

Sovellus satunnaisprojektiolle: korkeaulotteisen vektorin palauttaminen satunnaisista lineaarisista mittauksista

VISA MIKKOLA

PRO GRADU -TUTKIELMA
2014

OHJAAJA
SAMULI SILTANEN

TOINEN TARKASTAJA
PETRI OLA

HELSINGIN YLIOPISTO
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
MATEMATIIKAN JA TILASTOTIETEEN LAITOS
MATEMATIIKKA

Tiedekunta/Osasto — Fakultet/Sektion — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen		Matematiikan ja tilastotieteen laitos	
Tekijä — Författare — Author Visa Mikkola			
Työn nimi — Arbetets titel — Title Sovellus satunnaisprojektiolle: korkeaulotteisen vektorin palauttaminen satunnaisista lineaarisista mittauksista			
Oppiaine — Läroämne — Subject Matematiikka			
Työn laji — Arbetets art — Level Pro gradu -tutkielma		Aika — Datum — Month and year Syyskuu 2014	Sivumäärä — Sidoantal — Number of pages 69 s.
Tiivistelmä — Referat — Abstract			
<p>Satunnaisprojektiio on menetelmä korkeaulotteisen datamatriisin $X \in \mathbb{R}^{n \times m}$ ulottuvuuksien vähentämiseksi. Menetelmässä alkuperäinen n-ulotteinen data kuvataan satunnaisista suunnista virityvälle matalaulotteiselle, $k \ll n$, aliavaruudelle siten, että datavektorien väliset etäisyydet säilyvät approksimatiivisesti. Kuvauksen tekevän matriisin $R \in \mathbb{R}^{k \times n}$ alkiot voidaan valita esimerkiksi normaalijakaumasta ja siksi menetelmän laskennallinen toteutus saadaan hyvin kevyeksi. Tästä syystä monissa tutkimuksissa menetelmä on asetettu vastakkain perinteisen pääkomponenttianalyysin (engl. Principal component analysis, PCA) kanssa, joka on usein laskennallisesti huomattavasti vaativampi ulottuvuuksien vähentämismenetelmä. Samaa vastakkainasettelua käytetään tämän tutkielman satunnaisprojektiosta kertovassa johdantoluvussa, kun menetelmästä esitettyä teoriaa havainnollistetaan numeerisin testein.</p> <p>Satunnaisprojektion toimivuutta on jo tutkittu muun muassa kasvojentunnistuksen sekä kuva- ja tekstidatan ulottuvuuksien vähentämisen yhteydessä. Tutkielman pääaiheeksi on nostettu menetelmän sovellus käänteisongelmaan, jossa pyrkimyksenä on palauttaa tuntematon vektori $x \in \mathbb{R}^n$ approksimatiivisesti siitä tehdyistä satunnaisista mittauksista, kun x oletetaan heikkoon kuuluaan $wl_p(L)$ sisältyväksi vektoriksi. Oletus tarkoittaa, että järjestettäessä vektorin x koordinaatit itseisarvoiltaan laskevaan järjestykseen, x_{arr}, suppenevat ne kohti nollaa kuuluaan sädetä $L > 0$ ja suppenemisparametria $0 < p < 1$ noudattaen seuraavasti:</p> $ x_{\text{arr}}(l) \leq L \cdot l^{-\frac{1}{p}}, \quad \text{kun } 1 \leq l \leq n.$ <p>Tutkielma seuraa Emmanuel Candèsin ja Terence Taon artikkelissaan <i>Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?</i> (2006) antamaa ratkaisua ongelmalle: heikkoon kuuluaan $wl_p(L)$ sisältyvän vektorin $x \in \mathbb{R}^n$ ja mittaamatriisilla $R \in \mathbb{R}^{k \times n}$ tehdyistä mittauksista Rx ℓ_1-optimoinnilla palautetun ratkaisun \hat{x} välinen virhe noudattaa ylärajaa</p> $\ x - \hat{x}\ _{\ell_2} \leq C_{p,\alpha} \cdot L \cdot \left(\frac{k}{\beta}\right)^{\frac{1}{2} - \frac{1}{p}}.$ <p>Ylärajassa $C_{p,\alpha}$ on vakio ja β käytettyyn mittaamatriisiin liittyvä parametri. Tulosta ei siis suoraan johdeta yksittäiselle mittaamatriisityypille, vaan annetaan ominaisuudet, jotka omaavaa matriisia voidaan mittaamatriisina käyttää. Tutkielmassa erikseen todistetaan normaalijakaumasta kehitetyn satunnaismatriisin sopivan mittaamatriisiksi parametrin β arvolla $\log n$. Tällä matriisivalinnalla teoriaa myös esitellään numeerisesti. Esitelty virheen yläraja ja muut ongelmaa koskevat tulokset johdetaan tutkielmassa suoraan lineaarialgebran ja todennäköisyyslaskennan perusteista tuoden aiheen siten helpommin lähestyttäväksi; alkuperäisen artikkelin todistuksissa useat päättelyt on jätetty vaille tarkempia yksityiskohtia.</p>			
Avainsanat — Nyckelord — Keywords Satunnaisprojektiio, korkeaulotteinen data, käänteisongelma, ℓ_1 -optimointi, Compressed sensing.			
Säilytyspaikka — Förvaringsställe — Where deposited Kumpulan tiedekirjasto			
Muita tietoja — Övriga uppgifter — Additional information Tutkielman aiheen tekijä on saanut työskennellessään tutkimusapulaisena Suomen ilmatieteen laitoksella kesällä 2013 harjoittelun ohjaajalta Johan Siléniltä.			

Sisältö

1	Johdanto	7
2	Esitiedot	9
3	Satunnaisprojektiomenetelmä	19
4	Palautusongelman esittely	25
5	Vaatimukset mittausmatriisille	29
6	Ratkaisun vakaus normin ℓ_1 suhteen	33
7	Palautuslause: ratkaisun vakaus normin ℓ_2 suhteen	37
8	Mittausmatriisina normaalin satunnaismatriisi	43
9	Ratkaisun esittely numeerisesti	53
10	Tulosten esittely – Jälkisanat	59
11	Testitiedostot	61

Luku 1

Johdanto

Nykytutkimuksessa käsitellään monissa yhteyksissä kooltaan valtavia aineistoja. Tällöin perinteiset analysointimenetelmät eivät välttämättä ole riittävän tehokkaita ja siksi uusien menetelmien luominen ja käyttöönotto on tarpeen. Henkilökohtaisesti pääsin lähelle juuri tämän kaltaista tutkimusta, kun työskentelin tutkimusapulaisena Suomen ilmatieteen laitoksella kesällä 2013. Tutkimuksen aiheena oli uusien matemaattisten menetelmien soveltaminen ilmastodataan. Erityisesti kiinnostuksen kohteena oli aineiston ulottuvuuksien vähentämiseksi kehitetty satunnaisprojektiomenetelmä.

Satunnaisprojektiota on monissa artikkeleissa tutkittu perinteisemmän ulottuvuuksien vähennysmenetelmän pääkomponenttianalyysin (engl. Principal Component Analysis, PCA) haastajana. Tutkimuksia löytyy esimerkiksi kasvojen tunnistuksen [13], kemometrisen datan [24] sekä kuva- ja tekstidatan [3] parista. Vaikka satunnaisprojektiolla ei usein päästä yhtä tarkkoihin tuloksiin kuin PCA:n kanssa, on uuden menetelmän etuna sen laskennallinen keveys ja huomattavan yksinkertainen toteutus: ulottuvuuksien vähennys voidaan toteuttaa m :stä havainnosta koostuvalle n -ulotteiselle datamatriisille $X \in \mathbb{R}^{n \times m}$ kertomalla sitä esimerkiksi normaali-jakaumasta kehitetyllä matriisilla $R \in \mathbb{R}^{k \times n}$. Siten havaintojen projisoiminen alkuperäisestä ulottuvuudesta sen k -ulotteiselle aliavaruudelle saadaan suoritettua havaintodatasta riippumatta pelkällä matriisikertolaskulla ja otoksen kehittämällä normaali-jakaumasta. Tämä on huomattava ero PCA:han, jossa projektion tekevä matriisi muodostetaan datasidonnoisesti kyseisen datan ominaisvektorit etsimällä.

Satunnaisprojektion kehityksen taustalla on William Johnsonin ja Joram Lindenstraussin vuonna 1982 esittelemä tulos, jonka mukaan korkeaulotteiselle datamatriisille löytyy approksimaalisesti havaintojen väliset etäisyydet säilyttävä kuvaus matalaulotteiselle aliavaruudelle. Aliavaruuden ulottuvuuksien määrä riippuu vain logaritmisesti havaintojen määrästä; alkuperäisestä ulottuvuudesta se ei riipu ollenkaan. Klassisena esimerkkinä sopivasta kuvauksesta käytetään usein juuri normaali-jakaumasta kehitettyä satunnaismatriisia. Sen rinnalle on noussut myös muita matriisityyppejä, kuten Dimitris Achlioptaksen vuosituhaten alussa esittelemä harva matriisi [1], jonka alkioista jopa kaksi kolmasosaa on odotusarvoisesti nollija ja projisoinnissa vaaditun matriisikertolaskunkin sisätulot voidaan palauttaa yhteen- ja vähennyslaskuiksi.

Ilmatieteen laitoksella työskennellessäni satunnaisprojektiomenetelmän parissa tutkimus ei keskittynyt ainoastaa matriisin ulottuvuuksien madaltamiseen ja tästä projisoidusta datasta tehtävien analyysien suorittamiseen. Tutkimuksessa pohdittiin myös datan palauttamista satunnaisprojisoinnilla tehdyistä mittauksista. Tätä ongelmaa on tutkittu erityisesti tiiviyn havainnointiin (engl. compressed sensing) liittyvien käänteisongelmien parissa. Näissä ongelmissa tuntematon, kiinnostuksen kohteena oleva vektori $x \in \mathbb{R}^n$ oletetaan harvaksi (engl. sparse) eli sisältävän vain muutaman nollasta eroavan koordinaatin, ja pyrkimyksenä on rakentaa x hyvin pienestä määrästä siitä tehtyjä satunnaisia mittauksia joko eksaktisti tai edes approksimatiivisesti. Keskeisimmät tulokset aiheesta ovat syntyneet Emmanuel Candèsin ja Terence Taon välisen yhteistyön tuloksena. Esimerkiksi artikkelissaan [7] vuodelta 2006 he esittelevät tuloksen eksaktille palautukselle harvojen vektorien tapauksessa. Lisäksi samana vuonna ilmestyneessä artikkelissa [8] he todistavat approksimaalisen palautuksen onnistumisen, kun mittausten kohde kuuluu heikkoon kuulaan saat- taen siten sisältää muutaman suuremman arvon lisäksi useampia vain vähän nollas- ta eroavia lukuja. Näihin artikkeleihin minut tutustutti harjoitteluni ohjaaja Johan Silén, ja jälkimmäisen artikkelin, *Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?*, olen valinnut Pro Gradu -tutkielmani ai- heeksi.

Tutustuin artikkeliin katsoen sitä satunnaisprojektiomenetelmän näkökulmasta ja pyrkimykseni on tuoda aiheeni esitellyksi samalla tavalla myös lukijalle. Siten ennen varsinaiseen artikkeliin pureutumista esittelen satunnaisprojektiomenetelmää niin teoriassa kuin käytännössä: keskeisiä lauseita tulen havainnollistamaan sekä kuukausittaisia lämpötilahavaintoja sisältävällä ilmastoaineistolla että pääaiheena olevassa artikkelissa esiteltävillä heikon kuulan vektoreilla. Satunnaisprojektiota kä- sittelevään lukuun kuuluu myös lyhyt katsaus aiheen sovelluksiin.

Candèsin ja Taon artikkelia käsittelevissä luvuissa otan haasteeksi esitellä aiheen matalan tason käsitteillä johtaen tulokset suoraan lineaarialgebran ja todennäköi- syyslaskennan perusteista tuoden aiheen siten helpommin lähestyttäväksi; alkuperäi- sen artikkelin todistuksissa useat päättelyt on jätetty vaille tarkempia yksityiskoh- tia. Tutkielman luettavuutta tukemaan olen koonnut laajan esitieto-osuuden, josta esimerkiksi singulaariarvohajotelman, todennäköisyysteorian tai vakion piilottavan O -notaation kanssa mukavuusalueensa rajoilla liikkuva lukija voi tarkistaa keskei- simmät määritelmät. Jotta teorian tulokset eivät jäisi etäisiksi, tulen esittelemään palautuksen toimimista myös numeerisesti.

Luku 2

Esitiedot

Tutkielmassa käsiteltävät aiheet pohjaavat suurimmaksi osaksi lineaarialgebran ja todennäköisyyslaskennan perusteisiin. Tämän luvun tarkoituksena on esitellä nämä perusteet sekä niistä käyttämämme merkintätavat.

Luku on jaettu aihealueiden perusteella kolmeen esitietokokonaisuuteen. Ensimmäisessä kokonaisuudessa, lineaarialgebran esitiedoissa, käydään läpi normiavaruuksien perusteet, singulaariarvohajotelma sekä matriisinormi ja sen yhteys singulaariarvoihin. Näiden tietojen läpikäymistä yhdistää sama päämäärä: ylärajan muodostaminen ℓ_2 -normin muodossa esitetylle virheelle, joka syntyy palauttaessamme korkeaulotteisen vektorin pienestä määrästä lineaarisia mittauksia. Toisin sanoen lineaarialgebran esitietoihin pohjautuu koko tutkielman päälauseen, palautuslauseen 4.5, todistaminen.

Toisessa suuressa kokonaisuudessa esitellään todennäköisyysteorian perusteita aina todennäköisyysavaruuden määritelmästä tapahtumien riippumattomuuteen ja normaalijakauman ominaisuuksiin. Näiden esitietojen läpikäyminen antaa riittävästi työvälineitä, jotta voimme todistaa normaalijakauman otoksesta luodun satunnaismatriisin sopivan palautuslauseen 4.5 mainitsemaksi mittausmatriisiksi.

Viimeisen esitieto-osion aiheena on vakion piilottava O -notaatio. Tässä tutkielmassa notaatio esiintyy esimerkiksi palautuslauseen 4.5 voimassaolotodennäköisyydessä, josta notaatio piilottaa epäoleelliseksi käyvän vakiotermin eksponentiaalisesti pienevän termin rinnalta.

Lineaarialgebra

Lineaarialgebran esitiedot käydään läpi tutkielman keskeisenä lähteenä olevan Taon ja Candèsin artikkelin [8] ehdoilla. Siten käytetyt merkinnät sekä esiteltyjen määritelmien ja lauseiden soveltaminen pohjautuvat kyseiseen artikkeliin. Taustalla olevien matemaattisten määritelmien ja lauseiden sisältö löytyy esimerkiksi lähteistä [28, 22]. Tarkemmin eritellen lähteiden käyttöä, lähdeosa [28] kattaa esitiedot alun normin määritelmästä aina ortogonaalisuuden määritelmään (2.5) asti, ja tästä eteenpäin aiheet löytyvät teoksesta [22].

Määritelmä 2.1. (Normi) Kuvaus $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}_+$ ei-negatiivisille reaalityyppisille reaaliluvuille on normi avaruudessa \mathbb{R}^n , jos seuraavat ehdot ovat voimassa kaikilla vektoreilla $v, w \in \mathbb{R}^n$ ja vakioilla $a \in \mathbb{R}$:

$$(N1) \quad \|v + w\| \leq \|v\| + \|w\|,$$

$$(N2) \quad \|av\| = |a| \|v\|,$$

$$(N3) \quad \|v\| = 0 \Leftrightarrow v = 0.$$

Todetaan, että kuvaukset $\|\cdot\|_{\ell_p} : \mathbb{R}^n \rightarrow \mathbb{R}_+$ määrittelevät normit vektoriavaruuksille \mathbb{R}^n kaikilla $1 \leq p < \infty$, kun

$$(2.2) \quad \|v\|_{\ell_p} = \left(\sum_{k=1}^n |v(k)|^p \right)^{\frac{1}{p}}.$$

Lisäksi, jos $p = \infty$, on normina

$$(2.3) \quad \|v\|_{\ell_\infty} = \max_{k \in \mathbb{Z}_n} |v(k)|,$$

jossa $\mathbb{Z}_n := \{1, \dots, n\}$.

Merkitsemme näitä normiavaruuksia $(\mathbb{R}^n, \|\cdot\|_{\ell_p})$ symboleilla ℓ_p . Erityisesti avaruudessa ℓ_2 pätee vektoreille $v, w \in \ell_2$ tuttu Cauchy-Shwarzin epäyhtälö

$$(2.4) \quad |\langle v, w \rangle| \leq \|v\|_{\ell_2} \|w\|_{\ell_2},$$

jossa

$$\langle v, w \rangle = \sum_{k=1}^n v(k)w(k)$$

on vektorien välinen sisätulo.

Jotta voimme kätevästi laskea vektorin $v \in \mathbb{R}^n$ normin vain jonkun indeksijoukon $I \subset \mathbb{Z}_n$ suhteen, asetamme

$$\|v\|_{\ell_p(I)} := \left(\sum_{k \in I} |v(k)|^p \right)^{\frac{1}{p}}.$$

Tällöin erityisesti pätee $\ell_p = \ell_p(\mathbb{Z}_n)$. Vaihtoehtoisesti käytämme normin laskemiselle indeksijoukon I suhteen kyseisen joukon karakteristista funktiota

$$\mathbf{1}_I(k) = \begin{cases} 1, & \text{kun } k \in I \\ 0, & \text{kun } k \in I^c \end{cases},$$

jolloin saamme $\|v\|_{\ell_p(I)} = \|v\mathbf{1}_I\|_{\ell_p}$. Tässä merkinnässä on hyvä huomioida, että tulovektorin $v\mathbf{1}_I$ alkio on muotoa $v(k)\mathbf{1}_I(k)$ eli kyseessä ei ole vektorien sisätulo.

Karakteristisen funktion lisäksi käytämme merkkivektoria sgn , joka määriteltynä vektorille $v \in \mathbb{R}^n$ on muotoa

$$(\text{sgn}(v))(k) = \begin{cases} 1, & \text{kun } v(k) > 0 \\ 0, & \text{kun } v(k) = 0 \\ -1, & \text{kun } v(k) < 0 \end{cases}.$$

Käsitlemme seuraavaksi vektorien ortogonaalisuutta. Vektorien $v, w \in \mathbb{R}^n$ sanotaan olevan ortogonaaliset eli kohtisuorassa toisiaan vasten, kun niiden sisätulo saa arvon nolla eli

$$(2.5) \quad \langle v, w \rangle = 0.$$

Jos lisäksi ortogonaalisten vektorien normit ovat yksikköjä, kutsutaan vektoreja ortonormaaleiksi. Tällöin siis yhtälön (2.5) lisäksi vektoreille $v, w \in \mathbb{R}^n$ pätee

$$\langle v, v \rangle = \langle w, w \rangle = 1.$$

Kutsumme matriisia $M \in \mathbb{R}^{n \times m}$ ortonormaaliksi, mikäli sen sarakevektorit ovat keskenään ortonormaaleja. Eräs tärkeä esimerkki ortonormaalista neliömatriisista on identiteettimatriisi $I = \text{diag}(1, \dots, 1) \in \mathbb{R}^{n \times n}$, jonka diagonaalilta poikkeavat alkioit ovat nollia ja diagonaali-alkiot yksikköjä.

