HET SCHAAKSPEL OPGELOST

Cees Timmer

HOOFDSTUK 1: INLEIDING

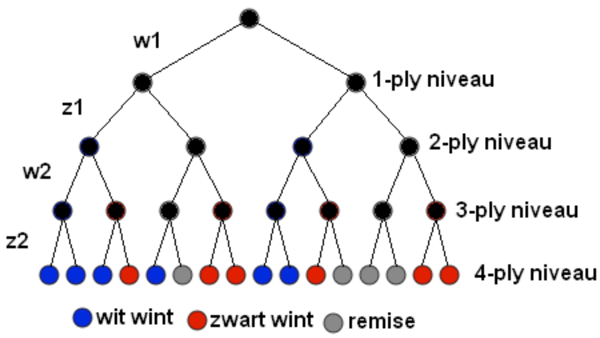
Onder deze provocerende titel wordt een niet al te technische uitleg gegeven hoe je het “schaakspel oplost”. Het schaakspel oplossen wil zeggen dat je na iedere zet van wit de beste zet van zwart vindt en omgekeerd. De “beste zet” wil zeggen de zet die zeker tot winst leidt of tenminste tot remise. Er zal niet altijd vanuit een gegeven stelling één beste zet bestaan; soms zijn er meerdere mogelijk.

Een menselijke speler zal vanuit een gegeven stelling trachten de beste zet te vinden door een combinatie van talent, inzicht, ervaring, intuïtie en rekenen. Toch is dit geen garantie dat de beste zet gevonden wordt. Een schaakprogramma rekent dieper maar mist menselijke eigenschappen zoals inzicht en intuïtie nodig voor de evaluatie van stellingen.

Dit artikel beschrijft een methode die gegarandeerd vanuit iedere stelling de beste zet oplevert. Het principe wordt eerst beschreven en daarna de mogelijkheden voor technische implementatie ervan.

HOOFDSTUK 2: SPELBOOM

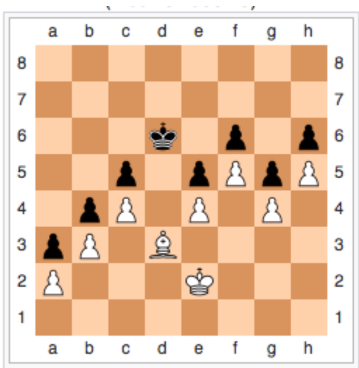
We starten vanuit de beginstelling waarin wit begint. Het proces wordt toegelicht aan de hand van Afbeelding 1.



Afbeelding 1: Evaluatie van de eindknopen (4-ply niveau)

Vanuit de beginstelling worden alle mogelijke zetten van wit uitgevoerd (w1). Dat zijn er in werkelijkheid 20. In Afbeelding 1 zijn dat er slechts 2. We gebruiken een sterk vereenvoudigd voorbeeld om het principe uit te leggen. Vervolgens voeren we na iedere mogelijke zet van wit alle mogelijke zwarte zetten uit (z1), enz. In Afbeelding 1 zijn er steeds 2 mogelijke zetten voor zwart na een zet van wit en omgekeerd. Je herhaalt dit proces totdat er geen zetten voor wit of zwart meer mogelijk zijn of een remise-spelregel van toepassing is. Dit is het geval als:

1) Wit mat staat (winst voor zwart).
2) Zwart mat staat (winst voor wit).
3) Wit pat staat (remise).
4) Zwart pat staat (remise).
5) Er heeft zich driemaal achterelkaar dezelfde stelling voorgedaan (remise).
6) Indien er sinds de laatste 50 zetten geen stuk is geslagen of een pion is opgespeeld (50-zetten regel, remise).
7) In een stelling waarin noch wit noch zwart door een serie legale zetten mat kan geven. Meestal is dit het geval als zowel wit als zwart te weinig materiaal hebben om de tegenstander mat te zetten maar in andere stellingen kan deze situatie ook voorkomen. (zie Afbeelding 2).



Afbeelding 2: Remise.

