

Sammanfattning av flervariabelanalysens grunder

Andreas Rejbrand, sommaren 2014

Innehållsförteckning

1	INLEDNING	5
2	DELMÄNGDER AV \mathbb{R}^n	6
2.1	UPPVÄRMNING	6
2.1.1	En väldigt trevlig notation.....	6
2.2	TOPOLOGISKA EGENSKAPER	7
2.2.1	Kompakta mängder.....	9
2.2.2	Rand, inre och slutet hölje (*).....	11
3	KURVOR OCH YTOR	13
3.1	KURVOR I PLANET	13
3.1.1	Fler kurvor	14
3.1.1.1	Ellipser	14
3.1.1.2	Räta linjer	15
3.1.1.3	Hyperbler.....	15
3.1.1.4	Hyperbler, igen! (*)	16
3.1.1.5	Parabler	17
3.1.1.6	Spiraler och annat (*)	17
3.2	KURVOR I RUMMET	18
3.3	MER OM KURVOR	19
3.4	YTOR I RUMMET	21
3.4.1	Fler ytor	21
3.4.1.1	Ellipsoider	21
3.4.1.2	Plan.....	22
3.4.1.3	Paraboloider	23
3.4.1.4	Koner	26
3.4.1.5	Cylindrar	26
3.4.1.6	Hyperboloider.....	27
3.4.1.7	Hyperbolisk paraboloid	28
3.4.2	Mer om ytor	29
3.5	SKÄRNINGSKURVOR.....	31
4	OMRÅDEN I PLANET OCH RUMMET	32
4.1	PLANA OMRÅDEN	32
4.2	OMRÅDEN I RUMMET	34
5	SKALÄRFÄLT	36
5.1	KURVOR OCH YTOR SOM NIVÅMÄNGDER	37
5.2	PARTIELLA DERIVATOR	38
5.3	RIKTNINGSDERIVATA	39
6	VEKTORFÄLT	40
7	SKALÄRFÄLT, FORTSÄTTNING	42
7.1	GRADIENT	42
7.2	GRADIENT OCH NIVÅMÄNGDER	44
7.2.1	Fler exempel	47
7.3	SAMMANSÄTTNINGAR AV FUNKTIONER – KURVOR I LANDSKAP.....	48
7.4	GRÄNSVÄRDEN	51

7.5	LOKALA MAX- OCH MINPUNKTER	57
7.5.1	<i>Repetition av envariabelfallet</i>	57
7.5.1.1	Funktioner definierade på äkta delmängder av \mathbb{R}	59
7.5.2	<i>Flervariabelfallet</i>	60
7.5.2.1	Olika typer av stationära punkter.....	61
7.5.2.1.1	Lokalt maximum	62
7.5.2.1.2	Lokalt minimum.....	62
7.5.2.1.3	Typisk sadel.....	62
7.5.2.1.4	Färlig sadel.....	63
7.5.2.1.5	Fåtölj?	63
7.5.2.1.6	Terrass	63
7.5.2.1.7	Dike/ås.....	64
7.5.2.2	Max och min på delmängder.....	64
7.6	LOKALA UNDERSÖKNINGAR – DET LÄTTASTE FALLET	65
7.7	LOKALA UNDERSÖKNINGAR – DET ALLMÄNNA FALLET.....	69
7.7.1	<i>Taylorutveckling</i>	69
7.7.2	<i>Undersökning av stationär punkt</i>	72
7.7.2.1	Ett förvånansvärt vanligt fel	76
7.7.2.2	Ett annat vanligt fel	77
8	OPTIMERING.....	78
8.1	REPETITION AV ENVARIABELFALLET	78
8.1.1	<i>Optimering på kompakt mängd</i>	79
8.2	FLERVARIABELFALLET	81
8.2.1	<i>Optimering på kompakta mängder</i>	82
8.3	EXEMPEL MED OPTIMERING I PLANET	82
8.3.1	<i>Max och min på kurva i planet</i>	83
8.3.2	<i>Max och min på tvådimensionellt område i planet</i>	86
8.3.3	<i>Repetition av resonemanget</i>	91
8.3.4	<i>Max och min på kurva i planet (igen)</i>	92
8.4	EXEMPEL MED OPTIMERING I RUMMET	93
8.4.1	<i>Max och min på kurva i rummet</i>	93
8.4.2	<i>Max och min på yta i rummet</i>	94
8.4.3	<i>Max och min på tredimensionellt område i rummet</i>	96
8.4.4	<i>Max och min på yta i rummet (igen)</i>	102
9	ETT NYTT SÄTT ATT OPTIMERA – MED BIVILLKOR.....	104
9.1	GEOMETRISK FÖRKLARING	105
9.2	OM ATT UTTRYCKA PARALLELLITET OCH LINJÄRT BEROENDE	105
9.3	ENKLA EXEMPEL	106
9.4	EXEMPEL.....	107
9.4.1	<i>Max och min på kurva i planet</i>	107
9.4.2	<i>Max och min på tvådimensionellt område i planet</i>	109
9.4.3	<i>Max och min på yta i rummet</i>	109
9.4.4	<i>Max och min på tredimensionellt område i rummet</i>	111
9.5	FLER EXEMPEL.....	113
9.5.1	<i>Ett mycket vanligt fel</i>	115
10	OPTIMERING PÅ ICKE-KOMPAKT MÄNGD.....	120
10.1	EXEMPEL I PLANET	121
11	INTEGRALER.....	127

11.1	ENVARIABELINTEGRALER	127
11.1.1	<i>Så... Vad är en integral?</i>	136
11.1.2	<i>Ett par extra exempel</i>	137
11.2	DUBBELINTEGRALER	139
11.3	TRIPPELINTEGRALER	140
11.4	MULTIPELINTEGRALER AV HÖGRE DIMENSION.....	141
11.5	ENKLA EXEMPEL	141
11.6	INTRESSANTA EXEMPEL	150
11.6.1	<i>Volymberäkningar</i>	150
11.6.2	<i>Areaberäkningar</i>	154
11.6.3	<i>Medelvärdesberäkningar</i>	155
11.6.3.1	Geometriskt centrum	156
11.6.3.2	Masscentrum.....	158
11.6.3.3	Fler exempel med medelvärden.....	158
11.7	SAMMANFATTNING AV INTEGRALTOLKNINGAR.....	160
12	GENERALISERADE INTEGRALER	161

1 Inledning

Det här dokumentet sammanfattar flervariabelanalysens grunder. Fokus ligger på grundläggande begrepp, på vad de betyder och hur de hänger ihop. Erfarenhet från matematikundervisning på högskolenivå visar nämligen att även om många studenter memorerar de tekniska beräkningsmoment som krävs för att lösa analysuppgifter, så brister det ibland vad beträffar den grundläggande förståelsen – studenten vet egentligen inte vad han gör!

Stort fokus ligger på tydlighet och på exempel. Dessutom finns gott om bilder som förtydligar. Vissa delar av dokumentet är överkurs. I den mån ett helt avsnitt är av sådan karaktär är rubriken markerad med en asterisk (*); i annat fall framgår det av brödtexten att materialet är överkurs.

Varje avsnitt (som inte är asteriskmarkerat) i dokumentet är viktigt, och läsare med avsikt att förstå flervariabelanalysen bör inte påbörja nästa avsnitt innan han helt och fullt förstått det aktuella avsnittet. Dessa läsare bör också göra alla övningsuppgifter vilka inte är asteriskmarkerade eller ligger i asteriskmarkerat avsnitt.

Fokus ligger som sagt på grundläggande begrepp. Men även praktisk räkning samt diskussioner om *god redovisning* framhävs. Däremot är det här dokumentet inte en fullständig kurs i flervariabelanalys. Det som framför allt utelämnas är de moment som de allra flesta studenter behärskar och som sällan missförstås. Till detta hör t.ex. grundläggande tillämpning av kedjeregeln samt variabelbyten i multipelintegarler. Här ger vi förvisso mängder av *exempel*, men ingen fullständig genomgång.

För att begränsa oss avstår¹ vi också från att behandla partiella differentialekvationer samt inversa och implicita funktionssatserna. Dessa moment brukar behandlas tillräckligt bra i den ordinarie kurslitteraturen och det finns ganska lite utrymme för färgfulla illustrationer och genuint intressanta tillämpningar.

Inom flervariabelanalysen studerar man delmängder av \mathbb{R}^n och funktioner av typen $\mathbb{R}^n \rightarrow \mathbb{R}^m$. Vi börjar därför med en genomgång av dessa begrepp.

¹ Författaren kan komma att lägga till kapitel om detta i framtiden.

2 Delmängder av \mathbb{R}^n

Läsaren är naturligtvis bekant med rummet \mathbb{R}^n ; det är inget annat än mängden av alla n -tupler (x_1, x_2, \dots, x_n) . Speciellt brukar vi tolka $\mathbb{R}^1 = \mathbb{R}$ som *tallinjen*, \mathbb{R}^2 som *planet* och \mathbb{R}^3 som *rummet*.

2.1 Uppvärmning

De vanligast förekommande delmängderna av \mathbb{R} är förmodligen de som består av ändligt många punkter – t.ex. $\{0, 1, 5\}$ som är lösningsmängden till ekvationen $x^3 - 6x^2 + 5x = 0$ – och *intervallen* – t.ex. $[-2, 1]$ som är lösningsmängden till olikheten $x^2 + x - 2 \leq 0$. Andra exempel på delmängder är heltalen \mathbb{Z} och de rationella talen \mathbb{Q} . Det finns också betydligt mer invecklade delmängder, som den så kallade Cantormängden, som den intresserade läsaren med överkursambitioner kan slå upp på egen hand.

Bland planets delmängder har vi förstås också ändliga mängder av punkter, t.ex. $\{(1,2), (3,5), (0, 10)\}$. Men vi har också linjer, rektanglar, ellipser, hyperbler och alla möjliga sorters kurvor, plana områden och blandningar av både det ena och andra. Läsaren bör utan problem kunna rita/beskriva följande mängder på egen hand:

1. $\{(x, y) \in \mathbb{R}^2: x = 0\}$
2. $\{(x, y) \in \mathbb{R}^2: x + 3y = 0\}$
3. $\{(x, y) \in \mathbb{R}^2: y \geq x\}$
4. $\{(x, y) \in \mathbb{R}^2: 0 \leq x \leq 1, 0 \leq y \leq 1\}$
5. $\{(x, y) \in \mathbb{R}^2: -\frac{1}{2} \leq y \leq \frac{1}{2}\}$
6. $\{(x, y) \in \mathbb{R}^2: 0 < x - 2y < 1\}$
7. $\{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 1\}$
8. $\{(x, y) \in \mathbb{R}^2: (x - 1)^2 + y^2 = 1\}$
9. $\{(x, y) \in \mathbb{R}^2: x^2 + 2y^2 \leq 2\}$
10. $\{(x, y) \in \mathbb{R}^2: x^2 \leq y \leq x\}$
11. $\{(x, y) \in \mathbb{R}^2: |x| + |y| < 1\}$
12. $\{(x, y) \in \mathbb{R}^2: \max(|x|, |y|) = 1\}$

(Med $\max(a, b)$ avses det största av talen a och b . Max är alltså (i det här fallet) en funktion $\mathbb{R}^2 \rightarrow \mathbb{R}$ som tar in två tal och ger ifrån sig det största av dem. Till exempel är $\max(5, 1) = 5$.)

Den student som vill se exempel på mer invecklade plana mängder kan slå upp Sierpinski mattan eller någon annan *fraktal*.

De enklaste delmängderna av \mathbb{R}^3 är återigen de som består av ändligt många punkter. Sedan har vi kurvor (t.ex. en rät linje eller en spiral), ytor (t.ex. xy -planet eller enhetssfären) och tre-dimensionella områden (t.ex. klot och rätblock). Och naturligtvis finns det även mer invecklade exempel, som t.ex. Mengers tvättsvamp.

2.1.1 En väldigt trevlig notation

När det gäller delmängder av \mathbb{R} har vi som bekant infört en speciell notation för de ofta förekommande s.k. *intervallen*, t.ex. $[a, b] = \{x \in \mathbb{R}: a \leq x \leq b\}$ och $[a, b[= \{x \in \mathbb{R}: a \leq x < b\}$. Detta underlättar naturligtvis författande av texter något ofantligt. När det gäller delmängder av $\mathbb{R}^2, \mathbb{R}^3$, o.s.v. kan man använda den så kallade *kartesiska produkten* – en allmän mängdoperation – för att på ett mycket koncist sätt med hjälp av intervall uttrycka axelparallella *linjer*, *rektanglar*, *plan* och *rätblock*. Ifall läsaren inte är bekant med mängdlärans elementära operationer kommer här en kort introduktion till den kartesiska produkten.

Låt A och B vara allmänna mängder. Då är den kartesiska produkten

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

I ord: den kartesiska produkten av de två mängderna A och B är mängden av alla ordnade par (a, b) där a är ett element i A och b är ett element i B .

Exempel. Låt $D = \{\text{kanin, råtta, hund}\}$ och $K = \{\text{♀, ♂}\}$. Då är

$$D \times K = \{(\text{kanin, ♀}), (\text{kanin, ♂}), (\text{råtta, ♀}), (\text{råtta, ♂}), (\text{hund, ♀}), (\text{hund, ♂})\}.$$

Exempel. $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.

Den kartesiska produkten av de *tre* mängderna A, B, C betecknas $A \times B \times C$ och är mängden av alla ordnade tripletter (a, b, c) där $a \in A, b \in B$ och $c \in C$. Generaliseringen av den kartesiska produkten till n faktorer är uppenbar.

Exempel. $\underbrace{\mathbb{R} \times \dots \times \mathbb{R}}_{n \text{ faktorer}} = \mathbb{R}^n$.

Anledningen till att vi (i det här dokumentet) är intresserade av den kartesiska produkten är att den kartesiska produkten av intervall (och andra delmängder av \mathbb{R}) är axelparallella linjer, rektanglar, plan, räbblock o.s.v.

Exempel. $[0,2] \times [0,1]$ är per definition mängden $\{(x, y) \in \mathbb{R}^2 : x \in [0,2], y \in [0,1]\} = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 2, 0 \leq y \leq 1\}$, d.v.s. en fylld rektangel i planet. $\mathbb{R} \times [0,1]$ är per definition $\{(x, y) \in \mathbb{R}^2 : x \in \mathbb{R}, y \in [0,1]\}$ som är ett horisontellt "band". $[0,1] \times \{0\}$ är per definition mängden $\{(x, y) \in \mathbb{R}^2 : x \in [0,1], y = 0\}$ som är det räta linjestycket från $(0,0)$ till $(1,0)$ i planet.

Exempel. $[0,1] \times [0,1] \times [0,1]$ är enhetskuben i rummet. $\{0\} \times \{0\} \times \mathbb{R}$ är z-axeln i rummet. $\mathbb{R} \times \mathbb{R} \times \{5\}$ är planet $z = 5$ i rummet.

Övningar

1. Rita eller beskriv följande mängder:
 - a. $[0,1] \times [1,2]$
 - b. $\{1\} \times]-1,1[$
 - c. $[-1,1] \times \{-1,0,1\}$
 - d. $\mathbb{R} \times [0,1[$
 - e. $[0,1] \times [0,1] \times [0,2]$
 - f. $[0,1] \times [0,1] \times \{0,2\}$
 - g. $\mathbb{R} \times \mathbb{R} \times [0,1]$
 - h. $\{-1,1\} \times \{-1,1\} \times \{-1,1\}$.
2. *Redogör för skillnaden mellan den n -ställiga kartesiska produkten $A_1 \times \dots \times A_n$ och den upprepade binära kartesiska produkten $((A_1 \times A_2) \times A_3) \times \dots \times A_n$. Är den senare associativ?

2.2 Topologiska egenskaper

Givet en punkt $\mathbf{a} \in \mathbb{R}^n$ och ett tal $r > 0$ så definierar vi

$$B(\mathbf{a}, r) := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{a}\| < r\}.$$

Om läsaren nu mot all förmodan förträngt det, så är

$$\|\mathbf{x} - \mathbf{a}\| = \sqrt{(x_1 - a_1)^2 + \dots + (x_n - a_n)^2}$$

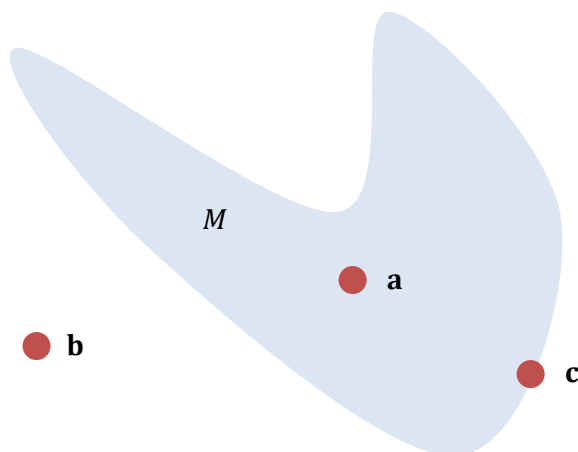
längden av vektorn $\mathbf{x} - \mathbf{a}$, d.v.s. *avståndet* mellan punkterna \mathbf{x} och \mathbf{a} . (Notera att högerledet reduceras till vanliga avståndsformeln i rummet, i planet och på tallinjen om $n = 3, 2$ respektive 1 .) $B(\mathbf{a}, r)$, som kallas för "det öppna klotet kring \mathbf{a} med radie r ", består alltså av alla punkter i \mathbb{R}^n som har avstånd mindre än r till \mathbf{a} . I rummet får vi ett vanligt klot, i planet får vi en disk och på tallinjen får vi ett intervall, i samtliga fall med punkten \mathbf{a} i mitten och med radien r .

En *omgivning* till en punkt $\mathbf{x} \in \mathbb{R}^n$ är en delmängd $U \subset \mathbb{R}^n$ som innehåller ett öppet klot kring \mathbf{x} .

Låt nu $M \subset \mathbb{R}^n$ vara en godtycklig delmängd; med M^c betecknar vi komplementet $\mathbb{R}^n \setminus M$. Om $\mathbf{x} \in \mathbb{R}^n$ är en punkt så gäller exakt ett av följande påståenden:

1. Det finns ett öppet klot kring \mathbf{x} vilket ligger helt i M .
2. Det finns ett öppet klot kring \mathbf{x} vilket ligger helt i M^c .
3. Varje öppet klot kring \mathbf{x} skär både M och M^c .

I första fallet säger vi att \mathbf{x} är en *inre* punkt till M ; i andra fallet säger vi att \mathbf{x} är en *yttre* punkt till M . I tredje fallet kallar vi \mathbf{x} för en *randpunkt* till M . I figuren nedan är \mathbf{a} en inre punkt till M medan \mathbf{b} är en yttre punkt och \mathbf{c} är en randpunkt.



Betrakta nu en fix mängd $M \subset \mathbb{R}^n$. Om alla randpunkter tillhör M så säger vi att M är *sluten*; om alla randpunkter tillhör M^c säger vi att M är *öppen*. Notera att en mängd kan vara *varken* öppen eller sluten, eller faktiskt *både* öppen och sluten. Alla fyra kombinationer är med andra ord möjliga (även om den sista är ovanlig i \mathbb{R}^n).

Exempel. Betrakta intervallet $M := [0, 1]$ som är en delmängd av den reella tallinjen. Varje punkt i $]0, 1[$ är en inre punkt till M eftersom det finns ett öppet klot (=ett intervall) kring punkten vilket helt ligger i M . Till exempel är 0.01 en inre punkt, eftersom det öppna klotet $]0.01 - 10^{-100}, 0.01 + 10^{-100}[$ ligger helt i M . På liknande sett ser vi att varje punkt i $]-\infty, 0[\cup]1, \infty[$ är en yttre punkt till M . Punkterna 0 och 1 är randpunkter till M , eftersom *varje* öppet klot till endera punkten nödvändigtvis skär både M och M^c . M är en sluten mängd, eftersom alla randpunkter (båda punkterna) tillhör mängden.

Rita följande exempelmängder, och var **mycket** noga med att du förstår exemplen.

Exempel. Mängden $[0,1] \times [0,2]$ är sluten medan $]0,2[\times]0,2[$ är öppen. Mängden $[0,1] \times]0,1[$ är varken öppen eller sluten. Det finns bara två delmängder i \mathbb{R}^n vilka är *både slutna och öppna*, nämligen \emptyset och \mathbb{R}^n .

Exempel. Disken $\{(x, y) \in \mathbb{R}^2: x^2 + y^2 < 1\}$ är öppen; disken $\{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq 1\}$ är sluten. I allmänhet är det "öppna klotet" $B(\mathbf{a}, r)$ öppet (som tur är!).

Exempel. Vi har sett att de inre punkterna till $[0,1] \subset \mathbb{R}$ är de i $]0,1[$ och att randpunkterna är 0 och 1. Betrakta nu den snarlika mängden $M := [0,1] \times \{0\} \subset \mathbb{R}^2$ i planet. Här finns *inga* inre punkter. Varför då? Jo, välj vilken punkt \mathbf{a} i mängden som helst och betrakta det öppna klotet $B(\mathbf{a}, r)$ kring punkten. Denna *disk* skär naturligtvis M^C , så punkten är inte en inre punkt. I stället är det en randpunkt.

Övningar

1. Visa att M är öppen omm varje punkt i M är en inre punkt, d.v.s. varje punkt i M har en omgivning som helt ligger i M .
2. Visa att M är sluten omm M^C är öppen.
3. Rita följande mängder och bestäm deras inre punkter, yttre punkter och randpunkter. Avgör om mängderna är öppna, slutna, ingetdera eller båda.
 - a. $[-1,1] \setminus \{0\}$
 - b. $[0,1] \cup]2,3[$
 - c. $\{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq 1, y > 0\}$
 - d. $\{(x, y) \in \mathbb{R}^2: 1 \leq x^2 + y^2 \leq 9, 0 \leq x \leq y\}$
 - e. $\{(x, y) \in \mathbb{R}^2: 0 \leq y \leq x^2\}$
 - f. $[0,1] \times \{0,1\}$
 - g. $\{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 < 1, z > 0\}$.

2.2.1 Kompakta mängder

En mängd $M \subset \mathbb{R}^2$ kallas för *begränsad* om det finns något tal $R > 0$ sådant att $M \subset B(\mathbf{0}, R)$, d.v.s. om M finns inuti *något* öppet klot kring origo. I princip betyder detta att mängden inte "sticker iväg mot oändligheten i någon riktning".

En mängd $M \subset \mathbb{R}^n$ kallas slutligen för *kompakt* omm den är sluten och begränsad. Egenskapen *kompakt* är mycket viktig, vilket vi kommer att se senare.

Exempel (*)

Betrakta familjen $(A_n)_{n=1}^\infty$ av delmängder av planet definierad av

$$A_n = \left\{ (x, y) \in \mathbb{R}^2: \frac{1}{n} \leq \sqrt{x^2 + y^2} \leq n \right\}.$$

(Rita några av mängderna. Notera att A_n är en *växande* följd av kompakta mängder; "växande" betyder att $A_n \subset A_{n+1}$ för varje $n \in \mathbb{Z}^+$.) Visa att det för varje kompakt mängd $K \subset \mathbb{R}^2 \setminus \{\mathbf{0}\}$ finns ett $N \in \mathbb{Z}^+$ sådant att $K \subset A_N$.

Lösning: Eftersom K är kompakt är K begränsad; det finns därför ett tal $R > 0$ sådant att $K \subset B(\mathbf{0}, R)$. Låt n_1 vara det minsta heltal som är större eller lika med R . Eftersom $\mathbf{0} \notin K$ kan $\mathbf{0}$ inte vara en inre punkt till K . Alltså är $\mathbf{0}$ antingen en randpunkt eller en yttre punkt. Men eftersom K är kompakt är K sluten, så den innehåller alla sina randpunkter. Eftersom $\mathbf{0} \notin K$ måste alltså $\mathbf{0}$ vara en yttre punkt. Per definition betyder detta att det finns ett tal $r > 0$ sådant att $B(\mathbf{0}, r) \subset K^c$. Låt n_2 vara det minsta positiva heltal sådant att $1/n_2 \leq r$.

Sätt $N := \max(n_1, n_2)$. Då är $K \subset B(\mathbf{0}, R) \subset B(\mathbf{0}, n_1) \subset B(\mathbf{0}, N)$ och $B(\mathbf{0}, 1/N) \subset B(\mathbf{0}, 1/n_2) \subset B(\mathbf{0}, r) \subset K^c$ (rita!) varför $K \subset A_N$, som önskat. ■

Övningar

1. Avgör vilka av mängderna i 3a–g i föregående avsnitt som är begränsade.
2. Avgör vilka av mängderna i 3a–g i föregående avsnitt som är kompakta.
3. *Inom den lite högre matematiken definieras begreppet *kompakt* på ett annat sätt. En mängd $M \subset \mathbb{R}^n$ säges vara kompakt omm varje öppen täckning har en ändlig deltäckning. En *täckning* av en mängd M är en familj av mängder i \mathbb{R}^n sådan att deras union är en supermängd till M (unionen "täcker" M). En *öppen täckning* är helt enkelt en täckning bestående av öppna mängder. Givet en täckning av M är en *deltäckning* en delmängd av familjen som också täcker M , och en *ändlig deltäckning* är en ändlig sådan delfamilj. Begreppet *kompakt* definieras på nästan precis det här sättet också mer allmänt för s.k. *topologiska rum* (speciellt för s.k. *metriska rum*). Man kan visa att dessa mer generella begrepp reduceras till "sluten och begränsad" i fallet \mathbb{R}^n (med den vanliga topologin respektive metriken).
 - a. Visa att \mathbb{R} inte är kompakt genom att betrakta den öppna täckningen $\{]k, k+2[: k \in \mathbb{Z}\}$. (Och \mathbb{R} är ju inte *begränsad*.)
 - b. Visa att $]0,1[$ inte är kompakt genom att betrakta den öppna täckningen $\{]1/k, 2[: k \in \mathbb{Z}^+\}$. (Och $]0,1[$ är ju inte *sluten*.)
 - c. Troliggör det faktum att $[0,1]$ är kompakt genom att betrakta några olika öppna täckningar – "försök och misslyckas" med att hitta en öppen täckning som *inte* har en ändlig deltäckning. (Och $[0,1]$ är ju sluten och begränsad.)
4. *En annan karaktärisering av kompakta delmängder av \mathbb{R}^n (eller, mer allmänt: metriska rum) är följande: en mängd är kompakt omm varje (oändlig) följd av element i mängden har en konvergent (oändlig) delföljd (där gränsvärdet tillhör mängden).
 - a. Visa att \mathbb{R} inte är kompakt genom att betrakta följderna $(x_k)_{k=1}^\infty \subset \mathbb{R}$ definierad av $x_k = k$.
 - b. Visa att $]0,1[$ inte är kompakt genom att betrakta följderna $(x_k)_{k=1}^\infty \subset]0,1[$ definierad av $x_k = 1/k$.
 - c. Troliggör det faktum att $[0,1]$ är kompakt genom att betrakta några olika följderna av element i $[0,1]$ – "försök och misslyckas" med att hitta en följd som *inte* har en konvergent delföljd. (Pröva $a_k = \frac{1}{k}$, $a_k = 1 - \frac{1}{k}$, $a_k = \frac{1}{2} [(-1)^k + 1]$ och några mer fantasifulla varianter.)

Anmärkning ():* I ett allmänt metriskt rum är begreppen "sluten" och "begränsad" fortfarande väldefinierade, men "kompakt" är där *inte* ekvivalent med "sluten och begränsad"; däremot är "kompakt" ekvivalent både med egenskapen i uppgift 3 och med egenskapen i uppgift 4. Begreppet är också

ekvivalent med "fullständig och totalt begränsad". I just det metriska rummet \mathbb{R}^n är alla dessa tre begrepp ekvivalenta med "sluten och begränsad".

2.2.2 Rand, inre och slutet hölje (*)

Om $M \subset \mathbb{R}^n$ så betecknar vi med ∂M mängden av alla randpunkter till M ; ∂M kallas för *randen* till M . Vi inför också $\bar{M} := M \cup \partial M$ som vi kallar M 's slutna hölje och $M^\circ := M \setminus \partial M$ som vi kallar för M 's inre. \bar{M} består alltså av M och alla M 's randpunkter medan M° består av alla inre punkter till M .

Exempel. Betrakta den öppna enhetsdisken $M := \{(x, y) \in \mathbb{R}^2: x^2 + y^2 < 1\}$. Då är $\partial M = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 1\}$ enhetscirkeln medan $\bar{M} = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq 1\}$ är den slutna enhetsdisken. Vi har också $M^\circ = M$.

Exempel. Betrakta den slutna enhetsdisken $M := \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq 1\}$. Då är $\partial M = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 1\}$ enhetscirkeln medan $M^\circ = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 < 1\}$ är den öppna enhetsdisken. Vi har också $\bar{M} = M$.

Exempel. Låt $M :=]0,1[\times]0,1[$. Då är ∂M rektangeln som består av de fyra "kanterna", $\bar{M} = [0,1] \times [0,1]$ och $M^\circ =]0,1[\times]0,1[$.

Exempel. Om $M :=]0,1[$ så är $\partial M = \{0,1\}$, $\bar{M} = [0,1]$ och $M^\circ =]0,1[$. Om $M :=]0,1[\times \{0\}$ så är $\partial M = [0,1] \times \{0\}$, $\bar{M} = [0,1] \times \{0\}$ och $M^\circ = \emptyset$.

Exempel. Betrakta den ändliga mängden $M := \{(0,0), (1,0), (1,1)\} \subset \mathbb{R}^2$. Vi har $\partial M = M$, $\bar{M} = M$ och $M^\circ = \emptyset$. M är alltså sluten och saknar inre punkter.

Om M är en sluten mängd är $\partial M \subset M$ varför $\bar{M} \stackrel{\text{def}}{=} M \cup \partial M = M$, vilket exemplifierades ovan. Omvänt, om $\bar{M} \stackrel{\text{def}}{=} M \cup \partial M = M$ följer det att $\partial M \subset M$ så att M är sluten. Vi har därför visat att

$$M \text{ är sluten} \iff \bar{M} = M.$$

Övningar

1. Visa att M är öppen omm $M^\circ = M$.
2. Visa att \bar{M} är sluten och att M° är öppen.
3. Visa att $\overline{M^\circ} = \bar{M}$ och att $(M^\circ)^\circ = M^\circ$. [Ledning: Använd de två föregående övningarna.]
4. Visa att $\partial M = \bar{M} \setminus M^\circ$. [Ledning: Rita ett Venndiagram över M och ∂M .]
5. Visa att om mängden M är ändlig så är $\bar{M} = M$, $M^\circ = \emptyset$ och $\partial M = M$.
6. Ge exempel på en mängd M sådan att $\overline{M^\circ} \neq \bar{M}$.
7. Visa att $\partial \mathbb{Z} = \mathbb{Z}$, $\bar{\mathbb{Z}} = \mathbb{Z}$ och $\mathbb{Z}^\circ = \emptyset$.
8. Visa att $\partial \mathbb{Q} = \mathbb{R}$, $\bar{\mathbb{Q}} = \mathbb{R}$ och $\mathbb{Q}^\circ = \emptyset$.
9. Visa att unionen av en samling öppna mängder är öppen.
10. Visa att snittet av *ändligt* många öppna mängder är öppet.
11. Visa att snittet av en samling slutna mängder är sluten. [Ledning: Använd föregående uppgift samt en av De Morgans lagar tillsammans med det faktum att en mängd är sluten omm dess komplement är öppet.]
12. Visa att unionen av *ändligt* många slutna mängder är sluten. [Samma ledning som i föregående uppgift.]

13. Visa att \bar{M} är snittet av alla slutna supermängder till M ; \bar{M} är alltså den *minsta* slutna supermängden till M , i den meningen att *varje* slutna supermängd till M innehåller \bar{M} .
14. Visa att M° är unionen av alla öppna delmängder till M ; M° är alltså den *största* öppna delmängden till M , i den meningen att *varje* öppen delmängd till M finns i M° .

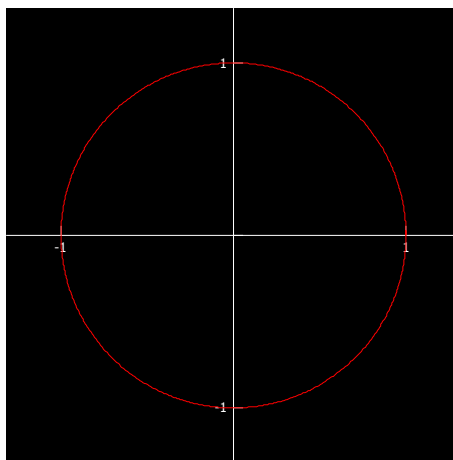
3 Kurvor och ytor

Man betraktar ofta kurvor i planet och rummet samt ytor i rummet. Kurvor och ytor är punktmängder, d.v.s. delmängder av \mathbb{R}^2 eller \mathbb{R}^3 (eller något annat rum). Kurvor är endimensionella "saker", som trådar; ytor är tvådimensionella "saker", som tygstycken.

3.1 Kurvor i planet

En kurva i planet kan ges av en *ekvation*. Till exempel säger vi att enhetscirkeln S^1 "är" $x^2 + y^2 = 1$. Vad som egentligen menas är då att enhetscirkeln är lika med *mängden* av punkter (x, y) i planet som uppfyller ekvationen, d.v.s.

$$S^1 = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 1\}.$$



Enhetscirkeln kan också parametreras, t.ex. med funktionen $\mathbf{r}: \mathbb{R} \rightarrow \mathbb{R}^2$ definierad av

$$\mathbf{r}(t) = \underline{\mathbf{e}} \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}.$$

Här menar vi att cirkeln är *värde*mängden till funktionen \mathbf{r} , d.v.s. mängden av punkter (x, y) som \mathbf{r} ger ifrån sig när t varierar. Vi har alltså $S^1 = V_{\mathbf{r}}$. Faktiskt räcker det med att t antar alla värden i $[0, 2\pi[$. S^1 är således mängden av punkter \mathbf{r} ger ifrån sig när vi stoppar in alla värden i $[0, 2\pi[$ i funktionen. Vi säger därför att S^1 är *bilden* av $[0, 2\pi[$ under parametreringsfunktionen \mathbf{r} och vi skriver detta $S^1 = \mathbf{r}([0, 2\pi[)$.

Ett alternativt sätt att parametrisera antingen övre eller undre halvan av enhetscirkeln är att använda x som parameter. Ekvationen $x^2 + y^2 = 1$ medför ju att $x = \pm\sqrt{1 - y^2}$, så t.ex. är bilden av intervallet $[-1, 1]$ under parametreringsfunktionen

$$\mathbf{r}(t) = \underline{\mathbf{e}} \begin{pmatrix} t \\ \sqrt{1 - t^2} \end{pmatrix}$$

inget annat än övre halvan (där $y \geq 0$) av enhetscirkeln.

Vad menas egentligen med att en kurva är "endimensionell"? Man *kan* ge en rigorös definition av dimensionsbegreppet av t.ex. en mångfald eller, mer allmänt, ett topologiskt rum. Men det tänker vi inte ge oss in i här! I stället nöjer vi oss med att konstatera följande:

Antag att du står på en punkt på enhetscirkeln (eller någon annan kurva). Då finns det *en* linjärt oberoende riktning du kan gå längs: framåt eller bakåt. Dessutom, för att ange en punkt på en kurva, räcker det med att ange *ett* tal (t.ex. avståndet från en fix punkt eller parametern t i en parametriseringsfunktion).

3.1.1 Fler kurvor

Vi beskriver i detta avsnitt alla plana kurvor man bör känna till (utöver enhetscirkeln).

3.1.1.1 Ellipser

Sedan tidigare matematikstudier vet läsaren att en cirkel med medelpunkt (x_0, y_0) och radie $r > 0$ har ekvationen

$$(x - x_0)^2 + (y - y_0)^2 = r^2.$$

Detta är uppenbart: vänsterledet i ekvationen är ju inget annat än kvadraten på avståndet mellan (x, y) och (x_0, y_0) . Denna cirkel är förstås $\mathbf{r}([0, 2\pi])$ där

$$\mathbf{r}(t) = \underline{\mathbf{e}} \begin{pmatrix} x_0 + r \cos t \\ y_0 + r \sin t \end{pmatrix}.$$

Mer allmänt är

$$\left(\frac{x - x_0}{a}\right)^2 + \left(\frac{y - y_0}{b}\right)^2 = 1$$

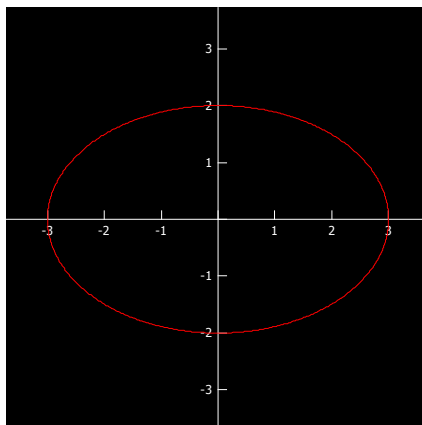
en ellips med medelpunkt i (x_0, y_0) och axlar parallella med koordinataxlarna och med halvaxellängderna a och $b > 0$. Denna är $\mathbf{r}([0, 2\pi])$ där

$$\mathbf{r}(t) = \underline{\mathbf{e}} \begin{pmatrix} x_0 + a \cos t \\ y_0 + b \sin t \end{pmatrix}$$

(pröva att stoppa in $\mathbf{r}(t)$ i ellipsens ekvation). Nedan visas ellipsen

$$\left(\frac{x}{3}\right)^2 + \left(\frac{y}{2}\right)^2 = 1$$

som är bilden $\mathbf{r}([0, 2\pi])$ under $\mathbf{r}(t) = (3 \cos t, 2 \sin t)$.



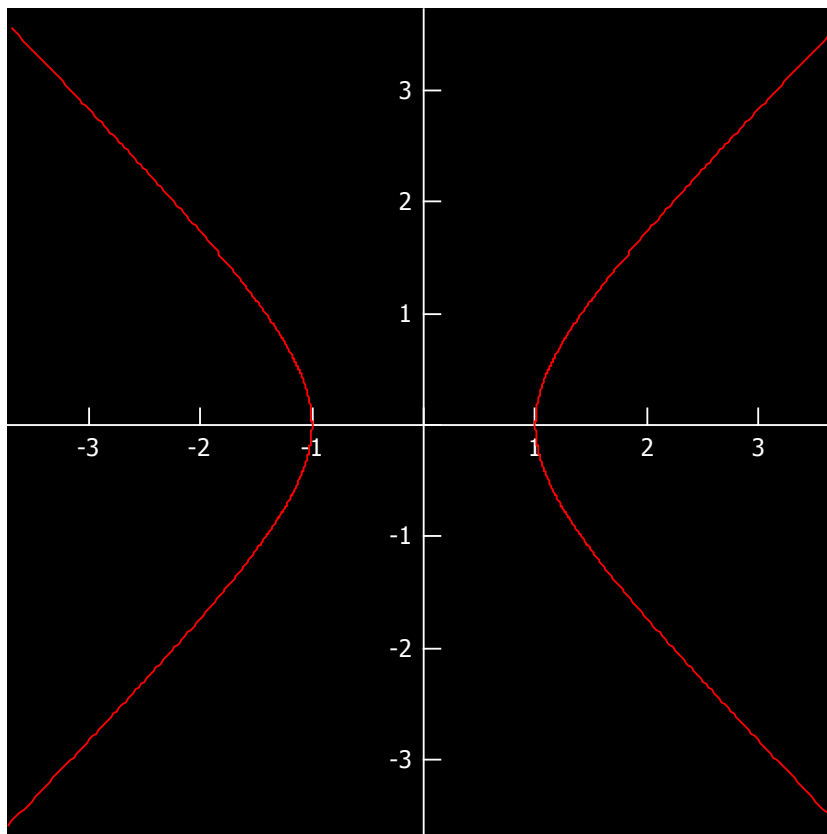
Planetbanor är ellipser. Till exempel rör sig jorden i en elliptisk bana kring solen, med solen i ena s.k. *brännpunkten* hos ellipsen.

3.1.1.2 Råta linjer

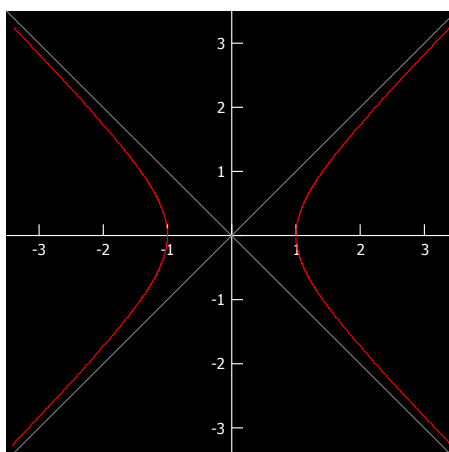
Innan vi fortsätter ge fler exempel på "intressanta" plana kurvor så påminner vi här om den råta linjen (som förstås också är en plan kurva). Vi har horisontella linjer ($y = c$) och vertikala linjer ($x = c$). Mer allmänt kan varje linje som inte är vertikal skrivas på formen $y = kx + m$. En helt allmän linje har ekvationen $ax + by = d$ och en normalvektor är (a, b) . Linjer kan trivialt parametreras. T.ex. är linjen $y = kx + m$ lika med $\mathbf{r}(\mathbb{R})$ där $\mathbf{r}(t) = (t, kt + m)$.

3.1.1.3 Hyperbler

Enhetscirkeln är $x^2 + y^2 = 1$, men vad är då $x^2 - y^2 = 1$? Den här kurvan kallas för en *hyperbel*.



Långt från origo är ettan i högerledet försumbar, så vi har $x^2 - y^2 \approx 0$, så $x \approx \pm y$. Linjerna $y = \pm x$ är asymptoter till kurvan, som för övrigt är den bana som en *komet* färdas längs när den slungas runt solen.



Den *högra* delen av hyperbeln $x^2 - y^2 = 1$ är bilden $\mathbf{r}(\mathbb{R})$ där $\mathbf{r}(t) = (\cosh t, \sinh t)$; den *vänstra* delen är $\mathbf{r}(\mathbb{R})$ där $\mathbf{r}(t) = (-\cosh t, \sinh t)$. (Pröva att stoppa in $\mathbf{r}(t)$ i ekvationen, och använd hyperboliska ettan.)²

Den mer allmänna hyperbeln $\left(\frac{x-x_0}{a}\right)^2 - \left(\frac{y-y_0}{b}\right)^2 = 1$ har förstås högra delen $\mathbf{r}(\mathbb{R})$ där

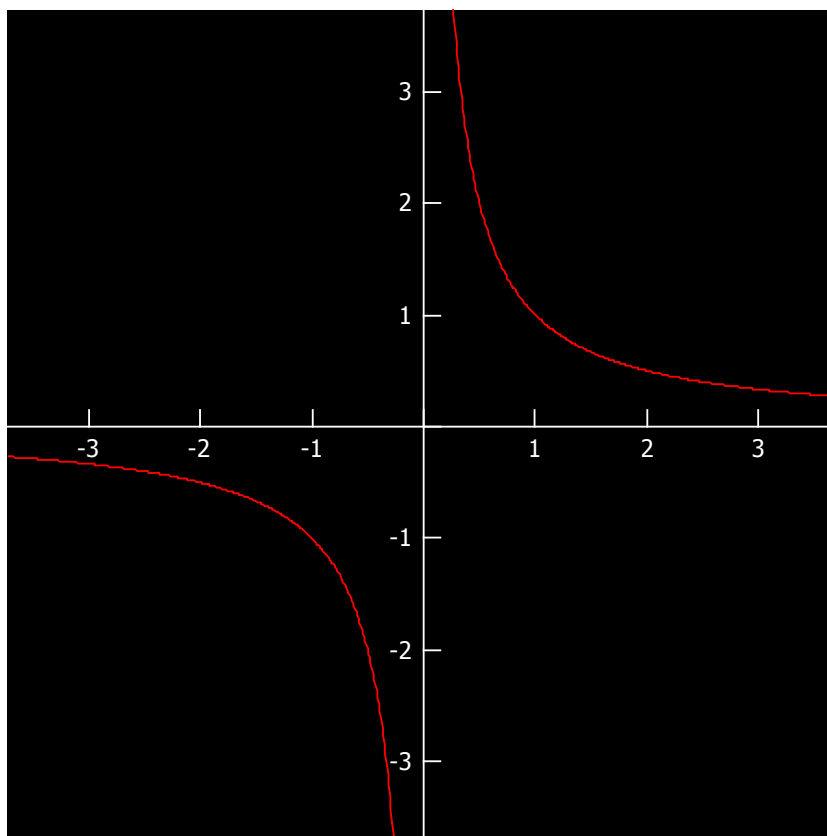
$$\mathbf{r}(t) = \underline{\mathbf{e}} \begin{pmatrix} x_0 + a \cosh t \\ y_0 + b \sinh t \end{pmatrix}.$$

Övning

1. *Vad har konstanterna a och b för geometrisk betydelse? Vad är asymptoterna till den mer allmänna hyperbeln?

3.1.1.4 Hyperbler, igen! (*)

Kurvan $xy = 1$ (d.v.s. $y = \frac{1}{x}$) har följande utseende:



Läsaren kan övertyga sig om att detta är en standardhyperbel vriden 45° moturs.

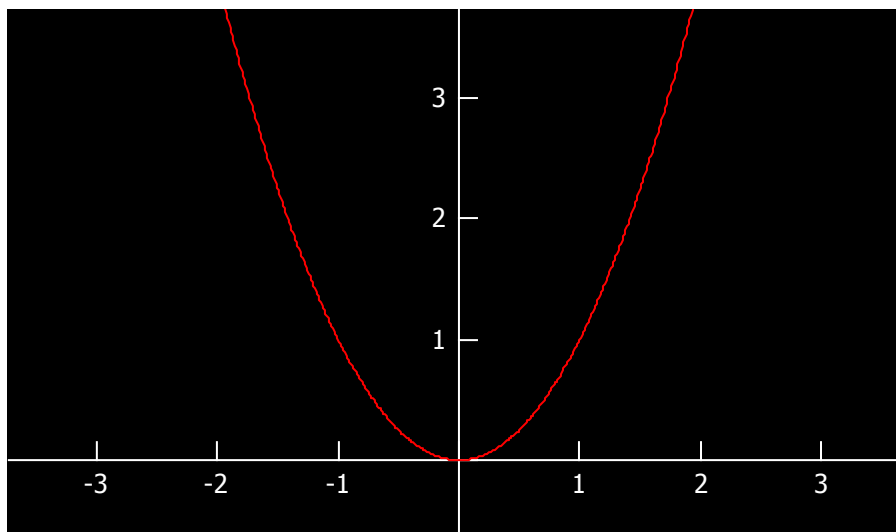
Övning

1. *Visa att $xy = 1$ är en standardhyperbel fast vriden genom att införa den nya basen $\underline{\mathbf{f}} = (\mathbf{f}_1 \quad \mathbf{f}_2)$ där $\mathbf{f}_1 = \frac{1}{\sqrt{2}}(1,1)$ och $\mathbf{f}_2 = \frac{1}{\sqrt{2}}(-1,1)$.

² Om läsaren inte är bekant med de hyperboliska funktionerna kan han med fördel konsultera <http://rejbrand.se/rejbrand/dokument/hyperboliska.pdf>.

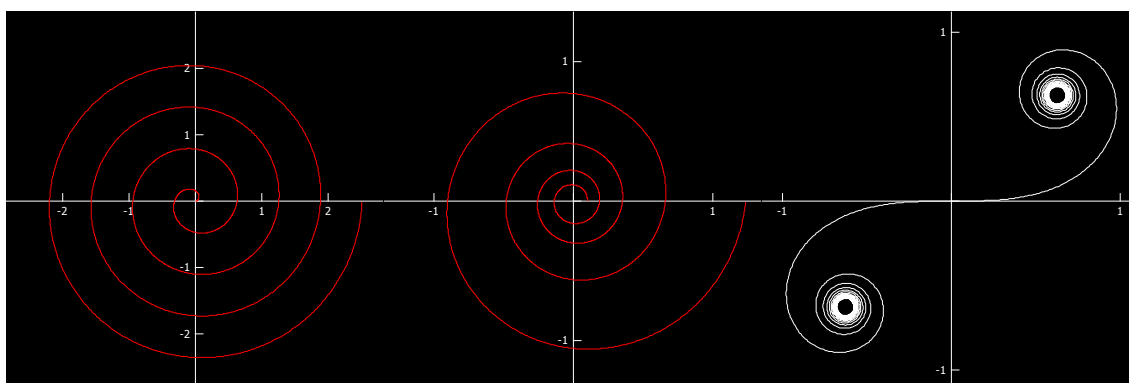
3.1.1.5 Parabler

Sedan gymnasiet vet läsaren att $y = x^2$ är en parabel. Den är bilden $\mathbf{r}(\mathbb{R})$ där $\mathbf{r}(t) = (t, t^2)$.



3.1.1.6 Spiraler och annat (*)

För skojs skull avslutar vi med några kurvor som inte har enkla ekvationer, men som är enkla att parametrisera. Först har vi den *Arkimediska spiralen* som är bilden $\mathbf{r}(\mathbb{R}^+)$ där $\mathbf{r}(t) = (t \cos t, t \sin t)$. Nedan, till vänster, visas $\mathbf{r}([0, 8\pi])$ där $\mathbf{r}(t) = \frac{1}{10}(t \cos t, t \sin t)$. Den *logaritmiska spiralen* är bilden $\mathbf{r}(\mathbb{R}^+)$ där $\mathbf{r}(t) = ae^{bt}(\cos t, \sin t)$; nedan, i mitten, visas $\mathbf{r}([0, 8\pi])$ för $a = b = 1/10$.

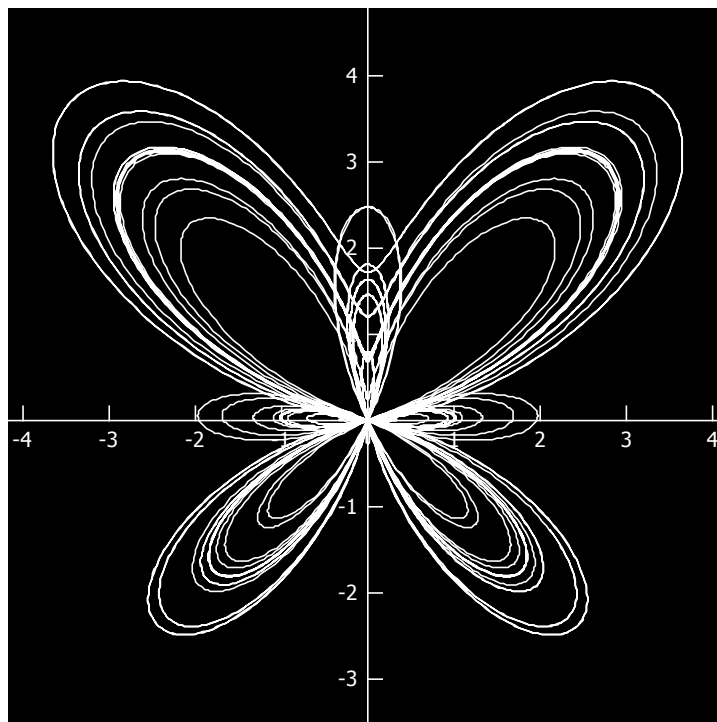


En intressantare spiral är *Eulerspiralen* som är bilden $\mathbf{r}(\mathbb{R})$ där $\mathbf{r}(t) = \left(\sqrt{\frac{2}{a}} C\left(\sqrt{\frac{a}{2}}t\right), \sqrt{\frac{2}{a}} S\left(\sqrt{\frac{a}{2}}t\right) \right)$ där de s.k. *Fresnelintegralerna* $C(x) := \int_0^x \cos t^2 dt$ och $S(x) := \int_0^x \sin t^2 dt$. Ovan, till höger, visas delen $\mathbf{r}([-10, 10])$ då $a = 2$. Den mest intressanta egenskapen hos Eulerspiralen är att dess s.k. *krökning* är proportionell mot kurvans längd från symmetripunkten, vilket gör att kurvan används för att förbinda raksträckor (med cirklar eller andra raksträckor) i järnvägar.

Till sist bjuder vi på följande kurva, som är bilden $\mathbf{r}([0, 100])$ under

$$\mathbf{r}(t) = \left(e^{\sin t} - 2 \cos 4t + \sin^5 \left(\frac{2t - \pi}{24} \right) \right) \underline{\mathbf{e}} \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

och går under namnet *fjärilskurvan*.



3.2 Kurvor i rummet

En kurva i rummet kan förstås inte ges av *en* ekvation, eftersom *en* ekvation i rummets koordinater ger en *yta*. Däremot kan en kurva i rummet parametreras av en funktion $\mathbb{R} \rightarrow \mathbb{R}^3$.

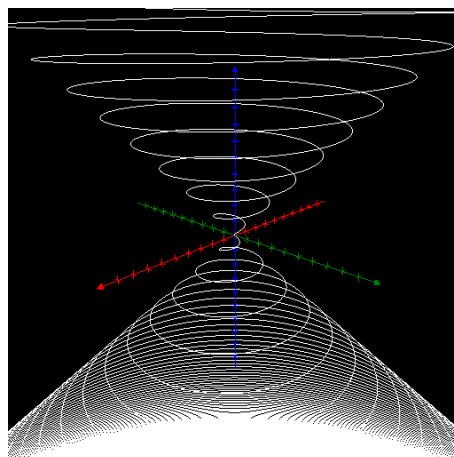
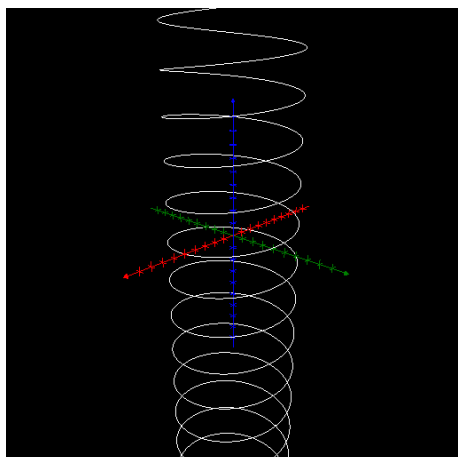
Den enklaste kurvan i rummet är den räta linjen som går genom punkten $\mathbf{r}_0 \in \mathbb{R}^3$ och har riktningsvektorn $\mathbf{v} \in \mathbb{R}^3$. Denna är bilden $\mathbf{r}(\mathbb{R})$ under parametreringsfunktionen $\mathbf{r}: \mathbb{R} \rightarrow \mathbb{R}^3$ definierad av

$$\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{v}.$$

Sedan är det förstås så att varje *plan* kurva kan "bäddas in" i något plan i rummet. Till exempel är enhetscirkeln i planet $z = 5$ bilden $\mathbf{r}([0, 2\pi])$ där $\mathbf{r}(t) = (\cos t, \sin t, 5)$.

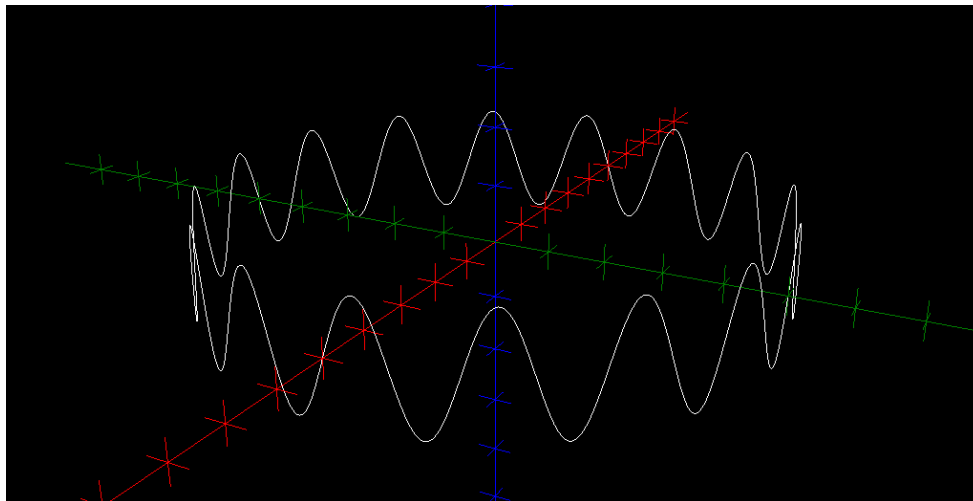
Något mer intressanta kurvor är den cirkulära helixen $\mathbf{r}(\mathbb{R})$ och den koniska helixen $\mathbf{p}(\mathbb{R})$ där

$$\mathbf{r}(t) = \underline{\mathbf{e}} \begin{pmatrix} \cos t \\ \sin t \\ t \end{pmatrix}, \quad \mathbf{p}(t) = \underline{\mathbf{e}} \begin{pmatrix} t \cos t \\ t \sin t \\ t \end{pmatrix}.$$



Som sista exempel ger vi "sinuskronan" $\mathbf{r}([0,2\pi])$ där

$$\mathbf{r}(t) = \underline{e} \begin{pmatrix} 5 \cos t \\ 5 \sin t \\ \sin 16t \end{pmatrix}.$$



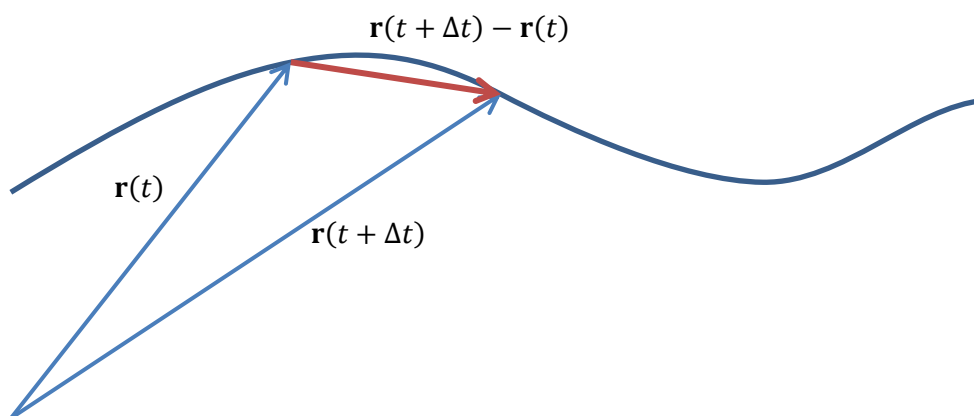
Övningar

1. *Konstanterna 5 och 16 är i princip godtyckliga och har valts för att ge en kurva med estetiskt tilltalande proportioner. Förklara vad konstanterna har för visuell effekt! Det är "viktigt" att 16 är ett heltal. Förklara varför! Vad händer om konstanten byts ut mot $k + 0.5$ för något heltal k ?
2. *Använd din favoritmjukvara (t.ex. www.algosim.se) för att experimentera med kurvor i rummet.

3.3 Mer om kurvor

Låt $\gamma = \mathbf{r}(I)$ vara en parametriserad kurva, där $I \subset \mathbb{R}$. Bilden av funktionen $\mathbf{r}: I \rightarrow \mathbb{R}^3$ kan tolkas som positionen för en partikel som flyttar sig i rummet. Till exempel kan vi tolka $\mathbf{r}(t)$ som en flygas position vid tiden $t \in I$.

Välj nu någon fix tidpunkt $t \in I$ och låt Δt vara en litet tal. Flugan är vid $\mathbf{r}(t)$ vid tiden t och vid $\mathbf{r}(t + \Delta t)$ vid tiden $t + \Delta t$, se bild.



Förflyttningen mellan dessa tidpunkter är $\mathbf{r}(t + \Delta t) - \mathbf{r}(t)$ så medelhastigheten hos flugan i intervallet $[t, t + \Delta t]$ är tydligen

$$\frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t}.$$

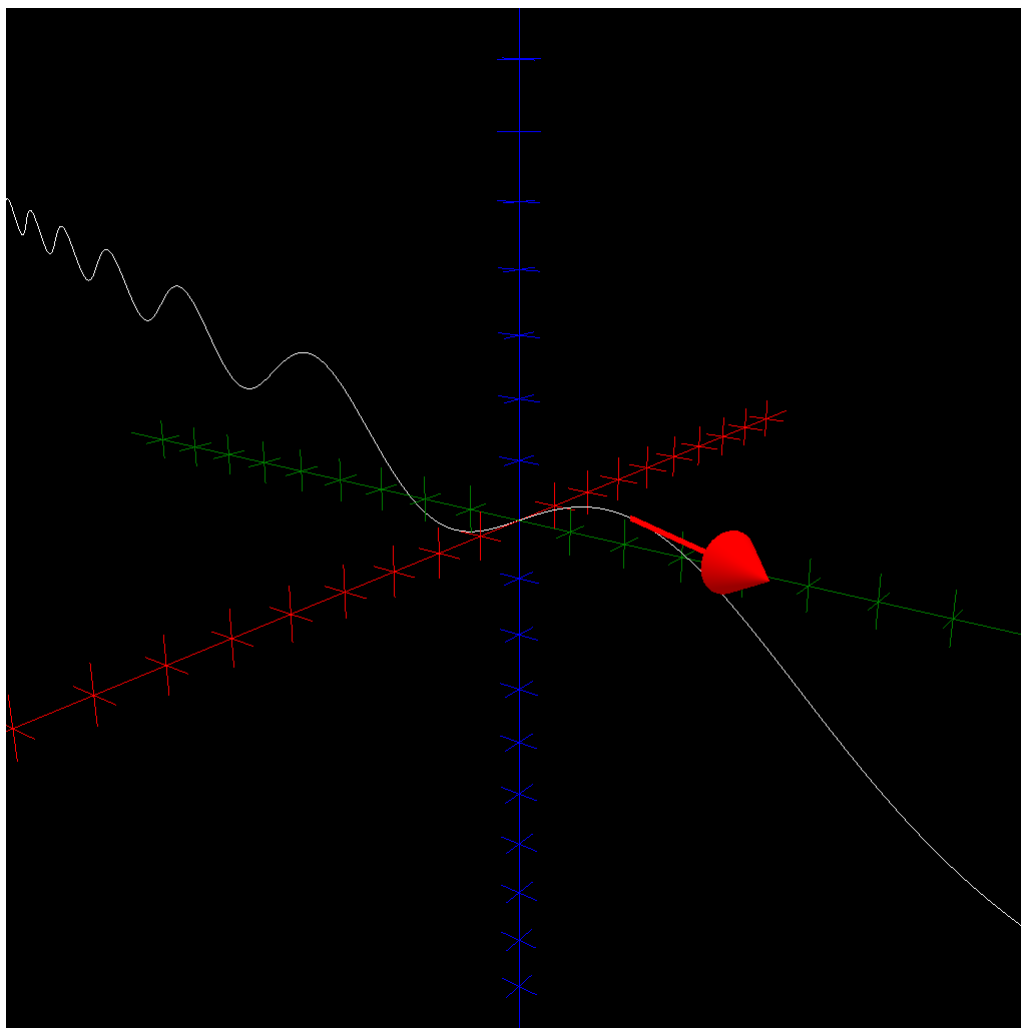
Det är naturligt att definiera flugans (momentan)hastighet vid tiden t som

$$\mathbf{v}(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t}.$$

I allmänhet definieras *derivatan* av en funktion $\mathbb{R} \rightarrow \mathbb{R}^n$ som detta gränsvärde. Detta är inte en "vanlig" derivata som läsaren är bekant med sedan envariabelanalysen (då vi definierar derivatan av en funktion av typen $\mathbb{R} \rightarrow \mathbb{R}$), men till *formen* är högerledet detsamma.

I praktiken använder man inte *definitionen* av den vektorvärda derivatan vid beräkningar. I stället deriverar man *komponentvis*. Det är nämligen så, att om $\mathbf{r}(t) = (x(t), y(t), z(t))$ i tre variabler (säg), så är $\mathbf{r}'(t) = (x'(t), y'(t), z'(t))$.

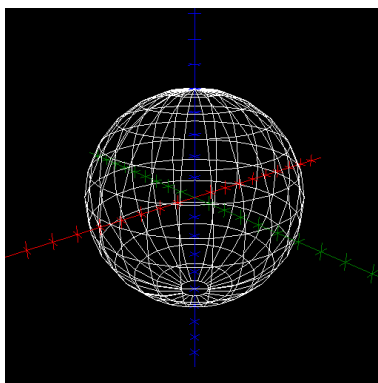
Exempel. En elektrons läge vid tiden t är $(t, 2t, \sin t)$. Då är partikelns hastighet vid tiden t lika med $(1, 2, \cos t)$. [Det kan också röra sig om ett flygplan som av okänd anledning oscillerar i höjded.]



3.4 Ytor i rummet

En yta i rummet ges ofta av en *ekvation*. Man säger t.ex. att $x^2 + y^2 + z^2 = 1$ "är" enhetssfären. Vad man egentligen menar då är att enhetssfären S^2 är mängden av punkter (x, y, z) i rummet vilka uppfyller ekvationen, d.v.s.

$$S^2 = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 = 1\}.$$



Enhetssfären kan också parametreras, t.ex. med funktionen $\mathbf{r}: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ definierad av

$$\mathbf{r}(\theta, \varphi) = \underline{\mathbf{e}} \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}.$$

Det betyder att enhetssfären är *värdemängden* till funktionen \mathbf{r} , d.v.s. mängden av punkter som ger ifrån sig när θ och φ varierar. Vi har alltså $S^2 = V_{\mathbf{r}}$. Faktiskt är enhetssfären bilden av rektangeln $[0, \pi] \times [0, 2\pi[$ i $\theta\varphi$ -planet, så $S^2 = \mathbf{r}([0, \pi] \times [0, 2\pi[)$.

Övre halvan ($z \geq 0$) av enhetssfären är också bilden $\mathbf{r}(\tilde{D})$ av den plana enhetsdisken $\tilde{D} := \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq 1\}$ under parametreringsfunktionen

$$\mathbf{r}(x, y) = \underline{\mathbf{e}} \begin{pmatrix} x \\ y \\ \sqrt{1 - x^2 - y^2} \end{pmatrix}.$$

En yta är *tvådimensionell*. Detta syns bland annat genom att man i varje punkt på en yta har *två* linjärt oberoende riktningar att gå längs: höger/vänster och uppåt/nedåt, säg. Det syns också på att det krävs *två* tal för att ange en position på en yta (t.ex. latitud och longitud på jordytan, eller parametrarna θ och φ i parametreringsfunktionen ovan).

3.4.1 Fler ytor

Vi beskriver här de enkla ytor man måste känna till.

3.4.1.1 Ellipsoider

Vi har betraktat enhetssfären. Mer allmänt är

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 + \left(\frac{z}{c}\right)^2 = 1$$

en *ellipsoid* kring origo med axelparallella axlar och halvaxellängderna a , b respektive c i x -, y - och z -led. Denna är bilden $\mathbf{r}([0, \pi] \times [0, 2\pi[)$ under

$$\mathbf{r}(\theta, \varphi) = \underline{\mathbf{e}} \begin{pmatrix} a \sin \theta \cos \varphi \\ b \sin \theta \sin \varphi \\ c \cos \theta \end{pmatrix}.$$

Ännu mer allmänt har vi samma ellipsoid fast centrerad kring (x_0, y_0, z_0) som uppfyller

$$\left(\frac{x-x_0}{a}\right)^2 + \left(\frac{y-y_0}{b}\right)^2 + \left(\frac{z-z_0}{c}\right)^2 = 1$$

och uppenbarligen är bilden $\mathbf{r}([0, \pi] \times [0, 2\pi])$ under

$$\mathbf{r}(\theta, \varphi) = \underline{\mathbf{e}} \begin{pmatrix} x_0 + a \sin \theta \cos \varphi \\ y_0 + b \sin \theta \sin \varphi \\ z_0 + c \cos \theta \end{pmatrix}.$$

Övning

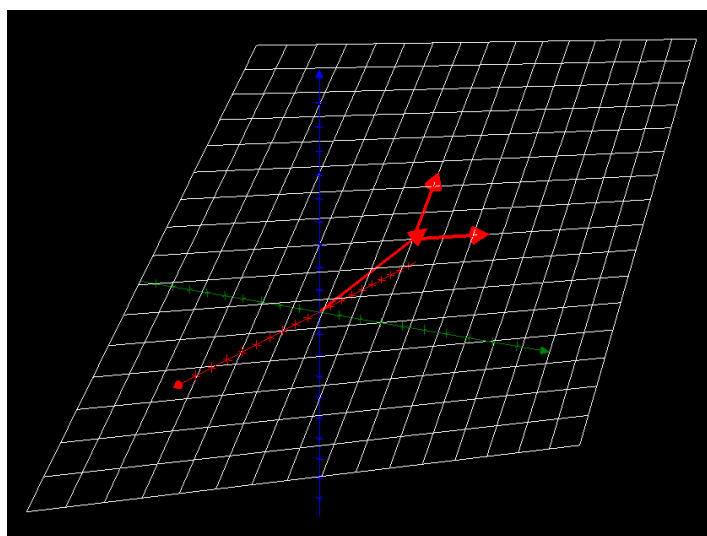
1. Ange ekvation och parametrisering för en sfär av radie r kring
 - a. origo
 - b. punkten (x_0, y_0, z_0) .

3.4.1.2 Plan

Innan vi ger fler exempel på "intressanta" ytor påminner vi läsaren om hans gamla kunskaper i linjär algebra. Ett allmänt *plan* i rummet har ekvationen $ax + by + cz = d$. En normalvektor till planet är (a, b, c) och planet innehåller origo om $d = 0$.

Ett plan som *inte* är vertikalt (d.v.s. som inte är parallellt med z -axeln) har $c \neq 0$ och dess ekvation kan skrivas $z = \frac{d}{c} - \frac{a}{c}x - \frac{b}{c}y$ och är uppenbarligen bilden $\mathbf{r}(\mathbb{R}^2)$ under $\mathbf{r}(x, y) = \left(x, y, \frac{d}{c} - \frac{a}{c}x - \frac{b}{c}y\right)$. (Om $c = 0$ kan man förstås parametrisera planet genom att lösa ut *en annan* av koordinaterna.)

Planet Π som går igenom punkten $\mathbf{p} \in \mathbb{R}^3$ och spänns upp av vektorerna $\mathbf{u} \in \mathbb{R}^3$ och $\mathbf{v} \in \mathbb{R}^3$ är bilden $\mathbf{r}(\mathbb{R}^2)$ under $\mathbf{r}(s, t) = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$.



Speciellt, om $\mathbf{p} = \mathbf{0}$, så är planet inget annat än det *linjära höljet* av vektorerna \mathbf{u} och \mathbf{v} (och planet är ett linjärt underrum till \mathbb{R}^3): $\Pi = [\mathbf{u}, \mathbf{v}]$.

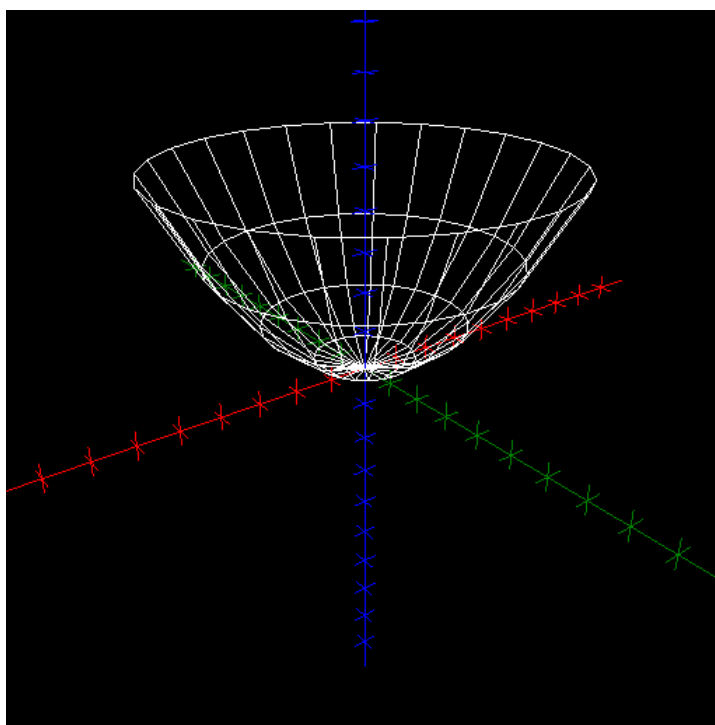
3.4.1.3 Paraboloider

Ytan $z = x^2 + y^2$ är en *paraboloid*.

Det är faktiskt mycket lätt att inse hur andragsytor som den här ser ut, genom att studera *plana snitt* av ytan. **Detta är mycket viktigt!** Först inser vi att om (x, y, z) tillhör ytan, så måste $z \geq 0$. Och den enda punkten i planet $z = 0$ som tillhör ytan är origo. Nu, till det intressanta: om vi snittar ytan med det horisontella planet $z = c$ för något $c > 0$ så får vi den plana kurvan $x^2 + y^2 = c$ som är en cirkel. *Varje horisontellt snitt av ytan är alltså en cirkel*. Frågan är hur denna cirkels radie varierar med z .

Om vi sätter $y = 0$, d.v.s. om vi tittar på snittet med xz -planet, så får vi den plana kurvan $z = x^2$ som är en *parabel*. (Samma resultat får vi så klart om vi sätter $x = 0$.) Paraboloiden erhålles sålunda om man roterar denna parabel kring z -axeln. Detta förklarar också ytans namn: roterar man en *parabel* kring z -axeln får man en *paraboloid*.

Nedan visas den del av paraboloiden som ligger under planet $z = 1$.



En paraboloid ser ut som en skål, fast det är en "jobbig" sorts skål, eftersom den förmodligen är väldigt vinglig om man försöker placera den på en plan yta, som ett bords ovansida.

Paraboloiden är bilden $\mathbf{r}(\mathbb{R} \times [0, 2\pi])$ under

$$\mathbf{r}(u, v) = \underline{\mathbf{e}} \begin{pmatrix} u \cos v \\ u \sin v \\ u^2 \end{pmatrix}.$$

Övning

1. Stoppa in $\mathbf{r}(u, v)$ i paraboloidens ekvation för att visa att $\mathbf{r}(\mathbb{R} \times [0, 2\pi]) \subset$ paraboloiden. Förklara sedan varför likhet råder.