Näytämme seuraavaksi, että ortonormaali matriisi ei kuvatessaan vektoreita muuta niiden euklidista pituutta eli ℓ_2 -normia. Olkoon $M \in \mathbb{R}^{n \times m}$ ortonormaali matriisi. Tällöin yllä mainituista sisätulon ominaisuuksista ortonormaaleille vektoreille seuraa yhtälö

$$M^T M = I \in \mathbb{R}^{m \times m},$$

jossa käytämme yleistä merkintää M^T matriisin M transpoosille. Olkoon nyt $v \in \mathbb{R}^m$ vektori. Tällöin saamme haluammamme tuloksen tutkimalla ℓ_2 -normin neliötä:

$$(2.6) \quad \|Mv\|_{\ell_2}^2 = \langle Mv, Mv \rangle = v^T M^T M v = v^T v = \langle v, v \rangle = \|v\|_{\ell_2}^2.$$

Siirrymme käsittelemään matriisien ominais- ja singulaariarvoja. Neliömatriisin $N \in \mathbb{R}^{n \times n}$ ominaisarvot $\lambda \in \mathbb{C}$ ja ominaisvektorit $v \in \mathbb{R}^n, v \neq 0$, määräytyvät yhtälöistä

$$(2.7) \quad Nv = \lambda v.$$

Ominaisvektorit ovat siis ne matriisiin N liittyvät vektorit, jotka eivät muuta suuntaa, kun niitä kuvataan kyseisellä matriisilla; vain vektorien euklidinen pituus muuttuu vektoriin liittyvän ominaisarvon mukaisesti.

Otamme annettuna tiedon, että yhteensä ominaisarvoja on $(n \times n)$ matriisilla n kappaletta. Mainittakoon tiedon lukumäärästä seuraavan, kun huomataan ominaisarvojen olevan determinantista $|N - \lambda I|$ määräytyvän n -asteisen polynomin nollakohtia.

Vastaavasti oletamme tunnetuksi matriisin singulaariarvohajotelman olemassaolon.

Määritelmä 2.8. (Singulaariarvohajotelma) Jokainen matriisi $M \in \mathbb{R}^{n \times m}$ voidaan ilmaista muodossa

$$M = U \Sigma V^T,$$

jossa matriisit $U \in \mathbb{R}^{n \times n}$ ja $V \in \mathbb{R}^{m \times m}$ ovat ortonormaaleja neliömatriiseja ja matriisi $\Sigma \in \mathbb{R}^{n \times m}$ on diagonaalimatriisi sisältäen diagonaalillaan laskevassa järjestyksessä positiiviset reaaliset luvut

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0,$$

jossa $p := \min\{n, m\}$. Näitä lukuja $\sigma_1, \dots, \sigma_p$ kutsutaan matriisin M singulaariarvoiksi.

Tutkimme seuraavaksi matriisia $R \in \mathbb{R}^{k \times n}$, $k \leq n$. Oletamme näin matriisin ulottuvuuksista, sillä tilanne tulee olemaan sama myöhemmin käsiteltävillä mittausmatriiseilla. Myöhemmin olemme kiinnostuneita myös matriisista RR^T . Singulaariarvohajotelmaa käyttämällä voimme esittää sen muodossa

$$RR^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma I \Sigma^T U^T = U\Sigma_n^2 U^T := U\Lambda U^T,$$

jossa $\Lambda := \Sigma_n^2$ on $(n \times n)$ diagonaalimatriisi sisältäen diagonaalialkioinaan laskevassa järjestyksessä R :n singulaariarvojen neliöt. Yllä matriisit V ja V^T hävisivät ortonormaaliudesta johtuen.

Kertoessamme tällä matriisilla yhtä matriisin U sarakevektoreista u_i , $1 \leq i \leq n$, saamme tehtyä sisätulon ominaisuuksia hyödyntäen päättelyn

$$(2.9) \quad RR^T u_i = U\Sigma_n^2 U^T u_i = U\Sigma_n^2 e_i = \sigma_i^2 U e_i = \sigma_i^2 u_i,$$

jossa

$$e_i(k) := \begin{cases} 1, & \text{kun } k = i, \\ 0 & \text{muulloin.} \end{cases}.$$

Voimme havaita, että saamamme yhtälö vastaa neliömatriisin RR^T ominaisarvot määräävää yhtälöä (2.7) ja siten matriisin ominaisarvot λ_i ovat R :n singulaariarvojen neliöt σ_i^2 ja niihin liittyvät ominaisvektorit ovat matriisin U sarakevektorit u_i .

Olemme näin todistaneet yhteyden ominaisarvojen ja singulaariarvojen välille. Näytämme vielä singulaariarvojen ja matriisinormin välillä olevan yhteyden esitellen samalla jatkossa tarvittavia arvioita matriisinormille. Aloitamme matriisinormin määritelmästä.

Määritelmä 2.10. Olkoon matriisi M lineaarinen kuvaus $\ell_2(\mathbb{Z}_n) \rightarrow \ell_2(\mathbb{Z}_m)$ joillain $n, m \in \mathbb{N}$. Tällöin kutsumme matriisin normiksi termiä

$$\|M\| := \max_{v \neq 0} \frac{\|Mv\|_{\ell_2}}{\|v\|_{\ell_2}}, \quad \text{jossa } v \in \mathbb{R}^n.$$

On hyvä huomata, että matriisinormi muodostuu matriisilla kuvatun vektorin ja alkuperäisen vektorin pituuden suhteesta. Siten esimerkiksi ortonormaalien matriisien normi on yksikkö, sillä ne eivät muuta kuvauksessa vektorien euklidista pituutta, kuten kohdassa (2.6) huomasimme.

Normin määritelmästä saamme suoraan jatkossa tarvittavan arvion

$$(2.11) \quad \|Mv\|_{\ell_2} \leq \|M\| \|v\|_{\ell_2} \quad \text{kaikilla } M \in \mathbb{R}^{n \times m} \quad \text{ja } v \in \mathbb{R}^n.$$

Saamme kätevästi arvion myös kahden matriisin tulon normille: Olkoon M ja v kuten edellä ja olkoon lisäksi $N \in \mathbb{R}^{k \times n}$. Käyttämällä äskeistä tulosta kahdesti saamme

$$\|NMv\|_{\ell_2} = \|N(Mv)\|_{\ell_2} \leq \|N\| \|Mv\|_{\ell_2} \leq \|N\| \|M\| \|v\|_{\ell_2}.$$

Tämä epäyhtälö pätee kaikilla v , joten jakaessamme sen v :n normilla päädyimme haluamaamme tulokseen

$$(2.12) \quad \|NM\| \leq \|N\| \|M\|.$$

On helppo nähdä, että tätä tulosta voimme ketjuttaa useammastakin matriisista koostuvalle kuvaukselle.

Olemme näin saaneet matriisinormille kaksi hyödyllistä arvioita. Näytämme vielä yhteyden normin ja singulaariarvojen välille. Huomataan ensin, että voimme ilmaista matriisinormin yhtäpitävästi hieman erilaisessa muodossa:

$$\|M\| = \max_{v \neq 0} \frac{\|Mv\|_{\ell_2}}{\|v\|_{\ell_2}} = \max_{v \neq 0} \left\| M \left(\frac{v}{\|v\|_{\ell_2}} \right) \right\|_{\ell_2} = \max_{\|v\|_{\ell_2}=1} \|Mv\|_{\ell_2}.$$

Tämän esityksen valossa diagonaalimatriisin $D = \text{diag}(d_1, \dots, d_n) \in \mathbb{R}^{n \times n}$ normin määrittäminen on helpompaa:

$$\|D\| = \max_{\|v\|_{\ell_2}=1} \|Dv\|_{\ell_2} = \max_{\|v\|_{\ell_2}=1} \sqrt{\sum_{i=1}^n (d_i v_i)^2} = \max_{1 \leq i \leq n} |d_i|.$$

Jatkamme matriisin $R \in \mathbb{R}^{k \times n}$, $k \leq n$, tutkimista tilanteesta, jossa sen kaikki singulaariarvot ovat aidosti positiivisia. Tämä lisäoletus tulee olemaan myöhemmin voimassa suurella todennäköisyydellä.

Koska osaamme laskea diagonaalimatriisin normin ja tiedämme, ettei ortonormaalilla matriisilla kertominen muuta vektorin pituutta, saamme matriisin R normin käyttämällä singulaariarvohajotelmaa:

$$(2.13) \quad \|R\| = \|U\Sigma V^T\| = \|\Sigma\| = \sigma_1(R) = \sqrt{\lambda_1(RR^T)}.$$

Tässä ilmoitamme singulaariarvojen ja ominaisarvojen jälkeen suluissa matriisin, johon ne liittyvät. Yllä olevasta voidaan todeta matriisin R transpoosin normin olevan sama kuin alkuperäisen matriisin, sillä

$$(2.14) \quad \|R^T\| = \|V\Sigma^T U^T\| = \|\Sigma^T\| = \sigma_1(R),$$

jossa ortonormaalien neliömatriisien transpoosit ovat luonnollisesti ortonormaaleja eikä matriisin Σ diagonaalialkiot vaihda paikkaa transponoinnissa.

Vastaavalla päättelyllä kuin matriisin R normin kohdalla, saamme tuloksen

$$(2.15) \quad \|RR^T\| = \lambda_1(RR^T).$$

Tarvitsemme jatkossa myös matriisin RR^T käänteismatriisin normia. Muistamalla, että matriisin M käänteismatriisi on M^{-1} , jos pätee $MM^{-1} = M^{-1}M = I$, ja käyttämällä singulaariarvohajotelmaa voimme kehittää halutun käänteismatriisin:

$$\begin{aligned} (RR^T)(U \text{diag}(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n})U^T) &= U \Lambda U^T U \text{diag}(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n})U^T \\ &= I \\ &= U \text{diag}(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n})U^T U \Lambda U^T \\ &= (U \text{diag}(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n})U^T)(RR^T). \end{aligned}$$

Siten pätee $(RR^T)^{-1} := U \operatorname{diag}(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}) U^T$. Yllä käänteislukujen ottamisesta ei synny ongelmia, sillä oletimme kaikki singulaariarvot aidosti positiivisiksi. Saamamme käänteismatriisin ominaisarvot ovat alkuperäisen matriisin ominaisarvojen käänteisluvut. Siten alkuperäisen matriisin RR^T pienintä ominaisarvoa λ_n vastaava käänteisluku on suurin kaikista ominaisarvojen käänteisluvuista. Kohdan (2.13) kaltaisella päättelyllä saamme nyt käänteismatriisin $(RR^T)^{-1}$ normiksi

$$(2.16) \quad \|(RR^T)^{-1}\| = \frac{1}{\lambda_n(RR^T)}.$$

Olemme tämän tuloksen myötä käyneet läpi kaikki lineaarialgebraa koskevat esitiedot.

Todennäköisyysteoria

Ainakin suurin osa tämän osion tuloksista löytynee useimmista todennäköisysteorian alkeita käsittelevistä kirjoista. Oma valintani koko osion lähteeksi on tuore ja laaja johdantoteos [27] todennäköisysteoriaan ja sen sovelluksiin.

Käytämme todennäköisyysavaruudesta merkintää $(\Omega, \mathcal{F}, \mathbb{P})$, jossa Ω on varsinainen todennäköisyysavaruus, \mathcal{F} on σ -algebran muodostava todennäköisyysavaruuden osajoukoista koostuva perhe, ja \mathbb{P} on todennäköisyysmitta. Määrittelemme vain viimeisimmän, sillä se kertoo meille edeltävistä riittävästi.

Määritelmä 2.17. *Todennäköisyysmitta* on kuvaus $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$, joka määrää jokaiselle tapahtumaperheen alkion $A \in \mathcal{F}$ arvon seuraavia aksioomia noudattaen:

(P1) *Positiivisuus:* $\mathbb{P}(A) \geq 0$ kaikilla $A \in \mathcal{F}$.

(P2) *Normalisointi:* $\mathbb{P}(\Omega) = 1$.

(P3) *Additiivisuus:* Jos tapahtumat $A, B \in \mathcal{F}$ ovat toisensa poissulkevia, eli $A \cap B = \emptyset$, niin $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

(P4) *Jatkuvuus:* Jos A_1, A_2, \dots on monotonisesti laskeva tapahtumajono, eli $A_k \supseteq A_{k+1}$ jokaisella k , ja pätee $\bigcap_{k \geq 1} A_k = \emptyset$, niin $\mathbb{P}(A_k) \rightarrow 0$, kun $k \rightarrow \infty$.

Huomautetaan heti, että mitan monotonisuus seuraa aksioomista välittömästi, sillä jos $A \subset B$, niin

$$(2.18) \quad \mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

Tästä seuraa edelleen, että jokaisen tapahtuman $A \subset \Omega$ todennäköisyys on rajoitettu sekä ylhäältä, että alhaalta:

$$1 = \mathbb{P}(\Omega) \geq \mathbb{P}(A) \geq \mathbb{P}(\emptyset) = 0.$$

Saamme siten todennäköisyyden myös tapahtuman $A \in \mathcal{F}$ komplementtitapahtumalle $A^c := \Omega \setminus A \in \mathcal{F}$:

$$(2.19) \quad 1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) \quad \text{eli} \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

Edelleen, saamme tapahtumajonolle A_1, A_2, \dots pareittain eroavia tapahtumia todennäköisyyden intuitiivisesti odotetun, muttei niin triviaalisti seuraavan identiteetin

$$(2.20) \quad \mathbb{P} \left(\bigcup_{k=1}^{\infty} A_k \right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

Vaikka tapahtumat $A, B \subset \Omega$ eivät olisi toisensa poissulkevia, saamme silti monotonisuuden nojalla arvion

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cup (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \leq \mathbb{P}(A) + \mathbb{P}(B).$$

Yhdistämällä tämän edelliseen saamme myöhemmin paljon käytetyn ylärajan mielivaltaisten tapahtumien A_1, A_2, \dots, A_n yhdisteen todennäköisyydelle:

$$(2.21) \quad \mathbb{P} \left(\bigcup_{k=1}^n A_k \right) \leq \sum_{k=1}^n \mathbb{P}(A_k).$$

Ositetaan seuraavaksi koko todennäköisyysavaruus numeroituvaan joukkoon toisensa poissulkevia tapahtumia eli olkoon $\bigcup_{k \geq 1} A_k = \Omega$, kun $A_i \cap A_j = \emptyset$ kaikilla $i \neq j$. Nyt jokaiselle tapahtumalle $B \in \mathcal{F}$ pätee identiteetin (2.20) nojalla

$$\mathbb{P}(B) = \mathbb{P}(B \cap \Omega) = \sum_{k \geq 1} \mathbb{P}(A_k \cap B)$$

Toisena määritelmänä esittelen ehdollisen todennäköisyyden.

Määritelmä 2.22. (Ehdollinen todennäköisyys) Olkoon A ja B todennäköisyysavaruuden tapahtumia, joista ainakin B :n todennäköisyys eroaa nolasta. Tällöin A :n todennäköisyys ehdolla, että B on tapahtunut määritellään todennäköisyytenä

$$(2.23) \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Huomautus 2.24. Määritelmän nojalla ehdolliset todennäköisyydet eivät noudata enää todennäköisyysmitan \mathbb{P} jakaumaa, vaan yllä määritelmän taustalla on uusi todennäköisyysmitta $\mathbb{P}_B : A \rightarrow \mathbb{P}(A|B)$, joka keskittää kaiken todennäköisyysmassan tapahtumaan B .

Määrittelemme vielä tapahtumien (pareittaisen) riippumattomuuden.

Määritelmä 2.25. (Riippumattomuus) Tapahtumien $A, B \in \mathcal{F}$ sanotaan olevan toisistaan riippumattomia, jos pätee ehto

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B).$$

Käsittelemme seuraavaksi lyhyesti satunnaismuuttujia. Menemättä tarkemmin mittateorian yksityiskohtiin todettakoon, että satunnaismuuttuja on johonkin todennäköisyysavaruuteen $(\Omega, \mathcal{F}, \mathbb{P})$ liittyvä reaaliarvoinen \mathcal{F} -mitallinen kuvaus $X : \Omega \rightarrow \mathbb{R}$.

Voimme valita käyttämäksemme σ -algebraksi \mathcal{F} Borel-joukkojen perheen, jolloin käytössämme oleviin todennäköisyystapahtumiin kuuluvat muun muassa kaikki avoimet, puoliavoimet ja suljetut reaalilukusuoran välit. Esimerkiksi puoliavoimen välin tapauksessa tapahtumat ovat muotoa $\{\omega \in \Omega : a \leq X(\omega) < b\} = X^{-1}[a, b)$, jossa $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ ja $a \leq b$.

Aiemmin esittelemämme todennäköisyysteorian määritelmät soveltuvat helposti myös satunnaismuuttujien avulla esitellyiksi. Esimerkiksi satunnaismuuttujat X ja Y ovat riippumattomat täsmälleen silloin, kun pätee

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B),$$

kaikilla Borel-joukoilla A, B .

Konkreettisenä esimerkkinä satunnaismuuttujasta esittelen standardinormaalijakaumaa noudattavan satunnaismuuttujan, joka tulee olemaan käytössä muun muassa satunnaisprojektioita luotaessa.

Määritelmä 2.26. (Standardinormaalijakauma): Satunnaismuuttujan Z sanotaan noudattavan standardinormaalijakaumaa, jos pätee

$$\mathbb{P}(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt, \quad \text{kun } z \in \mathbb{R}.$$

Tällöin merkitään $Z \sim N(0, 1)$.

Huomautus 2.27. Käytämme symbolia Z jatkossa vain standardinormaalijakaumaa noudattavalle satunnaismuuttujalle.

Tunnetusti tämän jakauman odotusarvona on 0 ja keskihajonnan neliönä eli varianssina on 1. En mene näiden jakaumakohtaisten tunnuslukujen syvällisempään määrittelyyn, sillä korkealla tasolla niiden nimet ovat riittävän itsensäselittäviä.

Standardinormaalijakaumaa noudattavasta satunnaismuuttujasta $Z \sim N(0, 1)$ saamme yleistä normaalijakaumaa $N(\mu, \sigma^2)$ odotusarvolla $\mu \in \mathbb{R}$ ja varianssilla $\sigma^2, \sigma > 0$, noudattavan satunnaismuuttujan X asettamalla

$$X := \frac{Z - \mu}{\sigma}.$$

Yleisemmin riippumattomien normaalijakaumaa noudattavien satunnaismuuttujien summamuuttuja noudattaa seuraavan lauseen antamaa jakaumaa.

Lause 2.28. *Olkoon X_1, \dots, X_n riippumattomia satunnaismuuttujia. Oletetaan jostaista indeksistä k vastaavan satunnaismuuttujan X_k noudattavan jakaumaa $N(\mu_k, \sigma_k^2)$. Tällöin*

$$(2.29) \quad \sum_{k=1}^n X_k \sim N\left(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2\right).$$

Tarvitsemme vielä todennäköisyyslärajan standardinormaalijakauman häntätodennäköisyyksille.

Lemma 2.30. *Olkoon Z standardinormaalijakaumaa noudattava satunnaismuuttuja. Tällöin*

$$\mathbb{P}(Z \leq z) \leq \frac{1}{z\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \text{jossa } z \leq 0.$$

Viimeisenä todennäköisyyslaskentaa sivuavana aiheena käsittelemme binomikerrointa

$$(2.31) \quad \binom{n}{m} := \begin{cases} \frac{n!}{m!(n-m)!}, & \text{kun } m \geq 0, \\ 0, & \text{kun } m < 0 \end{cases}.$$

Binomikerroin antaa tunnetusti niiden tapojen määrän, joilla m eri alkia voidaan valita n alkion joukosta, kun järjestyksellä ei ole merkitystä.

Arvioidessamme myöhemmin erään joukon alkioiden määrää tulemme tarvitsemaan seuraavaa ylärajaa:

$$(2.32) \quad \binom{n}{m} \leq e^{m+m \log(n/m)}.$$

Esitetään lyhyt perustelu ylärajalle. Havainnosta

$$1 \geq \binom{n}{m} \left(\frac{m}{n}\right)^m \left(1 - \frac{m}{n}\right)^{n-m} = \binom{n}{m} e^{-nH(m/n)},$$

jossa $H(q) = -q \log q - (1-q) \log(1-q)$, $0 \leq q \leq 1$, on e kantainen binäärientropiafunktio, seuraa epäyhtälö

$$(2.33) \quad \binom{n}{m} \leq e^{nH(m/n)}.$$

Tämä yläraja on poikkeuksellisesti lähteestä [2].

