KAPITEL 6 - FÖRDJUPAD TALTEORI

1. Fördjupat studium av restklasser som matematiska objekt

Vi har alltså introducerat *restklassen* hörande till ett visst modulos n som mängden av alla tal som ger en viss rest vid division med n. För ett visst n uppkommer då också n stycken restklasser som vi noterade med $\overline{0}, \overline{1}, \overline{2}, \ldots$. (Ordet "modulos" betyder bara ett fixt positivt heltal.)

Exempel:

1. Alla jämna tal: $\{\ldots, -4, -2, 0, 2, 4, 6, 8, \ldots\} = \{2k; k \in \mathbb{Z}\}$. Denna restklass skrivs då som $\overline{0}$, när vi har modulos 2 underförstått. (Jämna tal är ju de som är jämnt delbara med 2 som alltså ger rest 0 vid division med 2.)
2. Alla udda tal: $\{\ldots, -3, -1, 1, 3, 5, 7, \ldots\} = \{2k + 1; k \in \mathbb{Z}\}$. Denna restklass skrivs då som $\overline{1}$, när vi har modulos 2 underförstått. (Udda tal är de som inte är jämnt delbara med 2 som alltså ger rest 1 vid division med 2.)
3. Alla tal som ger rest 3 vid division med 5: $\{\ldots, -12, -7, -2, 3, 8, 13, 18, \ldots\} = \{5k + 3; k \in \mathbb{Z}\}$. Denna restklass skrivs $\overline{3}$, här är då modulus 5 underförstått.

Tillsammans med strecknotationen, alltså att vi drar ett streck ovanför det tal som representerar restklassen ($\overline{3}$), måste det alltid finnas ett bestämt n (som vi i exemplet ovan kallar ett "modulos"). Detta n var 2 i de två första delarna av exemplet och 5 i den sista delen. Det måste alltid finnas något sätt att veta vad n är, annars kan vi inte veta vad $\overline{x}$ betyder (där x används för att representera restklassen $\overline{x}$).

Vi betraktar återigen multiplikationstabellerna för $\mathbb{Z}_7$ och $\mathbb{Z}_{10}$:

$\mathbb{Z}_7$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\overline{6}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\overline{6}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{6}$	$\overline{1}$	$\overline{3}$	$\overline{5}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{6}$	$\overline{2}$	$\overline{5}$	$\overline{1}$	$\overline{4}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{1}$	$\overline{5}$	$\overline{2}$	$\overline{6}$	$\overline{3}$
$\overline{5}$	$\overline{0}$	$\overline{5}$	$\overline{3}$	$\overline{1}$	$\overline{6}$	$\overline{4}$	$\overline{2}$
$\overline{6}$	$\overline{0}$	$\overline{6}$	$\overline{5}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

$\mathbb{Z}_{10}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\overline{6}$	$\overline{7}$	$\overline{8}$	$\overline{9}$
$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$	$\overline{0}$
$\overline{1}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\overline{6}$	$\overline{7}$	$\overline{8}$	$\overline{9}$
$\overline{2}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{6}$	$\overline{8}$	$\overline{0}$	$\overline{2}$	$\overline{4}$	$\overline{6}$	$\overline{8}$
$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{6}$	$\overline{9}$	$\overline{2}$	$\overline{5}$	$\overline{8}$	$\overline{1}$	$\overline{4}$	$\overline{7}$
$\overline{4}$	$\overline{0}$	$\overline{4}$	$\overline{8}$	$\overline{2}$	$\overline{6}$	$\overline{0}$	$\overline{4}$	$\overline{8}$	$\overline{2}$	$\overline{6}$
$\overline{5}$	$\overline{0}$	$\overline{5}$	$\overline{0}$	$\overline{5}$	$\overline{0}$	$\overline{5}$	$\overline{0}$	$\overline{5}$	$\overline{0}$	$\overline{5}$
$\overline{6}$	$\overline{0}$	$\overline{6}$	$\overline{2}$	$\overline{8}$	$\overline{4}$	$\overline{0}$	$\overline{6}$	$\overline{2}$	$\overline{8}$	$\overline{4}$
$\overline{7}$	$\overline{0}$	$\overline{7}$	$\overline{4}$	$\overline{1}$	$\overline{8}$	$\overline{5}$	$\overline{2}$	$\overline{7}$	$\overline{6}$	$\overline{3}$
$\overline{8}$	$\overline{0}$	$\overline{8}$	$\overline{6}$	$\overline{4}$	$\overline{2}$	$\overline{0}$	$\overline{8}$	$\overline{6}$	$\overline{4}$	$\overline{2}$
$\overline{9}$	$\overline{0}$	$\overline{9}$	$\overline{8}$	$\overline{7}$	$\overline{6}$	$\overline{5}$	$\overline{4}$	$\overline{3}$	$\overline{2}$	$\overline{1}$

I kapitel 4 såg vi hur vi kunde introducera räkneoperationer på restklasser. Vi vill nu även införa division och för att kunna göra det på ett sätt som gör att läsaren ska kunna förstå hur det görs ska vi jämföra den situationen med alla andra situationer då vi infört räkneoperationer på andra matematiska objekt som tal, vektorer och matriser. Vi införde räkneoperationer som addition, subtraktion, olika sorters multiplikation. Alla dessa räkneoperationer har införts för restklasser och nu ska vi alltså införa division.

För att närma oss frågan med division ska vi först närmare studera vad subtraktion egentligen är.

Att *subtrahera* ett matematiskt objekt a från ett matematiskt objekt b innebär att finna det matematiska objekt x som uppfyller ekvationen

$$a + x = b.$$

Om det finns ett x som tillsammans med a och b uppfyller den har ekvationen så betecknar vi ett sådant x med $b - a$ och det här är förstås något vi gjort sedan vi varit små. Poängen här är att närmare begrunda vad vi egentligen gjort så att vi kan bättre förstå generaliseringarna så att detta kan vägleda oss när vi nu inför division.

Låt oss nu resonera på samma sätt beträffande division och se hur division hänger ihop med multiplikation.

Frågan om hur vi *dividerar* ett objekt med ett annat objekt upkommer då vi ställer oss frågan för ett givet a och b finns det ett objekt x som uppfyller ekvationen

$$a \cdot x = b.$$

Återigen om ett sådant x finns då betecknar vi det med by $\frac{b}{a}$. Och igen konstaterar vi att det här har vi sysslat med länge. Dock vet vi lite mer från universitetsmatematiken att om vi arbetar med matriser så när det gäller ekvationen

$$A \cdot X = B$$

så har den inte säkert lösningar för alla matriser A och B. Det finns lösningar för alla B bara om A är som det heter *inverterbar*. När matrisinversen fanns kunde lösningen skrivas

$$X = A^{-1}B$$

och matrisen A^{-1} kallades då *inversen* till A. Ett längre namn är den *multiplikativa inversen* till A. När vi är klara och infört division som "multiplikation med den multiplikativa inversen" så kommer vi även att skriva om $a \cdot x = b$ som

$$x = a^{-1} \cdot b$$

och uppfatta att denna ekvation har uppkommit från att vi multiplicerar båa led i $a \cdot x = b$ med just den här *multiplikativa inversen* (a^{-1}). Det är så här vi ska göra för att närma oss division av restklasser, vi talar alltså om division som *multiplikation med en multplikativ invers*.

Så vi betraktar ekvationen $a \cdot x = b$ i ljuset av hur multiplikationstabellerna för $\mathbb{Z}_7$ och $\mathbb{Z}_{10}$ ser ut och vi ställer alltså frågan när kan vi lösa ekvationen

$$a \cdot x = b?$$

När finns det alltså ett x som tillsammans med a och b uppfyller det här kravet då a, b och x är restklasser?

Det kommer att visa sig att för vissa a har den här ekvationen inga lösningar x som vi kan hitta i $\mathbb{Z}_{10}$, om vi till exempel sätter $b = \overline{4}$ och $a = \overline{5}$ så är det omöjligt att hitta $x \in \mathbb{Z}_{10}$ sådant att $a \cdot x = b$. Det skulle i så fall innebära att vi skulle ha $\overline{5} \cdot x = \overline{4}$ och det är omöjligt. (Varför?)

Vi begrundar alltså division i $\mathbb{Z}_{10}$ och vi formulerar det som att vi söker $x \in \mathbb{Z}_{10}$ som uppfyller

$$a \cdot x = b.$$

Vi kan då konstatera att det som är intressant att betrakta är huruvida den rad (i multiplikationstabellen) som bestäms av a har alla möjliga värden för ett b i sig eller inte. Vi tar ett exempel: kan vi lösa ekvationen

$$\overline{3} \cdot x = b$$

för alla möjliga värden på b? Raden som bestäms av $\overline{3}$ i multiplikationstabellen för $\mathbb{Z}_{10}$ är

$\overline{3}$	$\overline{0}$	$\overline{3}$	$\overline{6}$	$\overline{9}$	$\overline{2}$	$\overline{5}$	$\overline{8}$	$\overline{1}$	$\overline{4}$	$\overline{7}$

och *varje* möjligt b förekommer verkligen i den här raden. Om vi till exempel vill ha ett x så att $\overline{3} \cdot x = \overline{8}$ så tittar vi på den här raden i tabellen där elementet $\overline{8}$ förekommer. Det förekommer i kolumnen som bestäms av $\overline{6}$ (som vi avläser i tabellens översta rad) och det ger oss alltså att lösningen till ekvationen är $x = \overline{6}$. (Alltså $\overline{3} \cdot \overline{6} = \overline{8}$.)

Vid närmare inspektion av multiplikationstabellen för $\mathbb{Z}_{10}$, kan vi dra slutsatsen att ekvationen

$$a \cdot x = b$$

har en lösning $x \in \mathbb{Z}_{10}$ om och endast om $a = \overline{1}$ eller $a = \overline{3}$ eller $a = \overline{7}$ eller $a = \overline{9}$. Så vad är mönstret här? Vad behöver vi ställa för krav på a för att det ska finnas en lösning? Om vi minns hur vi införde multiplikation för matriser så var kravet att matrisen skulle vara inverterbar (som testas till exempel genom att beräkna en determininant) och här har vi alltså frågan som vi ställer i motsvarande situation för multiplikation av restklasser.

Så vad är kravet på a?

Vi kan få en ledtråd om vi studerar multiplikationstabellen för $\mathbb{Z}_7$, här innehåller *alla* rader *alla* element ur $\mathbb{Z}_7$. Det betyder att alla ekvationer av typen

$$a \cdot x = b$$

alltid har lösningar i $\mathbb{Z}_7$. (Utom förstås fallet $a = \overline{0}$.)

Till exempel om vi skulle vilja lösa $\overline{4} \cdot x = \overline{6}$ i $\mathbb{Z}_7$, vad blir då x? Ja vi tittar i raden som bestäms av $a = \overline{4}$ och försöker hitta $\overline{6}$. Och den finns kolumnen som bestäms av $\overline{5}$ (vi avläser det i tabellens översta rad) och det ger oss alltså att

$$\overline{4} \cdot \overline{5} = \overline{6}$$

så att lösningen till $\overline{4} \cdot x = \overline{6}$ är $x = \overline{5}$. Och den här processen, att söka en lösning till $a \cdot x = b$ kommer att ge ett x för alla a och b i $\mathbb{Z}_7$ ... men *inte* i $\mathbb{Z}_{10}$. Men i $\mathbb{Z}_{10}$ fungerar det för vissa a.

Satsen om lösningar till kongruensen $ax \equiv b \pmod{n}$ ger oss svaret. Den sade att den här kongruensen har en lösning på formen

$$x \equiv x_0 \pmod{n}$$

för alla b omm $\gcd(a, n) = 1$. Att lösningen till kongruensen $ax \equiv b \pmod{n}$ blir en hel restklass innebär att vi kan ge en ekvivalent formulering av satsen i termer av kongruensklasser. Den ekvivalenta formuleringen ger oss ett precist svar på vår fråga ovan. Av satsen om lösningar till kongruensen $ax \equiv b \pmod{n}$ (som gäller heltal) följer alltså den här satsen som rör restklasser:

Sats: Låt n vara ett fixt positivt heltal och betrakta $\mathbb{Z}_n$. Då gäller att ekvationen

$$ax = b$$

där $a, x, b \in \mathbb{Z}_n$ har en lösning för alla b om och endast om $\gcd(a, n) = 1$.