Ett alternativt sätt att parametrisera paraboloiden är att använda x och y som parametrar. Det är ju klart att paraboloiden är bilden $\mathbf{p}(\mathbb{R}^2)$ där

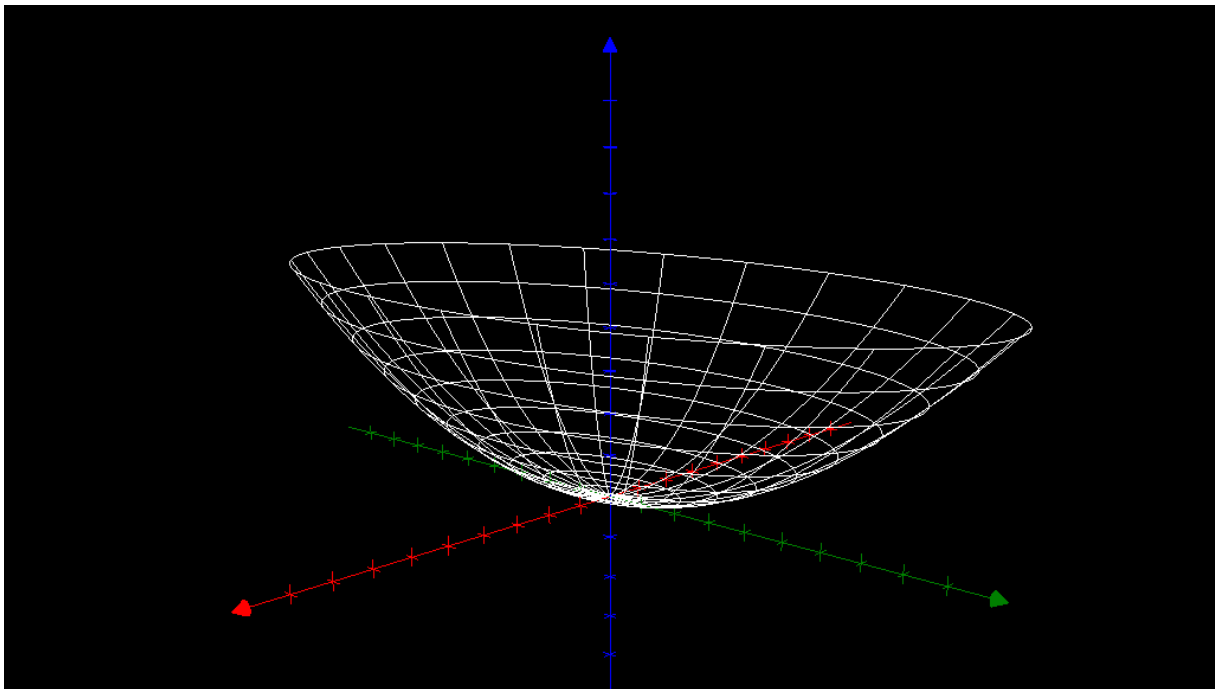
$$\mathbf{p}(x, y) = \underline{\mathbf{e}} \begin{pmatrix} x \\ y \\ x^2 + y^2 \end{pmatrix}.$$

En stor fördel med den första parametriseringen (d.v.s. \mathbf{r}) är att den är bättre anpassad till paraboloidens cirkulära symmetri. u är ju z -koordinaten medan v är "vinkeln" runt z -axeln; dessa är väldigt *naturliga* parametrar på ytan.

Om man t.ex. är intresserad av den del av paraboloiden som ligger under planet $z = 1$, som i bilden ovan, så är ju denna bilden $\mathbf{r}([0,1[\times [0,2\pi[)$ av en *rektangel* $[0,1[\times [0,2\pi[$ i det s.k. parameterplanet. Å andra sidan är den bilden $\mathbf{p}(\tilde{D})$ av en *disk* $\tilde{D} = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 < 1\}$. Se bilden på nästa uppslag.

Den "vanliga" paraboloiden kallas också för en *cirkulär paraboloid* eftersom snitten $z = c \neq 0$ ger cirklar. Lite mer allmän är alltså den *elliptiska paraboloiden*

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = z.$$

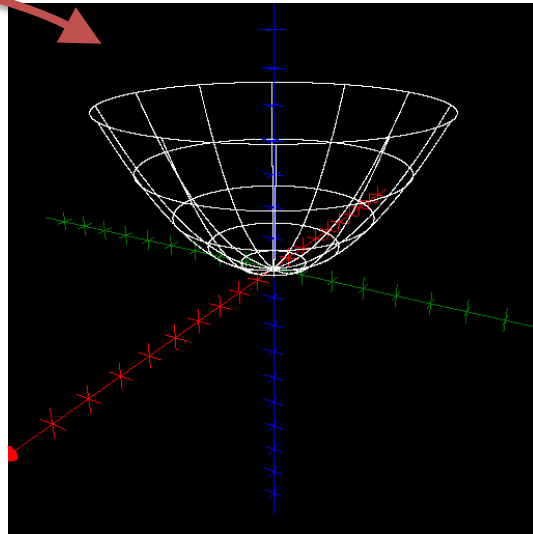
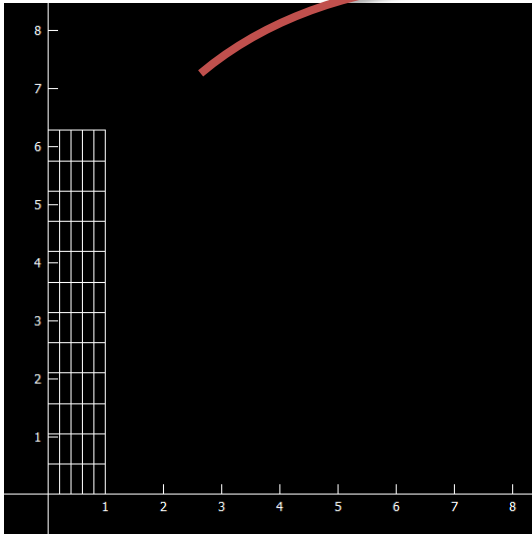


Det finns också en helt annan sorts paraboloid, men den återkommer vi till senare.

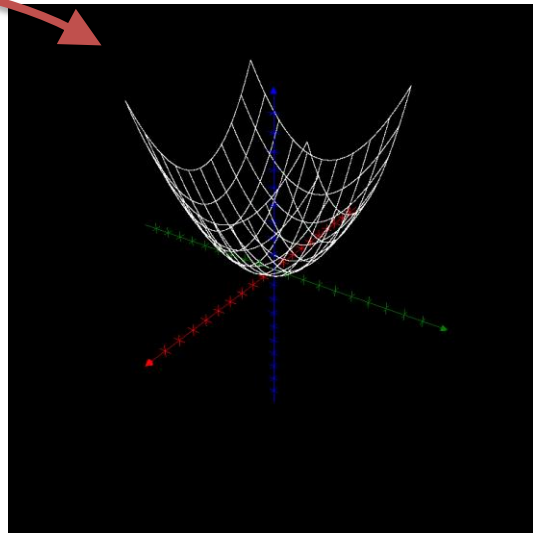
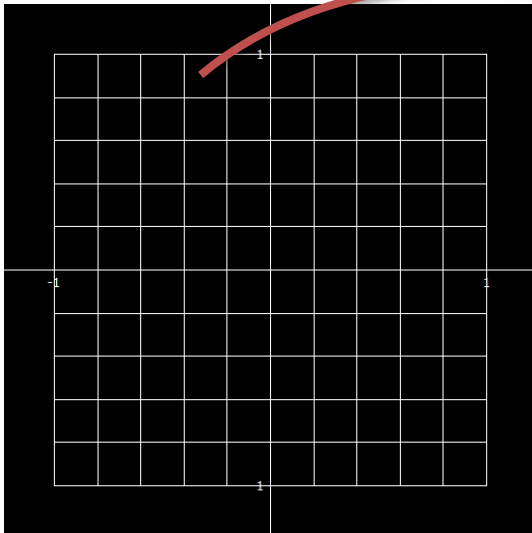
Övning

1. Parametrisera den allmänna elliptiska paraboloiden (på "båda sätten").

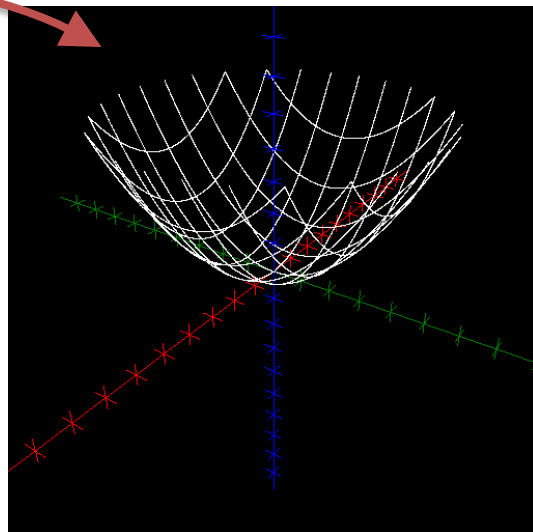
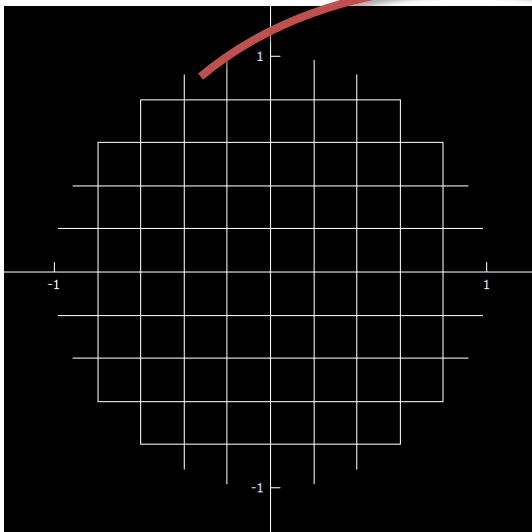
r



p

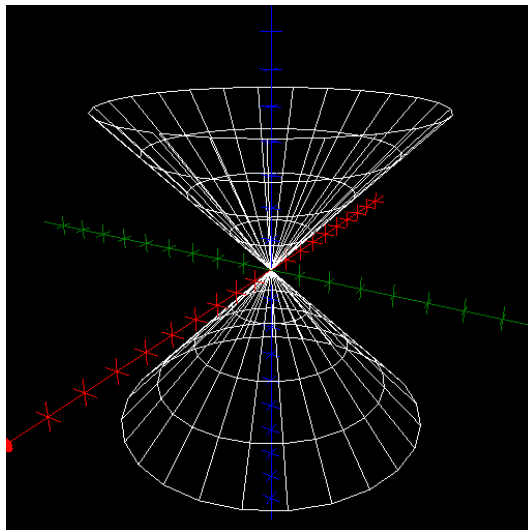


p



3.4.1.4 Koner

Ytan $x^2 + y^2 = z^2$ är en *kon*. Återigen ser vi att en skärning $z = c \neq 0$ ger en *cirkel* $x^2 + y^2 = c^2$ medan vi i t.ex. xz -planet $y = 0$ erhåller $x^2 = z^2$ vilket är ekvivalent med $z = \pm x$ som är två räta linjer. Nedan visas delmängden där $z \in [-1, 1]$.



Konen är bilden $\mathbf{r}(\mathbb{R} \times [0, 2\pi])$ där

$$\mathbf{r}(u, v) = \underline{\mathbf{e}} \begin{pmatrix} u \cos v \\ u \sin v \\ u \end{pmatrix}.$$

Man kan förstås också parametrisera den med x och y som parametrar, även om man då får ta överdelen ($z \geq 0$) och underdelen ($z \leq 0$) var för sig.

3.4.1.5 Cylindrar

Om $\Gamma \subset \mathbb{R}^2$ är en plan kurva så kallas $\Gamma \times \mathbb{R} \subset \mathbb{R}^3$ för en (rät) *cylinder*; studenter som tagit den kartesiska produkten till hjärtat har nu redan insett att $\Gamma \times \mathbb{R}$ är den yta som erhålles när kurvan Γ i xy -planet "dras ut" längs hela z -axeln. Om Γ är en ellips (cirkel) får vi en elliptisk (cirkulär) cylinder, om Γ är en parabel får vi en parabolisk cylinder, om Γ är en hyperbel får vi en hyperbolisk cylinder.

Den kanske mest välbekanta cylindern är den cirkulära. Varianten med radien $r > 0$ har ekvationen

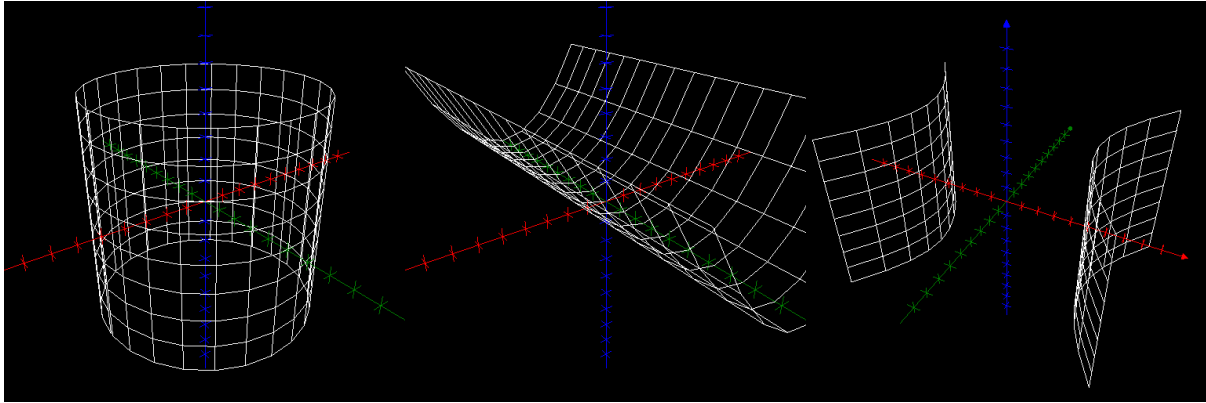
$$x^2 + y^2 = r^2.$$

Varje snitt $z = c$ ger $x^2 + y^2 = r^2$ som är en cirkel med radien r , som tydligen inte beror på z . (Och visst, t.ex. $y = 0$ ger ju $x = \pm r$. Alternativt kan man övertyga sig om att (x, y, z_1) uppfyller ekvationen omm (x, y, z_2) gör det. Vi får alltså samma kurva vid varje z -snitt.)

Cylindern är bilden $\mathbf{r}(\mathbb{R} \times [0, 2\pi])$ där

$$\mathbf{r}(z, \varphi) = \underline{\mathbf{e}} \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \\ z \end{pmatrix}.$$

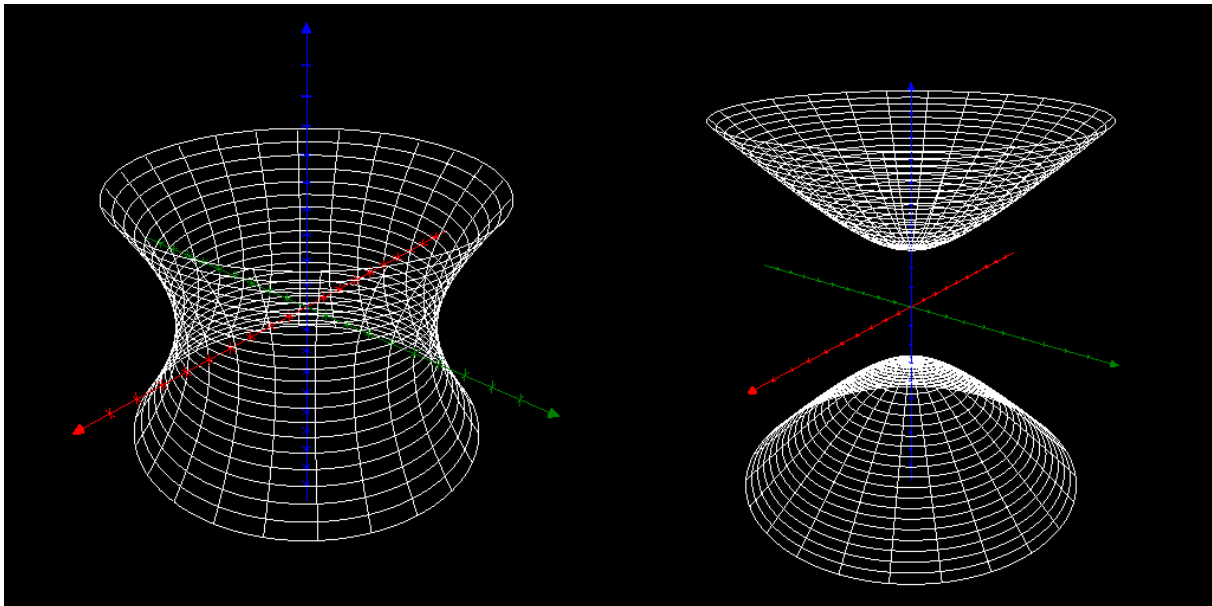
Andra typer av cylindrar är *paraboliska cylindrar* (t.ex. $z = x^2$) och *hyperboliska cylindrar* (t.ex. $x^2 - y^2 = 1$).



3.4.1.6 Hyperboloider

Kurvan $x^2 + y^2 - z^2 = 1$ kallas för en *enmantlad hyperboloid*. Ett snitt $z = c$ ger cirkeln $x^2 + y^2 = 1 + c^2$ så det rör sig om en rotationsyta. I xz -planet $y = 0$ har vi $x^2 - z^2 = 1$ som är en hyperbel, därav namnet.

Ytan $x^2 + y^2 - z^2 = -1$ kallas för en *tvåmantlad hyperboloid*. För fixt $z = c$ har vi kurvan $x^2 + y^2 = c^2 - 1$ som är tomma mängden om $|c| < 1$; sedan erhållet cirklar av ökande radie. Ytan består därför av två disjunkta delar, därav benämningen *tvåmantlad*. För t.ex. $y = 0$ erhålles $x^2 - z^2 = -1$ som är ekvivalent med $z^2 - x^2 = 1$ som är en hyperbel, vilket förklarar återstoden av ytans namn. Nedan visas dessa två vackra varelser.



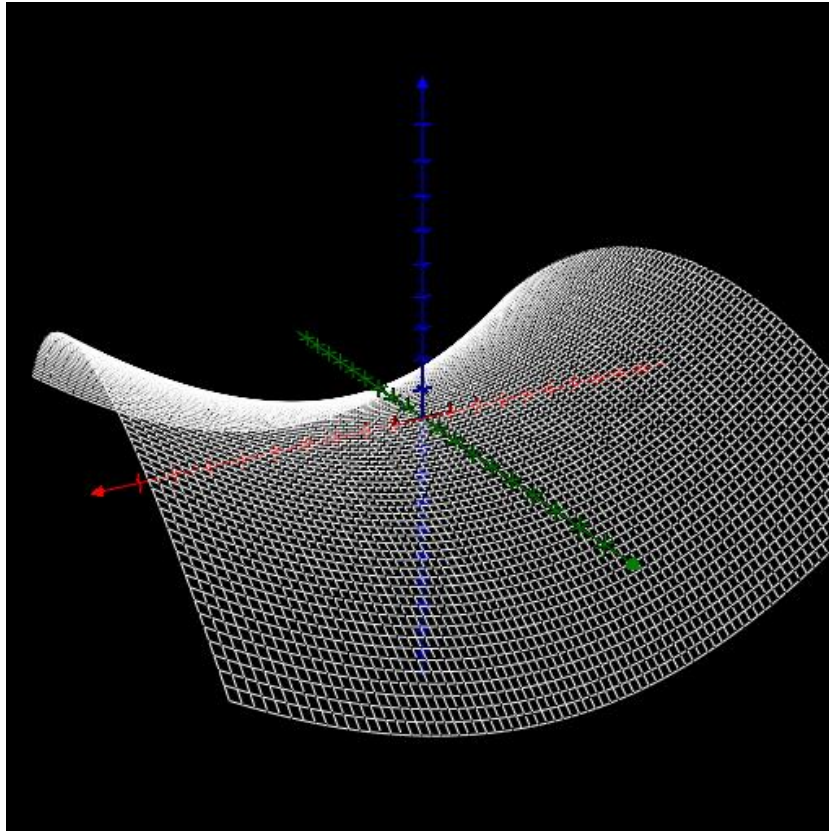
Övningar

1. *Parametrisera *hela* den enmantlade hyperboloiden på *två* sätt. [Ledning: Slå upp ytan på www.trecs.se.]
2. *Parametrisera övre delen av den tvåmantlade hyperboloiden på *två* sätt.

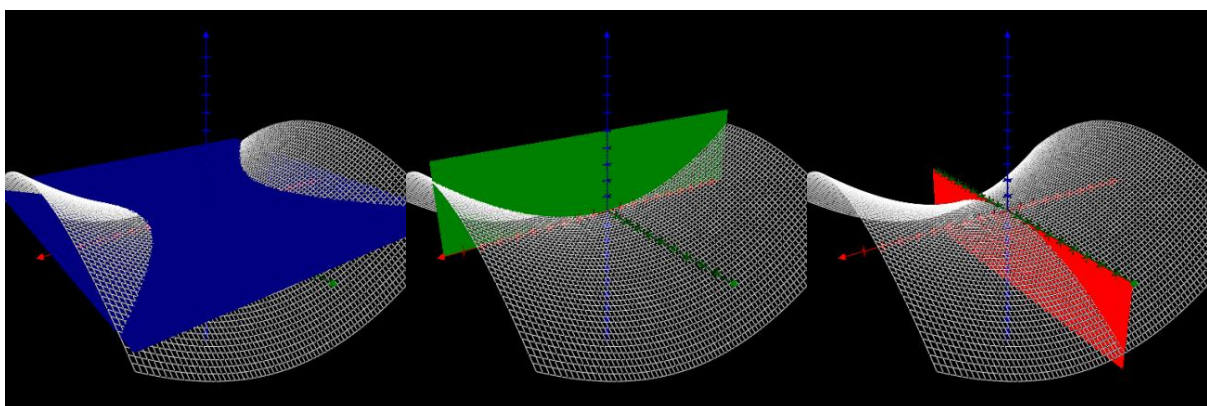
3.4.1.7 Hyperbolisk paraboloid

En oerhört fascinerande yta är den *hyperboliska paraboloiden*, i vilken man hittar både hyperbler och parabler, som namnet utlovar. Den hyperboliska paraboloiden är ytan med ekvationen $z = x^2 - y^2$; jämför med den vanliga (elliptiska) paraboloidens ekvation.

Snitt med plan $z = c$ ger $x^2 - y^2 = c$ som är hyperbler. Å andra sidan ger $y = 0$ parabeln $z = x^2$ medan $x = 0$ ger den spegelvända parabeln $z = -y^2$.



Notera de nämnda plana snitten (hyperbel, parabel, parabel):



Notera att den hyperboliska paraboloiden är formad som en *sadel*.

Övning

1. *Vad är $z = xy$ för yta?

3.4.2 Mer om ytor

Låt $\Sigma = \mathbf{r}(D)$ vara en parametriserad yta, där $D \subset \mathbb{R}^2$ är parameterområdet. Här är \mathbf{r} en funktion av typen $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ så vi kan skriva $\mathbf{r}(u, v) = (x(u, v), y(u, v), z(u, v))$ där (u, v) är koordinater i parameterområdet. Om vi håller v fix får vi en funktion $u \mapsto (x(u, v), y(u, v), z(u, v))$, d.v.s. en funktion av typen $\mathbb{R} \rightarrow \mathbb{R}^3$. Detta är alltså parametriseringsfunktionen för en *kurva* i rummet. Varje fix u ger alltså en kurva, så vi har en *familj* av kurvor.

På liknande sätt, om vi håller u fix får vi en funktion $v \mapsto (x(u, v), y(u, v), z(u, v))$ som är en annan parametriseringsfunktion för en rymdkurva. Varje val av v ger en kurva, så vi har återigen en *familj* av kurvor.

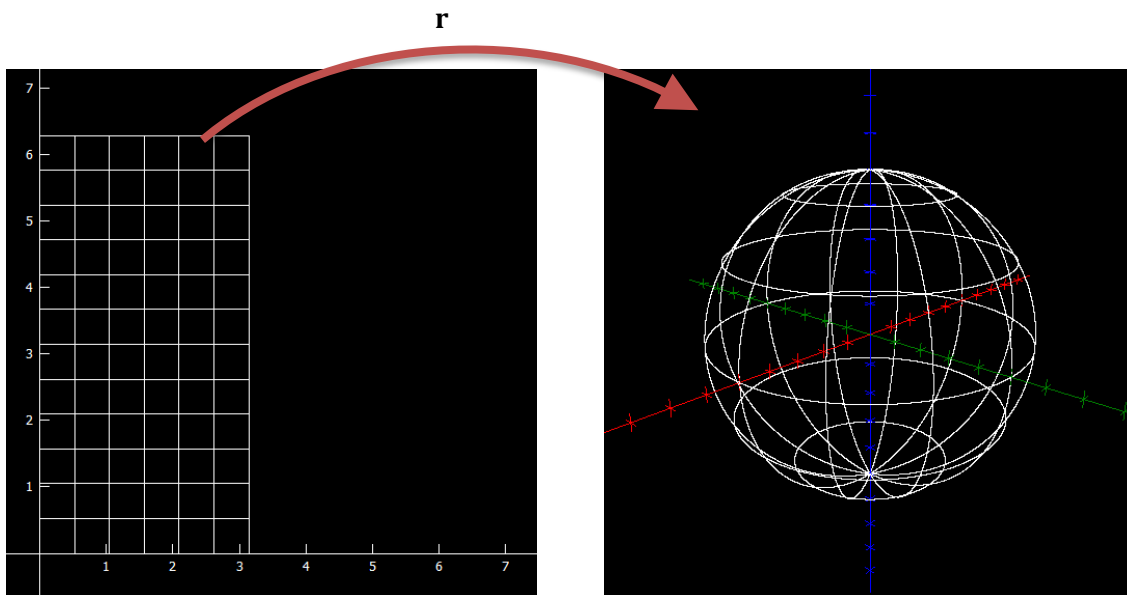
Vi har därmed *två* familjer av kurvor. Dessa kurvor kallas för parametriseringsfunktionens *parameterkurvor*.

För att rita en yta i rummet kan man rita några av dess parameterkurvor. Ofta väljer man då ett *rutnät* $N \subset D$ i parameterområdet, och ritar sedan bilden $\mathbf{r}(N)$. Faktiskt har vi gjort så i varje bild i hela dokumentet!

Låt oss ta enhetsfären som exempel. Enhetsfären är bilden $\mathbf{r}(D)$ av parameterområdet $D := [0, \pi] \times [0, 2\pi[$ under parametriseringsfunktionen

$$\mathbf{r}(\theta, \varphi) = \underline{\mathbf{e}} \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}.$$

Nedan visas ett rutnät $N \subset D$ av parameterområdet samt bilden $\mathbf{r}(N)$ av rutnätet.



Övningar

1. Gå tillbaka och titta på alla bilder på ytor i rummet. Notera att alla består av två familjer av kurvor. I de fall en parametriseringsfunktion för ytan anges, notera också att det är just parameterkurvorna som syns.

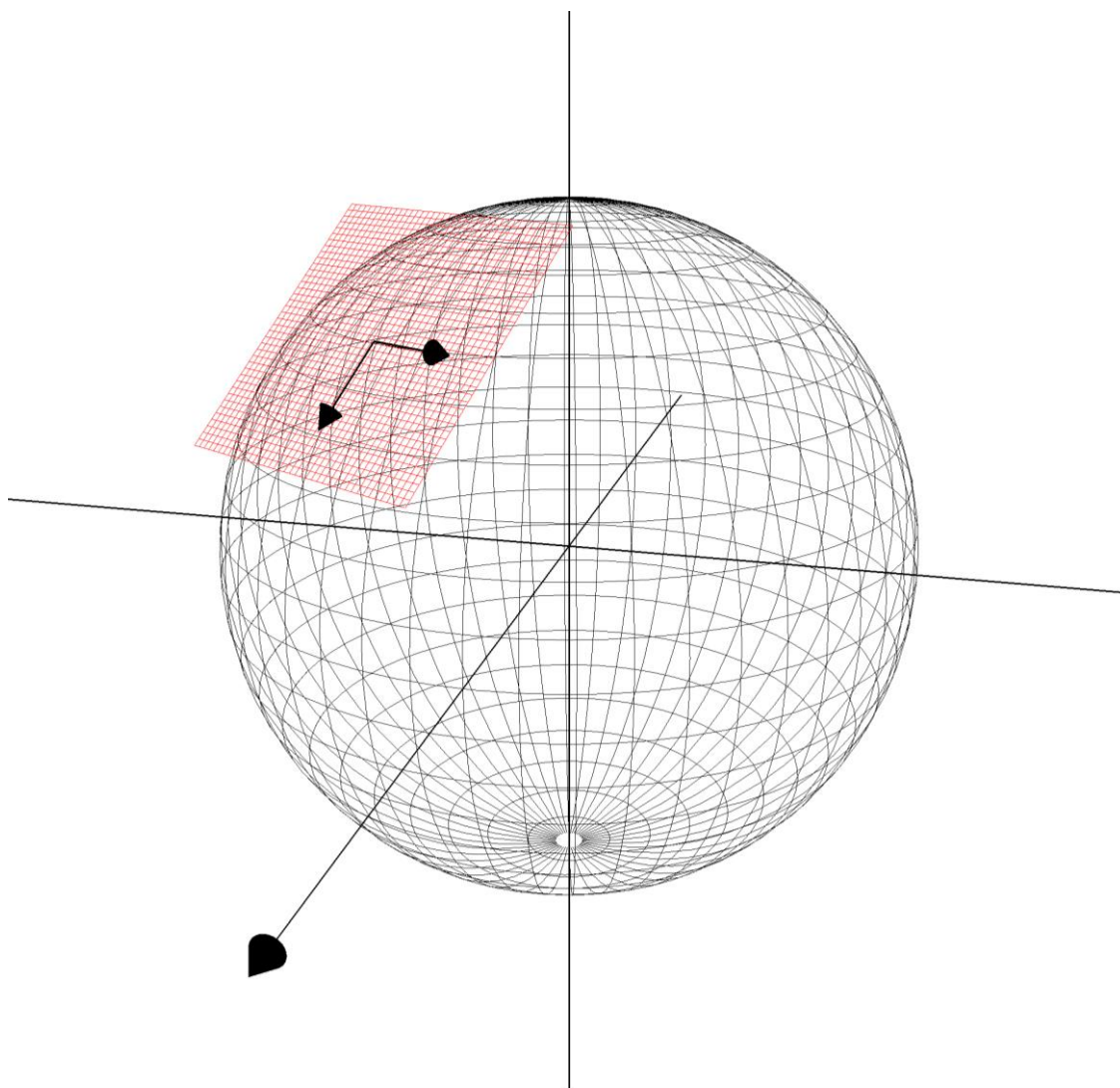
2. Parameterkurvorna är inte en egenskap hos en *yta*, utan hos en *parametriseringsfunktion*. (Och en yta kan ju parametreras på många olika sätt.) "Bevisa" detta genom att titta på parameterkurvorna för de två *olika* parametriseringsfunktionerna för *samma* yta (en paraboloid) i avsnitt 3.4.1.3.

Det är mycket enkelt att bestämma ett *tangentplan* till en parametriserad yta. Om parametriseringsfunktionen är \mathbf{r} och koordinaterna i parameterplanet är (u, v) så är ju $\mathbf{r}'_u(u, v)$ och $\mathbf{r}'_v(u, v)$ parameterkurvornas hastighetsvektorer i punkten $\mathbf{r}(u, v)$. Dessa är (i normalfallet) två linjärt oberoende vektorer som båda är parallella med ytan. Därför spänner de upp tangentplanet, som parametreras

$$(x, y, z) = \mathbf{r}(u, v) + s\mathbf{r}'_u(u, v) + t\mathbf{r}'_v(u, v).$$

Kryssprodukten $\mathbf{r}'_u(u, v) \times \mathbf{r}'_v(u, v)$ ger en normalvektor till ytan, och kan därför användas för att erhålla tangentplanets *ekvation*. (Om en normalvektor t.ex. är $(1, 4, 2)$ så är tangentplanets ekvation $x + 4y + 2z = d$ för något tal d som bestäms av det uppenbara villkoret att tangeringspunkten tillhör tangentplanet.)

Nedan visas tangentplanet i en punkt på en sfär (där parameterkurvorna ges av den vanliga parametriseringsfunktionen).



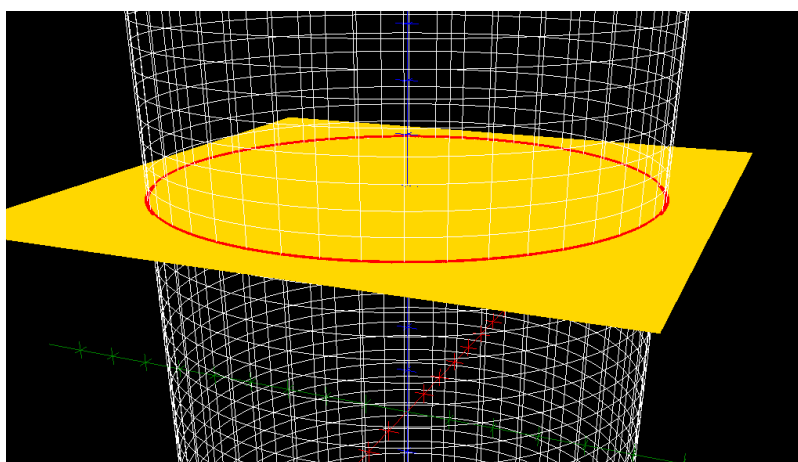
3.5 Skärningskurvor

Om vi är i planet och har *en* ekvation, t.ex. $x^2 + 3y^3 = 3$, så får vi en *kurva*. Om vi är i rummet och har *en* ekvation så får vi i stället en *yta*. Det är alltså normalt sett inte möjligt att ange en *kurva i rummet* med *en* ekvation (för du får ju då en yta, inte en kurva!). Hittills har vi därför bara *parametriserat* kurvor i rummet.

Däremot finns det ett naturligt sätt att ange *en del* rymdkurvor med *två* (icke-ekvivalenta) ekvationer; kurvan är i sådana fall *skärningskurvan* mellan de två ytor som ekvationerna beskriver.

Exempel. Enhetscirkeln γ i planet $z = 5$ är som bekant bilden $\mathbf{r}([0, 2\pi[)$ där $\mathbf{r}(t) = (\cos t, \sin t, 5)$. Men γ är också snittet mellan cylindern $C = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 = 1\}$ och planet $\Pi = \{(x, y, z) \in \mathbb{R}^3: z = 5\}$, d.v.s.

$$\gamma = \mathbf{r}([0, 2\pi[) = C \cap \Pi = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 = 1 \wedge z = 5\}.$$



Övningar

- *Utsagorna "en ekvation i planets koordinater ger en kurva" och "en ekvation i rummets koordinater ger en yta" är sanningar med viss modifikation. Betrakta t.ex. följande mängder. Är det naturligt att kalla dem för "kurvor"?
 - $\{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 0\}$
 - $\{(x, y) \in \mathbb{R}^2: \sin x = 2\}$
 - $\{(x, y) \in \mathbb{R}^2: (x - y)^2 + 2xy = x^2 + y^2\}$
 - $\{(x, y) \in \mathbb{R}^2: \chi_{\mathbb{Q}}(x)\chi_{\mathbb{Q}}(y) = 1\}$ där $\chi_{\mathbb{Q}}(t) = \begin{cases} 1 & \text{om } t \in \mathbb{Q} \\ 0 & \text{om } t \notin \mathbb{Q} \end{cases}$ är den så kallade *karakteristiska funktionen* för delmängden \mathbb{Q} av \mathbb{R} .
- *Utsagan "det är inte möjligt att ange en kurva i rummet med *en* ekvation" är en sanning med modifikation. Låt nämligen $\gamma \subset \mathbb{R}^3$ vara din favoritkurva i rummet och inför dess karakteristiska funktion $\chi_{\gamma}(x, y, z) = \begin{cases} 1 & \text{om } (x, y, z) \in \gamma \\ 0 & \text{om } (x, y, z) \notin \gamma \end{cases}$. Då är det trivialt så att $\gamma = \{(x, y, z) \in \mathbb{R}^3: \chi_{\gamma}(x, y, z) = 1\}$. Tänk på det.

4 Områden i planet och rummet

Vi började vår genomgång med att diskutera allmänna egenskaper hos mängder i \mathbb{R}^n (med "mängder i \mathbb{R}^n " menar vi förstås "delmängder av \mathbb{R}^n "). Sedan specialiserade vi oss på tre speciella typer av delmängder, nämligen kurvor i planet, kurvor i rummet och ytor i rummet. Notera att dessa delmängder samtliga är av lägre dimension än det omgivande rummet. Till exempel är en kurva i planet en endimensionell "sak" i det tvådimensionella rummet, medan en "yta" i rummet är en tvådimensionell sak i det tredimensionella rummet.

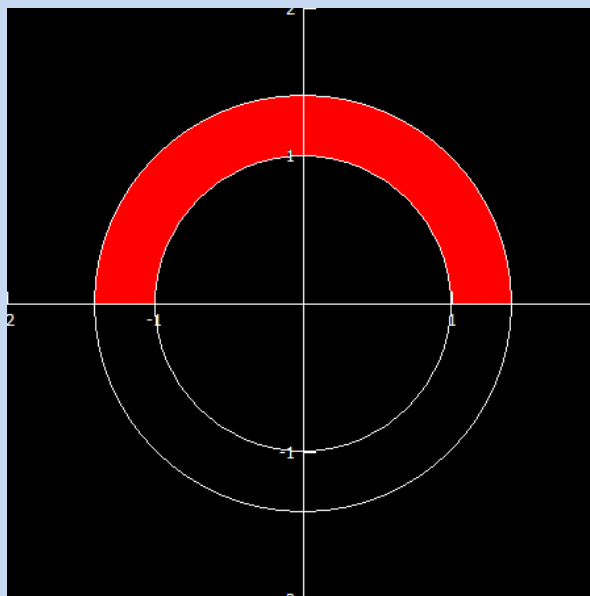
I det här avsnittet specialiserar vi oss i stället på delmängder av samma dimension som rummet, d.v.s. tvådimensionella områden i planet och tredimensionella områden i rummet. Typexemplen är diskar och (fyllda) rektanglar respektive klot och (fyllda) rätblock. Mer intressanta områden kan t.ex. beskrivas med hjälp av kurvor (i planet) och ytor (i rummet). Man kan nämligen beskriva en mängd genom att säga att man skall vara på *ena sidan* om en sådan kurva eller yta.

4.1 Plana områden

Exempel

Rita mängden $D := \{(x, y) \in \mathbb{R}^2: 1 < x^2 + y^2 < 2, y > 0\}$ och ange dess rand ∂D .

Lösning: Vi skall tydligen vara *utanför* cirkeln $x^2 + y^2 = 1$ men *innanför* cirkeln $x^2 + y^2 = 2$; dessa cirklar har radierna 1 respektive $\sqrt{2}$. Dessutom skall vi vara i det övre halvplanet $y > 0$. Vi har alltså följande område:



Randen ∂D består av fyra bitar. Först har vi de övre och undre cirkulära gränserna:

$$C_1 = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 2, y > 0\} \text{ respektive}$$

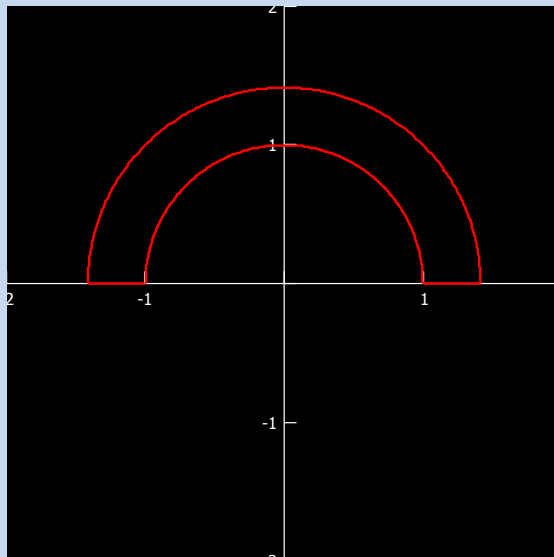
$$C_2 = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 1, y > 0\}.$$

Sedan har vi det vänstra och det högra linjestycket:

$$L_1 = [-\sqrt{2}, -1] \times \{0\} \quad \text{respektive} \quad L_2 = [1, \sqrt{2}] \times \{0\}.$$

Områdets rand är då

$$\partial D = C_1 \cup C_2 \cup L_1 \cup L_2.$$



Anmärkning: Lösningen i exemplet gör några godtyckliga val. Till exempel tillhör "hornpunkterna" $(\pm\sqrt{2}, 0)$ och $(\pm 1, 0)$ $L_1 \cup L_2$ men inte $C_1 \cup C_2$. Genom att byta ut " $y > 0$ " mot " $y \geq 0$ " och "[\cdot, \cdot]" mot " $], \cdot$]" får vi det motsatta förhållandet. Man kan också låta dessa fyra punkter tillhöra de båda tänkbara mängderna i varje fall. I vilket fall som helst blir ju unionen $C_1 \cup C_2 \cup L_1 \cup L_2$ densamma. Däremot går det inte att exkludera punkterna i båda mängderna; då saknas ju dessa i unionen. Man kan förstås också skriva $\partial D = C_1 \cup C_2 \cup L$ där $L = ([-\sqrt{2}, -1] \cup [1, \sqrt{2}]) \times \{0\}$ eller $L = ([-2, 2] \setminus]-1, 1[) \times \{0\}$. Alternativt kan helt skippa namngivningen av de olika "delarna" av randen och skriva direkt

$$\partial D = \{(x, y) \in \mathbb{R}^2: (x^2 + y^2 = 1 \wedge y > 0) \vee (x^2 + y^2 = 2 \wedge y > 0) \vee (-\sqrt{2} \leq x \leq -1 \wedge y = 0) \vee (1 \leq x \leq \sqrt{2} \wedge y = 0)\}.$$

(Matematik är som bekant en *konstform*, och det är upp till författaren att välja sina formuleringar så att konstverket blir så tydligt och vackert som möjligt.)

Notera att medan området är tvådimensionellt, så är dess rand en kurva, d.v.s. en endimensionell "sak". Denna minskning i dimension är i praktiken det normala.

Övningar

1. Rita följande plana områden:
 - a. $[0, \pi] \times [0, 2\pi[$
 - b. $\{(x, y) \in \mathbb{R}^2: x^2 \leq y \leq 1\}$
 - c. $\{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq 1, y \geq |x|\}$
 - d. $\{(x, y) \in \mathbb{R}^2: 1 \leq x^2 + y^2 \leq 4, 0 \leq x \leq y\}$
 - e. $\{(x, y) \in \mathbb{R}^2: 1 \leq x^2 - y^2 \leq 2, 1 \leq xy \leq 2\}$
 - f. $\{(x, y) \in \mathbb{R}^2: 0 \leq x + 2y \leq 3, 0 \leq x - y \leq 1\}$

g. $\{(x, y) \in \mathbb{R}^2: |x| \leq \frac{\pi}{2}, y \leq \cos x\}$.

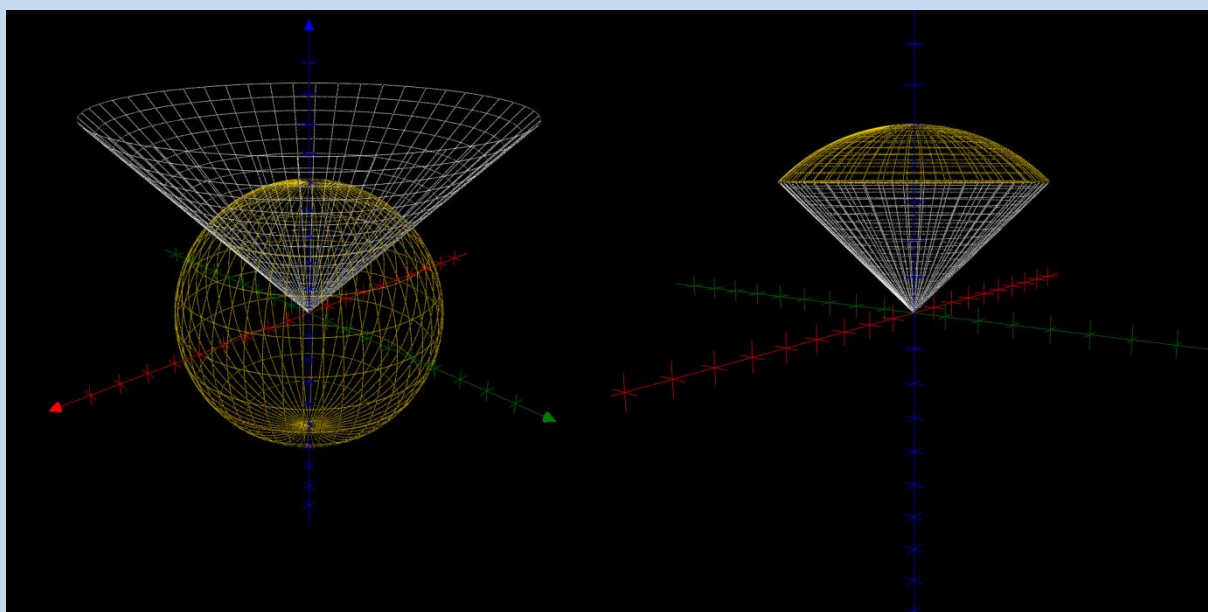
2. Ange randen till varje område i uppgift 1.

4.2 Områden i rummet

Exempel

Rita mängden $D := \{(x, y, z) \in \mathbb{R}^3: \sqrt{x^2 + y^2} \leq z \leq \sqrt{1 - x^2 - y^2}\}$ och ange dess rand ∂D .

Lösning: Eftersom $V_{\sqrt{\quad}} = [0, \infty[$ är det klart att $z \geq 0$ överallt i D . Första olikheten ger $x^2 + y^2 \leq z^2$ som (om vi tänker oss likhet mellan leden) är en dubbelkon, men med tanke på att $z \geq 0$ får vi bara övre delen. Riktningen på olikheten visar att vi skall vara "innanför eller på" konen. Andra olikheten medför $x^2 + y^2 + z^2 \leq 1$ så vi skall vara *innanför eller på* enhets sfären.



Randen delas naturligt in i två delar. Dels har vi den delen av konen $x^2 + y^2 = z^2$ som ligger mellan $z = 0$ och $z = z_s$, dels har vi den del av enhets sfären $x^2 + y^2 + z^2 = 1$ som ligger över $z = z_s$; här är z_s den höjd på vilken konen och sfären skär varandra.

Snittet mellan konen och sfären är uppenbarligen en cirkel på höjden z_s . För att bestämma denna höjd undersöker vi snittet. Om (x, y, z) tillhör både konen och sfären så är $x^2 + y^2 = z^2$ och $x^2 + y^2 + z^2 = 1$ vilket medför att $x^2 + y^2 = 1 - x^2 - y^2$ som är ekvivalent med $x^2 + y^2 = 1/2$. Snittet är alltså en cirkel med radie $1/\sqrt{2}$ på höjden $z = \sqrt{x^2 + y^2} = 1/\sqrt{2}$.

Randen är således $\partial D = C \cup S$ där konbiten

$$C = \left\{ (x, y, z) \in \mathbb{R}^3: x^2 + y^2 = z^2, 0 \leq z \leq \frac{1}{\sqrt{2}} \right\}$$

och sfärbiten

$$S = \left\{ (x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 = 1, z \geq \frac{1}{\sqrt{2}} \right\}.$$

Anmärkning: Sfären huggs alltså av längs en cirkel på höjden $z = 1/\sqrt{2}$, och konen huggs också av längs samma kurva. Denna kurva är som nämnt cirkeln

$$\gamma = \left\{ (x, y, z) \in \mathbb{R}^3: x^2 + y^2 = \frac{1}{2}, \quad z = \frac{1}{\sqrt{2}} \right\} = \mathbf{q}([0, 2\pi[)$$

där $\mathbf{q}(t) = \left(\frac{1}{\sqrt{2}} \cos t, \frac{1}{\sqrt{2}} \sin t, \frac{1}{\sqrt{2}} \right)$.

Även om det är frestande att kalla denna kurva för "randen" till sfärbiten S och konbiten C , så är det förstås inte sant enligt vår definition av rand. [Varje punkt på en yta i rummet är ju en randpunkt till ytan!] Om man däremot parametriserar endera ytan så kommer cirkeln vara bilden av en nyckeldel av randen till parameterområdet i parameterplanet. Till exempel är ju sfärbiten

$$S = \mathbf{r} \left(\left[0, \frac{\pi}{4} \right] \times [0, 2\pi[\right)$$

där $\mathbf{r}: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ är den vanliga parametreringsfunktionen för en sfär. Parameterområdet är rektangeln $\left[0, \frac{\pi}{4} \right] \times [0, 2\pi[$ i $\theta\varphi$ -planet och detta område har en rand som består av fyra räta linjer. Två av dessa, nämligen de med $\varphi = 0$ och $\varphi = 2\pi$, är i praktiken ofta ganska ointressanta eftersom de inte ger upphov till en "kant" på ytan (i princip kan vi ju låta φ minska under 0 eller öka över 2π – då går vi ett nytt varv runt ytan). Däremot är bilden av linjen med $\theta = \pi/4$ precis lika med kurvan γ .

Övningar

1. Rita följande områden i rummet:
 - a. $[0, 1] \times [0, \pi] \times [0, 2\pi[$
 - b. $\{(x, y, z) \in \mathbb{R}^3: x^2 - 4x + y^2 - 2y + z^2 \leq 0, \quad x \geq 2\}$
 - c. $\{(x, y, z) \in \mathbb{R}^3: 2x^2 + y^2 \leq z \leq 1\}$
 - d. $\{(x, y, z) \in \mathbb{R}^3: 1 \leq x^2 + y^2 + z^2 \leq 4\}$.
2. Ange randen till varje område i uppgift 1.

5 Skalärfält

Ett *skalärfält* är en funktion som till varje punkt i rummet eller planet (eller, mer allmänt: \mathbb{R}^n) ordnar ett tal. Det är alltså en funktion av typen $\mathbb{R}^n \rightarrow \mathbb{R}$. Ett typexempel på ett skalärfält är temperaturen i rummet eller planet. Till exempel kanske temperaturen i punkten (x, y, z) är lika med $T(x, y, z) = 2x^2 + yz$. Andra exempel på skalärfält är höjd över havet, densitet, laddningstäthet och en skalärpotential (t.ex. elektrisk potential eller gravitationspotential).

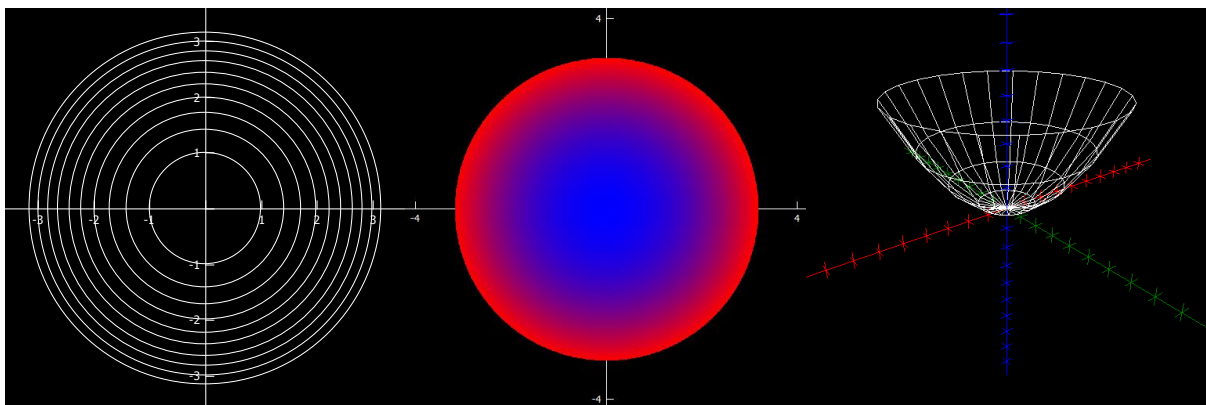
I envariabelanalysen, där vi studerar delmängder av \mathbb{R} och funktioner av typen $\mathbb{R} \rightarrow \mathbb{R}$, brukar vi *visualisera* funktioner genom att rita deras *grafer*. Grafen till en funktion $f: D_f \rightarrow \mathbb{R}$ ($D_f \subset \mathbb{R}$) är som bekant punktmängden $\{(x, y) \in \mathbb{R}^2: x \in D_f \wedge y = f(x)\}$, d.v.s. kurvan med ekvation $y = f(x)$. Vi ställer oss i det här avsnittet frågan hur man kan visualisera skalärfält i högre dimension.

Det finns tre särskilt användbara sätt på vilka man kan visualisera skalärfält. Alla tre fungerar ypperligt i planet, d.v.s. för skalärfält av typen $\mathbb{R}^2 \rightarrow \mathbb{R}$; de fungerar emellertid mindre bra i högre dimensioner.

Så, vi vill visualisera $F: \mathbb{R}^2 \rightarrow \mathbb{R}$. Ett sätt är att rita en väl vald familj av *nivåkurvor*, d.v.s. kurvor $F(x, y) = C$ för väl valda värden på konstanten C . (Om F är temperaturfältet är dessa kurvor *isotermer*.) Ett annat sätt är att rita ett "färgat plan", d.v.s. vi färgar pixeln (x, y) i planet enligt den färg som hör till funktionsvärdet $F(x, y)$ (kanske är blå pixel 0°C och röd pixel 100°C). Ett tredje sätt är att betrakta *graf*en $z = F(x, y)$, som är en yta i rummet. Den senare visualiseringsmetoden är förstås synnerligen lämplig om $F(x, y)$ är höjden över havet i punkten (x, y) – då får vi bokstavligt talat "landskapet".

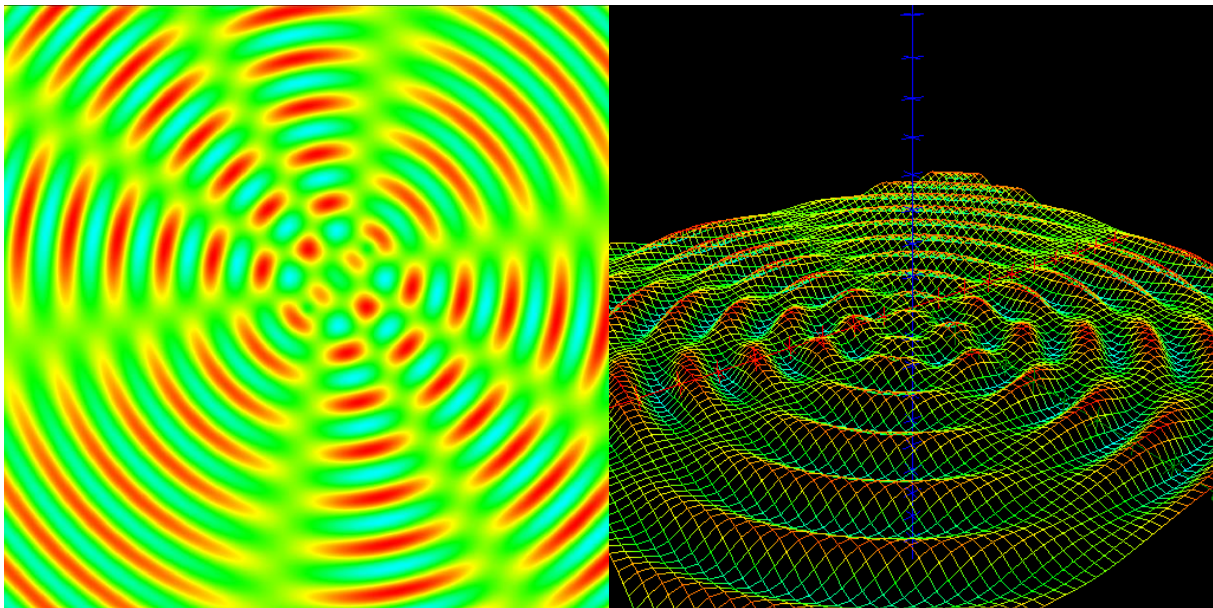
Som ett mycket enkelt exempel betraktar vi skalärfältet $F(x, y) = x^2 + y^2$. Nedan visas

- till vänster: nivåkurvorna med $C = 1, 2, 3, \dots, 10$; samtliga dessa är koncentriska cirklar kring origo. Notera att nivåkurvornas täthet är ett mått på hur snabbt funktionsvärdet ändras.
- i mitten: ett färgat plan där funktionsvärdet 0 svarar mot blå färg och funktionsvärdet 10 svarar mot röd färg. Däremellan ändras R- och B-komponenten i färgen rätlinjigt med funktionsvärdet.
- till höger: grafen $z = F(x, y)$ som tydligen är en (cirkulär) paraboloid.



Nedan visas ett något mer intressant exempel på dels ett "färgat plan", dels en graf till samma skalärfält. I det här fallet kan skalärfältet tänkas beskriva det vertikala utslaget hos en vattenyta efter att

två cirkulära vågor skapats en liten bit ifrån varandra (så vi får ett *interferensmönster*). (För att göra det hela ännu coolare är *grafan* också färgad.) Hur ser nivåkurvorna ut i det här fallet?



Om vi i stället har ett skalärfält i rummet, d.v.s. en funktion $F: \mathbb{R}^3 \rightarrow \mathbb{R}$, hur kan den visualiseras? Något som fungerar är att rita *nivåytor* till funktionen, d.v.s. ytor $F(x, y, z) = C$. Metoden med färgat plan är emellertid svår att översätta till en metod med ett "färgat rum", eftersom de färgade "pixlarna" skulle skymma varandra. Vidare, även om man rent tekniskt kan betrakta grafen $\{(x, y, z, w) \in \mathbb{R}^4: (x, y, z) \in D_F \wedge w = F(x, y, z)\}$, som i typfallet är en (tredimensionell) hyperyta i (det fyrdimensionella rummet) \mathbb{R}^4 , så saknar människor förmågan att föreställa sig sådana geometriska objekt mentalt.

Övning

1. Är enhetscirkeln *grafan* till någon funktion $\mathbb{R} \rightarrow \mathbb{R}$? Är enhetsfären grafen till någon funktion $\mathbb{R}^2 \rightarrow \mathbb{R}$?

5.1 Kurvor och ytor som nivå mängder

Vi har sett att ett skalärfält i planet eller i rummet ger upphov till en familj av kurvor respektive ytor, de så kallade nivåkurvorna respektive nivåytorna. (I allmänhet, för en funktion $\mathbb{R}^n \rightarrow \mathbb{R}$, så talar man om *nivå mängder*.)

Man kan också gå åt andra hållet. Givet en kurva i planet eller en yta i rummet, kan man ofta hitta ett skalärfält (i planet eller rummet) sådant att kurvan/ytan är en nivå mängd till skalärfältet. **Denna procedur är mycket viktig.** Detta är alltid möjligt ifall ytan eller kurvan är given av en ekvation: då flyttar man bara över högerledet (åtminstone alla icke-konstanta termer) till vänstra sidan (eller tvärtom), så att en konstant (som t.ex. noll) är allt som återstår i högerledet. Vänsterledet är då ett uttryck i planets/rummets koordinater, och ger alltså naturligt upphov till ett skalärfält. Hitta då bara på ett namn åt skalärfältet, t.ex. F eller någon följande bokstav, även om ϕ i princip också fungerar.

Exempel. Parabeln $y = x^2$ är nivåkurvan $F(x, y) = 0$ till skalärfältet $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ definierat av $F(x, y) = y - x^2$.

Exempel. Paraboloiden $z = x^2 + y^2$ är nivåytan $G(x, y, z) = 0$ till skalärfältet $G: \mathbb{R}^3 \rightarrow \mathbb{R}$ definierat av $G(x, y, z) = x^2 + y^2 - z$.

Exempel: Sfären $x^2 + y^2 + z^2 = 9$ är nivåytan $H(x, y, z) = 9$ till skalärfältet $H: \mathbb{R}^3 \rightarrow \mathbb{R}$ definierat av $H(x, y, z) = x^2 + y^2 + z^2$.

5.2 Partiella derivator

Om $F: \mathbb{R}^n \rightarrow \mathbb{R}$ är ett skalärfält så är det alltså en funktion som tar in n variabler och ger ifrån sig ett tal. Man kan förstås betrakta alla förutom en (den k :te, säg) av dessa variabler som konstanter; då återstår ett uttryck i en variabel (den k :te). Detta uttryck ger alltså en funktion $\mathbb{R} \rightarrow \mathbb{R}$ och om denna är deriverbar så kan man derivera den. Det erhållna uttrycket (när man återigen betraktar alla n variabler som just variabler) kallas för den k :te *partiella derivatan* av F .

Om man deriverar med avseende på variabeln x så betecknas den partiella derivatan F'_x , $\frac{\partial F}{\partial x}$ eller $\partial_x F$.

Exempel. Låt $F(x, y, z) = x^2y + ze^x$. Då är $F'_x(x, y, z) = 2xy + ze^x$, $F'_y(x, y, z) = x^2$ och $F'_z(x, y, z) = e^x$.

Notera speciellt att om F är ett skalärfält av typen $\mathbb{R}^3 \rightarrow \mathbb{R}$ och den partiella derivatan F'_x existerar, så är F'_x också ett skalärfält av typen $\mathbb{R}^3 \rightarrow \mathbb{R}$. F'_x är alltså precis samma *sorts* objekt som F . (Samma sak gäller samtliga partiella derivator.)

Vad gör man egentligen när man deriverar partiellt? Säg att vi har uttrycket $F(x, y, z)$ – det kan t.ex. vara temperaturen i punkten (x, y, z) i rummet. Vi beräknar nu $F'_x(x, y, z)$; då låter vi y och z vara fixa tal. Den erhållna funktionen $x \mapsto F(x, y, z)$ är då temperaturen vid koordinaten x på linjen med y och z som y - och z -koordinat. Därför är $F'_x(x, y, z)$ ett mått på hur snabbt temperaturen ändras (enhet: grader per meter) vid koordinaten x på denna linje.

Alltså: $F'_x(x, y, z)$ är ett mått på hur mycket temperaturen ökar per meter när man står i punkten (x, y, z) och går i x -led. $F'_y(x, y, z)$ är ett mått på hur mycket temperaturen ökar per meter när man står i punkten (x, y, z) och går i y -led. Och $F'_z(x, y, z)$ är ett mått på hur mycket temperaturen ökar per meter när man står i punkten (x, y, z) och går i z -led.

Om vi i stället är i planet blir det ännu lättare att formulera detta: Om $F(x, y)$ är temperaturen i punkten (x, y) så är $F'_x(x, y)$ så mycket temperaturen ökar (i enheten grader/meter) när man står i punkten (x, y) och går åt *höger*; $F'_y(x, y)$ är så mycket temperaturen ökar (i enheten grader/meter) när man står i punkten (x, y) och går *uppåt*.

Lägg märke till att de partiella derivatorna uppenbarligen också är skalärfält, d.v.s. beror på rummets/planets koordinater. Om vi t.ex. har en kamin i origo i planet så bör $F'_x(x, y)$ vara negativ då $x > 0$, och beloppet $|F'_x(x, y)|$ bör vara stort nära origo och bör avta när vi går ifrån origo.

Om $F: \mathbb{R}^n \rightarrow \mathbb{R}$ och standardbasen i \mathbb{R}^n är $\mathbf{e}_1, \dots, \mathbf{e}_n$ så har vi i formler

$$F'_{x_i}(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{F(\mathbf{x} + h\mathbf{e}_i) - F(\mathbf{x})}{h}.$$

Se till att du förstår denna formel.

5.3 Riktningderivata

Låt $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ vara ett skalärfält. Då är $F'_x(x, y)$ så mycket funktionsvärdet (initialt) ökar per längdenhet om man står i punkten (x, y) och går rakt åt höger, medan $F'_y(x, y)$ är motsvarande förändringshastighet rakt uppåt. Men om man går i någon annan riktning, då?

Vi kommer nu till begreppet *riktningderivata*. Låt $F: \mathbb{R}^n \rightarrow \mathbb{R}$ vara ett allmänt skalärfält och låt $\mathbf{p} \in \mathbb{R}^n$ vara en punkt i definitionsmängden. Låt också $\mathbf{v} \in \mathbb{R}^n$ vara en enhetsvektor som anger en riktning. Då inför vi

$$F'_v(\mathbf{p}) := \lim_{h \rightarrow 0} \frac{F(\mathbf{p} + h\mathbf{v}) - F(\mathbf{p})}{h}$$

som *riktningderivatan* av skalärfältet F i punkten \mathbf{p} och i riktningen \mathbf{v} . Det är uppenbart av definitionen att $F'_v(\mathbf{p})$ är ett tal, nämligen så mycket som F (initialt) ökar per längdenhet om man står i punkten \mathbf{p} och går i riktningen \mathbf{v} .

Riktningderivatan generaliserar de partiella derivatorna, som är riktningderivatorna i de riktningar som ges av koordinataxlarna (eller, ekvivalent, basvektorerna i \mathbb{R}^n), d.v.s. $F'_{\mathbf{e}_i}(\mathbf{p}) = F'_{x_i}(\mathbf{p})$. (Här betecknar, som bekant, $F'_{\mathbf{e}_i}(\mathbf{p})$ riktningderivatan i punkten \mathbf{p} och riktningen \mathbf{e}_i medan $F'_{x_i}(\mathbf{p})$ är den i :te partiella derivatan utvärderad i \mathbf{p} .)

Exempel (men se anmärkningen!)

Låt temperaturen i punkten (x, y) i planet vara $T(x, y) = x^2y$. Om du står i punkten $(1, 1)$ och går i riktningen som ges av vektorn $(1, 2)$, hur mycket ändras då (initialt) temperaturen per meter?

Lösning: En enhetsvektor i den givna riktningen är $\mathbf{v} = \frac{1}{\sqrt{5}}(1, 2)$ så att

$$\begin{aligned} F'_v(1, 1) &= \lim_{h \rightarrow 0} \frac{F\left(1 + \frac{h}{\sqrt{5}}, 1 + \frac{2h}{\sqrt{5}}\right) - F(1, 1)}{h} = \lim_{h \rightarrow 0} \frac{\left(1 + \frac{h}{\sqrt{5}}\right)^2 \left(1 + \frac{2h}{\sqrt{5}}\right) - 1}{h} = \\ &= \lim_{h \rightarrow 0} \left(\frac{4}{\sqrt{5}} + h + \frac{2h^2}{5\sqrt{5}}\right) = \frac{4}{\sqrt{5}}. \end{aligned}$$

Svar: Temperaturen ökar (initialt) med $\frac{4}{\sqrt{5}} \approx 1.8$ grader per meter.

Anmärkning: I praktiken finns det ett mycket smidigare sätt att räkna ut riktningderivator. Vi återkommer till detta efter att vi introducerat den så kallade *gradienten* till skalärfältet.

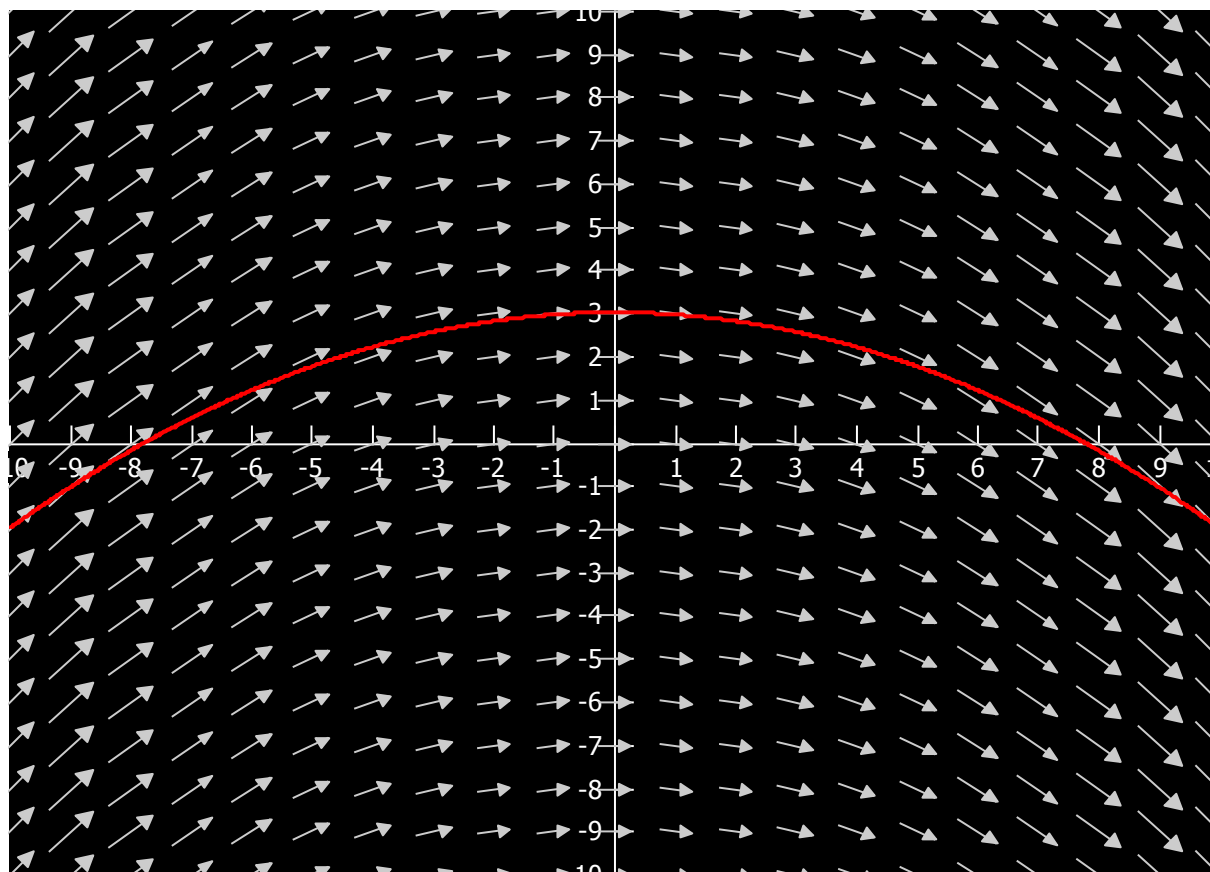
6 Vektorfält

Innan vi fortsätter att studera *skalärfält*, så introducerar vi här begreppet *vektorfält*. Ett skalärfält är som bekant en funktion som till varje punkt i rummet (\mathbb{R}^n , t.ex. planet) ordnar ett *tal*; ett vektorfält är analogt en funktion som till varje punkt i rummet ordnar en *vektor* (i rummets dimension).

Så medan ett skalärfält är en funktion $\mathbb{R}^n \rightarrow \mathbb{R}$ så är ett vektorfält en funktion $\mathbb{R}^n \rightarrow \mathbb{R}^n$.

Exempel. Betrakta vektorfältet $\mathbf{A}(x, y, z) = (x + y, x^2, e^{xz})$. Vektorn i punkten $(1, 2, 3) \in \mathbb{R}^3$ är då $(3, 4, e^3) \in \mathbb{R}^3$.

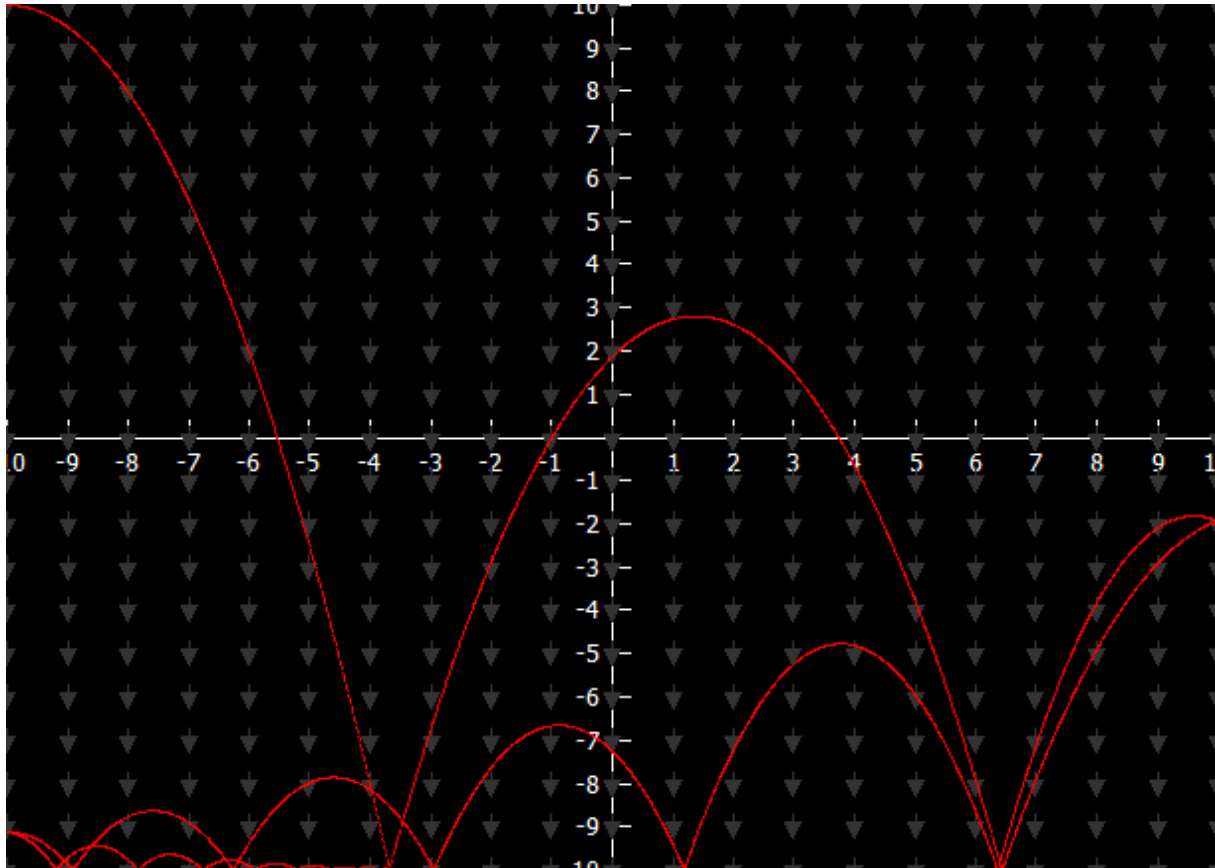
Exempel. Vektorfält är mycket viktiga inom fysiken. Betrakta t.ex. ett flöde av en fluid (gas eller vätska). Om flödet är "stationärt" (d.v.s. tidsberoende) så gäller att fluidens hastighetsvektor bara beror på positionen i rummet, d.v.s. hastigheten \mathbf{v} är ett vektorfält. Hastigheten i punkten (x, y, z) kanske är $\mathbf{v}(x, y, z) = (2, 0, -\frac{1}{5}x)$, t.ex. Nedan illustreras just detta vektorfält, där x -axeln pekar åt höger, y -axeln pekar rakt in mot pappret och z -axeln pekar rakt uppåt.