Määritellään apufunktio $h(q) = q + (1-q) \log(1-q)$. Funktion derivaataksi saamme $h'(q) = -\log(1-q)$. Koska derivaatta on positiivinen välillä $(0, 1)$ ja $h(0) = 0$, on myös h positiivinen välillä $(0, 1)$. Tästä seuraa epäyhtälö

$$-(1-q) \log(1-q) \leq q,$$

jota käyttämällä epäyhtälöön (2.33) saamme edelleen haluamamme arvion (2.32).

Asymptoottinen notaatio $O(\cdot)$

Seuraava määritelmä vakion piilottavalle O -notaatiolle on lähteestä [21].

Määritelmä 2.34. Olkoon $f, g : \mathbb{N} \rightarrow \mathbb{R}_+$ funktioita luonnollisilta luvuilta reaaliluvuille. Sanomme g :n olevan funktion f asymptoottinen yläraja ja merkitsemme tällöin $f(n) = O(g(n))$, jos on olemassa positiiviset kokonaisluvut C ja n_0 siten, että jokaisella $n \geq n_0$ pätee

$$f(n) \leq Cg(n).$$

Määritelmästä voi huomata, että yhtäsuuruusmerkin käyttö notaatiossa on jokseenkin kyseenalaista, sillä merkinnässä vasemmalla on yksittäinen funktio ja oikealla joukko funktioita. Merkintätapa on kuitenkin yleisesti käytössä.

Kuten mainittu, $O(g(n))$ samaistaa funktiot, jotka ovat vakiotekijää vaille pienempiä tai yhtäsuuria kuin g jostain n lähtien. Seuraava esimerkki on lähteestä [21]:

$$(5n^3 + 2n^2 + 22n + 6) = O(n^3).$$

Jatkossa tulemme käyttämään merkintää todennäköisyyksien yhteydessä. Tarkemmin ottaen käsittelemme tapahtumia, jotka ovat voimassa vähintään todennäköisyydellä

$$1 - O_\alpha(n^{-\rho/\alpha}),$$

jossa $\alpha, \rho > 0$ ovat vakioita. Poiketen lähteen [8] merkinnöistä ilmoitamme notaation O alaindeksissä vakion α , millä haluamme tähdentää, että annamme merkintään liittyvän vakion riippua α :sta.

Luku 3

Satunnaisprojektiomenetelmä

Oletetaan, että olemme kiinnostuneet datamatriisista $X = [x_1 \cdots x_m] \in \mathbb{R}^{n \times m}$, joka koostuu m :stä kappaleesta korkeaulotteisia vektoreja $x_i \in \mathbb{R}^n, 1 \leq i \leq m$. Oletetaan lisäksi, että haluaisimme analysoida dataa suorittamalla sille jonkin toimenpiteen tekävän algoritmin, jonka suoritus-aika riippuu datan ulottuvuuksien määrästä n .

Tämän kaltaisessa asetelmassa datan korkeaulotteisuudesta johtuen saatetaan usein joutua tilanteeseen, jossa algoritmin suoritus-aika kasvaa niin suureksi, ettei sitä ole enää mielekäästi suorittaa. Esimerkiksi monissa käytännönläheisissä ongelmissa parhaan tunnetun algoritmin suoritus-aika riippuu jopa eksponentiaalisesti ulottuvuuksien määrästä [1]. Eräs vaihtoehto alkuperäisessä ulottuvuudessa tapahtuvan laskennan sijaan on projisoida data alkuperäiseltä n -ulotteiselta avaruudelta sen k -ulotteiselle aliavaruudelle ja hakea tästä projisoidusta datasta approksimoiva ratkaisu alkuperäiselle ongelmalle.

Satunnaisprojektiomenetelmässä ideana on projisoida korkeaulotteinen data X täysin satunnaisista suunnista virittyvälle matalaulotteiselle aliavaruudelle [26]. Aluksi ajatus saattaa kuulostaa erikoiselta, mutta vertailemme tuloksia yleisesti käytetyn ulottuvuuksien vähentämismenetelmän kanssa saamme oikeutuksen idealle. Aloitamme määrittelemällä normaalin satunnaismatriisin käyttäen lähteen [8] esitystä.

Määritelmä 3.1. (Normaalinen satunnaismatriisi) Olkoon $R \in \mathbb{R}^{k \times n}$ matriisi, kun $1 \leq k \leq n$, ja olkoon sen alkiot $R(i, j)$ muotoa $(1/\sqrt{n})Z_{ij}$, jossa Z_{ij} :t ovat toisistaan riippumattomia standardinormaalijakaumaa noudattavia satunnaismuuttujia. Kutsumme tällöin matriisia R normaaliseksi satunnaismatriisiksi.

Näin määriteltynä kaikki mahdolliset suunnat avaruudessa \mathbb{R}^n ovat R :n rivivektoreille yhtä todennäköisiä [1] ja siten on perusteltua sanoa, että data X kuvautuu täysin satunnaisista suunnista viritetylle aliavaruudelle kuvauksessa RX .

Käyttäessämme ulottuvuuksia vähentävää menetelmää dataan useimmiten jotain tietoa häviää menetelmän soveltamisessa. Onkin tapauskohtaista, minkä tiedon haluamme säilyttää ulottuvuuksia vähennettäessä. Normaalin satunnaismatriisin vahvuus on havaintojen välisten euklidisten etäisyyksien säilyttäminen hyvällä todennäköisyydellä. Lähteiden [1, 26] tuloksista muotoiltu lause esittelee tätä normaalin satunnaismatriisin ominaisuutta teoriassa:

Lause 3.2. *Olkoon $x_i, x_j \in \mathbb{R}^n$ vektoreita ja $R \in \mathbb{R}^{k \times n}$ määritelmän 3.1 mukainen normaalin satunnaismatriisi. Tällöin jokaiselle $\epsilon > 0$ pätee*

$$\mathbb{P} \left((1 - \epsilon) \|x_i - x_j\|_{\ell_2}^2 \leq \frac{n}{k} \|Rx_i - Rx_j\|_{\ell_2}^2 \leq (1 + \epsilon) \|x_i - x_j\|_{\ell_2}^2 \right) \geq 1 - 2e^{-\left(\frac{\epsilon^2}{2} - \frac{\epsilon^3}{3}\right)k}.$$

Huomautus 3.3. Lauseessa termi n/k skaalaa etäisyydet toisiaan vastaaviksi: projisoidun yksikön pituisen vektorin $u \in \mathbb{R}^n$ odotusarvoinen pituus projektiossa Ru on $\sqrt{k/n}$ [3].

Havainnollistaaksemme normaalin satunnaismatriisin etäisyydet säilyttävää ominaisuutta laskemme sen kahteen eri testidataan aiheuttaman virheen ja vertaamme tuloksia toisen ulottuvuuksienvähentämimenetelmän, pääkomponenttianalyysin (engl. Principal Component Analysis, PCA), antamien tulosten kanssa. PCA:n lyhyt esittely on ensin tarpeen.

PCA on eräs suosituimmista menetelmistä ulottuvuuksien vähentämiseen. Siinä alkuperäiset n -ulotteiset havainnot kuvataan matriisin XX^T ensimmäisten k :n ominaisvektorin $U_k = [u_1 \ u_2 \ \dots \ u_k]$ osoittamaan suuntaan:

$$X_{\text{PCA}} = U_k^T X.$$

Näitä ominaisvektoreja vastaavien ominaisarvojen

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k (\geq \lambda_{k+1} \geq \dots \geq \lambda_n \geq 0)$$

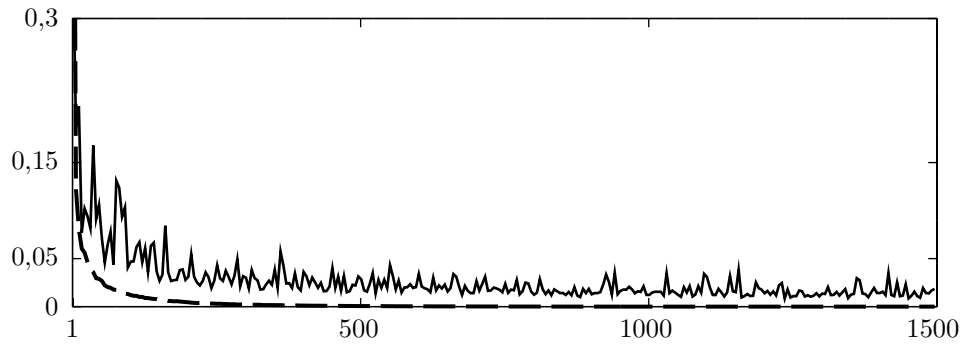
suuruus kuvaa datan vaihtelun suuruutta ominaisvektorin osoittamassa suunnassa. Mikäli siis ensimmäisten k :n ominaisvektorin suunnat selittävät suurimman osan datan vaihtelusta, voidaan usein loput suunnat jättää pois kärsien vain pienestä informaatiokadosta. Tarkemmin ottaen kuvasimme juuri niin kutsutun ”keskittämättömän” PCA:n, sillä usein ennen ominaisvektorien etsimistä datan sarakkeista poistetaan niiden keskiarvot. [11, 5]

PCA:n ja normaalin satunnaismatriisin vertailua varten käytetään ensimmäisenä testidatana kuukausittaisia pintalämpötiloja ympäri maapalloa [15]. Lähteen koko datan joukosta on valittu $m := 1000$ perättäistä kuukautta testidatamme sarakkeiksi, ja ulottuvuuksien vähennys kohdistuu sijainteihin, $n := 4608$. Kuvassa 3.1 esitetään 1000 satunnaisesta havaintoparista (x_i, x_j) lasketun suhteellisen virheen

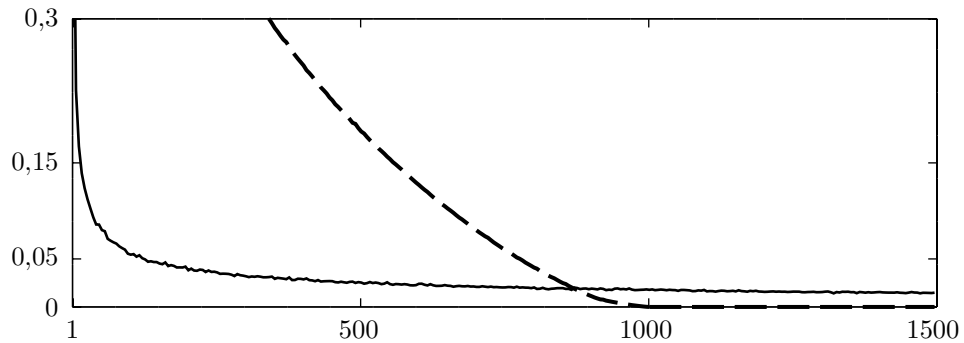
$$\left| \frac{\|x_i - x_j\|_{\ell_2} - \|Rx_i - Rx_j\|_{\ell_2}}{\|x_i - x_j\|_{\ell_2}} \right|$$

keskiarvo projektioulottuvuuden k funktiona sekä normaalin satunnaismatriisin että PCA:n tapauksessa. Satunnaismatriisin etäisyyksiä laskettaessa käytetään lisäksi lauseen 3.2 esittelemää skaalaustekijää $\sqrt{n/k}$. Tarkempi testin toteutus löytyy luvusta 11.

Kun dataksi vaihdetaan $m := 1000$ kappaletta seuraavissa luvuissa käsiteltäviä heikon kuulan $wl_p(100), p := 0,8, n := 4608$ ulotteisia vektoreita, saadaan kuvan 3.2 mukaiset tulokset. Heikon kuulan määritelmän numero on 4.3, ja testin tarkempi toteutus löytyy luvusta 11.



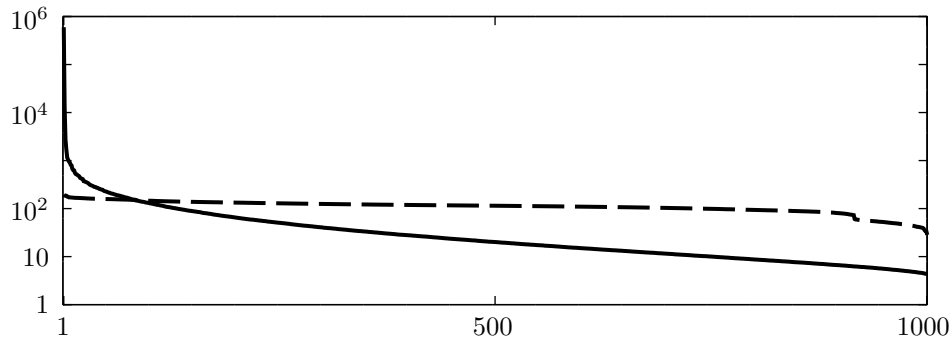
Kuva 3.1: Keskimääräinen suhteellinen virhe lämpötiladatan tapauksessa projisointiulottuvuuden k funktiona; normaalin satunnaismatriisi (—), PCA (---).



Kuva 3.2: Keskimääräinen suhteellinen virhe heikon kuulan vektoreiden tapauksessa projisointiulottuvuuden k funktiona; normaalin satunnaismatriisi (—), PCA (---).

Välittömästi on todettava, että PCA:lla päädyimme molemmilla datoilla riittävän suurilla k :n arvoilla huomattavasti pienempiin virheisiin kuin normaalilla satunnaismatriisilla: ilmastodatan tilanteessa keskimääräinen suhteellinen virhe on alle prosentin kaikilla $k \geq 126$ ja molemmilla datoilla virhetermit ovat nolliä, kun kaikkia positiivisia ominaisarvoja vastaavia ominaisvektoreja käytetään aliavaruuden virittämisessä eli $k \geq 1000$. Toisaalta, jos emme ole kiinnostuneita erittäin tarkoista arvioista, saamme satunnaismatriisilla heikonkuulan vektoreiden datalla jo paljon pienemmällä k huomattavasti parempia tuloksia kuin PCA:lla.

Nämä normaalin satunnaismatriisin antamat paremmat tulokset pienillä k heikon kuulan vektoreiden tapauksessa eivät johtuneet satunnaismatriisin soveltuvuudesta nimenomaan tähän datasettiin, saimmehan varsin saman kaltaisia tuloksia ilmastodatankin kohdalla. Taustalla on itse asiassa lause 3.2, joka takaa etäisyyden neliön suhteellisen virheen säilymiselle todennäköisyysalarajan mielivaltaisella vektoriparilla. Suhteelliset virheet säilyvät siis vähintään lauseen takaamalla todennäköisyydellä minkä tahansa datasetin vektoriparilla!



Kuva 3.3: Ilmastodatan (—) ja heikonkuulan vektorien muodostaman datan (---) ensimmäiset tuhat ominaisarvoa.

Tämä on huomattava etu suhteessa PCA:han, jonka suorituskyky poikkei rajusti datasettien välillä. PCA:n käyttäytymistä voimme selittää datasettien ominaisarvojen jakautumisten erilaisuudella: Kuten kuvasta 3.3 näemme, ilmastodatan kohdalla vain muutama ominaisarvo selittää suurimman osan datan vaihtelusta, jolloin näitä muutamaa ominaisarvoa vastaavien ominaisvektorien mukanaolo projektiomatriisissa riittää antamaan hyviä tuloksia. Sen sijaan heikon kuulan vektorien datasetin ominaisarvojen massa on jakautunut tasaisemmin ja siksi useampia ominaisvektoreja tarvitaan hyvien tulosten saavuttamiseksi.

Ulottuvuusienvähentämismenetelemiä vertaillessa todetun vääristymän lisäksi keskeinen kriteeri on menetelmän suorittamiseen kuluva aika. Esitettyjen menetelmien kohdalla normaalin satunnaismatriisi on huomattavasti kevyempi: Projektio-matriisin luominen vaatii vain lukujen generoimista standardinormaalijakaumasta ja yhden matriisien välisen kertolaskun kuuluen siten aikavaativuusluokkaan $O(knm)$. Sen sijaan PCA:n aikavaativuus on luokkaa $O(n^2m) + O(n^3)$. Näitä aikavaativuuksia ei tule pitää kuin neuvoa-antavina: esimerkiksi PCA:n toteuttamista voidaan nopeuttaa laskemalla koko singulaariarvohajotelman sijasta vain tarvittavat ominaisvektorit. Menetelmien laskennallisen keveyden vertailua löytyy tarkemmin artikkelista [3], jossa testidatana käytetään kuva- ja tekstidataa. [3]

Yllä asetimme PCA:n ja satunnaisprojektion vastakkain vertaillessamme niiden ominaisuuksia datan ulottuvuuksien vähentämiseen. PCA:ta käytetään myös datan muuttujien välisten suhteiden selvittämiseen [11] ja siksi menetelmien yhtäaikaisen käytön toimivuus on mielenkiintoinen kysymys: jos PCA:n suorittaminen suuren tietomäärän tilanteessa on liian raskas toimenpide, olisiko mahdollista approksimoida toimenpidettä laskemalla ulottuvuuksien määrää ensin satunnaisprojektiolla ja suorittamalla sitten PCA projisoidulle datalle? Tätä kysymystä pohdittiin ilmatieteen laitoksella työskennellessäni siellä tutkimusapulaisena kesällä 2013 ja lupaavista empiirisistä tuloksista johtuen aiheesta ilmestyy artikkeli [20].

Olemme luvussa käsitelleet satunnaisprojektion hyödyntämistä matriisiin, mutta lause 3.2 takaa etäisyyksien säilymisen vain yksittäisille vektoripareille. Seuraavaksi esiteltävä satunnaisprojektiomenetelmän teorian kannalta keskeinen tulos sen sijaan

koskettaa vektorijoukon kaikkien pareittaisten etäisyyksien samanaikaista säilymistä. Tulos on nimetty todistajiensa mukaan ja alkuperäisen esityksen, kirjassa [14], sijaan seuraavassa lemma on muotoiltu artikkelia [1] mukaillen:

Lemma 3.4. (Johnson-Lindenstrauss, 1982) *Olkoon $\epsilon > 0$, ja $X \in \mathbb{R}^{n \times m}$ joukko n -ulotteisen avaruuden vektoreja $x_i \in \mathbb{R}^n, i \in \mathbb{Z}_m$. Valitaan jokin kokonaisluku k noudattaen alarajaa $k \geq k_0 = O(\epsilon^{-2} \log m)$. Tällöin on olemassa sellainen kuvaus $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$, jolle kaikilla $i, j \in \mathbb{Z}_m$ pätee*

$$(1 - \epsilon) \|x_i - x_j\|_{\ell_2}^2 \leq \|f(x_i) - f(x_j)\|_{\ell_2}^2 \leq (1 + \epsilon) \|x_i - x_j\|_{\ell_2}^2.$$

Lemma takaa, että annetulle korkeulotteiselle vektorijoukolle löytyy pareittaiset etäisyydet mielivaltaisen tarkasti säilyttävä kuvaus aliavaruudelle, jonka ulottuvuuksien määrä on vain logaritminen vektorien määrään nähden eikä se riipu lainkaan alkuperäisestä ulottuvuuksien määrästä. Väitteen todistuksesta kiinnostuneiden lukijoiden kannattaa tutustua artikkeliin [10], jossa alkuperäistä todistusta on huomattavasti yksinkertaistettu samalla vaadittavaa ulottuvuuksien määrää k alentaen.

Varsin yksinkertaisilla laskuilla lausetta 3.2 hyödyntäen pystytään näyttämään, että normaalin satunnaismatriisi sopii lemmän 3.4 kuvaukseksi vähintään todennäköisyydellä $1 - m^{-\beta}$, kun $k_0 := (4 + 2\beta)(\epsilon^2/2 - \epsilon^3/3)^{-1} \log m$ [1]. Tässä $\beta > 0$ on vapaasti valittavissa. Ollessamme siis kiinnostuneita etäisyyksien säilymisestä voimme pitää lemmän 3.4 kaltaisen tuloksen toteuttamista sopivana ehtona satunnaisprojisointiin käytettävälle matriisille. Uusiksi esimerkeiksi satunnaisprojisointiin soveltuvista matriiseista saamme jakaumien

$$R(i, j) = \begin{cases} +1, & \text{todennäköisyydellä } 1/2 \\ -1, & \text{todennäköisyydellä } 1/2 \end{cases},$$

$$R(i, j) = \sqrt{3} \cdot \begin{cases} +1, & \text{todennäköisyydellä } 1/6 \\ 0, & \text{todennäköisyydellä } 4/6 \\ -1, & \text{todennäköisyydellä } 1/6 \end{cases},$$

riippumattomista otoksista kehitetyt matriisit R , sillä artikkelissa [1] todistetaan näiden matriisien noudattavan lemmää 3.4 samoin todennäköisyysrajoin kuin normaalin satunnaismatriisi.

