SAT $\leq_P$ 3SAT

Theoretische Informatica II

1 Inleiding

In dit tekstje beschrijven we een polynomiale reductie van SAT naar 3SAT. Omdat we reeds weten dat SAT NP-compleet is, betekent dit dat ook 3SAT NP-compleet is! (Waarom?)

2 Het begrip "subformule"

Zij ϕ een booleaanse formule. De verzameling *subformules* van ϕ wordt inductief gedefiniëerd als volgt:

- Als ϕ simpel een variabele x is, is de enige subformule van ϕ, ϕ zelf.
- Als ϕ van de vorm $(\phi_1 \wedge \phi_2)$ is, zijn de subformules van ϕ alle subformules van ϕ_1, alle subformules van ϕ_2, en ook ϕ zelf.
- Als ϕ van de vorm $\neg\phi_1$ is, zijn de subformules van ϕ alle subformules van ϕ_1, en ook ϕ zelf.

Merk op dat we hier enkel werken met de AND ($\wedge$) en de NOT ($\neg$). Inderdaad, de OR $(\phi_1 \vee \phi_2)$ kan gezien worden als een afkorting voor $\neg(\neg\phi_1 \wedge \neg\phi_2)$.

Voorbeeld 1 *Alle subformules van de formule*

$$\neg(x \wedge \neg y) \wedge (y \wedge \neg z)$$

zijn de volgende:

1) x
2) y
3) $\neg y$
4) $x \wedge \neg y$
5) $\neg(x \wedge \neg y)$
6) z
7) $\neg z$
8) $y \wedge \neg z$
9) $\neg(x \wedge \neg y) \wedge (y \wedge \neg z)$

Oefening 1 *Geef alle subformules van volgende formule:*

$$x \wedge \neg(\neg(y \wedge \neg z) \wedge \neg x)$$

3 De hulpformules ε_j

Een reductie van SAT naar 3SAT krijgt als input een booleaanse formule ϕ. Een naïeve poging zou erin bestaan ϕ gewoon te converteren naar 3CNF. Het is echter onmogelijk dit op een efficiënte wijze te doen. De CNF kan namelijk exponentiëel veel langer zijn dan de originele formule. Een voorbeeld is volgende formule:

$$(x_1 \wedge y_1) \vee (x_2 \wedge y_2) \vee \cdots \vee (x_{100} \wedge y_{100})$$

Probeer ze maar eens in CNF om te zetten; je zal al gauw merken dat je 2^{100} clauses nodig hebt!

We moeten dus iets slimmers proberen. We gaan een aantal hulpformules construeren, voor elke subformule eentje, die wél gemakkelijk te converteren zijn naar 3CNF. Zet alle subformules van ϕ op een rijtje:

$$\psi_1, \psi_2, \ldots, \psi_m$$

Neem voor elke $j = 1, \ldots, m$, met m dus het aantal verschillende subformules van ϕ, een nieuwe variabele u_j die nog niet voorkomt in ϕ.

Construeer nu voor elke $j = 1, \ldots, m$ de hulpformule ε_j als volgt:

- Als ψ_j een variabele is, stel x, dan definiëren we
$$\varepsilon_j \quad := \quad u_j \leftrightarrow x$$
- Als ψ_j van de vorm $(\psi_k \wedge \psi_\ell)$ is, dan definiëren we
$$\varepsilon_j \quad := \quad u_j \leftrightarrow (u_k \wedge u_\ell)$$
- Als ψ_j van de vorm $\neg\psi_k$ is, dan definiëren we
$$\varepsilon_j \quad := \quad u_j \leftrightarrow \neg u_k$$

Voorbeeld 2 *Als we Voorbeeld 1 verderzetten krijgen we volgende hulpformules:*

j	ε_j
1	$u_1 \leftrightarrow x$
2	$u_2 \leftrightarrow y$
3	$u_3 \leftrightarrow \neg u_2$
4	$u_4 \leftrightarrow (u_1 \wedge u_3)$
5	$u_5 \leftrightarrow \neg u_4$
6	$u_6 \leftrightarrow z$
7	$u_7 \leftrightarrow \neg u_6$
8	$u_8 \leftrightarrow (u_2 \wedge u_7)$
9	$u_9 \leftrightarrow (u_5 \wedge u_8)$

