Oavgörbara problem i elementär aritmetik

Erik Palmgren

31 augusti 2001

Många har nog slagits av hur svår talteori kan vara redan på den elementära nivån. Det är förhållandevis enkelt att råka skriva ner en Diofantisk ekvation där det inte finns någon känd metod för att bestämma om den har någon lösning. Fermats sista "sats" är ett bra exempel på detta, vilken man först efter 350 år lyckades bevisa. Ett objektivt skäl till att talteorin är svår är att den kan "koda" kombinatoriska problem. Detta upptäcktes, och användes på ett genialiskt sätt, av Kurt Gödel i hans berömda ofullständighetssats — det kanske viktigaste resultatet inom logiken.

Matematisk logik studerar väsentligen formerna för matematiska resonemang och beräkningar, och är en relativt ung gren av matematiken. Runt sekelskiftet 1900 genomgick matematiken den så kallade grundvalskrisen, som hade sitt upphov i paradoxer (Russells paradox, Burali-Fortis paradox) som upptäckts i alltför lättvindiga axiomatiseringar av mängdlära. För att säkerställa den nya axiomatiska metoden föreslog en av den tidens ledande matematiker, David Hilbert, att axiomatiska system och matematiska bevis själva skulle studeras med matematiska metoder. Syftet var att bevisa att inga paradoxer kunde uppstå i ett givet system, genom att undersöka formen hos de möjliga bevisen. För att detta skall vara meningsfullt måste naturligtvis de metoder med vilket detta görs själva vara fria från paradoxer eller motsägelser. Dödsstöten till detta Hilberts ambitiösa program kom när Kurt Gödel 1931 bevisade sin andra ofullständighetssats, som i grova drag säger att varje motsägelsefritt system som innehåller axiom för grundläggande talteori inte (ens) kan bevisa sin egen motsägelsefrihet. Ur den första ofullständighetssatsen framgick att redan elementär talteori kunde innehålla oavgörbara problem. Det senare resultatet stukade den rådande optimismen om matematikens obegränsade framstegsmöjligheter, som Hilbert hade formulerat i ordalagen *inom matematiken finns inget ovetbart.*

Vid den internationella matematikerkongressen 1900 presenterade Hilbert en lista av 23 problem för det kommande århundradet. Dessa problem har haft ett stort inflytande på matematikens utveckling. Tre av problemen

kan, åtminstone i efterhand, hänföras till logiken. Det andra, konsistensproblemet, kan i en mening sägas vara om inte löst, så "upplöst", av Gödels satser. Hilberts program lever dock vidare i modifierad form inom en gren av logiken, bevisteori, där man studerar om en teori är motsägelsefri relativt en annan, och försöker finna teorier vars motsägelsefrihet är evident i någon mening. Viktiga bidrag till detta progam har gjorts av den svenske logikern Per Martin-Löf. Det första problemet, om Cantors kontinuumhypotes, löstes 1963 av Paul Cohen genom hans bevis att hypotesen är oberoende av Zermelo-Fraenkels axiom (ZFC) för mängdläran. Det 10:e problemet i listan motstod länge lösningsförsöken:

> **10. Bestämning av lösbarheten av en Diofantisk ekvation.** *En Diofantisk ekvation med ett godtyckligt antal obekanta och med rationella heltalskoefficienter är given: ange ett förfarande, med vilket det efter ett ändligt antal operationer kan bestämmas om ekvationen är lösbar i rationella heltal.* (Författaren översättning.)

De (rationella) heltalen är $\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. I modernt språkbruk efterfrågar Hilbert en algoritm för att avgöra lösbarheten över $\mathbf{Z}$. Det Diofantiska problemet över $\mathbf{Z}$ med en obekant kan skrivas som

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$$

där koefficienterna är heltal. Man kan förstås använda allmänna metoder för lösning av sådana ekvationer över komplexa tal, men för heltalslösningar finns enklare knep att tillgå. En enkel uppskattning av storleken på lösningarna z ges av

$$|z| \leq 1 + |a_n|^{-1} \max(|a_n|, |a_{n-1}|, \ldots, |a_1|, |a_0|).$$

Det finns endast ändligt många heltal z som uppfyller olikheten. En metod för att för att avgöra lösbarheten i $\mathbf{Z}$ är helt enkelt att testa dessa heltal. Denna metod kan inte direkt användas för att bestämma lösbarheten över de rationella talen $\mathbf{Q}$ eftersom det finns oändligt många rationella tal z som uppfyller olikheten. Om vi dock betraktar en rationell lösning $z = k/m$ skriven på förkortad form, det vill säga största gemensamma delaren $\mathrm{sgd}(k, m) = 1$, så inser man genom en insättning i ekvationen att nämnaren m måste dela den ledande koefficienten a_n. Eftersom det endast finns ett begränsat antal z, där nämnaren uppfyller detta villkor, och olikheten, så kan vi använda metoden med systematisk testning.

Lösbarhet över $\mathbf{Z}$ kan även avgöras för en andragradsekvation med godtyckligt många obekanta. Detta använder ett djupt resultat av C.-L. Siegel från 1972.