Samassa artikkelissa kirjoittaja esittelee näitä jakaumia käytettäväksi relaatiotietokantojen hakujen toteuttamiseen. Esimerkiksi jälkimmäistä jakaumaa hyödynnetäessä matriisikertolaskun vaatimien pistetulojen laskemiseen tarvitaan vain yhteen- sekä vähennyslaskuja ja koko pistetulojen termeistä voi heti alkuksi jättää pois kaksi kolmasosaa! Tämä on huomattava etu suhteessa normaalin satunnaismatriisiin kaltaisiin täysiin reaalmatriiseihin, joiden käyttö tietokantojen hauissa osoittautuu monesti hankalaksi. [1]

Jo mainittujen tilanteiden lisäksi satunnaisprojektiota on tutkittu muun muassa kemometrisen datan [24], kasvojentunnistuksen [13], kolmiulotteisten esineiden tunnistuksen [17], ja puhetta sisältävistä äänidokumenteista koostuvien suurien tietokantojen analysoinnin [16] yhteydessä. Näistä viimeisessä tutkimuksessa vertailtiin satunnaisprojektiota PCA:n kanssa ja eduksi huomioitiin projektion tekevän matriisin yksinkertaisempi laajentaminen tietokannan kasvaessa.

Satunnaisprojektiomenetelmää käsitellään myös kirjassa [26], jossa yleisen teoria-pohjaisen esittelyn lisäksi käsitellään menetelmän soveltamista useisiin tietojenkäsittelytieteen klassisiin ongelmiin. Näistä ongelmista monien suorat ratkaisut riippuvat nykytietämyksen mukaan eksponentiaalisesti datan ulottuvuuksien määrästä, mutta käyttämällä satunnaisprojisointia sopivassa tilanteessa saadaan approksimoivia jopa polynomisessa ajassa toimivia ratkaisualgoritmeja. Sovelluksia ovat esimerkiksi verkon jakaminen k :hon osaan siten, että osien välille jää maksimaalinen määrä kaaria; lyhimmän etäisyyden laskeminen annetun pistejoukon ja pisteen välille; nopean matala-asteisen approksimaation laskeminen reaalmatriisille.

Siirymme seuraavassa luvussa käsittelemään tutkielman pääaihetta, joka on luonteeltaan varsin erilainen suhteessa tässä luvussa mainittuihin esimerkkeihin: tarkoituksemme on nimittäin approksimaalisesti palauttaa koko alkuperäinen datamatriisi satunnaismatriisilla projisoidusta datasta.

Luku 4

Palautusongelman esittely

Tutkielman pääaiheena oleva sovellus satunnaisprojektiolle esitellään tässä luvussa. Sovellus on läheisessä yhteydessä tiiviyn havainnoimiseen (engl. *compressed sensing*, CS) liittyvien käänteisongelmien kanssa ja siksi katson sopivaksi esitellä niitä ennen varsinaista ongelmaa. Yleinen piirre näissä CS-ongelmissa on etsiä niin kutsutusti ”harvoja” ratkaisuja. Harvuudella tarkoitetaan, että ratkaisussa on vain vähän nollasta eroavia arvoja. Esimerkiksi vektoria $x \in \mathbb{R}^n$ kutsutaan s -harvaksi, kun siinä on korkeintaan $s < n$ nollasta eroavaa koordinaattia.

Tyypillinen CS-ongelma on seuraavanlainen: olemme kiinnostuneet harvaksi oletetusta vektorista $x \in \mathbb{R}^n$ ja pyrkimyksemme on löytää se, kun tiedämme siitä satunnaisilla lineaarisilla mittauksilla tuotetut havainnot $y \in \mathbb{R}^k$ ja nämä havainnot tuottaneen matriisin $R \in \mathbb{R}^{k \times n}$. Meidän tulee siten ratkaista x yhtälöstä

$$(4.1) \quad y = Rx,$$

kun tiedämme y :n ja R :n.

Tällaisella yhtälöllä voi olla tunnetusti ääretönkin määrä ratkaisuja. Siksi etsimämme vektorin x löytäminen ei ole itsestään selvää. Klassinen keino on valita ratkaisujen joukosta sellainen vektori, joka minimoi jonkin normin; esimerkiksi euklidista pituutta vastaavan normin ℓ_2 .

Artikkelissa [4] havainnollistetaan, että etsiessämme harvaa ratkaisua ei ℓ_2 -normin minimoiminen tuota lähes koskaan hyvää tulosta. Sen sijaan minimoimalla ℓ_1 -normin eli löytämällä jonkin optimointiongelman

$$(4.2) \quad \min_{z \in \mathbb{R}^n} \|z\|_{\ell_1} \quad \text{ehdolla} \quad Rz = y$$

ratkaisuihin $\hat{x} \in \mathbb{R}^n$ saamme monissa tapauksissa harvoja ratkaisuja. Mainitussa artikkelissa esitetään jopa tuloksia, joiden mukaan riittävän harvoja vektoreita voidaan suurella todennäköisyydellä palauttaa eksaktisti pienestäkin määrästä satunnaisia lineaarisia mittauksia ℓ_1 -minimointia käyttämällä. Vastaavanlaisia tuloksia löytyy lähteestä [7].

Siirrymme tarkastelemaan tutkielman varsinaista ongelmaa. Poikkeuksena kohdan (4.1) yhteydessä määritellyyn ongelmaan on, ettemme oleta vektoria x yllä kuvatussa mielessä harvaksi. Voisi sanoa, että oletamme sen ennemminkin sisältävän

vain pienen määrän selvästi nolasta poikkeavia arvoja. Tarkemmin ottaen oletamme vektorin kuuluvan heikkoon L säteiseen ℓ_p -kuulaan:

Määritelmä 4.3. (Heikko kuula $wl_p(L)$) Olkoon $x \in \mathbb{R}^n$ vektori. Merkitään $x_{\text{arr}} \in \mathbb{R}^n$ vektoria, johon x :n alkiot on järjestetty itseisarvoiltaan laskevaan suuruusjärjestykseen.

Tällöin sanotaan, että x kuuluu heikkoon $L > 0$ säteiseen ℓ_p -kuulaan parametrilla $0 < p < 1$ eli $x \in wl_p(L)$, jos x_{arr} :n alkiot jokaisella $1 \leq l \leq n$ toteuttavat ehdon

$$(4.4) \quad |x_{\text{arr}}(l)| \leq L \cdot l^{-\frac{1}{p}}.$$

Oletamme seuraavaksi, että $x \in \mathbb{R}^n$ kuuluu heikkoon kuulaan $wl_p(L)$, ja $S \subset \mathbb{Z}_n$ on indeksijoukko, joka sisältää x :n s itseisarvoiltaan suurinta koordinaattia vastaavat indeksit. Tällöin $x\mathbb{1}_S$ on s -harva vektori. Voimme lisäksi valita s :n niin suureksi, että joukon $wl_p(L)$ määritelmän mukaisesti kaikki indeksejä S^c vastaavat x :n koordinaattien arvot ovat hyvin lähellä nollaa. Tästä seuraa, että joukossa S^c vektorin x ℓ_1 -normi on hyvin pieni. Mikäli optimointiongelma (4.2) antaisi edelleen tällä s :n arvolla hyvin s -harvaa vektoria $x\mathbb{1}_S$ vastaavan ratkaisun, on perusteltua kysyä antaisiko se edes kohtalaisen ratkaisun alkuperäisen vektorin $x \in wl_p(L)$ tapauksessa.

Tutkielmani keskeisenä pyrkimyksenä on vastata tähän kysymykseen myöntävästi näyttämällä, että ℓ_1 -minimoinnilla saadaan hyviä tuloksia myös heikon kuulan vektorien tapauksessa. Tarkemmin ottaen esitän luvussa 7 artikkelin [8] mukaisesti todistuksen seuraavalle lauseelle:

Lause 4.5. (Palautuslause): *Olkoon $R \in \mathbb{R}^{k \times n}$ mittausmatriisi, joka noudattaa UUP:tä parametrilla β_{UUP} ja ERP:tä parametrilla β_{ERP} . Asetetaan $\beta := \max(\beta_{UUP}, \beta_{ERP})$. Oletetaan, että $k \geq \beta$ ja vektori $x \in \mathbb{R}^n$ kuuluu heikkoon kuulaan $wl_p(L)$, jossa $0 < p < 1$ ja $L > 0$. Asetetaan vielä $r := 1/p - 1/2$.*

Tällöin jokaisella riittävän pienellä $\alpha > 0$ jokaisen mittauksista Rx saadun optimointiongelman (4.2) ratkaisun \hat{x} ja alkuperäisen vektorin x välisen erotuksen pituus noudattaa ylärajaa

$$\|x - \hat{x}\|_{\ell_2} \leq C_{p,\alpha} \cdot L \cdot \left(\frac{k}{\beta}\right)^{-r}$$

vähintään todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$. Ylärajassa $C_{p,\alpha}$ on positiivinen vakio, jonka arvo riippuu vakioista p ja α .

Lauseessa mainitut mittausmatriisilta vaaditut ominaisuudet UUP ja ERP esittelen seuraavassa luvussa 5. Luvussa 6 käsittelemme ERP:stä johdettavaa aputulosta. Palautuslauseen todistuksen lisäksi johdamme luvussa 7 UUP:n ominaisuuksista aputuloksen, jota tarvitsemme palautuslauseen todistamisessa. Jotta saisimme konkreettisen esimerkin sopivasta mittausmatriisista, todistamme luvussa 8 normaalin mittausmatriisiin täyttävän palautuslauseen asettamat vaatimukset. Tällä matriisivalinnalla esittelen teorian toimivuutta numeerisesti luvussa 9. Muita mahdollisia matriisivalintoja ja palautuslauseelle tehtäviä yleistyksiä käsittelem lyhyesti luvussa 10.

Tämän luvun alkupuolella esitelty CS-ongelmien teoria pohjaa jo edellä mainittuun lähteeseen [4]. Jälkimmäisessä osuudessa heikon kuulan määritelmä ja palautuslause ovat lähteestä [8]. Käytännön sovelluksista kiinnostuneet lukijat voivat löytää mielenkiintoiseksi artikkelin [19], jonka esittelemistä lukuisista sovelluksista mainittakoon kamerat, magneettikuvaus ja tietokoneverkkojen hallinta.

Luku 5

Vaatimukset mittausmatriisille

Lauseeksi 4.5 muotoiltu ratkaisu viime luvussa esiteltyyn käänteisongelmaan tekee kaksi oletusta mittausmatriisista. Tässä luvussa käymme läpi näiden ominaisuuksien määritelmät avaten niissä esiintyviä merkintöjä ja pohjustaen niiden käyttöä tulevissa luvuissa.

Määritelmiä varten tarvitsemme käyttöömmme rajoitekuvauksen:

Määritelmä 5.1. (Rajoitekuvaus Q_{nI}) Olkoon $I \subset \mathbb{Z}_n$ joukko indeksejä ja $i_1, \dots, i_{|I|}$ kyseiset indeksit nousevassa suuruusjärjestyksessä. Määritellään rajoitekuvaus $\ell_2(\mathbb{Z}_n) \rightarrow \ell_2(I)$ matriisina $Q_{nI} \in \mathbb{R}^{n \times |I|}$ asettamalla

$$Q_{nI}(j, i_j) = 1, \quad \text{kun } 1 \leq j \leq |I|,$$

ja täydentämällä muut alkiot nolliksi.

Näin määriteltynä rajoitekuvaus poistaa kuvattavasta vektorista $v \in \mathbb{R}^n$ kuvauksessa $Q_{nI}v$ kaikki indeksijoukon I ulkopuolella olevia indeksejä vastaavat ulottuvuudet jättäen I :n indeksejä vastaavat ulottuvuudet koskemattomiksi.

Esimerkki 5.2. Olkoon $n := 4$ ja $I := \{1, 3\} \subset \mathbb{Z}_4$ ja olkoon rajoitettavana vektorina $v \in \mathbb{R}^4$. Tällöin rajoite näyttää seuraavalta:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_3 \end{bmatrix}$$

Rajoitekuvauksella voimme myös valita indeksijoukkoa I vastaavat sarakkeet halutusta avaruuden $\mathbb{R}^{d \times n}$, $d \in \mathbb{N}$, matriisista:

Esimerkki 5.3. Olkoon jälleen $n := 4$ ja $I := \{1, 3\} \subset \mathbb{Z}_4$. Olkoon lisäksi $d := 3$ ja rajoitettavana matriisina $M \in \mathbb{R}^{3 \times 4}$. Kertomalla nyt matriisia M oikealta rajoitekuvauksen transpoosilla pystymme rajoittumaan matriisin M sarakkeisiin I :

$$MQ_{nI}^T = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{13} \\ m_{21} & m_{23} \\ m_{31} & m_{33} \end{bmatrix}$$

Tätä ominaisuutta hyödyntämällä määrittelemme jatkossa tarvittavan matriisin

$$R_{kS} := RQ_{nS}^T \in \mathbb{R}^{k \times |S|},$$

jossa on siis valittu indeksijoukon $S \subset \mathbb{Z}_n$ indeksejä vastaavat sarakkeet mittausmatriisista R .

Määrittelemme seuraavaksi molemmat mittausmatriisilta vaadittavat ominaisuudet. Tarkastelemme tämän jälkeen niissä olevia merkintöjä, mutta mainitaan jo nyt, että fraasia ”*kaikilla riittävän pienillä* $\alpha > 0$ ” käyttäessämme tarkoitamme jollain absoluuttisella vakiolla α_0 pätevän $\alpha \leq \alpha_0$.

Määritelmä 5.4. (UUP, Uniform Uncertainty Principle) Olkoon $R \in \mathbb{R}^{k \times n}$, $k \leq n$, matriisi, $S \subset \mathbb{Z}_n$ joukko, ja määritellään tapahtuma

$$(5.5) \quad E_S := \left\{ \frac{k}{2n} \leq \lambda_{\min}(R_{kS}^T R_{kS}) \leq \lambda_{\max}(R_{kS}^T R_{kS}) \leq \frac{3k}{2n} \right\}.$$

Matriisin R sanotaan noudattavan UUP:tä parametrilla β , jos *kaikilla riittävän pienillä* $\alpha > 0$ ja jollain kiinnitetyllä positiivisella vakiolla ρ pätee

$$\mathbb{P} \left(\bigcap_{S \in \mathcal{S}} E_S \right) \geq 1 - O_\alpha(n^{-\rho/\alpha}),$$

kun

$$(5.6) \quad \mathcal{S} := \left\{ S \subset \mathbb{Z}_n : 0 < |S| \leq \alpha \cdot \frac{k}{\beta} \right\}.$$

Määritelmä 5.7. (ERP, Exact Reconstruction Principle) Olkoon $R \in \mathbb{R}^{k \times n}$, $k \leq n$, matriisi. Olkoon lisäksi $S \subset \mathbb{Z}_n$ mielivaltainen indeksijoukko, jonka koko noudattaa rajoja

$$(5.8) \quad 0 < |S| \leq \alpha \cdot \frac{k}{\beta},$$

ja $\phi \in \mathbb{R}^n$ joukossa S kannatettu ($|\phi(s)| = 1, s \in S$) mielivaltainen merkkivektori. Matriisin R sanotaan noudattavan ERP:tä parametrilla β , jos *kaikilla riittävän pienillä* $\alpha > 0$ ja jollain kiinnitetyllä positiivisella vakiolla ρ löytyy vähintään todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$ vektori $v \in \mathbb{R}^n$ seuraavilla ominaisuuksilla:

- (i) $v(s) = \phi(s)$ kaikilla $s \in S$,
- (ii) v on lineaarikombinaatio matriisin R riveistä eli $v = R^T w$ jollain $w \in \mathbb{R}^k$,
- (iii) $|v(s)| \leq \frac{1}{2}$ kaikilla $s \in S^c$.

Katsoessamme näiden määritelmien merkitystä vain palautuslauseen 4.5 todistamisen kannalta voidaan kärjistetysti todeta, että normia $\|x - \hat{x}\|_{\ell_2}$ arvioidessamme UUP on tärkeässä osassa indeksijoukkoa S koskeissa arvioinneissa ja ERP:llä

johdamme arviot joukossa S^c . Suoraan emme kuitenkaan tule todistuksessa käyttämään kumpaakaan näistä määritelmistä, vaan johdamme myöhemmin molemmille omat apulauseet, joiden kautta määritelmien voima tulee näkyville.

Yleisesti ottaen UUP:tä voidaan pitää luonteeltaan saman kaltaisena kuin satunnaisprojektiomenetelmän yhteydessä esiteltyä Johnson-Lindenstrauss lemmaa. UUP:n vaatimus (5.5) voidaan nimittäin esittää muodossa

$$\frac{k}{2n} \|x\|_{\ell_2}^2 \leq \|Rx\|_{\ell_2}^2 \leq \frac{3k}{2n} \|x\|_{\ell_2}^2,$$

kun olemme rajoittuneet tutkimaan $(\alpha k/\beta)$ -harvoja tai harvempia vektoreita. ERP:n merkityksestä ongelmallemme kertoo artikkelissa [7] todistettu tulos, jonka mukaan saamme palautettua joukossa S tuetun vektorin x eksaktisti optimointi ongelmas- ta (4.2) täsmälleen silloin, kun ERP:n määritelmässä esitelty vektori v on olemas- sa merkkifunktiolle $\phi(s) = (\text{sgn}(x))(s)$. Tästä ominaisuudesta seuraa itse asiassa ERP:n nimi. [8]

Pureudumme vielä yksityiskohtaisemmin määritelmien merkintöihin. Ensinnä- kin molemmat määritelmät edellyttävät mittaamatriisilta ehtojensa toteuttamis- ta todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$, joka tulee edelleen periytymään palautus- lauseen 4.5 toteutumistodennäköisyydeksi. Yksinkertaistaen vaadittua todennäköi- syyttä voidaan pitää ”musertavan suurena”, kuten artikkelissa [8] asia ilmaistaan. Tätä suuruutta voinee selittää termillä $n^{-\rho/\alpha}$, joka vetää koko $O(\cdot)$ -termin lähelle nolaa käsitellessämme riittävän korkeaulotteisia matriiseja.

Yhteistä määritelmille on myös niissä esiintyvät määritelmiä ”konkretisoivat” vakiot. Kuten artikkelissa [8] todetaan, voisimme korvata ERP:n vektorin v ominai- suudessa (iii) esiintyvän painotusvakion $1/2$ jollain vakiolla avoimelta väliltä $(0, 1)$ ja UUP:n ominaisarvoja rajoittavat vakiot $1/2$ ja $3/2$ joillain vakioilla avoimelta väliltä $(0, \infty)$.

Juuri määritelmässä mainitun arvon sijaan ERP:n painotusvakiolle on oleellista, että v toimii likimääräisesti joukossa S kannatetun vektorin merkkifunktiona eli joukon ulkopuolelle jäävät arvot ovat yksikköä pienempiä, vaikka eivät aivan nolli- olisikaan. Tarkemmin tulemme havainnollistamaan painotusvakion mielivaltaisuutta omassa ongelmassamme, kun näytämme seurauslausetta 6.3 todistaessamme pääty- vämme oleellisesti samaan lopputulokseen muillakin painotusvakion valinnoilla. Ky- seinen seurauslause on sama aikaisemmin mainittu apulause, jonka kautta käytämme ERP:tä palautuslauseen todistuksessa.