Op deze wijze genereer je een *spelboom* (Eng. game tree) zoals is te zien in Afbeelding 1. Merk op dat niet iedere tak van de spelboom even lang hoeft te zijn zoals in ons voorbeeld wel het geval is. De bolletjes in de afbeelding zijn de ontstane stellingen na een zet van wit of zwart en de streepjes die de bolletjes verbinden zijn de halfzetten. Een zet van wit of zwart heet een *halfzet* (Eng. ply) en een zet van wit én zwart heet kortweg een *zet* (Eng. move). Het bovenste bolletje wordt de *wortel* (Eng. root) van de spelboom genoemd. De bolletjes heten *knopen* (Eng. nodes). De bolletjes van waaruit geen vertakkingen meer komen heten *eindknopen* (Eng. leaf nodes of terminal nodes).

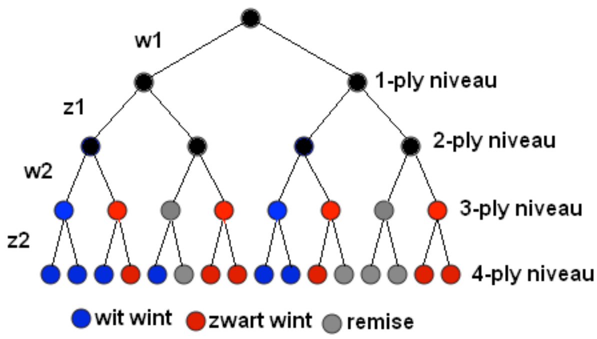
In de eindknopen zijn er slechts drie mogelijkheden: 1) zwart staat mat, dus wit wint (blauw bolletje), 2) wit staat mat, dus zwart wint (rood bolletje) of 3) remise (grijs bolletje).

Hoe bepalen we nu de beste zet voor wit? We starten met de onderste rij bolletjes (4-ply niveau). Voor de twee meest linkse eindknopen geldt dat allebei de zwarte zetten leiden tot winst voor wit, dus kleuren we het bolletje erboven blauw.

Bij de volgende twee eindknopen vanaf links gerekend zal zwart natuurlijk de zet kiezen naar de rode eindknoop en kleuren we het bolletje erboven rood.

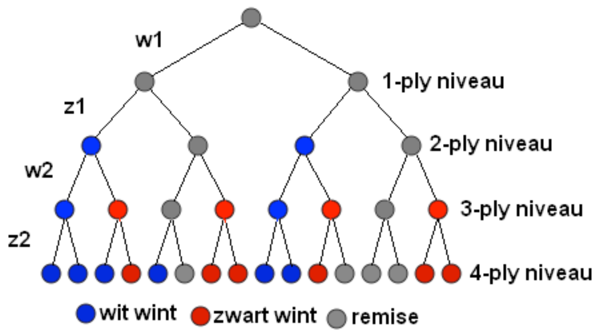
Bij het derde stel eindknopen vanaf links gerekend zal wit de zet kiezen naar het grijze bolletje en kleuren we het bolletje erboven grijs.

Op dezelfde wijze werken de overige bolletjes op de onderste rij af. Op 3-ply niveau hebben alle bolletjes nu een kleur gekregen (zie Afbeelding 3).



Afbeelding 3: Evaluatie op 3-ply niveau.

We herhalen bovenstaand proces voor alle bovenliggende rijen bolletjes totdat we Afbeelding 4 verkrijgen waarin alle bolletjes een kleur hebben gekregen.



Afbeelding 4: Evaluatie op 1-ply niveau.

De conclusie is dat wit vanuit de beginstelling maximaal remise kan krijgen. Dit algorithme waarbij we van achteren naar voren evalueren heet *retrograde analyse* of *achterwaartse* *inductie*. In werkelijkheid is de spelboom vanuit de beginpositie natuurlijk van ongekende proporties die ons voorstellingsvermogen ver te boven gaan. Toch zullen we in het volgende hoofstuk trachten de orde van grootte van de werkelijke spelboom te berekenen. Neemt allemaal niet weg dat de spelboom eindig is en een retrograde analyse dus theoretisch mogelijk is. Later zullen we de mogelijkheden van de praktische implementatie bekijken.