Den här satsen är tyvärr inte välformulerad. Minns att $a \in \mathbb{Z}_n$. Beteckningen n står visserligen för ett heltal men a, i den här satsen, är inte ett heltal! Det är ju en restklass! Det visar sig dock att alla heltal i en kongruensklass modulo n har *samma* största gemensamma delare med n. Det är förstås någonting som vi måste bevisa och det lämnas till läsaren som en övning. Vad det innebär är dock att vi kan *definiera* $\gcd(a, n)$, där $a \in \mathbb{Z}_n$ som $\gcd(z, n)$ där z kan väljas fritt i a. (En speciell övning som läsaren kan göra innehåller precis dessa steg.)

Vi ger en följdsats till satsen om lösningar till kongruensen $ax = b$ i $\mathbb{Z}_n$:

Sats: Om n är ett primtal så har ekvationen $ax = b$ i $\mathbb{Z}_n$ en lösning för alla $a \neq \overline{0}$ och alla $b \in \mathbb{Z}_n$.

Bevis: Detta är bara satsen omformulerad med ett primtal som n. Då gäller alltid $\gcd(a, n) = 1$ såvida inte $n|a \Leftrightarrow a = \overline{0}$. Beviset är klart.

Vi har tidigare talat om *multiplikativa inverser* när de är heltal och då har vi till exempel sagt att 7 är multiplikativ invers till 4 modulo 9 eftersom $7 \cdot 4 \equiv 1 \pmod{9}$. Men om vi har $x \cdot y \equiv 1 \pmod{n}$ så kommer samma kongruens gälla för alla tal i de restklasser som x och y är representanter för, vi kan alltså skriva

$$x \cdot y \equiv 1 \pmod{n} \Leftrightarrow \overline{x} \cdot \overline{y} = \overline{1}$$

så definitionen av multiplikativ invers kan utvidgas till kongruensklasser och vi kan göra följande definition:

Definition: Låt n vara ett positivt heltal och låt $a \in \mathbb{Z}_n$. Om $sa = \overline{1}$ för ett $s \in \mathbb{Z}_n$ så kallar vi s för den *multiplikativa inversen* för a i $\mathbb{Z}_n$.

Vi behöver egentligen också övertyga oss om att den multiplikativa inversen är entydig om vi vill kunna skriva "*den* multiplikativa inversen" men entydigheten följer lätt av resonemanget att om vi antar att vi har två multiplikativa inverser s_1 och s_2 till a, då gäller $s_1 a = \overline{1}$ och $s_2 a = \overline{1}$ och

$$s_1 = s_1 \cdot \overline{1} = s_1 \cdot (a \cdot s_2) = (s_1 \cdot a) \cdot s_2 = \overline{1} \cdot s_2 = s_2.$$

Vi tittar på en gammal tentafråga från April 2015:

Tillämpa Euklides utvidgade algoritm för att finna den multiplikativa inversen till $23 \pmod{17}$ och använd den för att lösa kongruensen $23x \equiv 337 \pmod{17}$.

Lösning: Vi utför upprepade applikationer av divisionsalgoritmen och får:

$$\begin{aligned} 23 &= 1 \cdot 17 + 6, \\ 17 &= 2 \cdot 6 + 5, \\ 6 &= 1 \cdot 5 + 1 \end{aligned}$$

så att $1 = 6 - 1 \cdot 5 = 6 - 1 \cdot (17 - 2 \cdot 6) = 3 \cdot 6 - 17 = 3 \cdot (23 - 17) - 17 = 3 \cdot 23 - 4 \cdot 17$. Det här betyder att den multikativa inversen för 23 (mod 17) är 3. Att multiplicera båda sidor av kongruensen med 3 kommer alltså att ge 1 som koefficient framför x, det vill säga vi har då löst kongruensen. Men det är också en bra idé att reducera 337 modulo 17 och sedan utföra multiplikationen med 3. Vi har då $337 \equiv 14$ och om vi gör allt detta får vi

$$23x \equiv 337 \pmod{17} \Leftrightarrow 23x \equiv 14 \pmod{17} \Leftrightarrow 3 \cdot 23x \equiv 3 \cdot 14 \pmod{17}$$

vilket kan skrivas $1 \cdot x \equiv 3 \cdot 14 \pmod{17} \Leftrightarrow x \equiv 8 \pmod{17}$.

Alternativ lösning: Det är faktiskt också en bra idé att även reducera 23 modulo 17. Vi har förstås $23 \equiv 6$ (mod 17) så innan vi överhuvudtaget sätter igång med Euklides utvidgade algoritm kan vi skriva

$$23x \equiv 337 \pmod{17} \Leftrightarrow 6x \equiv 337 \pmod{17} \Leftrightarrow 6x \equiv 14 \pmod{17}$$

och sedan arbeta därifrån. Det visar sig att siffrorna bli mindre då och lärdomen är att så fort vi kan reducera modulo det vi räknar med så gör vi det.

Vi studerar en annan formulering av samma problematik:

Exempel: Finn den multiplikativa inversen av $\overline{17}$ i $\mathbb{Z}_{19}$.

Lösning: Vi söker heltal s, t sådana att $s \cdot 17 + t \cdot 19 = 1$. Dessa heltal existerar säkert eftersom $17, 19$ är relativt prima. (Varför?) Euklides utvidgade algoritm ger då

$$19 = 1 \cdot 17 + 2 \qquad 17 = 8 \cdot 2 + 1 \qquad 1 = 19 - 8 \cdot (19 - 17) \Leftrightarrow 1 = 9 \cdot 17 - 8 \cdot 19$$

Detta betyder att $s = 9$ och $t = -8$ uppfyller $s \cdot 17 + t \cdot 19 = 1$ vilket betyder att den multiplikativa inversen av $\overline{17}$ i $\mathbb{Z}_{19}$ är $\overline{9}$.

Övningar

6.1.1 Läsaren kan själv välja två skilda primtal, p, q, vilka som helst, och öva på att hitta den multiplikativa inversen av $\overline{p}$ i $\mathbb{Z}_q$ men också den multiplikativa inversen av $\overline{q}$ i $\mathbb{Z}_p$. Gör detta för ett antal par av skilda primtal. (Kanske det är bra att välja $p, q = 11, 13$, $p, q = 11, 17$, $p, q = 23, 29$ och $p, q = 23, 37$.)

6.1.2 Studera närmare ett av de fall du hittade på i förra övningsuppgiften, där du hittade den multiplikativa inversen av $\overline{p}$ i $\mathbb{Z}_q$ för två skilda primtal p, q. Använd Euklides utvidgade algoritm för att hitta den multiplikativa inversen av $p + q$ i $\mathbb{Z}_q$ och reducera modulo q (så att du inte representerar den multiplikativa inversen med något annat än talen $1, 2, \ldots, q - 1$). Gör sedan samma sak för $p + 2 \cdot q$ och reducera modulo q. Kan du se ett mönster? Vad skulle du få om du sökte den multiplikativa inversen för $p + 10 \cdot q$?

2. Mer om multiplikativa inverser i $\mathbb{Z}_p$

Vi kan formulera resultaten från föregående avsnitt så här:

Sats: Låt n vara ett positivt heltal. Om $ac \equiv bc \pmod{n}$ och $\gcd(c, n) = 1$, så har vi $a \equiv b \pmod{n}$.

eller, om vi använder kongruensklassnotation kan vi uttrycka det så här:

Sats: Låt n vara ett positivt heltal och betrakta $\overline{a}, \overline{b}, \overline{c} \in \mathbb{Z}_n$. Om $\overline{a} \cdot \overline{c} = \overline{b} \cdot \overline{c}$ och $\gcd(c, n) = 1$, så har vi $\overline{a} = \overline{b}$.

Samma två satser uttryckta då n är ett primtal ($n = p$) får lydelserna:

Sats: Låt p vara ett primtal. Om $ac \equiv bc \pmod{p}$ och $c \not\equiv 0 \pmod{p}$ så har vi $a \equiv b \pmod{p}$.

(Eftersom p är ett primtal så har vi $c \not\equiv 0 \pmod{p} \Leftrightarrow p \nmid c \Leftrightarrow \gcd(c, p) = 1$ så den förra satsens förutsättningar är precis som de två föregående.)

Slutligen ger vi också formuleringen i $\mathbb{Z}_p$:

Sats: Låt p vara ett primtal och betrakta $\overline{a}, \overline{b}, \overline{c} \in \mathbb{Z}_p$. Om $\overline{a} \cdot \overline{c} = \overline{b} \cdot \overline{c}$ och $\overline{c} \neq \overline{0}$ så har vi $\overline{a} = \overline{b}$.

Den sista formuleringen är särskilt intressant. Det betyder att vi har delat båda led med $\overline{c}$, vi har "förkortat" eller "dividerat" med $\overline{c}$ som vi säger när det gäller vanliga tal skilda från 0. Alla dessa räkningar och egenskaper

betyder att $\mathbb{Z}_p$ fungerar i princip likadant som de vanliga reella talen: vi kan addera, subtrahera, multiplicera och dividera (med element som inte är 0).

Vi utforskar egenskaper hos $\mathbb{Z}_p$, där p är ett primtal. Där har alltså varje element som inte är $\overline{0}$ en multiplikativ invers. Det är fallet i till exempel $\mathbb{Z}_7$, elementen $\overline{1}, \overline{2}, \overline{3}, \overline{4}, \overline{5}, \overline{6}$ har alla multiplikativa inverser. Om vi betecknar dessa element med x, så kan vi alltså alltid finna ett s sådant att $sx = \overline{1}$. (Vad gäller för $\overline{7}$?)

För att hitta multiplikativa inverser till alla dessa element kan vi titta i multiplikationstabellen för $\mathbb{Z}_7$. Om vi gör det så ser vi att $\overline{1} \cdot \overline{1} = \overline{1}$ vilket visar att den multiplikativa inversen av $\overline{1}$ är $\overline{1}$ (självklart!), vidare ser vi att $\overline{2} \cdot \overline{4} = \overline{1}$ så att den multiplikativa inversen av $\overline{2}$ är $\overline{4}$.

Vi kan skapa en tabell:

$\mathbb{Z}_7$	Invers
$\overline{0}$	*Saknas*
$\overline{1}$	$\overline{1}$
$\overline{2}$	$\overline{4}$
$\overline{3}$	$\overline{5}$
$\overline{4}$	$\overline{2}$
$\overline{5}$	$\overline{3}$
$\overline{6}$	$\overline{6}$

Betrakta den här frågan: vilka element i $\mathbb{Z}_p$ har sig själva som multiplikativa inverser? Det vill säga vilka element $x \in \mathbb{Z}_p$ uppfyller ekvationen

$$x \cdot x = \overline{1}?$$

Om vi skriver $x \cdot x$ som x^2 så frågar vi alltså vilka $x \in \mathbb{Z}_p$ uppfyller ekvationen

$$x^2 = \bar{1}.$$

Eftersom p är ett primtal, så fungerar $\mathbb{Z}_p$ som de reella talen så vi kan lösa den här ekvationen som vi brukar lösa andragradsekvationer, genom att faktorisera polynomet:

$$x^2 = \bar{1} \Leftrightarrow x^2 - \bar{1}^2 = \bar{0} \Leftrightarrow (x + \bar{1}) \cdot (x - \bar{1}) = \bar{0} \Leftrightarrow x + \overline{1} = \overline{0} \vee x - \overline{1} = \overline{0} \Leftrightarrow x = \overline{-1} = \overline{p-1} \vee x = \overline{1}$$

så de enda två möjligheterna för ett element i $\mathbb{Z}_p$ att vara sin egen multiplikativa invers är om det är $\overline{1}$ eller $\overline{p-1}$. Det betyder att varje annat element i $\mathbb{Z}_p$, det vill säga de $p-3$ elementen $\overline{2}, \overline{3}, \ldots, \overline{p-2}$ alla har multiplikativa inverser som *inte* är dem själva.