Hieman muokkaamalla UUP:n konkretisoivia vakioita pystymme kirkastamaan määritelmän suhdetta satunnaisprojektiomenetelmän yhteydessä mainittuihin lau- seisiin. Mieltämällä nimittäin vakiot $1/2$ ja $3/2$ seuraukseksi, valinnalla $\epsilon := 1/2$, termeistä $(1-\epsilon)$ ja $(1+\epsilon)$, ja kertomalla termin k/n keskelle esityksessä (5.9) saamme rajat,

$$(5.9) \quad (1 - \epsilon) \|x\|_{\ell_2}^2 \leq \frac{n}{k} \|Rx\|_{\ell_2}^2 \leq (1 + \epsilon) \|x\|_{\ell_2}^2,$$

jotka vastaavat täsmälleen normaalille satunnaismatriisille esitettyjä rajoja lau- seessa 3.2. Mainitun lauseen tavoin myös UUP:n rajoihin liittyy toteutumistoden- näköisyys. Erityisesti UUP:n tapauksessa tämä toteutumistodennäköisyys kattaa

samanaikaisesti kaikki korkeintaan $\alpha k/\beta$ kokoiset joukot S joukkokohtaisen todennäköisyyden sijaan.

Tehdään vielä muutama havinto ylärajasta $\alpha k/\beta$ vähintään $\alpha k/\beta$ -harvojen vektoreiden tapauksessa. Jotta UUP:n ominaisuuksia voitaisiin soveltaa vähemmänkin harvoille vektoreille, on toiveena luonnollisesti mahdollisimman suuri yläraja. Käytännössä suuren ylärajan saaminen ei ole itsestään selvää, sillä α :n täytyy olla riittävän pieni, että käytettävä mittausmatriisi ylipäänsä toteuttaa UUP:n, ja mittausten määrä k halutaan jo lähtökohtaisesti pitää pienenä. α :n kohdalla pienuus lisäksi kasvattaa toteutumistodennäköisyyttä $1 - O_\alpha(n^{-\rho/\alpha})$. Siten jäljelle jää parametri β , jonka pienuus kasvattaa ylärajaa. Tämän parametriarvon pienuutta voikin käyttää eri mittausmatriisityyppien vertailuun. Normaalisen satunnaismatriisin kohdalla tulemme saamaan UUP:n tapauksessa β :lle arvon $\log n/k$ ja ERP:n tapauksessa vaatimattomamman arvon $\log n$.

Luku 6

Ratkaisun vakaus normin ℓ_1 suhteen

Palautusongelmassamme oletamme, että vektori $x \in \mathbb{R}^n$ kuuluu heikkoon kuulaan $wl_p(L)$. Vektorin heikkoon kuulaan sisältyminen voi nähdä oletuksena vektorin massan pakkautumisesta kooltaan pieneen joukkoon S . Tällä tarkoitamme, että x :n itseisarvojen summa eli ℓ_1 -normi kyseisen joukon ulkopuolella on pientä joukosta laskettuun ℓ_1 -normiin verrattuna.

Tässä luvussa todistamme, että suurella todennäköisyydellä ongelmamme millään ratkaisulla \hat{x} ei ole paljoa massaa joukon S ulkopuolella, jos alkuperäiselläkään vektorilla ei ole. Lisäksi tulemme saamaan ylärajan ratkaisun \hat{x} alkioiden itseisarvoille joukossa S^c . Tulosten saamiseksi meidän täytyy olettaa, että mittaukset on otettu ERP:n toteuttavalla mittausmatriisilla.

Todistamme massaa koskevan tuloksen ensin yleisille vektoreille $x \in \mathbb{R}^n$ ja johdamme siitä seurauksen heikon kuulan vektoreille.

Lemma 6.1. *Olkoon $x \in \mathbb{R}^n$ vektori, $R \in \mathbb{R}^{k \times n}$ ERP:tä noudattava matriisi ja S indeksijoukko, jonka koko noudattaa ERP:n määritelmän antamaa rajaa $|S| \leq \alpha \cdot \frac{k}{\beta}$. Jaetaan x osiin $x = x_S + h$ siten, että x_S on ainoastaan joukossa S kannatettu vektori.*

Tällöin jokainen ongelman (4.2) ratkaisu \hat{x} noudattaa rajoja

$$\|\hat{x} \cdot \mathbf{1}_{S^c}\|_{\ell_1} \leq 4 \|h\|_{\ell_1}$$

vähintään todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$.

Todistus. Luonnollisesti x toteuttaa itse optimointiongelman (4.2) vaatimuksen $Rz = Rx$. Tämän vaatimuksen toteuttavista vektoreista on ongelman ratkaisevalla vektorilla \hat{x} pienin ℓ_1 -normi ja siten täytyy päteä

$$(6.2) \quad \|\hat{x}\|_{\ell_1} \leq \|x\|_{\ell_1} \leq \|x_S + h\|_{\ell_1} \leq \|x_S\|_{\ell_1} + \|h\|_{\ell_1}.$$

Olkoon $\phi := \text{sgn}(x_S)$ merkkifunktio. Koska oletuksen mukaan x_S on kannatettu joukossa S , pätee $|\phi(s)| = 1$ kaikilla $s \in S$. Oletuksen mukaan R noudattaa ERP:tä ja siten todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$ on olemassa vektori $v \in \mathbb{R}^n$, että $v = R^T w$ jollain $w \in \mathbb{R}^k$ ja $v(s) = \phi(s)$ kaikilla $s \in S$ sekä $|v(s)| \leq \frac{1}{2}$ kaikilla $s \in S^c$.

Näillä tiedoilla saamme identiteetin ja kaksi arviota sisätulolle $\langle \hat{x}, v \rangle$. Kyseinen identiteetti rakentaa yhteyden alkuperäisen vektorin x ja ratkaisun \hat{x} välille:

$$\begin{aligned} \langle \hat{x}, v \rangle &= \langle \hat{x}, R^T w \rangle \\ &= \langle R\hat{x}, w \rangle \\ &= \langle Rx, w \rangle \\ &= \langle R(x_S + h), w \rangle \\ &= \langle x_S + h, R^T w \rangle \\ &= \langle x_S + h, v \rangle. \end{aligned}$$

Ensimmäisen arvion sisätulolle saamme käyttämällä tietoa

$$x_S(s)v(s) = \begin{cases} x_S(s)\phi(s), & \text{kun } s \in S \\ 0 \cdot v(s), & \text{kun } s \in S^c \end{cases} \implies \langle x_S, v \rangle = \|x_S\|_{\ell_1}$$

arvioidessamme identiteettiä alaspäin:

$$\langle \hat{x}, v \rangle = \langle x_S + h, v \rangle = \langle x_S, v \rangle + \langle h, v \rangle = \|x_S\|_{\ell_1} + \langle h, v \rangle \geq \|x_S\|_{\ell_1} - \|h\|_{\ell_1}.$$

Toisen arvion sisätulolle saamme, kun indeksijoukkoja S ja S^c tarkastellaan erikseen:

$$\begin{aligned} |\langle \hat{x}, v \rangle| &= \left| \sum_{s \in \mathbb{Z}_n} \hat{x}(s)v(s) \right| \\ &= \sum_{s \in \mathbb{Z}_n} |\hat{x}(s)| |v(s)| \\ &\leq \sum_{s \in S} |\hat{x}(s)| + \frac{1}{2} \sum_{s \in S^c} |\hat{x}(s)| \\ &= \sum_{s \in \mathbb{Z}_n} |\hat{x}(s)| - \frac{1}{2} \sum_{s \in S^c} |\hat{x}(s)| \\ &= \|\hat{x}\|_{\ell_1} - \frac{1}{2} \|\hat{x} \cdot \mathbf{1}_{S^c}\|_{\ell_1}. \end{aligned}$$

Olemme siten johtaneet sisätulolle ala- ja ylärajat

$$\|x_S\|_{\ell_1} - \|h\|_{\ell_1} \leq \langle \hat{x}, v \rangle \leq \|\hat{x}\|_{\ell_1} - \frac{1}{2} \|\hat{x} \cdot \mathbf{1}_{S^c}\|_{\ell_1}.$$

Tutkimalla vielä rajojen välistä epäyhtälöä voimme tehdä päättelyn

$$\frac{1}{2} \|\hat{x} \cdot \mathbf{1}_{S^c}\|_{\ell_1} \leq \|\hat{x}\|_{\ell_1} + \|h\|_{\ell_1} - \|x_S\|_{\ell_1} \stackrel{(6.2)}{\leq} (\|x_S\|_{\ell_1} + \|h\|_{\ell_1}) + \|h\|_{\ell_1} - \|x_S\|_{\ell_1} = 2 \|h\|_{\ell_1},$$

josta väite seuraa kertoessamme puolittain vakiolla 2. \square

ERP:n esittelyn yhteydessä mainitsimme määritelmän painotusvakion olevan $1/2$ konkretian takia. Äskeisen todistuksen lopusta voikin huomata, että päädyimme kertomaan vakiolla 2, jotta painotusvakion määräämä vakio $1/2$ häviäisi epäyhtälön

vasemmalta puolelta. Vaikka olisimme korvanneet lopun päättelyssä painotusvakion millä tahansa luvulla avoimelta väliltä $(0, 1)$ olisimme aina voineet tehdä oleellisesti vastaavat päätelmät ja kertoa lopussa painotusvakion määräämään vakion pois vasemmalta puolelta.

Käyttämällä saammaamme tulosta voimme johtaa siitä seurauksen heikon kuulan vektoreille. Tähän seuraukseen tiivistyy kaikki tarvittava ERP:n tarjoama tieto palautuslauseen 4.5 todistamista varten. Käytämme nimittäin ERP:tä palautuslauseen todistuksessa vain tämän seurauksen kautta.

Seuraus 6.3. *Olkoon $x \in wl_p(L)$ ja \hat{x} jokin ℓ_1 -minimoija ongelmalle (4.2). Tällöin lemmän 6.1 oletusten ollessa voimassa pätee ℓ_1 -arvio*

$$(6.4) \quad \|\hat{x} \cdot \mathbf{1}_{S^c}\|_{\ell_1} \leq \frac{4pL}{1-p} |S|^{1-1/p}.$$

Lisäksi saamme yksittäisiä alkiota koskevan arvion

$$(6.5) \quad |\hat{x}_{\text{arr}}(j)| \leq \frac{4pL}{1-p} \cdot |S|^{-1/p} \quad \text{jokaisella } j > 2|S|,$$

kun $\hat{x}_{\text{arr}} \in \mathbb{R}^n$ on vektori, johon \hat{x} :n alkiot on järjestetty itseisarvoltaan laskevaan suuruusjärjestykseen.

Todistus. Olkoon S indeksijoukko, joka sisältää $|S|$ vektorin $|x|$ suurinta arvoa. Asetetaan $x_S = x \cdot \mathbf{1}_S$ ja $h = x \cdot \mathbf{1}_{S^c}$ eli $x = x_S + h$ apulauseen 6.1 mukaisesti. Tutkimalla osaa h saamme

$$\|h\|_{\ell_1} = \|x \cdot \mathbf{1}_{S^c}\|_{\ell_1} = \sum_{s \in S^c} |x(s)| \leq \sum_{s=|S|+1}^n L \cdot s^{-1/p} \leq L \cdot \sum_{s=|S|+1}^{\infty} s^{-1/p}.$$

Voimme arvioida tätä yliharmonista summaa ylöspäin integraalin avulla seuraavasti:

$$\sum_{s=|S|+1}^{\infty} s^{-1/p} \leq \lim_{M \rightarrow \infty} \int_{|S|}^M s^{-1/p} ds \leq \lim_{M \rightarrow \infty} \int_{|S|}^M \frac{1}{1 - \frac{1}{p}} s^{1-\frac{1}{p}} = \frac{p}{1-p} |S|^{1-1/p}.$$

Siten saamme h :lle ylärajan

$$(6.6) \quad \|h\|_{\ell_1} \leq \frac{pL}{1-p} |S|^{1-1/p},$$

josta apulauseen 6.1 käyttämällä saadaan todistetuksi väitteen ensimmäinen osa:

$$(6.7) \quad \|\hat{x} \cdot \mathbf{1}_{S^c}\|_{\ell_1} \leq \frac{4pL}{1-p} |S|^{1-1/p}.$$

Jatkamme todistusta määrittelemällä joukon E_j , joka sisältää $j > 2|S|$ vektorin $|\hat{x}|$ suurinta arvoa vastaavat indeksit. Tällöin pätee luonnollisesti

$$|E_j \cap S^c| \geq j - |S|.$$

Siten saamme alarajan

$$\|\hat{x}\|_{\ell_1(E_j \cap S^c)} \geq (j - |S|) \cdot |\hat{x}_{\text{arr}}(j)| \geq |S| \cdot |\hat{x}_{\text{arr}}(j)|.$$

Käyttämällä nyt tulosta (6.7) saamme ylärajan

$$\|\hat{x}\|_{\ell_1(E_j \cap S^c)} \leq \|\hat{x} \cdot \mathbf{1}_{S^c}\|_{\ell_1} \leq \frac{4pL}{1-p} |S|^{1-1/p}.$$

Väite seuraa, kun yhdistetään ala- ja ylärajat ja jaetaan $|S|$:llä. □

Luvun molemmat todistukset pohjautuvat artikkelin [8] vastaavien tulosten todistuksiin. Olen lisännyt todistuksiin joitain yksityiskohtia, kuten seurauslauseessa vektorin h ylärajaa koskevan päättelyn. Kyseisessä kohdassa ja muissakin lähdeartikkelin todistuksissa on käytäntönä merkitä vakioita symbolilla C , mutta olen tutkielmassa pyrkinyt laskemaan ne eksaktisti.

Luku 7

Palautuslause: ratkaisun vakaus normin ℓ_2 suhteen

Todistamme tässä luvussa tutkielman päätuloksen eli luvussa 4 esitetyn palautuslauseen 4.5. Kuten mittausmatriisin ominaisuuksien UUP ja ERP esittelyn yhteydessä todettiin, ei kyseisiä ominaisuuksia käytetä palautuslauseen todistuksessa suoraan, vaan vain aputulosten välityksellä. Viime luvussa keskityttiin johtamaan tällainen aputulos, seuraus 6.3, ERP:lle. Seuraavaksi johdamme apulauseen UUP:n käyttöä varten. Apulauseen väite noudattaa pääosin artikkelin [8] vastaavan tuloksen muotoa, mutta väitteemme on heikompi. Tämä valinta ei johda päätuloksen väitteen heikkenemiseen, mutta lyhentää hieman aputuloksen todistusta.

Lemma 7.1. *Olkoon $R \in \mathbb{R}^{k \times n}$ UUP:tä noudattava matriisi. Tällöin seuraava väite on totta vähintään todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$: Jokaiselle vektorille $z \in \ell_2(S)$, jossa $S \subset \mathbb{Z}_n$ on UUP:n rajoja (5.6) noudattava indeksijoukko, löytyy laajennus $v \in \ell_2(\mathbb{Z}_n)$ ominaisuuksilla*

- (i) $Q_{nS}v = z$ eli laajennuksen rajoitus vastaa alkuperäistä vektoria.
- (ii) Laajennus on lineaarikombinaatio R :n riveistä eli $v = R^T w$ jollain $w \in \mathbb{R}^k$.
- (iii) Voimassa yläraja $\|v\|_{\ell_2(E)} \leq 3 \|z\|_{\ell_2(S)}$ jokaisella $E \subset \mathbb{Z}_n$, jolle $|E| \leq |S|$.

Todistus. Aloitamme todistuksen tekemällä matriisia $R_{kS} = RQ_{nS}^T$ koskevia normiarvioita, joiden avulla myöhemmin valittavalle laajennukselle v väitteen ominaisuuksien todistaminen sujuu helposti. Koska R on UUP:tä noudattava matriisi ja S noudattaa UUP:n määritelmässä mainittua ylärajaa $\alpha k/\beta$ ovat rajat

$$\frac{1}{2} \frac{k}{n} \leq \lambda_{\min}(R_{kS}^T R_{kS}) \leq \lambda_{\max}(R_{kS}^T R_{kS}) \leq \frac{3}{2} \frac{k}{n}.$$

voimassa todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$. Käytämme näitä rajoja normiarvioihin, mistä seuraa väitteessä mainittu toteutumistodennäköisyys. Rajojen lisäksi käytämme esitietojen kohdasta (2.15) alkaen esiteltyjä yhteyksiä matriisin normin, ominaisarvojen ja singulaariarvojen välillä.

Neliömatriisin normi on sama kuin sen suurin ominaisarvo eli

$$\|R_{kS}^T R_{kS}\| = \lambda_{\max}(R_{kS}^T R_{kS}) \leq \frac{3k}{2n}.$$

Koska UUP:n antamien rajojen mukaan matriisiin $R_{kS}^T R_{kS}$ kaikki ominaisarvot ovat aidosti positiivisia, tiedämme käänteismatriisin olevan olemassa ja edelleen tiedämme käänteismatriisin normin olevan alkuperäisen matriisin pienimmän ominaisarvon käänteisluku. Siten saamme käänteismatriisin normille ylärajan

$$\|(R_{kS}^T R_{kS})^{-1}\| \leq \frac{2n}{k}.$$

Matriisin R_{kS} normi on sama kuin sen suurin singulaariarvo, joka taas on matriisin $R_{kS}^T R_{kS}$ suurimman ominaisarvon neliöjuuri eli

$$\|R_{kS}\| \leq \sqrt{\frac{3k}{2n}}.$$

Sama arvio pätee myös matriisin R_{kE}^T normille todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$, sillä matriisin normi pysyy muuttumattomana transponoinnissa ja indeksijoukon E alkionäärä noudattaa ylärajaa $|E| \leq |S| \leq \alpha k/\beta$.

Olemme saaneet riittävästi tietoa normeista. Asetamme

$$v := R^T R_{kS} (R_{kS}^T R_{kS})^{-1} z$$

ja näytämme tällä valinnalla olevan aina vaaditut ominaisuudet (i)-(iii). Kaksi ensimmäistä ehtoa seuraavat suoraan v :n määritelmästä. Viimeinenkin ehto seuraa yksinkertaisesti tehdyistä normiarvioista:

$$\begin{aligned} \|v\|_{\ell_2(E)} &= \|Q_{nE} v\|_{\ell_2} \\ &\leq \|R_{kE}^T\| \|R_{kS}\| \|(R_{kS}^T R_{kS})^{-1}\| \|z\|_{\ell_2(S)} \\ &\leq \sqrt{\frac{3k}{2n}} \cdot \sqrt{\frac{3k}{2n}} \quad \frac{2n}{k} \quad \|z\|_{\ell_2(S)} \\ &= 3 \|z\|_{\ell_2(S)}. \end{aligned}$$

□

Voimme nyt todistaa tutkielman päätuloksen. Todistus on artikkelista [8], mutta olen lisännyt siihen useampia yksityiskohtia.

Lause. (Palautuslause 4.5): *Olkoon $R \in \mathbb{R}^{k \times n}$ mittausmatriisi, joka noudattaa UUP:tä parametrilla β_{UUP} ja ERP:tä parametrilla β_{ERP} . Asetetaan $\beta := \max(\beta_{UUP}, \beta_{ERP})$. Oletetaan, että $k \geq \beta$ ja vektori $x \in \mathbb{R}^n$ kuuluu heikkoon kuulaan $wl_p(L)$, jossa $0 < p < 1$ ja $L > 0$. Asetetaan vielä $r := 1/p - 1/2$.*

Tällöin jokaisella riittävän pienellä $\alpha > 0$ jokaisen mittauksista Rx saadun optimointiongelman (4.2) ratkaisun \hat{x} ja alkuperäisen vektorin x välisen erotuksen pituus noudattaa ylärajaa

$$\|x - \hat{x}\|_{\ell_2} \leq C_{p,\alpha} \cdot L \cdot \left(\frac{k}{\beta}\right)^{-r}$$

vähintään todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$. Ylärajassa $C_{p,\alpha}$ on positiivinen vakio, jonka arvo riippuu vakioista p ja α .

