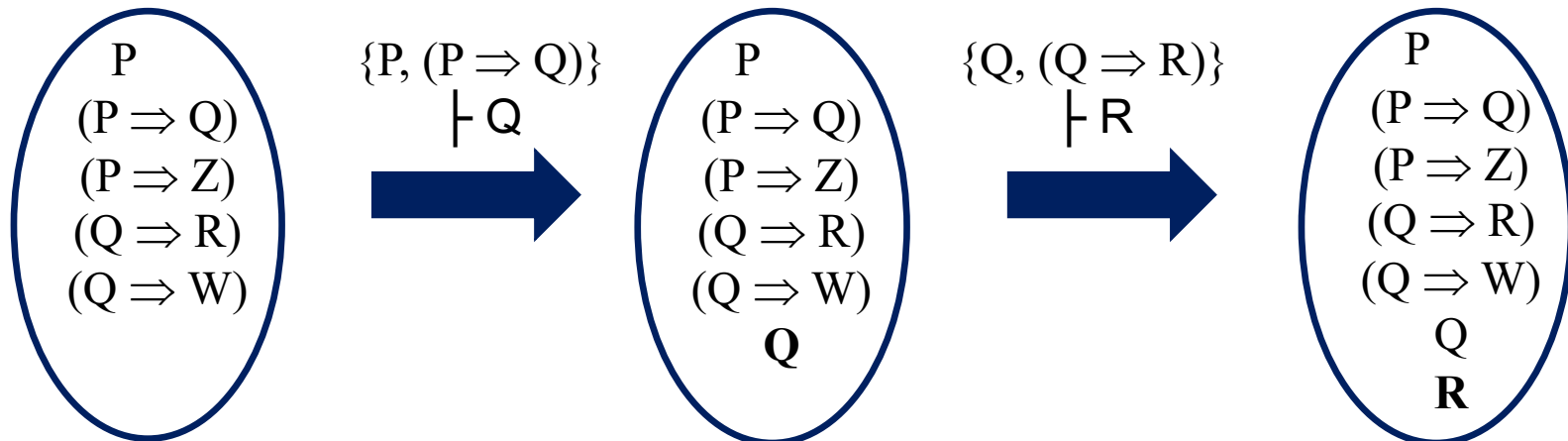


# Η αναζήτηση μπορεί να είναι πιο γρήγορη...

- Έστω ΒΓ:  $\{P, (P \Rightarrow Q), (P \Rightarrow Z), (Q \Rightarrow R), (Q \Rightarrow W)\}$
- Θέλουμε να δείξουμε:  $B\Gamma \models R$ .
- Ο TT-Entails? θα εξετάσει πίνακα με  $2^5$  γραμμές:

P	Q	R	Z	W	BΓ
T	T	T	T	T	T
F	T	T	T	T	F
...	...	...	...	...	...

- Με ιδανική ευρετική, η αναζήτηση απόδειξης (ή «απόδειξη θεωρημάτων») θα χρειαστεί μόνο 2 βήματα:



# Κανονική συζευκτική μορφή

- Κάθε πρόταση της ΠΛ είναι ταυτολογικά ισοδύναμη με μια πρόταση σε **κανονική συζευκτική μορφή (CNF)**:
  - $(l_{1,1} \vee \dots \vee l_{1,k_1}) \wedge \dots \wedge (l_{n,1} \vee \dots \vee l_{n,k_n})$
  - Κάθε  $l_{i,j}$  είναι **σύμβολο** (π.χ. P) ή **άρνηση συμβόλου** (π.χ.  $\neg P$ ).
- **Οπότε και η ΒΓ μπορεί να γραφτεί σε CNF.**
  - Μπορούμε να θεωρήσουμε ότι η ΒΓ είναι ένας μόνο τύπος, **μία μεγάλη σύζευξη όλων των τύπων που περιέχει η ΒΓ.**
  - Μετατρέπουμε τη μεγάλη σύζευξη σε CNF, οπότε η ΒΓ γίνεται **μία μεγάλη σύζευξη διαζεύξεων**, η κάθε μία διάζευξη της μορφής  $(l_{i,1} \vee \dots \vee l_{i,k_i})$ .
  - Μπορούμε κατόπιν να θεωρήσουμε ότι η ΒΓ περιέχει ως **ξεχωριστούς τύπους** όλες τις διαζεύξεις  $(l_{i,1} \vee \dots \vee l_{i,k_i})$  που προκύπτουν.