Oefening 2 *Doe hetzelfde met Oefening 1.*

4 Omzetting naar 3CNF

Er zijn 3 verschillende types van hulpformules ε_j: die van het type $u_j \leftrightarrow x$; die van het type $u_j \leftrightarrow (u_k \wedge u_\ell)$; en die van het type $u_j \leftrightarrow \neg u_k$. Elk type kunnen we gemakkelijk omzetten in 3CNF als volgt:

$$\begin{aligned} u_j \leftrightarrow x &\equiv (u_j \rightarrow x) \wedge (x \rightarrow u_j) \\ &\equiv (\neg u_j \vee x) \wedge (\neg x \vee u_j) \\ &\equiv (\neg u_j \vee x \vee x) \wedge (\neg x \vee u_j \vee u_j) \end{aligned}$$

$$\begin{aligned} u_j \leftrightarrow (u_k \wedge u_\ell) &\equiv (u_j \rightarrow (u_k \wedge u_\ell)) \wedge ((u_k \wedge u_\ell) \rightarrow u_j) \\ &\equiv (\neg u_j \vee (u_k \wedge u_\ell)) \wedge (\neg(u_k \wedge u_\ell) \vee u_j) \\ &\equiv (\neg u_j \vee u_k) \wedge (\neg u_j \vee u_\ell) \wedge (\neg u_k \vee \neg u_\ell \vee u_j) \\ &\equiv (\neg u_j \vee u_k \vee u_k) \wedge (\neg u_j \vee u_\ell \vee u_\ell) \wedge (\neg u_k \vee \neg u_\ell \vee u_j) \end{aligned}$$

$$\begin{aligned} u_j \leftrightarrow \neg u_k &\equiv (u_j \rightarrow \neg u_k) \wedge (\neg u_k \rightarrow u_j) \\ &\equiv (\neg u_j \vee \neg u_k) \wedge (u_k \vee u_j) \\ &\equiv (\neg u_j \vee \neg u_k \vee \neg u_k) \wedge (u_k \vee u_j \vee u_j) \end{aligned}$$

5 De reductie

We zijn nu klaar om onze reductie te beschrijven. Een van de subformules ψ_1, $\ldots$, ψ_m is ϕ zelf; onderstel dat dit ψ_m is. We definiëren nu volgende functie f van ϕ:

$$f(\phi) \quad := \quad \varepsilon_1 \wedge \cdots \wedge \varepsilon_m \wedge (u_m \vee u_m \vee u_m)$$

Onderstellend dat elke ε_j reeds in 3CNF is herschreven (we weten van hierboven dat dit gemakkelijk kan) is $f(\phi)$ een formule in 3CNF.

Hoe lang is $f(\phi)$? Elke ε_j heeft slechts een vaste lengte (bestaat uit hoogstens uit 3 variabelen!) Ook het stukje $(u_m \vee u_m \vee u_m)$ heeft een vaste lengte. De lengte van $f(\phi)$ is dus m keer een vaste lengte, waarbij m het aantal subformules is van ϕ. Dit aantal is duidelijk lineair in de lengte van ϕ zelf (een formule kan niet meer subformules hebben dan ze lang is!) We besluiten dat de lengte van $f(\phi)$ lineair is in de lengte van ϕ en dus zeker polynomiaal.[1]

Het is eenvoudig een efficiënt programmaatje te schrijven dat, op input ϕ, de formule $f(\phi)$ produceert. Onze reductie is dus polynomiaal berekenbaar.

[1] Eigenlijk is het $O(n \log n)$ in plaats van $O(n)$ omdat we $\log n$ bits nodig hebben om elke variabele u_j te schrijven.

6 Correctheid van de reductie

Het enige dat ons rest is aan te tonen dat onze reductie correct is. We moeten dus aantonen dat ϕ satisfiable is als en slechts $f(\phi)$ dat is. Dit is intuïtief duidelijk: de hulpformules ε_j introduceren hulpvariabelen u_j die de waarheidswaarde voorstellen van de subformules ψ_j. Omdat $f(\phi)$ bestaat uit de AND van al deze ε_j's, tesamen met u_m, die dus de waarheidswaarde van ψ_m voorstelt, wat ϕ zelf is (zie hierboven), is het dus duidelijk dat $f(\phi)$ in zekere zin "equivalent" is met ϕ.

Deze intuïtie maken we nu hard in de volgende twee lemma's.

Lemma 1 *Zij α een waarheidstoekenning op de variabelen voorkomend in ϕ. Onder deze α evalueert elke subformule ψ tot een waarde 1 of 0, die we noteren als $\alpha(\psi)$. Breid α uit naar de variabelen u_j als volgt:*

$$\alpha(u_j) := \begin{cases} 1 & \textit{als } \alpha(\psi_j) = 1 \\ 0 & \textit{als } \alpha(\psi_j) = 0 \end{cases}$$

Dan is elke hulpformule ε_j, onder deze uitgebreide waarheidstoekenning, voldaan.

Bewijs.

- Als ψ_j een variabele x is, dan is ε_j per definitie de formule $u_j \leftrightarrow x$. Deze is duidelijk voldaan onder α, omdat we $\alpha(u_j)$ op 1 hebben gesteld als en slechts als $\alpha(x)$ gelijk is aan 1.

- Als ψ_j van de vorm $(\psi_k \wedge \psi_\ell)$ is, dan is ε_j per definitie de formule $u_j \leftrightarrow (u_k \wedge u_\ell)$. We hebben $\alpha(u_j)$ op 1 gesteld als en slechts als ψ_j tot 1 evalueert onder α. Dit gebeurt als en slechts als zowel ψ_k en ψ_ℓ tot 1 evalueren onder α. En dit laatste is precies wanneer zowel $\alpha(u_k)$ en $\alpha(u_\ell)$ op 1 zijn gesteld. De formule ε_j is dus voldaan onder α.