Och visst är det så om vi tittar i tabellen ovan. De enda elementen i $\mathbb{Z}_7$ som är sina egna inverser är $\overline{1}$ och $\overline{6} = \overline{7-1}$.

Vi kommer att se en tillämpning av denna egenskap i beviset av *Wilsons Sats* i nästa avsnitt.

Övningar saknas men teorin kommer att tillämpas i andra sammanhang.

3. Några talteoretiska resultat

I detta avsnitt ska vi bevisa några viktiga resultat inom talteorin. Vi börjar med en enkel sats om primtalsfaktorer.

Sats: Låt $n > 1$ vara ett positivt heltal. Den minsta delaren $a > 1$ till n måste då alltid vara ett primtal.

Bevis: Självklart delar ju det positiva talet 1 varje annat heltal, det är därför vi behöver säga att $a > 1$ för att satsen ska bli sann. Talet 1 kallas då en *trivial* delare till varje annat heltal. (Vi bortser också från alla negativa delare.) Vi introducerar nu därför mängden av alla positiva *icketriviala* delare till n:

$$M = \{d \in \mathbb{Z} : d|n \wedge d > 1\}.$$

Vår uppgift är att visa att det minsta elementet i den här mängden är ett primtal. Vi konstaterar först att mängden inte är tom, faktiskt är n själv ett element i mängden. Talet n skulle också kunna uppfattas som en trivial delare, men vi väljer inte att säga så i just det här beviset.

Ok, M är en icketom mängd av positiva tal, då måste den ha ett minsta element. Om n själv är ett primtal ja då kan inte M innehålla några andra tal än just n själv och då blir n själv det minsta elementet i M. Så

om n är ett primtal är saken alltså klar. $(n = a)$. Vi antar därför att n *inte* är ett primtal. Vi kan då skriva

$$n = a \cdot b$$

där $a > 1$ är det minsta talet i M och $b > 1$ också ligger i M. Vad händer nu om a inte är ett primtal? För att se vad som händer då så antar vi att a inte är ett primtal. Då kan a faktoriseras i $p > 1, q > 1$, det vill säga $a = p \cdot q$. Men eftersom $n = a \cdot b = p \cdot q \cdot b$ så har vi tydligen $p|n$ så att *även* $p \in M$. Eftersom $p < a$ (eftersom $p \cdot q = a$ och $q > 1$) så blir detta en motsägelse, a skulle ju vara det *minsta* talet i M, men vi har hittat $p \in M$ som *ändå* är mindre än a. Antagandet om att det minsta talet $a \in M$ inte var ett primtal kan alltså inte vara sant, alltså är det minsta talet i M alltid ett primtal. Beviset är klart.

Från den här satsen följer en mycket viktig sats:

Sats: *(Euklides Sats.)* Det finns oändligt många primtal.

Bevis: Antag motsatsen, alltså att det bara finns ett ändligt antal primtal. Då kan vi skriva upp allihop i en lista: $p_1, p_2, \ldots, p_n$ där n något troligen mycket stort tal. Vi vet inte vilket stort tal det här är men poängen är att listan är ändlig och teoretiskt sätt kan skrivas upp och att den har ett slut. Vi vet att $p_1 = 2$ och $p_2 = 3$ och att det, enligt antagandet, inte finns något primtal större än p_n (vi antar också att de är givna i storleksordning). Bilda nu det speciella talet

$$N = p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1.$$

Det här talet är större än alla tal i listan så därför kan det inte vara ett primtal, det är ju inte med listan som innehåller *alla* primtal. Det betyder då (enligt förra satsen) att den minsta delaren till det här talet måste vara ett primtal. Eftersom vi har en fullständig lista på alla primtal så vet vi att det minsta talet som delar N måste finnas i den listan, det vill säga vi vet att

$$p_1|N \quad \text{eller} \quad p_2|N \quad \text{eller} \quad \ldots \quad \text{eller} \quad p_n|N.$$

Men inga av dessa delbarhetsförhållanden kan gälla eftersom N är konstruerat så att det alltid lämnar resten 1 vid division med varje tal ur listan. Vi har nått en motsägelse så antagandet om att det bara finns ändligt många primtal måste alltså vara falskt. Det finns alltså oändligt många primtal vilket fullbordar beviset.

Vi tittar på en annan sats:

Sats: *(Fermats Lilla Sats.)* Om p är ett primtal och $p \nmid c$, då gäller

$$c^{p-1} \equiv 1 \pmod{p}.$$

Bevis: Betrakta talen $1, 2, \ldots, (p-1)$. Om vi multiplicerar alla dessa tal med talet c så bildas talen $c, 2c, \ldots, (p-1)c$. Om två av talen i den andra listan är kongruenta modulo p, vad gäller då? Antag att dessa två tal har ordningsnummer j och k (i intervallet $1, 2, \ldots p-1$). Då har vi alltså

$$j \cdot c \equiv k \cdot c \pmod{p}.$$

Eftersom $p \nmid c$ så är $\gcd(c, p) = 1$ det vill säga det finns en multiplikativ invers till c modulo p, kalla den d. Men då får vi

$$j \cdot c \equiv k \cdot c \pmod{p} \Rightarrow j \cdot c \cdot d \equiv k \cdot c \cdot d \pmod{p} \Rightarrow j \cdot 1 \equiv k \cdot 1 \pmod{p} \Leftrightarrow j = k$$

där vi kan dra slutsatsen att $j = k$ eftersom både j och k är bland talen $1, 2, \ldots, (p-1)$. Men det här betyder alltså att i $\mathbb{Z}_p$ är elementen $\overline{c}, \overline{2c}, \ldots, \overline{(p-1)c}$ de samma som elementen $\overline{1}, \overline{2}, \ldots, \overline{c}$ men i en annan ordning. Det betyder att

$$\overline{c} \cdot \overline{2c} \cdot \ldots \cdot \overline{(p-1)c} = \overline{c} \cdot \overline{2} \cdot \overline{c} \cdot \ldots \cdot \overline{(p-1)} \cdot \overline{c} = \overline{1} \cdot \overline{2} \cdot \ldots \cdot \overline{p-1}$$

så att

$$c \cdot c \cdot \ldots \cdot c\overline{(p-1)!} = \overline{(p-1)!}$$

vilket ger efter division med $\overline{(p-1)!}$ att $c^{p-1} = \overline{1}$ vilket, uttryckt som en kongruens är $c^{p-1} \equiv 1 \pmod{p}$ vilket skulle bevisas.

I förra avsnittet arbetade vi med att para ihop elementen som inte var $\overline{0}$ eller $\overline{p-1}$ i $\mathbb{Z}_p$ med sina respektive multiplikativa inverser. Att det överhuvudtaget gick att göra så hängde på att p var ett udda primtal. Vi ska använda denna egenskap i beviset av nästa sats.

Sats: *(Wilsons Sats.)* Låt $n \in \mathbb{Z}$ vara ett godtyckligt positivt heltal. Då gäller

$$n \text{ är ett primtal} \Leftrightarrow (n-1)! \equiv -1 \pmod{n}.$$

Bevis: Antag att n är ett primtal. Om $n = 2$ så har vi

$$(2-1)! = 1 \equiv -1 \pmod 2$$

så att alltså $(n-1)! \equiv -1 \pmod n$ stämmer för $n = 2$. Om vi betraktar alla andra primtal så kommer de att vara udda. Då kan vi göra som i avsnittet om multiplikativa inverser i $\mathbb{Z}_p$ och rada upp hela produkten

$$\overline{1} \cdot \overline{2} \cdot \overline{3} \cdot \overline{4} \cdot \overline{5} \cdot \overline{6} = \overline{1} \cdot \overline{2} \cdot \overline{4} \cdot \overline{3} \cdot \overline{5} \cdot \overline{6} = \overline{1} \cdot \overline{1} \cdot \overline{1} \cdot \overline{6} = \overline{6}$$

och räkna ut produkten genom att para ihop de olika multiplikativa inverserna. Ovan är detta gjort i $\mathbb{Z}_7$ men rent generellt får kalkylen utseendet

$$\overline{1} \cdot \ldots \cdot \overline{p-1} = \overline{1} \cdot \overline{1} \cdot \ldots \cdot \overline{1} \cdot \overline{p-1} = \overline{p-1}$$

så att

$$\overline{(p-1)!} = \overline{p-1}$$

och detta omskrivet till en kongruens ger $(p-1)! \equiv p-1 \equiv -1 \pmod p$ vilket skulle bevisas.

Vi antar nu omvänt att $(n-1)! \equiv -1 \pmod n$. Antag också, för att nå en motsägelse att n *inte* är ett primtal. Då finns $a > 1$ och $b > 1$, positiva heltal sådana att $n = a \cdot b$. Detta insatt i $(n-1)! \equiv -1 \pmod n$ ger oss

$$(n-1)! = 1 \cdot 2 \cdot \ldots \cdot (a-1)a(a+1) \cdot \ldots \cdot (n-1) = -1 + k \cdot a \cdot b = -1 + k \cdot n$$

där k är något heltal. Det här är en motsägelse. Varför det? Jo, likheten kan skrivas om till

$$1 = k \cdot a \cdot b - 1 \cdot 2 \cdot \ldots \cdot (a-1)a(a+1) \cdot \ldots \cdot (n-1).$$

och högerledet kan skrivas som $a \cdot \textit{heltal}$ och slutsatsen blir att $a|1$ vilket är omöjligt. Antagandet om att n inte är ett primtal måste alltså vara falskt, det vill säga n måste vara ett primtal vilket skulle bevisas.

Övningar saknas med teorin kommer att tillämpas i andra sammanhang.

4. Aritmetikens fundamentalsats

Vi ska nu ge halva beviset för Aritmetikens Fundamentalsats som saknades förut. (Andra halvan kommer senare i kapitlet). Vi behöver dock en hjälpsats först.

Sats: Om a och b är två heltal och p är ett primtal som uppfyller $p|ab$ så gäller $p|a$ eller $p|b$.

Bevis: Antag att $p \nmid b$. Eftersom p är ett primtal har vi $\gcd(p, b) = 1$ och enligt Bezouts sats ger då

$$px + by = 1$$

för några heltal x, y. Antag vidare att $ab = pk$. Då har vi

$$a = a \cdot 1 = a \cdot (px + by) = apx + aby = apx + pky = p(ax + ky) = p \cdot \textit{heltal}$$

vilket alltså innebär att $p|a$. Vi har alltså visat $p \nmid b \Rightarrow p|a$ vilket är logiskt ekvivalent med $p|a$ eller $p|b$ och beviset är klart.

Som nämndes kommer vi bara att ge ena halva beviset av Aritmetikens Fundamentalsats. Av tidigare exempel är det ganska troligt att det finns standardmässiga primtalsfaktoriseringar av varje heltal, alltså faktoriseringar av typen

$$720 = 8 \cdot 9 \cdot 10 = 2^3 \cdot 3^2 \cdot 2 \cdot 5 = 2^4 \cdot 3^2 \cdot 5^1$$

där primtalen 2, 3 och 5 är de aktuella primatalsfaktorerna i det positiva heltalet 720. Aritmetikens Fundamentalsats säger att varje positivt heltal har exakt en sådan faktorisering. För att visa att det gäller måste vi visa två saker:

(1) För varje positivt heltal finns *minst* en standardmässig primtalsfaktorisering.
(2) För varje heltal finns *högst* en standardmässig primtalsfaktorisering.

Eftersom vi sett många exempel på primtalsfaktoriseringar så ska vi vänta med att visa att det alltid finns *minst* en sådan (första punkten) och visa att det finns *högst* en sådan (andra punkten). Senare i kapitlet visas den första punkten.

Andra punkten visas här nedan genom att vi antar att det finns två standardmässiga primtalsfaktoriseringar och sedan visa att dessa båda faktoriseringar måste vara identiska.

Aritmetikens Fundamentalsats: Låt N vara ett godtyckligt positivt heltal. Då finns en och endast en standardmässig primtalsfaktorisering av N.

