Esa V. Vesalainen[1]

Modulimuodot: lukuteorian aarteita

Modulimuodot ovat kaunis ja klassinen lukuteorian aihe, jolla on yhteyksiä hämmästyttävän moniin muihin aiheisiin. Seuraavassa tarkoituksena olisi hieman valottaa sitä, millaisia otuksia ne ovat, ja mitä iloa niistä on. Luonnollisesti tässä on mahdotonta tehdä täysin oikeutta niiden valtaisan suurelle teorialle ja lukuisille sovellutuksille.

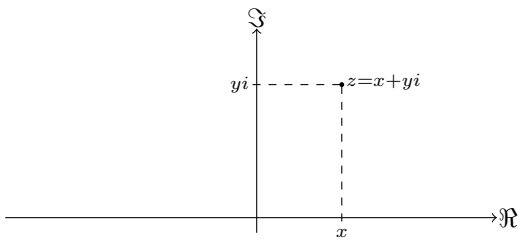
1 Hyperbolinen taso

Kaikkein klassisimmat modulimuodot ovat hyperbolisella tasolla määriteltyjä erittäin säännöllisiä ja söpöjä funktioita. On siis heti aluksi syytä sanoa jotakin tästä näyttämöstä, hyperbolisen tason epäeukleidisesta geometriasta.

Lukuteoriassa suosituin malli hyperboliselle tasolle on puolitasomalli: Määrittelemme *hyperbolisen tason* $\mathbb{H}^2$ kompleksitason ylemmäksi puolitasoksi:

$$\mathbb{H}^2 = \{z \in \mathbb{C} | \Im z > 0\}.$$

Merkitsemme hyperbolisen tason pisteitä z, ja kirjoitamme edelleen $z = x + yi$, missä tietenkin $x \in \mathbb{R}$ ja $y \in \mathbb{R}_+$.



Hyperbolinen taso $\mathbb{H}^2$ on kompleksitason ylempi puolitaso, ja merkitsemme sen pisteitä $z = x + yi$.

[1]Basque Center for Applied Mathematics

Tämä on tietenkin vain yksinkertainen karteesinen parametrisointi pisteille. Niiden varsinaisen geometrian kuvaaminen vaatii hieman enemmän. Ensinnäkin, infinitesimaalinen viivaelementti $\mathrm{d}s_{\mathbb{H}^2}$ pisteessä z saadaan yksinkertaisesti jakamalla vastaava eukleidinen viivaelementti imaginääriosalla:

$$\mathrm{d}s_{\mathbb{H}^2} = \frac{\mathrm{d}s_{\mathbb{R}^2}}{y} = \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y}.$$

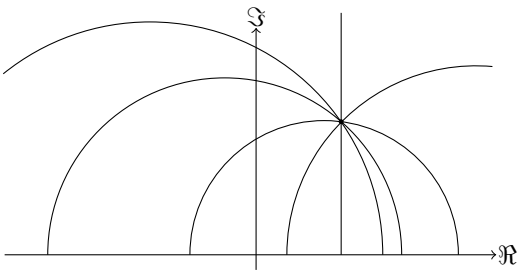
Toisin sanoen, käyrän hyperbolinen pituus saadaan niin, että vastaava eukleidinen pituus jaetaan imaginääriosalla y pisteen z lähellä. Käyrien leikatessa kulmat määritellään samoiksi kuin mitä ne ovat eukleidisessakin mielessä.

Pinta-alat lasketaan vastaavasti niin, että eukleidinen pinta-alamitta jaetaan neliöllä y^2 pisteen z lähellä, jolloin saadaan hyperbolisen tason luonnollinen pinta-alamitta

$$\mathrm{d}\mu = \frac{\mathrm{d}x\,\mathrm{d}y}{y^2}.$$

Pienessä mittakaavassa hyperbolinen geometria muistuttaa kovasti eukleidista. Esimerkiksi r-säteisen kiekon hyperbolinen pinta-ala on $\sim \pi r^2$ kun $r \longrightarrow 0+$. Sen sijaan isoilla r hyperbolisen geometrian varsin poikkeava luonne tulee kirkkaasti esiin: r-säteisen kiekon hyperbolinen pinta-ala on $\sim \pi e^r$ kun $r \longrightarrow \infty \ldots$

Suorien viivojen asemaa toimittavat ne puoliympyrät ja puolisuorat, jotka leikkaavat reaaliakselin kohtisuorasti. On helppo nähdä, että suoran ulkopuolisen pisteen kautta kulkee useampia ”yhdensuuntaisia” suoria, missä siis kaksi suoraa ajatellaan yhdensuuntaisiksi, jos ne eivät leikkaa missään tason $\mathbb{H}^2$ pisteessä.



Erään pisteen kautta kulkevia hyperbolisen tason suoria. Nämä kaikki kohtaavat reaaliakselin kohtisuorasti.

Tähän geometriaan liittyy luonnollisella tavalla hieman eksoottisempikin olio, *Laplace–Beltrami-operaattori*, joka mainitaan vielä myöhemmin:

$$-\Delta_{\mathbb{H}^2} = -y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).$$

Kannaltamme on varsin oleellista tietää, millaisia symmetrioita, eli geometrian säilyttäviä muunnoksia, tasolla $\mathbb{H}^2$ on. Eukleidisessa tasossa tällaisia pituudet ja kulmat säilyttäviä kuvauksia ovat esim. kierrot ja siirrot. Osoittautuu, että hyperbolisen tason $\mathbb{H}^2$ vastaus on yllättävä: pituudet ja kulmat suunnistuksineen säilyttäviä kuvauksia ovat *Möbius-kuvaukset*

$$z \longmapsto \frac{az+b}{cz+d},$$

missä a, b, c ja d ovat reaalilukuja niin, että $ad - bc = 1$, tai toisin sanoen

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$

2 Diskreetit aliryhmät

Ehkäpä yksinkertaisin tapa yrittää kuvailla modulimuotoja olisi sanoa, että ne ovat jossakin hyperbolisessa mielessä ”jaksollisia” tai ”melkein jaksollisia”. Pohtikaamme ensin, mitä jaksollisuus tarkoittaa eukleidisessa tasossa $\mathbb{R}^2$.