Byggandes på partiella resultat av Martin Davis, Hilary Putnam och Julia Robinson bevisade Juri Matijasevic 1971 att det allmänna problemet är algoritmiskt oavgörbart.

Efter ett par inledande avsnitt som förklarar fundamentala logiska begreppp, skall vi försöka ge en idé om beviset och dess talteoretiska ingredienser. Dock finns det överraskande positiva resultat om avgörbarhet i aritmetik och analytisk geometri (se avsnitt 4) som visar att automatisk problemlösning med dator teoretiskt sett har en stor räckvidd. Dessa metoder finns allmänt tillgängliga i vissa moderna datoralgebrasystem och kan tillämpas på enklare problem.

1 Elementära problem

Vi skall introducera begreppet elementärt (eller första ordningens) matematiskt språk. "Elementärt" syftar här på att de objekt som kan diskuteras i språket alla är element i en fixerad mängd D. Godtyckliga delmängder av D faller utanför språkets uttrycksmöjligheter. Vi kommer huvudsakligen att betrakta fallen då D är någon av följande mängder, de naturliga talen $\mathbf{N} = \{0, 1, 2, 3, \ldots\}$, mängden av hela tal $\mathbf{Z}$, mängden av rationella tal $\mathbf{Q}$ eller mängden av reella tal $\mathbf{R}$. I samtliga dessa mängder har de aritmetiska operationerna $+$ och $\cdot$ en överensstämmande mening. Ett påstående som

$$\textit{Det finns ett } x \in D \textit{ sådant att } y = x \cdot x \qquad (1)$$

har därför någon mening i var och en av de nämnda talmängderna, fast olika mening! För $D = \mathbf{N}, \mathbf{Z}$ säger påståendet att y är en perfekt kvadrat, för $D = \mathbf{R}$ betyder det endast att y är icke-negativ. Ett gemensamt elementärt språk bestäms av $L = \{+, \cdot, 0, 1\}$. En *term* över L är ett välbildat uttryck som kan skrivas med hjälp av operationerna, $+$ och $\cdot$, samt variabel-symboler $x, y, z, u, v, \ldots$, konstanterna $0, 1$ och parenteser. En sådan term kan betraktas som ett polynom med naturliga tal som koefficienter. Här är ett exempel på en term och motsvarande normaliserade polynom:

$$(x \cdot (x \cdot x)) + (((1 + 1) \cdot x) \cdot y) \qquad x^3 + 2xy$$

Ett *atomärt påstående* över L är en ekvation av formen $r = t$, där r och t är termer över L. De atomära påståenden kombineras sedan till *elementära påståenden* över L genom att använda logiska konnektiv och kvantifikatorer.

Vi påminner om symbolismen för dessa: $P \wedge Q$ betecknar P *och* Q, $P \vee Q$ betecknar P *eller* Q, $\neg P$ står för *icke* P. *Kvantifikatorerna* (eller *kvantorerna*) är $\forall$ (universalkvantorn) och $\exists$ (existenskvantorn). Existenspåståendet *det finns ett* $x \in D$ *så att* $P(x)$ uttrycks $(\exists x \in D)\, P(x)$, medan universalpåståendet *för alla* $x \in D$ *gäller* $P(x)$ uttrycks $(\forall x \in D)\, P(x)$. Detta är alla tillåtna sätt att bilda elementära påståenden över L. Påståendet (1) blir i logisk symbolism

$$(\exists x \in D)\, y = x \cdot x. \qquad (2)$$

En variabel som inte ligger inom en kvantifierad variabels räckvidd kallas *parameter*. Den enda parametern i (2) är y. Nya begrepp kan definieras i termer av elementära påståenden med parametrar:

$$\begin{aligned}
x < y \quad &\Leftrightarrow_{\text{def}} \quad (\exists z \in \mathbf{N})\, x + z + 1 = y \\
r = \text{Rest}(m, n) \quad &\Leftrightarrow_{\text{def}} \quad (\exists k \in \mathbf{N})\, m = kn + r \wedge (0 = r \vee 0 < r) \wedge r < n \\
k = \text{Kvot}(m, n) \quad &\Leftrightarrow_{\text{def}} \quad (\exists r \in \mathbf{N})\, m = kn + r \wedge (0 = r \vee 0 < r) \wedge r < n \\
p \text{ primtal} \quad &\Leftrightarrow_{\text{def}} \quad p > 1 \wedge (\forall m \in \mathbf{N})\, [1 < m < p \rightarrow \text{Rest}(p, m) > 0]
\end{aligned}$$