Bevis: *(Bevis av att det finns högst en.)* Vi antar att det finns två olika standardmässiga primtalsfaktoriseringar av N och betecknar dem med

$$N = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_n^{a_n} \qquad \text{respektive} \qquad N = q_1^{b_1} \cdot q_2^{b_2} \cdot \ldots \cdot q_m^{b_m}.$$

Här är alltså $E = \{p_1, p_2, \ldots, p_n\}$ mängden av primtalsfaktorer i den ena primtalsfaktoriseringen av N och $F = \{q_1, q_2, \ldots, q_m\}$ är mängden av primtalsfaktorer i den andra primtalsfaktoriseringen av N. (I exemplet ovan med 720 ovan fick vi primtalsfaktorerna $\{2, 3, 5\}$.) Vi ska visa att dessa mängder i själva verket är lika och att exponenterna också är lika, det vill säga att $n = m$, $p_i = q_i$ för alla i och $a_i = b_i$ för alla i.

Vi börjar med p_1. Uppenbarligen gäller $p_1|N$, det får vi från den ena faktoriseringen. Om vi då använder vår hjälpsats på utsagan

$$p_1 | q_1^{b_1} \cdot q_2^{b_2} \cdot \ldots \cdot q_m^{b_m} \quad (= N)$$

upprepade gånger kan vi dra slutsatsen att p_1 måste dela något av talen $q_1^{b_1}, q_2^{b_2}, \ldots, q_m^{b_m}$. Men det enda sättet som ett primtal (p_1) kan dela en potens av primtal ($q^{\text{någonting}}$) är om $p_1 = q$. Vi måste alltså ha $p = q_1$ eller $p = q_2$ eller ... eller $p_1 = q_m$. Det betyder att $p_1 \in F$. Precis samma argument kan göras för alla andra primtal åt båda hållen (för $p_2, \ldots, p_n$ respektive $q_1, \ldots, q_m$) så att vi får $E \subset F$ och $F \subset E$ och av detta kan vi dra slutsatsen $E = F$ varav följer $n = m$ och att $p_i = q_i$ för alla i. Det återstår att visa att $a_i = b_i$ för alla i. Vi har som sagt situationen

$$N = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_n^{a_n} = p_1^{b_1} \cdot p_2^{b_2} \cdot \ldots \cdot p_n^{b_n}$$

Om $a_1 < b_1$ så skulle vi kunna förkorta denna identitet och få

$$VL = p_2^{a_2} \cdot \ldots \cdot p_n^{a_n} = p_1^{b_1 - a_1} \cdot p_2^{b_2} \cdot \ldots \cdot p_n^{b_n} = HL$$

där vi alltså inte har någon primfaktor p_1 i vänster led (VL) men väl en primfaktor p_1 i höger led (HL). Vad detta betyder är alltså att $p_1|HL$. Eftersom $VL = HL$ måste vi alltså ha $p_1|VL$. Om vi här återigen använder hjälpsatsen så får vi att

$$p_1 | p_2^{a_2} \vee p_1 | p_3^{a_3} \vee \ldots \vee p_1 | p_n^{a_n}$$

som återigen leder till $p_1|p_2^{a_2} \vee p_1|p_3^{a_3} \vee \ldots \vee p_1|p_n^{a_n}$ vilket är en motsägelse eftersom vi jobbar med primtal och alla dessa primtal är olika (enda sättet för $p_1|p_2^{a_2}$ är om $p_1 = p_2$). Vi kan alltså inte ha $a_1 < b_1$ och på samma sätt inte heller $b_1 < a_1$, alltså måste $a_1 = b_1$ följa. På precis samma sätt kan vi visa $a_i = b_i$ för alla i. Beviset är klart eftersom vi visat att de båda primtalsfaktoriseringarna är identiska.

Det återstår som vi nämnt att visa att det överhuvudtaget finns någon primtalsfaktoriseringen av alla positiva heltal, det kommer att göras i avsnittet om så kallad *stark matematisk induktion*.

Övningar saknas med teorin kommer att tillämpas i andra sammanhang. Det har också funnits övningar på Aritmetikens Fundamentalsats i kapitel 4.

5. Matematisk Induktion

Den kraftfulla principen bakom så kallad *matematisk induktion* kan illustreras av en rad fallande dominobrickor. Vi har alla sett roliga filmer på nätet med jättemånga fallande dominobrickor som är uppställda så att varje dominobricka i följden är *tillräckligt nära* nästa dominobricka så att om en dominobricka faller så *medför* det att nästa dominibricka faller. Vi har alltså *implikationer* inbyggda i den här situationen. Eftersom alla dominobrickor står tillräckligt nära varandra så faller alla om den första får en liten knuff så att den faller på nästa dominobricka. Vi har alltså ett slags startimpuls som också måste ges för att alla ska falla.

När vi läser ovanstående beskrivning har vi förhoppningsvis en upplevelse av förståelse. Och all matematisk teoribyggnad vilar på något slags upplevelse av förståelse. Med vad är förståelse då? Vi lämnar den frågan men konstaterar att för att göra matematik behövs förståelse, annars blir det ett tomt manipulerande med formler.

Genom tiderna har människorna haft olika sorters talbegrepp och vi har nämnt några olika talmängder, de naturliga talen ($\mathbb{N}$), sedan kommer heltalen ($\mathbb{Z}$), sedan de rationella talen ($\mathbb{Q}$) sedan de reella talen ($\mathbb{R}$) och så vidare. Vi har också inklusionen $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ och varje inklusion representerar en utvidgning av talbegreppet. De naturliga talen kommer först och är $1, 2, 3, \ldots$. Associerat med de naturliga talen finns additionsoperationen, vi adderar två naturliga tal och får ett nytt naturligt tal. Efter det kommer heltalen och det nya jämfört med naturliga talen är talet 0 och de negativa talen, $-1, -2, -3, \ldots$ och subtraktionen,

vi kan addera men också subtrahera heltal och alltid få ett heltal. Nästa steg är de rationella talen och med dessa tal kan vi addera, subtrahera, multiplicera men även *dividera* och alltid få ett rationellt tal som resultat. (Med undantag för att vi aldrig får dividera med 0.) På liknande sätt involverar nästa steg, de reella talen en utvidgning av talbegreppet och en utvidgning av vad vi kan göra med talen. Och här kommer poängen: de första talen som nämndes i detta stycke, de naturliga talen, $\mathbb{N}$, var kommer de ifrån? Svaret kanske kommer att förvåna oss. Vi kan nämligen med fog hävda att de naturliga talen inte alls kommer från matematiken, de kommer från oss människor. Och vi skulle kunna säga att vi skapar matematiken med de naturliga talen som slags grund. Om det här stämmer kan man då ställa sig frågan var människan har fått de naturliga talen ifrån och genom människans utveckling genom många årtusenden skulle vi kunna förmoda att hon har vuxit upp med dem ... för 10000 år sedan kanske konversationer av typen

Om du får en häst av mig så kan väl jag få tre getter av dig?

kunde utspela sig. De naturliga talen fanns liksom i naturen och det är precis därför som de kallas de *naturliga* talen. Mänsklig begreppsbildning som involverar jämförelse av olika objekt och förmågan att konstatera att två objekt är av samma typ var kanske upphovet till de naturliga talen. ”3 stycken äpplen som vi kan äta”, ”12 fiskar som vi fångar (för att äta)”, ”för vintern kan vi lagra fettet från tre hjortar i den här grottan” och liknande, alla dessa överväganden handlar om överlevnad och egentligen så handlar kanske all utbildning och begreppbildning om detta: att leva bättre. Vi skulle kunna säga att det övergripande målet med ingenjörskunskap är att förbättra levnadsförhållandena.

Nog med filosofi! Vi har konstaterat att de naturliga talen principiellt inte har matematiskt ursprung, de finns snarare som en grund till matematiken. Det betyder att den användning som vi har av de naturliga talen har *mycket* djupa rötter. Det är kanske dessa djupa rötter som ligger bakom fascinationen med de fallande dominbrickorna. De modellerar de naturliga talen. De naturliga talen skulle ju kunna liknas vid en lång rad dominobrickor.

Det vi som vi nu ska göra är att studera hur principen för matematisk induktion kan användas för att formulera bevis av utsagor som består av kvantifierade predikat där variabeln som finns i predikatet löper över de naturliga talen. Ett exempel på en sådan utsaga skulle kunna vara

Summan av de n första udda naturliga talen är n^2.

För att se att det här är ett kvantifierat predikat kan vi formulera det på ett lite med klumpigt sätt:

För alla naturliga tal n gäller att om vi summerar den n första udda naturliga talen blir summan n^2.

Klumpigt som sagt, men vi kan lättare formulera detta matematiskt med en allkvantor:

$$\forall n \in \mathbb{N} : A(n), \text{ där } A(n) \text{ är predikatet } \sum_{k=1}^{n}(2k-1) = n^2.$$

Vi kan studera ett antal enskilda utsagor som bildas då vi sätter in de första naturliga talen $n = 1$, $n = 2$ och $n = 3$ och se att $A(1)$, $A(2)$ och $A(3)$ alla är sanna:

$$A(1) \Leftrightarrow 1 = 1^2, \textit{sant.}\ A(2) \Leftrightarrow 1 + 3 = 2^2 (= 4), \textit{sant.}\ A(3) \Leftrightarrow 1 + 3 + 5 = 3^2 (= 9), \textit{sant.}$$

Dessa observationer innebär förstås ett fullständigt acceptabelt matematisk bevis för att $A(1)$, $A(2)$ och $A(3)$ alla är sanna. Men nu vill vi visa att $A(n)$ är sann för *alla* naturliga tal och tar vi in principen för matematisk induktion som alltså är beskriven av de fallande dominobrickorna. Vi vill visa att alla utsagor

$$A(1), \quad A(2), \quad A(3), \quad A(4), \quad A(5), \quad \ldots$$

är sanna och om vi kan visa att alla implikationer $A(1) \Rightarrow A(2)$, $A(2) \Rightarrow A(3)$, $A(3) \Rightarrow A(4)$, ... är sanna ja då är alla utsagorna i samma situation som dominobrickorna, vi såg att $A(1)$ var sann (ovan). Om också alla implikationer $A(1) \Rightarrow A(2)$, $A(2) \Rightarrow A(3)$, $A(3) \Rightarrow A(4)$, ... är sanna så får vi

$$A(1) \text{ sann} \Rightarrow A(2) \text{ sann} \Rightarrow A(3) \text{ sann} \Rightarrow A(4) \text{ sann} \Rightarrow A(5) \text{ sann}, \ldots$$

det vill säga att $A(n)$ är sann för alla $n \in \mathbb{N}$. Sanningsvärdena uppstår precis på samma sätt som att alla dominorbrickor faller. Att första dominobrickan faller svarar mot att vi kontrollerade att $A(1)$ var sann. Att implikationenerna $A(1) \Rightarrow A(2)$, $A(2) \Rightarrow A(3)$, $A(3) \Rightarrow A(4)$, ... var sanna svarar då mot att dominobrickorna är placerade på ett sådant sätt att om en bricka faller så *medför* det att nästa bricka faller.

Strategin är att alltså att visa att alla utsagorna uppfyller två saker:

(1) $A(1)$ är sann. Detta är startvärdet som svarar mot att första dominbrickan fallller.

(2) För alla naturliga tal p är implikationen $A(p) \Rightarrow A(p+1)$ sann. Detta svarar mot att alla dominobrickor stå tillräckligt tätt tillsammans så att om en faller så faller nästa. Det är ju också i analogi med vad implikationen säger: om en utsaga är sann och implikationen är sann så bli också nästa utsaga sann.

(Varibeln p är bara införd för att poängtera att vi arbetar med en viss utsaga, det kommer från engelskans *particular*, alltså att vi då vi studerar sanningsvärdet hos implikationen gör detta för ett visst värde som vi kallar p. Vi skulle kunnat använda bokstaven n också i steg 2, men i den framställningen gör vi inte det.)

Vi ger nu ett formellt induktionsbevis för det påstående vi just gjort.

Sats: $\forall n \in \mathbb{N} : \sum_{k=1}^{n}(2k-1) = n^2$.

Bevis: *Matematisk induktion över n.* Vi inför predikatet $A(n) \Leftrightarrow \sum_{k=1}^{n}(2k-1) = n^2$. Det som ska visas kan då uttryckas $\forall n \in \mathbb{N} : A(n)$. Vi delar upp induktionsbeviset i tre delsteg.