Todistus. Oletamme S_0 :n indeksijoukoksi, joka sisältää t suurinta vektorin $|x|$ arvoa vastaavat indeksit, vastaavasti S_1 ja $|\hat{x}|$. Asetamme $S := S_0 \cup S_1$, jolle määrittelystä johtuen pätee $t \leq |S| \leq 2t$. Oletamme joukon S alkion määrän noudattavan lisäksi ylärajaa $\alpha k/\beta$, jolloin sekä UUP:n että ERP:n määritelmien kuvaamat ominaisuudet ovat voimassa vähintään todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$.

Todistamme väitteen arvioiden vektorin $(x - \hat{x})$ euklidista normia erikseen joukoissa S ja S^c . Päättelemme normille ylärajan ensin jälkimmäisessä indeksijoukossa. Käytämme tähän yleistä jokaiselle vektorille $v \in \mathbb{R}^n$ pätevää interpolointitulosta

$$\|v\|_{\ell_2}^2 = \sum_{k=1}^n v_k^2 = \sum_{k=1}^n |v_k| |v_k| \leq \sup_{k \in \mathbb{Z}_n} |v_k| \cdot \sum_{k=1}^n |v_k| = \|v\|_{\ell_\infty} \|v\|_{\ell_1}.$$

Arvioidaksemme siis vektorin $(x - \hat{x})$ euklidisen normin neliötä tulee meidän arvioida sen ℓ_∞ - ja ℓ_1 -normeja. ℓ_1 -normia koskevaa ylärajaa varten teemme ensin yksinkertaisen arvion

$$\|x - \hat{x}\|_{\ell_1(S^c)} \leq \|x - \hat{x}\|_{\ell_1(S_0^c)} \leq \|x\|_{\ell_1(S_0^c)} + \|\hat{x}\|_{\ell_1(S_0^c)}.$$

Tässä ylärajan vektorin x normille saamme seurauksen 6.3 epäyhtälöstä (6.6) ja vektorin \hat{x} normille saman seurauksen ensimmäisestä tuloksesta (6.4):

$$\begin{aligned} \|x - \hat{x}\|_{\ell_1(S^c)} &\leq \frac{pL}{1-p} |S_0|^{1-\frac{1}{p}} + \frac{4pL}{1-p} |S_0|^{1-\frac{1}{p}} \\ &= \frac{5pL2^{\frac{1}{p}}}{1-p} (2|S_0|)^{1-\frac{1}{p}} \\ (7.2) \qquad &\leq C_1 L \cdot |S|^{1-\frac{1}{p}}, \end{aligned}$$

jossa

$$C_1 := \frac{5p2^{\frac{1}{p}}}{1-p}.$$

ℓ_∞ -normia koskevan ylärajan johdamme samaan tapaan aloittaen arviolla

$$\|x - \hat{x}\|_{\ell_\infty(S^c)} \leq \|x\|_{\ell_\infty(S^c)} + \|\hat{x}\|_{\ell_\infty(S^c)} \leq \|x\|_{\ell_\infty(S_0^c)} + \|\hat{x}\|_{\ell_\infty(S_0^c)}.$$

Vektorin x yläraja joukossa S_0^c seuraa suoraan heikon kuulan määritelmästä:

$$\|x\|_{\ell_\infty(S_0^c)} \leq L \cdot |S_0|^{-\frac{1}{p}} = 2^{\frac{1}{p}} L \cdot |2S_0|^{-\frac{1}{p}} \leq 2^{\frac{1}{p}} L \cdot |S|^{-\frac{1}{p}}$$

Vektorin \hat{x} normille saamme ylärajan seurauksen 6.3 jälkimmäisestä tuloksesta, kun sovellamme sitä lähdeartikkelin toisen tekijän neuvosta [25] x :n $|S_0|/2$ suurimmalle alkionle:

$$\|\hat{x}\|_{\ell_\infty(S_0^c)} \leq \frac{4pL}{1-p} \cdot \left(\frac{|S_1|}{2}\right)^{-\frac{1}{p}} = \frac{4pL2^{\frac{2}{p}}}{1-p} \cdot (2|S_1|)^{-\frac{1}{p}} \leq \frac{4pL2^{\frac{2}{p}}}{1-p} \cdot |S|^{-\frac{1}{p}}.$$

Siten ℓ_∞ -normin ylärajaksi vektorille $(x - \hat{x})$ joukossa S^c saamme

$$(7.3) \qquad \|x - \hat{x}\|_{\ell_\infty(S^c)} \leq 2^{\frac{1}{p}} L \cdot |S|^{-\frac{1}{p}} + \frac{4pL2^{\frac{2}{p}}}{1-p} \cdot |S|^{-\frac{1}{p}} \leq C_2 L \cdot |S|^{-\frac{1}{p}},$$

jossa

$$C_2 := \frac{4p2^{\frac{2}{p}}}{1-p} + 2^{\frac{1}{p}}.$$

Yhdistämällä normiarviot (7.2) ja (7.3) saamme esiteltyä interpolointitulosta käyttämällä vektorin $(x - \hat{x})$ euklidiselle normille indeksijoukossa S^c ylärajan

$$\begin{aligned} \|x - \hat{x}\|_{\ell_2(S^c)} &\leq \|x - \hat{x}\|_{\ell_1(S^c)}^{\frac{1}{2}} \|x - \hat{x}\|_{\ell_\infty(S^c)}^{\frac{1}{2}} \\ &\leq \sqrt{C_1 C_2} L \cdot |S|^{\frac{1}{2} - \frac{1}{p}} \\ (7.4) \qquad \qquad &= \sqrt{C_1 C_2} L \cdot |S|^{-r}. \end{aligned}$$

Enää tulee näyttää vastaavanlainen raja vektorille $(x - \hat{x})$ joukossa S . Apulauseesta 7.1 seuraa, että löytyy vektori $v \in \ell_2(\mathbb{Z}_n)$, joka yhtyy vektoriin $(x - \hat{x})$ joukossa S , kuuluu R :n riviavaruuteen eli $v = R^T w$ jollain $w \in \mathbb{R}^k$, ja noudattaa ℓ_2 -ylärajaa

$$(7.5) \qquad \qquad \|v\|_{\ell_2(E)} \leq 3 \|x - \hat{x}\|_{\ell_2(S)}$$

jokaisella $|E| \leq |S|$.

Nyt v :n ominaisuuksista sekä optimointiongelman (4.2) rajoitteesta $Rx = R\hat{x}$ saamme

$$\langle x - \hat{x}, v \rangle = \langle x - \hat{x}, R^T w \rangle = \langle Rx - R\hat{x}, w \rangle = 0.$$

Jakamalla vielä vasemmanpuoleisimman sisätulon indeksien suhteen osiin S ja S^c ja muistamalla, että $v(s) = (x - \hat{x})(s)$ kaikilla $s \in S$, saamme yhtälön

$$(7.6) \qquad \qquad \sum_{s \in S} |x - \hat{x}|^2(s) = - \sum_{s \in S^c} (x - \hat{x})(s)v(s).$$

Tämän yhtälön johdosta voimme indeksien S sijasta keskittyä jälleen indekseihin S^c , sillä arvioidessamme oikean puolen itseisarvon pieneksi täytyy sitä olla myös vasemman puolen. Aloitamme jakamalla S^c :n $|S|$ kokoisein osiin.

Ositamme joukon S^c esittämällä sen alkiot vektorin $|x - \hat{x}|$ arvojen suhteen laskevaan suuruusjärjestykseen järjestettynä jonona $m_1, \dots, m_{n-|S|}$ ja muodostamalla tästä jonosta indeksijoukot

$$B_J := \{m_j : J \cdot |S| < j \leq (J+1) \cdot |S|\},$$

jossa $J \geq 0$. Luonnollisesti viimeinen joukoista saattaa olla vajaa eikä siten sisällä täyttä $|S|$ indeksiä. Määritellään vielä merkintä

$$I_J := \|x - \hat{x}\|_{\ell_2(B_J)} = \sqrt{\sum_{j=J \cdot |S| + 1}^{(J+1) \cdot |S|} |x - \hat{x}|^2(m_j)}.$$

Saamme nyt jokaista B_J kohden arvion

$$(7.7) \qquad \left| \langle x - \hat{x}, v \rangle_{(B_J)} \right| \leq \|x - \hat{x}\|_{\ell_2(B_J)} \|v\|_{\ell_2(B_J)} \leq I_J \cdot 3 \|x - \hat{x}\|_{\ell_2(S)},$$

kun käytämme ensin Cauchy-Schwarzin epäyhtälöä ja sitten apulauseen 7.1 v :lle antamaa ℓ_2 -ylärajaa (7.5).

Arvioimme seuraavaksi termejä I_J . Muistamalla, että vektorin $|x - \hat{x}|$ arvot ovat indeksijoukoissa B_J laskevassa järjestyksessä ja käyttämällä tietoa (7.3) vektorin arvojen ylärajalle, saamme ensimmäiselle termille arvion

$$I_0 = \sqrt{\sum_{j=1}^{|B_0|} |x - \hat{x}|^2(m_j)} \leq \sqrt{|S| \cdot |x - \hat{x}|^2(m_1)} \leq |S|^{\frac{1}{2}} \cdot C_2 L |S|^{-\frac{1}{p}} = C_2 L \cdot |S|^{-r}.$$

Vastaavasti saamme lopuille $I_J > 0$ arviot

$$I_J \leq |S|^{\frac{1}{2}} \cdot |x - \hat{x}|(m_{J|S|+1}) \leq |S|^{\frac{1}{2}} \cdot |S|^{-1} \|x - \hat{x}\|_{\ell_1(B_{J-1})} = |S|^{-\frac{1}{2}} \|x - \hat{x}\|_{\ell_1(B_{J-1})},$$

jossa jälkimmäinen epäyhtälö seuraa tiedosta

$$\|x - \hat{x}\|_{\ell_1(B_{J-1})} \geq |S| \cdot |x - \hat{x}|(m_{J|S|+1}).$$

Yhdistämällä nämä arviot saamme termien summalle ylärajan

$$\begin{aligned} \sum_{J \geq 0} I_J &\leq C_2 L \cdot |S|^{-r} + |S|^{-\frac{1}{2}} \sum_{J \geq 1} \|x - \hat{x}\|_{\ell_1(B_{J-1})} \\ &\leq C_2 L \cdot |S|^{-r} + |S|^{-\frac{1}{2}} \|x - \hat{x}\|_{\ell_1(S^c)} \\ &\leq C_2 L \cdot |S|^{-r} + |S|^{-\frac{1}{2}} \cdot C_1 L |S|^{1-\frac{1}{p}} \\ &= (C_1 + C_2) L \cdot |S|^{-r}, \end{aligned}$$

jossa kolmannessa arviossa käytettiin ℓ_1 -tulosta (7.2).

Voimme palata yhtälön (7.6) pariin. Käyttämällä juuri päätelemiämme ylärajoja saamme yhtälön oikeanpuolen itseisarvolle ylärajan,

$$\begin{aligned} \left| \sum_{s \in S^c} (x - \hat{x})(s)v(s) \right| &= \sum_{J \geq 0} \left| \sum_{j \in B_J} (x - \hat{x})(s)v(s) \right| \\ &\leq \sum_{J \geq 0} I_J \cdot 3 \|x - \hat{x}\|_{\ell_2(S)} \\ &\leq 3 \|x - \hat{x}\|_{\ell_2(S)} \cdot \sum_{J \geq 0} I_J \\ &\leq 3(C_1 + C_2)L \cdot |S|^{-r} \cdot \|x - \hat{x}\|_{\ell_2(S)}. \end{aligned}$$

Käyttämällä ylärajaa yhtälöön seuraa epäyhtälö

$$\begin{aligned} \|x - \hat{x}\|_{\ell_2(S)}^2 &= \sum_{s \in S} |x - \hat{x}|^2(s) \\ &= - \sum_{s \in S^c} (x - \hat{x})(s)v(s) \\ &\leq 3(C_1 + C_2)L \cdot |S|^{-r} \cdot \|x - \hat{x}\|_{\ell_2(S)}, \end{aligned}$$

jonka puolittain termillä $\|x - \hat{x}\|_{\ell_2(S)}$ jakamalla saamme kaivatun ℓ_2 -arvion indekseille S .

Yhdistämällä tähän kohdan (7.4) ℓ_2 -arvion indekseille S^c päädyimme väitteen vaatimaan ylärajaan vektorien x ja \hat{x} erotuksen pituudelle,

$$\begin{aligned} \|x - \hat{x}\|_{\ell_2} &\leq \|x - \hat{x}\|_{\ell_2(S)} + \|x - \hat{x}\|_{\ell_2(S^c)} \\ &\leq \left(3(C_1 + C_2) + \sqrt{C_1 C_2}\right) L \cdot |S|^{-r} \\ &\leq \left(3(C_1 + C_2) + \sqrt{C_1 C_2}\right) \alpha^{-r} \cdot L \cdot \left(\frac{k}{\beta}\right)^{-r} \\ &= C_{p,\alpha} \cdot L \cdot \left(\frac{k}{\beta}\right)^{-r}, \end{aligned}$$

joka pätee vähintään todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$ ja jossa vakio

$$C_{p,\alpha} := \left(3(C_1 + C_2) + \sqrt{C_1 C_2}\right) \alpha^{\frac{1}{2} - \frac{1}{p}}$$

riippuu vain vakioista p ja α . □

Luku 8

Mittausmatriisina normaalin satunnaismatriisi

Osoitamme tässä luvussa, että määritelmän 3.1 normaalin satunnaismatriisi noudattaa UUP:tä ja ERP:tä parametrilla $\beta = \log n$. Näin ollen saamme esimerkin matriisista, joka soveltuu palautuslauseessa 4.5 mainituksi mittausmatriisiksi.

Todistamme ensin normaalin satunnaismatriisin noudattavan UUP:tä, sillä tarvitsemme tätä tietoa näyttäessämme sen noudattavan myös ERP:tä. Molemmat todistukset seuraavat artikkelin [8] todistuksia, mutta olen muokannut todistuksien rakenteita ja lisännyt paljon yksityiskohtia. Ehkä merkittävimpänä muutoksena on konkretian vuoksi tehty parametrin α lisäys UUP:n todistukseen. Kyseistä parametria ei nimittäin mainita kertaakaan alkuperäisessä todistuksessa.

Ennen UUP:hen liittyvää todistusta esitän normaalin satunnaismatriisin singulaariarvoille todennäköisyyslärajan antavan apulauseen.

Lemma 8.1. *Olkoon $Z \in \mathbb{R}^{d \times m}$, $d \geq m$, matriisi, jonka alkiot ovat toisistaan riippumattomia ja noudattavat standardinormaalijakaumaa. Tällöin matriisin Z/\sqrt{d} suurimmalle ja pienimmälle singulaariarvolle pätevät seuraavat probabilistiset rajat:*

$$\begin{aligned}\mathbb{P}\left(\sigma_{\max}\left(\frac{1}{\sqrt{d}}Z\right) > 1 + \sqrt{\frac{m}{d}} + r\right) &\leq e^{-dr^2/2}, \\ \mathbb{P}\left(\sigma_{\min}\left(\frac{1}{\sqrt{d}}Z\right) < 1 - \sqrt{\frac{m}{d}} - r\right) &\leq e^{-dr^2/2}.\end{aligned}$$

Huomautus 8.2. Tässä itse asiassa pätee, että suurin singulaariarvo suppenee kohti arvoa $1 + \sqrt{m/d}$ ja vastaavasti pienimmälle singulaariarvolle.

Todistus. Tämän todistuksen sivuutamme. Todistuksesta kiinnostunut voi lukea sen artikkelista [23], johon myös lähdeartikkelissa [8] viitataan. \square

Pystymme käyttämään tätä apulauseita arvioidessamme todennäköisyyksiä, joilla UUP:n määritelmässä mainittu ominaisarvoja koskeva ehto (5.5) ei ole voimassa.

Seuraus 8.3. *Olkoon $R \in \mathbb{R}^{k \times n}$, $1 \leq k \leq n$, määritelmän 3.1 mukainen normaalin satunnaismatriisi ja $S \subset \mathbb{Z}_n$ joukko noudatten ehtoa $|S| \leq k/16$. Tällöin pätee*

$$\begin{aligned}\mathbb{P}\left(\lambda_{\min}(R_{kS}^T R_{kS}) < \frac{1}{2} \frac{k}{n}\right) &\leq e^{-\frac{k}{32}}, \\ \mathbb{P}\left(\lambda_{\max}(R_{kS}^T R_{kS}) > \frac{3}{2} \frac{k}{n}\right) &\leq e^{-\frac{k}{32}}.\end{aligned}$$

Todistus. Todistamme vain pienintä ominaisarvoa koskevan rajan, sillä vastaava suurinta ominaisarvoa koskevan tulos voidaan johtaa samalla tavalla.

Kuten esitetöjen kohdan yhteydessä (2.9) todetaan, on matriisin $(R_{kS}^T R_{kS})$ pienin ominaisarvo sama kuin matriisin R_{kS} pienimmän singulaariarvon neliö. Siten voimme yhtäpitävästi laskea todennäköisyyden tapahtumalle

$$\left\{ \sigma_{\min}(R_{kS}) < \sqrt{\frac{k}{2n}} \right\}.$$

Asettamalla $r := 1/4$ saamme apulauseen 8.1 pienintä singulaariarvoa koskevan tapahtuman rajoiksi

$$\begin{aligned} \sigma_{\min}\left(\frac{1}{\sqrt{k}}Z\right) &< 1 - \sqrt{\frac{|S|}{k}} - \frac{1}{4} \\ \Leftrightarrow \sqrt{\frac{n}{k}}\sigma_{\min}(R_{kS}) &< \frac{3}{4} - \sqrt{\frac{|S|}{k}} \\ \Leftrightarrow \sigma_{\min}(R_{kS}) &< \sqrt{\frac{k}{n}}\left(\frac{3}{4} - \sqrt{\frac{|S|}{k}}\right) \end{aligned}$$

ja todennäköisyydeksi

$$\mathbb{P}\left(\sigma_{\min}(R_{kS}) < \sqrt{\frac{k}{n}}\left(\frac{3}{4} - \sqrt{\frac{|S|}{k}}\right)\right) \leq e^{-\frac{k}{32}}.$$

Käyttämällä oletusta $|S| \leq k/16$ voimme huomata väitteen tapahtuman sisältyvän apulauseen antamaan tapahtumaan:

$$\begin{aligned} \left\{ \sigma_{\min}(R_{kS}) < \sqrt{\frac{k}{2n}} \right\} &\subset \left\{ \sigma_{\min}(R_{kS}) < \frac{1}{2}\sqrt{\frac{k}{n}} \right\} \\ &\subset \left\{ \sigma_{\min}(R_{kS}) < \sqrt{\frac{k}{n}}\left(\frac{3}{4} - \sqrt{\frac{|S|}{k}}\right) \right\}, \end{aligned}$$

Siten väite seuraa todennäköisyyksmitan monotonisuudesta, joka on esitelty kohdassa (2.18). \square

Äskeisen seurauksen avulla voimme todistaa normaalisen satunnaismatriisin noudattavan UUP:tä.