Tietenkin funktio $f\colon \mathbb{R}^2 \longrightarrow \mathbb{C}$ on jaksollinen, tai tarkemmin 1-jaksollinen, kun

$$f(x_1 + 1, x_2) = f(x_1, x_2 + 1) = f(x_1, x_2)$$

kaikilla $x = \langle x_1, x_2 \rangle \in \mathbb{R}^2$. Luonnollisesti tämä on yhtäpitävää sen kanssa, että

$$f(x + y) = f(x)$$

kaikilla $x \in \mathbb{R}^2$ ja jokaisella $y \in \mathbb{Z}^2$. Kuvaukset $x \longmapsto x + y$, missä $y \in \mathbb{Z}^2$ muodostavat tason symmetrioiden perheen, joka on tason kaikkien symmetrioiden ryhmän diskreetti aliryhmä.

Hyperbolisessa tasossa ”jaksollisuus”, tai oikeammin sanottuna automorfisuus, on sitä, että funktiolle $f\colon \mathbb{H}^2 \longrightarrow \mathbb{C}$ pätee

$$f\left(\frac{az+b}{cz+d}\right) = f(z)$$

kaikille $z \in \mathbb{H}^2$ ja joillekin Möbius-kuvauksille $z \longmapsto (az+b)/(cz+d)$, joiden olisi muodostettava jonkinlainen kaikkien symmetrioiden ryhmän diskreetti aliryhmä. Itse asiassa osoittautuu, että automorfisuus tässä yksinkertaisessa mielessä ja holomorfisuus eivät yhdessä salli muita kuin vakiofunktiot, mutta jos holomorfisuuden sijaan tyydytään reaalianalyyttisiin funktioihin, niin silloin jo saadaankin erittäin rikas teoria, mutta palaamme tähän lyhyesti myöhemmin.

Voidaksemme puhua ”jaksollisuudesta” hyperbolisessa mielessä meidän valittava diskreetti aliryhmä symmetrioita. Lukuteorian kannalta kiinnostavat aliryhmät muodostavat varsin rikkaita kokonaisuuksia, mutta keskitymme alla vain pariin hyvin erityiseen mutta kuitenkin tärkeään tapaukseen, jotka riittävät myöhemmin esitettäviin esimerkkeihin.

Kun $\lambda \in \{1, 2\}$, otamme tarkasteluun ryhmän $G(\lambda)$, jonka virittävät Möbius-kuvaukset

$$z \longmapsto z + \lambda \quad \text{ja} \quad z \longmapsto -\frac{1}{z},$$

jotka vastaavat matriiseita

$$\begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix} \quad \text{ja} \quad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

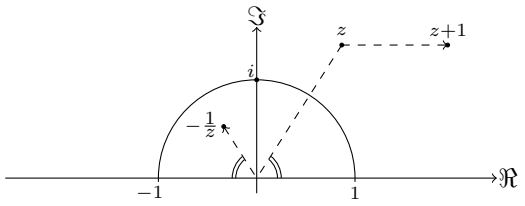
Erityisesti, kun $\lambda = 1$, on kyseessä oleva ryhmä $G(1)$ nimeltään *täysi moduliryhmä*, ja se koostuu Möbius-kuvauksista

$$z \longmapsto \frac{az+b}{cz+d},$$

missä

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2, \mathbb{Z}).$$

Ilmeisistä syistä ryhmää $G(1)$ vastaavia symmetrioita merkitään myös symbolilla $\mathrm{SL}(2,\mathbb{Z})$.



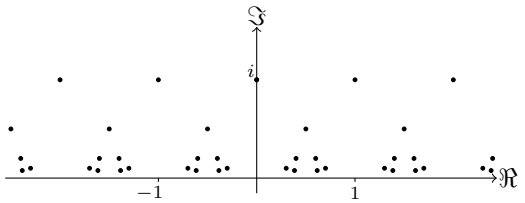
Miten hyperbolisen tason piste z kuvautuu Möbius-kuvauksissa $z \longmapsto z+1$ ja $z \longmapsto -1/z$. Jälkimmäisessä kuvauksessa siis oikeastaan peilataan imaginääriakselin suhteen, ja sen jälkeen etäisyys origosta muutetaan käänteisluvukseen. Kuvan yksikköympyrän kaari kuvautuu itselleen tässä jälkimmäisessä kuvauksessa, mutta sen oikeanpuoleinen kaari kuvautuu vasemmanpuoleiseksi kaareksi, ja kääntäen.

Toinen tapaus on siis $\lambda = 2$, *ϑ-ryhmä*, joka koostuu niistä täyden moduliryhmän Möbius-kuvauksista, joille pätee joko

$$a \equiv d \equiv 1 \pmod 2 \quad \text{ja} \quad b \equiv c \equiv 0 \pmod 2,$$

tai

$$a \equiv d \equiv 0 \pmod 2 \quad \text{ja} \quad b \equiv c \equiv 1 \pmod 2.$$



Pisteen i kuvia täyden moduliryhmän matriiseita vastaavissa Möbius-kuvauksissa. Reaaliakselia lähestyttäessä kuvapisteitä esiintyy eukleidisessa mielessä tiheämmässä ja tiheämmässä.

Ennen ensimmäistä modulimuotojen esittelyä on syytä mainita vielä *perusalueet*. Kun tarkastellaan eukleidisen tason siirtoja $x \longmapsto x+y$ vektoreilla $y \in \mathbb{Z}^2$, niin luonnollisesti näillä siirroilla päästään mistä tahansa tason pisteestä johonkin neliön $[0,1[^2$ pisteeseen, ja kääntäen. Toisaalta, mistään neliön $[0,1[^2$ pisteestä ei voi siirtyä mihinkään toiseen saman neliön pisteeseen näillä siirroilla. Kutsumme neliötä $[0,1[^2$ erääksi siirtojen ryhmän $\mathbb{Z}^2$ *perusalueeksi*.

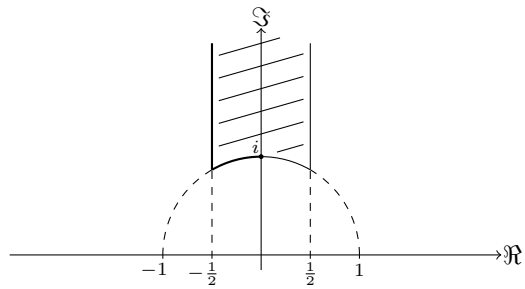
Samassa hengessä hyperbolisessa tasossa mistä tahansa pisteestä päästään ryhmän $G(\lambda)$ Möbius-kuvauksilla johonkin alueen $B(\lambda)$ pisteeseen, ja kääntäen, missä $B(\lambda)$ koostuu niistä pisteistä $z \in \mathbb{H}^2$, joille

$$|z| > 1 \quad \text{ja} \quad -\frac{\lambda}{2} \leqslant x < \frac{\lambda}{2},$$

tai joille

$$|z| = 1 \quad \text{ja} \quad -\frac{\lambda}{2} \leqslant x \leqslant 0.$$

Lisäksi mitkään kaksi eri alueen $B(\lambda)$ pistettä eivät koskaan ole kuvattavissa toisilleen ryhmän $G(\lambda)$ Möbius-kuvauksilla.