Steg 1. Visa att $A(1)$ är sann. Att visa att $A(n)$ är sant innebär att kontrollera att vänsterledet i påståendet som formerar $A(n)$ är lika med högerledet i påståendet som formerar $A(n)$ och vi ska göra detta för $n = 1$. Vi har redan gjort det ovan, men gör om det här för tydlighetens skull:

$$\text{Vänster led i } A(1) = \sum_{k=1}^{1}(2k-1) = 2 \cdot 1 - 1 = 2 - 1 = 1. \qquad \text{Högerled i } A(1) = 1^2 = 1.$$

Bara som ren bekvämlighet inför vi här den kortare notationen VL_n respektive HL_n för vänter respektive höger led i $A(n)$ och vi konstaterar att $VL_1 = 1$ och att $HL_1 = 1$ och alltså gäller $VL_1 = HL_1$ det vill säga $A(1)$ är sann. *(Första dominobrickan faller alltså!)*

Steg 2. I detta steg ska vi visa att alla de här implikationerna $A(p) \Rightarrow A(p+1)$ är sanna för alla p som är aktuella. (Som nämnt ovan skulle vi kunna använda variabeln n här istället för p men vi använder p just nu.)

Vi ska alltså visa implikationen

$$A(p) \Rightarrow A(p+1)$$

för alla $p \in \mathbb{N}$. Ett sätt att visa en implikation (och det vanligaste i induktionsbevis) är att anta att förledet är sant och visa att efterledet följer med hjälp av logiska och matematiska manövrar. Vi antar alltså att $A(p)$ är sann, det vill säga att vi för ett visst värde på $p \in \mathbb{N}$ har

$$A(p) \Leftrightarrow VL_p = HL_p \Leftrightarrow \sum_{k=1}^{p} = p^2.$$

Antagandet som vi gör i induktionsbevis kallas *induktionsantagandet* och observera att vi nu har en användbar formel. *Vi kan alltså använda oss av att den här likheten gäller.* Målet är att visa att $A(p+1)$ är sann och vi kommer att kunna göra det *med kraft* av induktionsantagandet. För att visa att $A(p+1)$ är sann studerar vi VL_{p+1} och HL_{p+1} och försöker se att de är lika. Och här ska vi alltså *använda* induktionsantagandet. Så vi får

$$VL_{p+1} = \sum_{k=1}^{p+1}(2k-1) = \sum_{k=1}^{p}(2k-1) + (2(p+1)-1) = \left[\sum_{k=1}^{p}(2k-1)\right] + \left[2(p+1)-1\right]$$

och här har vi än så länge bara satt in $p+1$ i uttrycket som definerar vad VL_{p+1} är lika med. Detta vänsterled (VL_{p+1}) är tydligen en summa med $p+1$ termer. Induktionsantagandet uttrycker någonting om de första p termerna i denna summa, därför är summan isärskriven ovan så att vi ser hela summan (med $p+1$ termen) som en summa av först p termen och sedan term nummer $p+1$. Vi arbetar vidare med det uttrycket och ersätter då summan av de första p termerna med det som vi enligt induktionsantagandet vet den summan är lika med, alltså:

$$VL_{p+1} = \left[\sum_{k=1}^{p}(2k-1)\right] + \left[2(p+1)-1\right] = p^2 + 2(p+1) - 1 = p^2 + 2p + 1.$$

och i det här steget har vi alltså *använt* induktionsantagandet och alltså ersatt $\left[\sum_{k=1}^{p}(2k-1)\right]$ med p^2. Det är *mycket viktigt* att vi poängterar detta, annars har vi inte en fullständig motivering för alla delsteg.

När vi använt induktionsantagandet är det lämpligt att undersöka vad HL_{p+1} är lika med, så vi sätter in $p+1$ i definitionen av HL_n och får:

$$HL_{p+1} = (p+1)^2 = p^2 + 2p + 1$$

och här kan vi konstatera att $VL_{p+1} = HL_{p+1}$ vilket är precis samma sak som att $A(p+1)$ är sann. Eftersom vi uppnådde detta med hjälp av induktionsantagandet $A(p)$ kan vi dra slutsatsen att implikationen

$A(p) \Rightarrow A(p+1)$ måste vara sann vilket fullbordar det andra steget i induktionsbeviset.

Nu kanske vi tänker oss att vi har fullbordat *hela* induktionsbeviset. Steg 1 och 2 ger oss kanske en inre syn av att hela situationen liknar dominobrickorna som faller och därmed blir alla utsagor $A(n)$ sanna eftersom de naturliga talen är i princip som en rad dominbrickor. Men vi vill faktiskt också ha denna sista sväng med i beviset, alltså en matematiskt korrekt formulering av "de naturliga talen är i princip som en rad dominbrickor".

Ett bevis är någonting som människorna har hittat på för att övertyga sig själva om vad som är sant respektive falskt. Ett bevis kräver fullständiga motiveringar och om vi läser orden "de naturliga talen är i princip som en rad dominbrickor" som ska vara det sista som krävs för att beviset ovan ska vara klart så får vi känslan av att de orden skrivs där just för att beviset inte är fullständigt ännu. Hade det varit fullständigt hade vi ju inte behövt skriva något alls. Men vi kan inte tala om "dominobrickor" i texten till ett formellt matematiskt bevis. Vi måste ha någonting mer precist.

Ett induktionsbevis är en matematisk utredning av frågan om för vilka naturliga tal n predikatet $A(n)$ är en sann utsaga. Om vi betecknar mängden av tal för vilket predikaten är sant med M så vill vi alltså visa att $M = \mathbb{N}$. Steg 1 och 2 i beviset ger oss att denna mängd har två egenskaper:

(1) $1 \in M$
(2) $p \in M \Rightarrow p+1 \in M$

och här finns det någonting som vi kan hänvisa till som kallas *induktionsaxiomet* som är den matematiska formuleringen av just de naturliga talens grundläggande princip. Induktionsaxiomet säger att om en mängd M har de två egenskaperna ovan så kan vi dra slutsatsen att mängden består av alla naturliga tal, det vill säga med hänvisning till induktionsaxiomet kan vi dra slutsatsen att $M = \mathbb{N}$ och det fullbordar beviset.

Den här långa matematisk-pedagogiska utredningen behöver förstås inte tas med. I vanliga formuleringar av induktionsbevis räcker det att skriva så här:

Steg 3. Steg 1 ger att $A(1)$ är sann. Steg 2 med $p = 1$ ger då $A(1)$ sann $\Rightarrow A(2)$ sann. Och upprepad användning av steg 2 ger

$$A(1) \text{ sann} \Rightarrow A(2) \text{ sann} \Rightarrow A(3) \text{ sann} \Rightarrow \ldots \Rightarrow \forall n \in \mathbb{N} : A(n)$$

och den sista slutsatsen kan dras tack vare steg 1 och 2 *och principen for matematisk induktion.* Beviset är klart.

I vårt formella förhållningssätt till matematiska bevis är det alltså väsentligt att formulera det sista steget. Som ett minimum måste vi skriva:

"*Steg 3.* Steg 1 och 2 och induktionsaxiomet fullbordar beviset."

och det kan anses godkänt men det kan också vara bra att artikulera sig mer fullständigt som i bevistexten som var given här.

Sammanfattningsvis: för att visa ett påstående av formen $\forall n \in \mathbb{N} : A(n)$ består alltså ett induktionsbevis av följande tre steg:

1. Kolla att påståendet håller för startvärdet, det vill säga kolla att $A(1)$ är sann.
2. Visa implikationen $A(p) \Rightarrow A(p+1)$ för alla $p \in \mathbb{N}$. För att visa den här implikationen antar vi först att $A(p)$ är sant för något visst värde på $p \in \mathbb{N}$. Det kallas *induktionsantagandet.* Sedan använder vi induktionsantagandet tillsammans med allt annat som vi vet är sant och visar att också $A(p+1)$ blir sant. Av detta drar vi slutsatsen att implikationen $A(p) \Rightarrow A(p+1)$ måste vara sann.
3. Som sista viktiga steg hänvisar vi till steg 1 och 2 principen för matematiskt induktion (alternativt kallat *induktionsaxiomet*). Vi kan här skriva ned följden av utsagor

   $$A(1) \text{ sann } \Rightarrow A(2) \text{ sann } \Rightarrow A(3) \text{ sann } \Rightarrow \ldots \Rightarrow A(n) \text{ sann för alla } n \in \mathbb{N}.$$

   för en mer utförlig formulering (men som ett minimum måste vi alltid ha en referens till steg 1 och 2 och induktionsaxiomet).

Vi studerar ytterligare ett exempel:

Exempel: Bevisa, med matematisk induktion att $7|5^{2n} - 2^{5n}$ för alla $n \geq 1$.

Bevis: *(Matematisk induktion över n.)* Vi introducerar predikatet $A(n)$ för uttalandet $7|5^{2n} - 2^{5n}$ där $n \in \mathbb{N}$. Uppgiften är att visa att $\forall n \in \mathbb{N} : A(n)$.

Steg 1. Kolla att $A(1)$ är sann. Är det så? Har vi att $5^{2\cdot 1} - 2^{5\cdot 1}$ är delbart med 7? Vi kan räkna ut det och vi får då

$$5^{2\cdot 1} - 2^{5\cdot 1} = 25 - 32 = -7 = 7 \cdot -1$$

och tydligen är $5^{2\cdot 1} - 2^{5\cdot 1}$ delbart med 7 så $A(1)$ är sann. Steg 1 är klart.

Steg 2. Vi ska nu visa implikationen $A(p) \Rightarrow A(p+1)$ för alla naturliga tal p och vi antar därför att $A(p)$ är sant för något naturligt tal p. Detta är induktionsantagandet och enligt det har vi alltså $7|5^{2p} - 2^{5p}$. Det betyder att det existerar ett heltal q sådant att $5^{2p} - 2^{5p} = 7q$. Med stöd av detta vill vi nu visa $A(p+1)$ det vill säga $7|5^{2(p+1)} - 2^{5(p+1)}$. Vi observerar därför att

$$5^{2(p+1)} - 2^{5(p+1)} = 5^{2p+2} - 2^{5p+5} = 25 \cdot 5^{2p} - 32 \cdot 2^{5p} = 25 \cdot 5^{2p} - 25 \cdot 2^{5p} - 7 \cdot 2^{5p}.$$

Men detta tal kan skrivas $25 \cdot (5^{2p} - 2^{5p}) - 7 \cdot 2^{5p}$, och om vi använder induktionstagandet som sade att $5^{2p} - 2^{5p} = 7q$ för något heltal q har vi

$$5^{2(p+1)} - 2^{5(p+1)} = 25 \cdot 7q - 7 \cdot 2^{5p} = 7 \cdot (25q - 2^{5p})$$

vilket helt klart är delbart med 7. Alltså har vi $7|5^{2(p+1)} - 2^{5(p+1)}$ vilket är precis $A(p+1)$. Vi har nu alltså visat implikationen $A(p) \Rightarrow A(p+1)$ vilket fullbordar steg 2.

Steg 3. De två tidigare stegen och principen för matematisk induktion gör att vi kan dra slutsatsen att

$$A(1) \text{ är sann (steg 1)} \Rightarrow A(2) \text{ sann (steg 2)} \Rightarrow \ldots \Rightarrow A(n) \text{ sann för alla } n \geq 1$$

vilket fullbordar beviset.

Vi *kan* också ha andra startvärden än 1. Matematisk induktion fungerar över hela $\mathbb{Z}$. Vi kommer att se det i senare exempel och i några av övningarna nedan.

Övningar

Finan: (Observera att Finan utelämnar hänvisning till induktionsaxiomet/induktionsprincipen i sina bevis vilket strängt taget är ofullständigt!) Exempel 11.1, Sats 11.1, Exempel 11.4, Exempel 11.5, Problem 11.1-11.12, Problem 14.7.