Lause 8.4. *Määritelmän 3.1 normaalin satunnaismatriisi $R \in \mathbb{R}^{k \times n}$, $1 \leq k \leq n$, noudattaa UUP:tä parametrilla $\beta := \log n$, kun $k \leq (c - \rho)^{-1}(\log n + 1)$, jossa $c := 1/32$ ja ρ on vakio avoimelta väliltä $(0, c)$.*

Todistus. Asetamme $m := \lfloor \alpha k / \beta \rfloor$, jossa $\lfloor \cdot \rfloor$ tarkoittaa pyöristystä alaspäin lähimpään kokonaislukuun. Päämääränämme on siten näyttää, että jokaisella riittävän pienellä $\alpha > 0$ ja jollain $\rho > 0$ pätee todennäköisyysalaraja

$$\mathbb{P}\left(\bigcap_{S \in \mathcal{S}_m} E_S\right) \geq 1 - O_\alpha(n^{-\rho/\alpha}),$$

jossa

$$\mathcal{S}_m = \{S \subset \mathbb{Z}_n : 0 < |S| \leq m\}.$$

Väitteen voidaan katsoa pätevän todennäköisyydellä 1, kun α sijaitsee avoimella välillä $(0, \beta/k)$. Tällöin nimittäin $m = 0$ eli \mathcal{S}_m on tyhjä. Oleellista on siten väitteen todistaminen tapauksessa

$$\alpha \geq \frac{\beta}{k}.$$

Aloitamme väitteen todistamisen tässä tilanteessa huomioimalla, että tapahtuman $\bigcap_{S \in \mathcal{S}_m} E_S$ komplementti on $\bigcup_{S \in \mathcal{S}_m} E_S^c$ eli voimme yhtäpitästi näyttää toteen todennäköisyysylärajan

$$(8.5) \quad \mathbb{P} \left(\bigcup_{S \in \mathcal{S}_m} E_S^c \right) \leq O_\alpha(n^{-\rho/\alpha}),$$

jossa $E_S^c = \{\lambda_{\min}(R_{kS}^T R_{kS}) < k/2n\} \cup \{\lambda_{\max}(R_{kS}^T R_{kS}) > 3k/2n\}$.

Tarkoituksenamme on siis arvioida komplementtitapahtuman todennäköisyyttä ylöspäin, kunnes saavutamme vaaditun O -notaation mukaisen muodon. Ensimmäiseen arviointiin käytämme esitietojen kohtaa (2.21), jonka mukaan tapahtumien yhdisteen todennäköisyys on korkeintaan yhtäsuurta kuin tapahtumien todennäköisyyksien summa. Siten

$$\mathbb{P} \left(\bigcup_{S \in \mathcal{S}_m} E_S^c \right) \leq \sum_{S \in \mathcal{S}_m} \mathbb{P}(E_S^c).$$

Tapahtumien E_S^c todennäköisyyksille saamme seurauksesta 8.3 ylärajan

$$(8.6) \quad \mathbb{P}(E_S^c) \leq 2e^{-ck}, \quad \text{jossa } c := 1/32,$$

kun joukkojen S alkiomäärä noudattaa ylärajaa

$$(8.7) \quad |S| \leq \frac{k}{16}.$$

Voimme käyttää arviota (8.6) itse asiassa jokaiselle $S \in \mathcal{S}_m$, sillä ainakin kaikilla $\alpha \leq 1/16$ myös m noudattaa ylärajaan $k/16$. Tutkittavien α joukko ei tällä ylärajalla jää tyhjäksi, koska mittausmäärän alarajan nojalla pätee $\log(n)/k < k/16$. Käyttämällä siten arviota jokaiselle joukolle saamme ylärajan

$$(8.8) \quad \mathbb{P} \left(\bigcup_{S \in \mathcal{S}_m} E_S^c \right) \leq 2e^{-ck} \cdot |\mathcal{S}_m|.$$

Seuraavaksi ryhdymme arvioimaan joukkoperheen \mathcal{S}_m alkioiden määrää. Jos kiinnitämme joukon S koon, pystymme valitsemaan $|S|$ alkioisen joukon n alkioisesta joukosta yhteensä $\binom{n}{|S|}$ eri tavalla. Tunnetusti binomikerroin on symmetrinen ja kasvava symmetriapisteeseen asti [27] eli

$$(8.9) \quad \binom{n}{l_1} \leq \binom{n}{l_2} \quad \text{kaikilla } 0 \leq l_1 \leq l_2 < \frac{n}{2}.$$

Koska riittävän pienillä α pätee $m \leq k/16$ ja toisaalta oletamme satunnaismatriisiin R ulottuvuuksien noudattavan ehtoa $k \leq n$, saamme

$$m \leq \frac{k}{16} \leq \frac{n}{16} < \frac{n}{2}.$$

Siten

$$(8.10) \quad \binom{n}{m} \geq \binom{n}{l} \quad \text{kaikilla } 1 \leq l \leq m,$$

josta saamme edelleen joukkoperheen \mathcal{S}_m alkioden määrälle arvion

$$|\mathcal{S}_m| = \sum_{l=1}^m \binom{n}{l} \leq m \binom{n}{m}.$$

Jäljelle jääneelle binomikertoimelle esitetöjen binäärientropiafunktion yhteydessä esitelty epäyhtälö (2.32) antaa ylärajan

$$\binom{n}{m} \leq e^{m+m \log(n/m)}.$$

Voimme nyt jatkaa väitteemme todennäköisyyden arviointia kohdasta (8.8) yhdistämällä siihen juuri päätelemämme tiedot joukkoperheen \mathcal{S}_m alkioden määrästä:

$$\mathbb{P} \left(\bigcup_{S \in \mathcal{S}_m} E_S^c \right) \leq 2e^{-ck} \cdot m \cdot e^{m+m \log(n/m)} = e^{\log 2 - ck + m + m \log(n/m) + \log m}.$$

Ottamalla puolittain luonnollisen logaritmin ja järjestelemällä uudelleen termejä päädyimme epäyhtälöön

$$(8.11) \quad \log \mathbb{P} \left(\bigcup_{S \in \mathcal{S}_m} E_S^c \right) \leq \log 2 - ck + m \left(\log \frac{n}{m} + 1 + \frac{1}{m} \log m \right).$$

Yllä sekä koko tutkielmassa oletamme logaritmien kantaluvuksi e :n, mikäli sitä ei ole erikseen logaritmin yhteyteen merkitty.

Tarkastelemalla epäyhtälön oikean puolen osaa

$$m \left(\log \frac{n}{m} + 1 + \frac{1}{m} \log m \right)$$

voimme nähdä sen m :n suhteen kasvavaksi funktioksi. Kohdassa $m := 1$ se saa arvon $\log n + 1$. Yhdistämällä nämä havainnot oletukseen mittausmäärän alarajasta huomaamme, että riittävän pienillä $\alpha > 0$ epäyhtälö

$$(8.12) \quad m \left(\log \frac{n}{m} + 1 + \frac{1}{m} \log m \right) \leq (c - \rho)k$$

on voimassa. Oletus mittausmäärän alarajasta varmistaa, että väitteemme pätee edes yhden alkion sisältäville joukoille S .

Käytämme tätä epäyhtälöä kohdan (8.11) arvioon. Tällöin

$$\log \mathbb{P} \left(\bigcup_{S \in \mathcal{S}_m} E_S^c \right) \leq \log 2 - \rho k.$$

Korottamalla n tämän epäyhtälön määäämiin potensseihin ja soveltamalla logaritmin laskusääntöjä

$$\log_b a = \frac{\log_d a}{\log_d b}, \quad \log_b d = \frac{1}{\log_d b}, \quad b^{\log_b a} = a,$$

jossa $a, b, d > 0$ ja $b, d \neq 0$, voimme tehdä päättelyn

$$n^{\log_e \mathbb{P}(\bigcup_{S \in \mathcal{S}_m} E_S^c)} = \left(n^{\log_n \mathbb{P}(\bigcup_{S \in \mathcal{S}_m} E_S^c)} \right)^{\frac{1}{\log_n e}} \leq n^{\log_e 2} n^{-\rho k}.$$

Muokkaamalla vielä termejä päädyimme kohdassa (8.5) vaadittuun todennäköisyyslärajaan:

$$\mathbb{P} \left(\bigcup_{S \in \mathcal{S}_m} E_S^c \right) \leq (n^{\log_e 2} n^{-\rho k})^{\log_n e} = n^{\frac{\log_e 2}{\log_e n}} n^{-\rho \frac{k}{\log_e n}} = 2n^{-\frac{\rho}{\alpha}} = O_\alpha \left(n^{-\frac{\rho}{\alpha}} \right),$$

joka pätee riittävän pienillä $\alpha \geq \beta/k$. \square

Olisimme voineet päätellä UUP:n parametrin β olevan verrannollinen jopa termin $\log n/k$ kanssa, mikäli olisimme tiputtaneet viimeisessä päättelyssä k :n pois ennen α :n käyttämistä. Saatu väite kuitenkin riittää meille, sillä päädyimme joka tapauksessa kyseiseen termiin ERP:n todistuksessa.

Lause 8.13. *Olkoon $R \in \mathbb{R}^{k \times n}$ määritelmän 3.1 mukainen normaalin satunnaismatriisi, joka noudattaa UUP:tä parametrilla $\hat{\beta} := \log n$. Tällöin R noudattaa ERP:tä parametrilla $\beta := \log n$.*

Todistus. Olkoon $S \subset \mathbb{Z}_n$ indeksijoukko, jonka alkioden määrälle pätee UUP:n ja ERP:n määritelmistä tuttu rajoitus

$$|S| \leq \alpha \frac{k}{\log n}.$$

Voimme olettaa vakion $\alpha > 0$ niin pieneksi, että UUP:n mukaisesti pätee arvio

$$(8.14) \quad \mathbb{P}(E_S) \geq 1 - O_\alpha \left(n^{-\frac{\hat{\rho}}{\alpha}} \right),$$

jossa $E_S = \{ \frac{1}{2} \frac{k}{n} \leq \lambda_{\min}(R_{kS}^T R_{kS}) \leq \lambda_{\max}(R_{kS}^T R_{kS}) \leq \frac{3}{2} \frac{k}{n} \}$ ja $\hat{\rho} > 0$. Olkoon vielä $\phi \in \mathbb{R}^n$ jokin joukossa S kannatettu merkkivektori.

Näillä oletuksilla väitteenämme on näyttää, että vähintään todennäköisyydellä $1 - O_\alpha \left(n^{-\frac{\hat{\rho}}{\alpha}} \right)$ on olemassa ERP:n määritelmässä mainitut ominaisuudet (i)-(iii) täyttävä vektori $v \in \mathbb{R}^n$. Asetamme

$$v := R^T R_{kS} (R_{kS}^T R_{kS})^{-1} Q_{nS} \phi$$

ja näytämme kyseisen valinnan täyttävän vaatimukset.

Ominaisuuden (i) mukaisesti vektorille v tulisi päteä yhtenevyys $v(s) = \phi(s)$ kaikilla $s \in S$, minkä voimme ilmaista rajoitekuvauksen avulla muodossa $Q_{nS}v = Q_{nS}\phi$. Saamme yhtäsuuruuden todistetuksi huomaamalla identiteetin

$$(Q_{nS}R^T) = (RQ_{nS}^T)^T = R_{kS}^T,$$

sillä nyt

$$Q_{nS}v = Q_{nS}R^T R_{kS}(R_{kS}^T R_{kS})^{-1}Q_{nS}\phi = (R_{kS}^T R_{kS})(R_{kS}^T R_{kS})^{-1}Q_{nS}\phi = Q_{nS}\phi.$$

Määrittelemme jatkoa varten vektorin

$$(8.15) \quad w := R_{kS}(R_{kS}^T R_{kS})^{-1}Q_{nS}\phi \in \mathbb{R}^k.$$

Voimme w :n avulla esittää vektorin $v = R^T w$ lineaarikombinaationa matriisin R riveistä. Tämä oli vektorilta v vaadittu ominaisuus (ii).

Enää meidän täytyy näyttää, että vektorilla v on ominaisuus (iii) vaaditulla todennäköisyydellä eli

$$\mathbb{P}(\{v(s) \leq \frac{1}{2} \text{ kaikilla } s \in S^c\}) \geq 1 - O_\alpha(n^{-\frac{\rho}{\alpha}}).$$

Yhtäpitävästi voimme todistaa komplementtitapahtumaa koskevan väitteen

$$(8.16) \quad \mathbb{P}(\{v(s) \leq \frac{1}{2} \text{ kaikilla } s \in S^c\}^c) \leq O_\alpha(n^{-\frac{\rho}{\alpha}}).$$

Päästäksemme käsiksi komplementtitapahtumaan määrittelemme

$$v^c := Q_{nS^c}v$$

ja esitämme komplementtitapahtuman sopivammassa muodossa:

$$\begin{aligned} \{|v(s)| \leq \frac{1}{2} \text{ kaikilla } s \in S^c\}^c &= \left(\bigcap_{s=1}^{|S^c|} \{|v^c(s)| \leq \frac{1}{2}\} \right)^c \\ &= \bigcup_{s=1}^{|S^c|} \{|v^c(s)| > \frac{1}{2}\} \\ &= \bigcup_{s=1}^{|S^c|} \{|v^c(s)| > \frac{1}{2}\} \end{aligned}$$

Arvioimme tämän tapahtuman todennäköisyyttä ylöspäin, kunnes päädyimme kohdan (8.16) mukaiseen O -notaatioon. Teemme ensimmäisen arvion jättämällä joukkojen päällekkäisyydet huomioimatta:

$$(8.17) \quad \mathbb{P} \left(\bigcup_{s=1}^{|S^c|} \{|v^c(s)| > \frac{1}{2}\} \right) \leq \sum_{s=1}^{|S^c|} \mathbb{P}(|v^c(s)| > \frac{1}{2}).$$

Jatkaaksemme arviointeja tarvitsemme tietoa vektorin v^c alkioden jakaumista. Esitämme siksi v^c :n matriisiin $R_{kS^c}^T$ ja kohdan (8.15) vektorin w tulona,

$$v^c = Q_{nS^c} v = Q_{nS^c} R^T w = R_{kS^c}^T w.$$

Matriisiin $R_{kS^c}^T$ ja vektorin w alkiot ovat molemmat funktioita matriisiin R riippumattomista normaalijakaumaa $N(0, 1/n)$ noudattavista satunnaismuuttujista. $R_{kS^c}^T$:n ja w :n alkiot eivät kuitenkaan sisällä samoja satunnaismuuttujia, sillä määritelmien-sä mukaisesti $R_{kS^c}^T$ sisältää vain R :n sarakkeiden S^c satunnaismuuttujia ja w vain sarakkeiden S satunnaismuuttujia. Siten $R_{kS^c}^T$ ja w ovat riippumattomia.

Vektorin w kiinnittäminen ei siis vaikuta $R_{kS^c}^T$:n alkioden jakaumaan. Tutkimalla täten vektorin v^c alkioita ehdolla w saamme

$$v^c(s) = \frac{1}{\sqrt{n}} \sum_{l=1}^k Z_{l,s} w_l = \sum_{l=1}^k \frac{w_l}{\sqrt{n}} Z_{l,s},$$

jossa satunnaismuuttujat $Z_{l,s}$ ovat riippumattomia ja noudattavat standardinormaalijakaumaa. Siten ehdollistettuina v^c :n alkiot ovat normaalijakautuneiden satunnaismuuttujien lineaarikombinaatioita eli esitietojen lauseen 2.28 mukaan ne noudattavat normaalijakaumaa

$$(8.18) \quad N\left(0, \frac{\|w\|_{\ell_2}^2}{n}\right).$$

Voidaksemme käyttää tätä ehdollista jakaumaa tulee meidän esittää epäyhtälön (8.17) ylärajan todennäköisyydet ehdollisina. Tähän pääsemme osittamalla todennäköisyysvaruuden w :n suhteen osiin $E_{w_i}, i \in I$, ja esittämällä tapahtumien $\{|v^c(s)| > 1/2\}$ todennäköisyydet summoina $\sum_{i \in I} \mathbb{P}(\{|v^c(s)| > 1/2\} \cap E_{w_i})$. Nämä leikkaukset voimme nimittäin muuttaa ehdollisiksi ehdollisen todennäköisyyden määritelmällä 2.22.

Tarvitsemme siis sopivan osituksen. Koska jakaumassa (8.18) esiintyy w :n ℓ_2 -normi, on luontevaa perustaa ositus siihen. Tutkimalla w :n määritelmää (8.15) voimme huomata pystyvämme rajoittamaan sen ℓ_2 -normia UUP:n antamalla arviolla (8.14) seuraavasti

$$\begin{aligned} \|w\|_{\ell_2} &\leq \|R_{kS}\| \|(R_{kS}^T R_{kS})^{-1}\| \|Q_{nS} \phi\|_{\ell_2} \\ &\leq \sqrt{\frac{3k}{2n}} \quad \frac{2n}{k} \quad \sqrt{|S|} \\ &= \sqrt{\frac{6n|S|}{k}}. \end{aligned}$$

Tätä epäyhtälöä käyttämällä muodostamme yksinkertaisen osituksen joukosta

$$E_w := \{\|w\|_{\ell_2} \leq \sqrt{6n|S|/k}\}$$

ja sen komplementista E_w^c . UUP:n käytöstä seuraa tapahtumalle E_w todennäköisyys-alaraja $1 - O_\alpha(n^{-\hat{\rho}/\alpha})$ ja komplementille E_w^c todennäköisyysyläraja $O_\alpha(n^{-\hat{\rho}/\alpha})$.

Epäyhtälön päättelyissä kahden ensimmäisen normin arviot seuraavat UUP:stä käyttämällä esitiedoissa alkaen kohdasta (2.13) mainittuja matriisinormin ja ominaisarvojen yhteyksiä, sillä

$$\lambda_{\max}(R_{kS}^T R_{kS}) \leq \frac{3k}{2n} \implies \|R_{kS}\| = \sigma_{\max}(R_{kS}) \leq \sqrt{\frac{3k}{2n}}$$

ja

$$\frac{k}{2n} \leq \lambda_{\min}(R_{kS}^T R_{kS}) \implies \|(R_{kS}^T R_{kS})^{-1}\| = \frac{1}{\lambda_{\min}(R_{kS}^T R_{kS})} \leq \frac{2n}{k}.$$

Viimeisessä termissä ℓ_2 -normi otetaan merkkifunktion $|S|$ pituisesta rajoittumasta eli

$$\|Q_{nS}\phi\|_{\ell_2} = \sqrt{\sum_{s \in S} \phi^2(s)} \leq \sqrt{\sum_{s \in S} 1} = \sqrt{|S|}.$$

Pystymme nyt ehdollistamaan ominaisuuden (iii) komplementtitapahtuman todennäköisyyksylärajan (8.17) tapahtumat. Käyttämällä lisäksi yksinkertaista ylärajaa $\mathbb{P}(E_w) \leq 1$ saamme tällöin

$$\begin{aligned} \mathbb{P}\left(\bigcup_{s=1}^{|S^c|} \{|v^c(s)| > \tfrac{1}{2}\}\right) &\leq \sum_{s=1}^{|S^c|} \left(\mathbb{P}(\{|v^c(s)| > \tfrac{1}{2}\} \cap E_w) + \mathbb{P}(\{|v^c(s)| > \tfrac{1}{2}\} \cap E_w^c)\right) \\ &\leq \sum_{s=1}^{|S^c|} \left(\mathbb{P}(|v^c(s)| > \tfrac{1}{2} | E_w) \mathbb{P}(E_w) + \mathbb{P}(|v^c(s)| > \tfrac{1}{2} | E_w^c) \mathbb{P}(E_w^c)\right) \\ &\leq \sum_{s=1}^{|S^c|} \left(\mathbb{P}(|v^c(s)| > \tfrac{1}{2} | E_w) \cdot 1 + 1 \cdot O_\alpha\left(n^{-\frac{\hat{\rho}}{\alpha}}\right)\right). \end{aligned}$$

Ehdollistettuina v^c :n alkiot ovat samoinjakautuneita ja pystymme siten vaihtamaan summan kertolaskuksi saaden

$$\begin{aligned} \mathbb{P}\left(\bigcup_{s=1}^{|S^c|} \{|v^c(s)| > \tfrac{1}{2}\}\right) &\leq |S^c| \left(\mathbb{P}(|v^c(s)| > \tfrac{1}{2} | E_w) + O_\alpha\left(n^{-\frac{\hat{\rho}}{\alpha}}\right)\right) \\ (8.19) \qquad \qquad \qquad &\leq n \mathbb{P}(|v^c(s)| > \tfrac{1}{2} | E_w) + O_\alpha\left(n^{1-\frac{\hat{\rho}}{\alpha}}\right). \end{aligned}$$

Tässä jälkimmäinen termi on lähes haluamaamme kohdan (8.16) muotoa. Keskitytään siis ensimmäisen termin saattaamiseen vastaavaan muotoon.