Mer allmänt kan man tala om en *struktur*

$$\mathcal{D} = \langle D; R_1, R_2, \ldots, f_1, f_2, \ldots, c_1, c_2, \ldots \rangle$$

vilken består av en mängd D och där R:n är relationer på D, f:n är operationer på D, och c:na är konstanter i D. Ett *elementärt påstående* över denna får då använda konstanterna, operationerna och variabler för att bygga upp termer, och de atomära påståendena får bildas med relationerna förutom $=$. Till exempel, om R_1 är relationen $<$ så är olikheten $r < t$ ett atomärt påstående närhelst r och t är termer. Vi skriver $\mathcal{D} \models P(x_1, \ldots, x_s)$ för att beteckna att ett elementärt påstående P över $\mathcal{D}$ är sant för parametrarna $x_1, \ldots, x_n$. I denna notation skriver man inte ut D i kvantoruttryck som $\forall x \in D$ och $\exists x \in D$, eftersom dessa kan utläsas ifrån $\mathcal{D}$. Exempelvis betyder $\langle \mathbf{N}; +, \cdot, 0, 1 \rangle \models (\exists z)\ x + z + 1 = y$ det samma som $(\exists z \in \mathbf{N})\ x + z + 1 = y$

En relation R på en mängd D brukar identifieras med en delmängd R av D^n för ett fixerat n. Man skriver omväxlande $R(x_1, \ldots, x_n)$ eller $(x_1, \ldots, x_n) \in R$ för att uttrycka att relationen R gäller mellan elementen $x_1, \ldots, x_n$. Relationen $R \subseteq D^n$ är *definierbar över* $\mathcal{D}$ om det finns ett elementärt påstående $P(x_1, \ldots, x_n)$ över $\mathcal{D}$ så att $R(x_1, \ldots, x_n)$ om, och endast om, $\mathcal{D} \models P(x_1, \ldots, x_n)$. Exempelvis är $<$ definierbar över $\langle \mathbf{N}; +, \cdot, 0, 1 \rangle$, se ovan. En funktion $f : D^n \rightarrow D$ är definierbar om relationen $f(x_1, \ldots, x_n) = x_{n+1}$ är definierbar.

Det fundamentala begreppet vi skall intressera oss för är detta. Vi säger att en struktur $\mathcal{D}$ är *avgörbar* om det finns en algoritm som för varje elementärt påstående P över $\mathcal{D}$, utan parametrar, avgör om P är sann eller falsk. Vi kommer att betrakta strukturerna $\langle \mathbf{N}; +, \cdot, 0, 1\rangle$, $\langle \mathbf{Z}; +, \cdot, 0, 1\rangle$, $\langle \mathbf{Q}; +, \cdot, 0, 1\rangle$, $\langle \mathbf{R}; <, +, \cdot, 0, 1\rangle$ och några till. Det är i dessa fall klart att elementära påståenden kan ges en precis syntax så att de kan läsas in i en dator och behandlas algoritmiskt.

2 Definierbarhet i de hela och rationella talen

Vad som kan uttryckas i det elementära språket över en struktur $\mathcal{D}$ bestämmer förstås hur svåra problem som kan formuleras. De tre talsystemen $\mathbf{N}$, $\mathbf{Z}$ och $\mathbf{Q}$ visar sig vara i det närmaste likvärdiga i det avseendet. Ett elementärt påstående över i $\langle \mathbf{Z}; +, \cdot, 0, 1\rangle$ kan översättas till ett ekvivalent påstående i $\langle \mathbf{N}; +, \cdot, 0, 1\rangle$ genom att betrakta ett heltal som ett par av naturliga tal (varvid deras skillnad är heltalet). På motsvarande sätt kan en översättning från $\langle \mathbf{Q}; +, \cdot, 0, 1\rangle$ till $\langle \mathbf{N}; +, \cdot, 0, 1\rangle$ erhållas genom att betrakta trippler (m, n, k) av naturliga tal som rationella tal $(m - n)/(k + 1)$. Detaljerna i detta är inte särskilt svåra. Översättningar i motsatt riktning kräver djupare resultat i talteori. Om man kan finna ett elementärt påstående $N(x)$ så att $\langle \mathbf{Z}; +, \cdot, 0, 1\rangle \models N(x)$ gäller precis då x är ett naturligt tal, kan vi enkelt översätta ett påstående i $\mathbf{N}$ till ett ekvivalent påstående i $\mathbf{Z}$, genom att systematiskt begränsa kvantorernas och parametrarnas utsträckning med $N(x)$. En sådant påstående $N(x)$ kan erhållas genom att utnyttja Lagranges berömda sats att varje naturligt tal är summan av fyra perfekta kvadrater.

$$N(x) =_{\text{def}} \exists x_1\, \exists x_2\, \exists x_3\, \exists x_4\; x_1^2 + x_2^2 + x_3^2 + x_4^2 = x.$$

(Observera att, exempelvis, talet 7 ej kan uttryckas som summan av tre perfekta kvadrater.) Vi kan även använda ett annat resultat av Lagrange om vad som brukar kallas *Pells ekvation*

$$x^2 - Dy^2 = 1.$$

(Denna ekvation spelar även en roll i lösningen av Hilberts 10:e problem.)