Eräs ryhmän $\mathrm{SL}(2,\mathbb{Z})$ perusalue, yllä määritelty $B(1)$. Alueen sisäosan lisäksi alueseen kuuluu myös vahvennetulla viivalla piirretty osuus reunasta piste i mukaan.

3 Holomorfiset modulimuodot

Hyperbolisessa tasossa määriteltyjä modulimuotoja on oleellisesti ottaen kahdenlaisia, holomorfisia modulimuotoja, ja ei-holomorfisia eli reaalianalyyttisiä modulimuotoja. Ihmettelemme holomorfisia muotoja ensimmäisinä.

Olkoon $\lambda \in \{1,2\}$. *Holomorfinen modulimuoto ryhmälle* $G(\lambda)$ on holomorfinen, eli kompleksianalyyttinen, funktio

$$f\colon \mathbb{H}^2 \longrightarrow \mathbb{C},$$

jolle

$$f(z+\lambda) = f(z) \quad \text{ja} \quad f\left(-\frac{1}{z}\right) = z^{\kappa}\, f(z)$$

kaikilla $z \in \mathbb{H}^2$ jollakin $\kappa \in \{0, 2, 4, \ldots\}$, ja jolle $|f(z)|$ on rajoitettu, kun $\Im z \longrightarrow \infty$. Merkitsemme tällöin $f \in M_\kappa(\lambda)$, mikä ei ole standardi merkintä, mutta sopii tämän artikkelin tarkoituksiin. Lukua κ kutsutaan modulimuodon f *painoksi*.

Näistä ehdoista seuraa, että $f \in M_\kappa(\lambda)$ voidaan esittää Fourier-sarjana

$$f(z) = \sum_{n=0}^{\infty} a(n)\, e\Big(\frac{nz}{\lambda}\Big),$$

missä Fourier-kertoimet $a(n)$ ovat kompleksilukuja, ja $e(w) = e^{2\pi i w}$ kaikilla $w \in \mathbb{C}$. Kun $\lambda = 1$ ja $a(0) = 0$, modulimuotoa $f(z)$ kutsutaan *kärkimuodoksi*.

Joitakin huomioita lienee syytä tehdä. Ensinnäkin, yllä annettu määritelmä ei ole suinkaan yleisin mahdollinen, vaan juuri niin yleinen, kuin pian vastaan tulevat esimerkit vaativat. Huomautettakoon myös, että Fourier-kertoimet usein ovat erittäin merkittäviä lukuteoreettisia otuksia, mistä esimerkit jäljempänä todistavatkin.

Lisäksi kärkimuotojen Fourier-kertoimet ovat hyvin haastavia matemaattisesti. Kaikkein klassisin esimerkki tästä lienee Delignen juhlittu arvio, jonka mukaan jollakin vakiolla $C \in \mathbb{R}_+$ pätee kaikilla $n \in \mathbb{Z}_+$

$$|a(n)| \leqslant C\, d(n),$$

missä $d(n)$ tarkoittaa luvun n positiivisten tekijöiden lukumäärää. Tämä tulos, jota toisinaan on luonnehdittu siksi matematiikan tuloksesi, jolle väitteen pituuden ja todistuksen pituuden suhde on pienin, perustui Delignen erittäin syvälliseen työhön algebrallisessa geometriassa Weilin konjektuureihin liittyen.

4 Modulimuotojen avaruudet

Esimerkkinä siitä, millaisia kokonaisuuksia avaruudet $M_\kappa(\lambda)$ ovat, käymme esimerkkinä läpi tapauksen $\lambda = 1$. Erityisen oleellista esimerkissä on se, että modulimuotojen muodostamat avaruudet ovat yllättävän hallittuja kokonaisuuksia.

Ensinnäkin, on suhteellisen helppo rakentaa esimerkkejä modulimuodoista: jos $\kappa \geqslant 4$ on parillinen kokonaisluku, niin

$$E_\kappa(z) = \frac{1}{2\zeta(\kappa)} \sum_{\langle m,n\rangle \in \mathbb{Z}^2 \setminus 0} \frac{1}{(mz+n)^\kappa}$$

on painoa κ oleva modulimuoto ryhmälle $\mathrm{SL}(2,\mathbb{Z})$, eli $E_\kappa \in M_\kappa(1)$. Osoittautuu myös, että näiden modulimuotojen, *holomorfisten Eisensteinin sarjojen*, Fourier-kertoimet ovat poikkeuksellisen yksinkertaisia; on nimittäin

$$E_\kappa(z) = 1 + (-1)^{\kappa/2} \frac{2\kappa}{B_\kappa} \sum_{n=1}^{\infty} \sigma_{\kappa-1}(n)\, e(nz),$$

missä B_κ on murtoluku nimeltä Bernoullin luku, ja $\sigma_{\kappa-1}(n)$ merkitsee luvun n positiivisten tekijöiden $(\kappa-1)$. potenssien summia:

$$\sigma_{\kappa-1}(n) = \sum_{d|n} d^{\kappa-1}.$$

Luonnollisesti tekijän $1/2\zeta(\kappa)$ ajatuksena on saada vakiotermiksi 1. Joka tapauksessa, ensimmäiset holomorfiset Eisensteinin sarjat ovat

$$E_4(z) = 1 + 240 \sum_{n=1}^{\infty} \sigma_3(n)\, e(nz),$$

$$E_6(z) = 1 - 504 \sum_{n=1}^{\infty} \sigma_5(n)\, e(nz),$$

sekä

$$E_8(z) = 1 + 480 \sum_{n=1}^{\infty} \sigma_7(n)\, e(nz).$$

Avaruuksien $M_\kappa(1)$ ymmärtämisen kannalta on erittäin kriittistä ymmärtää holomorfisten modulimuotojen nollakohtia. Kun $z_0 \in \mathbb{H}^2$ ja $f \in M_\kappa(1)$ ei häviä identtisesti, määrittelemme pisteen z_0 moninkertaisuuden $\nu_{z_0}(f)$ nollakohtana suurimmaksi $\nu \in \{0, 1, 2, \ldots\}$, jolle $f(z)/(z-z_0)^\nu$ on holomorfinen. Samassa hengessä määrittelemme myös pisteen ∞ moninkertaisuuden nollakohtana $\nu_\infty(f)$ yksinkertaisesti pienimmäksi $\nu \in \{0, 1, 2, \ldots\}$, jolle $a(\nu) \neq 0$.