# Μετατροπή σε CNF

Συμβουλή:  
Εφαρμόζετε τα  
βήματα πάντα με  
αυτή τη σειρά και  
μην παραλείπετε  
παρενθέσεις!

- **Μέθοδος** μετατροπής σε CNF:
  - Π.χ. ξεκινώντας από  $(B_{1,1} \Leftrightarrow (P_{1,2} \vee P_{2,1}))$ .
- Βήμα 1: **απαλοιφή**  $\Leftrightarrow$ .
  - $((B_{1,1} \Rightarrow (P_{1,2} \vee P_{2,1})) \wedge ((P_{1,2} \vee P_{2,1}) \Rightarrow B_{1,1}))$
- Βήμα 2: **απαλοιφή**  $\Rightarrow$ .
  - $((\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg(P_{1,2} \vee P_{2,1}) \vee B_{1,1}))$
- Βήμα 3: **μεταφορά**  $\neg$  στο εσωτερικό ως τα σύμβολα.
  - Με χρήση κανόνων De Morgan και διπλής άρνησης.
  - $((\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge ((\neg P_{1,2} \wedge \neg P_{2,1}) \vee B_{1,1}))$
- Βήμα 4: **επιμερισμός** των  $\vee$ .
  - $((\neg B_{1,1} \vee P_{1,2} \vee P_{2,1}) \wedge (\neg P_{1,2} \vee B_{1,1}) \wedge (\neg P_{2,1} \vee B_{1,1}))$

# Ο κανόνας της ανάλυσης (resolution)

- Μετατρέπουμε πρώτα τη **ΒΓ σε CNF** και χρησιμοποιούμε τις διαζεύξεις ως ξεχωριστούς τύπους.
  - Παίρνουμε ένα σύνολο από διαζευκτικούς τύπους, ο καθένας της μορφής  $(l_{i,1} \vee \dots \vee l_{i,k_i})$ .
  - Κάθε  $l_{i,j}$  («literal») είναι **σύμβολο** (π.χ. P) ή **άρνηση συμβόλου** (π.χ.  $\neg P$ ).
- **Ο κανόνας της ανάλυσης:**
  - Αν το  $l_i$  είναι η άρνηση του  $m_j$  ή αντίστροφα, τότε:
  - $(l_1 \vee \dots \vee l_k), (m_1 \vee \dots \vee m_n) \vdash (l_1 \vee \dots \vee l_{i-1} \vee l_{i+1} \vee \dots \vee l_k \vee m_1 \vee \dots \vee m_{j-1} \vee m_{j+1} \vee \dots \vee m_n)$
  - Ενώνουμε τις δύο διαζεύξεις και αφαιρούμε τα  $l_i$  και  $m_j$ .
  - Αφαιρούμε από τη διάζευξη που παράγεται τυχόν **πολλαπλά αντίγραφα** των ίδιων  $l$  και  $m$ .

# Ορθότητα του κανόνα της ανάλυσης

- Πρέπει να δείξουμε ότι ο κανόνας της ανάλυσης:  
 $(l_1 \vee \dots \vee l_k), (m_1 \vee \dots \vee m_n) \vdash (l_1 \vee \dots \vee l_{i-1} \vee l_{i+1} \vee \dots \vee l_k \vee m_1 \vee \dots \vee m_{j-1} \vee m_{j+1} \vee \dots \vee m_n)$   
είναι ορθός.
- Δηλαδή ότι **όποτε** (σε όποια μοντέλα) αληθεύουν τα  $(l_1 \vee \dots \vee l_k)$  και  $(m_1 \vee \dots \vee m_n)$ , τότε αληθεύει και το:  
 $(l_1 \vee \dots \vee l_{i-1} \vee l_{i+1} \vee \dots \vee l_k \vee m_1 \vee \dots \vee m_{j-1} \vee m_{j+1} \vee \dots \vee m_n)$