I bilden visas dessutom den *integralkurva* som börjar i punkten $(-10, -2)$. Detta är *den* kurva som börjar i denna punkt och vars tangentvektor (fysikaliskt: *hastighetsvektor*) i varje punkt är lika med vektorfältets värde i den punkten. Om vi tolkar vektorfältet som flödet av vatten är det alltså den kurva som en vattenmolekyl följer.

Inom fysiken förekommer vektorfält också som *kraftfält*, t.ex. det elektriska fältet kring en laddning, det magnetiska fältet kring en ledare eller gravitationsfältet kring solen. Om en partikel påverkas av

detta kraftfält så är dess *acceleration* alltid lika med en konstant gånger vektorfältets värde i punkten, enligt Newtons andra lag. Nedan visas gravitationsfältet ovanför markytan, och en boll som studsar i detta fält. (Bollen påverkas alltså bara av gravitationsfältet, så dess accelerationsvektor är överallt lika med vektorfältets värde, *förutom* i bildens kanter, där bollen studsar mot golv och väggar.)



7 Skälrfält, fortsättning

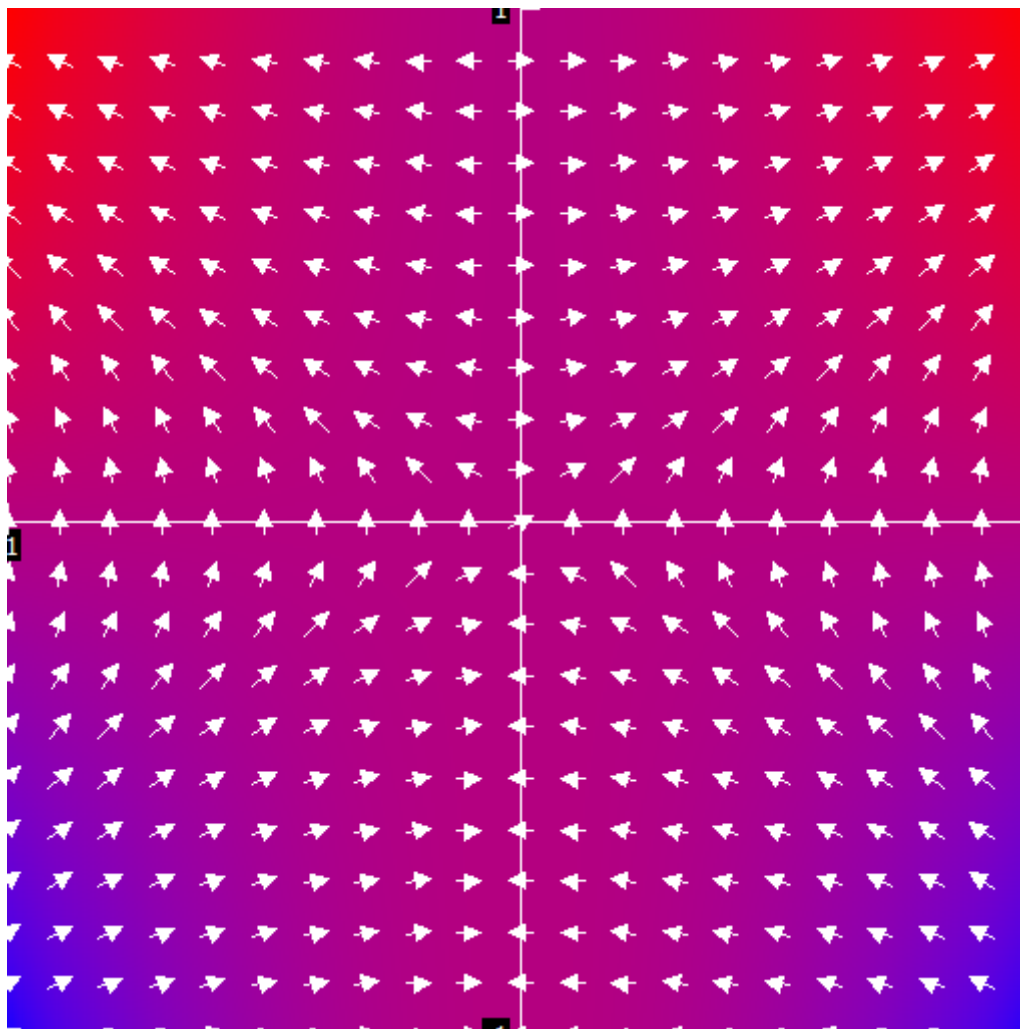
Vi fortsätter nu vår studie av skälrfält.

7.1 Gradient

Till varje skälrfält hör ett vektorfält, den så kallade *gradienten* till skälrfältet. Om $F: \mathbb{R}^n \rightarrow \mathbb{R}$ är ett skälrfält i \mathbb{R}^n så är dess gradient, som vi betecknar ∇F , ett vektorfält i \mathbb{R}^n , d.v.s. $\nabla F: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Hur definieras då ∇F ? Jo, $\nabla F(\mathbf{x})$ är den vektor som består av alla F :s partiella derivator i punkten \mathbf{x} . I formler har vi sålunda

$$\nabla F(\mathbf{x}) = \left(\frac{\partial F}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial F}{\partial x_n}(\mathbf{x}) \right).$$

Exempel. Låt $F(x, y) = x^2y$. Då är $\nabla F(x, y) = (2xy, x^2)$. Nedan visas dels ett färgat plan för F , dels vektorfältet ∇F . Skälrfältets minsta värde är -1 (blått) i nedre vänstra och högra hörnen. Längs koordinataxlarna är värdet 0 och skälrfältets största värde är $+1$ (rött) i övre vänstra och högra hörnen.



Hm... Det verkar som om $\nabla F(x, y)$ alltid pekar åt det "varmaste" hållet? Är detta en slump, eller är det alltid så att gradienten pekar åt det håll som skalärfältet växer mest? *Så är det alltid*. Dessutom är *beloppet* av vektorn lika med skalärfältets riktningsderivata i denna optimala riktning.

Allt detta följer av följande sats:

Sats. Om $F: \mathbb{R}^n \rightarrow \mathbb{R}$ är ett differentierbart skalärfält så är $F'_v(\mathbf{p}) = \nabla F(\mathbf{p}) \cdot \mathbf{v}$

I ord: "Riktningsderivatan av F i punkten \mathbf{p} och i riktningen \mathbf{v} är lika med skalärprodukten mellan F 's gradient utvärderad i \mathbf{p} och (den normerade) riktningsvektorn \mathbf{v} ." (Satsen följer direkt av kedjeregeln.)

Låt oss först göra om vårt gamla exempel med vår nyvunna kunskap.

Exempel

Låt temperaturen i punkten (x, y) i planet vara $T(x, y) = x^2y$. Om du står i punkten $(1, 1)$ och går i riktningen som ges av vektorn $(1, 2)$, hur mycket ändras då (initialt) temperaturen per meter?

Lösning: En enhetsvektor i den givna riktningen är $\frac{1}{\sqrt{5}}(1, 2)$ så att

$$T'_v(1, 1) = \nabla T(1, 1) \cdot \mathbf{v} = \frac{1}{\sqrt{5}} \mathbf{e} \begin{pmatrix} 2 \\ 1 \end{pmatrix} \cdot \mathbf{e} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{4}{\sqrt{5}}$$

Svar: Temperaturen ökar (initialt) med $\frac{4}{\sqrt{5}} \approx 1.8$ grader per meter.

Mycket smidigare!

Från satsen ser vi enkelt det faktum att riktningsderivatan generaliserar de partiella derivatorna, ty om $\mathbf{v} = \mathbf{e}_i$ så är $T'_v(\mathbf{p}) = \nabla T(\mathbf{p}) \cdot \mathbf{e}_i = T'_{x_i}$. Dessutom följer direkt påståendet att gradienten i varje punkt pekar åt det håll skalärfältet ökar mest. Vi har ju, om α är vinkeln mellan $\nabla T(\mathbf{p})$ och \mathbf{v} ,

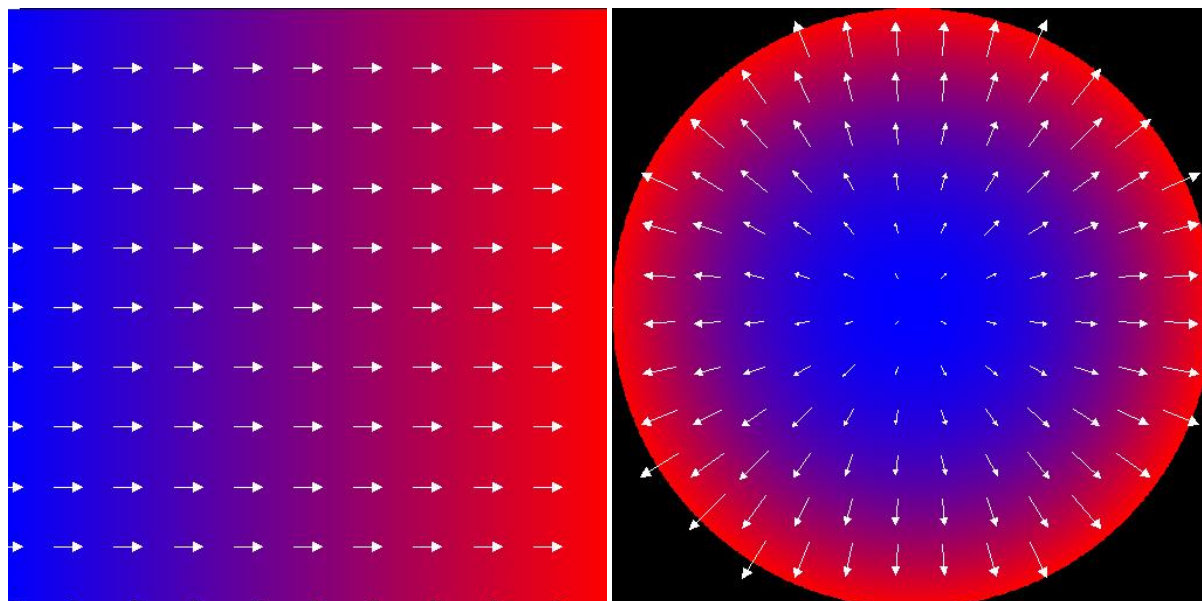
$$T'_v(\mathbf{p}) = \nabla T(\mathbf{p}) \cdot \mathbf{v} = |\nabla T(\mathbf{p})| |\mathbf{v}| \cos \alpha$$

som förstås är som störst då $\alpha = 0$, d.v.s. då \mathbf{v} är parallell med gradienten $\nabla T(\mathbf{p})$. Och i sådana fall är

$$T'_v(\mathbf{p}) = |\nabla T(\mathbf{p})| |\mathbf{v}| = |\nabla T(\mathbf{p})|$$

eftersom \mathbf{v} är en enhetsvektor. Vi ser alltså, som utlovat, att gradientens belopp är lika med just riktningsderivatan i den optimala riktningen.

Exempel. Betrakta skalärfältet $F(x, y) = x$. Det är klart att funktionsvärdet alltid ökar som mest rakt åt höger, och denna optimala riktningsderivata är 1. Detta gäller i *varje* punkt. Och mycket riktigt är gradienten $\nabla F(x, y) = (1, 0)$ det *konstanta* vektorfältet som *alltid* ger vektorn $(1, 0)$, som pekar rakt åt höger och har längd 1. Både skalärfältet och dess gradient visualiseras nedan till vänster.



Exempel. Betrakta skalärfältet $F(x, y) = x^2 + y^2$ (som vi visualiserade på alla tre sätten för ett par kapitel sedan). Det är *a priori* klart att F växer som mest rakt från origo (ty $F(x, y)$ är kvadraten på avståndet från origo!) och att riktningsderivatan i denna riktning *ökar* med ökat avstånd från origo. Och mycket riktigt är gradienten $\nabla F(x, y) = (2x, 2y)$ parallell med Ortsvektorn (x, y) samt har ett belopp som ökar med ökat avstånd från origo. Ovan, till höger, visas färgat plan och gradienten. (Här valde vi att bara visualisera den delen av F och ∇F där $x^2 + y^2 < 1$. Det färgade planet blir lättare att färgsätta då.)

7.2 Gradient och nivå mängder

Formeln $F'_V(\mathbf{p}) = \nabla F(\mathbf{p}) \cdot \mathbf{v}$ gör det uppenbart att $\mathbf{v} \perp \nabla F(\mathbf{p})$ medför $F'_V(\mathbf{p}) = 0$, d.v.s. går man *vin- kelrätt* mot gradienten så ändras funktions värde inte alls. Med andra ord går man längs en nivå- mängd. En vanligare formulering av denna observation är följande:

Gradienten i punkten \mathbf{p} är vinkelrät mot den nivå mängd som passerar genom \mathbf{p} .

Den här insikten har ganska stor praktisk betydelse i vissa problem.

Exempel

Bestäm på parameterform tangentlinjen och normallinjen till kurvan $x^4 + y^4 = 1 - xy^2 - yx^2$ i punkten $(1,0)$.

Lösning: (Vi noterar först att punkten verkligen *tillhör* kurvan.) Inför skalärfältet $F(x, y) := x^4 + y^4 + xy^2 + yx^2$; då är tydligen kurvan lika med nivåkurvan $F(x, y) = 1$. Eftersom

$$\nabla F(x, y) = (4x^3 + y^2 + 2xy, 4y^3 + 2xy + x^2)$$

i allmänhet så är speciellt

$$\nabla F(1,0) = (4, 1)$$

så att normallinjen till kurvan är bilden av $\mathbf{r}(t) = (1,0) + t(4,1)$. En tangentvektor till kurvan i aktuell punkt är uppenbarligen $(1, -4)$ så att tangentlinjen är bilden av $\mathbf{r}(t) = (1,0) + t(1, -4)$.

Svar: Normallinjen är bilden av $\mathbf{r}(t) = (1,0) + t(4,1)$; tangentlinjen är bilden av $\mathbf{r}(t) = (1,0) + t(1,-4)$.

Det är precis lika lätt att bestämma *ekvationer* för tangent och normal:

Exempel

Bestäm ekvationer för tangentlinjen och normallinjen till kurvan $x^4 + y^4 = 1 - xy^2 - yx^2$ i punkten $(1,0)$.

Lösning: (Vi noterar först att punkten verkligen *tillhör* kurvan.) Inför skalärfältet $F(x, y) := x^4 + y^4 + xy^2 + yx^2$; då är tydligen kurvan lika med nivåkurvan $F(x, y) = 1$. Eftersom

$$\nabla F(x, y) = (4x^3 + y^2 + 2xy, 4y^3 + 2xy + x^2)$$

i allmänhet så är speciellt

$$\nabla F(1,0) = (4, 1).$$

Eftersom detta är en normalvektor till kurvan så har *tangentlinjen* ekvationen $4x + y = d$ för något $d \in \mathbb{R}$. Eftersom $(1,0)$ tillhör linjen måste $d = 4$.

En tangentvektor till kurvan är uppenbarligen $(1, -4)$ så att *normallinjen* har ekvationen $x - 4y = d$ för något $d \in \mathbb{R}$. Eftersom $(1,0)$ tillhör linjen måste $d = 1$.

Svar: Normalen är $4x + y = 4$ och tangenten är $x - 4y = 1$.

Vi kan också studera *ytor* på motsvarande sätt:

Exempel

Bestäm tangentplanet till ytan $x^2 + 2y^2 + xy + z^2 + 3xz = 1$ i punkten $(1, 0, -3)$.

Lösning: (Vi noterar först att punkten verkligen *tillhör* ytan.) Inför skalärfältet $F(x, y, z) := x^2 + 2y^2 + xy + z^2 + 3xz$; då är ytan lika med nivåytan $F(x, y, z) = 1$. Gradienten

$$\nabla F(x, y, z) = (2x + y + 3z, 4y + x, 2z + 3x)$$

så att

$$\nabla F(1, 0, -3) = (-7, 1, -3)$$

är en normalvektor till ytan i punkten $(1, 0, -3)$. Ytans tangentplan i den här punkten har således ekvationen $-7x + y - 3z = d$ för någon konstant $d \in \mathbb{R}$. Eftersom $(1, 0, -3)$ tillhör planet måste $d = 2$.

Svar: Det sökta tangentplanet är $7x - y + 3z = -2$.

I dessa uppgifter ges vi alltså en kurva eller yta – d.v.s. en punktmängd – och sedan *inför* vi ett skalärfält i rummet sådant att mängden är en nivå mängd till funktionen. Slutligen bestämmer vi *skalärfältets* gradient, och använder insikten att gradientvektorn i en punkt på kurvan/ytan alltid är vinkelrät mot densamma.

Det är viktigt att man inte blandar ihop mängden med skalärfältet! Inte sällan ser man följande fel när man rättar skrivningar:

Ett ganska vanligt fel

Bestäm tangentplanet till ytan $x^2 + 2y^2 + xy + z^2 + 3xz = 1$ i punkten $(1, 0, -3)$.

Lösning: Ytans gradient är $(2x + y + 3z, 4y + x, 2z + 3x)$...

Detta är "obegripligt". Begreppet *gradient* är ju som bekant en egenskap hos ett *skalärfält*, d.v.s. hos en *funktion* (av typen $\mathbb{R}^n \rightarrow \mathbb{R}$). En *yta* (eller en kurva, för den delen), d.v.s. en (slags) delmängd av \mathbb{R}^n , har ingen "gradient".³

Felet här är alltså att man behandlar punktmängden som om den vore ett skalärfält. Även det omvända felet förekommer, att studenten behandlar ett skalärfält som om det vore en punktmängd:

Ett nästan lika vanligt fel

Bestäm tangentplanet till ytan $x^2 + 2y^2 + xy + z^2 + 3xz = 1$ i punkten $(1, 0, -3)$.

Lösning:

$$F(x, y, z) := x^2 + 2y^2 + xy + z^2 + 3xz$$

$$\nabla F(x, y, z) = (2x + y + 3z, 4y + x, 2z + 3x)$$

$$\nabla F(1, 0, -3) = (-7, 1, -3)$$

Tangentplan: $-7x + y - 3z = d$.

Eftersom $(1, 0, -3) \in F$ ser jag att $d = 2$.

Problemet här är " $(1, 0, -3) \in F$ ". Själva *yta*

$$S = \{(x, y, z) \in \mathbb{R}^3: x^2 + 2y^2 + xy + z^2 + 3xz = 1\}$$

är en punktmängd, närmare bestämd en delmängd till \mathbb{R}^3 , d.v.s. $S \subset \mathbb{R}^3$. En av punkterna i S är $(1, 0, -3)$, så vi kan skriva $(1, 0, -3) \in S$.

Men F är ju inte ens en mängd! Så F "innehåller" inte något alls! Det studenten skriver är ju att $(1, 0, -3)$ "tillhör" funktionen $F(x, y, z) := x^2 + 2y^2 + xy + z^2 + 3xz$. Vad i all sin dar betyder det?!

(Notera också den otydliga redovisningen. Man bör, som läsaren säkerligen redan känner till, använda svenska språket för att skapa en "röd tråd" genom lösningen; det mänskliga språket förbinder formlerna och anger deras relation till varandra. "Jag"-formen känns vidare något oäven.)

³ Visst, om kurvan eller ytan är given som en ekvation finns det ofta ett "naturligt" val av skalärfält (så när som på tecknet), men en kurva/yta måste ju inte vara angiven med en ekvation. Den kan t.ex. vara given på parameterform. I sådana fall har man förvisso (oftast) en väldefinierad normalriktning i varje punkt på mängden, men vektorns längd och riktning (åt ena sidan eller åt andra) är obestämd. Dessutom brukar ju ett vektorfält vara definierat i en hel öppen mängd i rummet (t.ex. hela \mathbb{R}^n), inte bara längs en yta eller kurva i rummet!

7.2.1 Fler exempel

Exempel

Bestäm de punkter på kurvan $x^2 + 2xy + y^2 + 2x = 1$ i vilka normallinjen passerar genom origo.

Lösning: Låt $F(x, y) := x^2 + 2xy + y^2 + 2x$ så att den givna kurvan γ är nivåkurvan $F(x, y) = 1$. En punkt på kurvan har en normallinje som passerar origo om kurvans normal i denna punkt är parallell med punktens Ortsvektor. Låt (a, b) vara en sådan punkt; då är alltså

$$\begin{aligned} \left\{ \begin{array}{l} (a, b) \in \gamma \\ \nabla F(a, b) \parallel (a, b) \end{array} \right. &\iff \left\{ \begin{array}{l} a^2 + 2ab + b^2 + 2a = 1 \\ \left| \begin{array}{cc} 2a + 2b + 2 & 2a + 2b \\ a & b \end{array} \right| = 0 \end{array} \right. \iff \\ &\iff \left\{ \begin{array}{l} a^2 + 2ab + b^2 + 2a = 1 \\ b^2 + b - a^2 = 0. \end{array} \right. \end{aligned}$$

Första ekvationen minus andra ger

$$\begin{aligned} 2a^2 + 2ab + 2a - b &= 1 \iff b(2a - 1) = 1 - 2a^2 - 2a \iff \\ \iff b &= \frac{1 - 2a^2 - 2a}{2a - 1} \end{aligned}$$

eftersom $a \neq 1/2$. Instöpat i andra ekvationen ger detta

$$\begin{aligned} \left(\frac{1 - 2a^2 - 2a}{2a - 1} \right)^2 + \frac{1 - 2a^2 - 2a}{2a - 1} - a^2 &= 0 \iff \\ \iff (1 - 2a^2 - 2a)^2 + (1 - 2a^2 - 2a)(2a - 1) - a^2(2a - 1)^2 &= 0 \iff \\ \iff 8a^3 - 3a^2 = 0 \iff a^2(8a - 3) = 0 \iff \\ \iff a = 0 \text{ eller } a = \frac{3}{8}. \end{aligned}$$

Respektive värde på a ger $b = -1$ och $b = \frac{1}{8}$.

Svar: Tangentlinjen går genom origo i punkterna $(0, -1)$ och $(\frac{3}{8}, \frac{1}{8})$.

Exempel

Visa att kurvan $x^3 + y^3 = 1$ skär koordinataxlarna under rät vinkel.

Lösning: Kurvan skär x -axeln när $y = 0$, vilket inträffar när $x = 1$; kurvan skär y -axeln när $x = 0$, d.v.s. när $y = 1$. Låt $F(x, y) := x^3 + y^3$ så att kurvan är nivåkurvan $F(x, y) = 1$. Eftersom $\nabla F(x, y) = (3x^2, 3y^2)$ så är kurvans normal i första skärningspunkten $\nabla F(1, 0) = (3, 0)$ som pekar rakt åt höger. Alltså är kurvans tangentlinje vertikal här, som önskat.

På samma sätt ser vi att $\nabla F(0, 1) = (0, 3)$ som pekar rakt uppåt, så att kurvans tangentlinje är horisontell här.

Exempel

Bestäm de punkter på ellipsoiden $x^2 + xy + 2y^2 + z^2 = 5$ i vilka tangentplanet är parallellt med planet $x + y = 0$.

Lösning: Inför $F(x, y, z) = x^2 + xy + 2y^2 + z^2$ så att den givna ytan S är nivåytan $F(x, y, z) = 5$. Låt (a, b, c) vara en intressant punkt. Detta betyder precis att

$$\begin{aligned} \left\{ \begin{array}{l} (a, b, c) \in S \\ \nabla F(a, b, c) \parallel (1, 1, 0) \end{array} \right. &\iff \left\{ \begin{array}{l} a^2 + ab + 2b^2 + c^2 = 5 \\ (2a + b, a + 4b, 2c) \times (1, 1, 0) = \mathbf{0} \end{array} \right. \iff \\ &\iff \left\{ \begin{array}{l} a^2 + ab + 2b^2 + c^2 = 5 \\ -2c = 0 \\ 2c = 0 \\ a - 3b = 0 \end{array} \right. \iff \left\{ \begin{array}{l} a^2 + ab + 2b^2 = 5 \\ c = 0 \\ a = 3b \end{array} \right. \iff \\ &\iff \left\{ \begin{array}{l} 14b^2 = 5 \\ c = 0 \\ a = 3b \end{array} \right. \iff (a, b, c) = \pm \left(3\sqrt{\frac{5}{14}}, \sqrt{\frac{5}{14}}, 0 \right). \end{aligned}$$

Svar: I punkterna $\pm \left(3\sqrt{\frac{5}{14}}, \sqrt{\frac{5}{14}}, 0 \right)$ är ellipsoidens tangentplan parallellt med planet $x + y = 0$.

7.3 Sammansättningar av funktioner - kurvor i landskap

Sedan tidigare matematikstudier är läsaren bekant med begreppet funktionsammansättning. Låt X , Y och Z vara mängder och antag att du har en funktion $f: X \rightarrow Y$ och en funktion $g: Y \rightarrow Z$. Välj ett $x \in X$. Du får då en bild $y := f(x)$. Eftersom $y \in Y$ kan du stoppa in detta värde i funktionen g , varvid du erhåller elementet $z := g(y) \in Z$. Totalt sett har du då beräknat $z = g(f(x))$, d.v.s. du har fått ett entydigt element $z \in Z$ givet ett element $x \in X$. Du har alltså en funktion $X \rightarrow Z$ definierad av $x \mapsto g(f(x))$. Denna funktion kallas för *sammansättningen* av f och g och betecknas $g \circ f$. Vi har alltså $g \circ f: X \rightarrow Z$ och kan skriva $z = (g \circ f)(x)$.

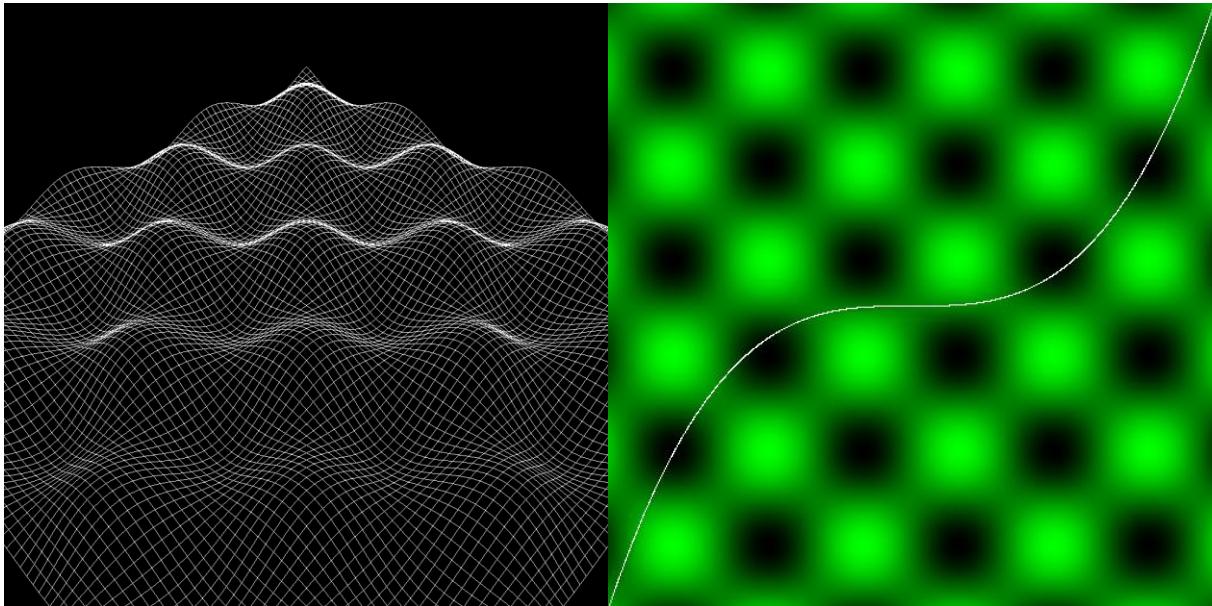
Exempel. Låt $f: \mathbb{R} \rightarrow \mathbb{R}$ och $g: \mathbb{R} \rightarrow \mathbb{R}$ vara definierade av $f(x) = x + 2$ och $g(x) = \sin x$. Då är $(f \circ g)(x) = f(g(x)) = f(\sin x) = \sin x + 2$ medan $(g \circ f)(x) = g(f(x)) = g(x + 2) = \sin(x + 2)$. Vi kan speciellt göra observationen att operationen funktionsammansättning *inte* är kommutativ. Funktionerna $x \mapsto \sin x + 2$ och $x \mapsto \sin(x + 2)$ är ju inte lika, ty det *inte* är sant att $\sin x + 2 = \sin(x + 2)$ för alla $x \in \mathbb{R}$.

Nu ett par typiska flervariabel exempel:

Exempel. Låt $T(x, y)$ vara temperaturen i punkten (x, y) i ett landskap; T är alltså ett skalärfält av typen $\mathbb{R}^2 \rightarrow \mathbb{R}$. Låt också $\mathbf{r}(t)$ vara en cyklists position i landskapet vid tiden t ; \mathbf{r} är alltså en funktion av typen $\mathbb{R} \rightarrow \mathbb{R}^2$. Sammansättningen $T \circ \mathbf{r}$ är då en funktion av typen $\mathbb{R} \rightarrow \mathbb{R}$ och ger temperaturen vid tiden t på cykelfärden.

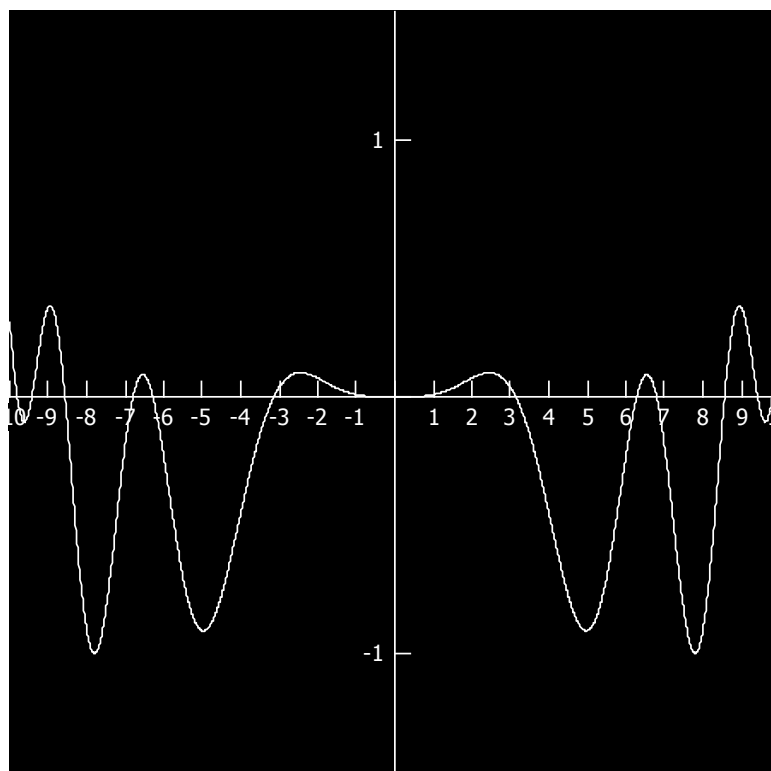
Exempel. Låt $z(x, y)$ vara höjden över havet i punkten (x, y) i ett landskap; z är alltså ett skalärfält av typen $\mathbb{R}^2 \rightarrow \mathbb{R}$. Låt också $\mathbf{r}(t)$ vara en cyklists geografiska position i landskapet vid tiden t ; \mathbf{r} är alltså en funktion av typen $\mathbb{R} \rightarrow \mathbb{R}^2$. Sammansättningen $z \circ \mathbf{r}$ är då en funktion av typen $\mathbb{R} \rightarrow \mathbb{R}$ och cykelfärdens *höjdprofil*, d.v.s. höjden över havet vid tiden t under cykelturen.

Exempel. Betrakta det böljande landskapet $z = \sin x \sin y$ (ytan är *graf*en till funktionen $F: (x, y) \mapsto \sin x \sin y$). Nedan, till vänster, visas detta landskap.

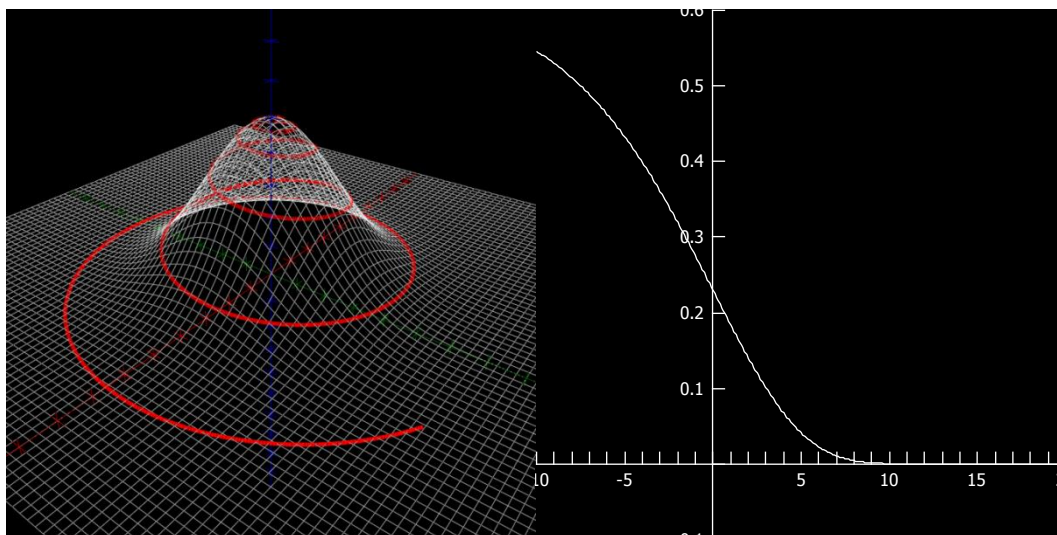


Betrakta också cykelturen $\mathbf{r}(t) = \left(t, \frac{1}{100}t^3\right)$ genom landskapet. Ovan, till höger, visas dels ett färgat plan för landskapsfunktionen F (svart: lågt; grönt: högt), dels cykelturen (i $[-10, 10]^2$). Notera att cykelturen främst sker vid låg höjd; dessutom är höjden nästan konstant under raksträcken i "mitt-partiet".

Sammanställningen $F \circ \mathbf{r}: \mathbb{R} \rightarrow \mathbb{R}$ ges av $F(\mathbf{r}(t)) = \sin t \sin\left(\frac{1}{100}t^3\right)$; detta är alltså höjden vid tiden t på cykelturen. Nedan visas grafen till $F \circ \mathbf{r}$, d.v.s. "höjdprofilen" för turen.



Exempel (*). Betrakta berget $z = 5e^{-\frac{x^2+y^2}{8}}$ som är grafen till skalärfältet $F: (x, y) \mapsto 5e^{-\frac{x^2+y^2}{8}}$. Det förefaller inte helt orimligt att en bilväg upp till bergets topp går som en spiral runt berget. Man kan t.ex. tänka sig att vägen är bilden $\mathbf{r}([-10, 20])$ där $\mathbf{r}(t) = (e^{t/10} \cos t, e^{t/10} \sin t)$. Nedan, till vänster, visas dels berget, dels bilvägen på berget, d.v.s. $\{(x, y, z) \in \mathbb{R}^3: (x, y) \in \mathbf{r}([-10, 20]), z = F(x, y)\}$. [”Bilvägen” $\mathbf{r}([-10, 20])$ är ju annars en *plan* kurva.]



Sammansättningen $F \circ \mathbf{r}: \mathbb{R} \rightarrow \mathbb{R}$ ger höjden vid tidpunkt t . Vi har, vilket illustreras ovan till höger,

$$F(\mathbf{r}(t)) = F(e^{t/10} \cos t, e^{t/10} \sin t) = 5e^{-\frac{e^{t/5}}{8}}.$$

Observera att funktionen $F \circ \mathbf{r}$ ger höjden vid *tiden* t under resan som parametriseras med \mathbf{r} . Grafen till den här funktionen beror alltså på den *hastighet* med vilken \mathbf{r} parametriserar kurvan (vägen).

Resan börjar (nästan) längst uppe på berget, och såväl den horisontella farten $|\mathbf{r}'(t)|$ som den riktiga farten ökar med tiden. Kom ihåg att *funktionen* \mathbf{r} innehåller mer information än *kurvan* $\mathbf{r}([-10, 20])$, som ju kan parametriseras på många olika sätt. *Funktionen* \mathbf{r} innehåller inte bara information om själva ”vägen” (d.v.s. bilden $\mathbf{r}([-10, 20])$), utan \mathbf{r} innehåller också information om kurvans *orientering* (i det här fallet från bergets topp och nedåt) och den hastighet med vilken kurvan genomlöps. ”Höjdprofilen” som visas ovan är alltså inte en egenskap hos skalärfältet och *kurvan*, utan en egenskap hos skalärfältet och *parametriseringsfunktionen*. Det går inte att erhålla ”höjdprofilen” i den högra bilden bara genom att titta på den vänstra bilden.

För att skapa en mer ”äkta” höjdprofil kan man parametrisera vägen på ett sådant sätt att farten (förslagsvis den tredimensionella) blir konstant. I sådana fall blir parametervärdet ett direkt mått på (proportionell mot) *båglängden*, d.v.s. längden längs kurvan. Vi får alltså ”höjden vid positionen s längs kurvan från bergets topp” i stället för ”höjden vid tidpunkt t på resan som bestäms av \mathbf{r} ”.

Övningar

- *Skriv ner den funktion $\mathbf{q}: \mathbb{R} \rightarrow \mathbb{R}^3$ som på ett naturligt sätt erhålles av \mathbf{r} och F och som ger bilvägen på berget (rymdkurva).
- *Bestäm dels den horisontella hastigheten $\mathbf{r}'(t)$, dels den riktiga hastigheten $\mathbf{q}'(t)$. Notera att farten ($|\mathbf{r}'(t)|$ respektive $|\mathbf{q}'(t)|$) *inte* är konstant.

7.4 Gränsvärden

Läsaren är väl bekant med gränsvärden för *envariabelfunktioner*.⁴ I det här avsnittet introducerar vi gränsvärdesbegreppet för *flervariabelfunktioner*. Både den intuitiva idén och de precisa definitionerna är helt analoga, så läsare som är förtrogna med envariabelfallet bör inte ha några större konceptuella problem med flervariabelfallet.

I envariabelfallet säger vi, t.ex., att " $f(x) \rightarrow A$ då $x \rightarrow a$ " omm det för varje $\epsilon > 0$ finns ett tal $\delta > 0$ sådant att avståndet $|f(x) - A| < \epsilon$ för alla $x \in]a - \delta, a + \delta[\setminus \{a\}$. Intervallet $]a - \delta, a + \delta[$ är en *öppen omgivning* med radie δ kring a , och kan också skrivas $B(a, \delta)$. Denna mängd består av alla punkter som har *avståndet* från a mindre än δ .

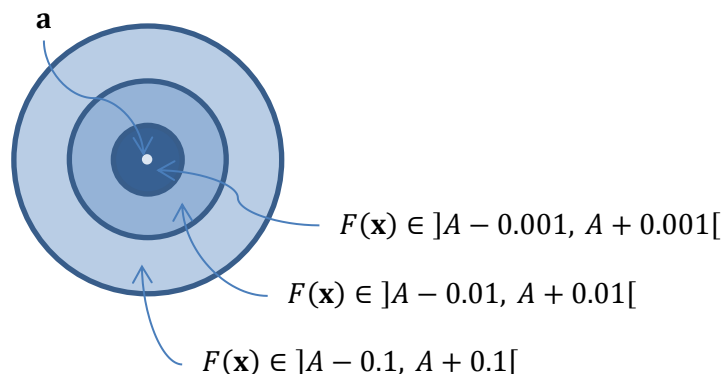
I flervariabelanalysen har vi i stället funktioner $F: \mathbb{R}^n \rightarrow \mathbb{R}$. Mängden av alla punkter som har avståndet från \mathbf{a} mindre än δ betecknas i allmänhet $B(\mathbf{a}, \delta)$. I planet är detta en (öppen) disk; i rummet är det ett (öppet) klot. Den senare benämningen används också generellt. Den uppenbara generaliseringen av gränsvärdesbegreppet är sålunda följande.

Definition. Låt $F: \mathbb{R}^n \rightarrow \mathbb{R}$. Vi säger att " $F(\mathbf{x}) \rightarrow A$ då $\mathbf{x} \rightarrow \mathbf{a}$ " omm det för varje $\epsilon > 0$ finns ett tal $\delta > 0$ sådant att $|F(\mathbf{x}) - A| < \epsilon$ för alla $x \in B(\mathbf{a}, \delta) \setminus \{\mathbf{a}\}$.

Definitionen ovan gäller tydligen bara för skalärfält F definierade på *hela* \mathbb{R}^n . För ett allmänt skalärfält definierat på en mängd $D_F \subset \mathbb{R}^n$ gör vi följande definition:

Definition. Låt $F: D_F \rightarrow \mathbb{R}$ med $D_F \subset \mathbb{R}^n$. Vi säger att " $F(\mathbf{x}) \rightarrow A$ då $\mathbf{x} \rightarrow \mathbf{a}$ " omm det för varje $\epsilon > 0$ finns ett tal $\delta > 0$ sådant att $|F(\mathbf{x}) - A| < \epsilon$ för alla $x \in (B(\mathbf{a}, \delta) \cap D_F) \setminus \{\mathbf{a}\}$.

En typisk situation illustreras nedan.



Exempel.⁵ Låt $F(x, y) := 4 + x^2 + y$. Det är uppenbart att $F(x, y) \rightarrow 4$ då $(x, y) \rightarrow 0$. Men vi introducerar ändå här en speciell teknik för att undersöka gränsvärden kring origo i planet, så att vi kan använda den i framtida, mindre triviala, problem. Tekniken går ut på att byta till *polära koordinater* (ρ, φ) i planet. Idén är att den polära koordinaten ρ precis är avståndet till origo, d.v.s. precis den storhet man talar om i gränsvärdesdefinitionen.

⁴ En någorlunda sympatisk introduktion till de precisa gränsvärdesdefinitionerna finns i författarens artikel om numeriska serier: <http://rejbrand.se/rejbrand/dokument/serier.pdf>.

⁵ Delar av exemplet kan eventuellt uppfattas som överkurs.

I polära koordinater är

$$F(\rho, \varphi) = 4 + \rho^2 \cos^2 \varphi + \rho \sin \varphi = 4 + \rho(\rho \cos \varphi + \sin \varphi)$$

så att avståndet

$$|F(\rho, \varphi) - 4| = \rho|\rho \cos \varphi + \sin \varphi| \leq \rho(1 + 1) = 2\rho$$

om $\rho \leq 1$. Välj nu något $\epsilon > 0$. Om vi väljer $\delta := \min(1, \epsilon/2)$ så har vi sålunda

$$\rho < \delta \quad \implies \quad |F(\rho, \varphi) - 4| \leq 2\rho < 2\delta \leq 2 \cdot \frac{\epsilon}{2} = \epsilon$$

d.v.s. $|F(x, y) - 4| < \epsilon$ för alla $(x, y) \in B(\mathbf{0}, \delta)$. Av definitionen av gränsvärde följer då att $F(x, y) \rightarrow 4$ då $(x, y) \rightarrow \mathbf{0}$.

I praktiken nöjer man sig vanligtvis med en något mindre detaljerad motivering, i stil med följande:

$$|F(\rho, \varphi) - 4| = \rho|\rho \cos \varphi + \sin \varphi| \stackrel{\rho \leq 1}{\leq} \rho(1 + 1) = 2\rho \rightarrow 0$$

då $\rho \rightarrow 0$. Alltså är $F(x, y) \rightarrow 4$ då $(x, y) \rightarrow \mathbf{0}$. Även följande, ännu mer kursiva, motivering är tänkbar:

$$F(\rho, \varphi) = 4 + \rho^2 \cos^2 \varphi + \rho \sin \varphi = 4 + \rho \underbrace{(\rho \cos \varphi + \sin \varphi)}_{\substack{\text{begränsad i en} \\ \text{omgivning av } \mathbf{0}}} \rightarrow 4$$

då $\rho \rightarrow 0$. Att parentesen är begränsad i någon omgivning av origo är viktigt. I vårt exempel gällde t.ex. att $|\rho \cos \varphi + \sin \varphi| \leq 2$ i $B(\mathbf{0}, 1)$, vilket var en avgörande ingrediens i vårt bevis. Att villkoret om begränsning är *tillräckligt* för att dra slutsats kan visas som följer: Antag att

$$F(x, y) = A + \rho B(x, y)$$

där B är begränsad i någon omgivning av origo. Det senare betyder att det finns tal $\delta' > 0$ och $M > 0$ sådana att $|B(x, y)| < M$ för alla $(x, y) \in B(\mathbf{0}, \delta')$. Men då gäller i denna omgivning att

$$|F(x, y) - A| = |\rho B(x, y)| = \rho|B(x, y)| < \rho M$$

varför det är uppenbart att vi givet $\epsilon > 0$ har $|F(x, y) - A| < \epsilon$ för alla $\rho < \delta := \min\left(\delta', \frac{\epsilon}{M}\right)$.

Exempel

Undersök gränsvärdet

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^3 + y^3}{x^2 + y^2}$$

Lösning: Vi har i polära koordinater

$$\frac{x^3 + y^3}{x^2 + y^2} = \frac{\rho^3 \cos^3 \varphi + \rho^3 \sin^3 \varphi}{\rho^2} = \rho \underbrace{(\cos^3 \varphi + \sin^3 \varphi)}_{\text{begränsad}} \rightarrow 0$$

då $\rho \rightarrow 0$.

$$\text{Svar: } \lim_{(x,y) \rightarrow (0,0)} \frac{x^3 + y^3}{x^2 + y^2} = 0.$$

Exempel

Undersök gränsvärdet

$$\lim_{(x,y) \rightarrow (0,0)} \frac{\sin x^2 + y^2}{x^2 + y^2}.$$

Lösning: Det är triviale att $x^2 + y^2 \rightarrow 0$ då $(x, y) \rightarrow (0, 0)$. Därför, enligt standardgränsvärde och sammansättningsregeln, är

$$\lim_{(x,y) \rightarrow (0,0)} \frac{\sin x^2 + y^2}{x^2 + y^2} = 1.$$

$$\text{Svar: } \lim_{(x,y) \rightarrow (0,0)} \frac{\sin x^2 + y^2}{x^2 + y^2} = 1.$$

I rummet fungerar sfäriska koordinater lika bra som polära koordinater i planet:

Exempel

Undersök gränsvärdet

$$\lim_{(x,y,z) \rightarrow (0,0,0)} \frac{x^3 + 3y^3 + 2z^4}{x^2 + y^2 + z^2}.$$

Lösning: Med sfäriska koordinater får vi

$$\begin{aligned} \frac{x^3 + 3y^3 + 2z^4}{x^2 + y^2 + z^2} &= \frac{r^3 \sin^3 \theta \cos^3 \varphi + 3r^3 \sin^3 \theta \sin^3 \varphi + 2r^4 \cos^4 \theta}{r^2} = \\ &= r \underbrace{(\sin^3 \theta \cos^3 \varphi + 3 \sin^3 \theta \sin^3 \varphi + 2r \cos^4 \theta)}_{\text{begränsad i en omgivning av origo}} \rightarrow 0 \end{aligned}$$

då $r \rightarrow 0$.

$$\text{Svar: } \lim_{(x,y,z) \rightarrow (0,0,0)} \frac{x^3 + 3y^3 + 2z^4}{x^2 + y^2 + z^2} = 0.$$

Exempel

Undersök gränsvärdet

$$\lim_{(x,y) \rightarrow (0,0)} \frac{2\sqrt{x^2 + y^2} + 3x^2}{\sqrt{x^2 + y^2}}.$$

Lösning: Med polära koordinater får vi

$$\frac{2\sqrt{x^2 + y^2} + 3x^2}{\sqrt{x^2 + y^2}} = \frac{2\rho + 3\rho^2 \cos^2 \varphi}{\rho} = 2 + 3\rho \underbrace{\cos^2 \varphi}_{\text{begr.}} \rightarrow 2$$

då $\rho \rightarrow 0$.

Svar: $\lim_{(x,y) \rightarrow (0,0)} \frac{2\sqrt{x^2+y^2}+3x^2}{\sqrt{x^2+y^2}} = 2.$

Det kan förstås inträffa att ett gränsvärde inte existerar. Oftast är det ganska lätt att motbevisa existensen av ett gränsvärde. Man studerar då några väl valda kurvor som passerar punkten i fråga och undersöker skalärfältets gränsvärden längs dessa kurvor (envariabelgränsvärden) när vi närmar oss punkten. Om vi lyckas hitta två sådana kurvor genom punkten längs vilka vi erhåller två *olika* gränsvärden (eller om någon kurva saknar gränsvärde) så kan vi dra slutsatsen att flervariabelfunktionen saknar gränsvärde i punkten.

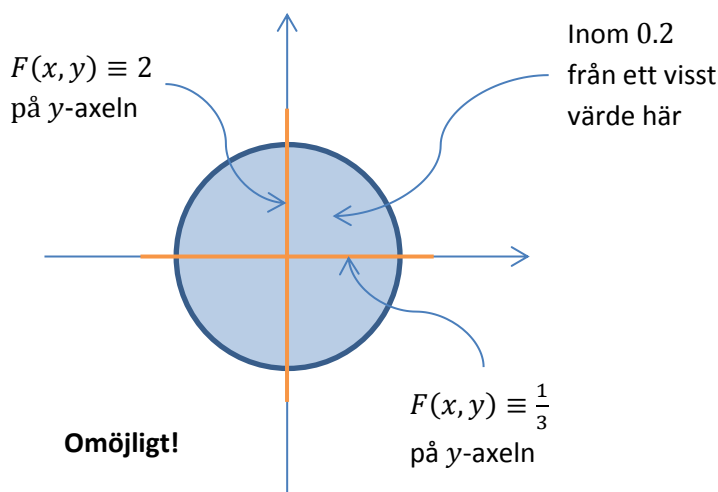
Exempel. Betrakta skalärfältet $F(x, y) = \frac{x^2+2y^2}{3x^2+y^2}$. Vi är intresserade av gränsvärdet i origo. Vi undersöker först kurvan $y = 0$, d.v.s. x -axeln. Denna kurva kan parametreras $x = t, y = 0$ och "temperaturen" vid tidpunkt t på vår resa längs x -axeln är

$$F(t) = \frac{t^2}{3t^2} = \frac{1}{3} \rightarrow \frac{1}{3}$$

då $t \rightarrow 0$. Om vi i stället går längs $x = 0$, d.v.s. längs y -axeln, så får vi vid tiden t "temperaturen"

$$F(t) = \frac{2t^2}{t^2} = 2 \rightarrow 2$$

då $t \rightarrow 0$. Alltså: *varje* omgivning av origo innehåller punkter där F är exakt $\frac{1}{3}$ och punkter där F är exakt 2. Detta motsäger existensen av ett gränsvärde, för om F skulle ha gränsvärdet A så skulle det finnas en omgivning kring origo sådan att $F(x, y)$ ligger mellan $A - 0.1$ och $A + 0.1$ (säg) i hela disken. Variationen av F är alltså mindre än 0.2 i den här omgivningen, men vi har ju nyss bevisat att variationen i *varje* omgivning av origo är minst $2 - \frac{1}{3} = \frac{5}{3}$.



Om läsaren *fortfarande* inte är övertygad om resonemanget här så kan han på egen hand markera punkterna $\frac{1}{3}$ och 2 på en tallinje och sedan försöka att rita ett *intervall* av längd 0.2 sådant att båda dessa punkter finns i intervallet.

Exempel

Undersök

$$\lim_{(x,y) \rightarrow (1,0)} \frac{x + 2y - 1}{x - y - 1}.$$

Lösning: Vi provar att gå längs kurvan $x = 1, y = t$. Detta ger funktionsvärdet

$$\frac{2t}{-t} = -2.$$

Vi provar att gå längs kurvan $x = t, y = 0$. Detta ger funktionsvärdet

$$\frac{t - 1}{t - 1} = 1.$$

Alltså antar skalärfältet både värdet -2 och värdet 1 i *varje* omgivning av $(1,0)$, så gränsvärde saknas.

Svar: Gränsvärde saknas.

Exempel

Undersök

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 + y}{x + y}.$$

Lösning: Vi provar att gå längs x -axeln ($x = t, y = 0$):

$$\frac{t^2}{t} = t \rightarrow 0$$

då $t \rightarrow 0$. Om det finns ett gränsvärde, måste det alltså vara 0.

Vi provar att gå längs y -axeln ($x = 0, y = t$):

$$\frac{t}{t} = 1.$$

Alltså saknas gränsvärde.

Svar: Gränsvärde saknas.

Exempel

Undersök

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 + xy + y^2}{x^2 + y^2}.$$

Lösning: Vi undersöker att gå längs x -axeln ($x = t, y = 0$):

$$\frac{t^2}{t^2} = 1.$$

Om det finns ett gränsvärde måste det alltså vara 1. Vi undersöker att gå längs y -axeln ($x = 0, y = t$):

$$\frac{t^2}{t^2} = 1.$$

Samma gränsvärde! Vi provar att gå längs $y = x$ ($x = t, y = t$):

$$\frac{t^2 + t^2 + t^2}{t^2 + t^2} = \frac{3t^2}{2t^2} = \frac{3}{2}.$$

Tydligen saknas gränsvärde.

Svar: Gränsvärde saknas.

Naturligtvis kan stryka en av undersökningarna som gav gränsvärdet 1. Bevisvärdet blir fullständigt ändå. Men på t.ex. en skrivning är det lätt hänt att man inte orkar "sudda ut" tidigare undersökningar. Men varför lät då artikelförfattaren den onödiga undersökningen stå kvar i det här dokumentet? Anledningen är rent pedagogisk. Författaren vill illustrera det faktum att man ibland får samma gränsvärde längs de två första kurvorna man undersöker, och då måste man (om man fortfarande tror att gränsvärde saknas) undersöka en tredje kurva. Om även denna ger samma gränsvärde måste man (om man fortfarande tror att gränsvärde saknas) undersöka en fjärde kurva, och så vidare.

Om man ägnat det senaste decenniet åt att undersöka olika kurvor och fått samma gränsvärde längs varje kurva, så bör man börja fundera på om det kanske rent av är så att gränsvärdet existerar. Då får man i stället försöka bevisa detta, t.ex. genom att införa polära koordinater.

Exempel

Undersök gränsvärdet

$$\lim_{(x,y) \rightarrow (0,0)} \frac{x^4 y^2}{(x^4 + y^2)^2}.$$

Lösning: Det är klart att skalärfältets värde är identiskt 0 både på x -axeln och på y -axeln (utanför origo). Så om gränsvärdet existerar så måste det vara lika med 0. Vi provat att gå längs en *annan* rät linje genom origo, d.v.s. längs $y = kx$ för något $k \in \mathbb{R} \setminus \{0\}$. Detta ger

$$\frac{k^2 x^6}{(x^4 + k^2 x^2)^2} = \frac{k^2 x^6}{x^8 + 2k^2 x^6 + k^4 x^4} = x^2 \frac{k^2}{x^4 + 2k^2 x^2 + k^4} \rightarrow 0$$

då $x \rightarrow 0$. Återigen samma gränsvärde!⁶

Vi provar nu att gå längs parabeln $y = x^2$. Detta ger

$$\frac{x^8}{(x^4 + x^4)^2} = \frac{x^8}{4x^8} = \frac{1}{4}.$$

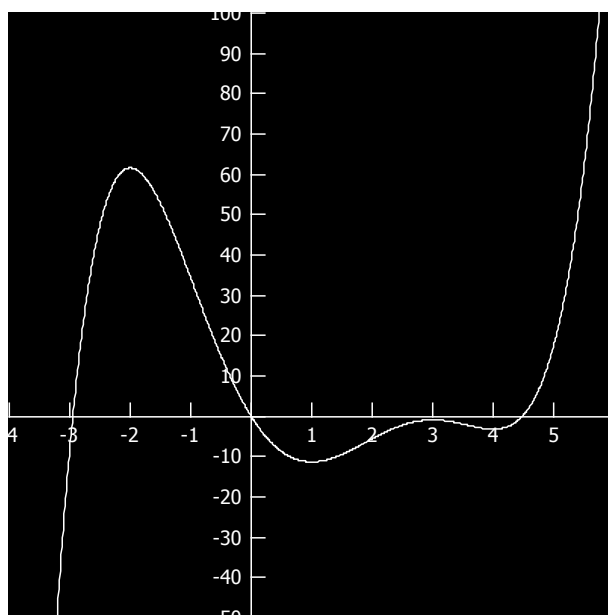
Alltså saknas gränsvärde.

Svar: Gränsvärde saknas.

7.5 Lokala max- och minpunkter

7.5.1 Repetition av envariabelfallet

Vi börjar med att repetera envariabelfallet, och dessutom situationen där funktionen är definierad på *hela* den reella axeln. Betrakta exemplet $f(x) = \frac{1}{5}x^5 - \frac{3}{2}x^4 + x^3 + 13x^2 - 24x$; dess graf visas nedan.



I bilden ser vi att f har ett "lokalt maximum" i -2 och ett i $+3$. Vi ser också att vi har ett "lokalt minimum" i $+1$ och ett i $+4$. Men vad exakt betyder dessa begrepp? Finns det någon precis formulering av begreppen, som preciserar den intuitiva idén om vad ett "lokalt extremvärde" är för något?

Svaret är "ja", och det är faktiskt väldigt enkelt. Läsaren som kommer ihåg sin envariabelanalys minns följande:

⁶ Om vi tolkar $F(x, y)$ som temperaturen i punkten (x, y) så har vi alltså situationen att temperaturen närmar sig 0°C längs *varje* raksträcka genom origo! Innebär inte detta att F *måste* ha gränsvärdet noll? Nej, tydligen inte, som vi kommer att se.

Definition (påminnelse). En funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ säges ha ett *lokalt maximum* i punkten $a \in \mathbb{R}$ omm det finns någon omgivning U av a sådan att $f(x) \leq f(a)$ för alla $x \in U$.

Om man påminner sig om vad begreppet "omgivning" betyder inser man att detta betyder precis följande:

Definition (påminnelse). En funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ säges ha ett *lokalt maximum* i punkten $a \in \mathbb{R}$ omm det finns ett tal $\epsilon > 0$ sådant att $f(x) \leq f(a)$ för alla $x \in]a - \epsilon, a + \epsilon[$.

Begreppet "lokalt minimum" definieras analogt. Själva punkten a kallas för en *lokal maximipunkt* (resp. *lokal minimipunkt*), och funktionsvärdet $f(a)$ kallas för ett *lokalt maximivärde* (resp. *lokalt minimivärde*). En punkt a som är en lokal maximipunkt eller en lokal minimipunkt kallas också för en *lokal extrempunkt*, och funktionsvärdet för ett *lokalt extremvärde*.

Vi talar dessutom om ett *strängt lokalt maximum* om vi har situationen " $f(x) < f(a)$ " i definitionen ovan.

Observera att ett lokalt extremvärde inte måste vara funktionens största eller minsta värde. Det rör sig bara om det största eller minsta värdet i *någon* omgivning av punkten, därav attributet "lokalt".

Övningar

1. Förvissa dig om att dessa precisa formuleringar passar den intuitiva uppfattningen om vad en "lokal extrempunkt" är för något. Titta på exempelfunktionen f ovan, och/eller hitta på egna exempel.
2. Inse att samtliga lokala extrempunkter i exemplet f ovan är *stränga* lokala extrempunkter.
3. Ge exempel på en funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ och en punkt $a \in \mathbb{R}$ sådan att a är ett lokalt maximum till g , men *inte* ett strängt lokalt maximum till g .
4. Ge exempel på en funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ och en punkt $a \in \mathbb{R}$ sådan att a är både ett lokalt maximum och ett lokalt minimum till g .
5. Visa att det inte finns någon funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ som har en punkt $a \in \mathbb{R}$ som både är ett *strängt* lokalt maximum och ett *strängt* lokalt minimum.
6. Ge ett enkelt exempel på en funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ som har oändligt många stränga lokala maximipunkter och oändligt många stränga lokala minimipunkter.
7. Ge ett enkelt exempel på en funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ som saknar lokala extrempunkter.
8. Definiera begreppen *globalt* maximum och *globalt* minimum (på ett "uppenbara" sättet).
9. Visa att ett globalt maximum nödvändigtvis är ett lokalt maximum, men att omvändningen inte gäller i allmänhet.

En mycket viktig insikt är följande:

Sats. Om funktionen $f: \mathbb{R} \rightarrow \mathbb{R}$ är deriverbar och $a \in \mathbb{R}$ är ett lokalt extremvärde, så är a en stationär punkt (=en punkt där derivatan är noll).

Detta är intuitivt uppenbart, ty om derivatan *inte* är noll så är den nollskild, t.ex. lika med 3. Det betyder att grafens lutning är 3 i punkten, så går man lite till höger får man större värden, och går man lite till vänster får man mindre värden. Alltså var punkten varken ett max eller ett min. (Här använde vi ett kontrapositivt resonemang.)

Satsen ger alltså ett *nödvändigt* villkor (derivatan lika med noll) för att en punkt skall vara en lokal extrempunkt, men det är INTE ett *tillräckligt* villkor. Det är alltså INTE så att en stationär punkt (en punkt där derivatan är noll) nödvändigtvis är ett lokalt extremvärde. Ett enkelt motexempel är funktionen $f(x) = x^3$ i origo. Här är derivatan noll, men origo är varken ett lokalt minimum eller ett lokalt maximum. (I stället säger vi att origo är en *terrasspunkt*.)

Övningar

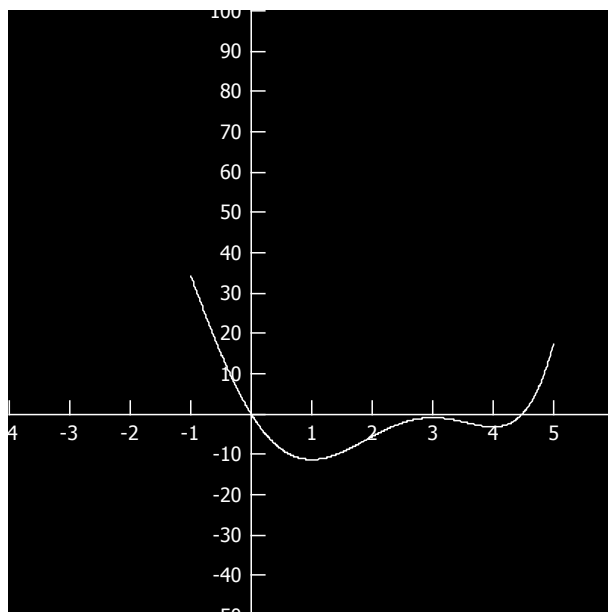
1. Påminn dig själv om hur derivatans teckenväxling kan användas för att avgöra om en stationär punkt är ett lokalt maximum eller minimum (eller ingetdera).
2. Påminn dig själv om hur *andradderivatan* i en stationär punkt i vissa fall kan användas för att avgöra om punkten är ett lokalt maximum eller minimum.
3. Undersök de två föregående frågorna i de konkreta exemplen $f(x) = x^2, -x^2, x^3, x^4$ och $-x^4$.

7.5.1.1 Funktioner definierade på äkta delmängder av \mathbb{R}

Notera att begreppen ovan endast definierades för funktioner $\mathbb{R} \rightarrow \mathbb{R}$, d.v.s. funktioner definierade på *hela* \mathbb{R} . För en funktion definierad bara på en delmängd av \mathbb{R} gör man analoga definitioner. Till exempel säger vi att funktionen $f: D_f \rightarrow \mathbb{R}$ (med $D_f \subset \mathbb{R}$) har ett lokalt maximum i punkten $a \in D_f$ om det finns ett tal $\epsilon > 0$ sådant att $f(x) \leq f(a)$ för alla $x \in]a - \epsilon, a + \epsilon[\cap D_f$.

Satsen om derivatan måste också förändras lite: om f har en lokal extrempunkt i en *inre punkt* i D_f så är derivatan noll i punkten. Betrakta t.ex. restriktionen av vår exempelfunktion till intervallet $[-1, 5]$, d.v.s. funktionen $f: [-1, 5] \rightarrow \mathbb{R}$ definierad av

$$f(x) = \frac{1}{5}x^5 - \frac{3}{2}x^4 + x^3 + 13x^2 - 24x, \quad \forall x \in [-1, 5].$$



Det är klart att båda ändpunkterna ($x = -1$ samt $x = 5$) nu är lokala maximipunkter, men derivatan är uppenbarligen inte noll här. I de *inre* extrempunkterna är derivatan däremot noll.

För att sammanfatta har vi

Sats. Om funktionen $f: D_f \rightarrow \mathbb{R}$, $D_f \subset \mathbb{R}$, är deriverbar och a är en *inre* punkt till D_f så gäller

$$a \text{ är en extrempunkt} \implies a \text{ är en stationär punkt.}$$

Detta är intuitivt uppenbart, ty om derivatan *inte* är noll så är den nollskild, t.ex. lika med 3. Det betyder att grafens lutning är 3 i punkten, så går man lite till höger får man större värden, och går man lite till vänster får man mindre värden. *Och eftersom punkten är en inre punkt så kan man gå både åt höger och åt vänster utan att lämna mängden!* Alltså var punkten varken ett max eller ett min.

7.5.2 Flervariabelfallet

Det är naturligt att göra följande definitioner:

Definition. Låt $F: D_F \rightarrow \mathbb{R}$, $D_F \subset \mathbb{R}^n$, vara ett skalärfält och låt $\mathbf{a} \in D_F$. Vi säger att \mathbf{a} är ett lokalt maximum till F om det finns en omgivning U av \mathbf{a} sådan att $F(\mathbf{x}) \leq F(\mathbf{a})$ för alla $\mathbf{x} \in U \cap D_F$.

Om man kommer ihåg vad begreppet "omgivning" betyder inser man att definitionen också kan skrivas så här:

Definition. Låt $F: D_F \rightarrow \mathbb{R}$, $D_F \subset \mathbb{R}^n$, vara ett skalärfält och låt $\mathbf{a} \in D_F$. Vi säger att \mathbf{a} är ett lokalt maximum till F om det finns ett tal $\epsilon > 0$ sådant att $F(\mathbf{x}) \leq F(\mathbf{a})$ för alla $\mathbf{x} \in B(\mathbf{a}, \epsilon) \cap D_F$.

I planet är omgivningen en öppen disk, i rummet ett öppet klot o.s.v.

På analogt sätt definieras begreppet "lokalt minimum", och vi definierar de "stränga" varianterna på de uppenbart sättet, liksom "maximipunkt", "maximivärde", "minimipunkt", "minivärde", "extrempunkt" och "extremvärde".

Övning

1. Vi diskuterade i början av kapitlet (på sidan 36) tre olika sätt att visualisera skalärfält av typen $\mathbb{R}^2 \rightarrow \mathbb{R}$. Hur identifierar man med hjälp av dessa metoder (med viss osäkerhet, som alltid när det gäller grafiska/numeriska metoder) lokala extrempunkter?

Precis som i envariabelfallet har vi ett enkelt *nödvändigt* krav för att en punkt skall vara en lokal extrempunkt:

Sats. Om skalärfältet $F: \mathbb{R}^n \rightarrow \mathbb{R}$ är partiellt deriverbart och $\mathbf{a} \in \mathbb{R}^n$ är en lokal extrempunkt så är $\nabla F(\mathbf{a}) = \mathbf{0}$.

En punkt där gradienten är noll kallas för en *stationär* punkt till skalärfältet. Satsen är (som vanligt!) intuitivt uppenbar, ty om gradienten *inte* är noll så är den en nollskild vektor, och därmed är skalärfältets riktningsderivata nollskild i denna riktning. I gradientens riktning ökar funktionsvärdet, och i motsatt riktning minskar funktionsvärdet (lika snabbt). Alltså var punkten varken ett minimum eller ett maximum.

Omvändningen av satsen behöver inte gälla – en stationär punkt behöver inte vara en extrempunkt. Vi kommer strax att ge exempel på detta.

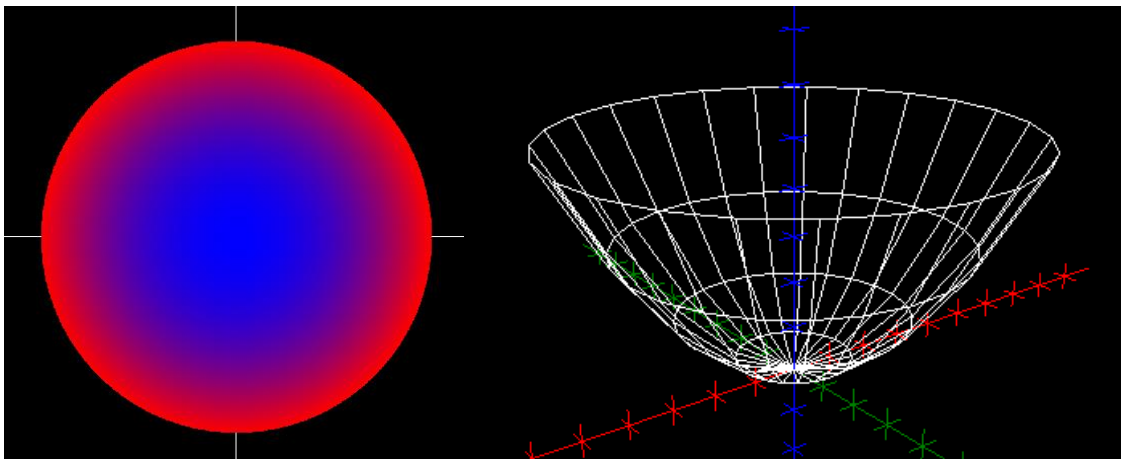
Vi studerar också ibland skalärfält definierade på områden som är äkta delmängder av \mathbb{R}^n , och i sådana fall måste vi i satsen – precis som i envariabelfallet – lägga till villkoret att punkten är en *inre* punkt:

Sats. Om skalärfältet $F: D_F \rightarrow \mathbb{R}$, $D_F \subset \mathbb{R}^n$, är partiellt deriverbart och \mathbf{a} är en *inre* punkt till D_F så gäller

$$\mathbf{a} \text{ är en extrempunkt} \implies \mathbf{a} \text{ är en stationär punkt.}$$

Detta är intuitivt uppenbart. För om gradienten är nollskild så har vi en positiv riktningsderivata i en riktning och en lika stor fast negativ riktningsderivata i motsatta riktningen. Och eftersom punkten är en *inre* punkt så **kan** vi gå i båda dessa riktningar utan att lämna mängden. Alltså var punkten varken en minimipunkt eller en maximipunkt.

Exempel. Låt $F: D \rightarrow \mathbb{R}$, där D är den fyllda enhetsdisken i xy -planet, vara definierad av $F(x, y) = x^2 + y^2$. Nedan visas dels ett färgat plan (blått: lägst, 0; rött: högst, 1), dels grafen till F .



Det är klart att $(0,0)$ är ett lokalt minimum – i själva verket gäller $F(0,0) \leq F(x, y)$ för alla $(x, y) \in D$, så origo är funktionens *globala* minimum. I origo, som är en *inre* punkt, är också mycket riktigt gradienten $(2x, 2y)$ lika med nollvektorn. Vidare är varje punkt på enhetscirkeln ett lokalt maximum, trots att gradienten inte är noll där. Men enhetscirkeln är ju också randen till D .

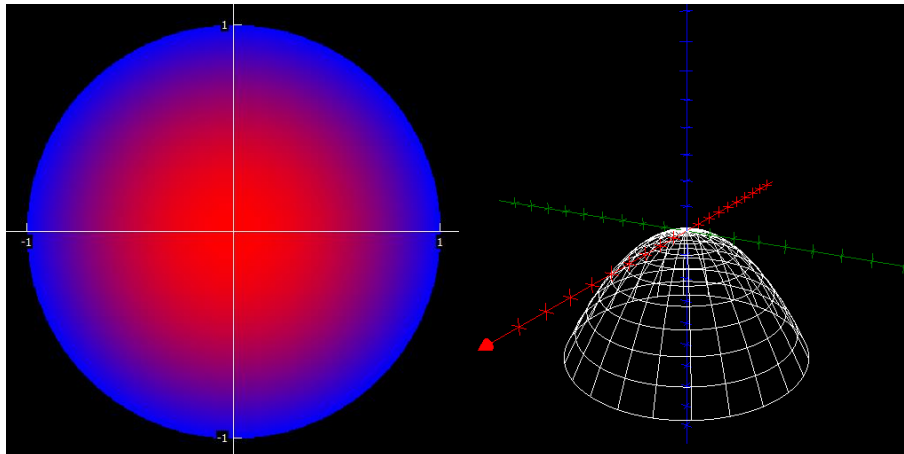
Punkten $(0.5, 0.5)$ är inte en lokal extrempunkt. Här är gradienten $(1, 1)$ så vi har en positiv riktningsderivata rakt utåt och en lika stor negativ riktningsderivata rakt inåt. Eftersom punkten är en *inre* punkt finns det en omgivning kring punkten vilken helt ligger i D . Och denna omgivning innehåller naturligtvis punkter både i gradientens riktning och i motsatta riktningen, d.v.s. både "varmare" och "kallare" punkter, så $(0.5, 0.5)$ är varken en lokal maxpunkt eller en lokal minpunkt.