Schaakprogramma's maken gebruik van retrograde analyse met een gedeeltelijke spelboom waarvan op slimme wijze hier en daar takken worden afgesnoeid. Behalve bij het vinden van een *mat-in-n* bereiken schaakprogramma's nooit de eindknopen omdat dit technisch (nog) niet mogelijk is.

HOOFDSTUK 3: AANTAL PARTIJEN

De spelboom *complexiteit* is het aantal eindknopen in de spelboom gerekend vanaf de beginstelling. Dit aantal is het totaal aantal mogelijke partijen. Het kan berekend worden als de vertakkingsfactor tot de macht van het aantal halfzetten nodig om het spel te voltooien, d.w.z. om de eindknopen te bereiken. Toegepast op ons voorbeeld in Afbeelding 1: $2^4 = 16$. Voor het schaakspel is dit aantal niet exact uit te rekenen omdat 1) de vertakkingsfactor niet vanuit iedere stelling hetzelfde is en 2) het aantal halfzetten tot de eindknopen niet een constante is, m.a.w. niet iedere tak aan de spelboom is even lang.

Om toch een schatting te maken nemen we aan dat na iedere zet er 35 mogelijke tegenzetten zijn en de gehele partij 40 zetten (dus 80 halfzetten) lang is. Dan is de complexiteit $35^{2*40} \approx 10^{123}$, dat is een 1 met 123 nullen! Claude Shannon[1)], een Engelse schaakcomputer pionier, schatte de complexiteit op 10^{120}. Ter vergelijking: het aantal atomen in het waarneembare heelal is ruwweg geschat op 10^{80}.

HOOFDSTUK 4: AANTAL STELLINGEN

Een interessant vraag is: hoeveel legale stellingen bestaan er? Een mogelijke berekening is de volgende uitgaande van 32 stukken. Het aantal manieren waarop je 32 verschillende stukken op 64 velden kan plaatsen is:

$$\frac{64!}{(64-32)!} = 4.82 \cdot 10^{53}$$

Echter niet alle 32 stukken zijn verschillend: zowel voor wit als zwart zijn er 8 pionnen, 2 torens, 2 lopers en 2 paarden hetzelfde. Daarom wordt de formule:

$$\frac{64!}{(64-32)!\cdot(8!\cdot 2!\cdot 2!\cdot 2!)^2} = 4.63\cdot 10^{42}$$

Deze mogelijke stellingen bevatten illegale stellingen zoals witte pionnen op de eerste of zwarte pionnen op de achtste rij. Bovendien bestaan er natuurlijk stellingen met minder dan 32 stukken. Een berekening[2)] die hiervoor tracht te corrigeren geeft als resultaat 10^{41}.

HOOFDSTUK 5: TABLEBASES

Een berekening van de volledige spelboom door een computer uitgaande van de beginstelling is onmogelijk. De spelboom is eenvoudigweg te groot zodat zelfs de meest krachtige computer deze klus niet kan klaren. Ook zal er nooit een computer komen die dat wel kan.

In plaats van de beginstelling met 32 stukken op het bord, kunnen we beginstellingen met minder stukken op het bord beschouwen in de hoop dat we deze wel kunnen "kraken". Hiermee komen we op het terrein van de *eindspel databases*, ook wel *tablebases* genoemd. Een tablebase bevat iedere mogelijke stelling met de beste zet voor wit of voor zwart[3)]. Een tablebase wordt gemaakt met behulp van retrograde analyse vanuit een spelboom zoals toegelicht in Hoofdstuk 2. In 2005 waren de tablebases tot en met 6 stukken gereed; in 2012 de tablebase voor 7 stukken. We beginnen nu tegen de grenzen van de hardware aan te lopen. In onderstaande tabel is de grootte van de tablebases aangegeven.

Tabel 1: Eindspel databases.