Nyt funktioteoriasta seuraa tärkeä tulos, *painokaava*: Jos $f \in M_\kappa(1)$, ja jos f ei häviä identtisesti, niin

$$v_\infty(f) + \frac{1}{2} v_i(f) + \frac{1}{3} v_\omega(f) + \sum_{\rho \in B(1) \setminus \{i,\omega\}} v_\rho(f) = \frac{\kappa}{12},$$

missä $\omega = (-1 + i\sqrt{3})/2$ on perusalueen $B(1)$ vasen alanurkka. Tästä kaavasta seuraa mm., että $M_0(1)$ sisältää vain vakiofunktiot, $M_2(1)$ sisältää

vain nollafunktion, ja että $M_4(1)$, $M_6(1)$, $M_8(1)$ ja $M_{10}(1)$ sisältävät kukin vain ja ainoastaan vastaavan Eisensteinin sarjan E_κ kompleksiset monikerrat.

Loppujen avaruuksien $M_\kappa(1)$ rakenteita voi ymmärtää näistä pienempipainoisista lähtien rekursiivisesti. Jos otetaan käyttöön tärkeä modulimuoto

$$\Delta(z) = e(z) \prod_{n=1}^{\infty} \big(1 - e(nz)\big)^{24},$$

jolle $\Delta \in M_{12}(1)$, ja joka on kärkimuoto, niin tällöin kuvaus

$$f \longmapsto \Delta\, f$$

on hyvä kuvaus $M_\kappa(1) \longrightarrow M_{\kappa+12}(1)$, ja vieläpä kuvaa avaruuden $M_\kappa(1)$ bijektiivisesti avaruuden $M_{\kappa+12}(1)$ kärkimuodoille. Toisin sanoen, jokainen $F \in M_{\kappa+12}(1)$ voidaan kirjoittaa muodossa

$$F = \alpha\, E_{\kappa+12} + \Delta\, f,$$

jollakin vakiolla $\alpha \in \mathbb{C}$ ja jollakin modulimuodolla $f \in M_\kappa(1)$. Lisäksi tästä näkee myös mukavasti avaruuksien $M_\kappa(1)$ dimensiot $\delta_\kappa = \dim M_\kappa(1)$:

κ	δ_κ	κ	δ_κ	κ	δ_κ	κ	δ_κ	$\cdots$
0	1	12	2	24	3	36	4	$\cdots$
2	0	14	1	26	2	38	3	$\cdots$
4	1	16	2	28	3	40	4	$\cdots$
6	1	18	2	30	3	42	4	$\cdots$
8	1	20	2	32	3	44	4	$\cdots$
10	1	22	2	34	3	46	4	$\cdots$

Itse asiassa osoittautuu, että

$$1728\,\Delta = E_4^3 - E_6^2,$$

ja että kaikki avaruuksien $M_\kappa(1)$ modulimuodot ovat modulimuotojen E_4 ja E_6 polynomeja.

Muita ja yleisempiä holomorfisten modulimuotojen avaruuksia voi ymmärtää samalla tavalla. Esimerkiksi avaruuden $M_\kappa(2)$ dimensio on

$$\dim M_\kappa(2) = 1 + \left\lfloor \frac{\kappa}{4} \right\rfloor.$$

Lisäksi jokaiselle $f \in M_\kappa(2)$ löytyy yksikäsitteiset kompleksiset kertoimet α_0, α_1, …, $\alpha_{\lfloor \kappa/4 \rfloor}$ niin, että

$$f = \sum_{\ell=0}^{\lfloor \kappa/4 \rfloor} \alpha_\ell\, \vartheta^{2\kappa-4\ell}\, E_4^\ell,$$

missä

$$\vartheta(z) = \sum_{n\in\mathbb{Z}} e\Big(\frac{n^2 z}{2}\Big).$$

5 Neliöiden summat

Tarkastelemme seuraavaksi neliöiden summia. Jos $k \in \mathbb{Z}_+$ ja $n \in \mathbb{Z}_+ \cup \{0\}$, niin $r_k(n)$ tarkoittaa niiden kokonaislukuvektoreiden $\langle x_1, x_2, \ldots, x_k\rangle$ lukumäärää, joille

$$x_1^2 + x_2^2 + \ldots + x_k^2 = n.$$

Modulimuodoilla voi todistaa luonnollisella tavalla kauniita kaavoja tällaisille lukumäärille. Tässä on pieni hahmotelma siitä, miten.

Tarkastellaan *ϑ-funktiota*

$$\vartheta_8(z) = \sum_{n=0}^{\infty} r_8(n)\, e\Big(\frac{nz}{2}\Big).$$

Välittömästi oleellisinta kannaltamme on, että $\vartheta_8 \in M_4(2)$. Tämän todistus on Poissonin summauskaavan sovellus.

Toisaalta voi osoittaa, että otukselle

$$\begin{aligned}\Theta(z) &= \frac{1}{15}\left(16\,E_4(z) - E_4\Big(\frac{z+1}{2}\Big)\right)\\ &= 1 + 16\sum_{n=1}^{\infty}\sum_{d\mid n}(-1)^{n-d}\,d^3\,e\Big(\frac{nz}{2}\Big),\end{aligned}$$

missä viimeisessä Σ-merkissä summataan luvun n positiivisten tekijöiden d yli, pätee samaten $\Theta \in M_4(2)$. Lisäksi modulimuotojen ϑ_8 ja Θ Fourier-kertoimet alkupäästä ovat yhtä suuria; onhan nimittäin

$$r_8(0) = 1, \quad r_8(1) = 16, \quad \text{ja} \quad r_8(2) = 112,$$

ja toisaalta

$$16\sum_{d\mid 1}(-1)^{1-d}\,d^3 = 16\,(-1)^{1-1}\,1^3 = 16 = r_8(1),$$

ja

$$\begin{aligned}16\sum_{d\mid 2}(-1)^{2-d}\,d^3 &= 16\left((-1)^{2-1}\,1^3 + (-1)^{2-2}\,2^3\right)\\ &= 16\,(-1+8) = 112 = r_8(2).\end{aligned}$$

Mutta jos kahdella modulimuodolla on samat parametrit, ja riittävän monta kerrointa alusta ovat yhteisiä, niin relevantin painokaavan ansiosta modulimuodot itse asiassa ovat *samat* ja *kaikki* kertoimet ovat samat! Osoittautuu, että modulimuodoille $M_4(2)$ riittää vain kolme yhteistä kerrointa...