Sats (Lagrange 1768) *Om D ej är en perfekt kvadrat, så har Pells ekvation en lösning där x och y är positiva heltal.*

Man kan nu utan besvär kontrollera att påståendet $N(z)$:

$$(\exists u\, z = u^2) \vee \exists x \exists y\, (x^2 = zy^2 + 1 \wedge \neg x = 0 \wedge \neg y = 0)$$

definierar de naturliga talen i $\langle \mathbf{Z}; +, \cdot, 0, 1 \rangle$.

En betydligt svårare reduktion är den från $\langle \mathbf{Z}; +, \cdot, 0, 1 \rangle$ till $\langle \mathbf{Q}; +, \cdot, 0, 1 \rangle$, som först bevisades av Julia Robinson, se Smoryński (1991). I denna utnyttjas djupare resultat om kvadratiska former (Hasse-Minkowskis sats).

3 Oavgörbara problem

Alan Turing gav 1936 ett övertygande argument för allt som kan beräknas eller avgöras algoritmiskt kan beräknas eller avgöras av hans Turingmaskiner. En Turingmaskin är en matematisk modell av en dator som består av en centralenhet och ett "sekundärminne" i form av ett oändligt långt band. Vid varje tidpunkt används dock bara en ändlig del av bandet. Indata skrivs på bandet. Maskinen startas, och om den stannar betraktas innehållet på bandet som utdata. En universell Turingmaskin tar en beskrivning e av en annan maskin M och indata x och simulerar sedan verkan av M med indata x. Man kan alltså betrakta e som (simulerings)programmet för M.

Sats (Turing 1936) *Stopproblemet är inte algoritmiskt avgörbart: det finns ingen Turingmaskin som avgör om Turingmaskinen e med indata x stannar efter ett ändligt antal steg.*

Beviset av Gödels ofullständighetssats innehöll många viktiga idéer om hur syntax och beräkningar kan kodas i talteori.

Gödels kodningslemma *Det finns en funktion $\beta : \mathbf{N}^3 \to \mathbf{N}$ definierbar över $\langle \mathbf{N}; +, \cdot, 0, 1 \rangle$ sådan att för varje ändlig följd $a_1, \ldots, a_n$ av naturliga tal, finns naturliga tal x och d så att*

$$\beta(x, d, i) = a_i \qquad (i = 1, \ldots, n).$$

Om vi antar att exponentialfunktionen x^y är definierbar över $\langle \mathbf{N}; +, \cdot, 0, 1 \rangle$, så kan β konstrueras genom $\beta(x, d, i) = \mathrm{Rest}(\mathrm{Kvot}(x, d^i), d)$. Sätt nu $x = a_n d^{n-1} + \cdots + a_2 d + a_1$ där $d = \max(a_1, \ldots, a_n) + 1$. Då gäller $\beta(x, d, i) = a_i$. Följden kodas alltså som siffror i ett nummer med basen d:

$$x = (a_n \; a_{n-1} \; \cdots \; a_2 \; a_1)_d.$$

Emellertid är det inte alls uppenbart att exponentialfunktionen är definierbar över $\langle \mathbf{N}; +, \cdot, 0, 1 \rangle$. Gödel använde istället den *kinesiska restsatsen* för sitt bevis. Låter man $d = (an)!$, där $a = \max(a_1, \ldots, a_n) + 1$, så är $1 + d, 1 + 2d, \ldots, 1 + nd$ parvis relativt prima. Den nämnda satsen ger då ett naturligt tal x med $x \equiv a_i \pmod{1 + id}$ för alla $i = 1, \ldots, n$. Vi har $a_i \leq d < 1 + id$, så $a_i = \mathrm{Rest}(x, 1 + id)$. Man kan alltså låta

$$\beta(x, d, i) = \mathrm{Rest}(x, 1 + id),$$

som uppenbarligen är definierbar över $\langle \mathbf{N}; +, \cdot, 0, 1 \rangle$.

Exponentialfunktionen kan nu definieras genom

$$\begin{aligned} y = b^n \quad \Leftrightarrow_{\mathrm{def}} \quad & (\exists x, d \in \mathbf{N})\, y = \beta(x, d, n + 1) \wedge \\ & \beta(x, d, 1) = 1 \wedge \\ & (\forall i < n)\, \beta(x, d, i + 2) = b \cdot \beta(x, d, i + 1). \end{aligned} \tag{3}$$

På liknande sätt kan andra rekursiva talföljder definieras, faktiskt alla så kallade *primitivt rekursiva funktioner*. Dessa funktioner kan identifieras med enkla datorprogram som endast använder addition, multiplikation, villkorssatser (IF-THEN-ELSE) och begränsade slingor (FOR-loopar) — ett sådant program stannar alltid! Det är välkänt att det finns en primitivt rekursiv funktion $T(e, x, y)$ så att $T(e, x, y) = 1$ om y beskriver en korrekt termierande beräkning i Turingmaskinen e med indata x, och $T(e, x, y) = 0$ i annat fall. Vi ser nu att stopproblemet kan uttryckas genom $(\exists y \in \mathbf{N})\, T(e, x, y) = 1$, så det är definierbart över $\langle \mathbf{N}; +, \cdot, 0, 1 \rangle$. Av detta följer från Turings sats

Sats Strukturen $\langle \mathbf{N}; +, \cdot, 0, 1 \rangle$ är oavgörbar.