Aantal stukken	**Grootte (MB)**	**Jaar**
3	0.06	1998
4	30	1998
5	7100	1998
6	1200000	2005
7	140000000	2012

Gegevens in de tablebase zijn gecomprimeerd volgens een format bedacht in 1998 door Eugene Nalimov, een Russisch schaakprogrammeur. Dit is tot op heden het meeste gebruikte format voor tablebases in schaakprogramma's. Een nieuwer opslagformat, genaamd Syzygy-format ontwikkeld door Ronald de Man (2013), een Nederlands wiskundige, geeft kleinere tablebases (factor 10-100) die al worden ondersteund door belangrijke schaakengines zoals Deep Fritz, Houdini en Stockfish. Syzygy-tablebases zijn er tot en met 6 stukken.

Je ziet dat de tablebase explosief groeit met ieder stuk erbij. Hoe groot zou een 32-stukken tablebase zijn? Ik heb op de *x*-as het aantal stukken uitgezet en op de *y*-as de 10-logarithme van de grootte van de tablebase in kB's. Zie Fig. 1.

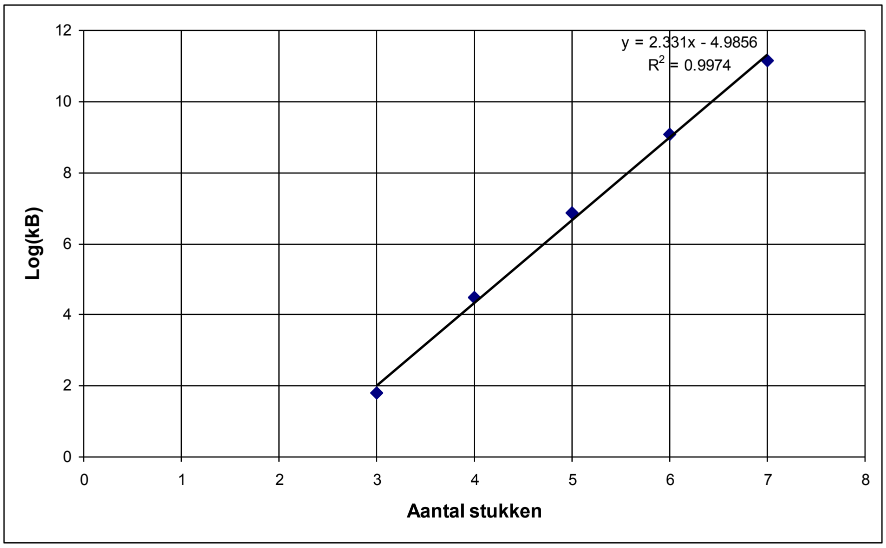


Fig. 1: Verband tussen aantal stukken en grootte van de tablebase.

Je kunt door de datapunten een rechte lijn trekken met vergelijking $y = 2.331x - 4.9856$. Nemen we $x = 32$ dan geeft dat $y = 69.6064$. Dus de geschatte grootte van een 32-stukken tablebase is afgerond 10^{69} kB = 10^{60} TB. "Mission impossible" zelfs met het compactere Syzygy-format. Om 10^{60} TB even in perspectief te plaatsen: de totale opslagcapaciteit bij grote data centers zoals Google, Facebook, Microsoft en Apple opgeteld, schat men in 2018 op ongeveer 10^{9} TB.

In mijn vorig artikel over eindspel databases[4)] gaf ik de voorspelling dat over 250 jaar de 32-stukken database klaar zou zijn. In het licht van bovenstaande bespiegelingen weten we nu dat dit niet gaat lukken. En dat is maar goed ook.

HOOFDSTUK 6: REFERENTIES

1) https://en.wikipedia.org/wiki/Shannon_number

2) http://www.nwchess.com/articles/misc/Chess_Board_Positions_article.pdf.

3) https://en.wikipedia.org/wiki/Endgame_tablebase

4) http://www.osseschaakvereniging.nl/artikelen/artikel1.php

Cees Timmer
Oss, 17 april 2018