# Ορθότητα του κανόνα της ανάλυσης

- Έστω ότι βρισκόμαστε σε μοντέλο όπου αληθεύουν τα  $(l_1 \vee \dots \vee l_k)$  και  $(m_1 \vee \dots \vee m_n)$ .
  - Και κάποιο  $l_i$  είναι η άρνηση κάποιου  $m_j$ .
- Αν το  $l_i$  είναι αληθές σε αυτό το μοντέλο,
  - τότε το  $m_j$  είναι ψευδές (αφού το ένα είναι η άρνηση του άλλου).
  - Αφού όμως αληθεύει  $(m_1 \vee \dots \vee m_n)$ , τότε είναι αληθές κάποιο από τα υπόλοιπα  $m$ .
  - Άρα αληθεύει το  $(m_1 \vee \dots \vee m_{j-1} \vee m_{j+1} \vee \dots \vee m_n)$ .
- Αν το  $l_i$  είναι ψευδές σε αυτό το μοντέλο,
  - τότε αφού αληθεύει  $(l_1 \vee \dots \vee l_k)$ , είναι αληθές κάποιο από τα υπόλοιπα  $l$ .
  - Άρα αληθεύει το  $(l_1 \vee \dots \vee l_{i-1} \vee l_{i+1} \vee \dots \vee l_k)$ .
- Άρα και στις δύο περιπτώσεις αληθεύει το:
  - $(l_1 \vee \dots \vee l_{i-1} \vee l_{i+1} \vee \dots \vee l_k) \vee (m_1 \vee \dots \vee m_{j-1} \vee m_{j+1} \vee \dots \vee m_n)$ .

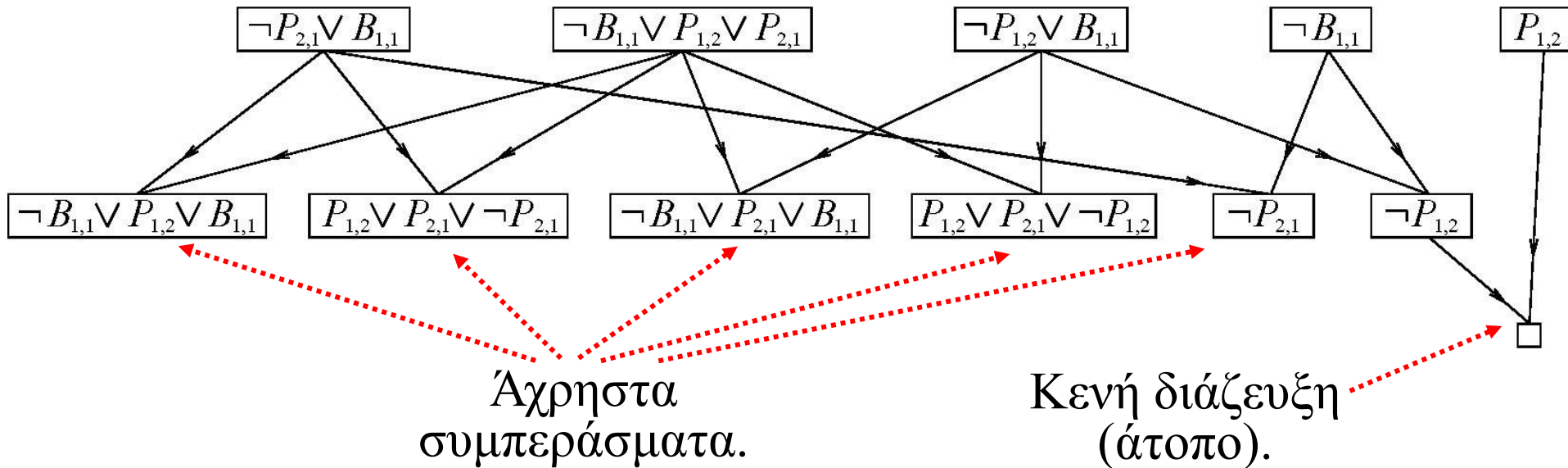
# Εξαγωγή συμπερασμάτων με ανάλυση

- Θέλουμε να εξετάσουμε αν  $B\Gamma \models \alpha$ .
- Μετατρέπουμε το  $(B\Gamma \wedge \neg\alpha)$  σε CNF.
  - Δηλαδή προσθέτουμε την άρνηση του προς απόδειξη τύπου στη  $B\Gamma$  και μετατρέπουμε τη  $B\Gamma$  σε CNF.
  - Και «σπάμε» τον τύπο CNF στις διαζεύξεις του.
- Εφαρμόζουμε τον κανόνα της ανάλυσης όσο είναι δυνατόν.
  - Μέχρι να μην μπορούμε να παραγάγουμε νέους τύπους.
- Απαντούμε  $B\Gamma \models_i \alpha$ , αν παραχθεί η κενή διάζευξη  $\square$ .
  - Η κενή διάζευξη δείχνει ότι καταλήξαμε σε άτοπο. Προκύπτει όταν έχουμε στη  $B\Gamma$  ένα σύμβολο και την άρνησή του.
- Απαντούμε  $B\Gamma \not\models_i \alpha$ , αν δεν παραχθεί  $\square$ .