Täten $\vartheta_8 = \Theta$ ja

$$r_8(n) = 16 \sum_{d|n} (-1)^{n-d} d^3$$

kaikille $n \in \mathbb{Z}_+$.

Aivan samassa hengessä, mutta hieman monimutkaisemmilla yksityiskohdilla ja yleisemmillä modulimuodoilla voisimme todistaa esimerkiksi Jacobin neljän neliön lauseen, jonka mukaan

$$r_4(n) = 8 \sum_{\substack{d|n, \\ 4\nmid d}} d,$$

missä summataan luvun n neljällä jaottomien positiivisten tekijöiden d yli (mistä erityisesti seuraisi, että jokainen n on kirjoitettavissa neljän neliön summana), tai voisimme todistaa vaikkapa, että

$$r_2(n) = 4 \sum_{\substack{d|n, \\ d\equiv 1(4)}} d - 4 \sum_{\substack{d|n, \\ d\equiv -1(4)}} d,$$

missä siis summataan niiden luvun n positiivisten tekijöiden d yli, joille $d \equiv \pm 1 \pmod 4$.

Parittoman monen neliön summat johtavat kyllä muuten kivoihin ϑ-funktioihin, mutta vaatii holomorfisten modulimuotojen teoriaa, missä painoksi sallitaan parittoman kokonaisluvun puolikas.

6 Koodit

Siirrytään sitten seuraavaksi toisenlaiseen ϑ-funktioiden sovellusalaan. *Koodi*, tai oikeammin *binäärinen lineaarinen lohkokoodi*, on vektoriavaruuden $\mathrm{GF}(2)^n$ vektorialiavaruus, missä $n \in \mathbb{Z}_+$ ja $\mathrm{GF}(2)$ on kahden elementin kunta $\mathbb{Z}/2\mathbb{Z}$, jonka elementtejä voi merkitä yksinkertaisesti 0 ja 1. Tällainen koodi on siis joukko n elementin vektoreita, kuten $\langle 0, 1, 1, 0, \ldots, 1\rangle$, joita lasketaan yhteen yksinkertaisesti komponenteittain modulo 2.

Sanomme, että koodi C on $\langle n, s, d\rangle$-koodi (missä $n, s, d \in \mathbb{Z}_+$), jos C sisältää s vektoria avaruudesta $\mathrm{GF}(2)^n$, ja mitkä tahansa kaksi koodin C vektoria poikkeavat vähintään d elementin verran. Tietenkin ajatus on siinä, että jos kommunikoidessa tai tallennettaessa vähemmän kuin $d/2$ komponenttia muuttuvat virheellisiksi, niin voimme silti pelastaa koko vektorin. Osoittautuu, että voi konstruoida koodeja, jotka selviävät hyvinkin suurista määristä virheitä hyvinkin pienellä määrällä ylimääräisiä bittejä. Lineaariset koodit ovat erityisen käytännöllisiä koodauksen ja dekoodauksen kannalta.

Antakaamme yksi esimerkki koodista. Seuraava $\langle 8, 16, 4\rangle$-koodi on nimeltään *laajennettu Hammingin koodi*:

00000000	01000111	10001101	11001010
00011011	01011100	10010110	11010001
00101110	01101001	10100011	11100100
00110101	01110010	10111000	11111111

Tässä koodissa siis on 16 koodisanaa, joista mitkä tahansa kaksi poikkeavat toisistaan ainakin neljän bitin osalta. Tämä koodi on selvästikin sellainen, että yhden virheellisen bitin pystyy aina korjaamaan. Koska koodisanoja on $16 = 2^4$, pystyy yhdellä koodisanalla siis välittämään neljä bittiä informaatiota. Jos saman virheenkorjausominaisuuden haluaisi esimerkiksi yksinkertaisesti toistamalla bittejä, niin tämä vaatisi jokaisen bitin lähettämisen kolmeen kertaan; eli kolminkertaisen määrän bittejä. Laajennettu Hammingin koodi saavuttaa saman virheenkorjausominaisuuden vain kaksinkertaisella bittimäärällä.

Määritellään vielä joitakin käsitteitä. Alla kuvattua teoriaa voi tehdä yleisemmillekin koodeille, mutta mieluummin esittelemme aihetta kevyesti ja klassisesti yleisyyden kustannuksella.

Yllä kuvatun kaltaiseen $\langle n, s, d\rangle$-koodiin C liittyy luonnollisella tavalla *duaalikoodi* $C^\perp$, joka määritellään näin:

$$C^\perp = \left\{ x \in \mathrm{GF}(2)^n \middle| x \cdot y = 0 \text{ kaikilla } y \in C \right\},$$

missä $x \cdot y$ merkitsee pistetuloa $x_1 y_1 + \ldots + x_n y_n$. Erään erityisen tärkeän koodien luokan muodostavat *itseduaalit koodit*, eli ne koodit C, joille $C = C^\perp$, joilla on merkillisiä ominaisuuksia,

joita kohta pohdimme. Laajennettu Hammingin koodi on itseduaali.

Esittelemme vielä yhden koodiin C liittyvän otuksen, jonka mainitsemme pian uudelleen. Koodisanan $x \in C$ *paino* $w(x)$ on siinä esiintyvien ykkösten lukumäärä. Koodin C *painoluetteloija* $W_C(X,Y)$ on kahden muuttujan X ja Y polynomi

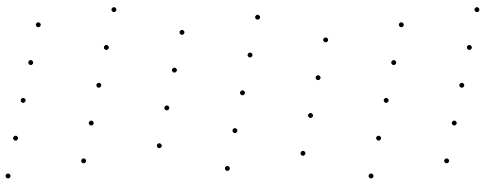
$$W_C(X,Y) = \sum_{x \in C} X^{n-w(x)}\, Y^{w(x)}.$$

Siis polynomissa $W_C(X,Y)$ termin $X^\alpha\, Y^\beta$ kerroin on niiden koodisanojen $x \in C$ lukumäärä, joissa esiintyy α nollaa ja β ykköstä. Esimerkiksi laajennetun Hammingin koodin painoluetteloija on $X^8 + 14\, X^4\, Y^4 + Y^8$.