6. Stark Matematisk Induktion

Matematisk induktion handlar om hela följder av utsagor som vi sett. Vi vill alltså visa att alla utsagor på formen

$$A(1), A(2), A(3), A(4), \ldots$$

är sanna. Och ett induktionsbevis där vi hänvisar till induktionsaxiomet baseras på att vi visat implikationen $A(p) \Rightarrow A(p+1)$ där kraften som ger att $A(p+1)$ är sann kommer från den närmast föregående utsagan alltså $A(p)$. Men målet är ju att etablera att alla $A(n)$ är sanna, det vill säga vi vill visa att

$$A(1) \text{ är sann}, A(2) \text{ är sann}, A(3) \text{ är sann}, A(4) \text{ är sann}, \ldots$$

och vi gör det i vanlig matematisk induktion genom att ta induktionssteget där som sagt $A(p+1)$ blir sann tack vare att $A(p)$ är sann: utsagan $A(p+1)$ blir alltså sann tack vare att den *närmast föregående* utsagan $A(p)$ är sann och vi får kaskaden

$$A(1) \text{ är sann}, A(2) \text{ är sann}, A(3) \text{ är sann}, A(4) \text{ är sann}, \ldots, A(p) \text{ är sann} \Rightarrow A(p+1) \text{ är sann}, \ldots$$

Men det skulle kunna finnas mer kraft i att utnyttja att *samtliga* föregående utsagor $A(1), \ldots, A(p)$ är sanna. Detta är principen bakom så kallad *stark matematisk induktion.*

Vi formulerar det som en alternativ induktionsprincip: (Vi har ett mer allmänt startvärde här, n_0, som skulle kunna vara 1, men vi ger en mer allmän formulering av stark matematisk induktion. På samma sätt skulle vanlig matematisk induktion kunnat starta på något annat heltal än 1.)

Principen för stark matematisk induktion: Givet ett uttalande $A(n)$ för varje heltal n. Om följande två villkor är uppfyllda:

1. $A(n_0)$ är sant för något heltal n_0;
2. För alla $p \geq n_0$ gäller implikationen $A(n_0) \wedge A(n_0+1) \wedge \ldots \wedge A(p) \Rightarrow A(p+1)$.

Då är $A(n)$ sann för alla $n \geq n_0$, det vill säga $\forall n \geq n_0 : A(n)$.

Vi ska använda den starka formen av matematisk induktion för att visa att alla ntaurliga tal större än 2 kan skrivas som en produkt av primtal. Vi har redan visat detta med Aritmetikens Fundamentalsats men vi studerar en annan formulering här.

Sats: Varje naturligt tal större än 2 kan skrivas som en produkt av primtal.

Bevis: Vi använder stark matematisk induktion och inför först predikatet

$$A(n) \Leftrightarrow n \text{ är en produkt av primtal.}$$

Det vi vill ska visa kan då uttryckas $\forall n \in \mathbb{N} : n \geq 2 \Rightarrow n$ är en produkt av primtal.

Vi tar nu de tre stegen i ett induktionsbevis:

1. Startvärdet är $n_0 = 2$ och $A(2)$ är sant eftersom 2 själv är ett primtal, därmed är det också en produkt av primtal innehållandes bara en faktor nämligen sig själv.

2. Fixera nu ett godtyckligt tal $p \geq 2$ antag att A gäller för alla naturliga tal, $2, 3, \ldots, p$, det vill säga anta att $A(2), A(3), \ldots, A(p)$ alla gäller. Då kanske $p+1$ redan är ett primtal och då blir $p+1$ (liksom 2 i steg 1) en produkt av primtal så då är $A(p+1)$ sann. Men om $p+1$ *inte* är ett primtal så finns det naturliga tal a, b med $2 \leq a, b < p+1$ sådana att $p+1 = a \cdot b$. Enligt induktionsantagandet gäller både $A(a)$ och $A(b)$ men det innebär ju att både a och b är produkter av primtal så blir även $p+1 = a \cdot b$ en produkt av primtal, det vill säga $A(p+1)$ är sann. Vi har nu alltså visat implikationen $A(2) \wedge A(3) \wedge \ldots \wedge A(p) \Rightarrow A(p+1)$ vilket fullbordar det andra steget i beviset.

3. Vi hänvisar nu till principen för stark matematisk induktion och steg 1 och steg 3 som ger att vi kan dra slutsatsen

   $$A(2) \text{ sann } \Rightarrow A(3) \text{ sann } \Rightarrow \ldots \Rightarrow A(n) \text{ sanna för alla } n \geq 2$$

   vilket fullbordar beviset.

Den här satsen är *svagare* än Aritmetikens Fundamentalsats eftersom den här satsen uttrycker att varje tal är en produkt av primtal men inte att den primtalsfaktoriseringen är *entydigt* bestämd som Aritmetikens Fundamental sats också uttryckte.

Övning

6.7.1 Visa att varje positivt heltal större än 7 kan skrivas som en summa

$$x \cdot 3 + y \cdot 5$$

där $x \geq 0$ och $y \geq 0$ är heltal. (Till exempel gäller $8 = 1 \cdot 3 + 1 \cdot 5$, $9 = 3 \cdot 3 + 0 \cdot 5$, $10 = 0 \cdot 3 + 2 \cdot 5$ osv.)

7. Matematisk induktion och välordningsprincipen

Då har vi alltså två sorters induktion: stark matematisk induktion och (vanlig) matematisk induktion. Det kommer att visa sig att dessa båda principer är ekvivalenta. Men för att se att det är ekvivalenta ska vi införa en tredje princip: välordningsprincipen.

Det här avsnittet är mer teoretiskt än de tidigare avsnitten och om läsaren har svårt att ta till sig detta avsnitt kan hen vänta med det här till senare. Det kan också hända att läraren till kursen anser att detta inte ingår i kursen.

Sats: *(Välordningsprincipen för de naturliga talen.)* Varje icketom delmängd av $\mathbb{N}$ har ett minsta element.

Principerna för matematisk induktion respektive stark matematisk induktion kan ges formuleringarna:

* *(Matematisk induktion.)* Om mängden $M \subset \mathbb{N}$ uppfyller villkoren
  1. $1 \in M$
  2. $p \in M \Rightarrow p+1 \in M$

  så gäller $M = \mathbb{N}$

* *(Stark matematisk induktion.)* Om mängden $M \subset \mathbb{N}$ uppfyller villkoren
  1. $1 \in M$
  2. $1 \in M \wedge 2 \in M \wedge \ldots \wedge p \in M \Rightarrow p+1 \in M$

  så gäller $M = \mathbb{N}$

(I den här formuleringen av stark matematisk induktion har vi valt att ha startvärdet 1, men formuleringar kring startvärden i de grundläggande formuleringarna av principerna spelar ingen roll. Vi kan lätt visa att den här formuleringen är ekvivalent med den tidigare.)

Vi ska nu visa att de här tre principerna alla är ekvivalenta. Vi kommer att göra detta i form av att de alla implicerar varandra, vi kommer att se att matematisk induktion $\Rightarrow$ välordningsprincipen $\Rightarrow$ stark matematisk induktion $\Rightarrow$ matematisk induktion.

Bevis 1: *(Matematisk induktion medför välordningsprincipen.)* Det här beviset blir ett vanligt induktionsbevis för att varje icketom delmängd av $\mathbb{N}$ har ett minsta element. Vi skapar först att bevis för alla ändliga delmängder av $\mathbb{N}$. Vi inför därför predikatet

$$A(n) \Leftrightarrow \text{alla delmängder av } \mathbb{N} \text{ som har } n \text{ element har ett minsta element}$$

och vi ska nu visa $\forall n \geq 1 : A(n)$.

Steg 1. Är $A(1)$ sann? Ja, $A(1)$ säger att alla mängder som består av *precis ett* element har ett minsta element och det är ju sant, det enda ingående elementet i mängden med ett element måste ju då också vara det minsta.

Steg 2. Antag att $A(p)$ är sant, det vill säga antag att alla mängder med p element i $\mathbb{N}$ har ett minsta element – detta är induktionsantagandet. Låt nu M vara en godtycklig delmängd av $\mathbb{N}$ med $p+1$ element. Välj ut ett element ur M, vilket som helst, vi kan kalla det x. Nu har vi alltså

$$M = \{x\} \cup (M - \{x\})$$

och mängden $M - \{x\}$ är en mängd med p element. Alltså har den ett minsta element enligt induktionsantagandet, kalla detta minsta element för y. Det är då klart att M har ett minsta element som blir det minsta av elementen x och y. Alltså har M ett minsta element och $A(p+1)$ är alltså sann. Vi drar slutsatsen att implikationen $A(p) \Rightarrow A(p+1)$ är sann vilket fullbordar steg 2 i beviset.

Steg 3. Steg 1 och 2 och principen för matematisk induktion fullbordar beviset av $\forall n \in \mathbb{N} : A(n)$, det vill säga alla ändliga icketomma delmängder av $\mathbb{N}$ har ett minsta element. Detta fullbordar induktionsdelen av beviset.

Steg 4. Steg 1-3 tar hand om alla *ändliga* delmängder av $\mathbb{N}$ men satsen uttrycker sig om alla mängder, vi måste alltså hantera fallet med oändliga mängder särskilt och visa att också de har ett minsta element. Men det är enkelt. Om M är en mängd med oändligt många element så kan vi bara välja något element $z \in M$ med egenskapen att det finns $x \in M$ med $x < z$. Det går eftersom M har oändligt många element. Bilda nu

$$E = M - \{t \in M : t < z\}$$

det vill säga alla element i M som är mindre än z. Mängden E är då icketom och ändlig (eftersom den ligger i *mathbbN* och är uppåt begränsad) vilket innebär enligt första delen av det här beviset att det finns ett minsta element i E. Det är klart att detta minsta element också måste vara minsta element i hela M. Alltså har M ett minsta element även i fallet då M har oändligt många element.

I båda fallet (då M är ändlig eller oändlig) har alltså M ett minsta element. Välordningsprincipen är därmed visad med hjälp av principen för matematisk induktion. Alltså kan vi uppfatta välordningsprincipen som en följd av principen för matematisk induktion vilket fullbordar bevis 1.

Bevis 2: *(Välordningsprincipen medför principen för stark matematisk induktion.)* Vi utgår nu från välordningsprincipen, det vill säga att varje ickteom delmängd av $\mathbb{N}$ alltid har ett minsta element och visar att från detta följer principen för stark matematisk induktion. Vi låter därför M vara en delmängd av $\mathbb{N}$ som uppfyller

1. $1 \in M$

2. $1 \in M \wedge 2 \in M \wedge \ldots \wedge p \in M \Rightarrow p+1 \in M$

för att principen för stark matematisk induktion ska gälla så ska vi nu visa att detta innebär att $M = \mathbb{N}$. Vi antar därför motsatsen, antag alltså att det finns element i $\mathbb{N}$ som *inte* ingår i M. Vi arbetar med $\mathbb{N}$ som universum här och det betyder att M^c är icketom. Men en icketom delmängd av $\mathbb{N}$ har ett minsta element, kalla detta minsta element för x. Att x är minsta element i M^c betyder förstås att alla element som är mindre än x *ingår* i M, vi har alltså situationen

1. $1 \in M$
2. $1 \in M \wedge 2 \in M \wedge \ldots \wedge x-1 \in M$.

Men då kan vi använda de egenskaperna för M som finns ovan för $p = x-1$ och alltså dra slutsatsen $p+1 = x-1+1 = x \in M$ vilket motsäger att $x \notin M$. Antagandet om att M^c skulle vara icketom måste alltså vara falskt vilket innebär att $M^c = \emptyset \Rightarrow M = \mathbb{N}$. Detta visar att principen for stark matematisk induktion gäller och eftersom vi visade den med hjälp av välordningsprincipen så kan vi uppfatta principen för stark matematisk induktion som en följd av välordningsprincipen. Detta fullbordar bevis 2.