7.5.2.1 Olika typer av stationära punkter

Låt $F: D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}^n$, vara ett skalärfält. En punkt där gradienten är noll kallas som nämnt för en *stationär* punkt. En sådan *kan* vara en lokal extrempunkt (bland de *inre* punkterna till D är det som bekant *bara* de stationära punkterna som *kan* vara extrempunkter), men den *måste* inte vara det. En stationär punkt som *inte* är en extrempunkt kallas oftast för en *sadelpunkt*, och motsvarar alltså begreppet *terrasspunkt* i envariabelanalysen. Men till skillnad från envariabelfallet finns det i flervariabelfallet flera olika *typer* av sadelpunkter. I avsnitten nedan ger vi exempel på de vanligaste typerna av stationära punkter.

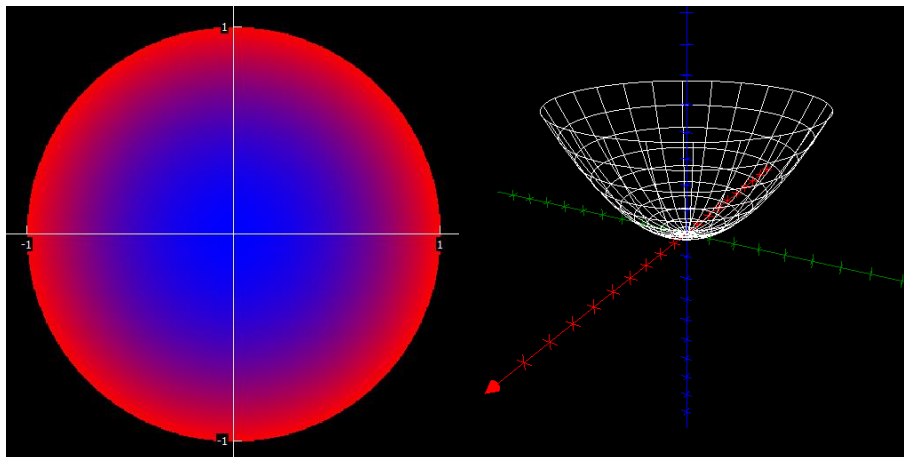
7.5.2.1.1 Lokalt maximum

Funktionen $F(x, y) = -x^2 - y^2$ har ett lokalt maximum i origo.



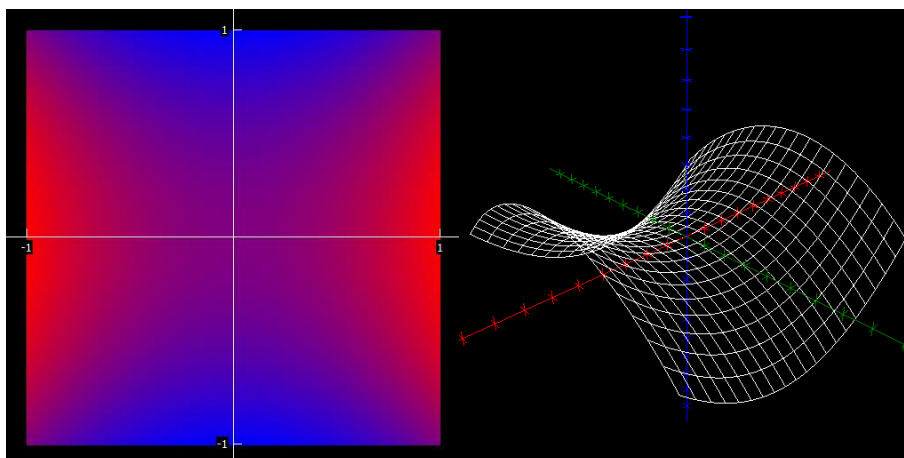
7.5.2.1.2 Lokalt minimum

Funktionen $F(x, y) = x^2 + y^2$ har ett lokalt minimum i origo.



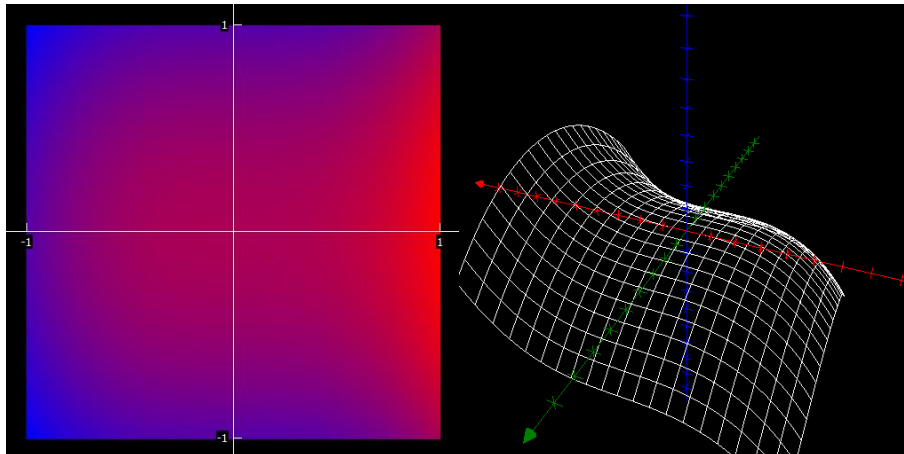
7.5.2.1.3 Typisk sadel

Funktionen $F(x, y) = x^2 - y^2$ har en stationär punkt i origo, fast det är *inte* en extrempunkt. I stället är det den mest typiska sortens sadelpunkt.



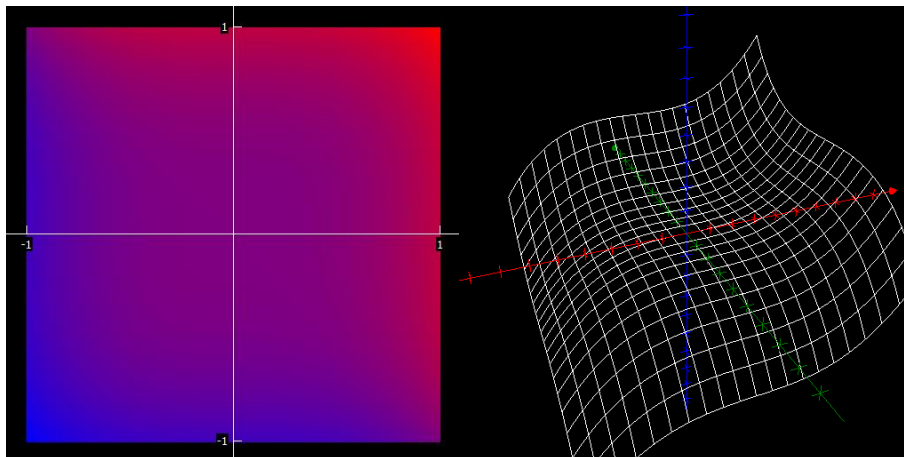
7.5.2.1.4 Farlig sadel

Funktionen $F(x, y) = x^3 - y^2$ har en annan sorts sadelpunkt i origo.



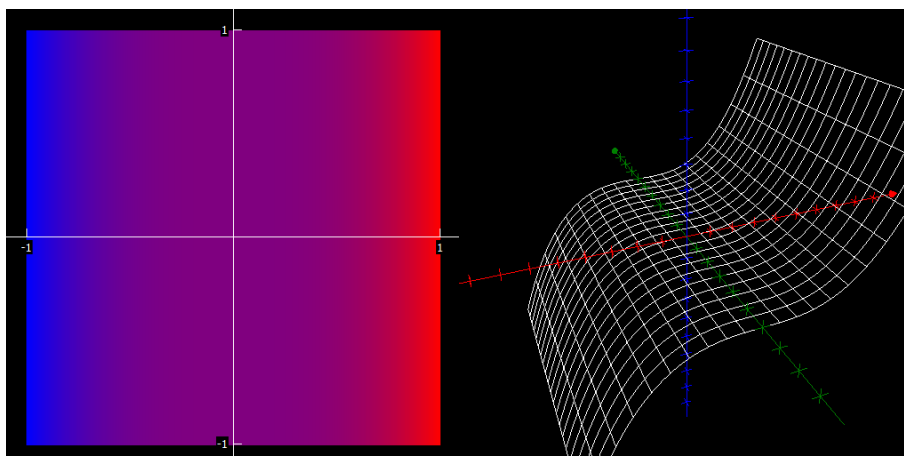
7.5.2.1.5 Fåtölj?

Funktionen $F(x, y) = x^3 - y^3$ har en tredje sorts sadelpunkt i origo.



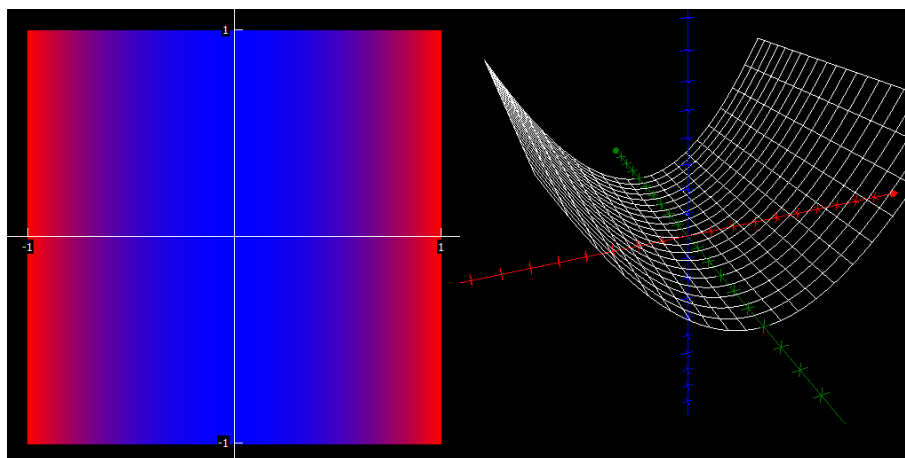
7.5.2.1.6 Terrass

Funktionen $F(x, y) = x^3$ ser ut som en terrass i landskapet (eller en obekväm parkbänk). Notera att grafen är en cylinder. Varje punkt på y -axeln är en stationär punkt.



7.5.2.1.7 Dike/ås

Funktionen $F(x, y) = x^2$ ser ut som ett dike i landskapet. (Och funktionen $F(x, y) = -x^2$ ser ut som en ås/avlång kulle.) Notera att grafen är en cylinder. Varje punkt på y -axeln är en stationär punkt.



Övningar

1. Se till att du förstår varför graferna ser ut som de gör. Undersök och beskriv, som alltid, snitten $x = \text{konst.}$, $y = \text{konst.}$ och $z = \text{konst.}$
2. Räkna ut skalärfältens gradienter och verifiera att de stationära punkterna är där de ska.
3. Notera att tangentplanet i en stationär punkt alltid är horisontellt. Bevisa också detta.
4. *Bestäm alla stationära punkter till funktionen $F(x, y) = \sin x \sin y$ som vi betraktade i avsnittet om sammansättningar av funktioner. Vilka av dessa är lokala maxima och minima?

7.5.2.2 Max och min på delmängder

I praktiken behöver man ibland hitta största och minsta värdet av ett skalärfält $F: D_F \rightarrow \mathbb{R}$, $D_F \subset \mathbb{R}^n$, på en *delmängd* av F 's definitionsmängd. Till exempel kan man vilja hitta största och lägsta värdet av $F(x, y)$ på en *kurva* i xy -planet. Om $F(x, y)$ är temperaturen i punkten (x, y) i planet, och kurvan parametreras av $\mathbf{r}: \mathbb{R} \rightarrow \mathbb{R}^2$, så är sammansättningen $F(\mathbf{r}(t))$ temperaturen vid tiden t längs kurvan. I det här avsnittet diskuterar vi sambandet mellan de lokala extremvärdena hos F och de lokala extremvärdena hos "temperaturprofilen" $F \circ \mathbf{r}$.

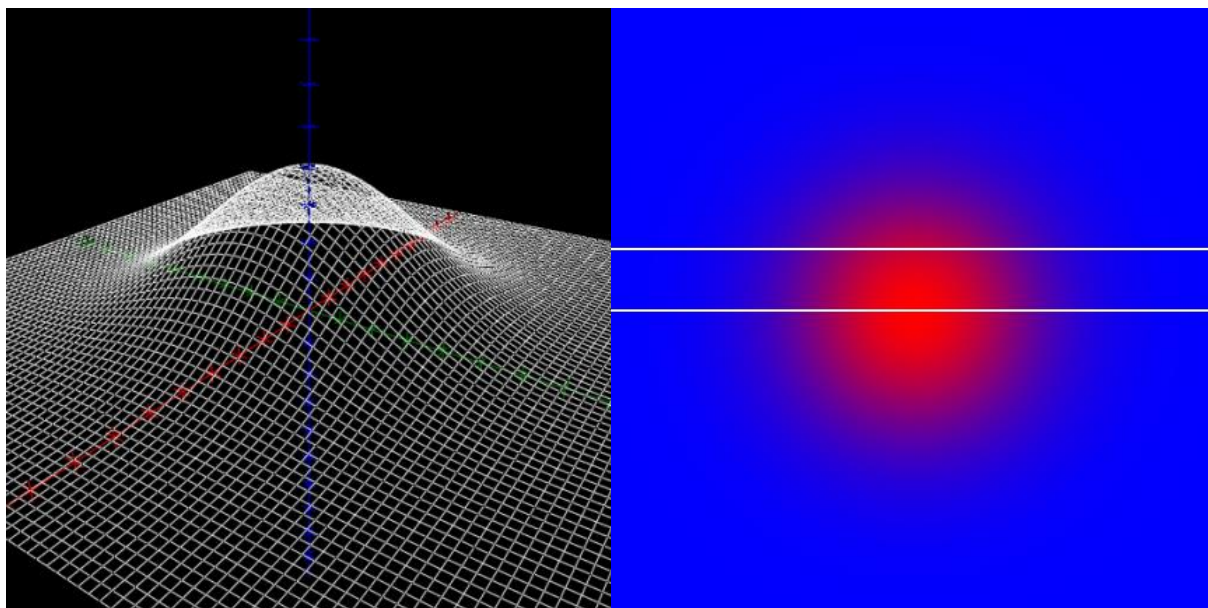
Vi ger ett exempel. Betrakta "kullen"

$$z = 4 e^{-\frac{x^2+y^2}{16}}.$$

Kullen är grafen till skalärfältet $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ definierat av $F(x, y) = 4 e^{-\frac{x^2+y^2}{16}}$. $F(x, y)$ tolkas alltså som *höjden över havet* vid punkten (x, y) . Betrakta också följande kurvor:

$$\begin{aligned} \Gamma_1 &:= \mathbf{r}_1(\mathbb{R}), & \mathbf{r}_1(t) &:= (t, 0) \\ \Gamma_2 &:= \mathbf{r}_2(\mathbb{R}), & \mathbf{r}_2(t) &:= (t, 2). \end{aligned}$$

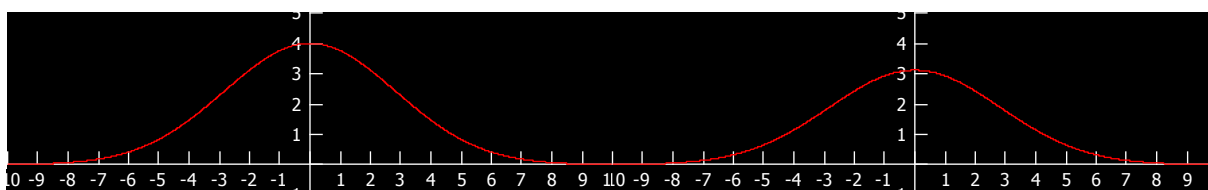
Nedan, till vänster, visas grafen till F , d.v.s. själva "kullen". Till höger visas ett färgat plan över F tillsammans med de två "cykelbanorna".



De sammansatta funktionerna $F \circ \mathbf{r}_1$ och $F \circ \mathbf{r}_2: \mathbb{R} \rightarrow \mathbb{R}$, vilka ger höjden över havet vid tidpunkt t längs respektive cykelbana, har följande formler och grafer:

$$F(\mathbf{r}_1(t)) = 4 e^{-t^2/16}$$

$$F(\mathbf{r}_2(t)) = 4 e^{-(t^2+4)/16}$$



Skalärfältet F har förstås ett lokalt maximum i origo, så det är inte konstigt att *varje* kurva som passerar genom origo får ett lokalt maximum i motsvarande tidpunkt. Detta exemplifieras i grafen till $F \circ \mathbf{r}_1$, där $t = 0$ ger positionen $(0, 0)$, d.v.s. mitt på berget.

Titta nu på den andra cykelbanan, $\mathbf{r}_2(\mathbb{R})$. Höjdprofilen $F \circ \mathbf{r}_2$ har tydligen också ett lokalt maximum i $t = 0$, vilket svarar mot punkten $(0, 2)$ på bergets sluttning. Men $(0, 2)$ är ju inte en bergstopp!

Följande slutsats är **mycket** viktig:

Observation. Varje cykelbana som passerar en bergstopp har ett lokalt maximum i bergstoppen. Men bara för att en cykelbana har ett lokalt maximum i någon punkt, så måste inte punkten vara en bergstopp!

Samma observation gäller – självfallet – också för lokala minimum.

7.6 Lokala undersökningar – det lättaste fallet

Vi betraktar i det här avsnittet skalärfält som ges av andragradspolynom, och undersöker om origo är en stationär punkt, och i sådana fall vilken *sorts* stationär punkt det rör sig om. Det visar sig nämligen att samma principiella undersökningsmetod kommer att fungera för allmänna skalärfält och allmänna punkter (inte bara origo).

För enkelhets skull begränsar vi oss till plana skalärfält. Det allmänna fallet fungerar på samma sätt. Vi betraktar alltså en funktion $F: \mathbb{R}^2 \rightarrow \mathbb{R}$ definierad av

$$F(x, y) = C + hx + ky + Kx^2 + Lxy + My^2.$$

Vi ser direkt att $F(0,0) = C$ samt att $\nabla F(0,0) = (h, k)$, så vi kan skriva

$$F(x, y) = F(0,0) + \nabla F(0,0) \cdot (x, y) + Kx^2 + Lxy + My^2.$$

Origo är tydligen en stationär punkt om $h = k = 0$, så vi intresseras oss endast för skalärfält på formen

$$F(x, y) = F(0,0) + Kx^2 + Lxy + My^2.$$

Detta kan skrivas

$$F(x, y) - F(0,0) = Kx^2 + Lxy + My^2$$

där högerledet är en *kvadratisk form*. Högerledet är förstas noll i origo, $x = y = 0$. Beroende på talen K, L och M kan för övrigt fem olika fall inträffa:

- Om $Kx^2 + Lxy + My^2 > 0$ utanför origo så kallas den kvadratiske formen för *positivt definit*. I detta fall är tydligen $F(x, y) - F(0,0) > 0$ överallt förutom i origo, d.v.s. $F(x, y) > F(0,0)$, och vi ser direkt att origo är ett (strängt) lokalt minimum.
- Om $Kx^2 + Lxy + My^2 < 0$ utanför origo så kallas den kvadratiske formen för *negativt definit*. I detta fall är tydligen $F(x, y) - F(0,0) < 0$ överallt förutom i origo, d.v.s. $F(x, y) < F(0,0)$, och vi ser direkt att origo är ett (strängt) lokalt maximum.
- Om $Kx^2 + Lxy + My^2 \geq 0$ överallt men likhet gäller även någonstans utanför origo, så kallas den kvadratiske formen för *positivt semidefinit*. I sådana fall finns i *varje* omgivning av origo punkter (x, y) sådana att $F(x, y) = F(0,0)$. Funktionen har ett lokalt minimum i origo.
- Om $Kx^2 + Lxy + My^2 \leq 0$ överallt men likhet gäller även någonstans utanför origo, så kallas den kvadratiske formen för *negativt semidefinit*. I sådana fall finns i *varje* omgivning av origo punkter (x, y) sådana att $F(x, y) = F(0,0)$. Funktionen har ett lokalt maximum i origo.
- Om $Kx^2 + Lxy + My^2$ antar både positiva och negativa värden kallas den kvadratiske formen för *indefinit*. I sådana fall finns i *varje* omgivning av origo punkter (x, y) sådana att $F(x, y) > F(0,0)$ samt punkter (x, y) sådana att $F(x, y) < F(0,0)$. Alltså är origo varken ett lokalt maximum eller ett lokalt minimum.

Mycket viktig övning

1. Förklara de tre sista punkterna, d.v.s. givet existensen av *en* punkt $(x, y) \in \mathbb{R}^2$ sådan att $Kx^2 + Lxy + My^2 > 0$ [eller < 0 eller $= 0$], hur vet man att *varje* omgivning av origo innehåller en sådan punkt? [Ledning: Låt (a, b) vara en sådan punkt, så $a^2 + Lab + Mb^2 > 0$, säg. Vad gäller för punkten $t(a, b)$?]

Genom att undersöka den kvadratiske formen $Kx^2 + Lxy + My^2$ kan man alltså i direkt se vad för sorts stationär punkt origo är. För att undersöka den kvadratiske formen kan man kvadratkomplettera, eller använda metoder från linjär algebra (spektralteori).

Exempel

Undersök om origo är en stationär punkt till funktionen $F(x, y) = 3 + x + x^2 + y^2$, och ange i sådana fall vad för sorts (maximum, minimum eller sadelpunkt).

Lösning: Eftersom $\nabla F(0,0) \neq \mathbf{0}$ är origo inte en stationär punkt.

Exempel. Undersök om origo är en stationär punkt till funktionen $F(x, y) = 3 + x^2 + xy + y^2$, och ange i sådana fall vad för sorts (maximum, minimum eller sadelpunkt).

Lösning: Origo är en stationär punkt eftersom $\nabla F(0,0) = \mathbf{0}$. Den kvadratiske formen

$$x^2 + xy + y^2 = \left(x + \frac{1}{2}y\right)^2 + \frac{3}{4}y^2 \geq 0$$

och likhet råder endast i origo, ty om likhet råder måste först $y = 0$, och sedan $x + \frac{1}{2}y = 0$ som ger $x = 0$, d.v.s. vi är i origo. Den kvadratiske formen är därför positivt definit så att origo är ett (strängt) lokalt minimum för F .

Svar: Origo är ett strängt lokalt minimum för F .

Exempel

Undersök om origo är en stationär punkt till funktionen $F(x, y) = 1 + x^2 + 3xy + y^2$, och ange i sådana fall vad för sorts (maximum, minimum eller sadelpunkt).

Lösning: Origo är en stationär punkt eftersom $\nabla F(0,0) = \mathbf{0}$. Den kvadratiske formen

$$x^2 + 3xy + y^2 = \left(x + \frac{3}{2}y\right)^2 - \frac{5}{4}y^2$$

är indefinit (t.ex. positiv då $x = 1, y = 0$ samt negativ då $x = -3/2, y = 1$). Därför är origo en sadelpunkt till F .

Svar: Origo är en sadelpunkt till F .

Exempel

Undersök om origo är en stationär punkt till funktionen $F(x, y) = 2x^2 - 5xy + y^2$, och ange i sådana fall vad för sorts (maximum, minimum eller sadelpunkt).

Lösning: Origo är en stationär punkt eftersom $\nabla F(0,0) = \mathbf{0}$. Den kvadratiske formen

$$2x^2 - 5xy + y^2 = 2 \left[x^2 - \frac{5}{2}xy + \frac{1}{2}y^2 \right] = 2 \left[\left(x - \frac{5}{4}y\right)^2 - \frac{17}{16}y^2 \right]$$

som är indefinit. Därför är origo en sadelpunkt till F .

Svar: Origo är en sadelpunkt till F .

Exempel

Undersök om origo är en stationär punkt till funktionen $F(x, y) = x^2 + 2xy + y^2$, och ange i sådana fall vad för sorts (maximum, minimum eller sadelpunkt).

Lösning: Eftersom $F(x, y) = (x + y)^2$ ser vi direkt att origo är ett lokalt minimum (men inte ett strängt sådant).

Övning

1. Beskriv eller skissa den paraboliska cylindern $z = x^2 + 2xy + y^2$ från föregående exempel.

Vi ger också ett par exempel med skalärfält i rummet.

Exempel

Undersök om origo är en stationär punkt till funktionen $F(x, y, z) = 2 + x^2 + y^2 + 2z^2 + xy + yz$, och ange i sådana fall vad för sorts (maximum, minimum eller sadelpunkt).

Lösning: Origo är en stationär punkt eftersom $\nabla F(0,0) = \mathbf{0}$. Den kvadratiske formen

$$\begin{aligned}x^2 + y^2 + 2z^2 + xy + yz &= \left(x + \frac{1}{2}y\right)^2 + \frac{3}{4}y^2 + 2z^2 + yz = \left(x + \frac{1}{2}y\right)^2 + \frac{3}{4}\left[y^2 + \frac{8}{3}z^2 + \frac{4}{3}yz\right] = \\ &= \left(x + \frac{1}{2}y\right)^2 + \frac{3}{4}\left[\left(y + \frac{2}{3}z\right)^2 + \frac{20}{9}z^2\right] = \left(x + \frac{1}{2}y\right)^2 + \frac{3}{4}\left(y + \frac{2}{3}z\right)^2 + \frac{5}{3}z^2\end{aligned}$$

är positivt definit (eftersom den är alltid ≥ 0 och likhet kräver $z = 0$, $y + \frac{2}{3}z = 0$ och $x + \frac{1}{2}y = 0$ vilket ger $y = 0$ och $x = 0$, d.v.s. $x = y = z = 0$). Origo är därför ett strängt lokalt minimum för F .

Svar: Origo är ett strängt lokalt minimum för F .

Exempel

Undersök om origo är en stationär punkt till funktionen $F(x, y, z) = 2 + x^2 - y^2 + z^2 - 2xz$, och ange i sådana fall vad för sorts (maximum, minimum eller sadelpunkt).

Lösning: Origo är en stationär punkt eftersom $\nabla F(0,0) = \mathbf{0}$. Den kvadratiske formen

$$x^2 - y^2 + z^2 - 2xz = (x - z)^2 - y^2$$

är indefinit (negativ t.ex. för $(x, y, z) = (0, 1, 0)$ och positiv t.ex. för $(x, y, z) = (1, 0, 0)$) så origo är en sadelpunkt för F .

Svar: Origo är en sadelpunkt för F .

7.7 Lokala undersökningar – det allmänna fallet

I föregående avsnitt avgjorde vi karaktären hos en stationär punkt till ett skalärfält F i specialfallet att punkten är origo och F ges av ett polynom av grad 2 i sina variabler. I det här avsnittet studerar vi det allmänna fallet, d.v.s. vi har ett *godtyckligt* skalärfält F och en *godtycklig* stationär punkt, och vi skall undersöka denna.

I princip kommer vi att överföra analysen till det enkla fallet i föregående avsnitt genom att *approximera* funktionen med ett andragradspolynom nära den stationära punkten, och sedan studera approximationen med de metoder vi använde i förra avsnittet. Detta kommer att ge rätt resultat (det är något som måste bevisas!) i samtliga fall förutom de två *semidefinita* fallen. I de fallen måste man alltså undersöka den stationära punkten på något annat sätt.

Den approximation vi använder är andra ordningens Taylorpolynom. För att kunna formulera approximationen måste vi först känna till begreppet *Hessian*. Om $F: \mathbb{R}^n \rightarrow \mathbb{R}$ är ett skalärfält så är Hessianen $n \times n$ -matrisen med alla dess partiella andraderivator (ifall dessa existerar, givetvis):

$$H(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1 x_1} & \cdots & \frac{\partial^2 F}{\partial x_1 x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n x_1} & \cdots & \frac{\partial^2 F}{\partial x_n x_n} \end{pmatrix}.$$

Exempel. Låt $F(x, y) = x^2 \sin y$. Då är

$$H(x, y) = \begin{pmatrix} 2 \sin y & 2x \cos y \\ 2x \cos y & -x^2 \sin y \end{pmatrix}.$$

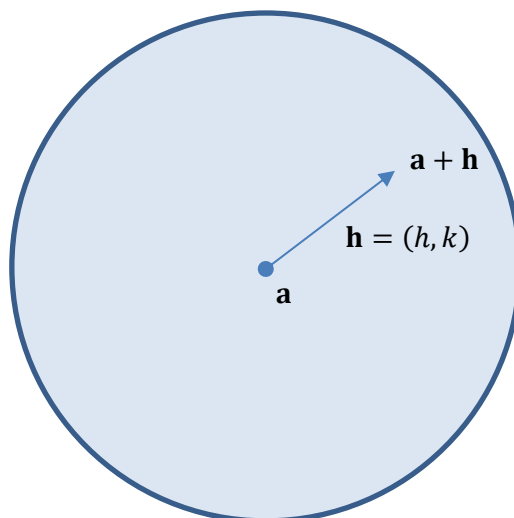
Notera att Hessianen kan betraktas som ett "matrisfält", d.v.s. som en funktion som till varje punkt i rummet ordnar en matris. Dessutom kan anmärkas att $H(\mathbf{x})$ är en symmetrisk matris i varje punkt om $F \in \mathcal{C}^2$ (vilket nästan alltid är fallet).

7.7.1 Taylorutveckling

Vi kan nu formulera satsen om Taylorutveckling, men låt oss först påminna om själva idén bakom Taylorutveckling (det är samma idé som i envariabelanalysen). Vi har alltså ett skalärfält $F: \mathbb{R}^n \rightarrow \mathbb{R}$ och vi vill approximera det med ett polynom i närheten av en given punkt $\mathbf{a} \in \mathbb{R}^n$. Anledningen är, precis som i envariabelfallet, att polynom är lätta att arbeta med. Så även om F ges av ett invecklat uttryck, så kan vi analysera funktionens beteende nära \mathbf{a} genom att betrakta dess Taylorpolynom i punkten.

Målet är således att hitta ett polynom som till varje punkt i någon omgivning U av $\mathbf{a} \in \mathbb{R}^n$ approximerar F 's värde i punkten. För att åstadkomma detta måste vi bestämma oss för en metod för att specificera punkter i U . Vi gör detta genom att införa en oberoende variabel $\mathbf{h} = (h_1, \dots, h_n)$ som ger *steget* från punkten \mathbf{a} . Målet är då att approximera $F(\mathbf{a} + \mathbf{h})$ med ett polynom i h_1, \dots, h_n .

I två respektive tre variabler brukar man beteckna komponenterna hos vektorn \mathbf{h} med (h, k) och (h, k, l) . I bilden nedan illustreras det tvådimensionella fallet.



Vi formulerar nu Taylorutvecklingen.

Sats. Antag att $F \in \mathcal{C}^3(U)$ där U är en öppen mängd i \mathbb{R}^n . Välj ett $\mathbf{a} \in U$. Då är

$$F(\mathbf{a} + \mathbf{h}) = F(\mathbf{a}) + \nabla F(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} + \mathcal{O}(|\mathbf{h}|^3)$$

för varje vektor $\mathbf{h} \in \mathbb{R}^n$.

Här gör vi *approximationen*

$$F(\mathbf{a} + \mathbf{h}) \approx F(\mathbf{a}) + \nabla F(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} \quad (1)$$

och det satsen säger är att *felet* då kan skrivas $\mathcal{O}(|\mathbf{h}|^3)$, d.v.s. felet kan skrivas på formen $B(\mathbf{h})|\mathbf{h}|^3$ där $B: \mathbb{R}^n \rightarrow \mathbb{R}$ är begränsad i någon omgivning av origo. Speciellt går felet mot noll då steget $\mathbf{h} \rightarrow \mathbf{0}$, d.v.s. då vi närmar oss punkten \mathbf{a} .

Läsaren kanske nu undrar var *polynomen* kommer in i bilden. Målet var ju hela tiden att approximera $F(\mathbf{a} + \mathbf{h})$ med ett *polynom* i \mathbf{h} 's komponenter, men vad i all sin dar är högerledet i (1)? Är det verkligen ett polynom i \mathbf{h} 's komponenter? Ja, det är det.

Exempel

Låt $F(x, y) = x^2 \sin y$. Bestäm Taylorutvecklingen av F kring punkten $(1, \frac{\pi}{2})$.

Lösning: Vi har $F(1, \frac{\pi}{2}) = 1$. Eftersom $\nabla F(x, y) = (2x \sin y, x^2 \cos y)$ är vidare $\nabla F(1, \frac{\pi}{2}) = (2, 0)$. Slutligen är Hessianen

$$H(x, y) = \begin{pmatrix} 2 \sin y & 2x \cos y \\ 2x \cos y & -x^2 \sin y \end{pmatrix}$$

i allmänhet och i vår punkt

$$H(1, \frac{\pi}{2}) = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}.$$

Därför är

$$\begin{aligned} F\left(1+h, \frac{\pi}{2}+k\right) &= F\left(1, \frac{\pi}{2}\right) + \nabla F\left(1, \frac{\pi}{2}\right) \cdot (h, k) + \frac{1}{2} \begin{pmatrix} h \\ k \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} + \mathcal{O}(|(h, k)|^3) = \\ &= 1 + 2h + 2h^2 - \frac{1}{2}k^2 + \mathcal{O}(|(h, k)|^3). \end{aligned}$$

Svar: $F\left(1+h, \frac{\pi}{2}+k\right) = 1 + 2h + 2h^2 - \frac{1}{2}k^2 + \mathcal{O}(|(h, k)|^3).$

Approximationen i exemplet är alltså $F\left(1+h, \frac{\pi}{2}+k\right) \approx 1 + 2h + 2h^2 - \frac{1}{2}k^2$ där högerledet är ett *polynom* i (h, k) , d.v.s. i steget vi går från punkten $\left(1, \frac{\pi}{2}\right)$.

Notera att ettan i approximationen är F 's värde i punkten $\left(1, \frac{\pi}{2}\right)$. Resten av högerledet står sålunda för *förändringen* när vi går steget (h, k) från den punkten. Och om steget är "litet", d.v.s. om beloppet $|(h, k)| = \sqrt{h^2 + k^2}$ är "litet", så bör approximationen vara god, eftersom feltermen är $\mathcal{O}(|(h, k)|^3)$ och därmed går mot noll då $(h, k) \rightarrow (0, 0)$.

Taylorapproximationen är ganska intuitiv, åtminstone "till 2/3". Det är ju klart att om $\mathbf{h} = (h, k)$ är ett litet steg så är

$$F(\mathbf{a} + \mathbf{h}) \approx F(\mathbf{a}).$$

(Approximationen så här långt hade varit exakt om F varit konstant.) En bättre approximation får vi om vi tar med första ordningens korrektion, d.v.s.

$$F(\mathbf{a} + \mathbf{h}) \approx F(\mathbf{a}) + \nabla F(\mathbf{a}) \cdot \mathbf{h}.$$

Skalärprodukten $\nabla F(\mathbf{a}) \cdot \mathbf{h}$ är ju lika med $\frac{\partial F}{\partial x}h + \frac{\partial F}{\partial y}k$ d.v.s. "så mycket F ökar per meter höger vi går gånger så många meter åt höger vi går plus så mycket F ökar per meter uppåt vi går gånger så många meter uppåt vi går". Ganska intuitivt. (Approximationen så här långt hade varit exakt om F hade haft konstant gradient.)

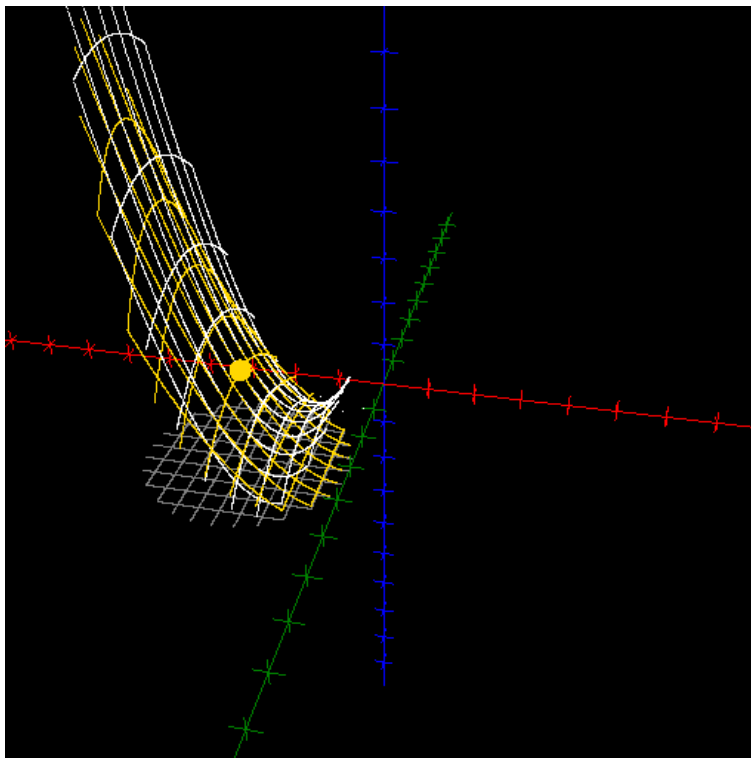
En ännu bättre approximation får vi om vi även tar med andra ordningens korrektion som i (1) ovan. (Här känner författaren emellertid inte till någon fullt lika enkel intuitiv förklaring, utan man får lita på det allmänna resonemanget bakom Taylorutveckling vilket studenten bör vara förtrogen med åtminstone i envariabelfallet.)

Exempel. I föregående exempel kom vi fram till att $F\left(1+h, \frac{\pi}{2}+k\right) \approx 1 + 2h + 2h^2 - \frac{1}{2}k^2$ för små steg (h, k) där $F(x, y) = x^2 \sin y$. Med ett enkelt variabelbyte ($x = 1 + h, y = \frac{\pi}{2} + k$) har vi alltså

$$F(x, y) \approx 1 + 2(x - 1) + 2(x - 1)^2 - \frac{1}{2}\left(y - \frac{\pi}{2}\right)^2$$

när (x, y) är nära $\left(1, \frac{\pi}{2}\right)$. Eftersom skalärfältet F och polynomet i högerledet är approximativt lika i närheten av $\left(1, \frac{\pi}{2}\right)$ bör deras visualiseringar påminna om varandra. (Precis som, i envariabelfallet, graferna $y = \sin x$ och $y = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!}$ är väldigt "lika" nära origo.) Nedan visas grafen dels till

F , dels till approximationen, över disken av radie 1 kring punkten $(1, \frac{\pi}{2})$. Tangeringspunkten $(1, \frac{\pi}{2}, 1)$ är markerad med en gul boll.



7.7.2 Undersökning av stationär punkt

I det här avsnittet använder vi (andra ordningens) Taylorutveckling som redskap för att undersöka naturen hos en *allmän* stationär punkt hos ett *allmänt* skalärfält. Eftersom (andra ordningens) Taylorutvecklingen är ett polynom av grad två är det åtminstone *tänkbart* att samma metoder som i det "enkla fallet" som vi studerade tidigare kan användas.

Säg nu att vi hittat en stationär punkt \mathbf{a} till skalärfältet $F: \mathbb{R}^n \rightarrow \mathbb{R}$ och att vi vill undersöka denna. Vi Taylorutvecklar därför F kring punkten \mathbf{a} , varvid vi erhåller

$$F(\mathbf{a} + \mathbf{h}) = F(\mathbf{a}) + \nabla F(\mathbf{a}) \cdot \mathbf{h} + \frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} + \mathcal{O}(|\mathbf{h}|^3)$$

där högerledet, som bekant, är ett polynom av grad 2 i komponenterna till \mathbf{h} , så när som på resttermen $\mathcal{O}(|\mathbf{h}|^3)$. Eftersom \mathbf{a} är en stationär punkt är $\nabla F(\mathbf{a}) = \mathbf{0}$ så att endast

$$F(\mathbf{a} + \mathbf{h}) = F(\mathbf{a}) + \frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} + \mathcal{O}(|\mathbf{h}|^3)$$

återstår. Den vanliga omflyttningen ger

$$F(\mathbf{a} + \mathbf{h}) - F(\mathbf{a}) = \frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} + \mathcal{O}(|\mathbf{h}|^3)$$

där högerledet, så när som på resttermen $\mathcal{O}(|\mathbf{h}|^3)$, är en kvadratisk form. Det är viktigt att komma ihåg att begreppen "lokal extrempunkt" och "sadelpunkt" är *lokala* till sin natur. Till exempel betyder ju "lokalt maximum" att det finns *någon* omgivning av punkten i vilken punkten är det största värdet;

denna omgivning kan vara hur liten som helst. Vi tittar alltså *nära* punkten \mathbf{a} , så \mathbf{h} är ett *litet* steg. Den kvadratiske formen $\frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h}$ är av storleksordningen "två små tal gånger varandra", medan resttermen $\mathcal{O}(|\mathbf{h}|^3)$ är av storleksordningen "tre små tal gånger varandra". Därför har vi

$$F(\mathbf{a} + \mathbf{h}) - F(\mathbf{a}) \approx \frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h}$$

om bara $|\mathbf{h}|$ är tillräckligt liten (vilket är allt vi är intresserade av!). Man kan därför misstänka att *samma* analys som vi gjorde i det "enkla fallet" ger rätt resultat även här (t.ex. positivt definit kvadratisk form ger strängt lokalt minimum).

Och det är nästan så väl. Det stora problemet uppstår om den kvadratiske formen $\frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h}$ är *semidefinit*. I sådana fall är ju $\frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h} = 0$ i en hel riktning (t.ex. då $\mathbf{h} = (t, 2t)$ i planet). Och jämfört med *exakt noll* är det klart att resttermen $\mathcal{O}(|\mathbf{h}|^3)$ inte är försumbar! I dessa två fall (positivt semidefinit och negativt semidefinit) fungerar alltså inte metoden med Taylorutveckling. I stället måste man undersöka den stationära punkten på något annat sätt.

Det är sålunda uppenbart att den "enkla" metoden att bara undersöka den kvadratiske formen $\frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h}$ *inte* fungerar om formen är semidefinit. Det faktum *att* metoden fungerar i de övriga fallen är något som måste bevisas (även om det förefaller troligt), och det görs säkerligen i läsarens ordinarie kursbok i flervariabelanalys. Dessa satser säger alltså t.ex. att om $\frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h}$ är positivt definit *så* är punkten \mathbf{a} ett strängt lokalt minimum för F (d.v.s., i princip: resttermen orsakar inget problem).

Exempel

Bestäm alla stationära punkter till $F(x, y) = x^4 - xy + y^4$ och undersök dessa.

Lösning: Gradienten är

$$\nabla F(x, y) = (4x^3 - y, 4y^3 - x)$$

så de stationära punkterna ges av

$$\begin{aligned} \nabla F(x, y) = \mathbf{0} &\iff \begin{cases} 4x^3 - y = 0 \\ 4y^3 - x = 0 \end{cases} \iff \\ &\iff \begin{cases} y = 4x^3 \\ x = 4(4x^3)^3 \end{cases} \iff \begin{cases} y = 4x^3 \\ x = 256x^9 \end{cases} \iff \\ &\iff (x, y) = (0, 0) \text{ eller } \begin{cases} y = 4x^3 \\ 1 = 256x^8 \end{cases} \iff \\ &\iff (x, y) = (0, 0) \text{ eller } (x, y) = \pm \left(\frac{1}{2}, \frac{1}{2}\right). \end{aligned}$$

Vi kommer att behöva Hessianen

$$H(x, y) = \begin{pmatrix} 12x^2 & -1 \\ -1 & 12y^2 \end{pmatrix}.$$

Vi undersöker nu de två stationära punkterna var för sig. I punkten $(x, y) = (0, 0)$ har vi

$$\frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} = \frac{1}{2} (h \ k) \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = -hk$$

som uppenbarligen är indefinit; origo är sålunda en sadelpunkt för F . (Detta kunde vi ha sett direkt. Hur?)

I båda punkterna $(x, y) = \pm \left(\frac{1}{2}, \frac{1}{2}\right)$ får vi

$$\begin{aligned} \frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} &= \frac{1}{2} (h \ k) \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = \frac{3}{2} h^2 - hk + \frac{3}{2} k^2 = \frac{3}{2} \left[h^2 - \frac{2}{3} hk + k^2 \right] = \\ &= \frac{3}{2} \left[\left(h - \frac{1}{3} k \right)^2 + \frac{8}{9} k^2 \right] \end{aligned}$$

som är positivt definit; alltså är båda punkterna stränga lokala minima.

Svar: F har en sadelpunkt i origo samt stränga lokala minima i $\pm \left(\frac{1}{2}, \frac{1}{2}\right)$; det finns inga andra stationära punkter.

Exempel

Bestäm alla stationära punkter till $F(x, y) = xye^{-x^2-y^2}$ och undersök dessa.

Lösning: Gradienten är

$$\nabla F(x, y) = (y - 2x^2y, x - 2xy^2)e^{-x^2-y^2}$$

så att de stationära punkterna ges av

$$\begin{aligned} \nabla F(x, y) = \mathbf{0} &\iff \begin{cases} y = 2x^2y \\ x = 2xy^2 \end{cases} \iff \\ &\iff (x, y) = (0, 0) \text{ eller } \begin{cases} 1 = 2x^2 \\ 1 = 2y^2 \end{cases} \iff \\ &\iff (x, y) \in \left\{ (0, 0), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \right\}. \end{aligned}$$

Hessianen är

$$H(x, y) = \begin{pmatrix} 4x^3y - 6xy & 1 - 2x^2 - 2y^2 + 4x^2y^2 \\ 1 - 2x^2 - 2y^2 + 4x^2y^2 & 4xy^3 - 6xy \end{pmatrix} e^{-x^2-y^2}.$$

I punkten $(x, y) = (0, 0)$ är

$$\frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} = \frac{1}{2} (h \ k) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = hk$$

som är indefinit, så origo är en sadelpunkt för F . I punkterna $(x, y) = \pm \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ är

$$\frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} = \frac{1}{2e} (h \ k) \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = -\frac{h^2}{e} - \frac{k^2}{e}$$

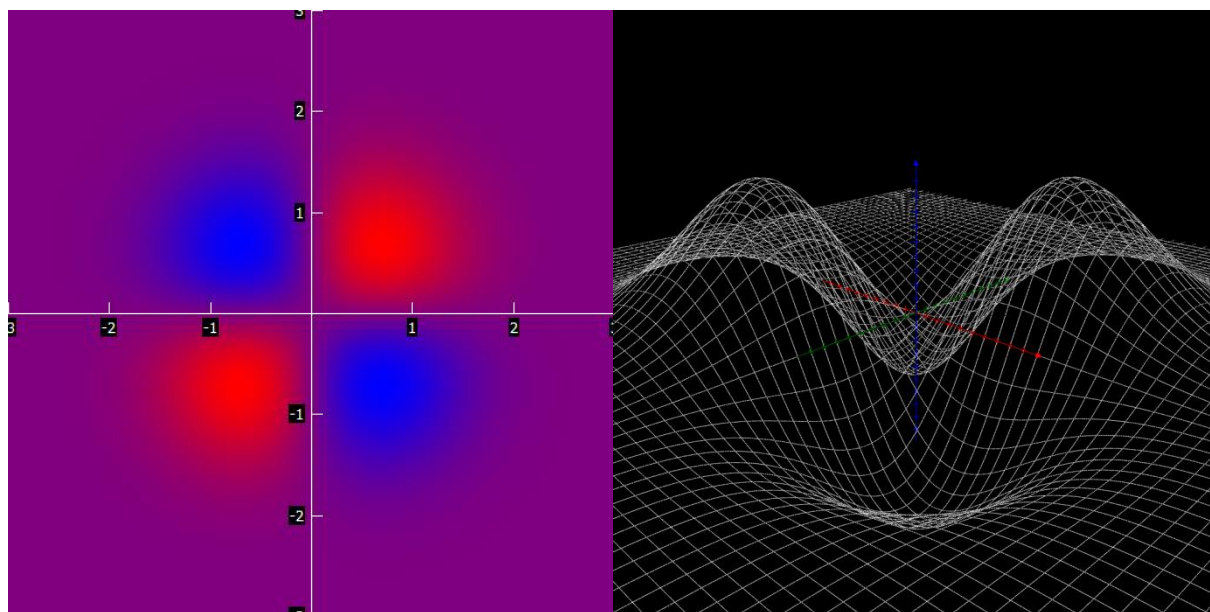
som är negativt definit, så dessa punkter är stränga lokala maxima. Slutligen, i punkterna $(x, y) = \pm \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ har vi

$$\frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} = \frac{1}{2e} (h \quad k) \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = \frac{h^2}{e} + \frac{k^2}{e}$$

som är positivt definit så att dessa två punkter är stränga lokala minima för F .

Svar: Origo är en sadelpunkt, $\pm \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ är stränga lokala maximipunkter och $\pm \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ är stränga lokala minimipunkter; det finns inga andra stationära punkter.

Nedan visas dels ett färgat plan, dels grafen till funktionen F i senaste exemplet (grafan är utdragen med en faktor 5 i z-led för att göra den tydligare).



Exempel

Bestäm alla stationära punkter till $F(x, y) = x^3 + x^2 + 4y^2 + 4xy$ och undersök dessa.

Lösning: Gradienten

$$\begin{aligned} \nabla F(x, y) = (3x^2 + 2x + 4y, 8y + 4x) = \mathbf{0} & \iff \\ \iff \begin{cases} 3x^2 + 2x + 4y = 0 \\ 2y + x = 0 \end{cases} & \iff \begin{cases} 3x^2 + 2x + 4y = 0 \\ x = -2y \end{cases} \iff \\ \iff \begin{cases} 12y^2 = 0 \\ x = -2y \end{cases} & \iff (x, y) = (0, 0) \end{aligned}$$

så origo är den enda stationära punkten till F . Hessianen

$$H(x, y) = \begin{pmatrix} 6x + 2 & 4 \\ 4 & 8 \end{pmatrix}$$

i allmänhet så i origo är

$$\frac{1}{2} \mathbf{h}^T H(\mathbf{a}) \mathbf{h} = \frac{1}{2} (h \quad k) \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix} = h^2 + 4hk + 4k^2 = (h + 2k)^2$$

som är positivt *semidefinit*. Alltså kan vi *inte* dra någon slutsats om den stationära punkten med hjälp av Taylorutvecklingen av andra ordningen. (Taylorutvecklingen kring origo hade vi förstås inte behövt härleda. Varför?)

Problemet är att den kvadratiske formen är noll längs linjen $h = -2k$ i hk -planet. Vi undersöker F 's beteende längs denna linje, som parametreras $(x, y) = (0, 0) + (-2t, t) = (-2t, t)$.

Vi får längs denna linje funktionsvärdena

$$F(-2t, t) = (-2t)^3 + (-2t)^2 + 4t^2 + 4(-2t)t = -8t^3$$

så att både positiva och negativa värden erhålles i *varje* omgivning av origo, där funktionsvärdet är noll. Tydligen är origo en sadelpunkt.

Svar: Den enda stationära punkten är origo, som är en sadelpunkt.

Anmärkning: Det är kanske inte så vanligt att studenter *direkt* inser det, men man kan också redan från början notera att $F(x, y) = (x + 2y)^2 + x^3$ och sedan på en gång påpeka att vi längs linjen $(x, y) = (-2t, t)$ har "profilen" $F(t) = t^3$ och därmed har att göra med en sadelpunkt i origo. Lite lättare hade kanske uppgiften " $x^2 + 2xy + y^2 + x^3$ " varit.

7.7.2.1 Ett förvånansvärt vanligt fel

Ett förvånansvärt vanligt fel studenter begår är att de behandlar likhetstecknet "=" felaktigt. Betrakta följande korrekta utdrag ur en lösning:

[...] I punkten (1,2) är den kvadratiske formen

$$2h^2 + 4hk + 6k^2 = 2[h^2 + 2hk + 3k^2] = 2[(h + k)^2 + 2k^2]$$

som är positivt definit, så (1,2) är ett strängt lokalt minimum för F , och här är $F(1,2) = 7$ (säg). [...]

Förvånansvärt ofta ser man följande **tokigheter** när man rättar skrivningar:

[...] I punkten (1,2) är den kvadratiske formen

$$2h^2 + 4hk + 6k^2 = h^2 + 2hk + 3k^2 = (h + k)^2 + 2k^2$$

som är positivt definit, så (1,2) är ett strängt lokalt minimum för F , och här är $(1,2) = 7$. [...]

Här har vi två fel: Första felet är " $2h^2 + 4hk + 6k^2 = h^2 + 2hk + 3k^2$ " som inte är sant. Om t.ex. $h = k = 1$ står det ju $12 = 6$ vilket uppenbarligen är fel. Förvisso är det sant att man, när man undersöker en kvadratisk form, kan multiplicera/dividera hela formen med ett positivt tal utan att formens *teckenkaraktär* (t.ex. positivt definit) ändras, men man kan ju inte för den sakens skull hävda att *likhet* som tal gäller!

Andra felet är " $(1,2) = 7$ " vilket om möjligt är ännu mer barockt. Likhetstecknet betyder ju att det som står till vänster och det som står till höger är *exakt samma matematiska objekt*, t.ex. är $1 + 1 = 2$ och $(1,3) = (1,3)$ sanna utsagor, medan $1 + 1 = 3$ och $(2,1) = (1,0)$ är falska utsagor.

Utsagan " $(1,2) = 7$ " är uppenbart falsk, eftersom en *vektor* aldrig kan vara exakt samma sak som ett *tal* (det är ju till och med *olika sorters* matematiska objekt)! Det studenten menar är i stället att $F(1,2) = 7$, d.v.s. det *tal* som funktionen F ger ifrån sig i punkten $(1,2)$ är lika med talet 7. $F(1,2)$ och 7 är alltså båda *tal*, och dessutom precis *samma* tal, så man kan skriva likhetstecken mellan dem och erhålla en sann utsaga.

Övning

1. Förklara varför följande utdrag är tokigt:

[...] Här är den kvadratiske formen $Q(h, k) = h^2 - k^2$ som är indefinit, ty t.ex. $(1,0) = 1$ och $(0,1) = -1$.

7.7.2.2 Ett annat vanligt fel

Ett annat vanligt fel är att studenter säger saker som "punkten $(1,2)$ är positivt definit", vilket är meningslöst (en *punkt* kan inte vara "positivt definit" – vad i all sin dar betyder det?!). Det studenten menar är "den kvadratiske formen (eller *Hessianen*) i punkten $(1,2)$ är positivt definit" eller liknande. "Positivt definit" är ju, som bekant, en egenskap hos en kvadratisk form (som t.ex. en Hessian).

8 Optimering

Nu är vi redo för ett av de två huvudnumren i flervariabelanalysen.

Optimering går ut på att (om möjligt) hitta det största och det minsta värdet hos ett skalärfält på en given mängd. Ett optimeringsproblem har därför två ingredienser:

1. ett skalärfält $F: D_F \rightarrow \mathbb{R}$ där $D_F \subset \mathbb{R}^n$
2. en mängd $U \subset D_F$.

Uppgiften är att hitta det största och minsta funktionsvärdet av $F(\mathbf{x})$ när $\mathbf{x} \in U$ – ifall dessa existerar – samt ange var dessa värden antas. Det kan hända att största eller minsta värde *saknas*.

Mer precist kallar vi $\max_{\mathbf{x} \in U} F(\mathbf{x})$ för funktionens *största värde* eller *globala maximum*, ifall det existerar. Symbolen " $\max_{\mathbf{x} \in U} F(\mathbf{x})$ " står för det största värdet i mängden $\{F(\mathbf{x}) : \mathbf{x} \in U\}$ ifall det existerar. Utsagan " $y = \max_{\mathbf{x} \in U} F(\mathbf{x})$ " betyder sålunda precis att (1) det finns ett $\mathbf{a} \in U$ sådant att $y = F(\mathbf{a})$ och (2) $F(\mathbf{x}) \leq F(\mathbf{a})$ för alla $\mathbf{x} \in U$.

En punkt $\mathbf{a} \in U$ kallas för en *global maximipunkt* om $F(\mathbf{a}) = \max_{\mathbf{x} \in U} F(\mathbf{x})$, vilket medför att $F(\mathbf{x}) \leq F(\mathbf{a})$ för alla $\mathbf{x} \in U$. På motsvarande sätt definieras funktionens *minsta värde* eller *globala minimum* samt *global minimipunkt*.

8.1 Repetition av envariabelfallet

Vi ger några enkla envariabelexempel som uppvärmning.

Exempel. Låt $f: \mathbb{R} \rightarrow \mathbb{R}$ vara definierad av $f(x) = x^2$ och låt $U = [-1, 1]$. I sådana fall är största värdet av f på U lika med 1, och både $x = -1$ och $x = 1$ är globala maximipunkter. Funktionen minsta värde på U är 0 och det finns precis en global minimipunkt, nämligen $x = 0$.

Exempel. Låt $f: \mathbb{R} \rightarrow \mathbb{R}$ vara definierad av $f(x) = x^2$ och låt $U =]-1, 1[$. Nu saknas största värde, ty givet varje tal $y < 1$ finns det ett tal $a \in U$ sådant att $f(a) > y$. Å andra sidan, om $y \geq 1$ så finns det *inget* tal $a \in U$ sådant att $f(a) = y$. Så *inget* tal är funktionens största värde. Däremot är det fortfarande sant att funktionens minsta värde är 0 och detta antages endast i $x = 0$.

Anmärkning ():* Även om största värdet $\max_{x \in U} f(x)$ inte existerar i senaste exemplet så existerar *supremum* $\sup_{x \in U} f(x)$ och är lika med 1. Ett tal M sådant att $f(x) \leq M$ för alla $x \in U$ kallas för en *majorant* till f på U , och supremum av f på U definieras som den minsta majoranten. Detta tal existerar alltid (med konventionen att supremum är " ∞ " om det inte finns någon majorant alls). Till exempel är det klart att $f(x) \leq 7$ för alla $x \in U$, precis som $f(x) \leq 3$ för alla $x \in U$ och $f(x) \leq 1$ för alla $x \in U$. Däremot är det inte sant att $f(x) \leq 0.7$ för alla $x \in U$. Alltså är talen 7, 3 och 1 – men inte talet 0.7 – majoranter till f på U . Mängden av majoranter är i det här fallet $[1, \infty[$. Den *minsta* majoranten är uppenbarligen 1, så $\sup_{x \in U} f(x) = 1$.

Motsvarigheten för "minsta värde" kallas för *infimum*. Ett tal M sådant att $f(x) \geq M$ för alla $f \in U$ kallas för en *minorant* till f på U , och *infimum* definieras som den största minoranten. Detta tal existerar alltid (med konventionen att infimum är " $-\infty$ " om det inte finns någon minorant alls).

Exempel. Låt $f: \mathbb{R} \rightarrow \mathbb{R}$ vara definierad av $f(x) = \arctan x$ och låt $U = \mathbb{R}$. Nu saknas både största och minsta värde av f på U . (Men infimum är $-\pi/2$ och supremum är $\pi/2$.)

Vi kan alltså inte stryka förutsättningen "kompakt" i följande, mycket viktiga, sats:

Sats. Antag att $f: D_f \rightarrow \mathbb{R}$ ($D_f \subset \mathbb{R}$) är *kontinuerlig* och att $U \subset D_f$ är *kompakt*. Då existerar största och minsta värdet av f på U .

Om man har en *kontinuerlig* funktion på en *kompakt* mängd vet man alltså att funktionen har både ett största och ett minsta värde på mängden. Detta gör optimeringsproblemet mycket lättare, som vi skall se.

8.1.1 Optimering på kompakt mängd

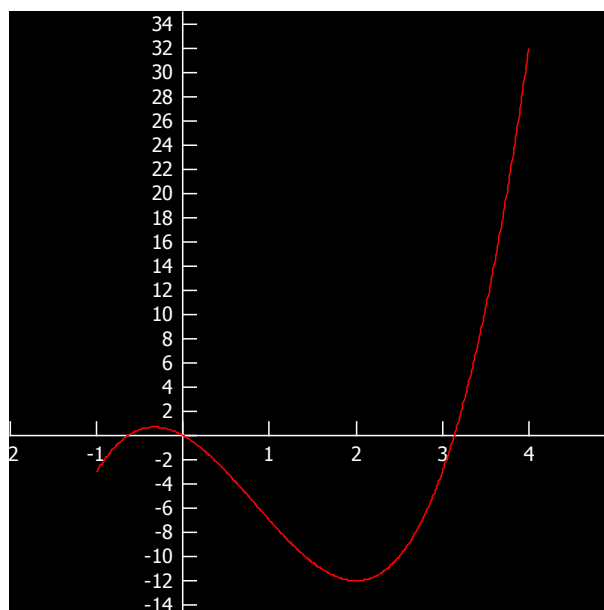
Säg att $f(x)$ är kontinuerlig på ett kompakt intervall⁷ I ; vi vet då att både största och minsta värde av f på I existerar, men hur bestämmer vi dessa?

Låt $x \in I$ vara en av funktionens globala maximipunkter (vi vet ju att det finns minst en sådan!) på I . Av definitionerna följer det att x då även är en *lokal* maximipunkt. Eftersom $x \in I$ är det klart att I antingen är en randpunkt eller en inre punkt till I . Om x är en inre punkt och f är deriverbar så måste x vara en stationär punkt, så $f'(x) = 0$. Detta intuitivt "uppenbara" resultat behandlade vi i avsnitt 7.5.1.1.

Alltså, om x är en maximipunkt så är x antingen en randpunkt eller en inre stationär punkt; dessa, oftast ganska få punkter, är de enda kandidaterna. Det vi gör i praktiken är att vi bestämmer dessa punkter och beräknar funktionsvärdet i dem. Sedan väljer vi bara det största av dem.

På motsvarande sätt bestämmer vi funktionens minsta värde.

Exempel. Betrakta funktionen $f(x) = 2x^3 - 5x^2 - 4x$ på $[-1, 4]$; grafen visas nedan.



⁷ Varför krångla till det?

Eftersom f är kontinuerlig och mängden är kompakt så finns garanterat största och minsta värde. Vi undersöker först de inre stationära punkterna. Derivatn

$$\begin{aligned} f'(x) = 6x^2 - 10x - 4 = 0 &\iff x^2 - \frac{5}{3}x - \frac{2}{3} = 0 &\iff \\ \iff \left(x - \frac{5}{6}\right)^2 - \frac{49}{36} = 0 &\iff \left(x - \frac{5}{6}\right)^2 = \frac{49}{36} &\iff \\ \iff x - \frac{5}{6} = \pm \frac{7}{6} &\iff x \in \left\{-\frac{1}{3}, 2\right\}. \end{aligned}$$

I dessa punkter är

$$f\left(-\frac{1}{3}\right) = \frac{19}{27}, \quad f(2) = -12.$$

Sedan undersöker vi randen. Randen är $\partial[-1, 4] = \{-1, 4\}$ och här är

$$f(-1) = -3, \quad f(4) = 32.$$

Av dessa fyra tal är 32 störst och -12 minst. Därför är 32 funktionens största värde på mängden, och detta antages (endast) i punkten $x = 4$; -12 är funktionens minsta värde på mängden, och detta antages (endast) i punkten $x = 2$.

Exempel

Bestäm (om de finns) största och minsta värde av $f(x) = 3x^4 + 4x^3 - 12x^2$ på $[-1, 1]$.

Lösning: Funktionen är kontinuerlig och mängden är kompakt, så största och minsta värde existerar. Vi undersöker först de inre stationära punkterna. Derivatn

$$\begin{aligned} f'(x) = 12x^3 + 12x^2 - 24x = 0 &\iff x^3 + x^2 - 2x = 0 &\iff \\ \iff x(x-1)(x+2) = 0 &\iff x \in \{-2, 0, 1\}. \end{aligned}$$

Det finns tydligen bara en inre stationär punkt, nämligen $x = 0$, medan randpunkten $x = 1$ också råkar vara en stationär punkt till f på \mathbb{R} . I punkten $x = 0$ är

$$f(0) = 0.$$

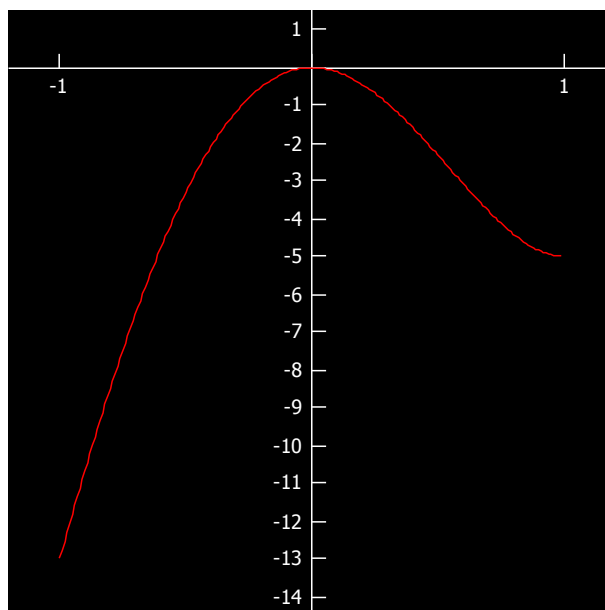
Randen är $\partial[-1, 1] = \{-1, 1\}$ och här är

$$f(-1) = -13, \quad f(1) = -5.$$

Av dessa värden är 0 störst och -13 minst.

Svar: Funktionen största värde är 0 och detta antages (endast) i punkten $x = 0$. Funktionen minsta värde är -13 och detta antages (endast) i punkten $x = -1$.

Nedan visas grafen till f i exemplet.



Övningar

1. Ge ett enkelt exempel på en funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ och ett kompakt intervall $I \subset \mathbb{R}$ sådant att f :s största och minsta värde på I båda antages på randen ∂I .
2. Ge ett enkelt exempel på en funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ och ett kompakt intervall $I \subset \mathbb{R}$ sådant att f :s största och minsta värde på I båda antages i inre stationära punkter.
3. Ge ett enkelt exempel på en funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ och ett slutet men obegränsat intervall $I \subset \mathbb{R}$ sådant att f saknar största värde på I .
4. Ge ett enkelt exempel på en funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ och ett begränsat men inte slutet intervall $I \subset \mathbb{R}$ sådant att f saknar största värde på I .
5. Ge ett enkelt exempel på en funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ och ett intervall I som *inte* är slutet, sådan att f ändå har både största och minsta värde på I .
6. Ge ett enkelt exempel på en funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ och ett intervall I som *inte* är begränsat, sådan att f ändå har både största och minsta värde på I .

8.2 Flervariabelfallet

Studera följande inledande exempel.

Exempel. Betrakta funktionen $F(x, y) = y$ på mängden $U = [0, 1] \times [0, 1[$ (rita!). Funktionen minsta värde är uppenbarligen 0 och detta antages längs x -axeln. Däremot saknas största värde, ty givet vilket $c < 1$ som helst finns det ett $\mathbf{a} \in U$ sådant att $f(\mathbf{a}) > c$, men om $c \geq 1$ så finns *inget* $\mathbf{a} \in U$ sådant att $f(\mathbf{a}) = c$. "Problemet" här är tydligen att "randbiten" $[0, 1] \times \{1\}$ inte tillhör mängden, d.v.s. att U inte är sluten.

Exempel. Betrakta funktionen $F(x, y) = x$ på \mathbb{R}^2 . Här saknas såväl största som minsta värde, ty givet vilket $c \in \mathbb{R}$ som helst så finns det ett $\mathbf{a} \in \mathbb{R}^2$ sådant att $F(\mathbf{a}) < c$ och ett $\mathbf{b} \in \mathbb{R}^2$ sådant att $F(\mathbf{b}) > c$. "Problemet" här är tydligen att mängden är obegränsad.

Ovanstående exempel visar att vi inte kan stryka förutsättningen "kompakt" i satsen nedan.

Sats. Antag att F är ett kontinuerligt skalärfält på den kompakta mängden $U \subset \mathbb{R}^n$. Då antar F både ett största och ett minsta värde på U .

Alltså: har man en *kontinuerlig* funktion på en *kompakt* mängd, så vet man helt säkert att funktionen har såväl ett största som ett minsta värde på U .

8.2.1 Optimering på kompakta mängder

Låt F vara ett kontinuerligt skalärfält definierat (åtminstone) på den kompakta mängden U . Vi vet att F har både ett största och ett minsta värde på U , men hur bestämmer vi dessa?

I princip på exakt samma sätt som i envariabelfallet. I fortsättningen utgår vi ifrån att F är partiellt deriverbar, så att vi kan beräkna dess gradient.

Låt $\mathbf{a} \in U$ vara en av funktionens globala maximipunkter (vi vet ju att det finns minst en sådan!). Det är klart att \mathbf{a} också är en lokal maximipunkt. Antingen är \mathbf{a} en randpunkt till U , eller så är det en inre punkt. I senare fallet måste \mathbf{a} vara en stationär punkt, eftersom det är en lokal maximipunkt. (Detta intuitivt "uppenbara" resultat behandlade vi i avsnitt 7.5.2.) Med andra ord är $\nabla F(\mathbf{a}) = \mathbf{0}$.

För att hitta en funktions globala maximum och minimum behöver man alltså, i praktiken, göra som mest två saker. Om mängden har inre punkter så börjar man med att sätta funktionens gradient till noll, för att hitta de stationära punkterna till skalärfältet (åtminstone de som ligger i det inre av mängden). Sedan undersöker man randen separat; detta är nästan alltid ett enklare problem. När man hittat alla inre stationära punkter samt (kandidater till) max och min på randen, så jämför man alla dessa punkter, och väljer ut de som ger största och de som ger minsta funktionsvärdet.

När man sätter gradienten till noll (vilket man bara gör om U har inre punkter!) finner man alla stationära punkter till F , även de som (eventuellt) ligger utanför mängden U som vi undersöker. Sådana kandidater är förstas irrelevanta för oss, och förkastas. Om en stationär punkt till F råkar ligga precis på randen till U , så det inte en *inre* stationär punkt, men den är ju ändå intressant för oss, så vi behöver inte förkasta den (men vi kan också välja att göra det, för vi kommer garanterat att upptäcka den på nytt när vi undersöker randen separat – kom ihåg att höjdprofilen för en cykelbana genom en bergstopp har ett maximum på bergstoppen).

När det är dags att undersöka randen brukar man parametrisera den, varvid man erhåller ett optimeringsproblem av lägre dimension. Detta är, åtminstone i teorin, ett enklare problem. (Ofta består randen av flera delar som naturligt undersökes var för sig.)

8.3 Exempel med optimering i planet

I det här avsnittet ger vi exempel på optimeringsproblem i planet. Dessa problem har två ingredienser:

1. ett plant skalärfält $F: D_F \rightarrow \mathbb{R}$ där $D_F \subset \mathbb{R}^2$
2. ett område $U \subset D_F$ i planet.

Uppgiften är att, om möjligt, bestämma största och minsta värdet av F på U samt ange var på U dessa värden antages.

8.3.1 Max och min på kurva i planet

Innan vi ger oss i kast med det fullständiga optimeringsmaskineriet värmer vi upp med det lättaste fallet, nämligen när U är en *kurva* (som saknar inre punkter). Eftersom en kurva är en endimensionell "sak" får vi i praktiken ett envariabelproblem efter att vi parametriserat kurvan. Låt nämligen kurvan vara bilden $\mathbf{r}([a, b])$ där $\mathbf{r}: \mathbb{R} \rightarrow \mathbb{R}^2$ är parametriseringsfunktionen. Då är $g(t) := F(\mathbf{r}(t))$ "temperaturen klockan t " på vår vandring längs kurvan, så vi behöver bara hitta max/min av envariabelfunktionen $g = F \circ \mathbf{r}$ på intervallet $[a, b]$.

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = x + y^2$ på $y = x^2$, $-1 \leq x \leq 1$.

Lösning: Låt Γ beteckna parabelbiten i fråga. (Eftersom F är kontinuerlig och Γ är kompakt så finns säkert⁸ största och minsta värde.)

Vi parametriserar Γ med x som parameter, d.v.s. Γ är bilden av $[-1, 1]$ under $(x, y) = (x, x^2)$. Vi erhåller funktionsvärdet

$$g(x) := F(x, x^2) = x + x^4.$$

Vi skall sålunda hitta max och min av $g(x)$ när $x \in [-1, 1]$. Derivatans

$$g'(x) = 1 + 4x^3 = 0 \quad \iff \quad x^3 = -\frac{1}{4} \quad \iff \quad x = -\frac{1}{\sqrt[3]{4}}$$

som ligger i $[-1, 1]$ och här är

$$g\left(-\frac{1}{\sqrt[3]{4}}\right) = -\frac{1}{\sqrt[3]{4}} + \frac{1}{\sqrt[3]{256}}.$$

I ändpunkterna $x = -1$ och $x = 1$ har vi

$$g(-1) = 0, \quad g(1) = 2.$$

Vi noterar att

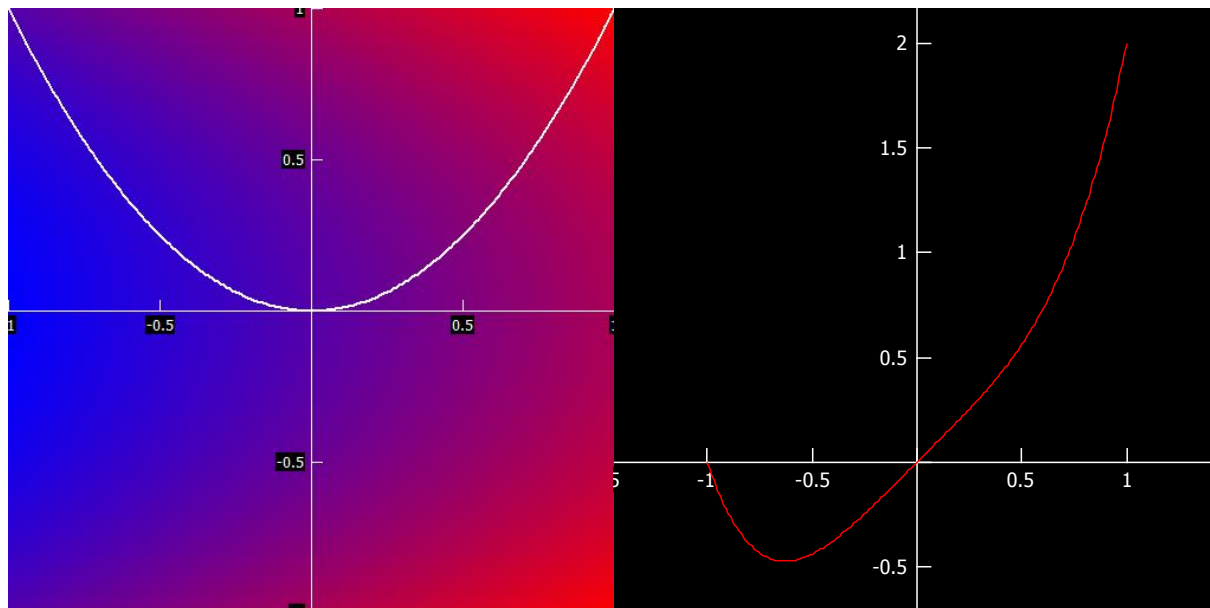
$$-\frac{1}{\sqrt[3]{4}} + \frac{1}{\sqrt[3]{256}} < 0 < 2$$

varvid vi är klara.