En mängd $S \subseteq \mathbf{N}^k$ kallas *rekursivt uppräknelig* om det finns en algoritm $n \mapsto s_n$ som räknar upp (eller genererar) alla element i S. (Upprepningar är tillåtna.)

Något överraskande är mängden S (stoppmängden) av par (e, x) sådana att Turingmaskinen e stannar med indata x rekursivt uppräknelig. Låt e_0 vara en maskin som stannar för varje indata. Följande algoritm generar S: Givet (e, x, y), om $T(e, x, y) = 1$ svara (e, x), annars svara (e_0, x).

Hilberts 10:e problem

En *Diofantisk relation* $R \subseteq \mathbf{N}^k$ är en relation som kan definieras genom

$$R(x_1, \ldots, x_k) \quad \Leftrightarrow_{\text{def}} \quad (\exists y_1, \ldots, y_n \in \mathbf{N}) \\ p(x_1, \ldots, x_k, y_1, \ldots, y_n) = q(x_1, \ldots, x_k, y_1, \ldots, y_n)$$

där p och q är polynom med koefficienter i $\mathbf{N}$. Alternativt kan vi säga att R är projektionen av lösningarna till ekvationen $p = q$ i de första k koordinaterna. En funktion $f : \mathbf{N}^k \to \mathbf{N}$ kallas Diofantisk om relationen $f(x_1, \ldots, x_k) = x_{k+1}$ är Diofantisk. Addition, multiplikation och restfunktionen $(m, n) \mapsto \text{Rest(m, n + 1)}$ är exempel på sådana funktioner. Relationen R kallas *exponentiellt Diofantisk* om p och q får vara godtyckliga termer bildade av variabler, konstanter $0, 1$, addition, multiplikation och exponentiering x^y. Exempel: $p = x_1^{1+x_1 \cdot y_1} + x_2$.

Martin Davis och Julia Robinson hade visat viktiga delresultat på vägen mot lösningen av Hilberts 10:e problem. Tillsammans med Hilary Putnam lyckades de visa

Davis-Putnam-Robinson satsen (1961) *Varje rekursivt uppräknelig mängd (eller relation) är exponentiellt Diofantisk.*

Det svåraste problemet återstod att lösa: att visa att exponentialfunktionen x^y är Diofantisk. (Definitionen (3) ovan är långt ifrån Diofantisk.) Uppfattningen bland vissa matematiker var att Davis-Putnam-Robinson satsen inte hade något med Hilberts 10:e problem att göra, eftersom exponentialfunktionen spelade en så central roll. Juri Matijasevic lyckades 1971 ge en Diofantisk definition av exponentialfunktionen. En metod som ligger nära hans ursprungliga är att använda sig av rötterna till Pells ekvation. Om $x_1, y_1 > 0$ är en lösning till $x^2 - Dy^2 = 1$, så är även alla x_n, y_n givna av

$$x_n + y_n\sqrt{D} = (x_1 + y_1\sqrt{D})^n \qquad (n = 1, 2, 3, \ldots)$$

lösningar. Väljer man $x_1, y_1 > 0$ att vara den lösning där $x_1 + y_1\sqrt{D}$ är minimal, så kan man visa att (x_n, y_n) utgör samtliga positiva lösningar. Det är lätt att se att lösningarna växer exponentiellt med n. Speciella egenskaper hos Pells ekvation för $D = a^2 - 1$, där $a \geq 2$, gör det nu möjligt att erhålla y_n som en Diofantiskt definierbar funktion av n. Från denna funktion kan den vanliga exponentialfunktionen erhållas. Konstruktionerna är intrikata, se (Smoryński 1991) för en detaljerad framställning.

Matijasevic sats (1971) *Varje rekursivt uppräknelig mängd (eller relation) är Diofantisk.*

Detta visar att Hilberts 10:e problem är oavgörbart eftersom stoppmängden S är rekursivt uppräknelig. Resultatet följer därmed av att stopproblemet är oavgörbart:

Följdsats *Hilberts 10:e problem är oavgörbart.*

Det är känt att man kan avgöra om ett system av linjära ekvationer har en heltalslösning. I kontrast till detta, visar satsen att det inte är möjligt att avgöra om ett *system* $p_1 = 0, \ldots, p_m = 0$ av andragradsekvationer med flera obekanta är lösbart, eftersom termer av högre totalgrad i en ekvation $q = 0$ kan ersättas av andragradstermer till kostnaden av att införa nya variabler och ekvationer. Detta system är i sin tur ekvivalent med den enda ekvationen $p_1^2 + \cdots + p_n^2 = 0$, vilket betyder att Hilberts 10:e problem är oavgörbart för fjärdegradspolynom. Vad gäller tredjegradsekvationer tycks problemet fortfarande öppet (Smoryński 1991).