Lopuksi, koodia kutsutaan *kahdesti parilliseksi*, jos sen jokaisen koodisanan paino on neljällä jaollinen. Laajennettu Hammingin koodi on selvästikin kahdesti parillinen.

7 Koodien ϑ-funktiot

Mutta miten koodit liittyvätkään modulimuotoihin? Siirrytään toiseen olioiden luokkaan, hiloihin. Jos $A \in \mathbb{R}^{n\times n}$ on kääntyvä matriisi, niin siihen liittyy n-ulotteinen *hila* $\Gamma \subset \mathbb{R}^n$, joka koostuu yksinkertaisesti pisteistä Ax, missä $x \in \mathbb{Z}^n$. Yksinkertaisin esimerkki hilasta on luonnollisesti $\mathbb{Z}^n$ itse.



Erään kaksiulotteisen hilan pisteitä.

Olkoon nyt C jokin $\langle n, s, d\rangle$-koodi. Voimme konstruoida siitä hilan $\Gamma_C \subset \mathbb{R}^n$ yksinkertaisesti ottamalla hilaan Γ_C mukaan ne vektorit $x/\sqrt{2}$, missä $x = \langle x_1, x_2, \ldots, x_n\rangle \in \mathbb{Z}^n$ ja joille

$$x_1 \equiv c_1, \quad x_2 \equiv c_2, \quad \ldots, \quad x_n \equiv c_n \pmod 2,$$

jollakin koodisanalla $\langle c_1, c_2, \ldots, c_n\rangle \in C$.

Jos meillä on hila Γ jonka vektoreiden pistetulot ovat parillisia kokonaislukuja, niin voimme määritellä siihen liittyvän ϑ-*funktion* sarjana

$$\vartheta_\Gamma(z) = \sum_{x \in \Gamma}^{\infty} e\!\left(\frac{(x\cdot x)\, z}{2}\right).$$

Osoittautuu, että kahdesti parillista itseduaalia koodia C vastaavan hilan Γ_C ϑ-funktio ϑ_{Γ_C} on modulimuoto! Tarkalleen ottaen $\vartheta_{\Gamma_C} \in M_{n/2}(1)$, ja lisäksi dimension n on oltava kahdeksalla jaollinen.

Määritellään pari pientä apu-ϑ-funktiota A ja B asettamalla

$$A(z) = \sum_{x \in 2\mathbb{Z}} e\!\left(\frac{x^2\, z}{4}\right),$$

missä siis sarja otetaan parillisten kokonaislukujen yli, sekä

$$B(z) = \sum_{x \in 2\mathbb{Z}+1} e\!\left(\frac{x^2\, z}{4}\right),$$

missä siis sarja otetaan parittomien kokonaislukujen yli. Näiden määritelmien taustalla on se mielenkiintoinen seikka, että

$$W_C(A,B) = \vartheta_{\Gamma_C}.$$

Koska ϑ-funktion ϑ_{Γ_C} paino on neljällä jaollinen, seuraa täyden moduliryhmän modulimuotojen avaruuksien rakenteesta, että ϑ_{Γ_C} on modulimuotojen E_4 ja Δ polynomi. Toisaalta osoittautuu, että

$$E_4 = A^8 + 14\, A^4\, B^4 + B^8$$

ja

$$16\,\Delta = A^4\, B^4 \left(A^4 - B^4\right)^4,$$

ja näin päädytään *Gleasonin lauseeseen*: kahdesti parillisen itseduaalin koodin painoluetteloija $W_C(X,Y)$ on aina lausekkeiden

$$X^8 + 14\, X^4\, Y^4 + Y^8 \quad \text{ja} \quad X^4\, Y^4 \left(X^4 - Y^4\right)^4$$

polynomi!

Samoilla työkaluilla voi osoittaa myös *MacWilliamsin identiteetin*:

$$W_C(X,Y) = W_C\!\left(\frac{X+Y}{\sqrt{2}}, \frac{X-Y}{\sqrt{2}}\right).$$

8 Ositukset

Neliöiden summien ohella lienee paikallaan mainita myös toinenkin hyvin klassinen aihe: ositukset.

Olkoon $n \in \mathbb{Z}_+$. Määrittelemme *ositusluvun* $p(n)$ niin, että se kertoo, kuinka monella eri tavalla luku n voidaan kirjoittaa positiivisten kokonaislukujen summana, kun termien järjestyksellä ei ole väliä ja summassa mikä tahansa termi saa esiintyä useamman kerran. Lisäksi määrittelemme $p(0) = 1$.

Esimerkiksi, tietenkin $p(1) = 1$. Koska $2 = 1 + 1$, on $p(2) = 2$. Koska $3 = 2 + 1 = 1 + 1 + 1$, on $p(3) = 3$. Edelleen, koska $4 = 3 + 1 = 2 + 2 = 2 + 1 + 1 = 1 + 1 + 1 + 1$, on $p(4) = 5$.

Hyvä, $p(n)$ on ilmeisen luonnollinen tutkimuskohde, mutta miten se liittyykään modulimuotoihin? On helppo vakuuttua siitä, että

$$\begin{aligned}\sum_{n=0}^{\infty} p(n)\, e(nz) &= \prod_{n=1}^{\infty} \big(1 + e(nz) + e(2nz) + e(3nz) + \ldots\big) \\ &= \prod_{n=1}^{\infty} \frac{1}{1 - e(nz)}.\end{aligned}$$

Siis osituksilla on ilmeinen yhteys *η-funktioon*, joka määritellään tulona

$$\eta(z) = e\Big(\frac{z}{24}\Big) \prod_{n=1}^{\infty} \big(1 - e(nz)\big)\,,$$

sillä onhan

$$\frac{1}{\eta(z)} = e\Big(\frac{-z}{24}\Big) \sum_{n=0}^{\infty} p(n)\, e(nz).$$

Tämä otus $\eta(z)$ on erittäin klassinen holomorfinen modulimuoto, vaikkakin monimutkaisempi kuin mitä yllä annettu yksinkertainen modulimuodon määritelmä sallii. Sen modulaarisesta luonteesta kuitenkin vahvasti todistaa se, että kärkimuoto $\Delta(z)$ on η-funktion 24. potenssi.