Svar: Funktionen minsta värde är $-\frac{1}{\sqrt[3]{4}} + \frac{1}{\sqrt[3]{256}}$ och detta antages (endast) i $\left(-\frac{1}{\sqrt[3]{4}}, \frac{1}{\sqrt[3]{16}}\right)$. Funktionen största värde är 2 och detta antages (endast) i $(1, 1)$.

Nedan visas dels den tvådimensionella situationen (med kurvan Γ och ett färgat plan för F) samt envariabelfunktionen g som ger "temperaturprofilen" längs "cykelbanan".

⁸ Ordet "säkert" är överflödigt och togs här endast med av pedagogiska skäl! I fortsättningen kommer vi prata "matematiska" och helt torrt konstatera att "största och minsta värde *existerar*".



Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = xy^3$ på enhetscirkeln.

Lösning: (Funktionen är kontinuerlig och enhetscirkeln är kompakt, så största och minsta värde existerar.) Enhetscirkeln kan parametriseras $(x, y) = (\cos t, \sin t)$ och funktionsvärdet vid t är

$$g(t) := F(\cos t, \sin t) = \cos t \sin^3 t.$$

Derivering ger

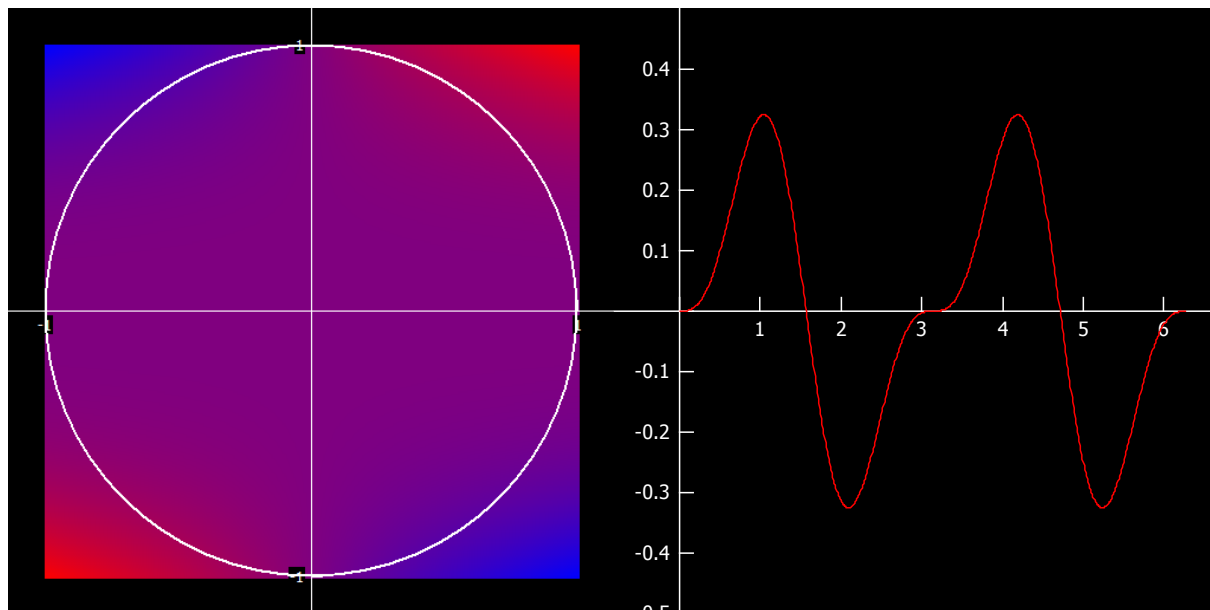
$$\begin{aligned} g'(t) &= 3 \cos^2 t \sin^2 t - \sin^4 t = \sin^2 t (3 \cos^2 t - \sin^2 t) = 0 && \iff \\ &\iff \sin t = 0 \text{ eller } \sin^2 t = 3 \cos^2 t && \iff \\ &\iff t = n\pi \text{ eller } \tan^2 t = 3 && \iff \\ &\iff t = n\pi \text{ eller } t = \frac{\pi}{3} + n\pi \text{ eller } t = -\frac{\pi}{3} + n\pi. \end{aligned}$$

I dessa stationära punkter (till g) har vi, under första varvet $[0, 2\pi[$,

$$\begin{aligned} g(0) &= 0, & g\left(\frac{\pi}{3}\right) &= \frac{3\sqrt{3}}{16}, & g\left(\frac{2\pi}{3}\right) &= -\frac{3\sqrt{3}}{16} \\ g(\pi) &= 0, & g\left(\frac{4\pi}{3}\right) &= \frac{3\sqrt{3}}{16}, & g\left(\frac{5\pi}{3}\right) &= -\frac{3\sqrt{3}}{16}. \end{aligned}$$

Svar: Största värdet $\frac{3\sqrt{3}}{16}$ antages (endast) i punkterna $\pm\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$. Minsta värdet $-\frac{3\sqrt{3}}{16}$ antages (endast) i punkterna $\pm\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)$.

Nedan visas dels den tvådimensionella situationen (med enhetscirkeln och ett färgat plan för F) samt envariabelfunktionen g som ger "temperaturprofilen" längs "cykelbanan".



Övning

1. Enhetscirkeln är ju (t.ex.) bilden av intervallet $[0, 2\pi[$ under $(x, y) = (\cos t, \sin t)$. Vi skall därför hitta max/min av $g(t)$ när $t \in [0, 2\pi[$. Varför behöver vi inte undersöka ändpunkterna $t = 0$ och $t = 2\pi$ separat? Varför räcker det med att sätta $g'(t) = 0$?

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = x - y$ då $x^2 + y^2 - 4x = -3$ och $x \geq 2$.

Lösning: Eftersom $x^2 + y^2 - 4x = -3 \Leftrightarrow (x - 2)^2 + y^2 = 1$ är mängden den (slutna) högra halvan av cirkeln med radie 1 kring $(2, 0)$. (Denna är kompakt och F är kontinuerlig, så största och minsta värde existerar.)

Halvcirkeln är bilden $\mathbf{r}\left(\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right)$ där $\mathbf{r}(t) = (2 + \cos t, \sin t)$ så vi skall hitta max/min av $g(t) := F(2 + \cos t, \sin t) = 2 + \cos t - \sin t$ på $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Derivatan

$$g'(t) = -\sin t - \cos t = 0 \Leftrightarrow \sin t = -\cos t \Leftrightarrow t = -\frac{\pi}{4} + n\pi$$

och av dessa stationära punkter är det endast $t = -\frac{\pi}{4}$ som tillhör $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Här är

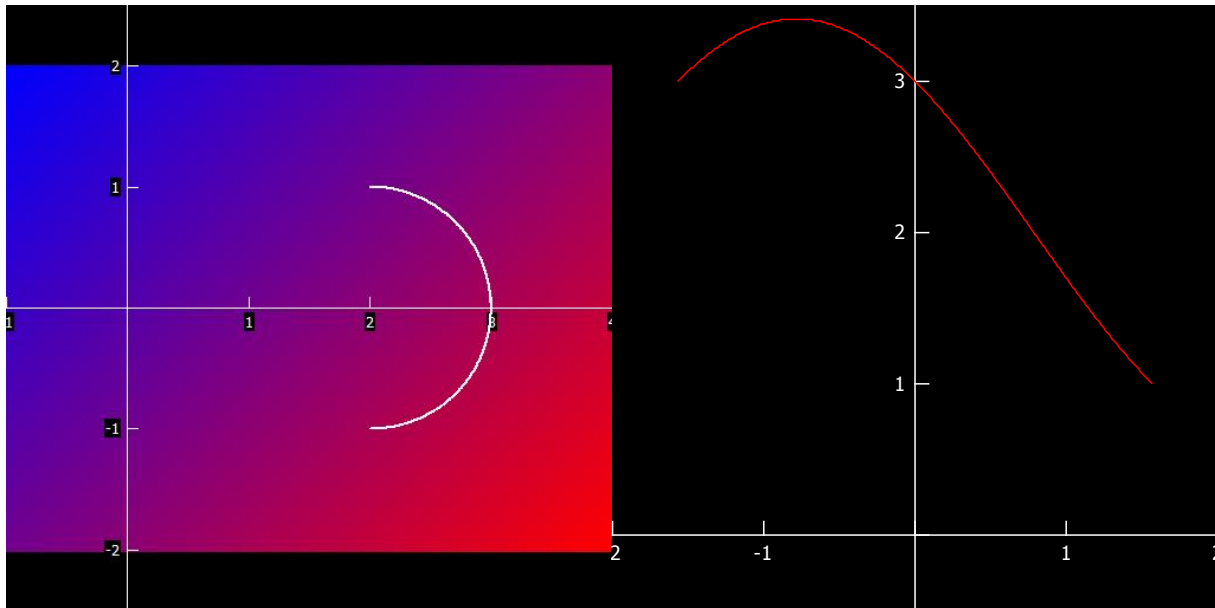
$$g\left(-\frac{\pi}{4}\right) = 2 + \sqrt{2} > 3.$$

I ändpunkterna har vi

$$g\left(-\frac{\pi}{2}\right) = 3, \quad g\left(\frac{\pi}{2}\right) = 1.$$

Svar: Största värdet $2 + \sqrt{2}$ erhålles (endast) i $\left(2 + \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$. Minsta värdet 1 erhålles (endast) i $(2, 1)$.

Nedan visas dels den tvådimensionella situationen (med halvcirkeln och ett färgat plan för F) samt envariabelfunktionen g som ger "temperaturprofilen" längs "cykelbanan".



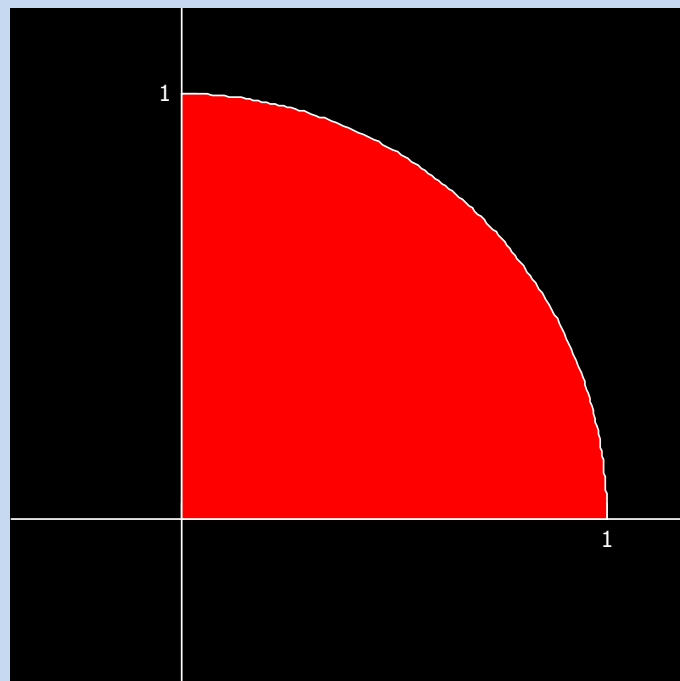
8.3.2 Max och min på tvådimensionellt område i planet

Vi betraktar nu fallet där mängden vi skal optimera över är ett *tvådimensionellt område* i planet. Nu behöver vi det fulla maskineriet.

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = (x + y)e^{-x^2 - y^2}$ på $D = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x^2 + y^2 \leq 1\}$.

Lösning: F är kontinuerlig och D är kompakt, så största och minsta värde existerar.



Vi bestämmer först de **inre stationära punkterna**. Gradienten

$$\begin{aligned} \nabla F(x, y) = (1 - 2x^2 - 2xy, 1 - 2y^2 - 2xy)e^{-x^2-y^2} = \mathbf{0} &\iff \\ \iff \begin{cases} 1 - 2x^2 - 2xy = 0 \\ 1 - 2y^2 - 2xy = 0 \end{cases} &\iff \begin{cases} 1 - 2x^2 - 2xy = 0 \\ x^2 = y^2. \end{cases} \end{aligned}$$

$x^2 = y^2$ betyder att $x = \pm y$. Om $x = y$ har vi $1 - 2x^2 - 2x^2 = 0$ som ger $x = \pm \frac{1}{2}$, så $\pm \left(\frac{1}{2}, \frac{1}{2}\right)$ är stationära punkter till F . Om $x = -y$ har vi å andra sidan $1 - 2x^2 + 2x^2 = 0$ som är ett falskt påstående, så $x \neq -y$. Därför är

$$\nabla F(x, y) = \mathbf{0} \iff (x, y) = \pm \left(\frac{1}{2}, \frac{1}{2}\right).$$

Av dessa stationära punkter är det emellertid endast $\left(\frac{1}{2}, \frac{1}{2}\right)$ som tillhör D . Här är funktionsvärdet

$$F\left(\frac{1}{2}, \frac{1}{2}\right) = e^{-1/2} = \frac{1}{\sqrt{e}}$$

Randen består av två linjestycken och en kvartscirkel, vilka vi nu undersöker var för sig.

Linjestycket $[0, 1] \times \{0\}$ kan parametreras $(x, y) = (t, 0)$ där $t \in [0, 1]$, och funktionsvärdet är här $F(t, 0) = te^{-t^2}$. Vi skall alltså hitta max och min av $g(t) := te^{-t^2}$ på $[0, 1]$. Detta är ett *envariabelproblem*. Derivatans $g'(t) = (1 - 2t^2)e^{-t^2} = 0 \iff t = \pm \frac{1}{\sqrt{2}}$. Av dessa två stationära punkter är det emellertid endast $t = \frac{1}{\sqrt{2}}$ som tillhör $[0, 1]$. Denna punkt svarar mot punkten $\left(\frac{1}{\sqrt{2}}, 0\right)$ och här är funktionsvärdet

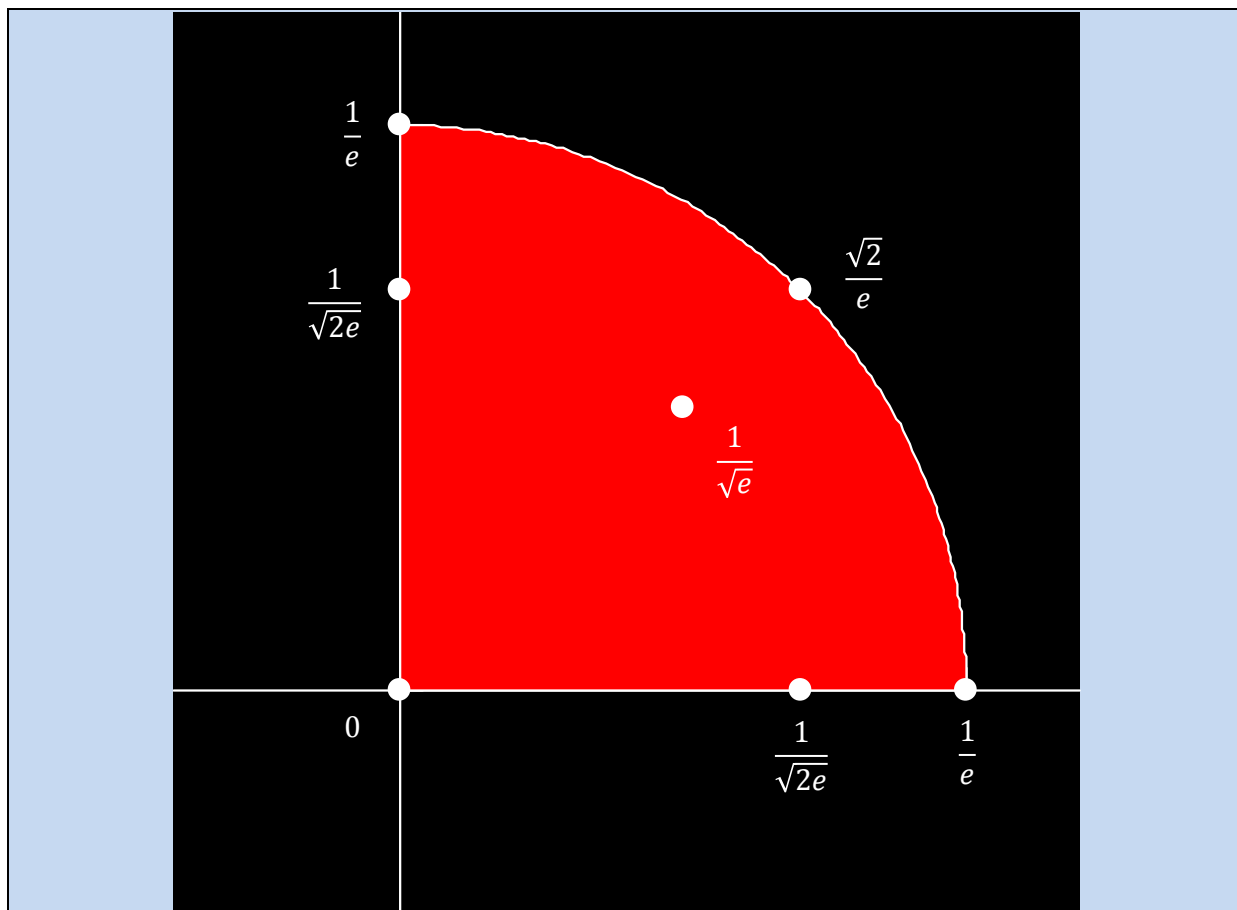
$$g\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}}e^{-1/2} = \frac{1}{\sqrt{2e}}$$

I ändpunkterna $t = 0$ och $t = 1$, vilka svarar mot $(0, 0)$ och $(1, 0)$, har vi $g(0) = 0$ samt $g(1) = 1/e$.

Linjestycket $\{0\} \times [0, 1]$ kan parametreras $(x, y) = (0, t)$ där $t \in [0, 1]$ och funktionsvärdet är här $F(0, t) = te^{-t^2}$. Helt analoga räkningar ger därför kandidaten $\left(0, \frac{1}{\sqrt{2}}\right)$ med funktionsvärdet $1/\sqrt{2e}$ samt ändpunkterna $F(0, 0) = 0$ (igen!!) och $F(0, 1) = 1/e$.

Kvartscirkeln $x^2 + y^2 = 1, x \geq 0, y \geq 0$ kan parametreras $(x, y) = (\cos t, \sin t)$ där $t \in [0, \pi/2]$, och funktionsvärdet är här $F(\cos t, \sin t) = (\cos t + \sin t)e^{-\cos^2 t - \sin^2 t} = (\cos t + \sin t)e^{-1}$. Vi skall alltså hitta max och min av *envariabelfunktionen* $h(t) := (\cos t + \sin t)e^{-1}$ på $[0, \pi/2]$. Derivatans $h'(t) = (-\sin t + \cos t)e^{-1} = 0 \iff \sin t = \cos t \iff t = \pi/4$. Denna punkt svarar mot $(x, y) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ och här är funktionsvärdet $h(\pi/4) = \sqrt{2}/e$. Ändpunkterna $t = 0$ och $t = \pi/2$ behöver i praktiken inte undersökas, eftersom dessa svarar mot de redan undersökta punkterna $(1, 0)$ respektive $(0, 1)$.

Samtliga kandidater är markerade i bilden nedan.



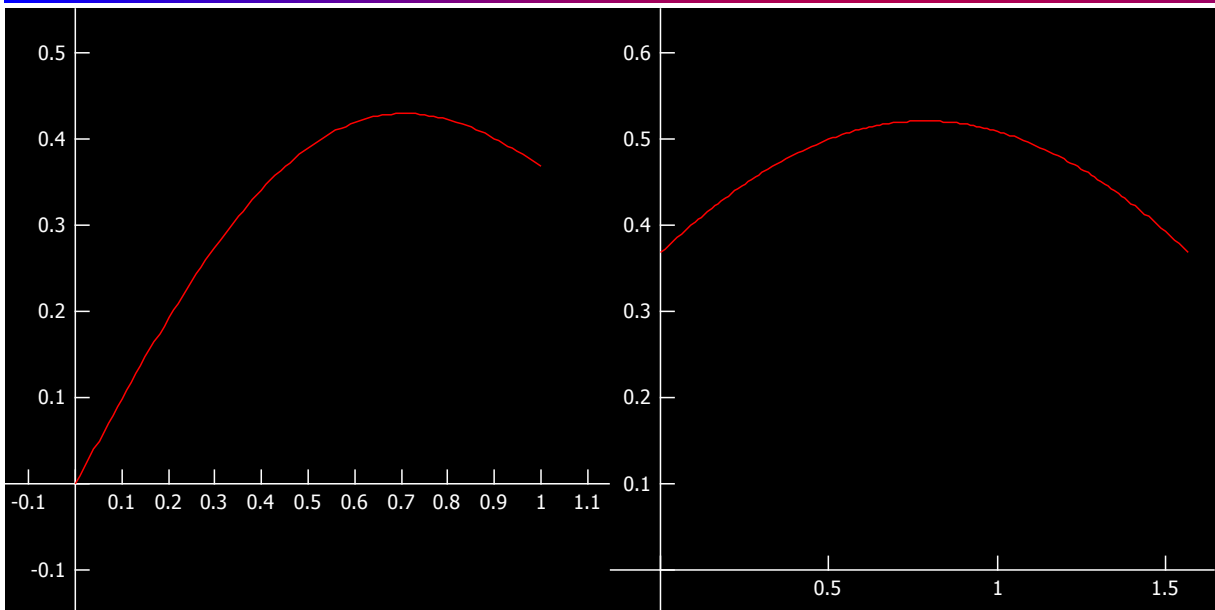
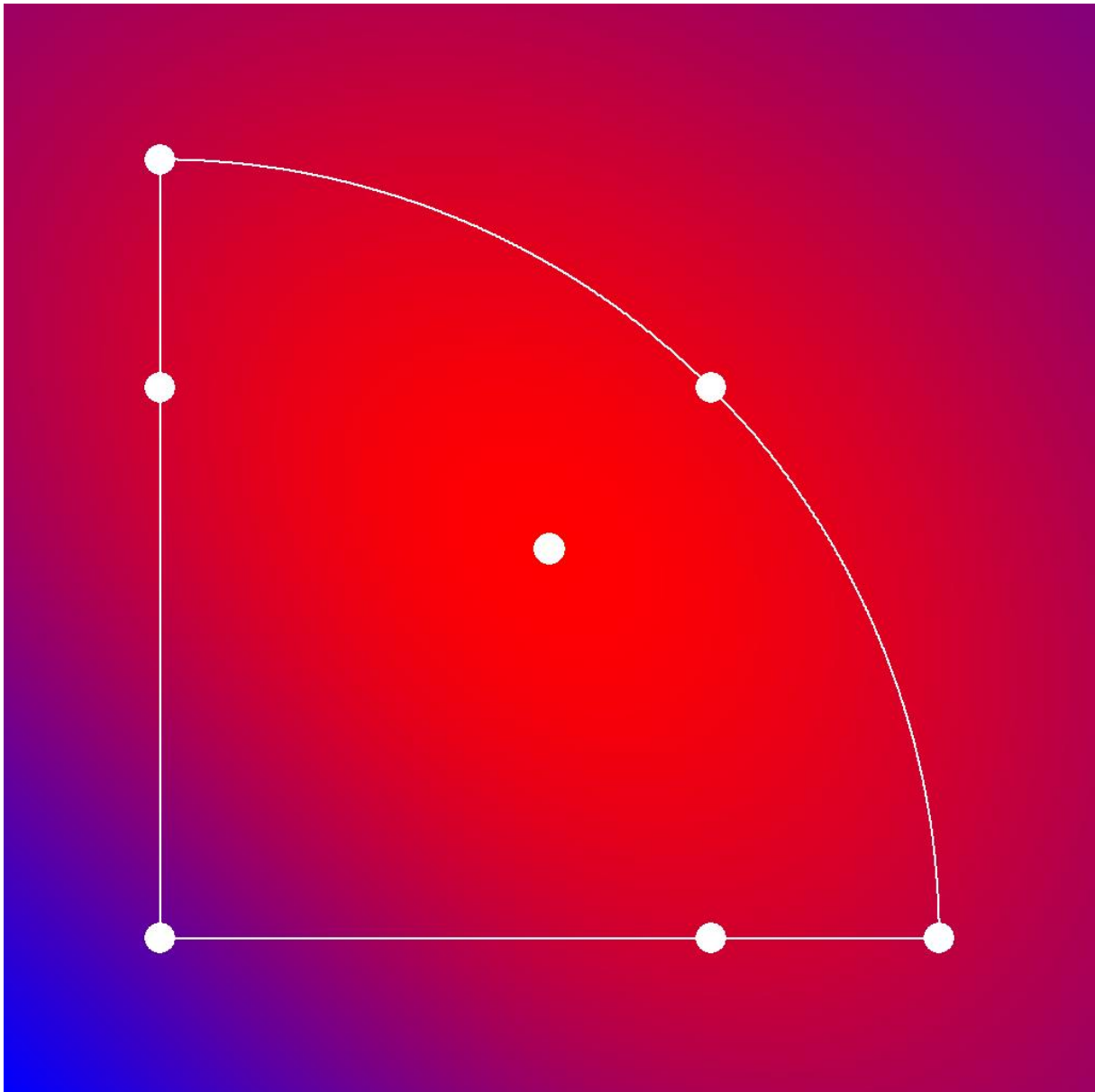
Jämförelse. Vi noterar att kandidatvärdena uppfyller

$$0 < \frac{1}{e} < \frac{1}{\sqrt{2e}} < \frac{\sqrt{2}}{e} < \frac{1}{\sqrt{e}}.$$

Svar: Funktionen minsta värde är 0 och detta antages (endast) i origo. Funktionen största värde är $\frac{1}{\sqrt{e}}$ och detta antages (endast) i $(\frac{1}{2}, \frac{1}{2})$.

Nedan, överst, visas ett färgat plan för F . I den här bilden "ser" vi att $(\frac{1}{2}, \frac{1}{2})$ är ett lokalt maximum för F , att de två räta linjestyckena samt kvartscirkeln får högsta värdet i deras "inre" kandidater (lokala maxima för "temperaturprofilerna"), samt att temperaturen är "på väg" att sjunka när vi går ifrån området (och längst nere till vänster hinner det bli ganska "kallt").

Underst, till höger, visas "temperaturprofilen" för de två räta linjerna (de har ju samma profil); till vänster visas "temperaturprofilen" för kvartscirkeln.



Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = xe^{y-x}$ på $D = \{(x, y) \in \mathbb{R}^2: x \geq 0, y \geq 0, x + y \leq 1\}$.

Lösning: F är kontinuerlig och D är kompakt, så största och minsta värde existerar.

Vi bestämmer först de inre stationära punkterna. Gradienten

$$\nabla F(x, y) = (1 - x, xe^{y-x})e^{y-x} = \mathbf{0} \iff x = 0 \text{ och } x = 1$$

så stationära punkter saknas. Randen består av tre linjestycken, vilka vi nu undersöker separat.

Linjestycket $[0, 1] \times \{0\}$ kan parametreras $(x, y) = (t, 0)$ där $t \in [0, 1]$ och här är funktionsvärdet $g(t) := te^{-t}$. Derivatan $g'(t) = (1 - t)e^{-t} = 0 \iff t = 1$ som är högra ändpunkten. Här är $g(1) = e^{-1}$. I vänstra ändpunkten är $g(0) = 0$.

Linjestycket $\{0\} \times [0, 1]$ kan parametreras $(x, y) = (0, t)$ där $t \in [0, 1]$ och här är funktionsvärdet identiskt noll.

Det sneda linjestycket $x \geq 0, y \geq 0, x + y = 1$ kan parametreras $(x, y) = (t, 1 - t)$ där $t \in [0, 1]$ och här är funktionsvärdet $h(t) := te^{1-2t}$. Derivatan $h'(t) = (1 - 2t)e^{1-2t} = 0 \iff t = 1/2$. Denna punkt svarar mot $(x, y) = (1/2, 1/2)$ och här är funktionsvärdet $h(1/2) = 1/2$. Ändpunkterna $t = 0$ och $t = 1$ behöver inte undersökas, eftersom dessa svarar mot de redan undersökta punkterna $(1, 0)$ och $(0, 1)$.

Svar: Funktionen minsta värde är 0 och detta antages (endast) på $\{0\} \times [0, 1]$. Funktionen största värde är $\frac{1}{2}$ och antages (endast) i punkten $(\frac{1}{2}, \frac{1}{2})$.

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = x^2 + y^2 + x$ på enhetsdisken $D = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq 1\}$.

Lösning: F är kontinuerlig och D är kompakt, så största och minsta värde existerar.

Vi bestämmer först de inre stationära punkterna. Gradienten

$$\nabla F(x, y) = (2x + 1, 2y) = \mathbf{0} \iff (x, y) = \left(-\frac{1}{2}, 0\right)$$

som är en inre punkt, och därmed en kandidat. Här är

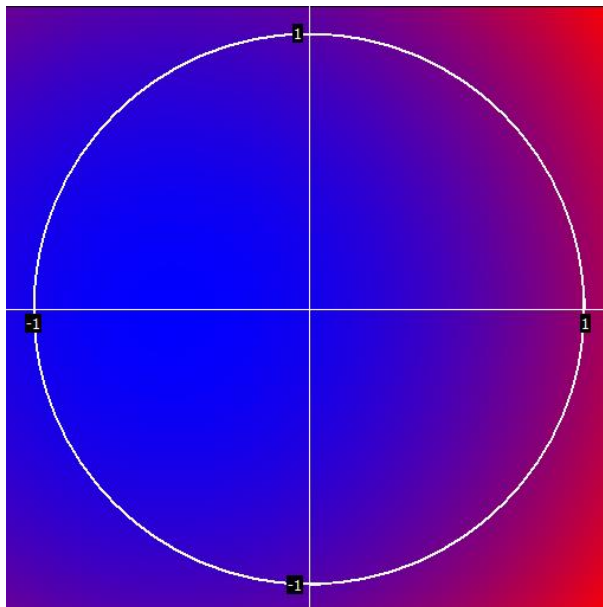
$$F\left(-\frac{1}{2}, 0\right) = -\frac{1}{4}.$$

Vi undersöker nu randen $x^2 + y^2 = 1$ (enhetscirkeln). Denna kan parametreras $\mathbf{r}(t) = (\cos t, \sin t)$ där $t \in [0, 2\pi[$, så vi vill hitta största och minsta värde av $g(t) := F(\cos t, \sin t) = \cos^2 t + \sin^2 t + \cos t = 1 + \cos t$ på $[0, 2\pi[$. Vi inser direkt att största värdet erhålles när $t = 0$,

d.v.s. i punkten $(1, 0)$, och detta värde är $g(0) = 2$, samt att minsta värdet erhålles när $t = \pi$, d.v.s. i punkten $(-1, 0)$, och detta värde är $g(\pi) = 0$.

Svar: Minsta värdet är $-\frac{1}{4}$ och erhålles (enbart) i $(-\frac{1}{2}, 0)$. Största värdet är 2 och erhålles (enbart) i $(1, 0)$.

Nedan visas ett färgat plan för F tillsammans med enhetsdisken.



8.3.3 Repetition av resonemanget

Vi har nu gett några inledande exempel på optimering av plana skalärfält F över kompakta mängder D . Tillvägagångssättet har alltid varit detsamma: först sätter vi skalärfältets gradient ∇F till noll för att finna inre stationära punkter, och sedan har vi undersökt randen ∂D (som ett nytt, separat, problem).

Hur vet vi att den här metoden fungerar? Vi har faktiskt gått igenom det ganska noggrant, men för säkerhets skull ger vi här en repetition av resonemanget.

Det resultat vi använder oss av är följande, där (x, y) är en godtycklig punkt i mängden:

$$(x, y) \text{ är en intressant punkt} \implies \begin{cases} (x, y) \text{ är en inre stationär punkt} \\ \text{eller} \\ (x, y) \text{ är en randpunkt.} \end{cases}$$

(Med "intressant punkt" menar vi en kandidat till extrempunkt.) För att bevisa denna implikation kan vi lika gärna bevisa dess kontrapositiva motsvarighet:

$$(x, y) \text{ är en inre punkt som inte är stationär} \implies (x, y) \text{ är en ointressant punkt.}$$

Men denna implikation är intuitivt "uppenbar", ty om (x, y) inte är en stationär punkt så är skalärfältets gradient $\nabla F(x, y)$ nollskild i punkten, så vi har en positiv riktningsderivata i gradientens riktning, t.ex. i riktningen $(1, 2)$. Då har vi också en lika stor negativ riktningsderivata i motsatta riktningen, t.ex. $(-1, -2)$. Så om vi går i riktningen $(1, 2)$ ökar "temperaturen", och om vi går i riktningen $(-1, -2)$ minskar "temperaturen". Därför innehåller varje omgivning av punkten både större och

mindre funktionsvärden. **Och eftersom punkten är en inre punkt så finns det en omgivning till punkten som ligger helt i mängden U .** Således finns både större och mindre funktionsvärden i U , så punkten är inte en global extrempunkt (eller ens en lokal sådan).

Ett lite mer handviftande (och betydligt sämre) resonemang är som följer: när vi sätter gradienten ∇F lika med noll så hittar vi skalärfältets egna bergstoppar- och dalar. De (av dessa) som ligger i mängden är förstas kandidater till att vara högsta respektive lägsta "höjden" i området vi undersöker. Sedan undersöker vi *hela* randen som en separat uppgift. Om vi t.ex. har en enskild bergstopp i mitten av området, så är denna högst upp, och lägsta höjden bör då inträffa någonstans på randen, där vi "hugger av" området.

8.3.4 Max och min på kurva i planet (igen)

Vi avslutar avsnittet om tvådimensionell optimering med att gå tillbaka till det enklaste fallet, när vi skall hitta max och min längs en *kurva* i planet. Anledningen är främst att författaren vill få en chans att kommentera ett av de allra vanligaste felen som studenter begår.

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = x^2 + y^2 + x$ på enhetscirkeln $S = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 1\}$.

Lösning: (F är kontinuerlig och S är kompakt, så största och minsta värde existerar.)

Enhetscirkeln kan parametriseras $(x, y) = (\cos t, \sin t)$ där $t \in [0, 2\pi[$, så vi vill hitta största och minsta värde av $g(t) := F(\cos t, \sin t) = \cos^2 t + \sin^2 t + \cos t = 1 + \cos t$ på $[0, 2\pi[$. Vi inser direkt att största värdet erhålles när $t = 0$, d.v.s. i punkten $(1, 0)$, och detta värde är $g(0) = 2$, samt att minsta värdet erhålles när $t = \pi$, d.v.s. i punkten $(-1, 0)$, och detta värde är $g(\pi) = 0$.

Svar: Minsta värdet är 0 och erhålles (enbart) i $(-1, 0)$. Största värdet är 2 och erhålles (enbart) i $(1, 0)$.

Det här ser man inte sällan när man rättar skrivningar:

Ett mycket vanligt fel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = x^2 + y^2 + x$ på enhetscirkeln $S = \{(x, y) \in \mathbb{R}^2: x^2 + y^2 = 1\}$.

Lösning: F är kontinuerlig och S är kompakt, så största och minsta värde existerar.

Vi bestämmer först de inre stationära punkterna. Gradienten

$$\nabla F(x, y) = (2x + 1, 2y) = \mathbf{0} \quad \iff \quad (x, y) = \left(-\frac{1}{2}, 0\right)$$

men den ligger inte på enhetscirkeln.

Vi undersöker nu randen. Enhetscirkeln kan parametriseras $(x, y) = (\cos t, \sin t)$ där $t \in [0, 2\pi[$, så vi vill hitta största och minsta värde av $g(t) := F(\cos t, \sin t) = \cos^2 t + \sin^2 t + \cos t = 1 + \cos t$ på $[0, 2\pi[$. Vi inser direkt att största värdet erhålles när $t = 0$, d.v.s. i punkten $(1, 0)$, och detta värde

är $g(0) = 2$, samt att minsta värdet erhålles när $t = \pi$, d.v.s. i punkten $(-1, 0)$, och detta värde är $g(\pi) = 0$.

Svar: Minsta värdet är 0 och erhålles (endast) i $(-1, 0)$. Största värdet är 2 och erhålles (endast) i $(1, 0)$.

Här börjar studenten med att sätta gradienten till nollvektorn, vilket är helt onödigt. Kom ihåg att vi sätter gradienten till noll för att hitta de *inre* stationära punkterna. Dessa punkter, tillsammans med randpunkterna, är ju de enda kandidaterna till max- och minpunkter. Men en kurva i planet, som t.ex. enhetscirkeln, har ju inga inre punkter (i stället är *varje* punkt en randpunkt)!! Att leta efter stationära inre punkter när det inte finns några inre punkter är som att leta efter en blå Volvo i en Saab-affär. Och sedan, när vi undersöker "randen" så undersöker vi ju faktiskt *hela* mängden, så det är klart att allt vi gjort tidigare har varit gjort i onödan!!

Man kan också se det så här: När vi sätter gradienten till noll hittar vi F :s alla stationära punkter – i synnerhet alla max och min – i hela planet. De allra flesta av dessa ligger förstås inte på enhetscirkeln. Visst, det *kan* hända att någon av dem faktiskt ligger på enhetscirkeln, men denna punkt kommer vi ju ändå att upptäcka i nästa steg när vi undersöker F på enhetscirkeln. Och vi måste naturligtvis undersöka F på enhetscirkeln eftersom "en cykelbana kan ha en högsta punkt som inte ligger på en bergstopp" (jfr avsnitt 7.5.2.2).

8.4 Exempel med optimering i rummet

Optimering i rummet fungerar på samma sätt som optimering i planet. Ett optimeringsproblem i rummet har två ingredienser:

1. ett skalärfält $F: D_F \rightarrow \mathbb{R}$ där $D_F \subset \mathbb{R}^3$
2. ett område $U \subset D_F$ i rummet.

Uppgiften är att, om möjligt, bestämma största och minsta värdet av F på U samt ange var på U dessa värden antages.

8.4.1 Max och min på kurva i rummet

Det lättaste är om mängden U är en kurva i rummet; via parametrisering erhåller vi då ett endimensionellt optimeringsproblem.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = zx^2 - 2yx + 4x - 3y$ på $\mathbf{r}([0, 1])$ där $\mathbf{r}(t) = (t, t^2, t^2)$.

Lösning: (F är kontinuerlig och mängden är kompakt, så max och min existerar.)

Funktionsvärdet vid parametervärdet t är

$$g(t) := t^4 - 2t^3 - 3t^2 + 4t$$

och derivatan

$$g'(t) = 4t^3 - 6t^2 - 6t + 4 = 0 \quad \iff \quad t \in \left\{-1, \frac{1}{2}, 2\right\}.$$

Av dessa stationära punkter ligger endast $t = \frac{1}{2}$ i intervallet $[0,1]$. Denna punkt svarar mot punkten $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ i rummet, och här är funktionsvärdet

$$g\left(\frac{1}{2}\right) = \frac{17}{16}.$$

I ändpunkterna $(0,0,0)$ respektive $(1,1,1)$ är

$$g(0) = 0, \quad g(1) = 0.$$

Svar: Minsta värdet är 0 och erhålles (endast) i $(0,0,0)$ och $(1,1,1)$. Största värdet är $\frac{17}{16}$ och erhålles (endast) i $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

8.4.2 Max och min på yta i rummet

Problemet att hitta max och min på en yta i rummet reduceras via parametrisering av ytan till problemet att hitta max och min på ett tvådimensionellt område, något vi redan behärskar.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x + y - z$ på $\Sigma := \{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2, z \leq 1\}$.

Lösning: Σ är den del av paraboliden $z = x^2 + y^2$ där $z \leq 1$. (Denna mängd är kompakt, och F är kontinuerlig, så största och minsta värde existerar.)

Paraboloden kan parametriseras med x och y som parametrar, d.v.s. via $(x, y, z) = (x, y, x^2 + y^2)$. Parameterområdet är då den slutna enhetsdisken $D \subset \mathbb{R}^2$. Vi skall alltså hitta max/min av

$$G(x, y) = x + y - x^2 - y^2$$

på D . G är kontinuerlig och D är kompakt, så största och minsta värde existerar, och kan endast antagas i inre stationära punkter samt på randen.

Gradienten

$$\nabla G(x, y) = (1 - 2x, 1 - 2y) = \mathbf{0} \quad \iff \quad (x, y) = \left(\frac{1}{2}, \frac{1}{2}\right)$$

som tillhör D . I denna punkt, vilken svarar mot $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}) \in \Sigma$, är funktionsvärdet

$$G(x, y) = \frac{1}{2}.$$

Randen ∂D är enhetscirkeln som kan parametriseras $(x, y) = (\cos t, \sin t)$. Funktionsvärdet vid parametervärdet t är

$$g(t) := \cos t + \sin t - \cos^2 t - \sin^2 t = \cos t + \sin t - 1 = \sqrt{2} \sin\left(t + \frac{\pi}{4}\right) - 1$$

och vi ser direkt att största värdet är $\sqrt{2} - 1$ och erhålles då $t = \pi/4$, vilket svarar mot punkten $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \in \partial D$, som i sin tur svarar mot $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1) \in \Sigma$; minsta värdet är $-\sqrt{2} - 1$ och erhålles då $t = -3\pi/4$, vilket svarar mot punkten $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) \in \partial D$, som i sin tur svarar mot $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 1) \in \Sigma$.

Svar: Minsta värdet är $-\sqrt{2} - 1$ och erhålles (endast) i $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 1)$. Största värdet är $\frac{1}{2}$ och erhålles (endast) i $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x^2 + 2y^2 + z^2 - z$ på enhetssfären.

Lösning: (Enhetssfären är kompakt och F är kontinuerlig, så största och minsta värde existerar.)

Enhetssfären är bilden $\mathbf{r}([0, \pi] \times [0, 2\pi[)$ av rektangeln $[0, \pi] \times [0, 2\pi[$ under

$$\mathbf{r}(\theta, \varphi) = \underline{\mathbf{e}} \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}.$$

Funktionsvärdet vid (θ, φ) är

$$G(\theta, \varphi) := F(\mathbf{r}(\theta, \varphi)) = 1 + \sin^2 \theta \sin^2 \varphi - \cos \theta$$

och vi skall alltså hitta max och min av $G(\theta, \varphi)$ på $[0, \pi] \times [0, 2\pi[$. Gradienten

$$\begin{aligned} \nabla G(\theta, \varphi) &= (2 \sin \theta \cos \theta \sin^2 \varphi + \sin \theta, 2 \sin^2 \theta \sin \varphi \cos \varphi) = \mathbf{0} && \iff \\ &\iff \begin{cases} \sin \theta (2 \cos \theta \sin^2 \varphi + 1) = 0 \\ \sin^2 \theta \sin \varphi \cos \varphi = 0. \end{cases} \end{aligned}$$

Om $\sin \theta = 0$ så befinner vi oss på randen till rektangeln i parameterplanet, och (faktiskt!) i de två polerna på sfären; här är funktionsvärdena $F(0,0,1) = 0$ respektive $F(0,0,-1) = 2$.

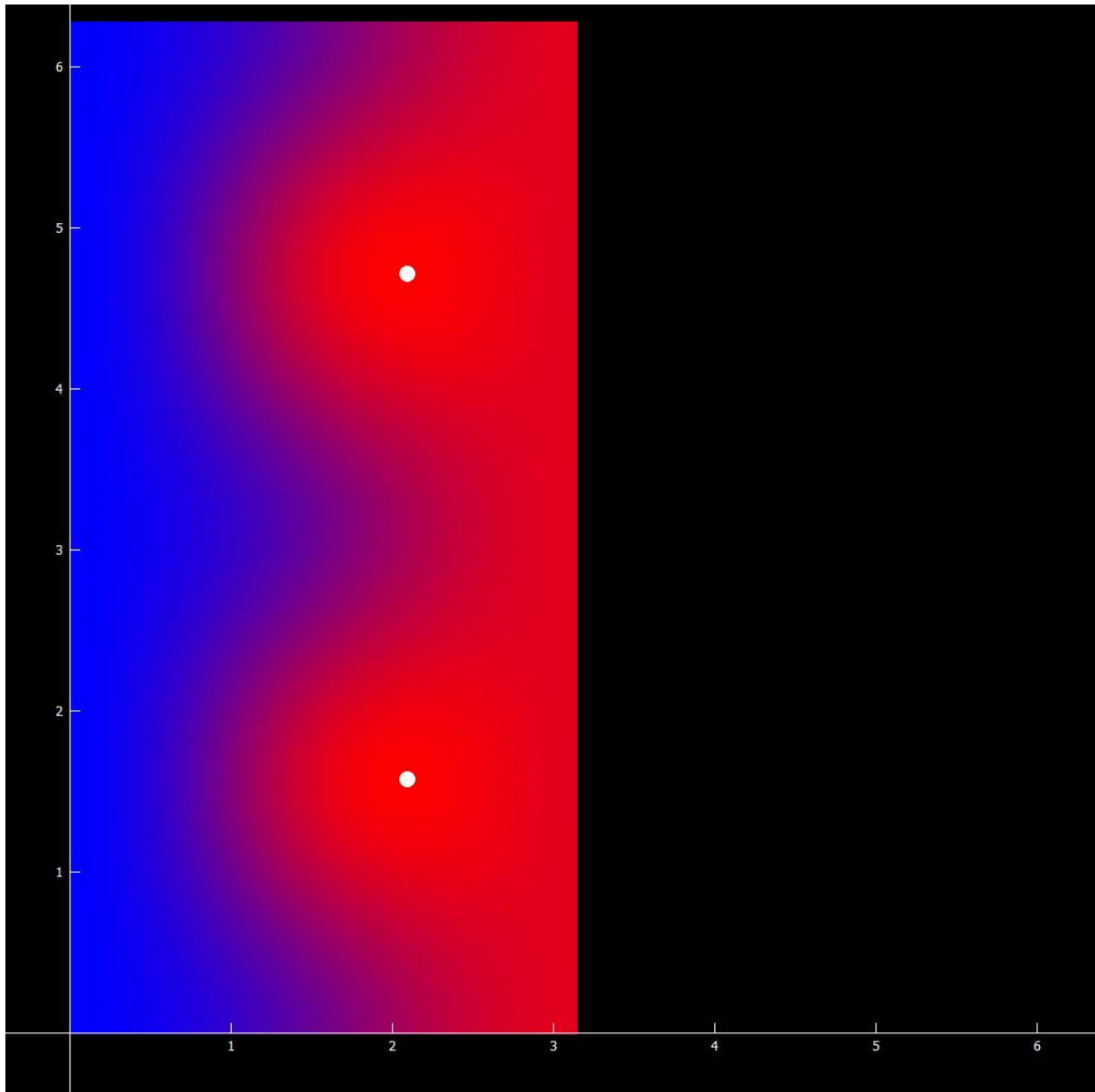
Om $\sin \theta \neq 0$ så är

$$\begin{cases} 2 \cos \theta \sin^2 \varphi + 1 = 0 \\ \sin \varphi \cos \varphi = 0. \end{cases}$$

Sista ekvationen kräver att φ är en heltalsmultipel av $\pi/2$, så vi får $\varphi \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Första ekvationen ger en motsägelse, emellertid, i fallen $\varphi \in \{0, \pi\}$ så i själva verket återstår bara $\{\frac{\pi}{2}, \frac{3\pi}{2}\}$. I båda fallen ger första ekvationen $\cos \theta = -1/2$, d.v.s. $\theta = 2\pi/3$. Vi har därför funnit de två inre stationära punkterna $(\frac{2\pi}{3}, \frac{\pi}{2})$ och $(\frac{2\pi}{3}, \frac{3\pi}{2})$. Dessa svarar mot punkterna $(0, \pm \frac{\sqrt{3}}{2}, -\frac{1}{2})$ på enhetssfären och här är funktionsvärdena $G(\frac{2\pi}{3}, \frac{\pi}{2}) = G(\frac{2\pi}{3}, \frac{3\pi}{2}) = \frac{9}{4} (> 2)$.

Svar: Funktionen minsta värde är 0 och erhålles (endast) i $(0,0,1)$. Funktionen största värde är $\frac{9}{4}$ och erhålles (endast) i $(0, \pm \frac{\sqrt{3}}{2}, -\frac{1}{2})$.

Nedan visas ett färgat plan för G . Nordpolen svarar mot $\theta = 0$, och här är "temperaturen" 0; sydpolen svarar mot $\theta = \pi$ och här är "temperaturen" 2.



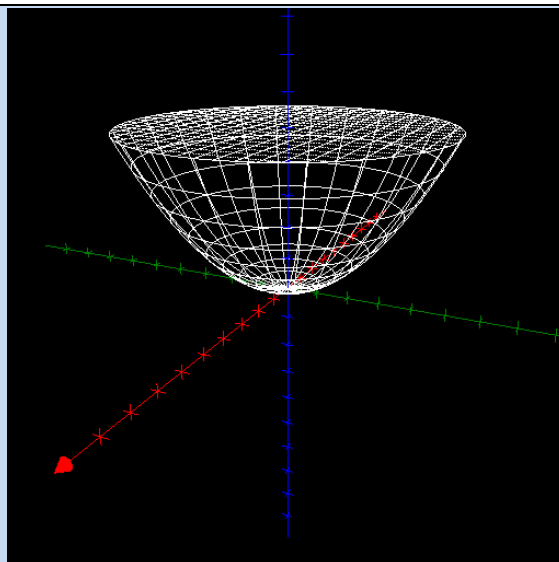
8.4.3 Max och min på tredimensionellt område i rummet

Nu kommer vi till "riktig" tredimensionell optimering.

Exempel

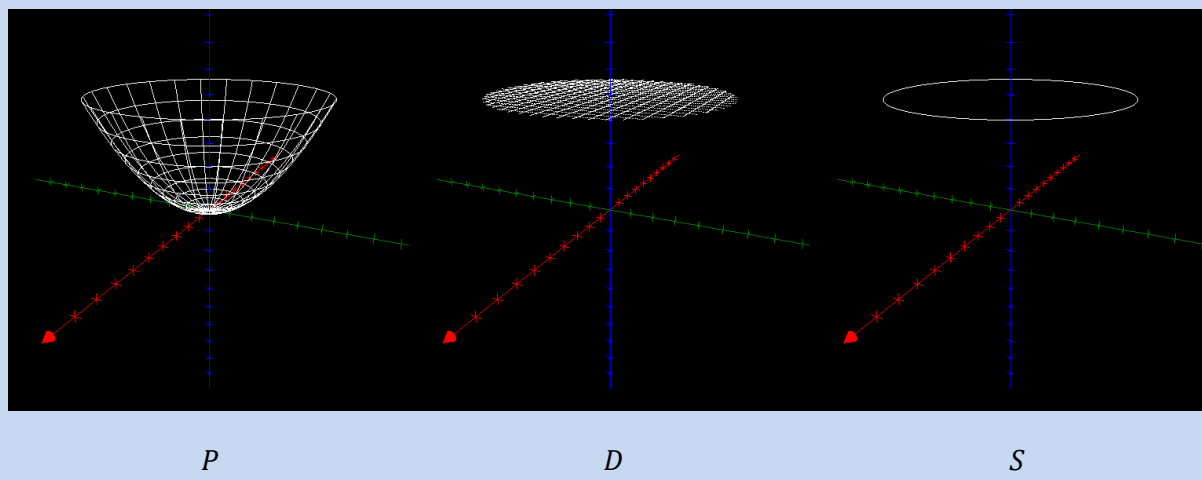
Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x^2 + y + z$ på $U := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq z \leq 1\}$.

Lösning: U är den del av rummet som ligger "innanför eller på" paraboloiden $z = x^2 + y^2$ och "under eller på" planet $z = 1$.



Mängden U är kompakt och F kontinuerlig, så största och minsta värde existerar och kan endast erhållas i inre stationära punkter eller på randen. Gradienten $\nabla F(x, y, z) = (2x, 1, 1)$ är aldrig lika med nollvektorn, så stationära punkter saknas.

Det återstår att bestämma största och minsta värdet av F på randen ∂U av området. ∂U kan på ett naturligt sätt delas in i två delar: $\partial U = P \cup D$ där $P = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 = z, z \leq 1\}$ är paraboloidbiten och $D = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 \leq 1, z = 1\}$ är disken som utgör områdets "tak". Notera att dessa *inte* är disjunkta, utan att de har cirkeln $S = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 = 1, z = 1\}$ som en gemensam "rand"⁹. Vi bestämmer kandidater till max och min på P och D separat; unionen av dessa kandidater utgör då en fullständig samling kandidater till det ursprungliga problemet att hitta max och min på *hela* randen.¹⁰



⁹ Notera att vi här *inte* använder ordet "rand" i dess vanliga betydelse som vi använt tidigare. (Till exempel är ju $\partial P = P$ och $\partial D = D$.) Däremot bör läsaren inse det naturliga i att kalla S för "randen" till P (och till D). Vi kommer inte att ge någon precis definition av det här nya randbegreppet, emellertid. Däremot bör läsaren notera att S är bilden av randen (i vanlig mening) till *parameterområdet* i respektive parametrisering av P och D som vi ger inom kort.

¹⁰ För att detta skall fungera kan vi mycket väl kräva att P och U är disjunkta, och därmed ta bort S från endera mängden. Men det gör vi inte, eftersom vi tycker det är trevligt att P och D är *kompakta* mängder. Däremot behöver vi i praktiken inte undersöka S två gånger.

Parabelbiten

Vi skall bestämma max och min av F på P . Vi börjar då, som alltid, med att parametrisera P . Vi väljer att använda x och y som koordinater, så att $P = \mathbf{r}(\tilde{P})$ där $\mathbf{r}: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ är definierad av $\mathbf{r}(x, y) = (x, y, x^2 + y^2)$ och där $\tilde{P} \subset \mathbb{R}^2$ är den slutna enhetsdisken i planet. Vi erhåller ($G := F \circ \mathbf{r}$) funktionsvärdena

$$G(x, y) = F(x, y, x^2 + y^2) = 2x^2 + y + y^2.$$

Vi skall alltså hitta max och min av $G(x, y)$ på \tilde{P} . Gradienten

$$\nabla G(x, y) = (4x, 1 + 2y) = \mathbf{0} \quad \iff \quad (x, y) = \left(0, -\frac{1}{2}\right).$$

Den här stationära punkten tillhör \tilde{P} och svarar mot punkten $\left(0, -\frac{1}{2}, \frac{1}{4}\right) \in U$. Här är funktionsvärdet

$$G\left(0, -\frac{1}{2}\right) = -\frac{1}{4}.$$

Vi behöver också undersöka randen $\partial\tilde{P}$, d.v.s. enhetscirkeln, separat. Bilden av denna är cirkeln S , som vi väljer att vänta med att undersöka.

Disken

Vi skall bestämma max och min av F på D . Vi börjar då, som alltid, med att parametrisera D . Vi väljer att använda x och y som parametrar, så att $D = \mathbf{p}(\tilde{D})$ där $\mathbf{p}: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ är definierad av $\mathbf{p}(x, y) = (x, y, 1)$ och där \tilde{D} är den slutna enhetsdisken. Vi erhåller ($H := F \circ \mathbf{p}$) funktionsvärdena

$$H(x, y) = F(x, y, 1) = x^2 + y + 1.$$

Vi skall alltså hitta max/min av $H(x, y)$ på \tilde{D} . Gradienten $\nabla H(x, y) = (2x, 1)$ som aldrig är noll; H har alltså inga stationära punkter.

Vi behöver också undersöka randen $\partial\tilde{D}$, d.v.s. enhetscirkeln, separat. Bilden av denna är återigen cirkeln S , som vi nu undersöker.¹¹

Enhetscirkeln kan parametriseras $(x, y) = (\cos t, \sin t)$ där $t \in [0, 2\pi[$ och funktionsvärdet vid parametervärdet t är

$$g(t) := H(\cos t, \sin t) = \cos^2 t + \sin t + 1.$$

Vi behöver hitta max/min av g på $[0, 2\pi[$. Derivatn

$$\begin{aligned} g'(t) = -2 \cos t \sin t + \cos t = 0 & \iff \cos t = 0 \quad \text{eller} \quad \sin t = \frac{1}{2} & \iff \\ \iff t \in \left\{ \frac{\pi}{2}, \frac{3\pi}{2}, \frac{\pi}{6}, \frac{5\pi}{6} \right\}. & & \end{aligned}$$

¹¹ Anledningen till att vi *inte* undersökte randen $\partial\tilde{P}$ förra gången var alltså att dess bild också är S , och det är onödigt att undersöka S två gånger, som nämnt ovan!

Dessa punkter svarar mot $(0, 1)$, $(0, -1)$, $(\frac{\sqrt{3}}{2}, \frac{1}{2})$ respektive $(-\frac{\sqrt{3}}{2}, \frac{1}{2})$ i \tilde{D} , d.v.s. mot $(0, 1, 1)$, $(0, -1, 1)$, $(\frac{\sqrt{3}}{2}, \frac{1}{2}, 1)$ respektive $(-\frac{\sqrt{3}}{2}, \frac{1}{2}, 1)$ i U . Funktionsvärdena är

$$g\left(\frac{\pi}{2}\right) = 2, \quad g\left(\frac{3\pi}{2}\right) = 0, \quad g\left(\frac{\pi}{6}\right) = \frac{9}{4}, \quad g\left(\frac{5\pi}{6}\right) = \frac{9}{4}.$$

Jämförelse av kandidater

Det är klart att $-\frac{1}{4} < 0 < 2 < \frac{9}{4}$.

Svar: Funktionen minsta värde är $-\frac{1}{4}$ och detta antages (endast) i $(0, -\frac{1}{2}, \frac{1}{4})$. Funktionen största värde är $\frac{9}{4}$ och detta antages (endast) i $(\pm\frac{\sqrt{3}}{2}, \frac{1}{2}, 1)$.

Redovisningen ovan är väldigt noggrann med att belysa den grundläggande teorin bakom optimering på kompakt mängd. Dessutom finns en massa pedagogiska förklaringar insprängda. I praktiken är den något behändigare redovisningen nedan lättare att läsa:

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x^2 + y + z$ på $U := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq z \leq 1\}$.

Lösning: Mängden U är kompakt och F kontinuerlig, så största och minsta värde existerar säkert.

Eftersom $\nabla F(x, y, z) = (2x, 1, 1)$ aldrig är noll saknas inre stationära punkter. Vi undersöker därför randen $\partial D = P \cup D$ där $P = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = z, z \leq 1\}$ är paraboloidbiten och $D = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, z = 1\}$ är disken som utgör området "tak". P och D har en gemensam "rand" i cirkeln $x^2 + y^2 = 1$ i planet $z = 1$.

Paraboloidbiten

Paraboloidbiten kan parametreras med x och y via $(x, y, z) = (x, y, x^2 + y^2)$ där (x, y) tillhör den slutna enhetsdisken. Här är $F(x, y, x^2 + y^2) = 2x^2 + y + y^2$ med gradienten $(4x, 1 + 2y) = \mathbf{0}$ omm $(x, y) = (0, -\frac{1}{2})$ som ligger i enhetsdisken och svarar mot punkten $(0, -\frac{1}{2}, \frac{1}{4})$. Här är funktionsvärdet $-\frac{1}{4}$.

Disken

På disken är $F(x, y, 1) = x^2 + y + 1$ med gradienten $(2x, 1)$ som aldrig är noll, så stationära punkter saknas i planet $z = 1$.

Cirkeln

Cirkeln kan parametreras $(x, y, z) = (\cos t, \sin t, 1)$ och funktionsvärdet vid parametervärdet t är $g(t) := \cos^2 t + \sin t + 1$ med derivatan $g'(t) = -2 \cos t \sin t + \cos t = 0$ omm $\cos t = 0$ eller

$\sin t = \frac{1}{2}$ d.v.s. omm $t \in \left\{ \frac{\pi}{2}, \frac{3\pi}{2}, \frac{\pi}{6}, \frac{5\pi}{6} \right\}$ vilket svarar mot punkterna $(0, 1, 1)$, $(0, -1, 1)$, $\left(\frac{\sqrt{3}}{2}, \frac{1}{2}, 1\right)$ respektive $\left(-\frac{\sqrt{3}}{2}, \frac{1}{2}, 1\right)$ i U . Funktionsvärdena är

$$g\left(\frac{\pi}{2}\right) = 2, \quad g\left(\frac{3\pi}{2}\right) = 0, \quad g\left(\frac{\pi}{6}\right) = \frac{9}{4}, \quad g\left(\frac{5\pi}{6}\right) = \frac{9}{4}.$$

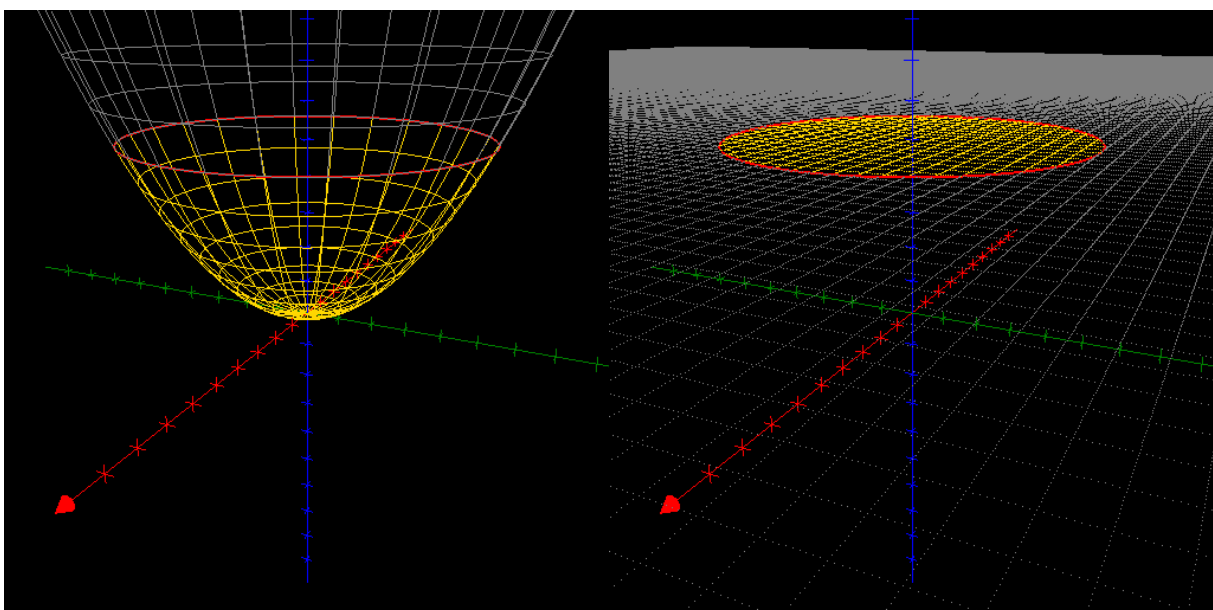
Det är klart att $-\frac{1}{4} < 0 < 2 < \frac{9}{4}$.

Svar: Funktionen minsta värde är $-\frac{1}{4}$ och detta antages (endast) i $\left(0, -\frac{1}{2}, \frac{1}{4}\right)$. Funktionen största värde är $\frac{9}{4}$ och detta antages (endast) i $\left(\pm\frac{\sqrt{3}}{2}, \frac{1}{2}, 1\right)$.

Förfarandet är tämligen naturligt. När vi sätter $\nabla F(x, y, z) = \mathbf{0}$ letar vi efter F :s egna stationära punkter, d.v.s. F :s riktiga "toppar och dalar" i hela rummet \mathbb{R}^3 . Eventuella stationära punkter utanför vårt område är vi inte intresserade av, så dessa förkastar vi. (I det här exemplet fanns inga stationära punkter alls.)

Men sedan måste vi också undersöka där vi hugger av området, d.v.s. på randen. Eftersom randen naturligt delas in i en paraboloid och en disk så undersöker vi dessa separat. När vi parametriserat paraboloiden och skall hitta max/min av funktionsvärdet på paraboloiden så sätter vi denna tvådimensionella gradient till noll, varvid vi hittar (bland annat) alla lokala max och min på paraboloiden. Eftersom vi bara är intresserade av den del av paraboloiden som ligger "under" $z = 1$ så förkastar vi eventuella kandidater som ligger utanför vår lilla del av paraboloiden. På samma sätt, när vi undersöker funktionsvärdena i planet $z = 1$ genom att sätta den tvådimensionella gradienten till noll så får vi (bland annat) alla lokala max och min i hela det planet, och återigen får vi förkasta eventuella kandidater utanför vårt område, d.v.s. utanför $x^2 + y^2 = 1$.

Slutligen måste vi undersöka cirkeln $x^2 + y^2 = 1, z = 1$ som är den gemensamma "randen" både till paraboloidbiten och till disken. Här hugger vi ju av dessa områden.



Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x + y - z^2$ på $U = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 \leq 1, z \geq 0\}$.

Lösning: U är (slutna) övre halvan av (slutna) enhetsklotet, som är en kompakt mängd. Samtidigt är F kontinuerlig, så största och minsta värde existerar.

Vi letar först efter inre stationära punkter. Gradienten $\nabla F(x, y, z) = (1, 1, -2z)$ är emellertid alltid nollskild, så F har inga stationära punkter i hela rummet.

Randen $\partial U = S' \cup D$ där $S' = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 = 1, z > 0\}$ är (slutna) övre halvan av enhetsfären och $D = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 \leq 1, z = 0\}$ är disken som utgör golvet. Dessa har en gemensam "rand" i cirkeln $S = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 = 1, z = 0\}$.

Halvsfären

Övre halvan av enhetsfären kan parametreras med x och y som parametrar.¹² Halvsfären är alltså bilden av slutna enhetsdisken i planet under $(x, y) \mapsto (x, y, \sqrt{1 - x^2 - y^2})$. Funktionsvärdet här är

$$G(x, y) := x + y - \left(\sqrt{1 - x^2 - y^2}\right)^2 = x^2 + x + y^2 + y - 1$$

och gradienten $\nabla G(x, y) = (2x + 1, 2y + 1) = \mathbf{0}$ omm $(x, y) = \left(-\frac{1}{2}, -\frac{1}{2}\right)$ som tillhör slutna enhetsdisken. Punkten, som svarar mot $\left(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{\sqrt{2}}\right)$, ger funktionsvärdet $G\left(-\frac{1}{2}, -\frac{1}{2}\right) = -\frac{3}{2}$.

(Vi behöver också undersöka G på randen till enhetsdisken, men bilden av denna rand (enhetscirkeln) sammanfaller med bilden av randen till disken D :s parameterområde, så vi väntar med den undersökningen.)

Disken

I planet $z = 0$ är funktionsvärdena $F(x, , 0) = x + y$ som uppenbarligen saknar stationära punkter. (Vi behöver bara inre stationära punkter i *enhetsdisken*, som är parameterområdet under den uppenbara parametreringen $(x, y, z) = (x, , 0)$ av D . Sedan behöver vi undersöka $x + y$ på randen av parameterområdet, vilket vi gör härnäst.)

Cirkeln

S kan parametreras $(x, y, z) = (\cos t, \sin t, 0)$ och vid parametervärdet t är funktionsvärdet $g(t) := F(\cos t, \sin t, 0) = \cos t + \sin t = \sqrt{2} \sin\left(t + \frac{\pi}{4}\right)$ som antar sitt största värde vid $t = \frac{\pi}{4}$ och sitt minsta värde vid $t = \frac{5\pi}{4}$. Dessa punkter svarar mot $\pm\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right) \in U$ och här är funktionsvärdena $g\left(\frac{\pi}{4}\right) = \sqrt{2}$ och $g\left(\frac{5\pi}{4}\right) = -\sqrt{2}$.

¹² Det är också lämpligt att välja just denna parametrering i det här fallet (och inte den trigonometriska parametreringen) med tanke på kvadraten på z i skalärfältets uttryck.

Jämförelse

Vi noterar slutligen att $\frac{3}{2} > \sqrt{2}$ ty $\left(\frac{3}{2}\right)^2 = \frac{9}{4} > \frac{8}{4} = 2 = (\sqrt{2})^2$. Alltså är $-\frac{3}{2} < -\sqrt{2}$.

Svar: Funktionen minsta värde är $-\frac{3}{2}$ och detta antages (endast) i $\left(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{\sqrt{2}}\right)$. Funktionen största värde är $\sqrt{2}$ och detta antages (endast) i $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x^2 + 2y^2 + z^2 - z$ på det slutna enhetsklotet.

Lösning: Det slutna enhetsklotet är kompakt och F är kontinuerlig, så största och minsta värde existerar. Gradienten

$$\nabla F(x, y, z) = (2x, 4y, 2z - 1) = \mathbf{0} \quad \iff \quad (x, y, z) = \left(0, 0, \frac{1}{2}\right)$$

som tillhör enhetsklotet. Här är funktionsvärdet

$$F\left(0, 0, \frac{1}{2}\right) = -\frac{1}{4}.$$

Vi behöver nu hitta max och min av F på enhetssfären. Men detta har vi gjort i en tidigare uppgift, och vi fann då att funktionens minsta värde på enhetssfären är 0 och erhålles (endast) i $(0, 0, 1)$.

Funktionens största värde på enhetssfären är $\frac{9}{4}$ och erhålles (endast) i $\left(0, \pm\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)$.

Svar: Funktionen minsta värde är $-\frac{1}{4}$ och erhålles (endast) i $\left(0, 0, \frac{1}{2}\right)$. Funktionen största värde är $\frac{9}{4}$ och erhålles (endast) i $\left(0, \pm\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)$.

8.4.4 Max och min på yta i rummet (igen)

Vi avslutar det här kapitlet med att ge ytterligare ett exempel på optimering på en yta i rummet. Anledningen är främst att författaren vill få en chans att kommentera ett av de allra vanligaste felen som studenter begår.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x^2 - y^2 + 2z$ på den del av planet $x + 2y - z = 1$ som ligger innanför pelaren $[0, 1] \times [0, 1] \times \mathbb{R}$.

Lösning: (Mängden är kompakt och funktionen är kontinuerlig, så största och minsta värde existerar.)

Vi parametriserar ytan med x och y som parametrar, så att den är bilden $\mathbf{r}([0, 1] \times [0, 1])$ av rektangeln $[0, 1] \times [0, 1]$ under $\mathbf{r}(x, y) = (x, y, x + 2y - 1)$. Funktionsvärdet vid (x, y) är då

$$G(x, y) := x^2 - y^2 + 2x + 4y - 2$$

Gradienten

$$\nabla G(x, y) = (2 + 2x, 4 - 2y) = \mathbf{0} \iff (x, y) = (-1, 2)$$

som ligger utanför området. Vi undersöker då randen av $[0,1] \times [0,1]$, som naturligt delas in i fyra linjestycken. Vi börjar med $[0,1] \times \{0\}$. Med x som parameter är funktionsvärdet här $x^2 + 2x - 2 = (x + 1)^2 - 3$ som har sin enda stationära punkt då $x = -1$ vilket ligger utanför vårt område. På $[0,1] \times \{1\}$ är funktionsvärdet $x^2 + 2x + 1 = (x + 1)^2$ som också saknar stationär punkt på $[0,1]$. På $\{0\} \times [0,1]$ är funktionsvärdet $-y^2 + 4y - 2 = -(y - 2)^2 + 2$ som har sin enda stationära punkt då $y = 2$, d.v.s. utanför $[0,1]$. Slutligen, på linjestycket $\{1\} \times [0,1]$ har vi värdet $-y^2 + 4y + 1 = -(y - 2)^2 + 5$ med samma slutsats.

Vi undersöker till sist hörnen (där vi hugger av linjestyckena). Här är

$$G(0,0) = -2, \quad G(1,0) = 1, \quad G(0,1) = 1, \quad G(1,1) = 4.$$

Dessa punkter svarar mot $(0,0,-1)$, $(1,0,0)$, $(0,1,1)$ respektive $(1,1,2)$.

Svar: Funktionen minsta värde är -2 och detta antages (endast) i punkten $(0,0,-1)$. Funktionen största värde är 4 och detta antages (endast) i punkten $(1,1,2)$.

Inte sällan ser man följande "tokighet" när man rättar skrivningar:

Ett mycket vanligt fel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x^2 - y^2 + 2z$ på den del av planet $x + 2y - z = 1$ som ligger innanför pelaren $[0,1] \times [0,1] \times \mathbb{R}$.

Lösning: Mängden är kompakt och funktionen är kontinuerlig, så största och minsta värde existerar.

Vi letar först efter inre stationära punkter. Gradienten $\nabla F(x, y, z) = (2x, -2y, 2)$ som aldrig är noll, så stationära punkter saknas.

Vi undersöker nu randen. Vi parametriserar ytan med x och y som parametrar...

Här börjar studenten med att sätta gradienten till nollvektorn, vilket är helt onödigt. Kom ihåg att vi sätter gradienten till noll för att hitta de *inre* stationära punkterna. Dessa punkter, tillsammans med randpunkterna, är ju de enda kandidaterna till max- och minpunkter. Men en yta i rummet, som t.ex. den i exemplet, har ju inga inre punkter (i stället är *varje* punkt en randpunkt)!! Att leta efter stationära inre punkter när det inte finns några inre punkter är som att leta efter en brun hund på ett katt-pensionat. Och sedan, när vi undersöker "randen" så undersöker vi ju faktiskt *hela* mängden, så det är klart att allt vi gjort tidigare har varit gjort i onödan!!

Man kan också se det så här: När vi sätter gradienten till noll hittar vi F 's alla stationära punkter – i synnerhet alla max och min – i hela rummet. De allra flesta av dessa ligger förstås inte på vår yta. Visst, det *kan* hända att någon eller några av dem faktiskt ligger på ytan, men dessa punkter kommer vi ju ändå att upptäcka i nästa steg när vi undersöker F på ytan. Och vi måste naturligtvis undersöka F på ytan eftersom högsta temperaturen på ytan inte måste vara högsta temperaturen i någon *tre*-dimensionell omgivning av punkten (jfr avsnitt 7.5.2.2).

9 Ett nytt sätt att optimera – med bivillkor

I det här kapitlet introducerar vi *en ny metod* för att lösa (stora delar av) optimeringsproblem. Vi löser samma sorts problem som i förra kapitlet, men med den nya metoden.