Det finns även positiva konsekvenser av Matijasevics resultat. Exempelvis finns ett polynom $r(x_1, \ldots, x_n)$ med heltalskoefficienter sådant att dess positiva värden, när $x_1, \ldots, x_n$ löper över de naturliga talen, är precis primtalen. De negativa värdena ignoreras. Det finns alltså en aritmetisk formel för att generera alla primtal utan att använda division! Mängden P av primtal är rekursivt uppräknelig eftersom det finns en välkänd algoritm som genererar dem: Erathostenes såll. Därmed är denna mängd Diofantisk, och det finns alltså ett polynom $p(x_1, \ldots, x_n, y)$ med heltalkoefficienter så att y primtal om, och endast om, $(\exists x_1, \ldots, x_n \in \mathbf{N})\, p(x_1, \ldots, x_n, y) = 0$. Nu kan r definieras genom ett enkelt trick:

$$r(x_1, \ldots, x_n, y) = (y+1)(1 - p(x_1, \ldots, x_n, y)^2) - 1.$$

Ett utförligt exempel på ett sådant polynom finns angivet i Davis mfl. (1976).

Hilberts 10:e problem där lösningar i rationella tal tillåts är fortfarande öppet (2000). Det är ej heller känt om $\mathbf{Z}$ kan definieras genom en Diofantisk relation över $\mathbf{Q}$.

4 Avgörbara problem

Aritmetik utan multiplikation. Stora framsteg gjordes på 1920-talet rörande avgörbarhetsproblem. M. Presburger och Thoralf Skolem bevisade oberoende av varandra att den elementära strukturen $\langle \mathbf{Z}; +, <, 0, 1\rangle$, dvs heltalen utan multiplikation, är avgörbar. Trots att multiplikation saknas kan många icke-triviala problem formuleras i dess elementära språk, till exempel lösbarheten av ett system av linjära olikheter över heltalen. Detta problem brukar gå under namnet *heltalsprogrammering* och har tillämpningar på olika typer av schemaläggning. Det är ett NP-fullständigt problem och klassificeras inom komplexitetsteorin som ett praktiskt ogörligt problem, åtminstone för stora mängder indata.

Den metod Presburger och Skolem utnyttjade för avgörbarhetsresultat var så kallad *kvantorelimination*. Att direkt kontrollera om $(\forall x \in D)\, P(x)$ är sann skulle innebära att testa om $P(x)$ gäller för varje $x \in D$. Detta är förstås inte möjligt när D är oändlig. Om man steg för steg kan eliminera kvantorerna $\forall$ och $\exists$ genom omskrivning, finns det möjlighet att slutresultatet är avgörbart.

Skolems metod bestod i att först betrakta en utvidgad struktur $\mathcal{Q} = \langle \mathbf{Q}; +, <, 0, 1, \{q\cdot\}_{q\in\mathbf{Q}}, \lfloor\ \rfloor\rangle$. Här är $q\cdot$ operationen att multiplicera med en *konstant* q ($q \in \mathbf{Q}$). Dessa benämns här *multiplikatorer*. Vi skriver ofta q för $q \cdot 1$. $\lfloor x \rfloor$ betecknar det största heltalet n sådant att $n \leq x$, ibland uttalat *golvet till* x. Notera att $\lfloor x \rfloor = x$ är ekvivalent med att $x \in \mathbf{Z}$, så att ett påstående om den ursprungliga strukturen $\langle \mathbf{Z}; +, <, 0, 1\rangle$ kan översättas till ett påstående om $\mathcal{Q}$. Man observerar först att en kvantorfri elementär utsaga (utan variabler) om $\mathcal{Q}$ helt enkelt består av likheter och olikheter mellan enkla aritmetiska uttryck (t.ex. $0+\lfloor 1/3\cdot 1+1\rfloor < (3/5)\cdot(1+1)$) kombinerade med logiska konnektiv. Sådana utsagor kan därmed avgöras med enkel satslogik. Problemet är nu visa att varje öppen utsaga $P(x_1, \ldots, x_n)$, där kvantorerna begränsas till $\mathbf{Z}$, är ekvivalent med en kvantorfri utsaga $R(x_1, \ldots, x_n)$. Kvantorerna elimineras inifrån, och genom logiska överväganden inser man att det räcker att lösa problemet för utsagor P av formen

$$(\exists y \in \mathbf{Z})\, A_1(x_1, \ldots, x_n, y) \wedge \cdots \wedge A_m(x_1, \ldots, x_n, y)$$

där varje A_i är ett atomärt påstående. Ett enkelt exempel på eliminationsprocessen är följande. Betrakta

$$(\exists y \in \mathbf{Z})\, 1 < \lfloor y/3 + x \rfloor \wedge y + \lfloor y/2 \rfloor < 3.$$

Låt $B(x, y)$ beteckna påståendet efter existenskvantorn. För att kunna lösa ut variabeln y noterar vi de heltal den divideras med i B. Produkten av dessa är $6 = 3 \cdot 2$. Vi kan nu skriva $(\exists y \in \mathbf{Z})\, B(x, y)$ som

$$(\exists y \in \mathbf{Z})\, B(x, 6y) \vee (\exists y \in \mathbf{Z})\, B(x, 6y + 1) \vee \cdots \vee (\exists y \in \mathbf{Z})\, B(x, 6y + 5)$$