Eräs esimerkki tyypillisestä analyyttisen lukuteorian kysymyksestä olisi: kuinka suuri $p(n)$ suurin piirtein on, kun n on iso? Sanotaan, että jos katsotaan osituslukujen taulukkoa, niin sen toinen laita näyttää kaukaa katsottuna vähän paraabelin kaarelta. Lukija voi itse päättää, onko tässä perää vai ei:

Taulukko ositusluvuista $p(1)$, $p(2)$, $p(3)$, …, $p(400)$. Tilan säästämisen nimissä taulukkoa on kierretty suoran kulman verran vasemmalle.

Joka tapauksessa, jos tässä näkee paraabelin kaaren niin silloin arvaisi, että olisi $\log p(n)$ jotakin sen kaltaista kuin $A\sqrt{n}$, jollakin positiivisella vakiolla A, ja siis $p(n)$ näyttäisi olevan suurin piirtein $e^{A\sqrt{n}}$, tai sinne päin. Kaunis klassinen tulos sanookin, että kun $n \longrightarrow \infty$, on

$$p(n) \sim \frac{e^{A\sqrt{n}}}{4\sqrt{3}\,n}, \quad \text{ts.} \quad \lim_{n\longrightarrow\infty} \frac{4\sqrt{3}\,n \cdot p(n)}{e^{A\sqrt{n}}} = 1,$$

missä $A = \pi\sqrt{2/3}$. Itse asiassa η-funktion teoria antaa tämän ohella huomattavan paljon tarkempaakin asymptoottista tietoa ja lisäksi sillä voi osoittaa kauniita *Ramanujanin kongruensseja*: luvuille $n \in \mathbb{Z}_+$ ja $\alpha \in \mathbb{Z}_+$ pätevät mm. Watsonin todistamat implikaatiot

$$24n \equiv 1 \,(\mathrm{mod}\, 5^\alpha) \implies 5^\alpha \mid p(n)$$

sekä

$$24n \equiv 1 \,(\mathrm{mod}\, 7^\alpha) \implies 7^{\lfloor \alpha/2 \rfloor + 1} \mid p(n).$$

9 Spektraaliteoria

Osoittautuu, että jokainen riittävän kiltti automorfinen $f\colon \mathbb{H}^2 \longrightarrow \mathbb{C}$, eli funktio, jolle

$$f\Big(\frac{az+b}{cz+d}\Big) = f(z), \text{ kaikille } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}(2,\mathbb{Z}),$$

voidaan esittää muodossa

$$f(z) = \text{vakio} + \text{diskreetti osa} + \text{jatkuva osa},$$

missä oikean puolen ensimmäinen termi on vain jokin kompleksivakio.

Diskreetti osa voidaan kirjoittaa Fourier-sarjan kaltaisena kehitelmänä $\sum_{j=1}^{\infty} a_n\, \psi_j(z)$, missä jokainen $\psi_j(z)$ on automorfinen Laplace–Beltrami-operaattorin ominaisfunktio. Eli, jokainen $\psi_j\colon \mathbb{H}^2 \longrightarrow \mathbb{C}$ on automorfinen reaalianalyyttinen funktio, jolle ψ_j on *neliöintegroituva* siinä mielessä, että

$$\int\limits_{B(1)} |\psi_j|^2 \,\mathrm{d}\mu < \infty,$$

ja se on operaattorin $-\Delta_{\mathbb{H}^2}$ ominaisfunktio, eli

$$-\Delta_{\mathbb{H}^2}\psi_j = \lambda_j\psi_j,$$

missä kertoimia $\lambda_1 \leqslant \lambda_2 \leqslant \dots$ kutsutaan operaattorin $-\Delta_{\mathbb{H}^2}$ *diskreetiksi spektriksi*. Funktioita ψ_j kutsutaan *Maassin muodoiksi*, tai *ei-holomorfisiksi kärkimuodoiksi*.

Jatkuva osa puolestaan on Fourier-integraalin kaltainen integraali, jossa integroidaan eksponenttifunktion sijaan *ei-holomorfisia Eisensteinin sarjoja* vasten. Viimeksi mainitut muistuttavat holomorfisia Eisensteinin sarjoja, mutta ne ovat vain reaalianalyyttisiä, eivät holomorfisia, ja ne toteuttavat Laplace–Beltrami-operaattorin ominaisarvoyhtälön olematta kuitenkaan neliöintegroituvia.

Kuten holomorfisilla modulimuodoillakin, näilläkin olioilla on Fourier-kehitelmänsä. Esimerkiksi, Maassin muodon ψ_j, joka vastaa ominaisarvoa λ_j, voi kirjoittaa Fourier-sarjana

$$\psi_j(z) = \sqrt{y}\sum_{n\neq 0} t_j(n)\, K_{i\sqrt{\lambda_j - 1/4}}(2\pi\,|n|\,y)\, e(nx),$$

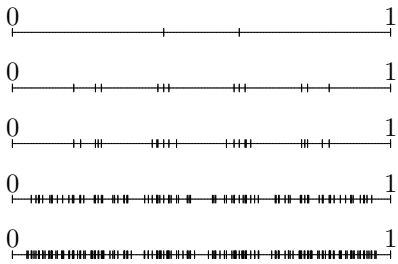
missä summaus on siis nollasta poikkeavien kokonaislukujen yli, ja K merkitsee tavallista K-Bessel-funktiota. Nämä kertoimet $t_j(n)$ ovat hyvin samantapaisia kuin holomorfisten kärkimuotojen Fourier-kertoimet, mutta niistä tiedetään vähemmän.

Luonnollisesti kaikkialla, myös lukuteoriassa, esiintyy paljon jaksollisia funktioita, ja siten klassiset Fourier-sarjat ovat lukuteoriassakin erittäin tärkeitä. Hieman yllättäen yllä kuvattu spektraalihajotelma *on osoittautunut hämmästyttävän voimakkaaksi työkaluksi analyyttisessä lukuteoriassa*, ja sitä on käytetty menestyksekkäästi niin Kloostermanin summien, aritmeettisten jonojen alkulukujen, Riemannin ζ-funktion ja monien muiden L-funktioiden kuin vaikkapa joidenkin tasanjakautuneisuusilmiöidenkin tutkimuksessa.