- Als ψ_j van de vorm $\neg\psi_k$ is, dan is ε_j per definitie de formule $u_j \leftrightarrow \neg u_k$. We hebben $\alpha(u_j)$ op 1 gesteld als en slechts als ψ_j tot 1 evalueert onder α. Dit gebeurt als en slechts als ψ_k tot 0 evalueert onder α. En dit laatste is precies wanneer $\alpha(u_k)$ op 0 is gesteld. De formule ε_j is dus voldaan onder α.

Het tweede lemma is een soort van omgekeerde van het eerste lemma:

Lemma 2 *Stel dat α een waarheidstoekenning is op alle variabelen in ϕ en alle variabelen u_j, zodat alle formules ε_j voldaan zijn onder α. Dan geldt voor elke j:*

$$\psi_j \textit{ is voldaan onder } \alpha \quad \Leftrightarrow \quad \alpha(u_j) = 1$$

Bewijs.

- Als ψ_j een variabele x is, dan

$$\begin{aligned}\psi_j \text{ voldaan onder } \alpha &\Leftrightarrow \alpha(x) = 1 \\ &\Leftrightarrow \alpha(u_j) = 1 \qquad \text{want } \varepsilon_j \text{ voldaan onder } \alpha\end{aligned}$$

  Immers, ε_j is $u_j \leftrightarrow x$.

- Als ψ_j van de vorm $(\psi_k \wedge \psi_\ell)$ is, dan

$$\begin{aligned}\psi_j \text{ voldaan onder } \alpha &\Leftrightarrow \psi_k \text{ en } \psi_\ell \text{ voldaan onder } \alpha \\ &\Leftrightarrow \alpha(u_k) = 1 \text{ en } \alpha(u_\ell) = 1 \qquad \text{per inductie} \\ &\Leftrightarrow \alpha(u_j) = 1 \qquad \text{want } \varepsilon_j \text{ voldaan onder } \alpha\end{aligned}$$

  Immers, ε_j is $u_j \leftrightarrow (u_k \wedge u_\ell)$.

- Als ψ_j van de vorm $\neg\psi_k$ is, dan

$$\begin{aligned}\psi_j \text{ voldaan onder } \alpha &\Leftrightarrow \psi_k \text{ niet voldaan onder } \alpha \\ &\Leftrightarrow \alpha(u_k) \neq 1 \qquad \text{per inductie} \\ &\Leftrightarrow \alpha(u_j) = 1 \qquad \text{want } \varepsilon_j \text{ voldaan onder } \alpha\end{aligned}$$

  Immers, ε_j is $u_j \leftrightarrow \neg u_k$.

Gewapend met onze twee lemma's zijn we klaar voor volgende

Stelling. *ϕ is satisfiable als en slechts als $f(\phi)$ satisfiable is.*

Bewijs. *Als.* Stel dat $f(\phi)$ satisfiable is. Er is dus een waarheidstoekenning α op de variabelen in $f(\phi)$ waaronder $f(\phi)$ voldaan is. De variabelen in $f(\phi)$ zijn precies de variabelen in ϕ plus alle variabelen u_j. Omdat $f(\phi)$ gelijk is aan

$$\varepsilon_1 \wedge \cdots \wedge \varepsilon_m \wedge (u_m \vee u_m \vee u_m),$$

en voldaan is onder α, is in het bijzonder elke ε_j voldaan onder α. Lemma 2 is dus toepasbaar, en vertelt ons dat ψ_m voldaan is onder α als en slechts als $\alpha(u_m) = 1$. Nu, dit laatste geldt zeker en vast, want $f(\phi)$ is voldaan onder α, en dus in het bijzonder ook u_m, en dus $\alpha(u_m) = 1$.

We besluiten dat ψ_m voldaan is onder α. Maar ψ_m is ϕ zelf! Dus ϕ is voldaan onder α en is dus satisfiable.

Slechts als. Stel dat ϕ satisfiable is. Er is dus een waarheidstoekenning α op de variabelen in ϕ waaronder ϕ voldaan is. Lemma 1 vertelt ons dat α kan uitgebreid worden naar de variabelen u_j zodat elke ε_j voldaan wordt. Deze uitbreiding, zoals beschreven in het lemma, stelt $\alpha(u_m)$ op 1 als en slechts als $\alpha(\psi_m) = 1$. Echter ψ_m is ϕ, en we weten dat ϕ voldaan is onder α dus inderdaad $\alpha(\psi_m) = 1$. We besluiten dat $\alpha(u_m) = 1$. Dus, elke ε_j is voldaan onder de uitbreiding, en ook $(u_m \vee u_m \vee u_m)$ is voldaan. Dus is heel $f(\phi)$ voldaan, en dus is $f(\phi)$ satisfiable.