Den nya metoden används till att hitta max och min av ett skalärfält $F: D_F \rightarrow \mathbb{R}$, $D_F \subset \mathbb{R}^n$ på en delmängd av D_F som är snittet mellan en eller flera nivåmängder (till *andra* skalärfält), vanligen kurvor och ytor.

Till exempel kanske vi har ett skalärfält i planet, $F: \mathbb{R}^2 \rightarrow \mathbb{R}$, och vill hitta max och min på kurvan $x^4 + y^4 = 3$. Denna kurva är ju nivåkurvan $G(x, y) = 3$ till skalärfältet $G(x, y) = x^4 + y^4$, så vår nya metod kommer att fungera.

Eller så har vi ett skalärfält i rummet, $F: \mathbb{R}^3 \rightarrow \mathbb{R}$, och vill hitta max och min på ytan $x^2 + y^2 + 2z^2 = 5$ (en ellipsoid). Ytan är nivåytan $G(x, y, z) = 5$ till skalärfältet $G(x, y, z) = x^2 + y^2 + 2z^2$ så vi kan använda vår nya metod.

Som tredje exempel, säg att vi vill hitta max och min av $F: \mathbb{R}^3 \rightarrow \mathbb{R}$ på skärningskurvan mellan cylindern $3x^2 + y^2 = 1$ och planet $x + y = 0$. (Denna kurva är en ellips.) De två ytorna är nivåytorna $G(x, y, z) = 1$ och $H(x, y, z) = 0$ till skalärfälten $G(x, y, z) = 3x^2 + y^2$ respektive $H(x, y, z) = x + y$, så återigen kan vi använda vår nya metod.

Om vi tittar tillbaka på de exempel vi behandlade i förra avsnittet inser vi att nästan alla av dem åtminstone till största delen kan lösas med hjälp av den nya metoden: vi kan hitta max/min av plana skalärfält på kurvor i planet samt av skalärfält i rummet på ytor och kurvor i rummet. Enda kravet är att ytorna och kurvorna beskrivs av *ekvationer* (parametrisering duger inte), så att vi kan betrakta dem som nivåmängder till skalärfält.

Faktiskt är den nya metoden i många fall effektivare än den "naiva" metoden med parametrisering som vi använde konsekvent i förra kapitlet.

Nu när vi gjort reklam för den nya metoden med bivillkor, låt oss titta på hur den fungerar, genom att första bara citera det centrala resultatet (vi förutsätter att alla funktioner är \mathcal{C}^1):

Sats. Antag att $\mathbf{a} \in D_F \cap D_G$ är en *inre* punkt (till $D_F \cap D_G$) som optimerar $F(\mathbf{x})$ under bivillkoret $G(\mathbf{x}) = 0$, där D_F och $D_G \subset \mathbb{R}^n$. Då är $\nabla F(\mathbf{a}) \parallel \nabla G(\mathbf{a})$.

I en kandidat till max eller min är alltså skalärfältets gradient parallell med gradienten som kommer från bivillkoret. Men om vi har två eller flera bivillkor (som när vi skall hitta max och min längs en skärningskurva mellan två (nivå)ytor i rummet)?

Sats. Antag att $\mathbf{a} \in D_F \cap \bigcap_{k=1}^N D_{G_k}$ är en *inre* punkt (till $D_F \cap \bigcap_{k=1}^N D_{G_k}$) som optimerar $F(\mathbf{x})$ under de N stycken bivillkoren $G_k(\mathbf{x}) = 0$, där D_F och varje $D_{G_k} \subset \mathbb{R}^n$. Då är de $N + 1$ vektorerna $\nabla F(\mathbf{a})$, $\nabla G_1(\mathbf{a})$, ..., $\nabla G_N(\mathbf{a})$ linjärt beroende.

I en kandidat till max eller min är alltså skalärfältets gradient och samtliga gradienter som kommer från bivillkoren linjärt beroende.

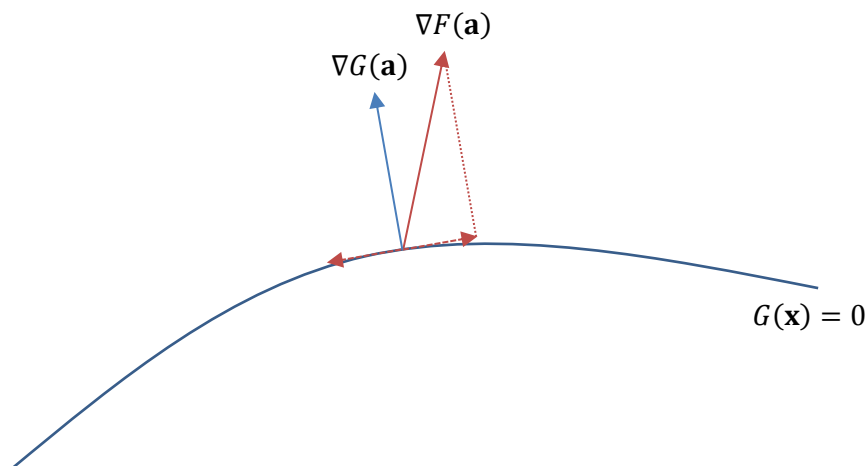
Övning

1. Visa att den första satsen är ett omedelbart specialfall av den andra. (D.v.s.: om två vektorer är linjärt beroende, så är de parallella.)

9.1 Geometrisk förklaring

Vi ger i det här avsnittet en mycket enkel geometrisk förklaring till den första satsen. Satsen säger att om \mathbf{a} är en inre punkt till $D_F \cap D_G$ som ger högsta eller lägsta värdet av $F(\mathbf{x})$ på $G(\mathbf{x}) = 0$, så måste $\nabla F(\mathbf{a}) \parallel \nabla G(\mathbf{a})$.

Vi troliggör satsen genom att troliggöra den *kontrapositiva* utsagan "Om $\nabla F(\mathbf{a}) \not\parallel \nabla G(\mathbf{a})$ så är \mathbf{a} inte en maximi- eller minimipunkt".¹³ Eftersom $\nabla F(\mathbf{a})$ inte är parallell med $\nabla G(\mathbf{a})$ så är $\nabla F(\mathbf{a})$ inte vinkelrät mot den nivå mängd (t.ex. nivåkurva eller nivåyta) som beskrivs av $G(\mathbf{x}) = 0$ i punkten \mathbf{a} . Därför är projektionen av $\nabla F(\mathbf{a})$ på nivå mängdens tangentplan (eller tangentlinje) en *nollskild* vektor. Sålunda har vi en positiv riktningsderivata i denna riktning, och en lika stor negativ riktningsderivata i motsatta riktningen. Alltså kan inte \mathbf{a} vara en extrempunkt, för längs $G(\mathbf{x}) = 0$ blir det "varmare" åt ena hållet och "kallare" åt andra hållet. I bilden nedan illustreras det plana fallet.



9.2 Om att uttrycka parallellitet och linjärt beroende

Kravet att två vektorer är parallella och kravet att N stycken vektorer är linjärt beroende är tydligen de centrala ingredienserna i den nya optimeringsmetoden. För att kunna lösa problem behöver vi därför *formulera* dessa krav på något sätt (t.ex. med hjälp av ekvationer). Detta bör vara välbekanta linjär algebra-kunskaper, men för säkerhets skull repeterar vi här de tre viktigaste insikterna.

Två vektorer \mathbf{u} och \mathbf{v} är, per definition, parallella om det finns ett tal $\lambda \in \mathbb{R}$ sådant att $\mathbf{u} = \lambda \mathbf{v}$ eller $\mathbf{v} = \lambda \mathbf{u}$.¹⁴ Detta skrives $\mathbf{u} \parallel \mathbf{v}$. I planet, d.v.s. om $\mathbf{u} = (u_x, u_y)$ och $\mathbf{v} = (v_x, v_y)$, så gäller

$$\mathbf{u} \parallel \mathbf{v} \iff \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = 0$$

¹³ Notera först att $\nabla F(\mathbf{a}) \not\parallel \nabla G(\mathbf{a})$ per definition (av begreppet "parallellitet") medför att båda vektorerna är nollskilda. Att $\nabla F(\mathbf{a})$ är nollskild betyder att vi *inte* befinner oss i en stationär punkt till F (trevligt!), och att $\nabla G(\mathbf{a})$ är nollskild medför (enligt implicita funktionssatsen) att nivå mängden $G(\mathbf{x}) = 0$ som innehåller \mathbf{a} i någon omgivning av \mathbf{a} kan parametriseras på ett kontinuerligt sätt i $n - 1$ av rummets koordinater.

¹⁴ Det går inte att formulera sig kortare genom att stryka "eller $\mathbf{v} = \lambda \mathbf{u}$ ". Varför? (Ledning: Nollvektorn är, per definition, parallell med varje vektor.)

(Läsaren kommer ihåg från linjära algebran att determinanten inte är något annat än \pm arean av den parallelogram som genereras av \mathbf{u} och \mathbf{v} , varför ekvivalensen är trivial.) I rummet, d.v.s. om $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, gäller att

$$\mathbf{u} \parallel \mathbf{v} \iff \mathbf{u} \times \mathbf{v} = \mathbf{0}.$$

Per definition är ju längden av kryssprodukten lika med $|\mathbf{u}||\mathbf{v}| \sin \alpha$ där α är vinkeln mellan faktorerna.

Tre vektorer $\mathbf{u} = (u_x, u_y, u_z)$, $\mathbf{v} = (v_x, v_y, v_z)$ och $\mathbf{w} = (w_x, w_y, w_z)$ i rummet är (per definition) linjärt beroende om det finns en icke-trivial lösning (λ, μ, ν) till ekvationen $\lambda\mathbf{u} + \mu\mathbf{v} + \nu\mathbf{w} = \mathbf{0}$, och detta är ekvivalent med att

$$\begin{vmatrix} u_x & u_y & u_z \\ v_x & v_y & v_z \\ w_x & w_y & w_z \end{vmatrix} = 0.$$

Att de tre vektorerna är linjärt beroende innebär ju precis att de ligger i samma plan, och determinanten ovan är inget annat än \pm volymen av den parallelepiped som de tre vektorerna genererar, så ekvivalensen är återigen uppenbar.¹⁵

Ett förvånansvärt vanligt fel

Ett förvånansvärt vanligt fel som studenter begår är att de skriver " $\mathbf{u} \parallel \mathbf{v} = \mathbf{0}$ " när de vill säga att \mathbf{u} och \mathbf{v} är parallella. Detta är fullkomligt obegripligt! $\mathbf{u} \parallel \mathbf{v}$ är ju en utsaga (d.v.s. ett påstående som antingen är sant eller falskt). Vad i all sin dar menas med att denna utsaga är lika med nollvektorn? Det är ungefär som att säga att "Katten är hemma är lika med hunden". Obegripligt.

9.3 Enkla exempel

Som uppvärmning ger vi ett par mycket enkla exempel.

Exempel. Säg att vi vill hitta största och minsta värdet av $F(x, y) = x$ på enhetscirkeln¹⁶. Enhetscirkeln $x^2 + y^2 = 1$ är nivåkurvan $G(x, y) = 1$ till skalärfältet $G(x, y) = x^2 + y^2$, så om (a, b) är en intressant punkt på cirkeln så är

$$\begin{aligned} \begin{cases} \nabla F(a, b) \parallel \nabla G(a, b) \\ G(a, b) = 1 \end{cases} &\iff \begin{cases} \begin{vmatrix} 1 & 0 \\ 2a & 2b \end{vmatrix} = 0 \\ a^2 + b^2 = 1 \end{cases} &\iff \\ &\iff \begin{cases} 2b = 0 \\ a^2 + b^2 = 1 \end{cases} &\iff (a, b) = (\pm 1, 0). \end{aligned}$$

I dessa punkter är $F(\pm 1, 0) = \pm 1$, så största värdet är $+1$ och detta antages (endast) i $(1, 0)$, medan minsta värdet är -1 och detta antages (endast) i $(-1, 0)$.

Exempel. Ofta vill man hitta max och min av ett skalärfält på *en del* av en kurva eller yta. I sådana fall kan bivillkorsmetoden (med parallella/linjärt beroende gradienter) användas för att hitta alla kandi-

¹⁵ Determinanten är faktiskt också lika med den så kallade *volymprodukten* $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$ (vilket gör det väldigt smidigt att testa linjärt (o)beroende i vissa datorsystem).

¹⁶ Självfallet är största värdet $+1$ (i $(1, 0)$) och minsta värdet -1 (i $(-1, 0)$). Exemplet är till för att illustrera den nya optimeringsmetoden.

dater *förutom* kurvans ändpunkter eller ytans "rand". Dessa punkter måste alltså undersökas separat (som vanligt). Säg att vi vill hitta max och min av $F(x, y) = x$ på $x^2 + y^2 = 1, x \geq 0$, d.v.s. på (slutna) *högra* halvan av enhetscirkeln. Vi använder då bivillkorsmaskineriet för att hitta kandidater då $x > 0$ (d.v.s. exkluderat ändpunkterna):

$$\begin{aligned} \begin{cases} \nabla F(a, b) \parallel \nabla G(a, b) \\ G(a, b) = 1 \\ a > 0 \end{cases} &\iff \begin{cases} \begin{vmatrix} 1 & 0 \\ 2a & 2b \end{vmatrix} = 0 \\ a^2 + b^2 = 1 \\ a > 0 \end{cases} &\iff \\ &\iff \begin{cases} 2b = 0 \\ a^2 + b^2 = 1 \\ a > 0 \end{cases} &\iff (a, b) = (1, 0) \end{aligned}$$

och i den enda kandidaten har vi $F(1,0) = 1$. I ändpunkterna är $F(0, -1) = 0$ och $F(0,1) = 0$. Så största värdet är $+1$ och detta antages (endast) i $(1,0)$, medan minsta värdet är 0 och detta antages (endast) i $(0, \pm 1)$.

9.4 Exempel

Vi belyser dels den nya metodens teknikaliteter, dels dess förträfflighet, genom att lösa om några av våra gamla exempelproblem med den nya metoden.

9.4.1 Max och min på kurva i planet

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = x + y^2$ på $y = x^2, -1 \leq x \leq 1$.

Lösning: Eftersom F är kontinuerlig och parabelbiten är kompakt så existerar största och minsta värde. Låt $G(x, y) := x^2 - y$ så att kurvan är nivåkurvan $G(x, y) = 0$. Låt (a, b) vara en intressant punkt som *inte* är en ändpunkt på parabelbiten. Då är

$$\begin{aligned} \begin{cases} \nabla F(a, b) \parallel \nabla G(a, b) \\ G(a, b) = 0 \end{cases} &\iff \begin{cases} \begin{vmatrix} 1 & 2b \\ 2a & -1 \end{vmatrix} = 0 \\ a^2 - b = 0 \end{cases} &\iff \\ &\iff \begin{cases} -1 - 4ab = 0 \\ b = a^2 \end{cases} &\iff \begin{cases} a^3 = -1/4 \\ b = a^2 \end{cases} &\iff \\ &\iff (a, b) = \left(-\frac{1}{\sqrt[3]{4}}, \frac{1}{\sqrt[3]{16}}\right) \end{aligned}$$

som vi ser ligger på vår del av parabeln. Här är

$$F\left(-\frac{1}{\sqrt[3]{4}}, \frac{1}{\sqrt[3]{16}}\right) = -\frac{1}{\sqrt[3]{4}} + \frac{1}{\sqrt[3]{256}}.$$

I ändpunkterna $x = -1$ och $x = 1$ har vi

$$g(-1) = 0, \quad g(1) = 2.$$

Vi noterar att

$$-\frac{1}{\sqrt[3]{4}} + \frac{1}{\sqrt[3]{256}} < 0 < 2$$

varvid vi är klara.

Svar: Funktionen minsta värde är $-\frac{1}{\sqrt[3]{4}} + \frac{1}{\sqrt[3]{256}}$ och detta antages (endast) i $(-\frac{1}{\sqrt[3]{4}}, \frac{1}{\sqrt[3]{16}})$. Funktionen största värde är 2 och detta antages (endast) i (1,1).

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = xy^3$ på enhetscirkeln.

Lösning: Funktionen är kontinuerlig och mängden är kompakt, så största och minsta värde existerar. Inför $G(x, y) := x^2 + y^2$ så att enhetscirkeln är nivåkurvan $G(x, y) = 1$. Låt (a, b) vara en intressant punkt. Då är

$$\begin{aligned} \begin{cases} \nabla F(a, b) \parallel \nabla G(a, b) \\ G(a, b) = 1 \end{cases} &\iff \begin{cases} \begin{vmatrix} b^3 & 3ab^2 \\ 2a & 2b \end{vmatrix} = 0 \\ a^2 + b^2 = 1 \end{cases} &\iff \\ &\iff \begin{cases} 2b^4 - 6a^2b^2 = 0 \\ a^2 + b^2 = 1. \end{cases} \end{aligned}$$

Om $b = 0$ så är $a = \pm 1$, så vi har kandidaterna $(\pm 1, 0)$. Om $b \neq 0$ har vi

$$\begin{aligned} \begin{cases} b^2 - 3a^2 = 0 \\ a^2 + b^2 = 1 \end{cases} &\iff \begin{cases} b^2 = 3a^2 \\ 4a^2 = 1 \end{cases} &\iff \\ &\iff (a, b) \in \left\{ \left(\frac{1}{2}, \frac{\sqrt{3}}{2} \right), \left(\frac{1}{2}, -\frac{\sqrt{3}}{2} \right), \left(-\frac{1}{2}, \frac{\sqrt{3}}{2} \right), \left(-\frac{1}{2}, -\frac{\sqrt{3}}{2} \right) \right\}. \end{aligned}$$

I våra kandidatpunkter är

$$\begin{aligned} F(1, 0) = 0, \quad F\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = \frac{3\sqrt{3}}{16}, \quad F\left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = -\frac{3\sqrt{3}}{16} \\ F(-1, 0) = 0, \quad F\left(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) = \frac{3\sqrt{3}}{16}, \quad F\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) = -\frac{3\sqrt{3}}{16}. \end{aligned}$$

Svar: Största värdet $\frac{3\sqrt{3}}{16}$ antages (endast) i punkterna $\pm \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$. Minsta värdet $-\frac{3\sqrt{3}}{16}$ antages (endast) i punkterna $\pm \left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right)$.

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = x - y$ då $x^2 + y^2 - 4x = -3$ och $x \geq 2$.

Lösning: Eftersom $x^2 + y^2 - 4x = -3 \iff (x - 2)^2 + y^2 = 1$ är mängden den (slutna) högra halvan av cirkeln med radie 1 kring $(2, 0)$. Denna är kompakt och F är kontinuerlig, så största och minsta värde existerar.

Inför nu skalärfältet $G(x, y) = x^2 + y^2 - 4x$ så att kurvan är nivåkurvan $G(x, y) = -3$. Låt (a, b) vara en intressant punkt som *inte* är en av de två ändpunkterna på halvcirkeln. Då är

$$\begin{aligned} \begin{cases} \nabla F(a,b) \parallel \nabla G(a,b) \\ G(a,b) = -3 \\ a > 2 \end{cases} &\iff \begin{cases} \begin{vmatrix} 1 & -1 \\ 2a-4 & 2b \end{vmatrix} = 0 \\ a^2 + b^2 - 4a = -3 \\ a > 2 \end{cases} &\iff \\ &\iff \begin{cases} 2b + 2a - 4 = 0 \\ a^2 + b^2 - 4a = -3 \\ a > 2 \end{cases} &\iff \begin{cases} a + b = 2 \\ a^2 + b^2 - 4a = -3 \\ a > 2 \end{cases} &\iff \\ &\iff \begin{cases} (a,b) = \left(2 \pm \frac{1}{\sqrt{2}}, \mp \frac{1}{\sqrt{2}}\right) \\ a > 2 \end{cases} &\iff (a,b) = \left(2 + \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right). \end{aligned}$$

I denna punkt, samt i ändpunkterna $(2, \pm 1)$, är

$$F(2,1) = 1, \quad F(2,-1) = 3, \quad F\left(2 + \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = 2 + \sqrt{2}.$$

Svar: Största värdet $2 + \sqrt{2}$ erhålles (endast) i $\left(2 + \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$. Minsta värdet 1 erhålles (endast) i $(2,1)$.

9.4.2 Max och min på tvådimensionellt område i planet

Bestäm (om möjligt) största och minsta värdet av $F(x,y) = x^2 + y^2 + x$ på enhetsdisken $D = \{(x,y) \in \mathbb{R}^2: x^2 + y^2 \leq 1\}$.

Lösning: F är kontinuerlig och D är kompakt, så största och minsta värde existerar.

Vi bestämmer först de inre stationära punkterna. Gradienten

$$\nabla F(x,y) = (2x+1, 2y) = \mathbf{0} \iff (x,y) = \left(-\frac{1}{2}, 0\right)$$

som är en inre punkt, och därmed en kandidat. Här är

$$F\left(-\frac{1}{2}, 0\right) = -\frac{1}{4}.$$

Vi undersöker nu randen $x^2 + y^2 = 1$ (enhetscirkeln). Låt $G(x,y) = x^2 + y^2$ så att cirkeln är nivå-kurvan $G(x,y) = 1$. Låt (a,b) vara en intressant punkt. Då är

$$\begin{aligned} \begin{cases} \nabla F(a,b) \parallel \nabla G(a,b) \\ G(a,b) = 1 \end{cases} &\iff \begin{cases} \begin{vmatrix} 2a+1 & 2b \\ 2a & 2b \end{vmatrix} = 0 \\ a^2 + b^2 = 1 \end{cases} &\iff \\ &\iff \begin{cases} 4ab + 2b - 4ab = 0 \\ a^2 + b^2 = 1 \end{cases} &\iff (a,b) = (\pm 1, 0). \end{aligned}$$

I dessa punkter är $F(1,0) = 2$ och $F(-1,0) = 0$.

Svar: Minsta värdet är $-\frac{1}{4}$ och erhålles (endast) i $\left(-\frac{1}{2}, 0\right)$. Största värdet är 2 och erhålles (endast) i $(1,0)$.

9.4.3 Max och min på yta i rummet

Bestäm (om möjligt) största och minsta värde av $F(x,y,z) = x + y - z$ på $\Sigma := \{(x,y,z) \in \mathbb{R}^3: z = x^2 + y^2, z \leq 1\}$.

Lösning: Σ är den del av paraboloiden $z = x^2 + y^2$ där $z \leq 1$. Denna mängd är kompakt, och F är kontinuerlig, så största och minsta värde existerar.

Låt (a, b, c) vara en intressant punkt på paraboloiden, sådan att $c < 1$. Inför $G(x, y, z) := x^2 + y^2 - z$ så att paraboloiden är nivåytan $G(x, y, z) = 0$. Då är

$$\begin{aligned} \begin{cases} \nabla F(a, b, c) \parallel \nabla G(a, b, c) \\ G(a, b, c) = 0 \\ c < 1 \end{cases} &\iff \begin{cases} (1, 1, -1) \times (2a, 2b, -1) = \mathbf{0} \\ a^2 + b^2 = c \\ c < 1 \end{cases} &\iff \\ &\iff \begin{cases} 2b - 1 = 0 \\ 1 - 2a = 0 \\ 2b - 2a = 0 \\ a^2 + b^2 = c \\ c < 1 \end{cases} &\iff (a, b, c) = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right). \end{aligned}$$

Vi har

$$F\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2}.$$

Vi måste också undersöka F på ytans "rand" $x^2 + y^2 = 1, z = 1$. Sätt därför $H(x, y, z) := z$. Vi skall alltså hitta max och min av $F(x, y, z)$ under de två bivillkoren $G(x, y, z) = 0$ och $H(x, y, z) = 1$. Om (a, b, c) är en intressant punkt så är

$$\begin{aligned} \begin{cases} \nabla F(a, b, c), \nabla G(a, b, c), \nabla H(a, b, c) \text{ lin. ber.} \\ G(a, b, c) = 0 \\ H(a, b, c) = 1 \end{cases} &\iff \\ &\iff \begin{cases} \begin{vmatrix} 1 & 1 & -1 \\ 2a & 2b & -1 \\ 0 & 0 & 1 \end{vmatrix} = 0 \\ G(a, b, c) = 0 \\ H(a, b, c) = 1 \end{cases} &\iff \begin{cases} 2b - 2a = 0 \\ a^2 + b^2 - c = 0 \\ c = 1 \end{cases} &\iff \\ &\iff (a, b, c) = \left(\pm \frac{1}{\sqrt{2}}, \pm \frac{1}{\sqrt{2}}, 1\right). \end{aligned}$$

Vi har

$$F\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 1\right) = \sqrt{2} - 1, \quad F\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 1\right) = -\sqrt{2} - 1.$$

Svar: Minsta värdet är $-\sqrt{2} - 1$ och erhålles (endast) i $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 1\right)$. Största värdet är $\frac{1}{2}$ och erhålles (endast) i $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x^2 + 2y^2 + z^2 - z$ på enhetsfären.

Lösning: Enhetsfären är kompakt och F är kontinuerlig, så största och minsta värde existerar.

Låt (a, b, c) vara en intressant punkt. Inför $G(x, y, z) := x^2 + y^2 + z^2$ så att enhetsfären är nivåytan $G(x, y, z) = 1$. Då är

$$\begin{aligned} \begin{cases} \nabla F(a, b, c) \parallel \nabla G(a, b, c) \\ G(a, b, c) = 1 \end{cases} &\iff \begin{cases} (2a, 4b, 2c - 1) \times (a, b, c) = \mathbf{0} \\ a^2 + b^2 + c^2 = 1 \end{cases} &\iff \\ &\iff \begin{cases} 2bc + b = 0 \\ -a = 0 \\ -2ab = 0 \\ a^2 + b^2 + c^2 = 1. \end{cases} \end{aligned}$$

Vi ser direkt att $a = 0$. En möjlighet är sedan att $b = 0$ vilket ger $c = \pm 1$, så vi har kandidaterna $(0, 0, \pm 1)$, d.v.s. polerna. Här är $F(0, 0, 1) = 0$ och $F(0, 0, -1) = 2$. Om $b \neq 0$ är $2c + 1 = 0$ och $b^2 + c^2 = 1$ vilket är ekvivalent med $(a, b, c) = (0, \pm \frac{\sqrt{3}}{2}, -\frac{1}{2})$. I dessa två punkter är $F(0, \pm \frac{\sqrt{3}}{2}, -\frac{1}{2}) = \frac{9}{4} (> 2)$.

Svar: Funktionen minsta värde är 0 och erhålles (endast) i $(0, 0, 1)$. Funktionen största värde är $\frac{9}{4}$ och erhålles (endast) i $(0, \pm \frac{\sqrt{3}}{2}, -\frac{1}{2})$.

9.4.4 Max och min på tredimensionellt område i rummet

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x^2 + y + z$ på $U := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq z \leq 1\}$.

Lösning: Mängden U är kompakt och F kontinuerlig, så största och minsta värde existerar.

Eftersom $\nabla F(x, y, z) = (2x, 1, 1)$ aldrig är noll saknas inre stationära punkter. Vi undersöker därför randen $\partial D = P \cup D$ där $P = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = z, z \leq 1\}$ är paraboloidbiten och $D = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1, z = 1\}$ är disken som utgör områdets "tak". P och D har en gemensam "rand" i cirkeln $S: x^2 + y^2 = 1, z = 1$.

Inför $G(x, y, z) := x^2 + y^2 - z$ och $H(x, y, z) = z$. Då är P den del av nivåytan $G(x, y, z) = 0$ där $z \leq 1$, medan D är den del av nivåytan $H(x, y, z) = 1$ där $x^2 + y^2 \leq 1$. Sista men inte minst är S snittet mellan nivåytorna $G(x, y, z) = 0$ och $H(x, y, z) = 1$.

Vi undersöker först paraboloidbiten. Om (a, b, c) är en intressant punkt där $c < 1$ är

$$\begin{aligned} \begin{cases} \nabla F(a, b, c) \parallel \nabla G(a, b, c) \\ G(a, b, c) = 0 \\ c < 1 \end{cases} &\iff \begin{cases} (2a, 1, 1) \times (2a, 2b, -1) = \mathbf{0} \\ a^2 + b^2 - c = 0 \\ c < 1 \end{cases} &\iff \\ &\iff \begin{cases} -1 - 2b = 0 \\ 4a = 0 \\ 4ab - 2a = 0 \\ a^2 + b^2 - c = 0 \\ c < 1 \end{cases} &\iff (a, b, c) = (0, -\frac{1}{2}, \frac{1}{4}). \end{aligned}$$

Här är $F(0, -\frac{1}{2}, \frac{1}{4}) = -\frac{1}{4}$.

Vi undersöker sedan disken. Om (a, b, c) är en intressant punkt där $a^2 + b^2 < 1$ så är

$$\begin{cases} \nabla F(a, b, c) \parallel \nabla H(a, b, c) \\ H(a, b, c) = 1 \\ a^2 + b^2 < 1 \end{cases} \iff \begin{cases} 1 = 0 \\ -2a = 0 \\ 0 = 0 \\ c = 1 \\ a^2 + b^2 < 1 \end{cases}$$

vilket är absurt. Alltså finns ingen sådan intressant punkt. Slutligen undersöker vi F på cirkeln S . Om (a, b, c) är en intressant punkt här så är

$$\begin{cases} \nabla F(a, b, c), \nabla G(a, b, c), \nabla H(a, b, c) \text{ lin. ber.} \\ G(a, b, c) = 0 \\ H(a, b, c) = 1 \end{cases} \iff \begin{cases} \left| \begin{array}{ccc} 2a & 1 & 1 \\ 2a & 2b & -1 \\ 0 & 0 & 1 \end{array} \right| = 0 \\ a^2 + b^2 + c = 0 \\ c = 1 \end{cases} \iff \begin{cases} 4ab - 2a = 0 \\ a^2 + b^2 - c = 0 \\ c = 1 \end{cases}$$

som direkt ger $c = 1$. En möjlighet är sedan att $a = 0$ i vilket fall $b = \pm 1$. I dessa punkter är $F(0, 1, 1) = 2$ respektive $F(0, -1, 1) = 0$. Om $a \neq 0$ är $4b - 2 = 0$ och $a^2 + b^2 - 1 = 0$ som ger $b = \frac{1}{2}$ och $a = \pm \frac{\sqrt{3}}{2}$; här är $F\left(\pm \frac{\sqrt{3}}{2}, \frac{1}{2}, 1\right) = \frac{9}{4}$.

Svar: Funktionen minsta värde är $-\frac{1}{4}$ och detta antages (endast) i $\left(0, -\frac{1}{2}, \frac{1}{4}\right)$. Funktionen största värde är $\frac{9}{4}$ och detta antages (endast) i $\left(\pm \frac{\sqrt{3}}{2}, \frac{1}{2}, 1\right)$.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x + y - z^2$ på $U = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 \leq 1, z \geq 0\}$.

Lösning: U är (slutna) övre halvan av (slutna) enhetsklotet, som är en kompakt mängd. Samtidigt är F kontinuerlig, så största och minsta värde existerar.

Vi letar först efter inre stationära punkter. Gradienten $\nabla F(x, y, z) = (1, 1, -2z)$ är emellertid alltid nollskild, så F har inga stationära punkter i hela rummet.

Randen $\partial U = S' \cup D$ där $S' = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 = 1, z > 0\}$ är (slutna) övre halvan av enhets sfären och $D = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 \leq 1, z = 0\}$ är disken som utgör golvet. Dessa har en gemensam "rand" i cirkeln $S = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 = 1, z = 0\}$.

Inför $G(x, y, z) := x^2 + y^2 + z^2$ och $H(x, y, z) = z$. Då är S' den del av nivåytan $G(x, y, z) = 1$ där $z > 0$, D den del av nivåytan $H(x, y, z) = 0$ där $x^2 + y^2 \leq 1$ och S snittet mellan nivåytorna $G(x, y, z) = 1$ och $H(x, y, z) = 0$.

Halvsfären. Låt (a, b, c) vara en intressant punkt på halvsfären, där $z > 0$. Då är

$$\begin{cases} \nabla F(a, b, c) \parallel \nabla G(a, b, c) \\ G(a, b, c) = 1 \\ z > 0 \end{cases} \iff \begin{cases} (1, 1, -2c) \times (a, b, c) = \mathbf{0} \\ a^2 + b^2 + c^2 = 1 \\ c > 0 \end{cases}$$

$$\begin{aligned} &\iff \begin{cases} c + 2bc = 0 \\ -2ac - c = 0 \\ b - a = 0 \\ a^2 + b^2 + c^2 = 1 \\ c > 0 \end{cases} \iff \begin{cases} a = b \\ 1 + 2a = 0 \\ 2a^2 + c^2 = 1 \\ c > 0 \end{cases} \iff \\ &\iff (a, b, c) = \left(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{\sqrt{2}}\right). \end{aligned}$$

Här är $F\left(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{\sqrt{2}}\right) = -\frac{3}{2}$.

Disken. Låt (a, b, c) vara en intressant punkt på disken, där $x^2 + y^2 < 1$. Då är

$$\begin{aligned} &\begin{cases} \nabla F(a, b, c) \parallel \nabla H(a, b, c) \\ H(a, b, c) = 0 \\ a^2 + b^2 < 1 \end{cases} \iff \begin{cases} (1, 1, -2c) \times (0, 0, 1) = \mathbf{0} \\ c = 0 \\ a^2 + b^2 < 1 \end{cases} \iff \\ &\iff \begin{cases} 1 = 0 \\ -1 = 0 \\ 0 = 0 \\ c = 0 \\ a^2 + b^2 < 1 \end{cases} \end{aligned}$$

vilket är absurt, så någon sådan intressant punkt finns inte.

Cirkeln. Låt slutligen (a, b, c) vara en intressant punkt på snittet mellan nivåytorna $G(a, b, c) = 1$ och $H(a, b, c) = 0$. Då är

$$\begin{aligned} &\begin{cases} \nabla F(a, b, c), \nabla G(a, b, c), \nabla H(a, b, c) \text{ lin. ber.} \\ G(a, b, c) = 1 \\ H(a, b, c) = 0 \end{cases} \iff \\ &\iff \begin{cases} \begin{vmatrix} 1 & 1 & -2c \\ a & b & c \\ 0 & 0 & 1 \end{vmatrix} = 0 \\ a^2 + b^2 + c^2 = 1 \\ c = 0 \end{cases} \iff \begin{cases} b - a = 0 \\ a^2 + b^2 + c^2 = 1 \\ c = 0 \end{cases} \iff \\ &\iff (a, b, c) = \pm \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right). \end{aligned}$$

I dessa två punkter är $F\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right) = \sqrt{2}$ respektive $F\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right) = -\sqrt{2}$.

Svar: Funktionen minsta värde är $-\frac{3}{2}$ och detta antages (endast) i $\left(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{\sqrt{2}}\right)$. Funktionen största värde är $\sqrt{2}$ och detta antages (endast) i $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$.

9.5 Fler exempel

I förra avsnittet såg vi hur metoden med bivillkor "strömlinjeformar" lösningen av optimeringsproblem. Man slipper helt parametrisera kurvor och ytor; i stället får man direkt ett ekvationssystem som man kan lösa med standardhantverk (i den mån exakt lösning för hand över huvud taget är möjlig, förstås). Det enda som krävs är att kurvorna och ytorna kan beskrivas av ekvationer, men detta är mer regel än undantag.

I förra avsnittet gjorde vi om tidigare lösta exempelproblem med det nya maskineriet. Dessa problem hade det gemensamt, att det var lätt att parametrisera relevanta kurvor och ytor. Även om det vanligtvis är *möjligt* att parametrisera, så kan det emellertid vara *omständligt* i många fall. Då innebär det nya maskineriet, där man slipper parametrisera, en förenkling. Vi ger några ytterligare exempel.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = x^2 - 3y$ då $x^4 + y^2 \leq 1$.

Lösning: Funktionen är kontinuerlig och mängden är kompakt¹⁷, så största och minsta värde existerar.

Gradienten $\nabla F(x, y) = (2x, -3)$ är aldrig noll, så F saknar stationära punkter. För att undersöka randen av området, inför $G(x, y) := x^4 + y^2$ så att randen är nivåkurvan $G(x, y) = 1$. Låt nu (a, b) vara en intressant punkt på denna kurva. Då är

$$\begin{aligned} \begin{cases} \nabla F(a, b) \parallel \nabla G(a, b) \\ G(a, b) = 1 \end{cases} &\iff \begin{cases} \begin{vmatrix} 2a & -3 \\ 4a^3 & 2b \end{vmatrix} = 0 \\ a^4 + b^2 = 1 \end{cases} &\iff \\ &\iff \begin{cases} 4ab + 12a^3 = 0 \\ a^4 + b^2 = 1. \end{cases} \end{aligned}$$

En möjlighet är att $a = 0$ i vilket fall $b = \pm 1$. Om $a \neq 0$ så är $b + 3a^2 = 0$ och $a^4 + b^2 = 1$ vilket ger $10a^4 = 1$, d.v.s. $a = \pm \frac{1}{\sqrt[4]{10}}$ och $b = -\frac{3}{\sqrt{10}}$. Dessa tre kandidater ger värdena

$$F(0,1) = -3, \quad F(0,-1) = 3, \quad F\left(\frac{1}{\sqrt[4]{10}}, -\frac{3}{\sqrt{10}}\right) = \sqrt{10}, \quad F\left(-\frac{1}{\sqrt[4]{10}}, -\frac{3}{\sqrt{10}}\right) = \sqrt{10}.$$

Svar: Funktionen minsta värde är -3 och detta antages (endast) i $(0,1)$. Funktionen största värde är $\sqrt{10}$ och detta antages (endast) i $\left(\pm \frac{1}{\sqrt[4]{10}}, -\frac{3}{\sqrt{10}}\right)$.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y, z) = x - y + z$ när $z \geq 2x^2 + y^2$ och $x + 3y + z \leq 2$.

Lösning: Vi skall tydligen vara "innanför eller på" paraboloiden $z = 2x^2 + y^2$ och "under eller på" planet $z = 2 - x - 3y$, så mängden är kompakt. Samtidigt är F kontinuerlig, så största och minsta värde existerar.

Gradienten $\nabla F(x, y, z) = (1, -1, 1)$ är aldrig noll, så F saknar stationära punkter.

Låt $G(x, y, z) := 2x^2 + y^2 - z$ och $H(x, y, z) := x + 3y + z$ så att de två begränsningsytorna är delar av nivåytorna $G(x, y, z) = 0$ respektive $H(x, y, z) = 2$.

Låt (a, b, c) vara en intressant punkt på den del av **paraboloiden** $G(x, y, z) = 0$ där $H(x, y, z) < 2$. Då är

¹⁷ Förklara varför!

$$\begin{cases} \nabla F(a, b, c) \parallel \nabla G(a, b, c) \\ G(a, b, c) = 0 \\ H(a, b, c) < 2 \end{cases} \iff \begin{cases} (1, -1, 1) \times (4a, 2b, -1) = \mathbf{0} \\ 2a^2 + b^2 - c = 0 \\ a + 3b + c < 2 \end{cases} \iff$$

$$\iff \begin{cases} 1 - 2b = 0 \\ 4a + 1 = 0 \\ 2b + 4a = 0 \\ 2a^2 + b^2 - c = 0 \\ a + 3b + c < 2 \end{cases} \iff (a, b, c) = \left(-\frac{1}{4}, \frac{1}{2}, \frac{3}{8}\right).$$

Här är $F\left(-\frac{1}{4}, \frac{1}{2}, \frac{3}{8}\right) = -\frac{3}{8}$.

Låt nu (a, b, c) vara en intressant punkt på den del av **planet** $H(x, y, z) = 2$ där $G(x, y, z) < 0$. Då är

$$\begin{cases} \nabla F(a, b, c) \parallel \nabla H(a, b, c) \\ H(a, b, c) = 2 \\ G(a, b, c) < 0 \end{cases} \iff \begin{cases} (1, -1, 1) \times (1, 3, 1) = \mathbf{0} \\ a + 3b + c = 2 \\ 2a^2 + b^2 - c < 0 \end{cases}$$

där första ekvationen är en absurditet, så intressant punkt här saknas.

Låt till sist (a, b, c) vara en intressant punkt på **skärningskurvan** $G(a, b, c) = 0, H(a, b, c) = 2$. Då är

$$\begin{cases} \nabla F(a, b, c), \nabla G(a, b, c), \nabla H(a, b, c) \text{ lin. ber.} \\ G(a, b, c) = 0 \\ H(a, b, c) = 2 \end{cases} \iff$$

$$\iff \begin{cases} \begin{vmatrix} 1 & -1 & 1 \\ 4a & 2b & -1 \\ 1 & 3 & 1 \end{vmatrix} = 0 \\ 2a^2 + b^2 - c = 0 \\ a + 3b + c = 2 \end{cases} \iff \begin{cases} 2b + 3 + 4a + 1 + \\ \quad + 12a - 2b = 0 \\ 2a^2 + b^2 - c = 0 \\ a + 3b + c = 2 \end{cases} \iff$$

$$\iff \begin{cases} a = -1/4 \\ 2a^2 + b^2 - c = 0 \\ a + 3b + c = 2 \end{cases} \iff (a, b, c) = \left(-\frac{1}{4}, -\frac{3}{2} \pm \frac{\sqrt{70}}{4}, \frac{27}{4} \mp \frac{3\sqrt{70}}{4}\right).$$

I de här punkterna är

$$F\left(-\frac{1}{4}, -\frac{3}{2} \pm \frac{\sqrt{70}}{4}, \frac{27}{4} \mp \frac{3\sqrt{70}}{4}\right) = 8 \mp \sqrt{70}.$$

Svar: Funktionen minsta värde är $-\frac{3}{8}$ och detta antages (endast) i $\left(-\frac{1}{4}, \frac{1}{2}, \frac{3}{8}\right)$. Funktionen största värde är $8 + \sqrt{70}$ och detta antages (endast) i $\left(-\frac{1}{4}, -\frac{3}{2} - \frac{\sqrt{70}}{4}, \frac{27}{4} + \frac{3\sqrt{70}}{4}\right)$.

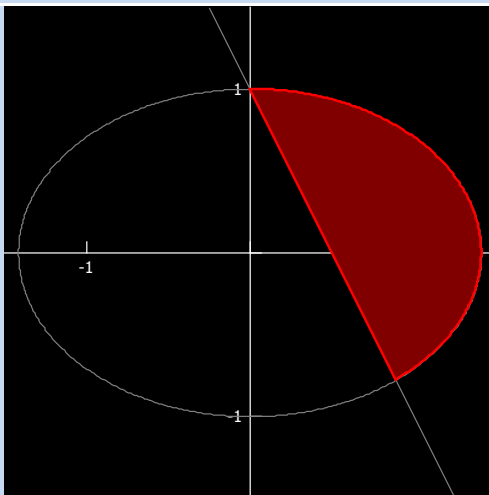
9.5.1 Ett mycket vanligt fel

Låt oss avsluta kapitlet om kompakt optimering med att diskutera ett vanligt fel. Betrakta följande (korrekt lösta) problem.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = x + y$ då $x^2 + 2y^2 \leq 2$ och $2x + y \geq 1$.

Lösning: Området är den del av den fyllda ellipsen $x^2 + 2y^2 \leq 1$ som ligger "över eller på" linjen $y = 1 - 2x$.



Gradienten $\nabla F(x, y) = (1, 1)$ är aldrig noll så stationära punkter saknas. För att undersöka randen, inför $G(x, y) := x^2 + 2y^2$ och $H(x, y) := 2x + y$ så att ellipsen och linjen är nivåkurvorna $G(x, y) = 2$ respektive $H(x, y) = 1$. Vi undersöker först **ellipsskärvan**; låt (a, b) vara en ("inre") intressant punkt här. Då är

$$\begin{cases} \nabla F(a, b) \parallel \nabla G(a, b) \\ G(a, b) = 2 \\ H(a, b) > 1 \end{cases} \iff \begin{cases} \begin{vmatrix} 1 & 1 \\ 2a & 4b \end{vmatrix} = 0 \\ a^2 + 2b^2 = 2 \\ 2a + b > 1 \end{cases} \iff \\ \iff \begin{cases} a = 2b \\ a^2 + 2b^2 = 2 \\ 2a + b > 1 \end{cases} \iff (a, b) = \left(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right).$$

Här är $F\left(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = \sqrt{3}$. Vi undersöker närmast **linjestycket**. Om (a, b) är en ("inre") intressant punkt här så är

$$\begin{cases} \nabla F(a, b) \parallel \nabla H(a, b) \\ H(a, b) = 1 \\ G(a, b) < 2 \end{cases} \iff \begin{cases} \begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} = 0 \\ 2a + b = 1 \\ a^2 + 2b^2 < 2 \end{cases}$$

vilket är absurt, så intressant punkt här saknas. Slutligen betraktar vi de två "hörn" $G(x, y) = 2, H(x, y) = 1$, där vi "hugger av" såväl ellipsskärvan som linjestycket. Vi har

$$\begin{cases} G(x, y) = 2 \\ H(x, y) = 1 \end{cases} \iff (x, y) \in \left\{ (0, 1), \left(\frac{8}{9}, -\frac{7}{9}\right) \right\}$$

och i dessa två punkter är

$$F(0, 1) = 1, \quad F\left(\frac{8}{9}, -\frac{7}{9}\right) = \frac{1}{9}.$$

Svar: Funktionen minsta värde är $\frac{1}{9}$ och detta antages (endast) i $(\frac{8}{9}, -\frac{7}{9})$. Funktionen största värde är $\sqrt{3}$ och detta antages (endast) i $(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}})$.

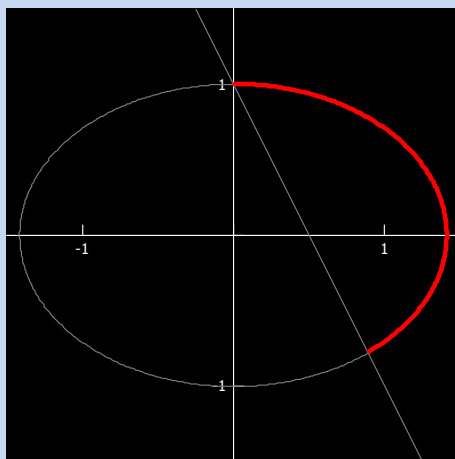
Notera hur vi först undersöker det tvådimensionella området $G < 2, H > 1$ (med $\nabla F = \mathbf{0}$), sedan det endimensionella området $G = 2, H > 1$, sedan det endimensionella området $H = 1, G < 2$, och sist det nolldimensionella området $G = 2, H = 1$. Om ett bivillkor används med *likhet* kallas det för "aktivt", och om ett bivillkor används med *olikhet* kallas det för "passivt", åtminstone av somliga författare. Det typiska är alltså att ett aktivt bivillkor minskar dimensionen (beskriver en hyperyta), medan ett passivt bara säger att man skall befinna sig på ena sidan av någon hyperyta. Exemplet ovan är ganska typiskt för optimeringsproblem i planet över ett plant område (eller i rummet över ett tredimensionellt område) på det sättet att en fullständig undersökning innehåller "alla möjliga kombinationer" av aktiva och passiva bivillkor. Ibland när man rättar skrivningar får man känslan av att studenten tagit fasta på det faktumet i stället för att ta till sig den bakomliggande teorin och de resonemang som ligger bakom metoderna. Detta är inte bra, speciellt inte i problem där man *inte* skall undersöka "alla möjliga kombinationer" av aktiva och passiva bivillkor.

För att ge ett exempel betraktar vi följande, enklare problem, där vi bara vill hitta max och min på *ellipsskärvan* $x^2 + 2y^2 = 2, 2x + y \geq 1$.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = x + y$ då $x^2 + 2y^2 = 2$ och $2x + y \geq 1$.

Lösning: Området är den del av ellipsskärvan $x^2 + 2y^2 = 2$ som ligger "över eller på" linjen $y = 1 - 2x$.



Inför $G(x, y) := x^2 + 2y^2$ och $H(x, y) := 2x + y$ så att ellipsen och linjen är nivåkurvorna $G(x, y) = 2$ respektive $H(x, y) = 1$. Vi vill sålunda hitta max och min av $F(x, y)$ då $G(x, y) = 2$ och $H(x, y) \geq 1$. Låt (a, b) vara en intressant punkt annan än ändpunkterna. Då är

$$\begin{cases} \nabla F(a, b) \parallel \nabla G(a, b) \\ G(a, b) = 2 \\ H(a, b) > 1 \end{cases} \iff \begin{cases} \begin{vmatrix} 1 & 1 \\ 2a & 4b \end{vmatrix} = 0 \\ a^2 + 2b^2 = 2 \\ 2a + b > 1 \end{cases} \iff$$

$$\iff \begin{cases} a = 2b \\ a^2 + 2b^2 = 2 \\ 2a + b > 1 \end{cases} \iff (a, b) = \left(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right).$$

Här är $F\left(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = \sqrt{3}$.

Slutligen betraktar vi de två ändpunkterna $G(x, y) = 2, H(x, y) = 1$, där vi "hugger av" ellipsskärvan. Vi har

$$\begin{cases} G(x, y) = 2 \\ H(x, y) = 1 \end{cases} \iff (x, y) \in \left\{(0, 1), \left(\frac{8}{9}, -\frac{7}{9}\right)\right\}$$

och i dessa två punkter är

$$F(0, 1) = 1, \quad F\left(\frac{8}{9}, -\frac{7}{9}\right) = \frac{1}{9}.$$

Svar: Funktionen minsta värde är $\frac{1}{9}$ och detta antages (endast) i $\left(\frac{8}{9}, -\frac{7}{9}\right)$. Funktionen största värde är $\sqrt{3}$ och detta antages (endast) i $\left(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$.

(Att vi får samma svar i slutändan är förstas en ren slump.) Som vi redan långt tidigare anmärkt vore det väldigt tokigt om man började med att sätta $\nabla F(x, y) = \mathbf{0}$; se avsnitt 8.3.4 (och 8.4.4). Men det misstag vi vill belysa här är i stället följande:

Ett mycket vanligt fel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = x + y$ då $x^2 + 2y^2 = 2$ och $2x + y \geq 1$.

Lösning: Inför $G(x, y) := x^2 + 2y^2$ och $H(x, y) := 2x + y$ så att ellipsen och linjen är nivåkurvorna $G(x, y) = 2$ respektive $H(x, y) = 1$. Vi undersöker först **ellipsskärvan**; låt (a, b) vara en intressant punkt här. Då är

$$\begin{aligned} \begin{cases} \nabla F(a, b) \parallel \nabla G(a, b) \\ G(a, b) = 2 \\ H(a, b) > 1 \end{cases} &\iff \begin{cases} \begin{vmatrix} 1 & 1 \\ 2a & 4b \end{vmatrix} = 0 \\ a^2 + 2b^2 = 2 \\ 2a + b > 1 \end{cases} \iff \\ &\iff \begin{cases} a = 2b \\ a^2 + 2b^2 = 2 \\ 2a + b > 1 \end{cases} \iff (a, b) = \left(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right). \end{aligned}$$

Här är $F\left(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right) = \sqrt{3}$. Vi undersöker härnäst **linjestycket**. Om (a, b) är en intressant punkt här så är

$$\begin{cases} \nabla F(a, b) \parallel \nabla H(a, b) \\ H(a, b) = 1 \\ G(a, b) < 2 \end{cases} \iff \begin{cases} \begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} = 0 \\ 2a + b = 1 \\ a^2 + 2b^2 < 2 \end{cases}$$

vilket är absurt, så intressant punkt här saknas. Slutligen betraktar vi de två "hörnen" $G(x, y) = 2, H(x, y) = 1$. Vi har

$$\begin{cases} G(x, y) = 2 \\ H(x, y) = 1 \end{cases} \iff (x, y) \in \left\{(0, 1), \left(\frac{8}{9}, -\frac{7}{9}\right)\right\}$$

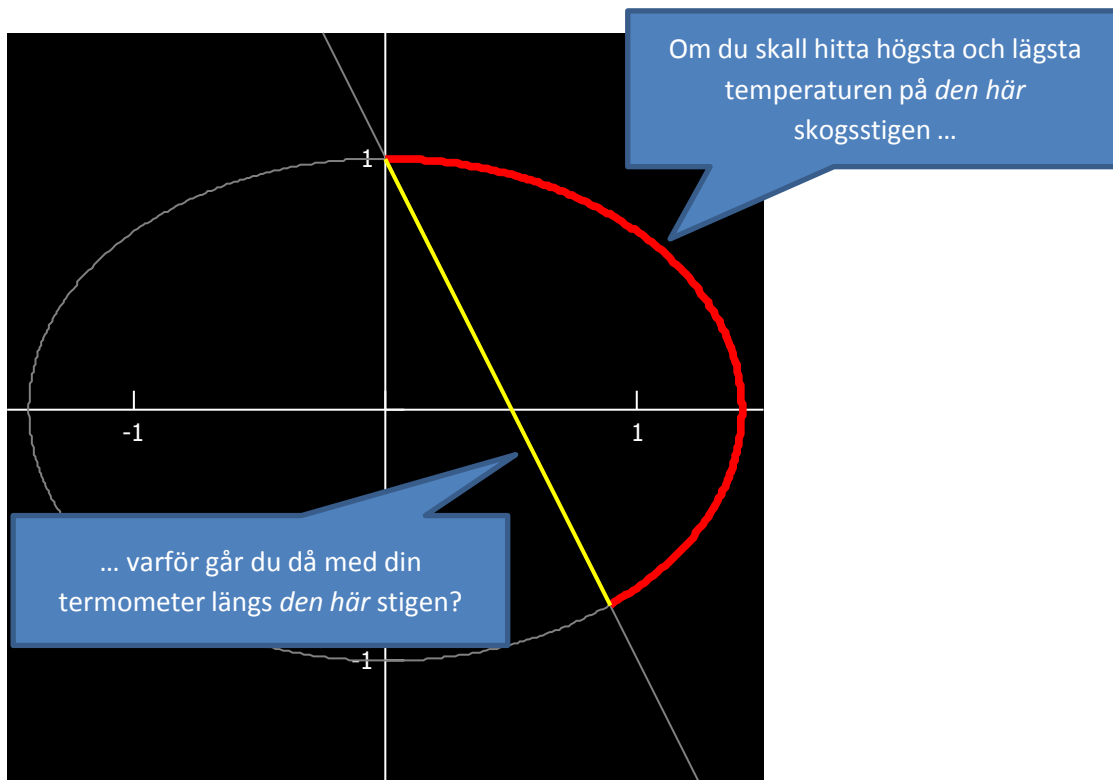
och i dessa två punkter är

$$F(0,1) = 1, \quad F\left(\frac{8}{9}, -\frac{7}{9}\right) = \frac{1}{9}.$$

Svar: Funktionen minsta värde är $\frac{1}{9}$ och detta antages (endast) i $\left(\frac{8}{9}, -\frac{7}{9}\right)$. Funktionen största värde är $\sqrt{3}$ och detta antages (endast) i $\left(\frac{2}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$.

Detta är mycket "tokigt". Den andra undersökningen, där studenten sätter $\nabla F(a, b) \parallel \nabla H(a, b)$, ämnar ju till att hitta högsta och lägsta värdet på *linjestycket* $H = 1, G \leq 2$. Men uppgiften gick ju ut på att hitta högsta och lägsta värdet på *ellipsskärvan* $G = 2, H \geq 1$!

En vardagsanalogi är följande: säg att vi har två skogsstigar mellan punkt A och B , och att en person fått i uppdrag att bestämma högsta och lägsta temperaturen på den enda av dessa. Efter att personen gått med termometer längs denna stig, så beger han sig till *den andra* stigen och går med termometern längs hela den stigen. Skämt åsido antyder redovisningar som den ovan att studenten egentligen inte vet vad han gör, vilket kan uppfattas som dåligt.



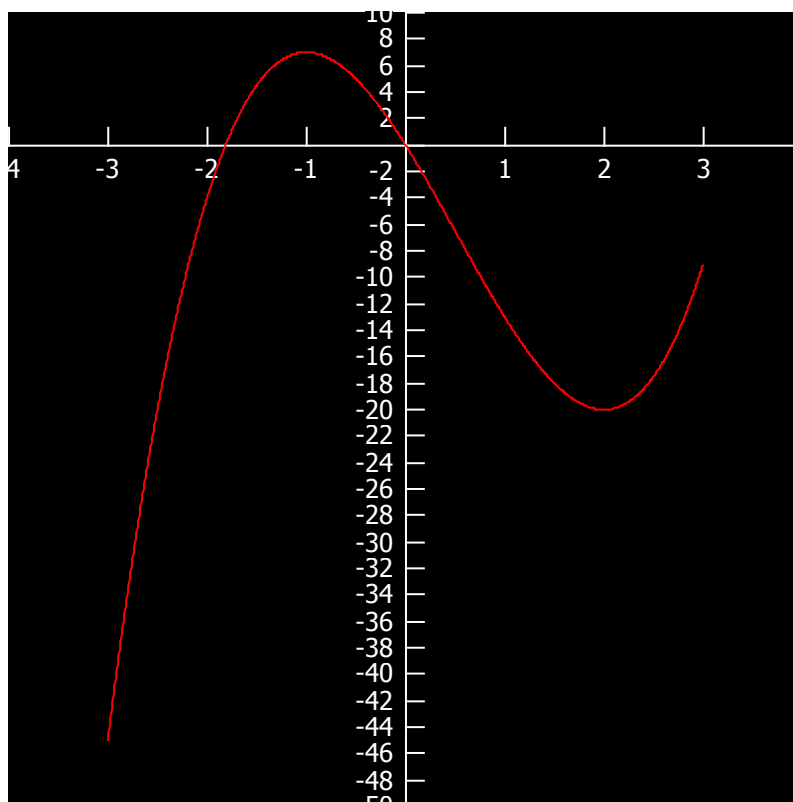
10 Optimering på icke-kompakt mängd

Hittills har vi bara studerat optimeringsproblem på *kompakta* mängder. Detta faktum har underlättat för oss, eftersom vi då kunnat *förutsätta* existensen av maximi- och minimipunkter.

Låt oss repetera varför detta underlättar. Säg att vi vill hitta max av C^1 -envariabelfunktionen f på det kompakta intervallet $I = [a, b]$. Vi vet att största värde existerar, eftersom f är kontinuerlig och I kompakt. Det finns alltså minst en maximipunkt. Låt p beteckna en av dem. Eftersom p är ett lokalt maximum måste $f'(p) = 0$ eller så tillhör p randen: $p \in \{a, b\}$. Vi behöver då bara undersöka funktionsvärdena i alla stationära punkter samt i randpunkterna (vanligtvis en handfull kandidater), och välja ut största värdet. Då är vi klara.

Exempel. Bestäm största och minsta värde av $f(x) = 2x^3 - 3x^2 - 12x$ på $[-3, 3]$.

Lösning: f är kontinuerlig och $[-3, 3]$ är kompakt, så största och minsta värde existerar. Derivatans $f'(x) = 6x^2 - 6x - 12 = 0 \Leftrightarrow x \in \{-1, 2\}$ och $f(-1) = 7$, $f(2) = -20$. I ändpunkterna är $f(-3) = -45$ respektive $f(3) = -9$. Största värdet är alltså 7 och detta antages (endast) i $x = -1$; minsta värdet är -45 och detta antages (endast) i $x = -3$.



Vad kan gå fel om vi *inte* kan vara säkra på existensen av max och min? Det är ju fortfarande så, att om t.ex. p är ett globalt maximum, *så* är p en stationär punkt till funktionen eller en randpunkt till mängden. Jo, visst, men p kanske inte existerar! Som exempel, betrakta samma f som i exemplet ovan, fast med mängden lika med hela \mathbb{R} eller $[-3, \infty[$.

Om mängden inte är kompakt räcker det alltså inte med att lista kandidater, utan man måste angripa problemet på ett annat sätt. Vi ger nedan exempel på flervariabeloptimering på icke-kompakta mängder.

10.1 Exempel i planet

Exempel

Bestäm (om möjligt) största och minsta värdet av $F(x, y) = (x + y)e^{-x^2 - y^2}$ på \mathbb{R}^2 .

Lösning: Vi bestämmer först skalärfältets stationära punkter. Gradienten

$$\begin{aligned} \nabla F(x, y) = (1 - 2x^2 - 2xy, 1 - 2y^2 - 2xy)e^{-x^2 - y^2} = \mathbf{0} & \iff \\ \iff \begin{cases} 1 - 2x^2 - 2xy = 0 \\ 1 - 2y^2 - 2xy = 0 \end{cases} & \iff \begin{cases} 1 - 2x^2 - 2xy = 0 \\ x^2 = y^2. \end{cases} \end{aligned}$$

$x^2 = y^2$ betyder att $x = \pm y$. Om $x = y$ har vi $1 - 2x^2 - 2x^2 = 0$ som ger $x = \pm \frac{1}{2}$, så $\pm \left(\frac{1}{2}, \frac{1}{2}\right)$ är stationära punkter till F . Om $x = -y$ har vi å andra sidan $1 - 2x^2 + 2x^2 = 0$ som är ett falskt påstående, så $x \neq -y$. Därför är

$$\nabla F(x, y) = \mathbf{0} \iff (x, y) = \pm \left(\frac{1}{2}, \frac{1}{2}\right).$$

I dessa punkter är

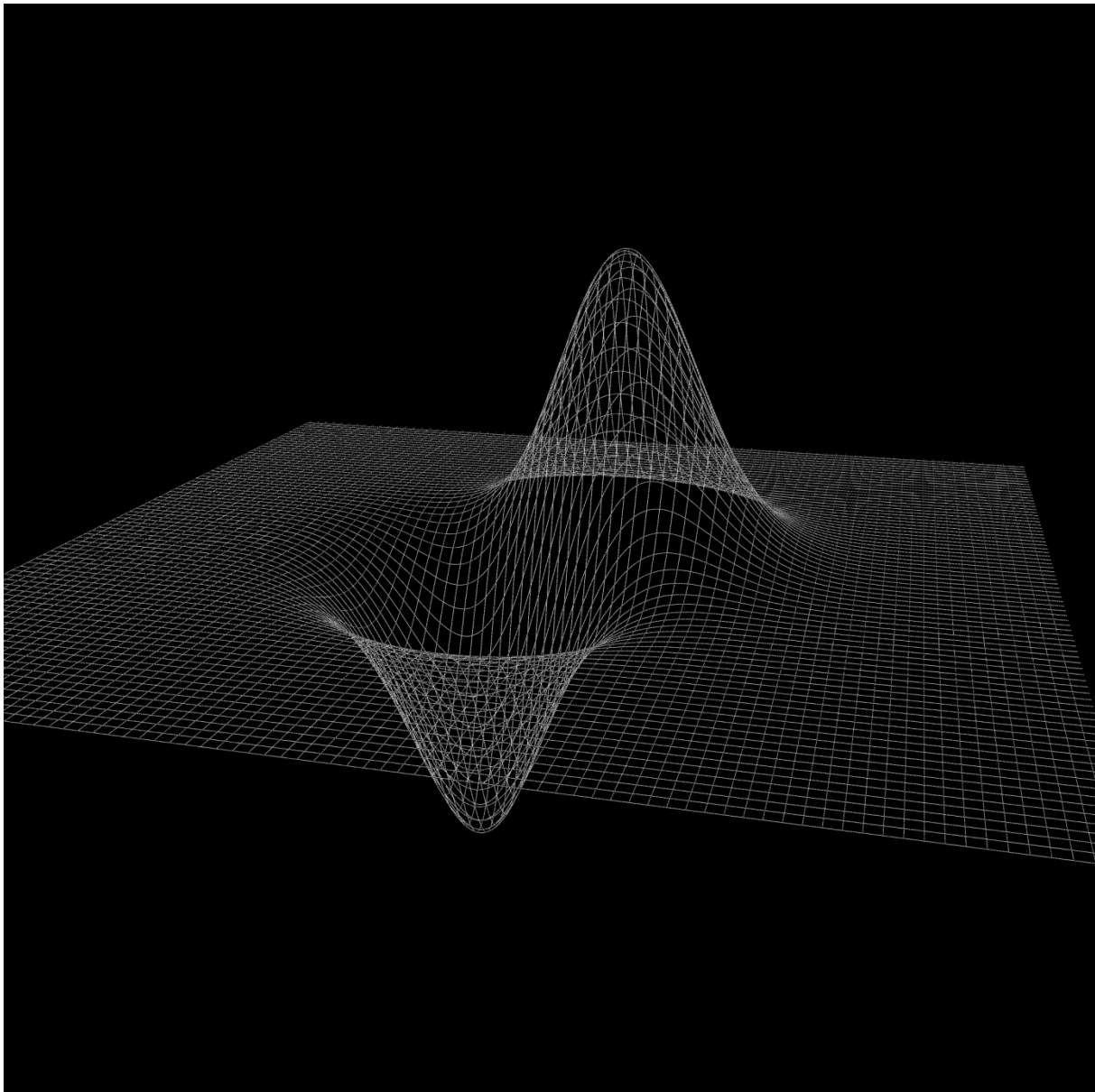
$$F\left(\pm \left(\frac{1}{2}, \frac{1}{2}\right)\right) = \pm \frac{1}{\sqrt{e}}$$

Sedan noterar vi att $F(x, y) \rightarrow 0$ då $\sqrt{x^2 + y^2} \rightarrow \infty$; i polära koordinater har vi ju $F(\rho, \varphi) = (\cos \varphi + \sin \varphi)\rho e^{-\rho^2}$ som uppenbarligen $\rightarrow 0$ då $\rho \rightarrow \infty$. Speciellt finns det ett tal $R > 100$ (R är alltså större än avståndet från origo till de stationära punkterna) sådant att $|F(x, y)| < 0.01$ för alla $(x, y) \in \mathbb{R}^2$ med $\sqrt{x^2 + y^2} \geq R$. Låt nu D beteckna den slutna disken kring origo med radie R . Detta är en *kompakt* mängd, så F har såväl ett största som ett minsta värde på D , och extrempunkterna måste vara stationära punkter eller randpunkter. De enda stationära punkterna är $\pm \left(\frac{1}{2}, \frac{1}{2}\right)$, och här är $F\left(\pm \left(\frac{1}{2}, \frac{1}{2}\right)\right) = \pm \frac{1}{\sqrt{e}}$. På randen är $|F(x, y)| < 0.01 (< \left|\pm \frac{1}{\sqrt{e}}\right|)$, så största värdet av F på D är $\frac{1}{\sqrt{e}}$ och antages i $\left(\frac{1}{2}, \frac{1}{2}\right)$ medan minsta värdet är $-\frac{1}{\sqrt{e}}$ och antages i $-\left(\frac{1}{2}, \frac{1}{2}\right)$.

Men överallt utanför D är $|F(x, y)| < 0.01$, så F 's största värde här måste vara mindre än största värdet av F på D , och motsvarande för minsta värdet. Därför måste de globala extrempunkterna av F på D också vara funktionens globala extrempunkter på hela \mathbb{R}^2 .

Svar: Funktionen minsta värde är $-\frac{1}{\sqrt{e}}$ och detta antages (endast) i $-\left(\frac{1}{2}, \frac{1}{2}\right)$. Funktionen största värde är $\frac{1}{\sqrt{e}}$ och detta antages (endast) i $\left(\frac{1}{2}, \frac{1}{2}\right)$.

I det här fallet existerade tydligen såväl största som minsta värde. Nedan visas grafen till F (något utdragen i z-led för att göra den tydligare).



Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = \frac{xy}{1+(x^2+y^2)^2}$ på $[0, \infty[^2$.

Lösning: Det är klart att $F(x, y) \geq 0$ för alla $(x, y) \in [0, \infty[^2$ samt att $F(x, y) = 0$ omm $x = 0$ eller $y = 0$. Alltså är funktionens minsta värde 0 och detta antages (endast) på koordinataxlarna.

I det *inre* $]0, \infty[^2$ gäller att gradienten

$$\begin{aligned} \nabla F(x, y) &= \frac{1}{1+(x^2+y^2)^2} \mathbf{e} \begin{pmatrix} y(y^4 - 2x^2y^2 - 3x^4 + 1) \\ x(x^4 - 2x^2y^2 - 3y^4 + 1) \end{pmatrix} = \mathbf{0} && \iff \\ \iff & \begin{cases} x^4 - 2x^2y^2 - 3y^4 + 1 = 0 \\ y^4 - 2x^2y^2 - 3x^4 + 1 = 0 \end{cases} && \iff \\ \iff & \begin{cases} x^4 - 2x^2y^2 - 3y^4 + 1 = 0 \\ 4x^4 - 4y^4 = 0 \end{cases} && \iff \end{aligned}$$

$$\begin{aligned} &\iff \begin{cases} x^4 - 2x^2y^2 - 3y^4 + 1 = 0 \\ x = y \end{cases} \iff \\ &\iff \begin{cases} 4x^4 = 1 \\ x = y \end{cases} \iff (x, y) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \end{aligned}$$

där vi använde dels att x och y är nollskilda i $]0, \infty[^2$, dels att de har samma tecken här. I den här stationära punkten är

$$F\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \frac{1}{4}.$$

I polära koordinater är

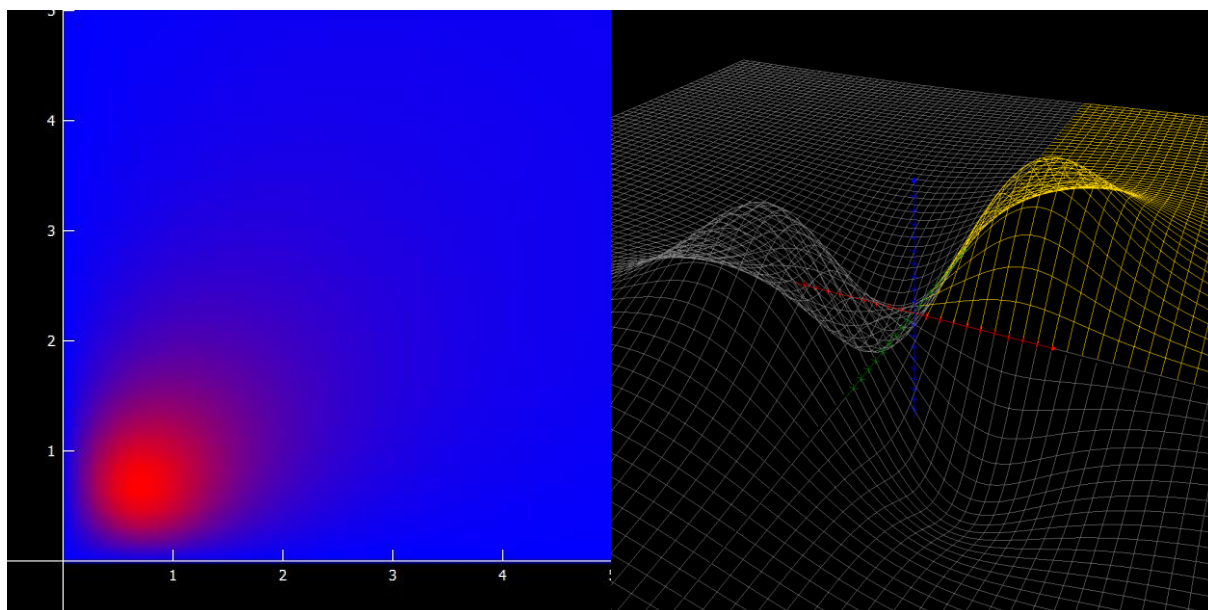
$$|F(\rho, \varphi)| = \left| \frac{\rho^2 \cos \varphi \sin \varphi}{1 + \rho^4} \right| \leq \frac{\rho^2}{1 + \rho^4} \rightarrow 0$$

då $\rho \rightarrow \infty$ så det finns ett tal $R > 100$ sådant att $F(x, y) < 0.01$ så snart $\sqrt{x^2 + y^2} \geq R$. Området $[0, \infty[^2 \cap \overline{B(\mathbf{0}, R)}$, där $\overline{B(\mathbf{0}, R)}$ är den slutna disken kring origo med radie R , är kompakt och på randen är funktionsvärdet $F(x, y) < 0.01 (< 1/4)$. Alltså är minsta värdet av F här 0, vilket antages på koordinataxlarna, medan största värdet är $\frac{1}{4}$, vilket antages i den stationära punkten.

Men utanför $\overline{B(\mathbf{0}, R)}$ är $0 < F(x, y) < 0.01 < 1/4$, så det följer att dessa extrempunkter i $[0, \infty[^2 \cap \overline{B(\mathbf{0}, R)}$ också är extrempunkter i hela $[0, \infty[^2$.

Svar: Funktionen minsta värde är 0 och detta antages (endast) på koordinataxlarna. Funktionen största värde är $\frac{1}{4}$ och detta antages (endast) i $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

Även i det här exemplet existerade tydligen både största och minsta värde. Nedan visas ett färgat plan för F . Till höger visas också grafen, där även F 's beteende i resten av planet visas i grått, som en bonus.



Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = \frac{xy}{1+x^2+y^2}$ på $[0, \infty]^2$.

Lösning: Det är klart att $F(x, y) \geq 0$ för alla $(x, y) \in [0, \infty]^2$ samt att $F(x, y) = 0$ om $x = 0$ eller $y = 0$. Alltså är funktionens minsta värde 0 och detta antages (endast) på koordinataxlarna.