# Παράδειγμα στον κόσμο του Wumpus

Προσοχή: Οι τύποι πρέπει να είναι **διαζεύξεις**. Αν ένας τύπος είναι σύζευξη διαζεύξεων, τον σπάμε σε ξεχωριστούς τύπους-διαζεύξεις.

Προσοχή: Σε κάθε «ζευγάρι» τύπων, φεύγει πάντα **ένα μόνο σύμβολο** από τον έναν τύπο και **ένα μόνο σύμβολο** από τον άλλον.



Ο αλγόριθμος **τερματίζει πάντα**, γιατί υπάρχει πεπερασμένος αριθμός διαζεύξεων (χωρίς πολλαπλά αντίγραφα του ίδιου όρου) που μπορούν να κατασκευαστούν από τα πεπερασμένα σύμβολα της αρχικής ΒΓ και της προς απόδειξη πρότασης.



# Αλγόριθμος απόδειξης με ανάλυση

Οι διαζεύξεις της CNF μορφής.

**function** PL-RESOLUTION( $KB, \alpha$ ) **returns** *true* or *false*

$clauses \leftarrow$  the set of clauses in the CNF representation of  $KB \wedge \neg\alpha$

$new \leftarrow \{ \}$

**loop do**

**for each**  $C_i, C_j$  **in**  $clauses$  **do**

$resolvents \leftarrow$  PL-RESOLVE( $C_i, C_j$ )

**if**  $resolvents$  contains the empty clause **then return** *true*

$new \leftarrow new \cup resolvents$

**if**  $new \subseteq clauses$  **then return** *false*

$clauses \leftarrow clauses \cup new$

Εφαρμογή του κανόνα resolution.

Αν δεν καταφέραμε να παραγάγουμε καμία νέα διάζευξη, σταματάμε και απαντάμε ΒΓ~~χι~~ α.

# Ορθότητα και πληρότητα

- **Ορθότητα** (αν  $B\Gamma \vdash_i \alpha$ , τότε  $B\Gamma \vDash \alpha$ ).
  - Αν  $B\Gamma \vdash_i \alpha$ , τότε παρήχθη  $\square$  με διαδοχικές εφαρμογές του κανόνα της ανάλυσης, ξεκινώντας από  $(B\Gamma \wedge \neg\alpha)$ .
  - Οπότε  $(B\Gamma \wedge \neg\alpha) \vDash \square$ , λόγω της ορθότητας του κανόνα της ανάλυσης. Άρα σε όλα τα μοντέλα όπου αληθεύει η  $(B\Gamma \wedge \neg\alpha)$  αληθεύει και η  $\square$ . Αλλά η  $\square$  δεν αληθεύει ποτέ.
  - Άρα η  $(B\Gamma \wedge \neg\alpha)$  δεν αληθεύει σε κανένα μοντέλο, δηλαδή είναι μη ικανοποιήσιμη. Οπότε,  $B\Gamma \vDash \alpha$ .
  - Γιατί, σύμφωνα με τα προηγούμενα,  $B\Gamma \vDash \alpha$  ανν  $(B\Gamma \wedge \neg\alpha)$  μη ικανοποιήσιμη.
- **Πληρότητα** (αν  $B\Gamma \vDash \alpha$ , τότε  $B\Gamma \vdash_i \alpha$ ).
  - Δηλαδή αν  $B\Gamma \vDash \alpha$ , τότε παράγεται η κενή διάζευξη.
  - Η απόδειξη παραλείπεται.

# Βιβλιογραφία

- Russel & Norvig (4<sup>η</sup> έκδοση): ενότητα 7.5 ως και υπο-ενότητα 7.5.2, χωρίς την υπο-ενότητα «πληρότητα της ανάλυσης»
  - Όσοι ενδιαφέρονται μπορούν να διαβάσουν προαιρετικά και την υπο-ενότητα «πληρότητα της ανάλυσης».
- Βλαχάβας κ.ά: ενότητες 9.1.1, 9.1.2.