10 Tasanjakautumisilmiöitä

Näin lopuksi alla on pari kaunista esimerkkiä tasanjakautumisilmiöistä, joissa modulimuodot näyttelevät tärkeitä osia.

Eräs neliönjäännösten teorian perustulos sanoo, että jos p on alkuluku, jolle $p \equiv 1 \pmod 4$, niin löytyy $x \in \mathbb{Z}$, jolle $x^2 \equiv -1 \pmod p$. Tällainen ratkaisu voidaan valita täsmälleen kahdella eri tavalla niin, että $0 \leqslant x < p$, jolloin $x/p \in [0,1[$. Kun $T \in \mathbb{R}_+$, tarkastellaan näiden osamäärien joukkoa X_T niille alkuluvuille p, joille $p \leqslant T$ ja $p \equiv 1 \pmod 4$. Duke, Friedlander ja Iwaniec todistivat, että joukon X_T luvut jakautuvat tasaisesti välille $[0,1[$, kun $T \longrightarrow \infty$.



Joukot X_{10}, X_{50}, X_{100}, X_{500} ja X_{1000}.

Vaihtoehtoisesti tuloksen voisi myös muotoilla näin: jos $f\colon [0,1] \longrightarrow \mathbb{C}$ on jatkuva, niin

$$\frac{1}{\#X_T}\sum_{x\in X_T} f(x) \xrightarrow[T\longrightarrow\infty]{} \int_0^1 f.$$

Tämän todistus perustuu paitsi alkulukujen teoriaan ja seulateoriaan, myös yllä kuvattuun spektraaliteoriaan, eli Maassin muotoihin ja epäholomorfisiin Eisensteinin sarjoihin.

Toinen, vieläkin kauniimpi, esimerkki on Linnikin ongelma. Olkoon N niiden lukujen $n \in \mathbb{Z}_+$ joukko, joille $n \equiv 1$, 2, 3, 5 tai 6 (mod 8).

Klassinen Legendren lause takaa, että jos $n \in N$, niin n on kolmen neliön summa. Tämä puolestaan tarkoittaa sitä, että $\sqrt{n}$-säteisen origokeskisen pallonkuoren pinnalla on kokonaislukukoordinaattisia pisteitä. Projisoidaan nämä pisteet yksikköpallonkuorelle säteittäisesti, jolloin saadaan joukko

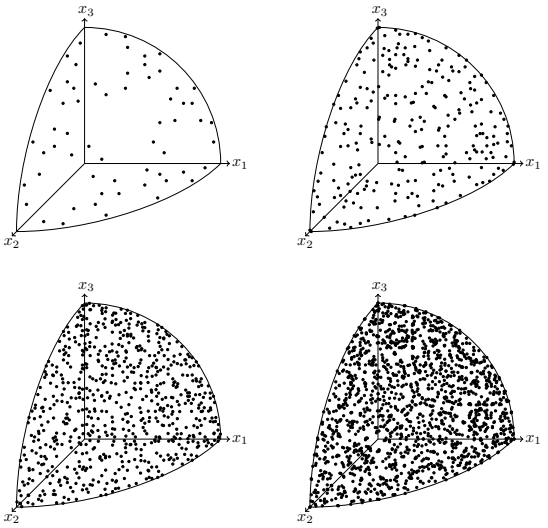
$$\Omega_n = \left\{ \frac{x}{|x|} \middle| x \in \mathbb{Z}^3, |x|^2 = n \right\}.$$

Linnik kysyi, ovatko joukon Ω_n pisteet tiheämmässä ja tiheämmässä yksikköpallonkuorella $S^2 = \{x \in \mathbb{R}^3 \,|\, |x| = 1\}$, kun $n \in N$ kasvaa rajatta, ja jakautuvatko ne sille tasaisesti? Duke todisti, että näin on: jos $f\colon S^2 \longrightarrow \mathbb{C}$ on jatkuva, niin

$$\frac{1}{\#\Omega_n}\sum_{x\in\Omega_n} f(x) \xrightarrow[n\longrightarrow\infty]{n\in N} \frac{1}{4\pi}\int_{S^2} f,$$

missä 4π on tietenkin vain pallonkuoren S^2 ala.

Tämä tulos perustuu ϑ-funktioihin, jotka parittoman monen neliön ollessa kyseessä ovat puolikokonaispainoisia holomorfisia modulimuotoja. Todistus perustuukin oleelliselta osin puolikokonaispainoisten holomorfisten kärkimuotojen Fourier-kertoimien arviointiin.



Ylempänä joukkojen Ω_{1001} ja Ω_{10001} ja alempana joukkojen Ω_{100001} ja $\Omega_{1000001}$ ne pisteet, joiden koordinaatit ovat ei-negatiivisia.

Lähteet

Seuraavassa on mainittu vain ne lähteet, joille tämä artikkeli ja sen kirjoittaja ovat eniten velkaa.

[1] Ebeling, W.: *Lattices and Codes. A Course Partially Based on Lectures by Friedrich Hirzebruch*, Springer, 2013.

[2] Ernvall-Hytönen, A.-M.: *Johdatus modulimuotoihin ja Linnikin ongelmaan*, luentomuistiinpanot, Helsingin yliopisto, kevät 2012.

[3] Gunning, R. C.: *Lectures on Modular Forms*, Annals of Mathematics Studies, 48, Princeton University Press, 1962.

[4] Jutila, M.: *Modulimuodot*, luentomuistiinpanot, Turun yliopisto, syksy 2002.

[5] Knopp, M. I.: *Modular Functions in Analytic Number Theory*, AMS Chelsea Publishing, 2008.

[6] Koecher, M., ja A. Krieg: *Elliptische Funktionen und Modulformen*, Springer, 2007.

[7] Kowalski, E.: *Un cours de théorie analytique des nombres*, Cours spécialisés, 13, Société Mathématique de France, 2004.

[8] Motohashi, Y.: *Spectral Theory of the Riemann Zeta-Function*, Cambridge Tracts in Mathematics, 127, Cambridge University Press, 1997.

[9] Newman, D. J.: *Analytic Number Theory*, Graduate Texts in Mathematics, 177, Springer, 2000.

[10] Ogg, A.: *Modular Forms and Dirichlet Series*, W. A. Benjamin, 1969.

[11] Serre, J.-P.: *Cours d'arithmétique*, Presses universitaires de France, 1970.