Bevis 3: *(Principen för stark matematisk induktion medför principen för (vanlig) matematisk induktion.)* I detta sista bevis utgår vi från principen för stark matematisk induktion och visar att av den följer principen för matematisk induktion. Vi antar alltså att följande gäller:

(Stark matematisk induktion.) Om mängden $M \subset \mathbb{N}$ uppfyller villkoren

1. $1 \in M$
2. $1 \in M \wedge 2 \in M \wedge \ldots \wedge p \in M \Rightarrow p+1 \in M$

så gäller $M = \mathbb{N}$

utgående från detta ska vi visa

(Matematisk induktion.) Om mängden $M \subset \mathbb{N}$ uppfyller villkoren

(a) $1 \in M$
(b) $1 \in M \wedge 2 \in M \wedge \ldots \wedge p \in M \Rightarrow p+1 \in M$

så gäller $M = \mathbb{N}$

Vi har ändrat benämningen av punkterna som definierar matematisk induktion till *(a)* och *(b)* för tydlighetens skull.

Vi ska alltså visa att principen för matematisk induktion är uppfylld så vi låter M vara en godtycklig delmängd av $\mathbb{N}$ som uppfyller *(a)* och *(b)*. Att *(a)* är uppfyllt betyder att krav 1 i principen för stark matematisk induktion är uppfyllt. Vidare, för att undersöka om också 2:an gäller, antag att mängden M uppfyller $1 \in M \wedge 2 \in M \wedge \ldots \wedge p \in M$. Vi är föstås fria att bara släppa de $p-1$ första utsagorna så att detta medför $p \in M$. Men då får vi $p+1 \in M$ enligt *(b)* och detta betyder alltså att implikationen

$$1 \in M \wedge 2 \in M \wedge \ldots \wedge p \in M \Rightarrow p+1 \in M$$

är uppfylld och det är krav 2 i principen för stark matematisk induktion. Den ger oss då alltså att $M = \mathbb{N}$ enligt principen för stark matematisk induktion som vi har antagit gäller. Men då har vi alltså utgått från de två kraven *(a)* och *(b)* i principen för matematisk induktion och visat $M = \mathbb{N}$. Eftersom det också är slutsatsen för matematisk induktion så gäller även matematisk induktion. Vi har alltså visat att principen för matematisk induktion följer av principen för stark matematisk induktion vilket fullbordar bevis 3.

De tre bevisen tagna tillsammans visar att alla tre principer är ekvivalenta. (Så "stark" matematisk induktion är egentligen inte starkare än vanlig induktion.)

Med hjälp av välordningsprincipen kan vi ge ett bevis för divisionsalgoritmen:

Sats 3.6 *(Divisionsalgoritmen.)* Låt n vara ett givet heltal. För varje positivt heltal d finns då entydigt bestämda tal q och r med $0 \leq r \leq d-1$ sådana att

$$n = q \cdot d + r.$$

Bevis: Vi kommer att använda en speciell variant av välordningsprincipen för heltalen som lyder så här:

Välordningsprincipen för heltalen: Varje nedåt begränsad icketom delmängd av $\mathbb{Z}$ har ett minsta element.

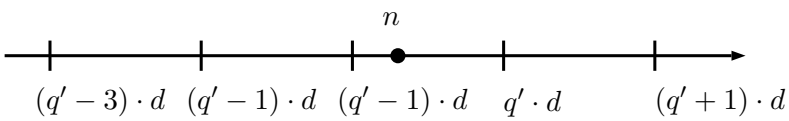
Låt nu n vara heltalet i satsens förutsättningar och bilda mängden

$$E = \{t \in \mathbb{Z} : n < t \cdot d\}.$$

Den här mängden är icketom eftersom t kan väljas hur stort som helst och alltid garanterat kunna uppfylla $n < t \cdot d$. Vidare är mängden nedåt begränsad eftersom n är ett fixt tal. Om t väljs tillräckligt stort negativt kommer olikheten $n < t \cdot d$ till slut att inte vara uppfylld. Vidare kommer den inte heller vara uppfylld för några mindre tal än detta stora negativa tal. Eftersom E är icketom och nedåt begränsad finns ett minsta tal i E som vi kallar q'. Att q' är det minsta talet i E betyder att

$$n < q' \cdot d \quad \text{och} \quad n \geq (q' - 1) \cdot d.$$

Vi illustrerar situationen i en figur:



alla multiplar av d ligger på jämna avstånd längs med hela tallinjen och någonstans finns talet n mellan två av dessa multiplar (eller exakt på). Med välordningsprincipen tar vi fasta på q' som är den minsta multipeln som är större än n, den finns tack vare välordningsprincipen.

Om vi nu sätter $q = q' - 1$ och $r = n - q \cdot d$ har vi alltså funnit q och r som uppfyller kraven i satsen. Det återstår att visa att q och r är entydigt bestämda av kravet $0 \leq r \leq d - 1$ och därför antar vi att det finns två uppsättningar av tal, q_1, r_1 och q_2, r_2 som uppfyller kraven

$$n = q_1 \cdot d + r_1 \quad (0 \leq r_1 \leq d - 1) \quad \text{och} \quad n = q_2 \cdot d + r_2 \quad (0 \leq r_2 \leq d - 1).$$

Vi visar två implikationer: $q_1 = q_2 \Rightarrow r_1 = r_2$ och $r_1 = r_2 \Rightarrow q_1 = q_2$.

1. Om $r_1 = r_2 = r$ så får vi
$$0 = n - n = (q_1 \cdot d + r) - (q_2 \cdot d + r) = (q_1 - q_2) \cdot d \Rightarrow q_1 - q_2 = 0 \Rightarrow q_1 = q_2.$$
2. Om å andra sidan $q_1 = q_2 = q$ så få vi
$$0 = n - n = (q \cdot d + r_1) - (q \cdot d + r_2) = (q - q) \cdot d + r_1 - r_2 \Rightarrow r_1 - r_2 = 0 \Rightarrow r_1 = r_2.$$

Vi kan alltså dra slutsaten att antingen gäller $r_1 = r_2$ och $q_1 = q_2$ och då är satsen fullständigt visad, eller så gäller $r_1 \neq r_2$ och $q_1 \neq q_2$. Det gäller alltså att visa att detta sista fall inte kan gälla, så vi studerar det fallet i detalj. Vi får då

$$0 = n - n = (q_1 \cdot d + r_1) - (q_2 \cdot d + r_2) = (q_1 - q_2) \cdot d + (r_1 - r_2) \Rightarrow (q_1 - q_2) \cdot d = (r_2 - r_1) \Rightarrow d | r_2 - r_1$$

men det här är en motsägelse eftersom talet $r_2 - r_1$ måste ligga i intervallet $-d + 1, \ldots, d - 1$ och här finns bara 0 som multipel av d. Det enda sättet som vi kan ha $d | r_2 - r_1$ är om $r_2 - r_1 = 0$ vilket innebär att $r_2 = r_1$ vilket motsäger vilket motsäger att både $q_1 \neq q_2$ och $r_1 \neq r_2$ ska gälla. Beviset är fullbordat.

8. Rekursivt definierade talföljder

Vi ska nu använda matematisk induktion och talteori för att studera rekursivt definierade följder av tal. Vi skapar en sådan genom att ange starttalen och sen anger vi hur alla talen som kommer efter beror på de tal som kommer tidigare. Vi ger ett exempel.

Exempel: Låt $a_0 = 1$ och $a_{n+1} = (n + 1) \cdot a_n$ för alla $n \geq 0$. Då blir a_1 definerad av $1 \cdot a_0 = 1 \cdot 1 = 1$. När a_1 är bestämd kan vi räkna ut nästa som är a_2 och det blir $a_2 = 2 \cdot a_1 = 2 \cdot 1 = 2$. Processen forstätter förstås i all oändlighet:

$$a_3 = 3 \cdot 2 = 6, \qquad a_4 = 4 \cdot 3 = 24, \qquad a_5 = 5 \cdot 24 = 120, \ldots$$

och det allmänna mönstret blir

$$a_n = n \cdot (n - 1) \cdot \ldots \cdot 2 \cdot 1.$$

Den här talföljden, eller den här processen, att multiplicera alla tal från och med n ner till 1 är en viktig process och vi illustrerar alltså här en viktig talföljd. Talet $n \cdot (n - 1) \cdot \ldots \cdot 2 \cdot 1$ kallas *fakulteten* av n, eller n*-fakultet*. (På engelska kallas det *factorial*.) Vi betecknar n-fakultet med symbolen $n!$ och vi ser alltså av det

ovanstående exemplet att $a_n = n!$.

Notera att när vi definierar en talföljd rekursivt så innehåller definitionen en referens till det föregående talet. Det här kommer att passa mycket bra ihop med induktion där andra steget ju också handlar om att referera till föregående fall. Vi tar ett exempel på det.

Exempel: Visa att för alla $n \geq 4$ gäller $n! \geq 2^n$.

Lösning: Vi använder matematisk induktion: introducera predikatet

$$A(n) \Leftrightarrow n! \geq 2^n$$

för alla naturliga tal n. Det som vi ska visa kan formuleras $\forall n \geq 4 : n! \geq 2^n$. Nu tar vi de tre stegen i ett induktionsbevis:

(1) Startvärdet är 4 så kontrollera alltså att $A(4)$ är sann. Vi ska alltså säkerställa att $4! \geq 2^4$. Vi räknar ut båda leden och får att detta är ekvivalent med $24 \geq 16$ ($4! = 24$ och $2^4 = 16$) och det här är förstås sant.

(2) Vi ska nu visa implikationen $A(p) \Rightarrow A(p+1)$ för alla $p \geq 4$ så vi gör induktionsantagandet att $A(p)$ är sant för ett visst $p \geq 4$. Då gäller alltså $A(p) \Leftrightarrow p! \geq 2^p$. Med hjälp av detta ska vi nu visa att $A(p+1)$ också är sann. Vi sätter upp vänster och höger led av $A(p+1)$:

$$VL_{p+1} = (p+1)! = (p+1) \cdot p! \qquad HL_{p+1} = 2^{p+1} = 2 \cdot 2^p$$

Kan vi nu se att $VL_{p+1} \geq HL_{p+1}$ med hjälp av induktionsantagandet som var $VL_p = p! \geq 2^p = HL_p$? Ja här kan vi se hur matematisk induktion passar bra ihop med rekursion, VL_{p+1} innehåller precis $p!$ som ju förekommer i induktionsantagandet och att induktionsantagandet gäller innebär att vi kan skatta $VL_{p+1} = (p+1) \cdot p!$ nedåt med $(p+1) \cdot 2^p$. Så vi har

$$VL_{p+1} = (p+1) \cdot p! \geq (p+1) \cdot 2^p$$

tack vare $A(p) \Leftrightarrow p! \geq 2^p$. Men vi har ju också att $p \geq 4$ så vi kan fortsätta göra skattningen $(p+1) \cdot 2^p \geq (4+1) \cdot 2^p = 5 \cdot 2^p \geq 2 \cdot 2^p = 2^{p+1} = HL_{p+1}$. Så sammantaget ger induktionsantagandet *och* $p \geq 4$ att $VL_{p+1} \geq 2^{p+1}$ håller, och detta är ju förstås $A(p+1)$. Vi har med det här visat att för $p \geq 4$ gäller $A(p) \Rightarrow A(p+1)$ vilket fullbordar induktionssteget.

(3) De två ovanstående stegen innebär att

$$A(4) \text{ är sann (steg 1)} \Rightarrow A(5) \text{ är sann } (A(4)\text{+steg 2}) \Rightarrow A(6) \text{ är sann } (A(5)\text{+steg 2}) \Rightarrow \ldots$$

$$\Rightarrow \forall n \geq 4 : A(n)$$

och detta resonemang håller tack vare steg 1 och 2 och principen för matematisk induktion.

Det här var ett induktionsbevis där vi inte hade startvärdet 1, vi startade på 4, men principen för matematisk induktion gäller egentligen för alla heltal och utsagor som är formulerade i form av en rad av påståenden med ett startvärde. Det är *strukturen* på hur utsagorna hänger ihop som är viktig (som att dominobrickorna ska stå tillräckligt nära varandra), det är inte viktigt hur vi numrerar dem.

Vi ska nu studera två speciella typer av rekursivt definierade talföljder. Vi ger dessa i samma definition. I definitionen inför vi också ett skrivsätt för en talföljd. Vi skriver så här

$$\{a_n\}_{n=1}^{\infty}$$

då vi betecknar en följd av tal där talen får namnen a_1, a_2, a_3 och så vidare.