I polära koordinater är, i hela området,

$$F(\rho, \varphi) = \frac{\rho^2 \cos \varphi \sin \varphi}{1 + \rho^2} = \frac{\rho^2 \sin 2\varphi}{2(1 + \rho^2)} \leq \frac{\rho^2}{2(1 + \rho^2)} = \frac{1}{2} - \frac{1}{2(1 + \rho^2)} \leq \frac{1}{2}$$

så ett eventuellt globalt maximum är nödvändigtvis $\leq 1/2$. Samtidigt gäller att

$$F\left(\rho, \frac{\pi}{4}\right) = \frac{\rho^2 \sin 2\varphi}{2(1 + \rho^2)} = \frac{\rho^2}{2(1 + \rho^2)} = \frac{1}{2} - \frac{1}{2(1 + \rho^2)} \rightarrow \frac{1}{2}$$

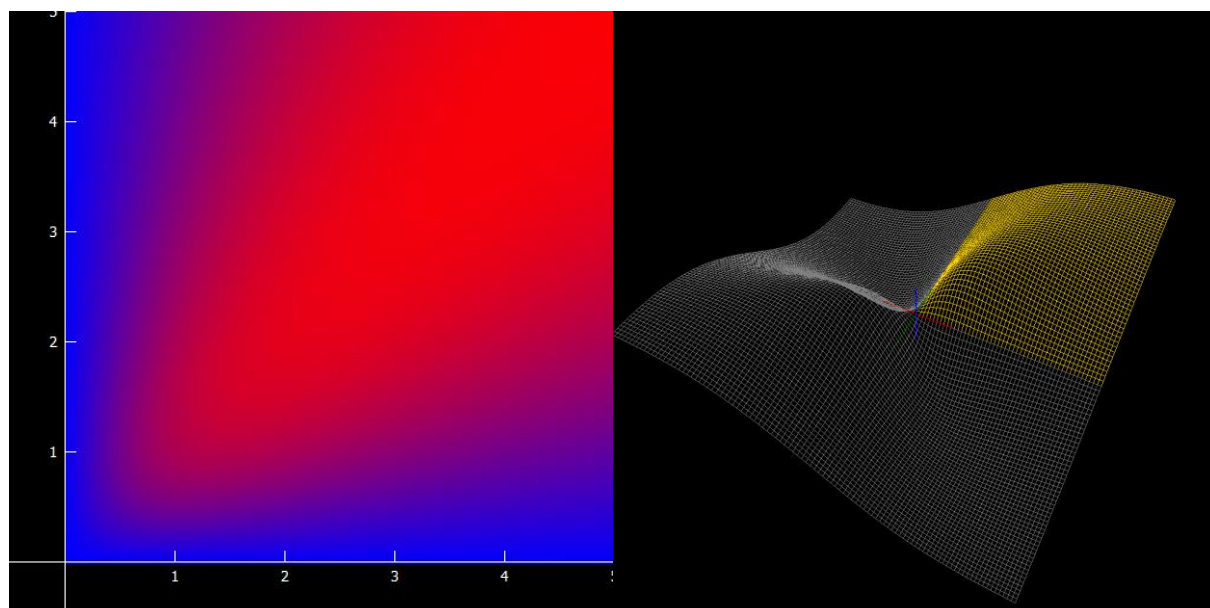
då $\rho \rightarrow \infty$, så ett eventuellt globalt maximum måste vara $\geq \frac{1}{2}$. Sålunda, om det finns ett globalt maximum, måste det vara lika med $1/2$. Låt (a, b) vara en global maximipunkt; då är alltså

$$\begin{aligned} F(a, b) = \frac{1}{2} &\iff \frac{ab}{1 + a^2 + b^2} = \frac{1}{2} &\iff \\ \iff 2ab = 1 + a^2 + b^2 &\iff a^2 - 2ab + b^2 = -1 &\iff \\ \iff (a - b)^2 = -1 && \end{aligned}$$

vilket är omöjligt. Sålunda saknas global maximipunkt.

Svar: Funktionens minsta värde är 0 och detta antages (endast) på koordinataxlarna. Funktionen saknar största värde (*men funktionens *supremum* är $1/2$ och vi närmar oss detta värde när vi går från origo längs den räta linjen $x = y$).

Nedan visas ett färgat plan för F , samt grafen. Grafen visar också F :s beteende i *hela* planet, som en bonus.



Låt oss pröva en lätt variant av föregående exempel:

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = \frac{xy}{1+x^2+y^2}$ på $]0, \infty[^2$.

Lösning: Det är klart att $F(x, y) > 0$ för alla $(x, y) \in]0, \infty[^2$ samt att $F(x, y) \rightarrow 0$ när vi närmar oss koordinataxlarna. Alltså saknas minsta värde.¹⁸

Samma analys som i föregående exempel visar att även största värde saknas.

Svar: Funktionen saknar minsta värde (*men funktionens *infimum* är 0 och vi närmar oss detta värde när vi närmar oss koordinataxlarna, t.ex. vinkelrätt). Funktionen saknar också största värde (*men funktionens *supremum* är $1/2$ och vi närmar oss detta värde när vi går från origo längs den räta linjen $x = y$).

Låt oss avsluta med ett par enkla exempel.

Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = \arctan(x^2 + y^2)$ i hela planet.

Lösning: Eftersom $x^2 + y^2 \geq 0$ överallt och $\arctan t \geq 0$ då $t \geq 0$ är det klart att $F(x, y) \geq 0$ överallt. Samtidigt är $F(x, y) = 0$ om $(x, y) = (0, 0)$ så funktionens minsta värde är 0 och detta antages (endast) i origo.

Eftersom $V_{\arctan} =]-\frac{\pi}{2}, \frac{\pi}{2}[$ är det vidare klart att $F(x, y) < \frac{\pi}{2}$ överallt. Dessutom gäller att $F(x, y) \rightarrow \frac{\pi}{2}$ när $x^2 + y^2 \rightarrow \infty$. Alltså saknas största värde.

Svar: Funktionen minsta värde är 0 och detta antages (endast) i origo. Funktionen saknar största värde (*men *supremum* är $\frac{\pi}{2}$ och vi närmar oss detta värde när $x^2 + y^2 \rightarrow \infty$).

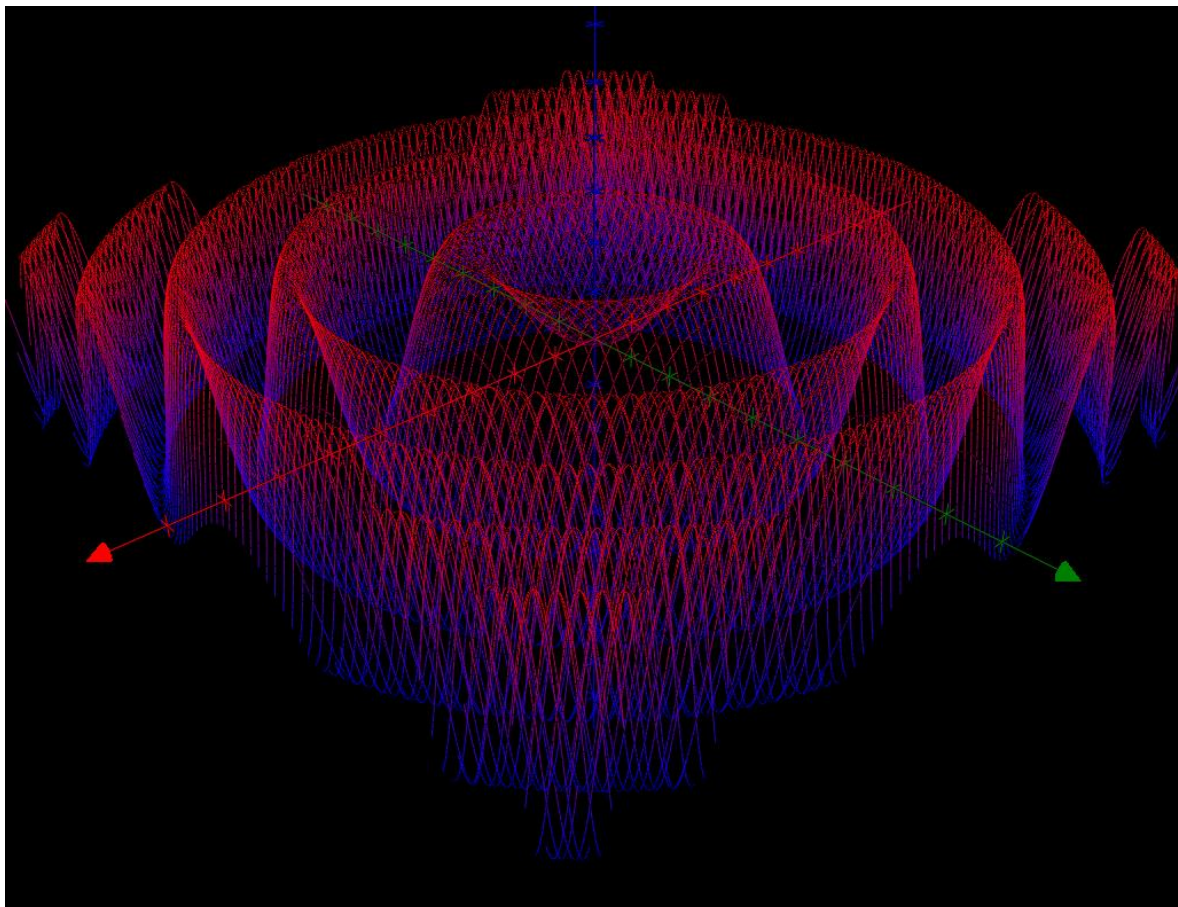
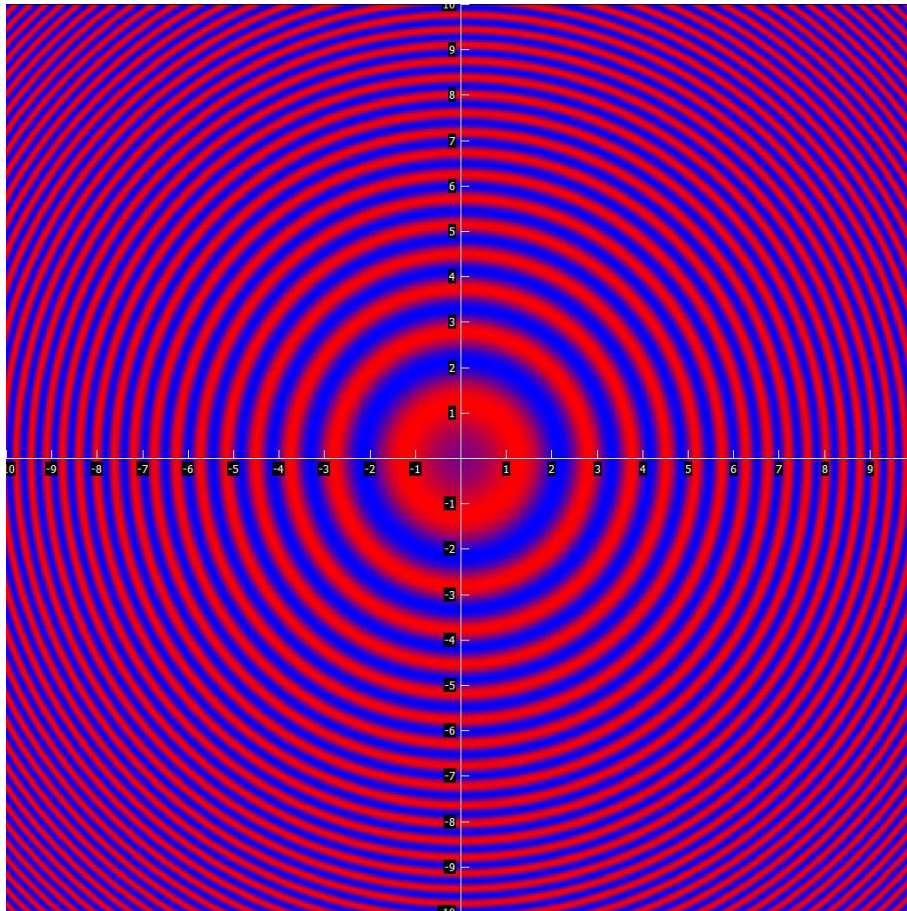
Exempel

Bestäm (om möjligt) största och minsta värde av $F(x, y) = \sin(x^2 + y^2)$ i hela planet.

Lösning: Vi ser direkt att största värdet är 1 och detta erhålles precis i de koncentriska cirkelarna $x^2 + y^2 = \frac{\pi}{2} + n \cdot 2\pi$ där $n \in \mathbb{N}$. Samtidigt är minsta värdet -1 och detta erhålles precis i de koncentriska cirkelarna $x^2 + y^2 = \frac{3\pi}{2} + n \cdot 2\pi$ där $n \in \mathbb{N}$.

Svar: Funktionen minsta värde är -1 och detta erhålles precis i $\bigcup_{n=0}^{\infty} \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = \frac{3\pi}{2} + n \cdot 2\pi\}$. Funktionen största värde är $+1$ och detta erhålles precis i $\bigcup_{n=0}^{\infty} \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = \frac{\pi}{2} + n \cdot 2\pi\}$.

¹⁸ Se till att du kan motivera detta i detalj. (Ledning: minsta värdet kan inte vara ≤ 0 ty ...? Och det kan inte vara > 0 ty ...?)



11 Integraler

Nu kommer vi till det andra huvudmomentet inom flervariabelanalysen, nämligen *integration*. Detta torde också vara det flervariabelmoment som naturverkar- och ingenjörstudenter kommer att använda mest i sina fortsatta studier.

Det här kapitlet innehåller inte en fullständig behandling av integraler. När man introducerar ett (någorlunda omfattande) matematiskt objekt gör man ofta fyra saker:

1. Man motiverar behovet av det nya objektet, och ger den grundläggande idén bakom det.
2. Man definierar objektet på ett rigoröst sätt.
3. Man ger satser som beskriver det nya objektet.
4. Man ger exempel på tillämpningar med/av det nya objektet.

I synnerhet gäller listan ovan i allra högsta grad för integraler. I det här kapitlet kommer vi att motivera integralbegreppet, och ge den grundläggande idén bakom det. Vi kommer också att definiera integralbegreppet rigoröst i envariabelfallet. (Rigorösa definitioner i flervariabelfallet kan läsaren studera i hans ordinarie kurslitteratur.) Däremot kommer vi *inte* att behandla några satser om integraler. Anledningen är bland annat att det här dokumentet är tänkt att behandla grunderna i flervariabelanalysen på ett mycket noggrant sätt, och då är definitioner och grundläggande begrepp viktigare än satser. Dessutom säger författarens erfarenhet att många studenter är tämligen skickliga på att beräkna integraler i praktiken (t.ex. med hjälp av analysens huvudsats och satserna om itererad enkelintegration och variabelbyte), men har lite svårare för att förstå vad en integral egentligen *är* för något.

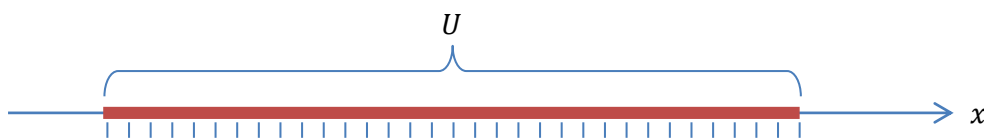
Även om vi inte behandlar t.ex. analysens huvudsats, satsen om itererad enkelintegration och satsen om variabelbyte i multipelintegral i kapitlet så avslutar vi kapitlet med att *tillämpa* dessa satser i en mängd (intressanta) exempel.

11.1 Envariabelintegraler

Vi börjar med att repetera envariabelintegraler. Framför allt vill författaren ställa frågan "Vad *är* en integral?". Ett integrationsproblem i en variabel har två ingredienser:

1. en funktion $f: D_f \rightarrow \mathbb{R}$
2. en begränsad mängd $U \subset D_f$ på vilken f är begränsad.

Mängden är nästan alltid ett *kompakt intervall* och funktionen är vanligtvis åtminstone styckvis kontinuerlig. Själva idén bakom integralbegreppet är som följer. Man börjar med att dela upp mängden i N stycken (små) delintervall (t.ex. 1 000 stycken lika stora intervall).



Tag nu *en* av dessa bitar. Om N är *stort* och varje enskild bit *liten*, så f 's värde är (om f är snäll nog) nästan konstant på biten. Tag något av f 's värden på biten (t.ex. det i mitten, eller det i vänstra eller högra ändpunkten) och gångra det med bitens längd. Gör detta för *varje* bit, och *summera* dessa

produkter. Om vi kallar bitarna för U_1, U_2, \dots, U_N och för varje $k = 1, \dots, N$ väljer ett $x_k \in U_k$ så får vi då talet

$$I_1 = f(x_1)L(U_1) + f(x_2)L(U_2) + \dots + f(x_N)L(U_N)$$

där $L(U_k)$ betecknar *längden* av intervallet U_k . Summan av produkter i högerledet kallas för en *Riemannsumma* för f på U ; dess exakta värde beror förstås på indelningen och valet av x_k i varje bit, men om bitarna är många och små bör vi hamna ganska *nära* ett visst tal oberoende av den precisa indelningen och valen av x_k .

Om vi upprepar denna procedur med finare och finare indelning (d.v.s. fler och fler bitar som blir mindre och mindre – t.ex. hundra lika stora bitar, tusen lika stora bitar, tiotusen lika stora bitar, ...) så får vi en följd av tal: $I_1, I_2, I_3, I_4, \dots$. Själva idén är att *integralen* av f över U skall vara "gränsvärdet" av en sådan följd (och det skall inte spela någon roll exakt *hur* man gör indelningen, bara den blir finare och finare – detta är det normala ifall funktionen f är snäll). Det här talet betecknas¹⁹

$\int_U f(x)dx$. Om U är ett kompakt intervall $[a, b]$ brukar man skriva $\int_a^b f(x)dx$ i stället.

Notera i synnerhet att integralen av den konstanta funktionen 1 ger U 's *längd*, d.v.s. $\int_a^b 1dx = b - a$ (förklara – utan att använda primitiva funktioner!).

Anmärkning ()*: Det finns flera sätt att precis definiera integralbegrepp. En del av dessa ger upphov till *samma* begrepp, medan andra ger upphov till *andra* integralbegrepp. Det vanligaste elementära integralbegreppet är Riemannintegralen som kan definieras som följer: Med en *indelning* av $[a, b]$ (där $b > a$) menas en följd av tal $(t_k)_{k=0}^n$ sådan att $a = t_0 < t_1 < \dots < t_n = b$; vi delar här in $[a, b]$ i de n stycken "småbitarna" $U_1 := [t_0, t_1]$, $U_2 := [t_1, t_2]$, ..., $U_n := [t_{n-1}, t_n]$. Indelningens *finhet* är $\max_{k=0}^{n-1} (t_{k+1} - t_k)$, d.v.s. längden av största "småbiten". Alltså: låg finhet betyder att uppdelningen är "fin", d.v.s. den består av (många) små bitar. För varje $k = 1, \dots, n$ väljer vi sedan en punkt $x_k \in U_k$. Med ett sådant val har vi en *taggad indelning* (t_k, x_k) . Om det finns ett tal I sådant att det för varje $\epsilon > 0$ finns ett $\delta > 0$ sådant att det för *varje* taggad uppdelning (t_k, x_k) med finhet mindre än δ gäller att avståndet mellan Riemannsumman och I är $|\sum_{k=1}^n f(x_k)(t_k - t_{k-1}) - I| < \epsilon$ så säger vi att f är Riemannintegrerbar över $[a, b]$, och vi definierar Riemannintegralen $\int_a^b f(x)dx := I$.

Den intresserade läsaren uppmanas att på egen hand undersöka (t.ex. genom studie av böcker och Internet) varför man inte definierar Riemannintegralen på ett enklare sätt, som t.ex. genom att kräva att indelningarna består av *lika stora* intervall, och att man hela tiden väljer det *mellersta* värdet i varje bit, och sedan bara säga att integralens värde är gränsvärdet av följderna av Riemannsummorna när man tar 1, 2, 3, ... stycken bitar.

Ett alternativt sätt att konstruera Riemannintegralen (d.v.s. exakt samma integralbegrepp) bygger på begreppen *över-* och *undersummor*. Inför återigen en indelning $(t_k)_{k=0}^n$ av $[a, b]$, men den här gången väljer vi inte längre valfritt värde i varje bit, utan det "största" värdet (eller, egentligen: *supremum*). I formler, låt $M_k := \sup_{x \in U_k} f(x)$ för varje $k = 1, \dots, n$. Vi säger att $\sum_{k=1}^n M_k L(U_k)$ är en *översumma* till f på $[a, b]$. Begreppet *undersumma* definieras analogt, fast med infimum i stället för supremum. Det är klart att en översumma alltid är större eller lika med en undersumma (för att se detta, inför en gemensam, finare, indelning). Vi vill förstås att varje översumma [resp. undersumma] skall vara större [resp. mindre] eller lika med integralens värde. Dessutom "borde" det (om f är tillräckligt snäll) finnas såväl över- som undersummor vilka kommer godtyckligt nära integralens värde (om indelningarna är tillräckligt fina) och därmed godtyckligt nära varandra. Låt \mathcal{M} vara mängden av alla översummor (när indelningen varierar), och låt på samma sätt \mathcal{m} vara mängden av alla under-

¹⁹ Notera hur själva notationen $\int_U f(x)dx$ "antyder" att det rör sig om (gränsvärdet) av en Riemannsumma: tecknet \int står för summa, och det som "summeras" är just funktionsvärdet $f(x)$ gånger en liten sträcka " dx ".

summor. Sätt sedan $\bar{I} := \inf \mathcal{M}$ och $\underline{I} := \sup m$. Ifall $\bar{I} = \underline{I}$ säger vi att f är Riemannintegrerbar över $[a, b]$, och vi definierar $\int_a^b f(x) dx := \bar{I}$. Vi har nu åter erhållit Riemannintegralen (man kan visa att det är exakt samma begrepp som det vi introducerade i föregående stycke). [Ett alternativt sätt att formulera utsagan $\bar{I} = \underline{I}$ är som följer: till varje $\epsilon > 0$ finns det en översumma Φ och en undersumma θ sådana att $\Phi - \theta < \epsilon$.]

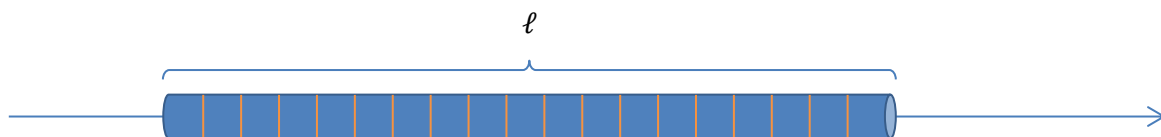
Inom högre matematik används sällan Riemannintegralen över huvud taget. I stället används den så kallade *Lebesgueintegralen*. Detta är ett annat integralbegrepp, som definieras med hjälp av s.k. måtteori. I praktiken är skillnaden mellan de två integralbegreppen att det finns funktioner som är Lebesgueintegrerbara men som inte är Riemannintegrerbara. (Dessa funktioner är ganska "osnälla". Till exempel är ju såväl varje styckvis kontinuerlig som varje monoton funktion Riemannintegrerbar.) Ett exempel är den karaktäristiska funktionen $\chi_{\mathbb{Q}}$ för \mathbb{Q} på \mathbb{R} , d.v.s. $\chi_{\mathbb{Q}}(x) = 1$ om $x \in \mathbb{Q}$ och $\chi_{\mathbb{Q}}(x) = 0$ annars. Lebesgueintegralen av $\chi_{\mathbb{Q}}$ över ett kompakt intervall $[a, b]$ är 0, medan den motsvarande Riemannintegralen inte existerar. Det senare faktumet kan den intresserade läsaren lätt kontrollera själv (genom att visa att $\bar{I} = b - a$ medan $\underline{I} = 0$). Däremot gäller att om en funktion är Riemannintegrerbar så är den också Lebesgueintegrerbar, och de två integralerna har i så fall samma värde. Så t.ex. för styckvis kontinuerliga funktioner på kompakta intervall är de två begreppen lika. Anledningen till att man inom den högre matematiken föredrar Lebesgueintegralen framför Riemannintegralen är att Lebesgueintegralen har trevligare teoretiska egenskaper (och den går också att generalisera på fler sätt).

Först pratade vi om själva *idén* bakom integralbegreppet (Riemannsummor), sedan om precisa integraldefinitioner (t.ex. via över- och undersummor). I praktiken används mycket sällan dessa konstruktioner vid praktisk *beräkning* av integraler (enda nämnvärda undantaget torde vara numerisk integration på dator). Då används i stället andra tekniker, som t.ex. primitiva funktioner, variabelbyten och partiell integration.

Det är trots detta mycket viktigt att studenten åtminstone är välbekant med idén bakom integralbegreppet, eftersom det är denna som säger vad en integral "är" för något, och det är denna man lutar sig mot i tillämpningar när man inser att det är en integral man har med att göra.

Så... Vad är poängen med integraler? Har man någon nytta av dem? Ja, massor. Se följande exempel.

Exempel (massa hos tråd). Antag att en tråd av längd ℓ ligger mellan $x = 0$ och $x = \ell$ och att dess densitet [enhet: g/cm] är $\rho(x)$ i punkten $x \in [0, \ell]$.²⁰ Vad är hela trådens massa?



Ett naturligt sätt att uppskatta trådens massa är att dela in $[0, \ell]$ i 10 lika stora delintervall. Varje delintervall är då litet, så $\rho(x)$ är ungefär konstant på intervallet. Om vi låter x_k vara mittpunkten på det k :te intervallet U_k så är trådens massa i det här intervallet ungefär $\rho(x_k)L(U_k)$ [enhet: g/cm · cm = g] och hela trådens massa är ungefär

$$I_1 = \rho(x_1)L(U_1) + \rho(x_2)L(U_2) + \dots + \rho(x_{10})L(U_{10}).$$

En bättre approximation bör vi få om vi i stället delar in tråden i 100 lika stora bitar, o.s.v. Det är sålunda rimligt att tolka $\int_{[0, \ell]} \rho(x) dx$ som hela trådens massa.

²⁰ Detta är en *modell* av trådens massfördelning. I praktiken innebär modellen att om talet δ är "mycket litet" så är massan av tråden i $[x_0 - \delta/2, x_0 + \delta/2]$ lika med $\delta\rho(x_0)$.

Exempel (area under graf). Låt $f: [a, b] \rightarrow \mathbb{R}^+$ vara en kontinuerlig och positiv funktion. Vad är *arean* under grafen till f , d.v.s. vad är arean av mängden $\{(x, y) \in \mathbb{R}^2: a \leq x \leq b, 0 \leq y \leq f(x)\}$?

Ett naturligt angreppssätt är att dela in intervallet $[a, b]$ i 10 lika stora delintervall. I varje del är $f(x)$ ungefär konstant, så arean under grafen i den här delen är ungefär lika med detta funktionsvärde gånger delens längd (en rektangelarea). Om vi t.ex. låter x_k vara mittpunkten på det k :te intervallet U_k så är arean under grafen över U_k ungefär $f(x_k)L(U_k)$ och hela arean är ungefär

$$I_{10} = f(x_1)L(U_1) + f(x_2)L(U_2) + \dots + f(x_{10})L(U_{10}).$$

En bättre approximation bör vi få om vi i stället delar in $[a, b]$ i 100 lika stora bitar, o.s.v. Det är sålunda rimligt att tolka $\int_{[a,b]} f(x)dx$ som arean under grafen till f mellan a och b . (I själva verket kan man med fördel *definiera* begreppet "area under graf" på det här sättet.) På nästa sida visas de allt bättre approximationer som erhålles när $n = 1, 2, 4, 8, 20, 50$ i det enkla exemplet $f(x) = 4 - x^3/120 + x^2/10 + x$ på intervallet $[4, 16]$. I exemplet erhålles följande värden (där integralen är beräknad med hjälp av primitiv funktion):

I_1	I_2	I_4	I_8	I_{20}	I_{50}	$\int_4^{16} f(x)dx$
188	171.8	167.75	166.7375	166.454	166.40864	166.4

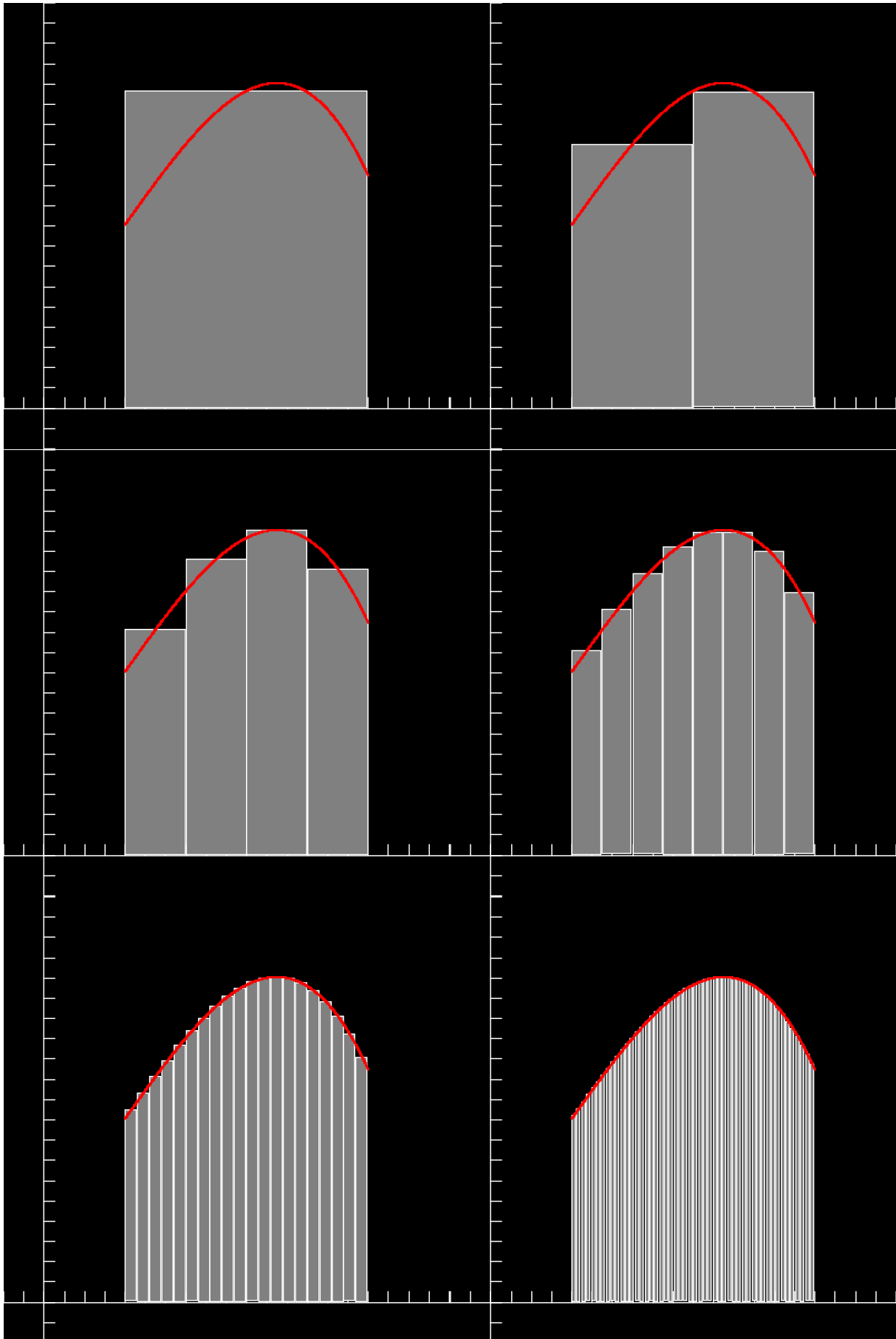
Övning

- *Fundera på hur bilderna skulle se ut om vi valde största eller minsta funktionsvärdet i varje del i stället för funktionsvärdet i mitten av delen. Vi erhåller då över- och undersummor till integralen. Verkar det som om $\bar{I} = \underline{I}$ (med den notation som användes i anmärkningen ovan)?

När man väl lärt sig vad en integral är för något brukar man motivera praktiska integraluppställningar på ett mer kursivt sätt. Som exempel kan vi betrakta den nyligen studerade areaberäkningen: vi vill hitta arean under grafen $y = f(x)$ mellan $x = a$ och $x = b$. Vi delar då in x -axeln i massa småbitar av längd dx (som är "oändligt små"). Rektangeln över biten vid x har höjden $f(x)$, så dess area är $dA := f(x)dx$. Hela arean under grafen får vi genom att summera dessa smårektanglars areor, varvid vi erhåller den totala arean $\int_a^b dA = \int_a^b f(x)dx$. Klart. Det här resonemanget är väldigt handvif-tande (bland annat nämns inte gränsövergången alls, och vi låtsas som om " dx " vore ett "oändligt litet" tal, och "summan" är inte en riktig summa, utan en integral!), och det är alltid underförstått att det egentligen står för det mer precisa resonemang vi gjort tidigare.

Övning

- Se till att du ser kopplingen mellan det riktiga resonemanget och den kursiva motiveringen, så att du känner dig övertygad om giltigheten i den kursiva motiveringen.



Exempel (area under graf). Vi vill beräkna arean A innanför en cirkel med radie r . För att åstadkomma detta betraktar vi cirkelkurvan $x^2 + y^2 = r^2$ kring origo. Övre delen av denna kan skrivas $y = \sqrt{r^2 - x^2}$, och är alltså grafen av en funktion. Arean under den här grafen (som ligger mellan $x = -r$ och $x = r$) är precis halva den sökta cirkelarean, som sålunda är

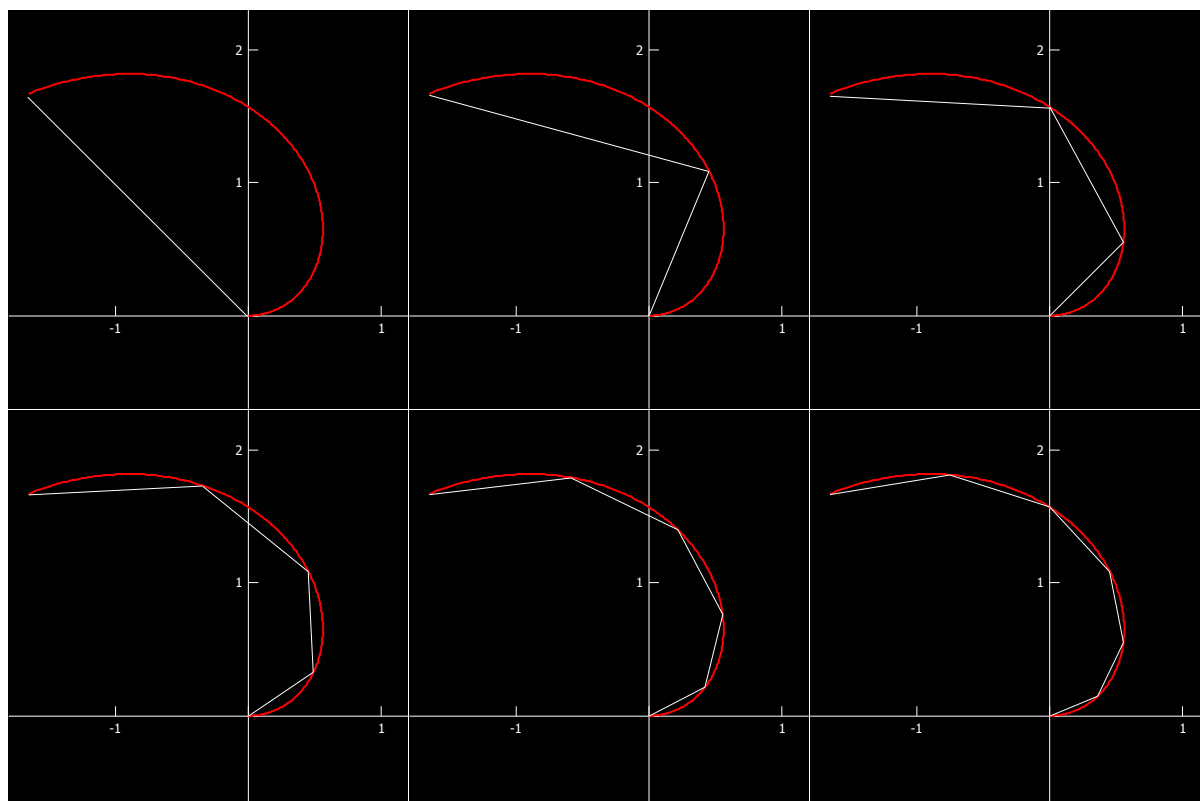
$$\begin{aligned} A &= 2 \int_{-r}^r \sqrt{r^2 - x^2} dx = 2r \int_{-r}^r \sqrt{1 - \left(\frac{x}{r}\right)^2} dx = \left[\begin{array}{l} t := \frac{x}{r} \\ dx = r dt \end{array} \right] = 2r^2 \int_{-1}^1 \sqrt{1 - t^2} dt = \\ &= \left[\begin{array}{l} t = \sin v \\ v \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ dt = \cos v dv \end{array} \right] = 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} |\cos v| \cos v dv = 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 v dv = \\ &= \left[\begin{array}{l} \text{cosinus för} \\ \text{dubbla vinkeln} \end{array} \right] = 2r^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\frac{1}{2} + \frac{1}{2} \cos 2v\right) dv = 2r^2 \left[\frac{v}{2} + \frac{1}{4} \sin 2v \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = r^2 \pi. \end{aligned}$$

Läsarens högstadielärare hade alltså rätt.

Övning

1. Beräkna arean innanför en ellips med halvaxellängderna a och b . Kontrollera ditt resultat genom att sätta $a = b$.

Exempel (längd av kurva). Vi vill beräkna längden $L(\Gamma)$ av kurvan $\Gamma := \mathbf{r}\left(\left[0, \frac{4\pi}{3}\right]\right)$ där $\mathbf{r}(t) := t(\cos t, \sin t)$. Ett naturligt sätt att angripa problemet är att dela in "restiden" $\left[0, \frac{4\pi}{3}\right]$ i N stycken lika stora intervall, och sedan approximera kurvan i varje delintervall med en rät linje från bilden av första till bilden av sista tidpunkten i intervallet, se bilderna nedan.



Om indelningen är fin, så är den räta linjen i varje delintervall en väldigt god approximation till den riktiga kurvan i delintervallet, som därför nästan har konstant hastighetsvektor där. Alltså är längden av kurvan i delintervallet ungefär $|\mathbf{r}'(t)|\Delta t$ där t är *någon* tidpunkt i intervallet, och Δt är tidintervalllets längd. Om vi summerar alla dessa sträckor erhåller vi ungefär hela kurvans längd, och när indelningens finhet går mot noll bör vi få hela kurvans längd (i själva verket kan vi *definiera* begreppet kurvlängd på det här sättet), d.v.s. kurvans längd²¹

$$L(\Gamma) = \int_0^{\frac{3\pi}{4}} |\mathbf{r}'(t)| dt = \int_0^{\frac{3\pi}{4}} \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2} dt = \int_0^{\frac{3\pi}{4}} \sqrt{1 + t^2} dt \approx 3.81.$$

Exempel (volym av rotations kropp: "skivformeln"). Vi vill beräkna volymen V av en (rät cirkulär) kon med höjd h och radie r . Vi kan producera en sådan kon genom att rita den räta linjen $y = \frac{r}{h}x$ från $x = 0$ till $x = h$, och sedan rotera området mellan x -axeln och linjestycket kring x -axeln. För att beräkna volymen av den erhållna rotationskroppen (d.v.s. konen) delar vi in x -axeln mellan 0 och h i en massa småbitar. En allmän av dessa ligger vid koordinaten x och har bredd dx och höjden från x -axeln upp till kurvan är här $y = rx/h$. Denna del ger alltså under rotationen upphov till en skiva med volymen $dV = y^2 \pi dx = (rx/h)^2 \pi dx$, varför hela konens volym är

$$V = \int_0^h dV = \frac{r^2}{h^2} \pi \int_0^h x^2 dx = \frac{r^2}{h^2} \pi \cdot \frac{1}{3} h^3 = \frac{r^2 h \pi}{3}.$$

Detta är precis en tredjedel av volymen av "motsvarande" cylinder, just som vi fick lära oss på högskolastadiet.

Exempel (volym av rotations kropp: "skivformeln"). Låt oss även bestämma volymen V av ett klot med radien r . Denna kropp erhålles när området under kurvan $y = \sqrt{r^2 - x^2}$ roteras ett varv kring x -axeln, så volymen är

$$\begin{aligned} V &= \int_{-r}^r y^2 \pi dx = \int_{-r}^r (\sqrt{r^2 - x^2})^2 \pi dx = \pi \int_{-r}^r (r^2 - x^2) dx = \pi \left[r^2 x - \frac{1}{3} x^3 \right]_{-r}^r = \\ &= \pi \left(r^3 - \frac{1}{3} r^3 + r^3 - \frac{1}{3} r^3 \right) = \frac{4\pi r^3}{3}. \end{aligned}$$

Exempel (volym av rotations kropp: "rörformeln"). Om området $0 \leq x \leq 4$, $e^x \leq y \leq e^4$, roteras ett varv kring y -axeln uppkommer en rotationskropp. Vi vill bestämma dess volym V . Vi delar in x -axeln från 0 till 4 i en massa småbitar av längd dx . Den del av området som ligger ovanför en sådan bit, vid koordinaten x , har arean $dA = (e^4 - e^x) dx$ och volymen av det "rör" som uppkommer när denna del roteras kring y -axeln är ungefär $dV = 2\pi x dA = 2\pi x(e^4 - e^x) dx$. Hela rotationskroppens volym är sålunda

$$V = \int_0^4 dV = \int_0^4 2\pi x(e^4 - e^x) dx = [\text{Partiell integration}] = 2\pi(5e^4 - 1).$$

Exempel (area av rotationsyta). Låt oss beräkna arean A av en kon med höjd h och radie r . Utgå från samma upplägg som i tidigare konexemplet. Dela in x -axeln från $x = 0$ till $x = h$ i småbitar. Linjebi-

²¹ Det är inga problem att beräkna integralen exakt med standardmetoder från envariabelanalysen.

ten ovanför småbiten vid koordinaten x har kurvlängden $ds = \sqrt{1 + r^2/h^2} dx$ så arean av den "remsa" som uppkommer när den här linjebiten roterar kring x -axeln är

$$dA = 2\pi y ds = \frac{2\pi r x}{h} \sqrt{1 + \frac{r^2}{h^2}} dx.$$

Hela arean blir därför

$$A = \int_0^h dA = \frac{2\pi r}{h} \sqrt{1 + \frac{r^2}{h^2}} \int_0^h x dx = \frac{2\pi r}{h} \sqrt{1 + \frac{r^2}{h^2}} \cdot \frac{1}{2} h^2 = \pi r h \sqrt{1 + \frac{r^2}{h^2}} = \pi r \sqrt{h^2 + r^2}.$$

Exempel (area av rotationsyta). Låt oss också beräkna arean av en sfär med radie r . Sfären uppkommer när kurvan $y = \sqrt{r^2 - x^2}$ roterar ett varv kring x -axeln. Vi delar som vanligt in x -axeln från $x = -r$ till $x = r$ i småbitar. Delen vid koordinaten x har bredden dx och kurvan ovanför biten har längden

$$ds = \sqrt{1 + y'(x)^2} dx = \sqrt{1 + \left(-\frac{x}{\sqrt{r^2 - x^2}}\right)^2} dx = \sqrt{1 + \frac{x^2}{r^2 - x^2}} dx = \frac{r}{\sqrt{r^2 - x^2}} dx$$

vilket ger upphov till delarean $dA = 2\pi y \cdot ds = 2\pi \sqrt{r^2 - x^2} \cdot \frac{r}{\sqrt{r^2 - x^2}} dx = 2\pi r dx$ (Gud vad smidigt!!) så att hela arean blir

$$A = \int_{-r}^r dA = \int_{-r}^r 2\pi r dx = 2\pi r \int_{-r}^r dx = 4\pi r^2.$$

Exempel (integral som medelvärde). Antag att temperaturen i en lokal är $T(t)$ [°C] vid tidpunkten t [h]. Vad är medeltemperaturen under ett givet dygn, t.ex. då $t \in [0, 24[$?

Det här är förstås till viss del en definitionsfråga (vad *menas* med "medelvärdet" av en funktion på ett intervall?), så som i tidigare exempel kan den formel vi resonerar oss fram till tjäna som själva definitionen av den nämnda sortens "medelvärde".

Ett naturligt sätt att ta fram ett rimligt "medelvärde" är att *sampla* temperaturen vid jämna mellanrum, t.ex. varje timme. Vi erhåller då en *ändlig följd* av tal:

$$T(0), \quad T(1), \quad \dots, \quad T(22), \quad T(23).$$

Vi kan sedan beräkna det vanliga *aritmetiska* medelvärdet $\frac{1}{24} \sum_{t=0}^{23} T(t)$ av dessa tal. En bättre approximation av det "sanna" medelvärdet under dygnet bör vi få om vi samplar oftare, t.ex. varje minut, eller varje sekund.

Antag att vi väntar tiden Δt [h] mellan varje sampling, och att vi totalt gör n samplings med start då $t = 0$. Vi får då (med några rimliga val) sambandet $n\Delta t = 24$. Medelvärdet \bar{T}_n av alla mätvärden blir

$$\begin{aligned}\bar{T}_n &:= \frac{T(0) + T(\Delta t) + T(2\Delta t) + \dots + T((n-1)\Delta t)}{n} = \\ &= \frac{1}{24} \left[T(0) \cdot \frac{24}{n} + T(\Delta t) \cdot \frac{24}{n} + T(2\Delta t) \cdot \frac{24}{n} + \dots + T((n-1)\Delta t) \cdot \frac{24}{n} \right] = \\ &= \frac{1}{24} [T(0) \cdot \Delta t + T(\Delta t) \cdot \Delta t + T(2\Delta t) \cdot \Delta t + \dots + T((n-1)\Delta t) \cdot \Delta t].\end{aligned}$$

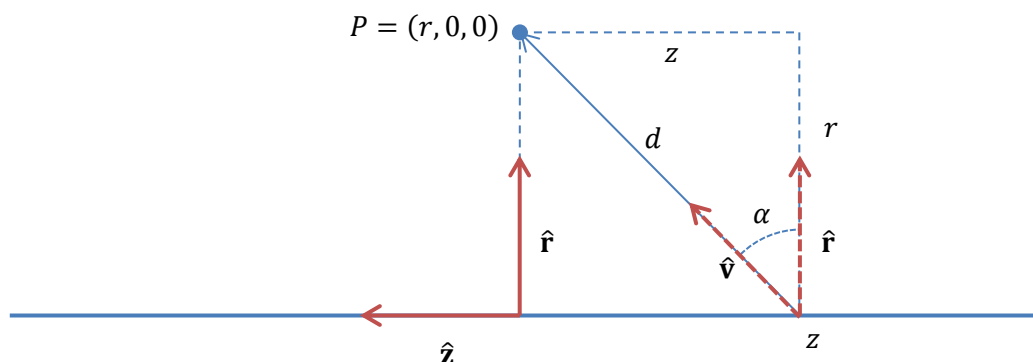
Men uttrycket inom parentes är ju en Riemannsumma för T över $[0, 24]$, så när $\Delta t \rightarrow 0$ (d.v.s. $n \rightarrow \infty$) får vi

$$\bar{T}_n \rightarrow \frac{1}{24} \int_0^{24} T(t) dt.$$

Av den här anledningen brukar man *definiera* medelvärdet av $f: [a, b] \rightarrow \mathbb{R}$ som $\frac{1}{b-a} \int_a^b f(x) dx$. Detta är alltså den "naturliga" kontinuerliga motsvarigheten till det diskreta aritmetiska medelvärdesbegreppet.

Exempel (*, elektrisk fältstyrka). Som sista exempel hämtar vi (från en av författarens fysiktexter) en trevlig beräkning från elektrostatiken. Säg att vi har en oändligt lång ledare med konstant laddningstäthet ρ [enhet: coulomb/meter]. Vad är den elektriska fältstyrkan runt ledaren?

Inför ett cylindriskt koordinatsystem sådant att ledaren ligger kring z -axeln. Av symmetriskäl är det klart att det elektriska fältet inte kan bero på de cylindriska koordinaterna z eller φ , och inte heller kan fältvektorn ha någon komponent i \hat{z} - eller $\hat{\varphi}$ -riktningen. Alltså är det elektriska fältet $\mathbf{E} = E(r)\hat{r}$. Betrakta nu punkten $P := (r, \varphi, z) = (r, 0, 0)$. Av våra symmetribetraktelser räcker det med att beräkna fältvektorn i denna punkt.



Betrakta en liten del dz av tråden, belägen vid koordinaten z . Avståndet mellan den här delen av tråden och fältpunkten P är $d(z) := \sqrt{z^2 + r^2}$ så den här delen av tråden, som har laddningen $dQ = \rho dz$, kommer enligt Coulombs lag²² att bidra med den elektriska fältstyrkan

$$d\mathbf{E}(z) = \frac{dQ}{4\pi\epsilon_0 d(z)^2} \hat{\mathbf{v}}(z) = \frac{\rho dz}{4\pi\epsilon_0 d(z)^2} \hat{\mathbf{v}}(z)$$

²² På gymnasiet brukar storheten $1/4\pi\epsilon_0$ kallas för "k", kort och gott.

där $\hat{\mathbf{v}}(z)$ är den normerade riktningsvektorn från trådbiten till fältpunkten. Eftersom vi redan insett att nettofältet i P (eller, för den delen, *någonstans*) bara har en radiell komponent, räcker det att betrakta enbart den radiella projektionen av varje $d\mathbf{E}(z)$. Denna är

$$dE_r(z) = \frac{\rho dz}{4\pi\epsilon_0 d(z)^2} \hat{\mathbf{v}}(z) \cdot \hat{\mathbf{r}} = \frac{\rho dz}{4\pi\epsilon_0 d(z)^2} \cos \alpha.$$

Men vinkeln

$$\alpha := \angle(\hat{\mathbf{v}}, \hat{\mathbf{r}}) = \arctan \frac{z}{r}$$

varför

$$\cos \alpha = \cos \arctan \frac{z}{r} = \frac{1}{\sqrt{1 + \frac{z^2}{r^2}}} = \frac{r}{\sqrt{r^2 + z^2}}$$

så att

$$dE_r(z) = \frac{r\rho dz}{4\pi\epsilon_0 (z^2 + r^2)^{3/2}}.$$

Det elektriska nettofältet i P är därför (det här är en *generaliserad* Riemannintegral)

$$E_r(r) = \int_{\mathbb{R}} dE_r(z) = \int_{\mathbb{R}} \frac{r\rho dz}{4\pi\epsilon_0 (z^2 + r^2)^{3/2}} = \frac{\rho}{2\pi\epsilon_0 r},$$

d.v.s.

$$\mathbf{E} = \frac{\rho}{2\pi\epsilon_0 r} \hat{\mathbf{r}}.$$

(Det här resultatet kan härledas betydligt smidigare med hjälp av metoder från vektoranalysen.)

11.1.1 Så... Vad är en integral?

I föregående avsnitt introducerade vi integralbegreppet från envariabelanalysen. Dels gav vi den bakomliggande idén bakom begreppet – en integral är gränsvärdet av en Riemannsumma –, dels gav vi ett par precisa definitioner av (Riemann)integralen. Sedan bjöd vi på ett smörgåsbord – nästan en matematisk orgie – av exempel på tillämpningar.

Poängen med föregående avsnitt var att få läsaren att inse vad en integral egentligen "är" för något: gränsvärdet av en Riemannsumma. Tyvärr är det nämligen så att många studenter såväl på gymnasienivå som på högskolenivå betraktar begreppet "integral" som synonymt med "area under graf". Förvisso är det sant integraler är det ideala verktyget för plana areaberäkningar, och en av de exempelapplikationer vi gav var just beräkning av plan area. Men i många andra integraltillämpningar är det högst onaturligt att tala om "area under graf".

Till exempel, när vi beräknade massan hos en tråd med variabel densitet, så tänkte vi inte så mycket på "area under graf" (även om det naturligtvis är sant att trådens massa är samma tal som arean under densitetsgrafen). Och när vi beräknade längden av kurvan $\Gamma := \mathbf{r} \left(\left[0, \frac{4\pi}{3} \right] \right)$ där $\mathbf{r}(t) :=$

$t(\cos t, \sin t)$ så landade vi i integralen $\int_0^{\frac{3\pi}{4}} \sqrt{1+t^2} dt$, men vi betraktade ingen plan area. Visst,

längden av kurvan är samma tal som arean under grafen $y = \sqrt{1+x^2}$, men vad har det med saken

att göra? Det är knappast så att man direkt inser "jaha, längden av kurvan $\Gamma := \mathbf{r} \left(\left[0, \frac{4\pi}{3} \right] \right)$ där $\mathbf{r}(t) := t(\cos t, \sin t)$ är så klart lika med arean under grafen $y = \sqrt{1+x^2}$ ".

De övriga exemplen i avsnittet saknar också koppling till begreppet "area under graf". Slutsaten torde vara att man i de flesta integraltillämpningar inte alls tänker på areor under grafer. Det man tänker på i stället är den grundläggande *idén* bakom integralbegreppet: uppdelning i småbitar, bildandet av Riemannsumma, och gränsövergång (om läsaren ursäktar Oxfordkommat).

Att tänka på integraler som "areor under grafer" kan sålunda vara handikappande. Dessutom kan en integral vara negativ (vilket inte en area kan vara); tolkningen att $\int_a^b f(x)dx$ är "arean under grafen $y = f(x)$ " gäller ju bara om f är en positiv funktion.

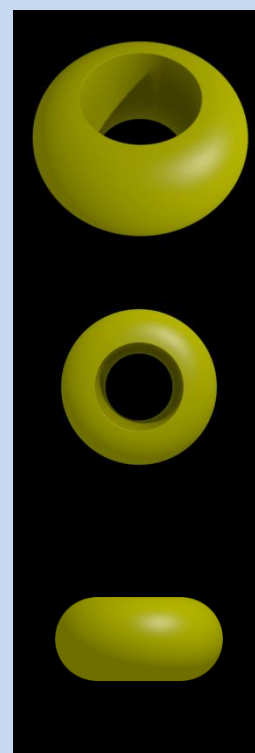
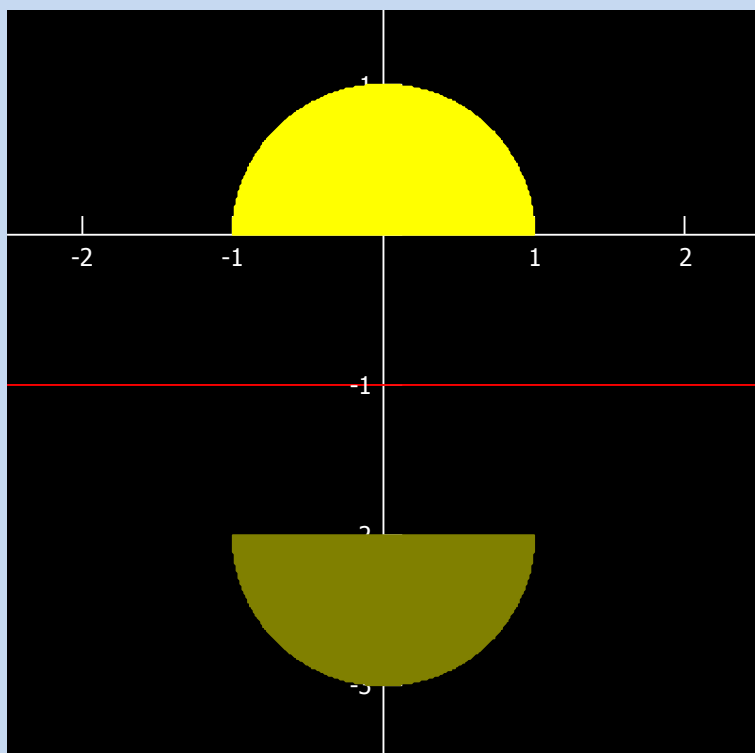
En tredje – och ganska central – anledning till att areaassociationen är olämplig på högskolenivå är att generaliseringen av integralbegreppet till högre dimensioner är fullständigt trivial om man tänker på integraler som gränsvärden av Riemannsummor, men väldigt knepig om man tänker på integraler som "areor under grafer".

11.1.2 Ett par extra exempel

Vi avslutar med två extra exempel på tillämpningar av envariabelintegraler, eftersom resultaten är intressanta och problemen egentligen kan betraktas som flervariabelproblem som "råkar" vara hanterbara med "rena" envariabelmetoder.

Exempel

Beräkna volymen av den kropp som uppkommer när området $D := \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq \sqrt{1-x^2}\}$ roteras ett varv kring linjen $y = -1$. (Notera att den erhållna kroppen är en osedvanligt otyplig servettring.)



Lösning: Vi delar in x -axeln mellan $x = -1$ och $x = 1$ i småbitar av längd dx . Småbiten vid koordinaten x ger vid rotationen upphov till en kropp av volymen $A(x)dx$ där tvärsnittsarean $A(x) := (\sqrt{1-x^2} + 1)^2 \pi - 1^2 \pi$ så att hela den sökta volymen är

$$\begin{aligned} V &:= \int_{-1}^1 A(x)dx = \pi \int_{-1}^1 \left((\sqrt{1-x^2} + 1)^2 - 1 \right) dx = \pi \int_{-1}^1 (1 - x^2 + 2\sqrt{1-x^2}) dx = \\ &= \pi \int_{-1}^1 (1 - x^2)dx + 2\pi \int_{-1}^1 \sqrt{1-x^2} dx \end{aligned}$$

där

$$\int_{-1}^1 (1 - x^2)dx = \left[x - \frac{1}{3}x^3 \right]_{-1}^1 = 1 - \frac{1}{3} + 1 - \frac{1}{3} = \frac{4}{3}$$

och

$$\begin{aligned} \int_{-1}^1 \sqrt{1-x^2} dx &= \left[\begin{array}{l} x = \sin t \\ t \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ dx = \cos t dt \end{array} \right] = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{1 - \sin^2 t} \cos t dt = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{\cos^2 t} \cos t dt = \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} |\cos t| \cos t dt = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 t dt = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\frac{1}{2} \cos 2t + \frac{1}{2} \right) dt = \\ &= \left[\frac{1}{4} \sin 2t + \frac{1}{2} t \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = \frac{\pi}{2} \end{aligned}$$

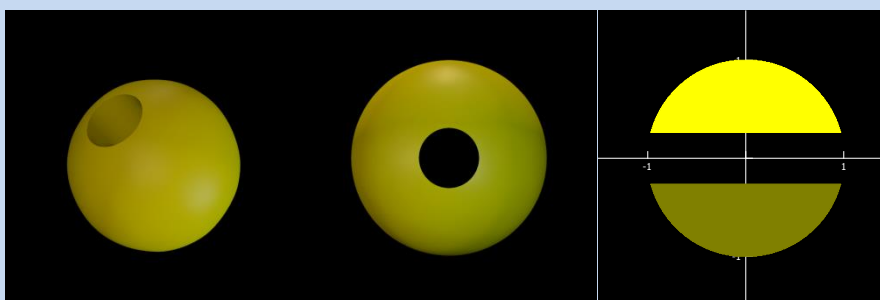
så att

$$V = \pi \cdot \frac{4}{3} + 2\pi \cdot \frac{\pi}{2} = \frac{4\pi}{3} + \pi^2$$

Svar: Volymen är $\frac{4\pi}{3} + \pi^2$.

Exempel

Ur ett klot av radie R borrar ett cylindriskt hål, rakt genom klotets mittpunkt, med radie $r < R$. Beräkna volymen av det som återstår av klotet. (Notera att även den här kroppen kan betraktas som en servettring om hålet är stort. Om hålet är litet rör det sig i stället om en ljusstake som inte kan stå ordentligt på en plan yta (varför?). Om hålet är ännu mindre rör det sig om en pärla som man kan trä en tråd igenom.)



Lösning: Återstoden av klotet är den kropp som uppkommer när området

$$D := \{(x, y) \in \mathbb{R}^2 : r \leq y \leq \sqrt{R^2 - x^2}\}$$

roteras ett varv kring x -axeln; detta sträcker sig från $x = -\sqrt{R^2 - r^2}$ till $x = \sqrt{R^2 - r^2}$ så att den sökta volymen är

$$\begin{aligned} V &:= \int_{-\sqrt{R^2-r^2}}^{\sqrt{R^2-r^2}} \left((\sqrt{R^2-x^2})^2 \pi - r^2 \pi \right) dx = \pi \int_{-\sqrt{R^2-r^2}}^{\sqrt{R^2-r^2}} (R^2 - x^2 - r^2) dx = \\ &= \pi \left[R^2 x - \frac{1}{3} x^3 - r^2 x \right]_{-\sqrt{R^2-r^2}}^{\sqrt{R^2-r^2}} = 2\pi \left((R^2 - r^2) \sqrt{R^2 - r^2} - \frac{1}{3} (R^2 - r^2)^{\frac{3}{2}} \right) = \\ &= \frac{4\pi}{3} (R^2 - r^2)^{3/2}. \end{aligned}$$

Svar: Volymen är $\frac{4\pi}{3} (R^2 - r^2)^{3/2}$.

Övningar

1. Kontrollera rimligheten i det sista resultatet genom att pröva $r = 0$ och $r = R$.
2. De två senaste exemplen löstes med skivformeln. Pröva att lösa dem med rörformeln!

11.2 Dubbelintegraler

Vi kommer i det här dokumentet inte att ge någon formell definition av begreppet dubbelintegral som vi gjorde i envariabelfallet i en asteriskmarkerad anmärkning (läsaren är i stället hänvisad till ordinarie kurslitteratur), men vi kommer att ge den grundläggande idén. Eftersom dubbelintegraler är för skalärfält i planet vad enkelintegraler är för funktioner på tallinjen, kommer läsaren att känna igen sig. Mycket.

Om $F: D_F \rightarrow \mathbb{R}$ är ett skalärfält i planet (d.v.s. $D_F \subset \mathbb{R}^2$) och $D \subset D_F$ så är dubbelintegralen $\iint_D F(x, y) dx dy$, om den existerar, ett tal. Precis som i envariabelfallet är integralen inget annat än "gränsvärdet av en Riemannsumma". Idén är att man delar in D i N stycken småbitar. I varje småbit är $F(x, y)$ ungefär konstant (om F är snäll nog). I varje bit väljer vi något funktionsvärde och gångrar detta med bitens *area*; sedan summerar vi dessa produkter, varvid vi erhåller ett tal. Integralen $\iint_D F(x, y) dx dy$ är gränsvärdet av dessa s.k. Riemannsummor när indelningens finhet går mot noll (d.v.s. när vi tar fler och fler, mindre och mindre, bitar).

Mer specifikt, säg att vi delar in D i de N stycken småbitarna $(D_k)_{k=1}^N$. För varje $k = 1, \dots, N$ väljer vi en punkt $(x_k, y_k) \in D_k$. Riemannsumman är då

$$F(x_1, y_1)A(D_1) + F(x_2, y_2)A(D_2) + \dots + F(x_N, y_N)A(D_N)$$

där $A(D_k)$ är arean av småbiten D_k .

Notera i synnerhet att integralen av den konstanta funktionen 1 ger D 's *totala area*, d.v.s.

$$\iint_D 1 dx dy = A(D).$$

Exempel (massa av skiva). Antag att en plastskiva upptar rektangeln $D := [0, 1] \times [0, 1]$ i xy -planet och att densiteten i punkten (x, y) är lika med $\rho(x, y)$ [enhet: g/cm^2]. Vad är hela skivans massa?

Ett naturligt angreppssätt är att dela in rektangeln i en miljon småbitar (t.ex. lika stora kvadrater, 1000×1000 stycken); kalla dessa småbitar $D_1, \dots, D_{1000000}$. I varje småbit är densiteten $\rho(x, y)$ ungefär konstant. Om vi väljer någon punkt $(x_k, y_k) \in D_k$ för varje k så är därför massan i biten D_k ungefär lika med $\rho(x_k, y_k)A(D_k)$ [enhet: $\text{g/cm}^2 \cdot \text{cm}^2 = \text{g}$]. Hela skivans massa är således ungefär

$$\rho(x_1, y_1)A(D_1) + \rho(x_2, y_2)A(D_2) + \dots + \rho(x_{1000000}, y_{1000000})A(D_{1000000}).$$

Vi bör få en bättre approximation om vi gör om samma sak fast med finare indelning, d.v.s. det är rimligt att tolka $\iint_D \rho(x, y) dx dy$ som den totala massan hos skivan.

Exempel (volym under graf). Låt $F: \mathbb{R}^2 \rightarrow \mathbb{R}^+$ vara ett kontinuerligt och positivt skalärfält i planet, och låt $D \subset \mathbb{R}^2$. Vad är volymen under grafen $z = F(x, y)$? Ett naturligt sätt att angripa problemet är att dela in D i N stycken småbitar D_k . I varje småbit är "takhöjden" $F(x, y)$ ungefär konstant, så volymen av den del av "byggnaden" som ligger över D_k är ungefär $F(x_k, y_k)A(D_k)$ där (x_k, y_k) är en godtyckligt vald punkt i D_k . Hela "byggnadens" volym är sålunda, approximativt,

$$F(x_1, y_1)A(D_1) + F(x_2, y_2)A(D_2) + \dots + F(x_N, y_N)A(D_N).$$

Vi bör få en bättre approximation om vi gör om samma sak fast med finare indelning, d.v.s. det är rimligt att tolka $\iint_D F(x, y) dx dy$ som volymen under grafen $z = F(x, y)$ över D .

Idén med Riemannsummor säger inget om hur man beräknar dubbelintegraler i praktiken, men precis som i envariabelfallet är det ändå viktigt att studenten förstår idén bakom begreppet. När man i praktiken skall beräkna en dubbelintegral använder man tekniker som itererad enkelintegration och variabelbyte.

11.3 Trippelintegraler

Trippelintegraler är för skalärfält i rummet vad dubbelintegraler är för skalärfält i planet.

Om $F: D_F \rightarrow \mathbb{R}$ är ett skalärfält i rummet (d.v.s. $D_F \subset \mathbb{R}^3$) och $D \subset D_F$ så är trippelintegralen $\iiint_D F(x, y, z) dx dy dz$, om den existerar, ett tal. Idén är att man delar in D i N stycken småbitar. I varje småbit är $F(x, y, z)$ ungefär konstant (om F är snäll nog). I varje bit väljer vi något funktionsvärde och gångrar detta med bitens volym; sedan summerar vi dessa produkter, varvid vi erhåller ett tal. Integralen $\iiint_D F(x, y, z) dx dy dz$ är gränsvärdet av dessa s.k. Riemannsummor när indelningens finhet går mot noll (d.v.s. när vi tar fler och fler, mindre och mindre, bitar).

Mer specifikt, säg att vi delar in D i de N stycken småbitarna $(D_k)_{k=1}^N$. För varje $k = 1, \dots, N$ väljer vi en punkt $(x_k, y_k, z_k) \in D_k$. Riemannsumman är då

$$F(x_1, y_1, z_1)V(D_1) + F(x_2, y_2, z_2)V(D_2) + \dots + F(x_N, y_N, z_N)V(D_N)$$

där $V(D_k)$ är volymen av småbiten D_k .

Notera i synnerhet att integralen av den konstanta funktionen 1 ger D 's totala volym, d.v.s. $\iiint_D 1 dx dy dz = V(D)$.

Exempel (massa av kropp). Antag att ett föremål upptar området D i rummet och att densiteten i punkten (x, y, z) är lika med $\rho(x, y, z)$ [enhet: g/cm^3]. Vad är hela kroppens massa?

Ett naturligt angreppssätt är att dela in D i N stycken småbitar; kalla dessa småbitar D_1, \dots, D_N . I varje småbit är densiteten $\rho(x, y, z)$ ungefär konstant. Om vi väljer någon punkt $(x_k, y_k, z_k) \in D_k$ för varje k så är därför massan i biten D_k ungefär lika med $\rho(x_k, y_k, z_k)V(D_k)$ [enhet: $\text{g/cm}^3 \cdot \text{cm}^3 = \text{g}$]. Hela kroppens massa är således ungefär

$$\rho(x_1, y_1, z_1)V(D_1) + \rho(x_2, y_2, z_2)V(D_2) + \dots + \rho(x_N, y_N, z_N)V(D_N).$$

Vi bör få en bättre approximation om vi gör om samma sak fast med finare indelning, d.v.s. det är rimligt att tolka $\iiint_D \rho(x, y, z) dx dy dz$ som den totala massan hos kroppen.

Idén med Riemannsummor säger inget om hur man beräknar trippelintegraler i praktiken, men det är ändå viktigt att studenten förstår idén bakom begreppet. När man i praktiken skall beräkna en trippelintegral använder man tekniker som itererad enkelintegration och variabelbyte.

11.4 Multipelintegraler av högre dimension

På liknande sätt som i planet och rummet definieras multipelintegraler i \mathbb{R}^n i allmänhet. För att t.ex. approximera $\iiint_D F(x, y, z, w) dx dy dz dw$ över $D \subset \mathbb{R}^4$ kan man dela in D i småbitar, och i varje bit gånga $F(x, y, z, w)$ med bitens hypervolym; sedan summerar man, och integralen är gränsvärdet när indelningen blir finare. Som med dubbel- och trippelintegraler beräknar man i praktiken (om möjligt) allmänna multipelintegraler med hjälp av tekniker som itererad enkeltintegration och variabelbyte.

11.5 Enkla exempel

Vi förutsätter i exemplen nedan att läsaren är bekant med standardteknikerna för beräkning av multipelintegraler: itererad enkelintegration och variabelbyte.

Exempel

Beräkna $\iint_D (1 + 2x + y) dx dy$ där D är (den fyllda) triangeln med hörn i $(0,0)$, $(1,0)$ och $(0,1)$.

Lösning: x sträcker sig från 0 till 1 och för varje fixt x går y från 0 till $1 - x$, så integralen är

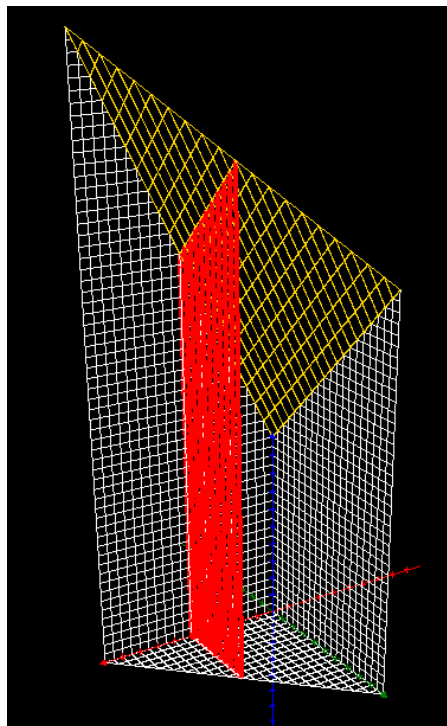
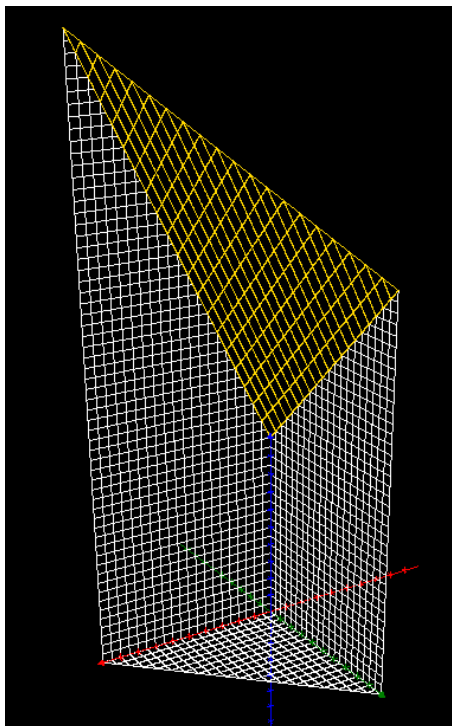
$$\begin{aligned} \iint_D (1 + 2x + y) dx dy &= \int_0^1 \left(\int_0^{1-x} (1 + 2x + y) dy \right) dx = \int_0^1 \left(\left[y + 2xy + \frac{1}{2}y^2 \right]_{y=0}^{1-x} \right) dx = \\ &= \int_0^1 \left(-\frac{3}{2}x^2 + \frac{3}{2} \right) dx = \left[-\frac{1}{2}x^3 + \frac{3}{2}x \right]_0^1 = -\frac{1}{2} + \frac{3}{2} = 1. \end{aligned}$$

Svar: $\iint_D (2x + y) dx dy = 1$ där D är (den fyllda) triangeln med hörn i $(0,0)$, $(1,0)$ och $(0,1)$.

Anmärkning 1: Eftersom skalärfältet $F(x, y) := 1 + 2x + y$ är positivt på D är integralen $\iint_D (1 + 2x + y) dx dy$ lika med volymen under grafen $z = 1 + 2x + y$ ovanför D . Vi kan alltså tolka detta värde som volymen av en trekantig byggnad med lutande tak, se bilden nedan till vänster.

Den här bilden erbjuder också ett mycket trevligt troliggörande av formeln för itererad enkelintegration. Byggnaden sträcker sig tydligen från 0 till 1 i x -led. Välj något $x_0 \in [0,1]$ och snitta byggnaden med planet $x = x_0$, se bilden nedan, till höger. Arealen av den här (tänkta) "väggen" är $A(x_0) := \int_0^{1-x_0} (1 + 2x_0 + y) dy$ (area under graf!) och dess tjocklek är dx , så dess volym är $dV := A(x_0) dx = \left(\int_0^{1-x_0} (1 + 2x_0 + y) dy \right) dx$. Hela byggnadens volym erhålles om vi summerar dessa väggars voly-

mer, d.v.s. hela volymen är $\int_0^1 \left(\int_0^{1-x_0} (1 + 2x_0 + y) dy \right) dx_0$ vilket är precis integralen ovan. (Tänk på Riemannsummorna till den yttre integralen för att se att detta är rätt och rimligt.)



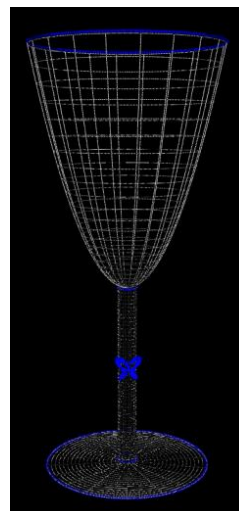
Anmärkning 2: I exemplet beräknade vi integralen $\iint_D (1 + 2x + y) dx dy$ där $D = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1 - x\}$ med hjälp av itererad enkelintegration. Men även utan den beräkningen kan vi mycket lätt uppskatta integralens värde. Minsta värdet av $F(x, y) := 1 + 2x + y$ på D är ju 1 (och erhålles i origo), medan största värdet är 3 (och erhålles i $(1,0)$). D 's area är $\frac{1}{2}$, så vi ser direkt att

$$\frac{1}{2} \leq \iint_D (1 + 2x + y) dx dy \leq \frac{3}{2}.$$

Vi kan i själva verket stänga in integralen i ett hur litet intervall som helst genom att dela in D i mindre bitar, och i varje bit välja största och minsta värdet av F , och bilda motsvarande över- och undersummor.

Exempel

Insidan av ett dryckesglas har formen av paraboloiden $z = x^2 + y^2$ och glaset är fyllt med Coca-Cola upp till $z = 5$ cm. Vad är volymen av drycken i glaset?