Dvs man delar upp existenspåståendet i 6 olika fall beroende på modulus. $B(x, 6y + k)$ är (för y heltal) ekvivalent med $1 < \lfloor 2y \rfloor + \lfloor k/3 + x \rfloor \wedge 6y + k + \lfloor 3y \rfloor + \lfloor k/2 \rfloor < 3$ dvs $(1 - \lfloor x + k/3 \rfloor)/2 < y \wedge y < (3 - (k + \lfloor k/2 \rfloor))/9$. Men vi kan nu eliminera existenskvantorn genom att utnyttja att $(\exists n \in \mathbf{Z})\, a < n < b$ är ekvivalent med $\lfloor a \rfloor + 1 < b$. Följaktligen är $(\exists y \in \mathbf{Z})\, B(x, 6y + k)$ ekvivalent med

$$\lfloor (1 - \lfloor x + k/3 \rfloor)/2 \rfloor + 1 < (3 - (k + \lfloor k/2 \rfloor))/9.$$

Det ursprungliga existenspåståendet gäller alltså om, och endast om, denna olikhet gäller för något $k = 0, 1, 2, 3, 4, 5$. Den oändliga kvantifikationen över $\mathbf{Z}$ har alltså ersatts av en kvantifikation över $\{0, 1, 2, 3, 4, 5\}$, vilken kan uttryckas med en satslogisk disjunktion. För en framställning av det allmänna fallet se Smoryński (1991). Förfarandet upprepas tills dess att inga kvantorer återstår. Märk hur golvfunktionen och nya multiplikatorer introduceras under dess lopp. Komplexiteten hos påståendena kan dessvärre växa exponentiellt när kvantorerna elimineras.

Elementär analytisk geometri. Alfred Tarski bevisade under 1930-talet att elementära påståenden över strukturen $\langle \mathbf{R}; +, \cdot, <, 0, 1 \rangle$ är avgörbara. Detta kan tyckas vara ett förvånande resultat eftersom $\mathbf{R}$ har en mycket mer komplicerad uppbyggnad än till exempel $\mathbf{Z}$. Emellertid kan $\mathbf{Z}$ inte definieras i strukturen. Tarskis metod för bevisa avgörbarheten var återigen kvantorelimination. Vi skall kort beskriva några av idéerna i denna. Över de reella talen gäller som bekant satsen om mellanliggande värden: för kontinuerliga funktioner f på intervallet $[a, b]$

$$f(a) \leq 0 \leq f(b) \vee f(b) \leq 0 \leq f(a) \Longrightarrow (\exists x)\, f(x) = 0 \wedge a \leq x \leq b. \tag{4}$$

Betrakta fallet då f är ett polynom utan multipla nollställen. Om man vet att f har högst ett nollställe i intervallet, så gäller även den omvända implikationen i (4). Detta eliminerar existenskvantorn. En välkänd algoritmisk metod enligt Sturm ger ett sätt att isolera samtliga rötter till $f(x) = 0$ inom rationella interval $[a_1, b_1], [a_2, b_2], \ldots, [a_n, b_n]$, så att varje interval innehåller högst en rot. Med hjälp av denna kan kvantorn i $\exists x\, f(x) = 0$ elimineras, även då x är obegränsad. Fallet med multipla rötter kan behandlas genom

att betrakta derivatorna av f. Den generella metoden är ganska komplicerad. Stora förbättringar av Tarskis ursprungliga metod har gjorts av George Collins, se Caviness och Johnson (1998). Tidsåtgången för den bästa kända algoritmen för kvantorelimination (se Marker 1996) är dock dubbelt exponentiell i antalet kvantorväxlingar $\forall\exists$ och $\exists\forall$ det finns i problemet när det ställs på Prenex-form.

Följande problem och satser kan tämligen direkt formuleras som elementära utsagor över $\langle \mathbf{R}; +, \cdot, <, 0, 1\rangle$ och är därmed enligt Tarskis resultat algoritmiskt lösbara eller bevisbara.

(a) Triangelolikheten och Cauchy-Schwarz' olikhet för det Euklidiska rummet $\mathbf{R}^n$.

(b) Bestäm det minsta tal c sådant att

$$(x-a)(b-x) \le c(b-a)^2 \qquad (a \le x \le b).$$

(c) Bestämning av asymptoter: Givet polynom $p(x)$ och $q(x)$, finn λ sådan att

$$(\forall \varepsilon > 0)\,(\exists d > 0)\,(\forall x > d)\,\left|\frac{p(x)}{q(x)} - \lambda\right| < \varepsilon.$$

(d) Finn villkor på talen p, q och r så att

$$(\forall x)\, x^4 + px^2 + qx + r \ge 0.$$

(e) Givet en n-dimensionell sfär av radie 1. Hur många punkter kan placeras ut på denna så att inget par av punkter ligger närmare varandra än 1? (Euler, $n = 3$)

(f) Hur många enhetssfärer kan packas i en kubisk låda med sidan 10? (Sfärpackningsproblemet)

(g) Ett labyrintiskt konstgalleri skall övervakas. Antag att begränsningsytorna bestäms av ett system av algebraiska olikheter. Hur många över-vakningskameror ($\le n$) behövs för att varje skrymsle av galleriet skall kunna ses från någon av dem?