Definition: En följd av tal $\{a_n\}_{n=1}^{\infty}$ kallas en *aritmetisk talföljd* om det finns en konstant d sådan att $a_{n+1} = a_n + d$ för alla $n \in \mathbb{N}$. En följd av tal $\{a_n\}_{n=1}^{\infty}$ kallas en *geometrisk talföljd* om det finns en konstant a sådan att $a_{n+1} = a \cdot a_n$ för alla $n \in \mathbb{N}$. Talet a kallas då talföljdens *kvot*.

Den aritmetriska följden skapas alltså genom att successivt addera ett visst tal d för att skapa nästa tal i följden. Den geometriska följden skapas genom att successivt multiplicera att visst tal a för att skapa nästa tal i följden. När det gäller den geometriska följden så får talet a namnet *kvot*, men talet d i den aritmetiska följden saknar namn. (Valet av bokstaven d för tanken till ordet "distans" och i en aritmetisk följd ligger talen

på en konstant distans från varandra.)

Vi ger en sats om dessa typer av följder.

Sats: Låt $\{a_n\}_{n=1}^{\infty}$ vara en aritmetisk talföljd och låt $\{b_n\}_{n=1}^{\infty}$ vara en geometrisk talföljd. För alla $n \in \mathbb{N}$ gäller då $a_n = a_1 + (n-1) \cdot d$ och $b_n = b_1 \cdot b^{(n-1)}$, där b är kvoten för $\{b_n\}$.

Bevis: *Vi visar endast påståendet om $\{a_n\}$ och lämnar påståendet om $\{b_n\}$ som en övning.* Detta är ett induktionsbevis så vi introducerar predikatet

$$A(n) \Leftrightarrow a_n = a_1 + (n-1) \cdot d.$$

Vår uppgift är att visa $\forall n \in \mathbb{N} : A(n)$. Vi tar nu de tre stegen som är nödvändiga för induktionen.

1. $A(1) \Leftrightarrow a_1 = a_1 + (1-1) \cdot d \Leftrightarrow a_1 = a_1$ vilket är uppenbart sant.

2. Vi går nu vidare och visar implikationen $A(p) \Rightarrow A(p+1)$ och antar därför att $A(p)$ är sant för ett visst $p \in \mathbb{N}$. Det är ekvivalent med att

   $$VL_p = a_p = a_1 + (p-1) \cdot d = HL_p.$$

   Baserat på detta ska vi nu visa $A(p+1)$, det vill säga $VL_{p+1} = HL_{p+1}$ så vi studerar VL_{p+1}:

   $$VL_{p+1} = a_{p+1} = \{\textit{enligt definitionen av aritmentisk talföljd}\} = a_p + d = VL_p + d$$

   men detta tal är $HL_p + d$, *enligt induktionsantagandet* ($A(p) \Leftrightarrow VL_p = HL_p$), så hela uttrycket är $a_1 + (p-1) \cdot d + d = a_1 + p \cdot d = a_1 + (p+1-1) \cdot d = HL_{p+1}$ så att $VL_{p+1} = HL_{p+1}$ vilket precis är $A(p+1)$. Vi har med detta visat att implikationen $A(p) \Rightarrow A(p+1)$ alltid gäller för alla $p \in \mathbb{N}$.

3. Steg 1 och 2 tillsammans med principen för matematisk induktion ger att vi kan dra slutsatsen $\forall n \in \mathbb{N} : A(n)$ vilket fullbordar beviset.

Övningar

Finan: Exempel 11.4, Exempel 11.6, Problem 11.13, Exempel 22.1, Exempel 22.3, Exempel 22.4, Exempel 22.5, Exempel 22.6, Exempel 22.7, Exempel 22.9, Exempel 22.10, Exempel 22.11, Exempel 22.13, Problem 22.1, Problem 22.2, Problem 22.3, Problem 22.6, Problem 22.7, problem 22.8, Problem 22.11, Problem 22.12, Problem 22.13.

9. Differensekvationer

Tänk tillbaka ett tag på talföljden som definerades i början av förra avsnittet. Vi hade där

$$a_0 = 1 \qquad \text{och} \qquad a_{n+1} = (n+1) \cdot a_n$$

och vi kunde se ett mönster så att den rekursivt definerade talföljdens tal kunde uttryckas explicit, alltså inte rekursivt och vi kunde då skriva $a_n = n!$ där alltså högerledet – $n!$ – inte innehåller några uttryck med a_n. Vi ska nu studera hur det kan göras allmänt för en viss typ av rekursivt definerade talföljder. Tekniken kommer att vara fullständigt analog med tekniken för att lösa differentialekvationer och här kommer vi att kalla det hela för att "lösa differensekvationer" men det kommer bara handla om att hitta en explicit formel för en rekursivt definerad talföljd. Vi gör en definition för skapa klarhet.

Definition: Identiteten $a_{n+2} = A \cdot a_{n+1} + B \cdot a_n$, där $\{a_n\}_{n=1}^{\infty}$ är en talföljd och A och B är konstanta tal, kallas en *linjär homogen differensekvation med konstansta koefficienter av andra ordningen.* Vi kan omväxlande välja att ange startvärdena (som då kan anges med till exempel $a_0 = a$ och $a_1 = b$ för några tal a och b) eller bara betrakta identiteten utan starttal. Processen att omvandla den rekursiva definitionen av differensekvationen till en explicit formel som anger vad varje tal a_n är (utan referens till tidigare tal i följden) kallas att *lösa* differensekvationen.

Att *lösa* differensekvationen $a_{n+2} = A \cdot a_{n+1} + B \cdot a_n$ kommer då involvera den så kallade *karakteristiska ekvationen* som för andra gradens differensekvationer på ovanstående form får utseendet

$$x^2 = A \cdot x + B$$

och den allmänna lösningen, alltså det explicita för vad a_n är ges av

$$a_n = C \cdot r_1^n + D \cdot r_2^n \qquad \text{respektive} \qquad a_n = (C \cdot n + D) \cdot r^n$$

beroende på om den karakteristiska ekvationen har två enkelrötter (r_1 och r_2) eller en dubbelrot (r). Konstanterna C och D bestäms av starttalen $a_0 = a$ och $a_1 = b$ som är helt analogt med det som kallas "begynnelsevillkor" i teorin för differentialekvationer. (Även de båda lösningsformlerna för de explicita uttrycken för a_n är i analogi med teorin för differentialekvationer.)

Vi tar två typiska exempel:

Exempel: Finn ett explicit uttryck för talen i talföljden som defineras av $a_0 = 3, a_1 = 2, a_{n+2} = 3a_{n+1} - 2a_n$.

Lösning: Vi ska alltså lösa differensekvationen $a_{n+2} = 3a_{n+1} - 2a_n$ som tydligen har den karakteristiska ekvationen $x^2 = 3 \cdot x - 2 \Leftrightarrow x = 1 \vee x = 2$, det vill säga vi har två enkelträtter $r_1 = 1$ och $r_2 = 2$. Alltså är en explicit formel för a_n given av

$$a_n = C \cdot 1^n + D \cdot 2^n.$$

Om vi $n = 0$ och $n = 1$ så få vi uttrycken för a_0 respektive a_1 och startvärdena $a_0 = 3, a_1 = 2$ ger oss då de båda ekvationerna $C + D = 3$ och $C + 2 \cdot D = 2$ som är ett linjärt ekvationssystem med två variabler och två ekvationer som har lösningarna $C = 4$ och $D = -1$. Detta insatt i formeln för $\{a_n\}$ ger att lösningen till differensekvationen är $a_n = 4 - 2^n$.

Och här kommer ett exempel med en dubbelrot:

Exempel: Lös differensekvationen $a_{n+2} = 4a_{n+1} - 4a_n$, $n \geq 0$, där $a_0 = 1$ och $a_1 = 4$.

Lösning: Den karakteristiska ekvationen har utseendet $x^2 = 4x - 4 \Leftrightarrow (x-2)^2 = 0$ som har dubbelroten $x = 2$. Alltså är den allmänna lösningen till differensekvationen given av $a_n = (C \cdot n + D) \cdot 2^n$ och om vi sätter $n = 0$ och $n = 1$ uppstår ekvationerna $(C \cdot 0 + D) \cdot 2^0 = a_0 = 1$ och $(C \cdot 1 + D) \cdot 2^1 = 4$ från starttalen och dessa kan skrivas

$$D = 1 \wedge C + D = 2 \Leftrightarrow C = D = 1$$

så den explicita lösningen till den givna differensekvationen är $a_n = (C \cdot n + D) \cdot 2^n = (n+1) \cdot 2^n$.

Vi ger ett klassiskt exempel:

Exempel: Talföljden definierad av

$$a_0 = 1, a_1 = 1, a_{n+2} = a_{n+1} + a_n, \text{ för alla } n \geq 0.$$

kallas *Fibonaccis talföljd* och om vi skriver ut talen ser det ut så här:

$$1, 1, 2, 3, 5, 8, 13, 21, \ldots.$$

Följden bildas alltså genom att sätta de två första talen till 1 och sedan bildas de följande talen genom att varje tal sätts till summan av de två föregående. Vi hittar ett explicit uttryck för a_n genom att betrakta det hela som en differensekvation som vi löser.

Den karakteristiska ekvationen har utseendet $x^2 = x + 1$ som har lösningarna $x = \frac{1 \pm \sqrt{5}}{2}$. Det här är två enkelrötter så den allmänna lösningen har formen

$$a_n = C \cdot \left(\frac{1+\sqrt{5}}{2}\right)^n + D \cdot \left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Om vi sätter $n = 0, n = 1$ så får vi från startvärdena ekvationerna $C+D = 1$ respektive $C \cdot (\frac{1+\sqrt{5}}{2}) + D \cdot (\frac{1-\sqrt{5}}{2}) = 1$ och det här ekvationssystemet har lösningarna $C = \frac{1}{\sqrt{5}} \frac{1+\sqrt{5}}{2}$, $D = -\frac{1}{\sqrt{5}} \frac{1-\sqrt{5}}{2}$, och detta ger den explicita lösningen

$$a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}.$$

Det är ganska anmärkningsvärt att det explicita uttrycket för talen är ganska komplicerade med potenser och en massa rotuttryck men att ändå, när vi sätter in $n = 0, 1, 2, \ldots$, så försvinner alla rotuttryck och allt går jämnt ut och bildar den vackra följden av naturliga tal som alltså utgör Fibonaccis talföljd:

$$1, 1, 2, 3, 5, 8, \ldots$$

(*I vissa andra framställningar anges de första talen i följden som 0 och 1 och inte som här 1 och 1, det ger inga större skillnader mot den här framställningen, alla tal blir bara förskjutna ett steg.*)

Övningar

Finan: Problem 22.14, Problem 22.15.

6.9.1 Det är i själva verket väldigt lätt att skapa övningsuppgifter för differensekvationer. Gör så här:

1. Välj två heltal, till exempel 2 och 3.
2. Teckna andragradsekvationen som har dessa två tal som rötter, i det här fallet $(x-2)(x-3)=0 \Leftrightarrow x^2-5x+6=0$.
3. Byt x^2 mot a_{n+2}, x mot a_{n+1} och 1 mot a_n och lös ut a_{n+2} uttryckt i a_{n+1} och a_n. I detta fall fås $x^2-5x+6=0 \rightarrow a_{n+2}-5a_{n+1}+6a_n=0 \Leftrightarrow a_{n+2}=5a_{n+1}-6a_n$.
4. Den resulterande ekvationen tillsammans med startvillkor utgör sedan en differensekvation som kan lösas. Ofta väljer vi startvillkoren så att lösningen blir enkel att hantera. Till exempel kan vi sätta $a_0=2$ och $a_1=5$. Detta ger oss $a_n=2^n+3^n$. Det går förstås lika lätt att skapa problem med dubbelrötter.

Ta en studiekamrat till hjälp och skapa differensekvationer åt varandra.Var och en gå igenom de 4 stegen ovan och ger den andra personen en differensekvation att lösa. Samråd om era lösningar. (Din kamrat kommer alltid att ha facit!)

Blandade övningar

Gå direkt på gamla tentauppgifter – det finna många heltalsproblem särskilt med induktion.