Lösning: Om D betecknar området i glaset där drycken är, d.v.s. om $D = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 \leq z \leq 5\}$, så är volymen $V(D) := \iiint_D dx dy dz$. Det finns två rimliga angreppssätt att beräkna $V(D)$.

Metod 1: Skivor

Vi kan dela in z -axeln i småbitar av höjd dz . Drycken vid höjd z är då formad som en "skiva" (eg. cylinder) av höjd dz och radie $r(z) := \sqrt{z}$, och därmed tvärsnittsarean $A(z) := r(z)^2\pi = z\pi$; skivans volym är alltså $dV = A(z)dz = z\pi dz$ så att hela vätskans volym är

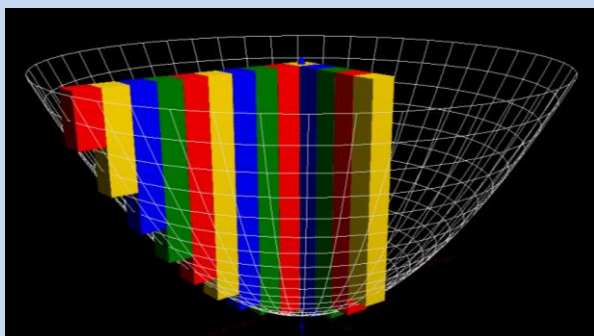
$$V(D) = \iiint_D dx dy dz = \int_0^5 dV = \int_0^5 A(z) dz = \int_0^5 z\pi dz = \frac{25\pi}{2}.$$



Metod 2: Staplar

Projektionen av D på xy -planet är disken $\tilde{D} := \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 \leq \sqrt{5}, z = 0\} \simeq \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq \sqrt{5}\}$. Vi kan dela in denna i smårektanglar av area $dA = dx dy$. Vätskan ovanför en sådan rektangel, vid koordinaterna x och y , är en pelare som sträcker sig från $z = x^2 + y^2$ upp till $z = 5$, så att dess höjd är $h(x, y) := 5 - x^2 - y^2$ och dess volym är $dV = h(x, y) dx dy = (5 - x^2 - y^2) dx dy$. Hela vätskans volym är sålunda

$$\begin{aligned} V(D) &= \iiint_D dx dy dz = \iint_{\tilde{D}} dV = \iint_{\tilde{D}} h(x, y) dx dy = \iint_{\tilde{D}} (5 - x^2 - y^2) dx dy = \\ &= \left[\begin{array}{l} x = \rho \cos \varphi \\ y = \rho \sin \varphi \\ J = \rho \end{array} \right] = \iint_E (5 - \rho^2) \rho d\rho d\varphi = \int_0^{\sqrt{5}} (5\rho - \rho^3) d\rho \int_0^{2\pi} d\varphi = \\ &= \frac{25}{4} \cdot 2\pi = \frac{25\pi}{2}. \end{aligned}$$



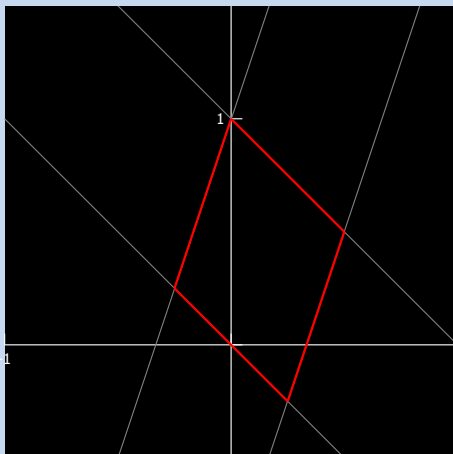
Svar: Vätskans volym är $25\pi/2 \text{ cm}^3$.

Exempel

Beräkna $\iint_D (x - y) dx dy$ där D är den (fyllda) parallelogrammen med hörn i $(0,1)$, $(\frac{1}{2}, \frac{1}{2})$ och $\pm(\frac{1}{4}, -\frac{1}{4})$.

Lösning: Genom att bestämma ekvationerna för de fyra linjer som begränsar D ser vi att

$$D = \{(x, y) \in \mathbb{R}^2: 0 \leq x + y \leq 1, -1 \leq 3x - y \leq 1\}.$$



Det är sålunda lämpligt att utföra variabelbytet

$$\begin{aligned} u &:= x + y \\ v &:= 3x - y \end{aligned}$$

ty då gäller

$$\begin{cases} 0 \leq x + y \leq 1 \\ -1 \leq 3x - y \leq 1 \end{cases} \iff \begin{cases} 0 \leq u \leq 1 \\ -1 \leq v \leq 1, \end{cases}$$

d.v.s. i nya koordinaterna svarar D mot rektangeln $E := [0,1] \times [-1,1]$. Funktionaldeterminanten för funktionen $(x, y) \mapsto (u, v)$ är -4 ; eftersom variabelbytet är "åt fel håll" är funktionaldeterminanten vi behöver infoga i integranden lika med $\frac{1}{4}$ till beloppet.

Vi ser direkt att $u + v = 4x$ och $v - 3u = -4y$ så att det inversa variabelbytet är $x = \frac{1}{4}(u + v)$, $y = \frac{1}{4}(3u - v)$. Speciellt är $x - y = \frac{1}{2}(v - u)$ varför

$$\iint_D (x - y) dx dy = \frac{1}{8} \iint_E (v - u) dx dy = \frac{1}{8} \int_0^1 \left(\int_{-1}^1 (v - u) dv \right) du = -\frac{1}{8}.$$

Svar: $\iint_D (x - y) dx dy = -\frac{1}{8}$ där D är den (fyllda) parallelogrammen med hörn i $(0,1)$, $(\frac{1}{2}, \frac{1}{2})$ och $\pm(\frac{1}{4}, -\frac{1}{4})$.

Anmärkning: Uppgiften kan också lösas utan variabelbyte, men man behöver då dela in området i tre delar (oavsett om man väljer att integrera med x eller med y innerst). Varför?

Exempel

Beräkna $\iint_D x^2 dx dy$ där $D := \{(x, y) \in \mathbb{R}^2: x^2 + 4y^2 \leq 4\}$.

Lösning: Villkoret för D kan skrivas

$$\left(\frac{x}{2}\right)^2 + y^2 \leq 1$$

så D är en fylld ellips med halvaxlar parallella med koordinataxlarna vilka sträcker sig från -2 till 2 i x -led och från -1 till 1 i y -led (rita!). Vi inför därför "elliptiska koordinater" enligt

$$\begin{aligned}x &= 2\rho \cos \varphi \\y &= \rho \sin \varphi\end{aligned}$$

så att D svarar mot rektangeln $E := [0, 1] \times [0, 2\pi[$ i (ρ, φ) -planet. Funktionaldeterminanten är 2ρ , så

$$\iint_D x^2 dx dy = \iint_E 4\rho^2 \cos^2 \varphi \cdot 2\rho d\rho d\varphi = 8 \int_0^1 \rho^3 d\rho \int_0^{2\pi} \cos^2 \varphi d\varphi = 8 \cdot \frac{1}{4} \cdot \pi = 2\pi.$$

Svar: $\iint_D x^2 dx dy = 2\pi$ där $D = \{(x, y) \in \mathbb{R}^2: x^2 + 4y^2 \leq 4\}$.

Anmärkning: Variabelbytet i föregående uppgift kan göras i två steg. Först kan vi skala koordinatsystemet enligt $u = x/2$, $v = y$, och sedan kan vi införa vanliga polära koordinater $u = \rho \cos \varphi$, $v = \rho \sin \varphi$. Vad är de två funktionaldeterminanterna? Vad är deras produkt? Vilket område får vi i uv -planet? Vilket område får vi i $\rho\varphi$ -planet? Prova gärna att lösa uppgiften på det sättet!

Exempel

Beräkna $\iint_D (x^2 + y^2) dx dy$ där $D = \{(x, y) \in \mathbb{R}^2: x^2 - 2x + y^2 \leq 0\}$.

Lösning: Eftersom

$$x^2 - 2x + y^2 \leq 0 \quad \iff \quad (x - 1)^2 + y^2 \leq 1$$

är D den fyllda cirkeln med radie 1 kring $(1, 0)$. Med tanke på området²³ inför vi förskjutna polära koordinater (ρ, φ) enligt

$$\begin{aligned}x &= 1 + \rho \cos \varphi \\y &= \rho \sin \varphi\end{aligned}$$

så att området D svarar mot rektangeln $E := [0, 1] \times [0, 2\pi[$ i $\rho\varphi$ -planet. Funktionaldeterminanten är ρ och vi ser att $x^2 + y^2 = 1 + 2\rho \cos \varphi + \rho^2 \cos^2 \varphi + \rho^2 \sin^2 \varphi = 1 + 2\rho \cos \varphi + \rho^2$ så att

²³ Om man tittar på *integranden* är det lämpligt att införa vanliga polära koordinater ($x = \rho \cos \varphi$, $y = \rho \sin \varphi$) ty då blir $x^2 + y^2$ helt enkelt ρ^2 ; däremot blir inte området trivialt. Om man i stället tittar på området är det mycket lämpligt att införa polära koordinater kring punkten $(1, 0)$, d.v.s. $x = 1 + \rho \cos \varphi$, $y = \rho \sin \varphi$, ty då blir området i nya koordinatsystemet rektangeln $[0, 1] \times [0, 2\pi[$; däremot blir inte integranden lika trivial. Vi måste göra ett val, och vi väljer det senare alternativet, som visar sig ge ganska enkla räkningar.

$$\begin{aligned} \iint_D (x^2 + y^2) dx dy &= \iint_E (1 + 2\rho \cos \varphi + \rho^2) \rho d\rho d\varphi = \int_0^1 \left(\int_0^{2\pi} (\rho + 2\rho^2 \cos \varphi + \rho^3) d\varphi \right) d\rho = \\ &= \dots = \frac{3\pi}{2}. \end{aligned}$$

Svar: $\iint_D (x^2 + y^2) dx dy = \frac{3\pi}{2}$ där $D = \{(x, y) \in \mathbb{R}^2: x^2 - 2x + y^2 \leq 0\}$.

Anmärkning: Variabelbytet i föregående uppgift kan göras i två steg. Först kan vi translatera koordinatsystemet enligt $u = x - 1$, $v = y$, och sedan kan vi införa vanliga polära koordinater $u = \rho \cos \varphi$, $v = \rho \sin \varphi$. Vad är de två funktionaldeterminanterna? Vad är deras produkt? Vilket område får vi i uv -planet? Vilket område får vi i $\rho\varphi$ -planet? Prova gärna att lösa uppgiften på det här sättet!

Exempel

Beräkna $\iint_D \sqrt{x^2 + y^2} dx dy$ där $D = \{(x, y) \in \mathbb{R}^2: x^2 - 2x + y^2 \leq 0\}$.

Lösning: Eftersom

$$x^2 - 2x + y^2 \leq 0 \iff (x - 1)^2 + y^2 \leq 1$$

är D den fyllda cirkeln med radie 1 kring $(1, 0)$. Med tanke på *integranden*²⁴ inför vi vanliga polära koordinater (ρ, φ) enligt $x = \rho \cos \varphi$, $y = \rho \sin \varphi$; $\sqrt{x^2 + y^2}$ blir då ρ . För att beskriva området i polära koordinater noterar vi först att varje punkt i D har $\varphi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ och att r -koordinaterna i D för varje sådan vinkel sträcker sig från 0 till cirkelns "bortre kant". Cirkeln har ekvationen $x^2 - 2x + y^2 = 0$ som i polära koordinater lyder $\rho^2 - 2\rho \cos \varphi = 0$ vilket ger $\rho = 0$ (origo) eller $\rho - 2 \cos \varphi = 0$, d.v.s. $\rho = 2 \cos \varphi$, som alltså är det sökta bortre avståndet för varje fixt $\varphi \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Med

$$E := \left\{ (\rho, \varphi) \in \mathbb{R}^2: -\frac{\pi}{2} \leq \varphi \leq \frac{\pi}{2}, \rho \leq 2 \cos \varphi \right\}$$

har vi därför

$$\begin{aligned} \iint_D \sqrt{x^2 + y^2} dx dy &= \iint_E \rho \cdot \rho d\rho d\varphi = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\int_0^{2 \cos \varphi} \rho^2 d\rho \right) d\varphi = \frac{1}{3} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} [\rho^3]_{\rho=0}^{2 \cos \varphi} d\varphi = \\ &= \frac{8}{3} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^3 \varphi d\varphi = \frac{8}{3} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 \varphi \cos \varphi d\varphi = \frac{8}{3} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (1 - \sin^2 \varphi) \cos \varphi d\varphi = \\ &= \left[s := \sin \varphi \right] = \frac{8}{3} \int_{-1}^1 (1 - s^2) ds = \frac{32}{9} \end{aligned}$$

varför vi är klara.

Svar: $\iint_D \sqrt{x^2 + y^2} dx dy = \frac{32}{9}$ där $D = \{(x, y) \in \mathbb{R}^2: x^2 - 2x + y^2 \leq 0\}$.

²⁴ Se föregående fotnot. Prova gärna att byta koordinater så att området blir fint i stället. Hur blir integranden?

Exempel

Beräkna arean innanför en cirkel med radie r med hjälp av lämpligt variabelbyte i dubbelintegral.

Lösning: Inför disken $D := \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq r^2\}$ så att den sökta arean är $A(D) = \iint_D dx dy$.

Med polärt variabelbyte

$$\begin{aligned}x &= \rho \cos \varphi \\y &= \rho \sin \varphi \\J &= \rho\end{aligned}$$

(där J står för funktionaldeterminantens belopp) svarar disken D mot rektangeln $E := [0, r] \times [0, 2\pi[$ i (ρ, φ) -planet, varför

$$A(D) = \iint_D dx dy = \iint_E \rho d\rho d\varphi = \int_0^r \rho d\rho \int_0^{2\pi} d\varphi = \frac{1}{2}r^2 \cdot 2\pi = r^2\pi.$$

Exempel

Antag att densiteten i punkten (x, y) på enhetsdisken $x^2 + y^2 \leq 1$ är $\rho(x, y) := 2 - x^2 - y^2$. Vad är hela enhetsdiskens massa?

Lösning: Om D betecknar enhetsdisken så är massan

$$\begin{aligned}M &:= \iint_D \rho(x, y) dx dy = \iint_D (2 - x^2 - y^2) dx dy = \left[\begin{array}{l} x = \rho \cos \varphi \\ y = \rho \sin \varphi \\ J = \rho \\ (\rho, \varphi) \in E := [0, 1] \times [0, 2\pi[\end{array} \right] = \\ &= \iint_E (2 - \rho^2) \rho d\rho d\varphi = \int_0^1 (2\rho - \rho^3) d\rho \int_0^{2\pi} d\varphi = \left[\rho^2 - \frac{1}{4}\rho^4 \right]_0^1 \cdot 2\pi = \\ &= 2\pi \left(1 - \frac{1}{4} \right) = \frac{3\pi}{2}.\end{aligned}$$

Svar: Massan är $\frac{3\pi}{2}$.

Exempel

Beräkna volymen av ett klot av radie R med hjälp av lämpligt variabelbyte i trippelintegral.

Lösning: Med byte till sfäriska koordinater svarar klotet $K = \{(x, y, z) \in \mathbb{R}^3: x^2 + y^2 + z^2 \leq R^2\}$ mot rätklocket $[0, R] \times [0, \pi] \times [0, 2\pi[$ i $r\theta\varphi$ -rummet. Vi får volymen

$$\begin{aligned}V(K) &= \iiint_K dx dy dz = \left[\begin{array}{l} x = r \sin \theta \cos \varphi \\ y = r \sin \theta \sin \varphi \\ z = r \cos \theta \\ J = r^2 \sin \theta \\ (r, \theta, \varphi) \in E := [0, R] \times [0, \pi] \times [0, 2\pi[\end{array} \right] = \iiint_E r^2 \sin \theta dr d\theta d\varphi = \\ &= \int_0^R r^2 dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\varphi = \frac{1}{3}R^3 \cdot 2 \cdot 2\pi = \frac{4\pi R^3}{3}.\end{aligned}$$

Exempel

Betrakta kroppen $K := \{(x, y, z) \in \mathbb{R}^3 : 1 \leq x^2 + y^2 + z^2 \leq 4, z \geq 0\}$ och antag att densiteten i punkten (x, y, z) är $\rho(x, y, z) := x^2 + y^2 + z^2$. Vad är K 's massa?

Lösning: Massan är

$$\begin{aligned}
 M &:= \iiint_K \rho(x, y, z) dx dy dz = \iiint_K (x^2 + y^2 + z^2) dx dy dz = \\
 &= \left[\begin{array}{l} x = r \sin \theta \cos \varphi \\ y = r \sin \theta \sin \varphi \\ z = r \cos \theta \\ J = r^2 \sin \theta \\ (r, \theta, \varphi) \in E := [1, 2] \times [0, \pi/2] \times [0, 2\pi[\end{array} \right] = \iiint_E r^2 r^2 \sin \theta dr d\theta d\varphi = \\
 &= \int_1^2 r^4 dr \int_0^{\pi/2} \sin \theta d\theta \int_0^{2\pi} d\varphi = \frac{31}{5} \cdot 1 \cdot 2\pi = \frac{62\pi}{5}.
 \end{aligned}$$

Svar: Massan är $\frac{62\pi}{5}$.

Exempel

Antag att densiteten i punkten (x, y) på enhetsdisken $x^2 + y^2 \leq 1$ är $\rho(x, y) := e^{-x^2 - y^2}$. Vad är hela enhetsdiskens massa?

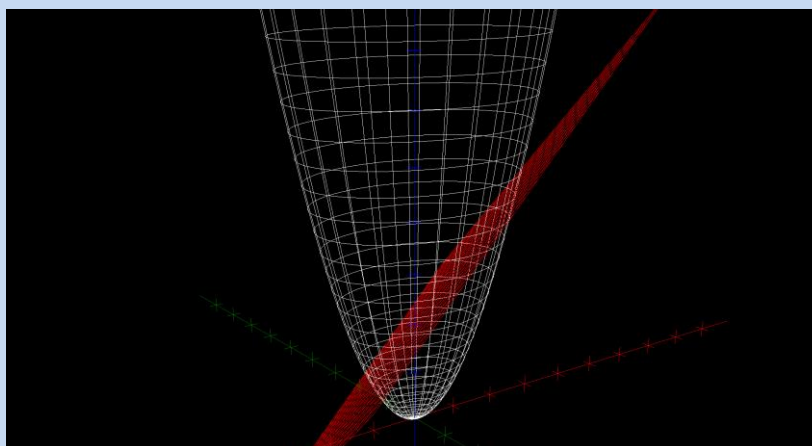
Lösning: Med polära koordinater (ρ, φ) svarar enhetsdisken D mot rektangeln $E := [0, 1] \times [0, 2\pi[$ så att massan är

$$\iint_D e^{-x^2 - y^2} dx dy = \iint_E e^{-\rho^2} \rho d\rho d\varphi = \int_0^1 \rho e^{-\rho^2} d\rho \int_0^{2\pi} d\varphi = 2\pi \left[-\frac{1}{2} e^{-\rho^2} \right]_0^1 = \pi(1 - e^{-1}).$$

Svar: Massan är $\pi(1 - e^{-1})$.

Exempel

Beräkna volymen av den del av inandömet av paraboloiden $z = x^2 + 2y^2$ som ligger under planet $x - y + z = 2$.



Lösning: Eftersom planet kan skrivas $z = 2 - x + y$ så skär det paraboloiden. Vi skall alltså beräkna volymen $V(D)$ av det icke-triviala området $D := \{(x, y, z) \in \mathbb{R}^3: x^2 + 2y^2 \leq z \leq 2 - x + y\}$.

Vi använder metoden med "staplar". Vi behöver då projektionen av D på xy -planet, och för att bestämma denna bestämmer vi dess rand, vilken uppenbarligen är projektionen av skärningskurvan mellan paraboloiden och planet. Skärningskurvan ges av

$$\begin{aligned} \begin{cases} z = x^2 + 2y^2 \\ z = 2 - x + y \end{cases} &\implies x^2 + 2y^2 = 2 - x + y \implies \\ &\implies \left(x + \frac{1}{2}\right)^2 + 2\left(y - \frac{1}{4}\right)^2 = \frac{19}{8} \end{aligned}$$

så att skärningskurvan tillhör den elliptiska cylindern $\left(x + \frac{1}{2}\right)^2 + 2\left(y - \frac{1}{4}\right)^2 = \frac{19}{8}$; projektionen på xy -planet är därmed ellipsen $\left(x + \frac{1}{2}\right)^2 + 2\left(y - \frac{1}{4}\right)^2 = \frac{19}{8}, z = 0$. D 's projektion på xy -planet är alltså $\left(x + \frac{1}{2}\right)^2 + 2\left(y - \frac{1}{4}\right)^2 \leq \frac{19}{8}, z = 0$. Eftersom

$$\left(x + \frac{1}{2}\right)^2 + 2\left(y - \frac{1}{4}\right)^2 \leq \frac{19}{8} \iff \left(\frac{x + 1/2}{\sqrt{19/2\sqrt{2}}}\right)^2 + \left(\frac{y - 1/4}{\sqrt{19/4}}\right)^2 \leq 1$$

rör det sig om en fylld koordinataxelparallell ellips med medelpunkt i $\left(-\frac{1}{2}, \frac{1}{4}\right)$ och halvaxellängderna $\frac{\sqrt{19}}{2\sqrt{2}}$ respektive $\frac{\sqrt{19}}{4}$. Med

$$\tilde{D} := \left\{ (x, y) \in \mathbb{R}^2: \left(x + \frac{1}{2}\right)^2 + 2\left(y - \frac{1}{4}\right)^2 \leq \frac{19}{8} \right\}$$

är den sökta volymen $\iint_{\tilde{D}} h(x, y) dx dy$ där $h(x, y)$ är stapelhöjden ovanför (x, y) . Eftersom $h(x, y) = 2 - x + y - x^2 - 2y^2$ får vi volymen

$$\iint_{\tilde{D}} h(x, y) dx dy = \iint_{\tilde{D}} (2 - x + y - x^2 - 2y^2) dx dy.$$

I de elliptiska koordinaterna (variabelbytet har funktionaldeterminanten $\frac{19}{8\sqrt{2}}\rho$)

$$\begin{aligned} x &= -\frac{1}{2} + \frac{\sqrt{19}}{2\sqrt{2}}\rho \cos \varphi \\ y &= \frac{1}{4} + \frac{\sqrt{19}}{4}\rho \sin \varphi \end{aligned}$$

svarar \tilde{D} mot rektangeln $E := [0, 1] \times [0, 2\pi[$ varför

$$\begin{aligned} V(D) &= \iint_{\tilde{D}} h(x, y) dx dy = \iint_{\tilde{D}} (2 - x + y - x^2 - 2y^2) dx dy = \frac{19}{8\sqrt{2}} \iint_E \left(\frac{19}{8} - \frac{19}{8}\rho^2\right) \rho d\rho d\varphi = \\ &= \frac{361}{64\sqrt{2}} \int_0^1 (\rho - \rho^3) d\rho \int_0^{2\pi} d\varphi = \frac{361}{64\sqrt{2}} \cdot \frac{1}{4} \cdot 2\pi = \frac{361\pi}{128\sqrt{2}} \end{aligned}$$

Svar: Volymen är $\frac{361\pi}{128\sqrt{2}} \approx 6.27$.

11.6 Intressanta exempel

Nu när vi värmt upp med en rad enkla exempel tillåter vi oss att ge några mer intressanta exempel.

11.6.1 Volymberäkningar

Exempel

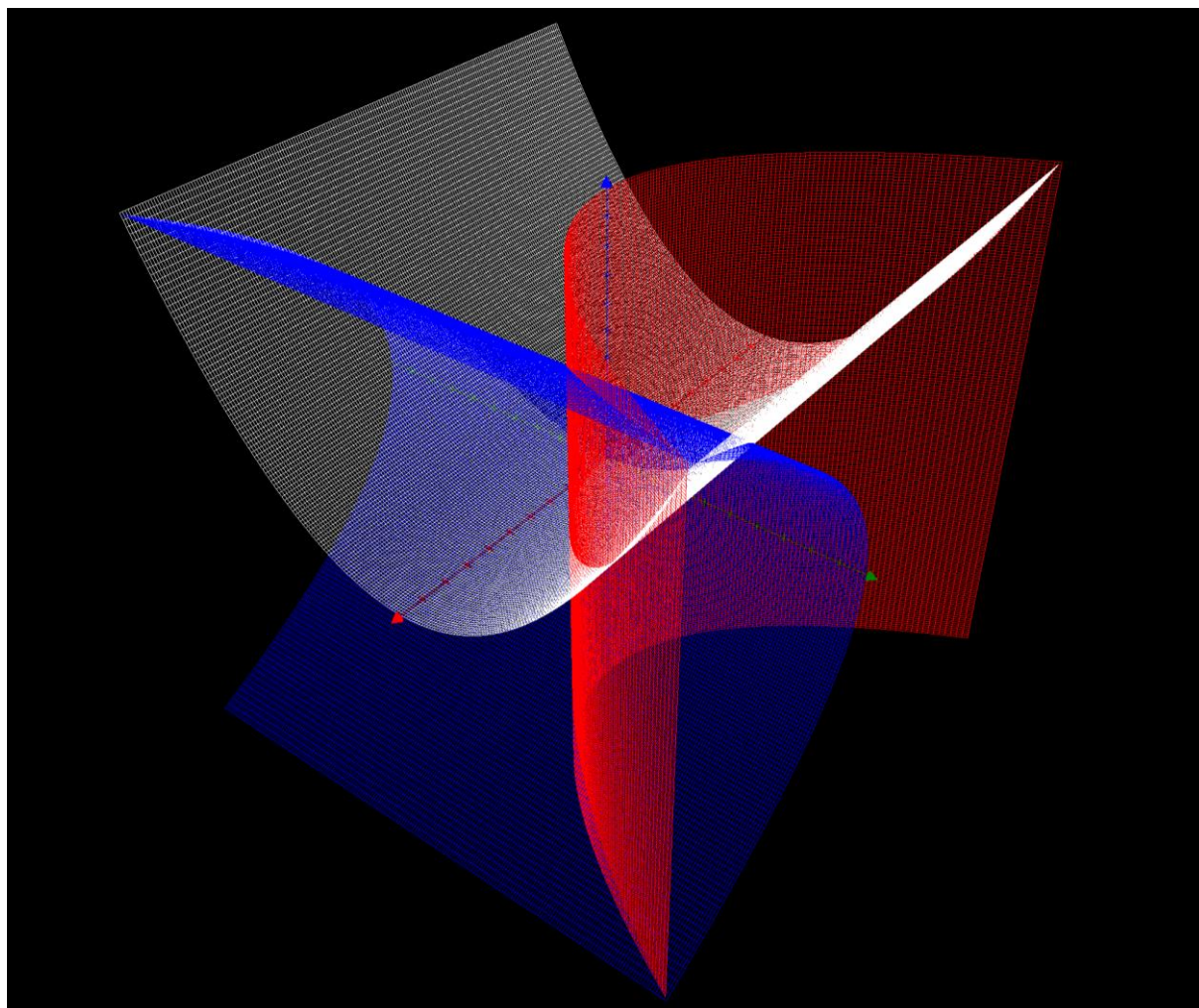
Bestäm volymen av "frökapseln" $D := \{(x, y, z) \in \mathbb{R}^3 : z \geq y^2 \wedge y \geq x^2 \wedge x \geq z^2\}$ som begränsas av de tre paraboliska cylindrarna $z = y^2$, $y = x^2$ och $x = z^2$.

Lösning: Vi noterar först att D ligger helt i den slutna positiva oktanten ty $x \geq z^2 \geq 0$ och motsvarande för y och z . Sedan noterar vi att $x \geq z^2 \geq y^4 \geq x^8$ vilket medför att $x \leq 1$ och motsvarande för y och z . Sålunda är $D \subset [0, 1]^3$ (så den sökta volymen $V(D) \in [0, 1]$). Vi ser också att x antar alla värden mellan 0 och 1, ty t.ex. $(0, 0, 0) \in D$ och $(1, 1, 1) \in D$. För varje $x \in [0, 1]$ gäller att y går från x^2 upp till $\sqrt{z} = \sqrt[4]{x}$. Slutligen, för varje fixt (x, y) gäller att z går från y^2 till \sqrt{x} . Därför är

$$V(D) = \int_0^1 \left(\int_{x^2}^{\sqrt[4]{x}} \left(\int_{y^2}^{\sqrt{x}} dz \right) dy \right) dx = \frac{1}{7}.$$

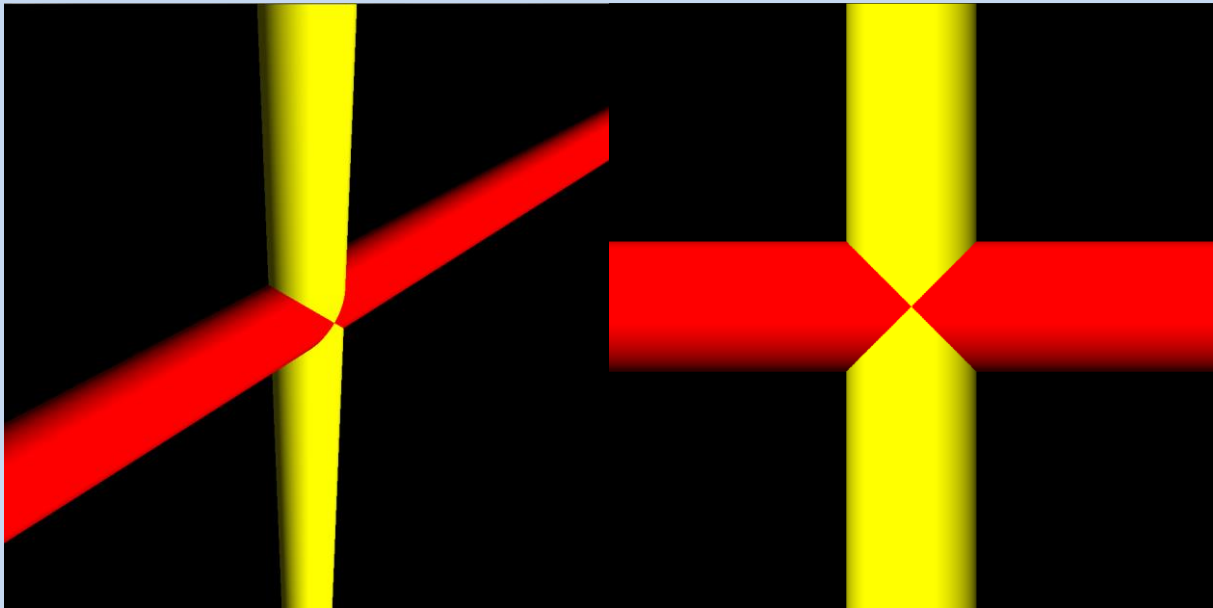
Svar: Volymen är $\frac{1}{7}$.

Nedan visas en bild av de tre paraboliska cylindrarna som omsluter "frökapseln".

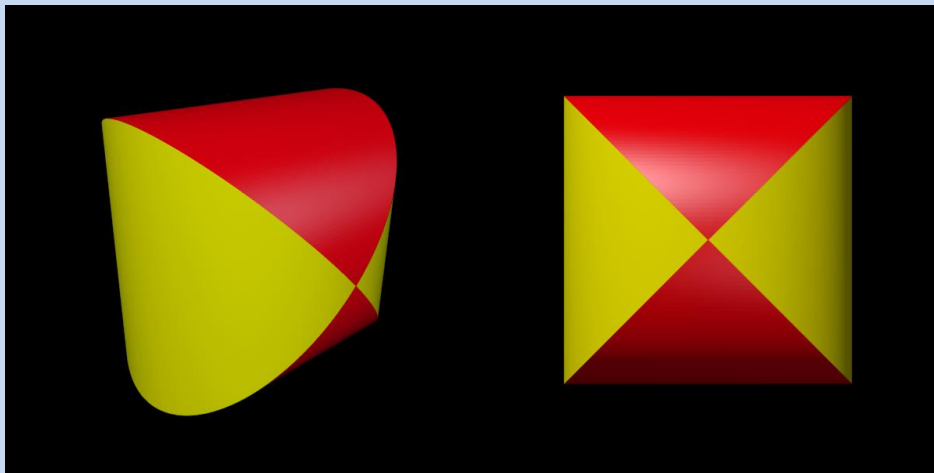


Exempel

Beräkna volymen av snittet mellan cylindrarna $y^2 + z^2 \leq 1$ (röd) och $x^2 + z^2 \leq 1$ (gul).



Observera att snittet inte syns i bilderna ovan, eftersom det ligger gömt inuti unionen av cylindrarna. Här kommer ett par bilder av snittet:

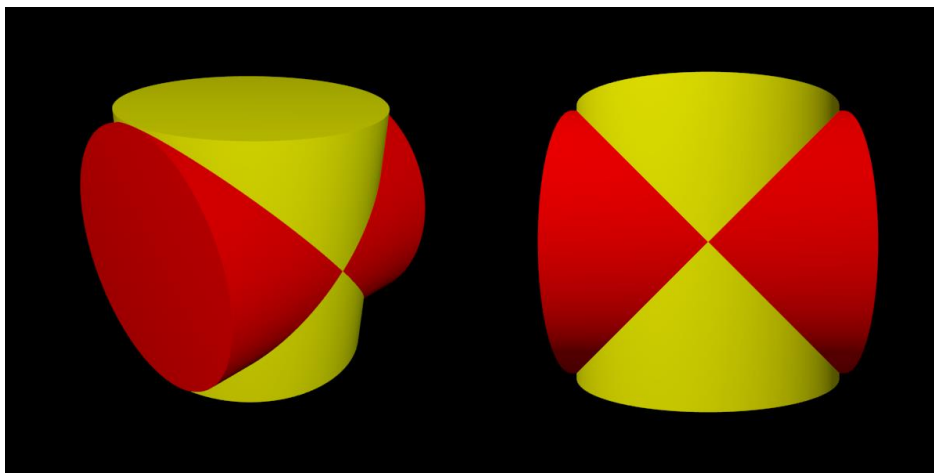


Lösning: Av symmetriskäl räcker det med att beräkna volymen av snittet i en av de åttondelar som syns i bilden till höger, t.ex. den åttondel där $x \geq 0$ och $0 \leq y \leq x$. Här går x från 0 till 1 och för varje fixt sådant x går y från 0 till x . För varje (x, y) går slutligen z från $-\sqrt{1-x^2}$ till $\sqrt{1-x^2}$ (d.v.s. mellan de gula begränsningsytorna). Därför är den sökta volymen

$$\begin{aligned} V &:= 8 \int_0^1 \left(\int_0^x \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dz \right) dy \right) dx = 16 \int_0^1 \left(\int_0^x \sqrt{1-x^2} dy \right) dx = 16 \int_0^1 (x\sqrt{1-x^2}) dx = \\ &= 16 \left[-\frac{1}{3}(1-x^2)^{3/2} \right]_0^1 = \frac{16}{3}. \end{aligned}$$

Svar: Den sökta volymen är $\frac{16}{3}$.

Att i stället beräkna volymen av *unionen* mellan cylindrarna vore förstås fånigt, eftersom den inte är ändlig. Däremot kan vi beräkna volymen av den del av unionen som ligger i $[-1,1]^3$:



Exempel (*)

Beräkna volymen av den del av unionen av de två cylindrarna $y^2 + z^2 \leq 1$ (röd) och $x^2 + z^2 \leq 1$ (gul) vilken ligger i $[-1,1]^3$.

Lösning (metod 1): Skillnaden jämfört med föregående beräkning är att vi nu, för varje fixt (x, y) , låter z gå mellan de röda begränsningsytorna i stället för mellan de gula. Vi får alltså volymen

$$\begin{aligned}
 V &:= 8 \int_0^1 \left(\int_0^x \left(\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dz \right) dy \right) dx = 16 \int_0^1 \left(\int_0^x \sqrt{1-y^2} dy \right) dx = \left[\begin{array}{l} y = \sin t \\ t \in [0, \pi/2] \\ dy = \cos t dt \end{array} \right] = \\
 &= 16 \int_0^1 \left(\int_0^{\arcsin x} \sqrt{1-\sin^2 t} \cos t dt \right) dx = \\
 &= 16 \int_0^1 \left(\int_0^{\arcsin x} |\cos t| \cos t dt \right) dx = 16 \int_0^1 \left(\int_0^{\arcsin x} \cos^2 t dt \right) dx = \\
 &= 16 \int_0^1 \left(\int_0^{\arcsin x} \left(\frac{1}{2} \cos 2t + \frac{1}{2} \right) dt \right) dx = 16 \int_0^1 \left(\left[\frac{1}{4} \sin 2t + \frac{1}{2} t \right]_{t=0}^{\arcsin x} \right) dx = \\
 &= 16 \int_0^1 \left(\left[\frac{1}{2} \sin t \cos t + \frac{1}{2} t \right]_{t=0}^{\arcsin x} \right) dx = \\
 &= 8 \int_0^1 (\sin \arcsin x \cos \arcsin x + \arcsin x) dx = \\
 &= 8 \int_0^1 (x \sqrt{1-\sin^2 \arcsin x} + \arcsin x) dx = 8 \int_0^1 (x \sqrt{1-x^2} + \arcsin x) dx = \\
 &= 8 \left[-\frac{1}{3} (1-x^2)^{3/2} + \sqrt{1-x^2} + x \arcsin x \right]_0^1 = 8 \left(\frac{\pi}{2} + \frac{1}{3} - 1 \right) = 4\pi - \frac{16}{3}
 \end{aligned}$$

eftersom

$$\int \arcsin x dx = \int \underset{\uparrow}{1} \cdot \underset{\downarrow}{\arcsin x} dx = x \arcsin x - \int \frac{x}{\sqrt{1-x^2}} dx = x \arcsin x + \sqrt{1-x^2} + C.$$

Svar: Den sökta volymen är $4\pi - \frac{16}{3}$.

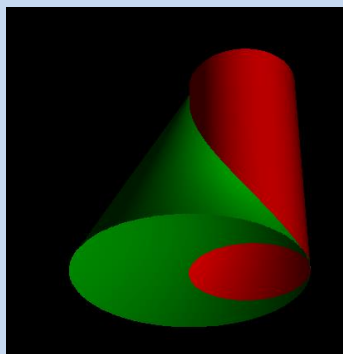
Lösning (metod 2): Låt $C_1 := \{(x, y, z) \in \mathbb{R}^3 : y^2 + z^2 \leq 1, -1 \leq x \leq 1\}$ och $C_2 := \{(x, y, z) \in \mathbb{R}^3 : x^2 + z^2 \leq 1, -1 \leq y \leq 1\}$. Då är $V(C_1 \cup C_2) = V(C_1) + V(C_2) - V(C_1 \cap C_2) = 2\pi + 2\pi - \frac{16}{3}$ där $V(C_1 \cap C_2)$ kommer från föregående exempel.

Svar: Den sökta volymen är $4\pi - \frac{16}{3}$.

Exempel

Ur en kon av radie R och höjd h borrar ett cylindriskt hål av radie $R/2$ på ett sådant sätt att borrhålets axel är parallell med konens axel och borrhålets rand innehåller konens symmetriaxel. Vad är volymen av det som sedan återstår av konen?

Lösning: Vi bestämmer volymen av snittet mellan konen och borrhålet. Låt konen ligga längs z -axeln med spetsen i origo och bottenytan i planet $z = -h$, så att dess ekvation är $x^2 + y^2 \leq \frac{R^2}{h^2}z^2$, $-h \leq z \leq 0$, och låt borrhålet vara $(x - \frac{R}{2})^2 + y^2 \leq \frac{R^2}{4}$, se bild.



Vi använder metoden med staplar. Låt $\tilde{D} := \{(x, y) \in \mathbb{R}^2 : (x - \frac{R}{2})^2 + y^2 \leq \frac{R^2}{4}\}$ vara projektionen av snittet på xy -planet, fast betraktat som en plan delmängd. Stapeln ovanför $(x, y) \in \tilde{D}$ går från $z = -h$ till $z = -\frac{h}{R}\sqrt{x^2 + y^2}$, så volymen av snittet är

$$\begin{aligned} V_{\text{snitt}} &:= \iint_{\tilde{D}} \left(-\frac{h}{R}\sqrt{x^2 + y^2} - (-h) \right) dx dy = h \iint_{\tilde{D}} \left(1 - \frac{1}{R}\sqrt{x^2 + y^2} \right) dx dy = \\ &= h \iint_{\tilde{D}} dx dy - \frac{h}{R} \iint_{\tilde{D}} \sqrt{x^2 + y^2} dx dy = hA(\tilde{D}) - \frac{h}{R} \iint_{\tilde{D}} \sqrt{x^2 + y^2} dx dy = \\ &= \frac{R^2 h \pi}{4} - \frac{h}{R} \cdot \frac{4R^3}{9} = \frac{R^2 h \pi}{4} - \frac{4R^2 h}{9} \end{aligned}$$

där vi skipade uträkningen av $\iint_{\tilde{D}} \sqrt{x^2 + y^2} dx dy$ eftersom vi gjorde i princip exakt denna i ett tidigare exempel. Den sökta volymen är sålunda

$$V := \frac{R^2 h \pi}{3} - V_{\text{snitt}} = \frac{R^2 h \pi}{3} - \frac{R^2 h \pi}{4} + \frac{4R^2 h}{9} = R^2 h \left(\frac{\pi}{12} + \frac{4}{9} \right).$$

Svar: Den sökta volymen är $R^2 h \left(\frac{\pi}{12} + \frac{4}{9} \right)$.

11.6.2 Areaberäkningar

Låt $\Sigma := \mathbf{r}(D)$ vara en yta parametriserad av $\mathbf{r}: D \rightarrow \mathbb{R}^3$ där $D \subset \mathbb{R}^2$. Vad är arean av Σ ? Ett naturligt tillvägagångssätt är att man delar in D i en massa små kvadrater. Betrakta den kvadrat som sträcker sig från (u, v) till $(u + \Delta u, v + \Delta v)$. Bilden av denna kvadrat är i allmänhet en buktig yta, men den approximeras väl – om Δu och Δv är små och \mathbf{r} är snäll nog – av tangentplansbiten $\mathbf{q}([\Delta u, \Delta v]^2)$ där

$$\mathbf{q}(s, t) = \mathbf{r}(u, v) + s\mathbf{r}'_u(u, v) + t\mathbf{r}'_v(u, v).$$

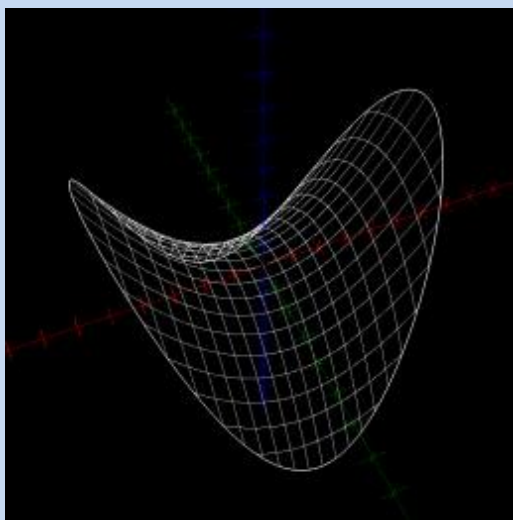
Denna parallelogram har arean $|(\Delta u \mathbf{r}'_u(u, v)) \times (\Delta v \mathbf{r}'_v(u, v))| = |\mathbf{r}'_u(u, v) \times \mathbf{r}'_v(u, v)|\Delta u\Delta v$. Om indelningen av D i kvadrater görs finare bör en bättre approximation av Σ 's area erhållas, varför det är rimligt att tolka

$$\iint_D |\mathbf{r}'_u(u, v) \times \mathbf{r}'_v(u, v)| dudv$$

som arean av Σ . I själva verket kan vi med fördel *definiera* arean $A(\Sigma)$ av en parametriserad yta Σ på det här sättet.

Exempel

Beräkna arean av sadeln $z = x^2 - y^2$ där $x^2 + y^2 \leq 1$.



Lösning: Ytan är bilden $\Sigma := \mathbf{r}(D)$ där $\mathbf{r}(u, v) := (u, v, u^2 - v^2)$ och D är den fyllda enhetsdisken. Dess area är sålunda

$$\begin{aligned} A(\Sigma) &= \iint_D |\mathbf{r}'_u(u, v) \times \mathbf{r}'_v(u, v)| dudv = \iint_D \left| \mathbf{e} \begin{pmatrix} 1 \\ 0 \\ 2u \end{pmatrix} \times \mathbf{e} \begin{pmatrix} 0 \\ 1 \\ -2v \end{pmatrix} \right| dudv = \iint_D \left| \mathbf{e} \begin{pmatrix} -2u \\ 2v \\ 1 \end{pmatrix} \right| dudv = \\ &= \iint_D \sqrt{4u^2 + 4v^2 + 1} dudv = \int_{E := [0,1] \times [0,2\pi[} \begin{matrix} u = \rho \cos \varphi \\ v = \rho \sin \varphi \\ J = \rho \end{matrix} \rho \sqrt{4\rho^2 + 1} d\rho d\varphi = \\ &= \int_0^1 \rho \sqrt{4\rho^2 + 1} d\rho \int_0^{2\pi} d\varphi = 2\pi \left[\frac{1}{12} (4\rho^2 + 1)^{3/2} \right]_0^1 = \frac{\pi}{6} (5^{3/2} - 1). \end{aligned}$$

Svar: Arean av sadeln är $\frac{\pi}{6} (5^{3/2} - 1)$.

Övningar

1. Beräkna arean av paraboloiden $z = x^2 + y^2$ innanför cylindern $x^2 + y^2 = 1$.
2. Beräkna arean i föregående uppgift på ett alternativt sätt (ledning: paraboloiden är en rotationsyta).
3. I avsnittet om enkelvariabelintegraler beräknade vi arean av en sfär genom att utnyttja det faktum att sfären är en rotationsyta. Bestäm arean av en sfär med den allmänna metoden med parametrisering.

11.6.3 Medelvärdesberäkningar

Säg att $F: D \rightarrow \mathbb{R}$ är ett skalärfält på D . Vad är medelvärdet av F på D ? Det kan t.ex. hända att D är ett område i planet, och $F(x, y)$ är temperaturen i punkten $(x, y) \in D$. Frågan är då vad *medeltemperaturen* är i området. Precis som i envariabelfallet måste vi egentligen *definiera* vad vi menar med "medelvärdet" av ett skalärfält på en mängd, så det uttryck vi resonerar oss fram till är snarare en definition än en sats.

Ett naturligt tillvägagångssätt är att dela in D i N stycken lika stora småbitar $(D_k)_{k=1}^N$. Om småbitarnas storlek är $\Delta\mu$ så är $N\Delta\mu = \mu(D)$ där $\mu(D)$ är måttet (längd, area, volym, hypervolym) av D . I varje småbit är – om F är snäll nog och biten liten nog – F ungefär konstant, så med god approximation kan vi välja något funktionsvärde $F_k := F(\mathbf{x})$, $\mathbf{x} \in D_k$, i biten och säga att $F(\mathbf{x}) \approx F_k$ för alla $\mathbf{x} \in D_k$.

Sedan kan vi bilda talet $\hat{F} := \frac{1}{N} \sum_{k=1}^N F_k$ vilket kan betraktas som ett approximativt medelvärde av F på D . Men

$$\hat{F} := \frac{1}{N} \sum_{k=1}^N F_k = \frac{1}{\mu(D)} \sum_{k=1}^N F_k \frac{\mu(D)}{N} = \frac{1}{\mu(D)} \sum_{k=1}^N F_k \Delta\mu \rightarrow \frac{1}{\mu(D)} \int_D F(\mathbf{x}) d\mathbf{x}$$

när indelningen av D görs finare. $\int_D F(\mathbf{x}) d\mathbf{x}$ står här för multipelintegralen av F över D ; t.ex. i planet menar vi $\iint_D F(x, y) dx dy$ och i rummet $\iiint_D F(x, y, z) dx dy dz$. Det är alltså rimligt att definiera

$$\hat{F} := \frac{1}{\mu(D)} \int_D F(\mathbf{x}) d\mathbf{x}$$

som medelvärde av F på D . Notera att detta är en uppenbar generalisering av medelvärdesbegreppet för envariabelfunktioner som vi introducerade i det avsnittet.

Exempel

Låt $F(x, y) := x^2$ och $D := \{(x, y) \in \mathbb{R}^2: x^2 + 4y^2 \leq 4\}$. Bestäm medelvärdet \hat{F} av F på D .

Lösning: I ett tidigare exempel fann vi att $\iint_D x^2 dx dy = 2\pi$ och att D är en ellips med halvaxellängderna 2 respektive 1 längs x -axeln respektive y -axeln. D har alltså arean 2π så att

$$\hat{F} = \frac{1}{A(D)} \iint_D F(x, y) dx dy = \frac{1}{2\pi} \cdot 2\pi = 1.$$

Svar: $\hat{F} = 1$.

11.6.3.1 Geometriskt centrum

Givet N stycken partiklar med positioner $(\mathbf{r}_k)_{k=1}^N = (x_k, y_k, z_k)_{k=1}^N$ definieras deras *geometriska centrum* som punkten $(\hat{x}, \hat{y}, \hat{z})$ där $\hat{x} := \frac{1}{N} \sum_{k=1}^N x_k$ är (det aritmetiska) medelvärdet av partiklarnas x -koordinater och \hat{y} och \hat{z} definieras analogt.

Men om vi har *oändligt* många punkter, t.ex. en (oändlig men begränsad) plan mängd $D \subset \mathbb{R}^2$? Kan vi prata om "geometriskt centrum" även här? Ja, det är, vilket läsaren själv kan kontrollera rimligheten i, naturligt att definiera det geometriska centret som punkten $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_n)$ där \hat{x}_k är medelvärdet av skalärfältet $(x_1, \dots, x_n) \mapsto x_k$ på D .

Innan vi gör några intressanta beräkningar av geometriskt centrum värmer vi upp med några trivialberäkningar för att se att medelvärdesformeln "fungerar".

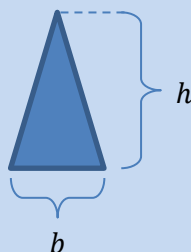
Exempel. Vad är det geometriska centrumet hos enhetsdisken $D \subset \mathbb{R}^2$? Naturligtvis är det origo, men låt oss ändå pröva vår formel. \hat{x} är medelvärdet av skalärfältet x på D , så $\hat{x} = \frac{1}{A(D)} \iint_D x dx dy = \frac{1}{\pi} \cdot 0 = 0$ av symmetriskäl. Men låt oss *ändå* göra beräkningen i detalj, av pedagogiska skäl.

$$\begin{aligned} \hat{x} &= \frac{1}{A(D)} \iint_D x dx dy = \left[\begin{array}{l} x = \rho \cos \varphi \\ y = \rho \sin \varphi \\ J = \rho \\ E := [0,1] \times [0,2\pi] \end{array} \right] = \frac{1}{\pi} \iint_E \rho^2 \cos \varphi d\rho d\varphi = \frac{1}{\pi} \int_0^1 \rho^2 d\rho \int_0^{2\pi} \cos \varphi = \\ &= \frac{1}{\pi} \cdot \frac{1}{3} \cdot 0 = 0. \end{aligned}$$

På precis analogt sätt ser vi att $\hat{y} = 0$, så det geometriska centrumet är origo, som väntat. Nog med tråkigheter, låt oss göra något roligt i stället!

Exempel

Beräkna det geometriska centrumet hos en likbent triangel med bas b och höjd h .



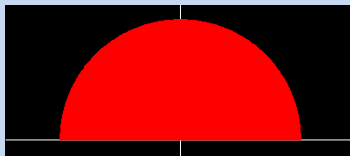
Lösning: Låt triangeln D vara området till vänster om $y = \frac{2h}{b}x$, till höger om $y = -\frac{2h}{b}x$ och under $y = h$. Det är trivialt att $\hat{x} = 0$ och vi finner

$$\begin{aligned} \hat{y} &= \frac{1}{A(D)} \iint_D y dx dy = \frac{2}{bh} \int_0^h \left(\int_{-\frac{b}{2h}y}^{\frac{b}{2h}y} y dx \right) dy = \frac{2}{bh} \int_0^h y \left(\int_{-\frac{b}{2h}y}^{\frac{b}{2h}y} dx \right) dy = \frac{2}{h^2} \int_0^h y^2 dy = \\ &= \frac{2}{h^2} \cdot \frac{1}{3} h^3 = \frac{2}{3} h. \end{aligned}$$

Svar: Det geometriska centrumet ligger på symmetriaxeln, $\frac{2}{3}$ av höjden från spetsen.

Exempel

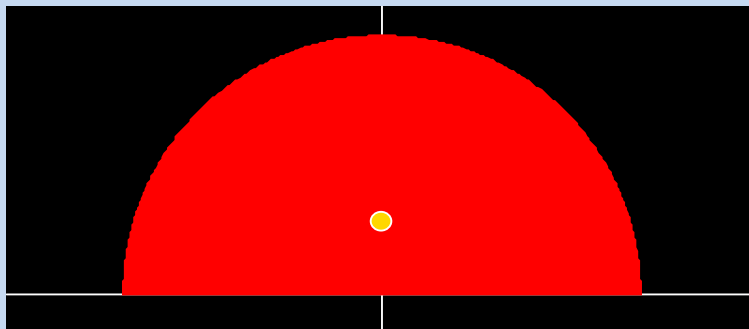
Bestäm det geometriska centrumet hos en halv disk av radie R .



Lösning: Låt disken vara $D := \{(x, y) \in \mathbb{R}^2: x^2 + y^2 \leq R^2, y \geq 0\}$; det är då trivialt att $\hat{x} = 0$. Vi har

$$\begin{aligned} \hat{y} &= \frac{1}{A(D)} \iint_D y dx dy = \frac{2}{R^2 \pi} \iint_D y dx dy = \left[\begin{array}{l} x = \rho \cos \varphi \\ y = \rho \sin \varphi \\ J = \rho \\ E := [0, R] \times [0, \pi] \end{array} \right] = \frac{2}{R^2 \pi} \iint_E \rho^2 \sin \varphi dx dy = \\ &= \frac{2}{R^2 \pi} \int_0^R \rho^2 d\rho \int_0^\pi \sin \varphi d\varphi = \frac{2}{R^2 \pi} \cdot \frac{1}{3} R^3 \cdot 2 = \frac{4}{3\pi} R. \end{aligned}$$

Svar: Det geometriska centrumet ligger på symmetriaxeln, $\frac{4}{3\pi}$ av radien från bottenlinjen, d.v.s. ungefär 42 % av radien från motsvarande hela disks medelpunkt.



Exempel

Bestäm det geometriska centrumet hos $\{(x, y) \in \mathbb{R}^2: -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}, 0 \leq y \leq \cos x\}$.

Lösning: Låt mängden heta D . Det är klart att $\hat{x} = 0$ och att

$$\hat{y} = \frac{1}{A(D)} \iint_D y dx dy = \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\int_0^{\cos x} y dy \right) dx = \frac{1}{4} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 x dx = \dots = \frac{\pi}{8}.$$

Svar: Det geometriska centrumet är $(0, \frac{\pi}{8})$.

Övningar

1. I planet:
 - a. Bestäm det geometriska centrumet hos en rätvinklig triangel.
 - b. Bestäm det geometriska centrumet hos $\{(x, y) \in \mathbb{R}^2: x^2 \leq y \leq 1\}$.
 - c. Bestäm det geometriska centrumet hos $\{(x, y) \in \mathbb{R}^2: -1 \leq x \leq 1, 0 \leq y \leq e^x\}$.

2. I rummet:

- Bestäm det geometriska centrumet hos en kon med radie r och höjd h .
- Bestäm det geometriska centrumet hos ett halvklot med radie R .
- Bestäm det geometriska centrumet hos en pyramid (kvadratisk botten) med botten-sida s och höjd h .

11.6.3.2 Masscentrum

Om en fysisk kropp K har konstant densitet är dess geometriska centrum också dess fysikaliska *masscentrum*. Om i stället densiteten $\rho(x, y, z)$ varierar definieras masscentrumet som punkten $(\hat{x}, \hat{y}, \hat{z})$ där

$$\hat{x} = \frac{1}{M(K)} \iiint_K x dM = \frac{\iiint_K x dM}{\iiint_K dM} = \frac{\iiint_K x \rho(x, y, z) dx dy dz}{\iiint_K \rho(x, y, z) dx dy dz}$$

där $M(K)$ är totala massan hos kroppen och analogt för \hat{y} och \hat{z} . (Rent matematiskt rör det sig om *viktade* medelvärden.)

11.6.3.3 Fler exempel med medelvärden

Exempel

Bestäm medelavståndet till origo på enhetsdisken.

Lösning: Vi skall alltså hitta medelvärdet av skalärfältet $F(x, y) = \sqrt{x^2 + y^2}$ på enhetsdisken D . Detta är

$$\begin{aligned} \frac{1}{A(D)} \iint_D F(x, y) dx dy &= \frac{1}{\pi} \iint_D \sqrt{x^2 + y^2} dx dy = \left[\begin{array}{l} x = \rho \cos \varphi \\ y = \rho \sin \varphi \\ J = \rho \\ E := [0, 1] \times [0, 2\pi[\end{array} \right] = \frac{1}{\pi} \iint_E \rho^2 d\rho d\varphi = \\ &= \frac{1}{\pi} \int_0^1 \rho^2 d\rho \int_0^{2\pi} d\varphi = \frac{1}{\pi} \cdot \frac{1}{3} \cdot 2\pi = \frac{2}{3}. \end{aligned}$$

Svar: Medelavståndet är $\frac{2}{3}$.

Exempel

Bestäm medelavståndet till origo på enhetsklotet.

Lösning: Om D betecknar enhetsklotet så är medelavståndet

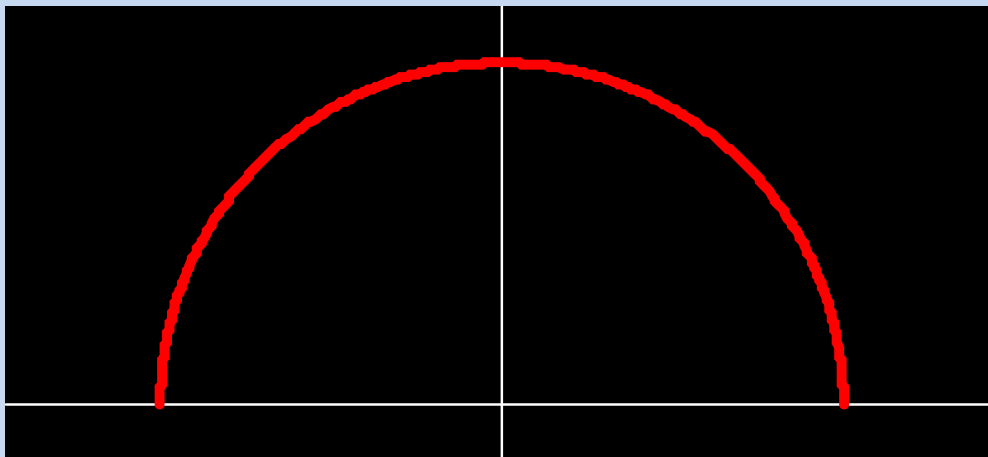
$$\begin{aligned} \frac{1}{V(D)} \iiint_D \sqrt{x^2 + y^2 + z^2} dx dy dz &= \left[\begin{array}{l} x = r \sin \theta \cos \varphi \\ y = r \sin \theta \sin \varphi \\ z = r \cos \theta \\ J = r^2 \sin \theta \\ E := [0, 1] \times [0, \pi] \times [0, 2\pi[\end{array} \right] = \frac{3}{4\pi} \iiint_E r^3 \sin \theta dr d\theta d\varphi = \\ &= \frac{3}{4\pi} \int_0^1 r^3 dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\varphi = \frac{3}{4\pi} \cdot \frac{1}{4} \cdot 2 \cdot 2\pi = \frac{3}{4}. \end{aligned}$$

Svar: Medelavståndet är $\frac{3}{4}$.

Hittills har vi beräknat medelvärden av skalärfält i \mathbb{R}^n på n -dimensionella områden. Men man kan också tala om medelvärden på andra sorters mängder, t.ex. medelvärdet av $F(x, y)$ på en *kurva* (d.v.s. en endimensionell "sak") i planet, eller medelvärdet av $F(x, y, z)$ på en *kurva* eller *yta* i rummet. Med ett resonemang liknande de vi gett tidigare ser man det naturliga i att definiera dessa tre medelvärdsbegrepp som s.k. kurv- eller ytintegraler dividerat med *kurvans längd* eller *ytans area*. Vi ger ett enkelt exempel.

Exempel

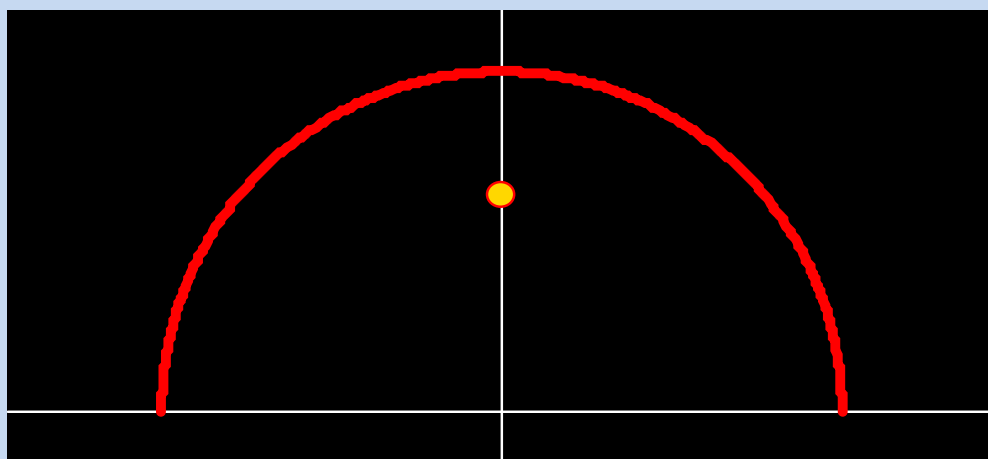
Bestäm masscentrumet hos en oböjlig homogen metalltråd formad som en halvcirkel med radie r .



Lösning: Låt tråden sträcka sig längs $\gamma := \{(x, y) \in \mathbb{R}^2: x^2 + y^2 = r^2, y \geq 0\}$. Det är då klart att $\hat{x} = 0$ och med parametriseringen $(x, y) = (r \cos t, r \sin t)$, $t \in [0, \pi]$, av γ får vi båglängden $ds = \sqrt{\dot{x}^2 + \dot{y}^2} dt = r dt$ och sålunda

$$\hat{y} = \frac{1}{L(\gamma)} \int_{\gamma} y ds = \frac{1}{r\pi} \int_0^{\pi} r \sin t \cdot r dt = \frac{r}{\pi} \int_0^{\pi} \sin t dt = \frac{2r}{\pi}.$$

Svar: Masscentrumet ligger på symmetriaxeln, på avståndet $\left(1 - \frac{2}{\pi}\right)r$ från trådens "mittpunkt" (d.v.s. ungefär 36 % av radien in mot motsvarande hela cirkels medelpunkt, eller ungefär 64 % av radien från motsvarande hela cirkels medelpunkt).



11.7 Sammanfattning av integraltolkningar

Låt oss avsluta kapitlet om integraler med att sammanfatta de typiska tolkningarna av dessa objekt.

	$\int_D f(x)dx$	$\iint_D f(x,y)dxdy$	$\iiint_D f(x,y,z)dxdydz$
I allmänhet	Gränsvärde av Riemannsumma	Gränsvärde av Riemannsumma	Gränsvärde av Riemannsumma
Om $f(\mathbf{x}) \geq 0$ på D	Arean under grafen $y = f(x)$	Volymen under grafen $z = f(x,y)$	Hypervolymen under grafen $w = f(x,y,z)$
Om $f(\mathbf{x}) \equiv 1$ på D	Längden av D	Arean av D	Volymen av D

12 Generaliserade integraler

Vi har lärt oss att integrera *begränsade* funktioner på *begränsade* mängder. Precis som i envariabel-fallet kan vi sedan, via gränsvärden, utöka teorin till att även tillåta integration av obegränsade funktioner på obegränsade mängder.²⁵

Det lättaste – och egentligen mest intressanta – fallet är när integranden är icke-negativ. Antag därför att $f(\mathbf{x}) \geq 0$ överallt på den öppna mängden $D \subset \mathbb{R}^n$. Vi tillåter nu att f är obegränsad, och att D är obegränsad. För att definiera $\int_D f(\mathbf{x})d\mathbf{x}$ (där \int står för \int , \iint , \iiint , ...) inför vi samlingen \mathcal{D} av alla *begränsade* kvadrerbara delmängder D' av D sådana att restriktionen av f till D' är *begränsad*.

Varje D' är alltså en *begränsad* (kvadrerbar) delmängd av D , där man också skurit bort områden där f är obegränsad.

Integralen $\int_{D'} f(\mathbf{x})d\mathbf{x}$ existerar då (i vanlig mening) för varje $D' \in \mathcal{D}$. Det är naturligt att definiera $\int_D f(\mathbf{x})d\mathbf{x} := \sup_{D' \in \mathcal{D}} \int_{D'} f(\mathbf{x})d\mathbf{x}$. Vi säger att den här *generaliserade* integralen är konvergent omm supremumet är ändligt.

I praktiken kan man beräkna värdet av en generaliserad integral med hjälp av en s.k. *uttömmande svit*. En sådan är en familj $(D_k)_{k=1}^\infty$ av element i \mathcal{D} som "närmar sig" D när $k \rightarrow \infty$, i följande mening:

1. $D_{k+1} \supset D_k$ för varje $k \in \mathbb{Z}^+$.
2. För varje $D' \in \mathcal{D}$ finns det ett $K \in \mathbb{Z}^+$ sådant att $D' \subset D_K$ (och därmed är $D' \subset D_k$ för varje $k \geq K$).

Punkt (1) säger att mängderna "växer"; punkt (2) säger att de till slut blir större än varje element i \mathcal{D} .

Det är inte svårt att inse att om $(D_k)_{k=1}^\infty$ är en uttömmande svit för D och integranden så är

$$\int_D f(\mathbf{x})d\mathbf{x} \left(\stackrel{\text{def}}{=} \sup_{D' \in \mathcal{D}} \int_{D'} f(\mathbf{x})d\mathbf{x} \right) = \lim_{k \rightarrow \infty} \int_{D_k} f(\mathbf{x})d\mathbf{x},$$

d.v.s. det räcker med att bestämma den vanliga integralen $I_k := \int_{D_k} f(\mathbf{x})d\mathbf{x}$ för varje k och sedan bara låta $k \rightarrow \infty$, d.v.s. – i princip – låta " $D_k \rightarrow D$ ". Förfarandet är egentligen mycket naturligt och tämligen enkelt. Se följande exempel.

Exempel

Beräkna $\iint_{\mathbb{R}^2} e^{-x^2-y^2} dx dy$.

Lösning: Notera att integranden är icke-negativ. Låt $D_k := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq k^2\}$ vara disken av radie k kring origo. Då är alltså D_k en uttömmande svit för \mathbb{R}^2 och integranden; speciellt existerar den *vanliga* integralen av integranden på varje D_k . Vi har

²⁵ Om läsaren tycker att texten nedan är svår kan han förenkla den något genom att tänka bort ordet "kvadrerbar", som egentligen är väsentligt men i praktiken nästan överflödigt eftersom de mängder man studerar i inledande analyskurser så gott som alltid är kvadrerbara. Om läsaren (som är en förstaårsstudent som saknar intresse för fraktaler) föreställer sig en mängd i planet eller rummet (t.ex.) så är den nästan säkert kvadrerbar.

$$\begin{aligned} \iint_{D_k} e^{-x^2-y^2} dx dy &= \left[\begin{array}{l} x = \rho \cos \varphi \\ y = \rho \sin \varphi \\ J = \rho \\ E_k := [0, k] \times [0, 2\pi[\end{array} \right] = \iint_{E_k} \rho e^{-\rho^2} d\rho d\varphi = \int_0^k \rho e^{-\rho^2} d\rho \int_0^{2\pi} d\varphi = \\ &= 2\pi \left[-\frac{1}{2} e^{-\rho^2} \right]_0^k = \pi(1 - e^{-k^2}) \rightarrow \pi \end{aligned}$$

då $k \rightarrow \infty$.

Svar: $\iint_{\mathbb{R}^2} e^{-x^2-y^2} dx dy = \pi$.

Det som gjorde att integralen i föregående exempel var generaliserad var att området var obegränsat. I nästa exempel är i stället "problemet" att integranden är en obegränsad funktion.

Exempel

Beräkna $\iint_D \frac{1}{\sqrt{x^2+y^2}} dx dy$ där $D := \{(x, y) \in \mathbb{R}^2 : 0 < x^2 + y^2 < 1\}$ är den punkterade enhetsdisken.

Lösning: Notera att integranden är icke-negativ. Låt $D_k := \{(x, y) \in \mathbb{R}^2 : \frac{1}{k^2} < x^2 + y^2 < 1\}$ vara den del av D som undviker en omgivning av radie $1/k$ kring origo. Vi har då

$$\iint_{D_k} \frac{1}{\sqrt{x^2+y^2}} dx dy = \left[\begin{array}{l} x = \rho \cos \varphi \\ y = \rho \sin \varphi \\ J = \rho \\ E_k := \left] \frac{1}{k}, 1 \right[\times [0, 2\pi[\end{array} \right] = \iint_{E_k} dx dy = A(E_k) = \pi - \frac{1}{k^2} \pi \rightarrow \pi$$

då $k \rightarrow \infty$.

Svar: $\iint_D \frac{1}{\sqrt{x^2+y^2}} dx dy = \pi$ där $D = \{(x, y) \in \mathbb{R}^2 : 0 < x^2 + y^2 < 1\}$.

Exempel

Beräkna $\iint_{\mathbb{R}^2 \setminus \{0\}} \frac{1}{\sqrt{x^2+y^2}} dx dy$.

Lösning: Notera att integranden är icke-negativ. Inför $D_k := \{(x, y) \in \mathbb{R}^2 : \frac{1}{k^2} < x^2 + y^2 < k^2\}$. Eftersom

$$\iint_{D_k} \frac{1}{\sqrt{x^2+y^2}} dx dy = \left[\begin{array}{l} x = \rho \cos \varphi \\ y = \rho \sin \varphi \\ J = \rho \\ E_k := \left] \frac{1}{k}, k \right[\times [0, 2\pi[\end{array} \right] = \iint_{E_k} dx dy = A(E_k) = k^2\pi - \frac{1}{k^2}\pi \rightarrow \infty$$

då $k \rightarrow \infty$ är den generaliserade integralen divergent.

Svar: $\iint_{\mathbb{R}^2 \setminus \{0\}} \frac{1}{\sqrt{x^2+y^2}} dx dy$ är divergent.

Exempel (*)

Beräkna $\iiint_{\mathbb{R}^3} e^{-x^2-y^2-z^2} dx dy dz$.

Lösning: Först behöver vi bevisa ett klassiskt envariabelresultat. Vi såg i ett tidigare exempel att $\iint_{\mathbb{R}^2} e^{-x^2-y^2} dx dy = \pi$. När vi gjorde den beräkningen använde vi koncentriska diskar i vår uttömmade svit. En alternativ uttömmande svit är $D_k := [-k, k]^2$ så att

$$\begin{aligned} \pi &= \iint_{\mathbb{R}^2} e^{-x^2-y^2} dx dy = \lim_{k \rightarrow \infty} \iint_{D_k} e^{-x^2-y^2} dx dy = \lim_{k \rightarrow \infty} \iint_{D_k} e^{-x^2} e^{-y^2} dx dy = \\ &= \lim_{k \rightarrow \infty} \int_{-k}^k e^{-x^2} dx \int_{-k}^k e^{-y^2} dy = \lim_{k \rightarrow \infty} \left(\int_{-k}^k e^{-x^2} dx \right)^2 = \left(\lim_{k \rightarrow \infty} \int_{-k}^k e^{-x^2} dx \right)^2 = \\ &= \left(\int_{-\infty}^{\infty} e^{-x^2} dx \right)^2 \end{aligned}$$

varför vi erhållit det klassiska resultatet²⁶

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Låt oss nu gå tillbaka till vår trippelintegral. Inför $D_k := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq k^2\}$. Då är

$$\begin{aligned} \iiint_{D_k} e^{-x^2-y^2-z^2} dx dy dz &= \left[\begin{array}{l} x = r \sin \theta \cos \varphi \\ y = r \sin \theta \sin \varphi \\ z = r \cos \theta \\ J = r^2 \sin \theta \\ E_k := [0, k] \times [0, \pi] \times [0, 2\pi[\end{array} \right] = \iiint_{E_k} e^{-r^2} r^2 \sin \theta dx dy dz = \\ &= \int_0^k r^2 e^{-r^2} dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\varphi = 4\pi \int_0^k r^2 e^{-r^2} dr = 4\pi \int_0^k \underbrace{r}_{\downarrow} \cdot \underbrace{r e^{-r^2}}_{\uparrow} dr = \\ &= 4\pi \left(\left[-\frac{1}{2} r e^{-r^2} \right]_0^k - \int_0^k \left(-\frac{1}{2} e^{-r^2} \right) dr \right) = 2\pi \left(\int_0^k e^{-r^2} dr - k e^{-k^2} \right) = \\ &= 2\pi \left(\frac{1}{2} \int_{-k}^k e^{-r^2} dr - k e^{-k^2} \right) = \pi \int_{-k}^k e^{-r^2} dr - 2\pi k e^{-k^2} \rightarrow \pi^{3/2} \end{aligned}$$

då $k \rightarrow \infty$.

Svar: $\iiint_{\mathbb{R}^3} e^{-x^2-y^2-z^2} dx dy dz = \pi^{3/2}$.

²⁶ Funktionen $x \mapsto e^{-x^2}$ är ju ett klassiskt exempel på en funktion vars primitiv inte kan uttryckas i elementära funktioner, så utan "extra kunskap" är det svårt att bestämma värdet av $\int_{-\infty}^{\infty} e^{-x^2} dx$.