(h) Rörelseplanering: En robot skall patrullera galleriet i räta linjer. Hur många vändpunkter ($\le n$) behövs för att den under vägen skall kunna övervaka alla utrymmen och samtidigt undvika att krocka med väggar eller föremål.

Vi påpekar att n skall vara ett fixerat positivt heltal i (a), (e), (g) och (h). För problem (e) och (f) är det möjligt att finna övre gränser för antalet sfärer och punkter genom att utnyttja Dirichlets lådprincip.

Kvantorelimination i praktiken. Kvantoreliminationsmetoden för reella tal $\langle \mathbf{R}; +, \cdot, <, 0, 1 \rangle$ kan programmeras för bruk på persondatorer. Det finns ett antal experimentella implementationer (Caviness och Johnson 1998). Sådana rutiner ingår även i det populära datoralgebrasystemet *Mathematica* (Strzebonski 2000). Exempelvis kan problemen (b), (c) och (e) (för cirklar) snabbt lösas automatiskt med hjälp av detta.

```
In[3]:=
<< Developer`; << Experimental`;

In[4]:=
Resolve[ForAll[{a, b, x}, a ≤ x && x ≤ b && c ∈ Reals,
  (x - a) * (b - x) ≤ c (b - a)^2], {c}]
```

$$\text{Out[4]=}\quad c \geq \frac{1}{4}$$

Transcendentala funktioner. En intressant fråga är om Tarskis resultat kan utvidgas till transcendentala funktioner. Detta är ett aktivt forskningsområde med anknytning till svåra algebraiska problem. Ett enkelt negativt resultat för strukturen $\langle \mathbf{R}; +, \cdot, <, 0, 1, \sin \rangle$ är följande. Nollställena till $\sin x$ är heltalsmultipplerna av π, så varje rationellt tal kan skrivas som kvoten av två sådana nollställen. Följande påstående definierar de rationella talen över $\mathbf{R}$: a rationell om och endast om

$$(\exists x, y, z \in \mathbf{R})\ \sin^2 x + \sin^2 y + (x - z^2 - 1)^2 + (ax - y)^2 = 0.$$

Därav följer att strukturen oavgörbar.

Det är ännu ett öppet problem om utvidgningen $\langle \mathbf{R}; +, \cdot, <, 0, 1, \exp \rangle$ med exponentialfunktionen exp är avgörbar. Macintyre och Wilkie (1996) har dock visat avgörbarhet under antagande att en välkänd förmodan inom transcendental talteori är sann:

Schanuels förmodan. *Om $\alpha_1, \ldots, \alpha_n$ är reella tal, linjärt oberoende över* $\mathbf{Q}$*, så är transcendensgraden av kroppsutvidgningen* $\mathbf{Q}(\alpha_1, \ldots, \alpha_n, e^{\alpha_1}, \ldots, e^{\alpha_n})$ *över* $\mathbf{Q}$ *minst* n*.*

Märkligt nog är ingen generell algoritm känd (Marker 1996) för att avgöra variabelfria påståenden över $\langle \mathbf{R}; +, -, \cdot, <, 0, 1, \exp \rangle$, till exempel likheter som

$$e^{e^2-1} - e^5 = e^{6+e^{-3}}.$$

(Just denna är dock falsk.)

Litteratur

B.F. Caviness och J.R. Johnson (red.) (1998). *Quantifier Elimination and Cylindrical Algebraic Decomposition.* Springer-Verlag.

Martin Davis, Yuri Matiyasevich och Julia Robinson (1976). Hilbert's tenth problem. Diophantine equations: positive aspects of a negative solution. I: F.E. Browder (red.) *Mathematical developments arising from Hilbert problems.* American Mathematical Society, Rhode Island.

Angus Macintyre och Alex J. Wilkie (1996). On the decidability of the real exponential field. I: P. Odifreddi (red.) *Kreiseliana: about and around Georg Kreisel.* A.K. Peters.

David Marker (1996). Model Theory and Exponentiation. *Notices of the American Mathematical Society* 43:7, pp 753-759.

Yuri Matiyasevich (1993). *Hilbert's Tenth Problem.* The MIT Press, London.

Tryggve Nagell (1951). *Introduction to Number Theory.* Almqvist och Wiksell, Uppsala.

Craig Smoryński (1991). *Logical Number Theory I.* Springer-Verlag.

Adam Strzebonski (2000). Solving Algebraic Inequalities. *The Mathematica Journal 7:4,* pp 525 – 541.

Alfred Tarski (1951) *A Decision Method for Elementary Algebra and Geometry.* University of California Press. Nytryck i Caviness och Johnson (1998).