Käyttämällä tietoa v^c :n alkiodien ehdollisesta jakaumasta (8.18) saamme

$$\mathbb{P}(|v^c(s)| > \tfrac{1}{2} | E_w) = \mathbb{P}\left(\left|\frac{\|w\|_{\ell_2}}{\sqrt{n}} Z\right| > \tfrac{1}{2} | E_w\right) = \mathbb{P}\left(|Z| > \frac{\sqrt{n}}{2\|w\|_{\ell_2}} | E_w\right).$$

Ehdon E_w ollessa voimassa oletuksestamme $|S| \leq \alpha k / \log n$ seuraa w :n ℓ_2 -normille yläraja

$$\|w\|_{\ell_2} \leq \sqrt{\frac{6n|S|}{k}} \leq \sqrt{\frac{6\alpha n}{\log n}},$$

josta edelleen seuraa sisältyvyys

$$\left\{|Z| > \frac{\sqrt{n}}{2\|w\|_{\ell_2}}\right\} \subset \left\{|Z| > \frac{\sqrt{n}}{2} \sqrt{\frac{\log n}{6\alpha n}}\right\} = \left\{|Z| > \frac{1}{2} \sqrt{\frac{\log n}{6\alpha}}\right\}.$$

Siten todennäköisyysmitan monotonisuudesta (2.18) saamme arvion

$$\mathbb{P}(|v^c(s)| > \frac{1}{2} |E_w) \leq \mathbb{P}\left(|Z| > \frac{1}{2} \sqrt{\frac{\log n}{6\alpha}}\right).$$

Koska normaalijakauma on symmetrinen origon suhteen, voimme käsitellä itseisarvon sijaan häntätodennäköisyyksiä

$$\mathbb{P}\left(|Z| > \frac{1}{2} \sqrt{\frac{\log n}{6\alpha}}\right) = 2\mathbb{P}\left(Z \leq \frac{1}{2} \sqrt{\frac{\log n}{6\alpha}}\right).$$

Näille standardinormaalijakauman todennäköisyyksille esitietojen kaava

$$\mathbb{P}(Z \leq z) \leq \frac{1}{z\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \text{jossa } z \leq 0,$$

antaa meidän tapauksessamme arvion

$$\mathbb{P}\left(Z \leq \frac{1}{2} \sqrt{\frac{\log n}{6\alpha}}\right) \leq \frac{2\sqrt{6\alpha}}{\sqrt{2\pi \log n}} e^{\frac{1}{2} \left(\frac{1}{2} \sqrt{\frac{\log n}{6\alpha}}\right)^2} \leq e^{-\frac{1}{48\alpha} \log n} = n^{-\frac{1}{48\alpha}},$$

missä jälkimmäinen epäyhtälö pätee riittävän pienillä $\alpha > 0$.

Saamamme yläraja on nyt ominaisuuden (iii) komplementtitapahtuman todennäköisyysylärajalta vaadittua muotoa, joten voimme yhdistää arviomme kohdan (8.19) arvion kanssa. Tällöin

$$\begin{aligned} \mathbb{P}\left(\bigcup_{s=1}^{|S^c|} \{|v^c(s)| > \frac{1}{2}\}\right) &\leq n \cdot 2n^{-\frac{1}{48\alpha}} + O_\alpha\left(n^{1-\frac{\rho}{\alpha}}\right) \\ &= 2n^{1-\frac{1}{48\alpha}} + O_\alpha\left(n^{1-\frac{\rho}{\alpha}}\right). \end{aligned}$$

Kun tässä $\alpha > 0$ on riittävän pieni, voimme valita vakion $\rho > 0$, jolla pätee

$$\mathbb{P}\left(\bigcup_{s=1}^{|S^c|} \{|v^c(s)| > \frac{1}{2}\}\right) \leq O_\alpha\left(n^{-\frac{\rho}{\alpha}}\right).$$

Olemme näin saaneet ominaisuuden (iii) komplementtitapahtumalle vaaditun todennäköisyysylärajan eli myös itse ominaisuus on vektorilla v vaaditulla todennäköisyydellä. Siten v täyttää ERP:n määritelmässä esitellyn vektorin ominaisuudet eli R noudattaa ERP :tä parametrilla $\beta = \log n$. \square

Luku 9

Ratkaisun esittely numeerisesti

Tutkielman palautuslauseeseen 4.5 tiivistyvän teorian mukaan heikkoon kuuluu $wl_p(L)$ kuuluvan vektorin $x \in \mathbb{R}^n$ ja siitä normaalilla satunnaismatriisilla tehdyistä mittauksista ℓ_1 -minimoinnilla palautetun ratkaisun $\hat{x} \in \mathbb{R}^n$ välinen virhetermi noudattaa ylärajaa

$$\|x - \hat{x}\|_{\ell_2} \leq C_{p,\alpha} \cdot L \cdot \left(\frac{k}{\log n} \right)^{\frac{1}{2} - \frac{1}{p}}$$

vähintään todennäköisyydellä $1 - O_\alpha(n^{-\rho/\alpha})$. Vaikka palautuksen onnistumisen todentaminen on jo itsessään mielenkiintoista, keskitymme tässä luvussa tutkimaan riippuuko virhetermi $\|x - \hat{x}\|_{\ell_2}$ ilmiöön liittyvistä suureista palautuslauseen kuvaamalla tavalla. Siten potentiaalisia tutkimuskohteita ovat alkuperäiseen vektoriin liittyvät vakiot n , L , p sekä mittausmatriisi R ja siihen liittyvät vakiot k , β , ρ , α . Myös virhetermin tarkka yläraja ja toteutumistodennäköisyyden tarkka alaraja ovat mahdollisia tutkimuskohteita. Näiden rajojen tutkimisen hylkäämme kuitenkin siksi, että niiden saavuttamiseksi olemme todistuksissa tehneet karkeita arvioita, joiden seurauksena ainakin vakiotermit ovat kasvaneet todennäköisesti kohtuuttoman suuriksi. Emme myöskään vertaile mittausmatriisien eroja, sillä olemme todistaneet tutkielmassa vain normaalin satunnaismatriisin kelvolliseksi, mikä tekee siitä ainoan varteenotettavan vaihtoehdon testiemme mittausmatriisiksi. Lisäksi se yhdistää palautusongelman tutkielman suurempaan aiheeseen eli satunnaisprojektioon ja sen hyödyntämiseen.

Tutkimuskohteiksi valitsemme heikon kuulun suppenemisparametrin p ja mitausten määrän k . Niiden valitsemista puoltaa molempien termien palautuslauseen väittämä huomattava vaikutus virhetermin suuruuteen: molemmat ovat mukana virhetermiä ylhäältä eksponentiaalisesti rajoittavassa termissä. Tähän luokkaan kuuluisi myös alkuperäisen vektorin ulottuvuus n , joka vaikuttaa virheen lisäksi toteutumistodennäköisyyteen, mutta jätämme n :n testien ulkopuolelle testien laajuuden kurissapitämiseksi. Samasta syystä emme tutki myöskään heikon kuulun säteen vaikutusta virhetermiin.

Ennen testien aloittamista määrittelemme tarkemmin testiasetelman: kaikki testit tulemme suorittamaan numeerisen laskennan kielellä GNU Octave [12]; mittausmatriisina käytämme vain jo mainittua normaalista satunnaismatriisia; testivektorien heikon kuulun säteeksi kiinnitämme $L := 1$; testivektorien ulottuvuuksien mää-

räksi kiinnitämme $n := 300$. Erityisesti viimeisen kiinnityksen pienuus vaatii perusteluja. Valinnan olen tehnyt laskennan raskaudesta johtuen: testit olen ajanut kotikoneellani ja pitääkseni testien suoritusajat kohtuullisina olen jotunut karsimaan testivektorien ulottuvuuksista, jotta testit on voitu suorittaa kohtuullisen suurella määrällä testivektoreja ja useammilla eri suppenemisparametrin arvoilla sekä mitausten määrällä.

Testejä varten täytyy myös päätää, kuinka luomme käytännössä satunnaisia vektoreja valituilla parametreilla heikosta kuulasta. Saadaksemme sopivan otoksen juuri tietyllä suppenemisparametrin p arvolla päädyin vaatimaan testivektoreiden alkioita heikon kuulun määritelmässä esiintyvän ylärajan

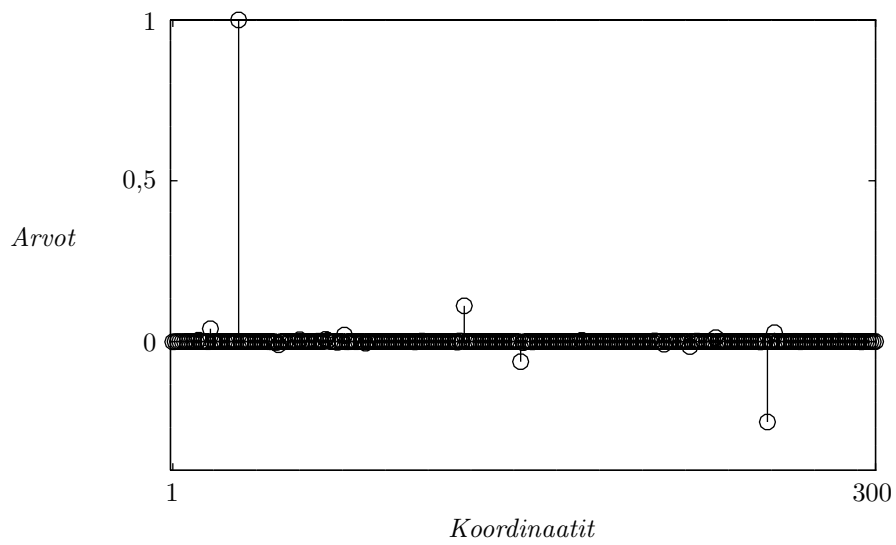
$$|x_{\text{arr}}(l)| \leq L \cdot l^{-\frac{1}{p}}$$

sijasta yhtäsuuruutta sen kanssa. Mikäli sallisimme epäyhtälön, saattaisi arvottu vektori kuulua toteutustavasta riippuen valitun heikon kuulun lisäksi paljon pienisäteisempään ja suppenemisvauhdiltaan nopeampaan kuulaankin. Siten eri p :n arvoilla kehitettyjä testivektoreja ei voitaisi ainakaan pienillä vektorimäärillä niin selvästi vertailla keskenään kuin eksaktia yhtäsuuruutta vaatiessamme. Yhtäsuuruudesta johtuen vektorien luominenkin pysyy yksinkertaisena: meidän riittää kehittää kaikille tietyä suppenemisparametria p noudattaville vektoreille yhteiset arvot sisältävä vektori x_{arr} , josta arpomalla arvoille uudet etumerkit ja indeksit saamme satunnaisen testivektorin. Kirjoittamani Octave-funktio `weaklp.m` luo tällä tavalla $wl_p(L)$ vektoreja ja sen tarkka sisältö löytyy luvusta 11. Esimerkiksi käskyllä `weaklp(n, L, p, randU)`, jossa $n := 300$, $L := 1$ ja $p := 0.5$, saamme kuvan 9.1 vektorin.

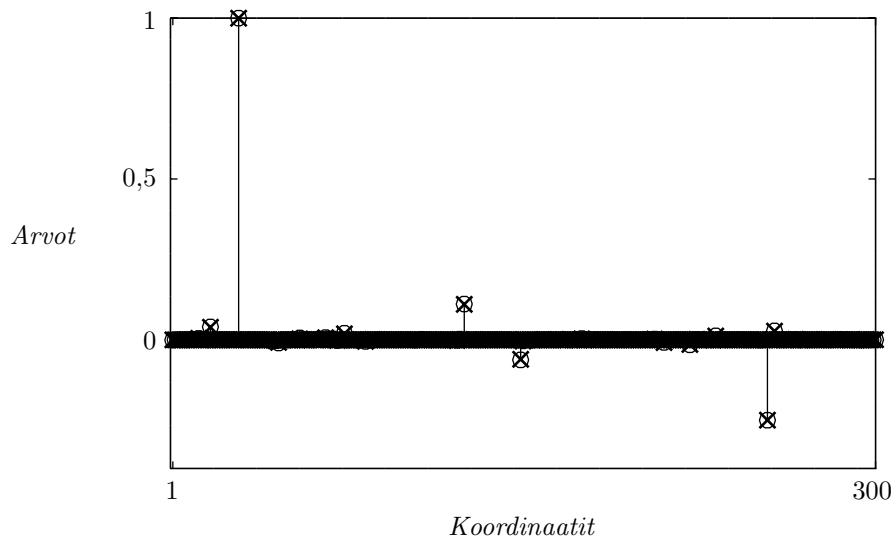
Kokeilemme ennen varsinaisia testejä palauttaa juuri luomamme vektorin: Ensin valitsemme mittausten määräksi 100 ja muodostamme mittaukset tekevän normaalin satunnaismatriisi $R \in \mathbb{R}^{100 \times 300}$. Sitten laskemme mittausten tulokset Rx ja etsimme ratkaisun optimointiongelmaa (4.2). Saamamme ratkaisu on piirretty kuvaan 9.2 ruksein. Palautus näyttäisi onnistuneen hyvin, sillä kaikki ruksit ovat alkuperäisen vektorin palloin kuvattujen arvojen päällä. Virhetermin arvoksi saamme

$$\|x - \hat{x}\|_{\ell_2} \approx 0,008.$$

Vektorin palautuksen tekemiseen käytetyn Octave-tiedoston `mkTesti1.m` sisältö löytyy luvusta 11. Tiedosto käyttää optimointiongelman ratkaisemiseen Emmanuel Candesin ja Justin Rombergin artikkelin [6] yhteydessä saatavaa ratkaisualgoritmia `l1eq_pd.m`. Algoritmi on kirjoitettu MATLAB-kielellä [18], mutta se toimii ilman muutoksia Octavella versiota 3.8.1. Samaa algoritmia `l1eq_pd.m` hyödynnän kaikissa tulevissakin palautuksissa.



Kuva 9.1: Satunnainen vektori $x \in wl_p(1), p := 0,5$ funktiolla *weaklp.m*.
 Funktiokutsu suoritettu tiedostosta *mkTesti1.m*



Kuva 9.2: Palautus sadasta mittauksesta funktiolla *mkTesti1.m*.
 Palautetun vektorin arvot on piirretty ruksein alkuperäisen vektorin arvojen päälle.

Mittausmäärän ja suppenemisparametrin vaikutus keskimääräiseen virheeseen

Saatuamme esimakua teorian toimivuudesta aloitamme suppenemisparametria ja mittausmäärää koskevat testit. Palautuslauseen mukaan suppenemisparametrin pienentäminen ja mittausmäärän kasvattaminen vähentävät virheen suuruutta. Tätä väitettä tulemme testeillämme kokeilemaan.

Valitsemme tutkittaviksi suppenemisparametrin p arvoiksi $\frac{1}{10}, \frac{3}{10}, \frac{6}{10}, \frac{9}{10}$, ja luomme jokaista p kohti 40 vektorin $x_{(m,p)} \in \mathbb{R}^{300}, m \in \mathbb{Z}_{40}$, ryhmän kuulusta $wl_p(1)$. Palautukset teemme jokaiselle vektorille kaikilla mittausmäärillä $k \in \mathbb{Z}_{299}$ ja palautuksessa jokaista paria $(x_{(m,p)}, k)$ kohti käytämme aina uutta mittausmatriisia.

Yksittäisten vektorien virheiden vertailemisen sijaan vertailemme suhteellisten virheiden keskiarvoja eri ryhmien välillä. Siten laskemme jokaista paria (k, p) kohti palautuksessa syntyneiden suhteellisten virheiden keskiarvon ja esitämme se mittausmäärän k funktiona jokaisella p :

$$\text{err}_p(k) := \frac{1}{40} \sum_{m=1}^{40} \frac{\|x_{(m,p)} - \hat{x}_{(m,p,k)}\|_{\ell_2}}{\|x_{(m,p)}\|_{\ell_2}},$$

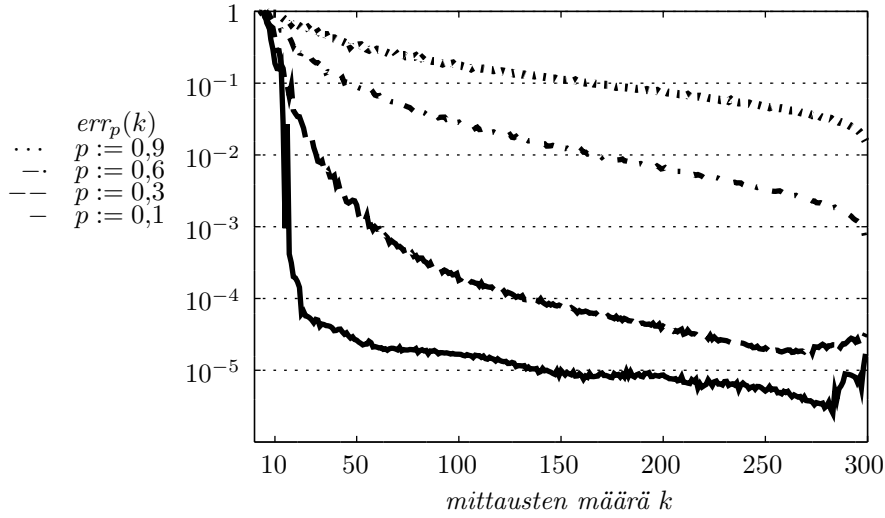
jossa $x_{(m,p)} \in wl_p(1)$ ja $\hat{x}_{(m,p,k)}$ on k :sta mittauksesta $Rx_{(m,p)}$ palautettu vektori. Yllä suhteellisella virheellä tarkoitamme muodostuneen virheen pituuden ja alkuperäisen vektorin pituuden suhdetta

$$\frac{\|x - \hat{x}\|_{\ell_2}}{\|x\|_{\ell_2}},$$

jota käyttämällä normiltaan eri pituisten eli eri p :n arvoilla luotujen vektorien palautuksien onnistuneisuutta on helpompi vertailla keskenään.

Eräällä testikerralla syntyneet virhefunktiot err_p on esitetty kuvassa 9.3. Testin suorittamiseen käytetty tiedosto mkTesti2.m löytyy luvusta 11. Kuvassa virhe on esitetty logaritmisella asteikolla, jotta sen käyttäytyminen tulee paremmin esille myös suuremmilla mittausmäärillä.

Voimme havaita sekä p :n että k :n vaikuttavan palautuksen onnistumiseen, kuten palautuslause 4.5 antoi olettaa: pienillä p ja suurilla k saamme keskimäärin pienempiä virheitä. Erityisesti p näyttäisi määräävän, kuinka pienen mittaluokan virheitä voimme olettaa saavamme mittausten määrän ollessa riittävä: isoilla p virhe kyllä laskee jatkuvasti, mutta ei koskaan pääse samalle tasolle kuin pienemillä p . Muistamalla p :n yhteyden vektorin selvästi nolasta eroavien arvojen määrään ja sitä kautta vektorin harvuuteen voi tässä nähdä yhteyden vektorin harvuuden ja palautuksen tuloksen välillä.



Kuva 9.3: Suhteellisten virheiden keskiarvo mittausmäärän funktiona eri suppenemisparametrien arvoilla. Kuvan tulokset on muodostettu tiedostolla `mkTesti2.m`.

Palautuksen onnistumisen todennäköisyys

Yllä käsitelimme keskimääräisen virheen suhdetta suppenemisparametriin ja mitausten määrään. Seuraavaksi tutkimme vastaavasti palautuksen onnistumistodennäköisyyttä. Luomme siis otoksen vektoreita, palautamme ne ja laskemme, kuinka monen vektorin kohdalla palautus onnistui suhteessa kaikkien vektorien määrään.

Aluksi meidän tulee määritellä mitä tarkoittaa onnistunut palautus. Sovitaan, että kutsumme palautusta onnistuneeksi, kun virheen pituus suhteessa palautettavan vektorin pituuteen on alle prosentin eli

$$\frac{\|x - \hat{x}\|_{\ell_2}}{\|x\|_{\ell_2}} < 0,01.$$

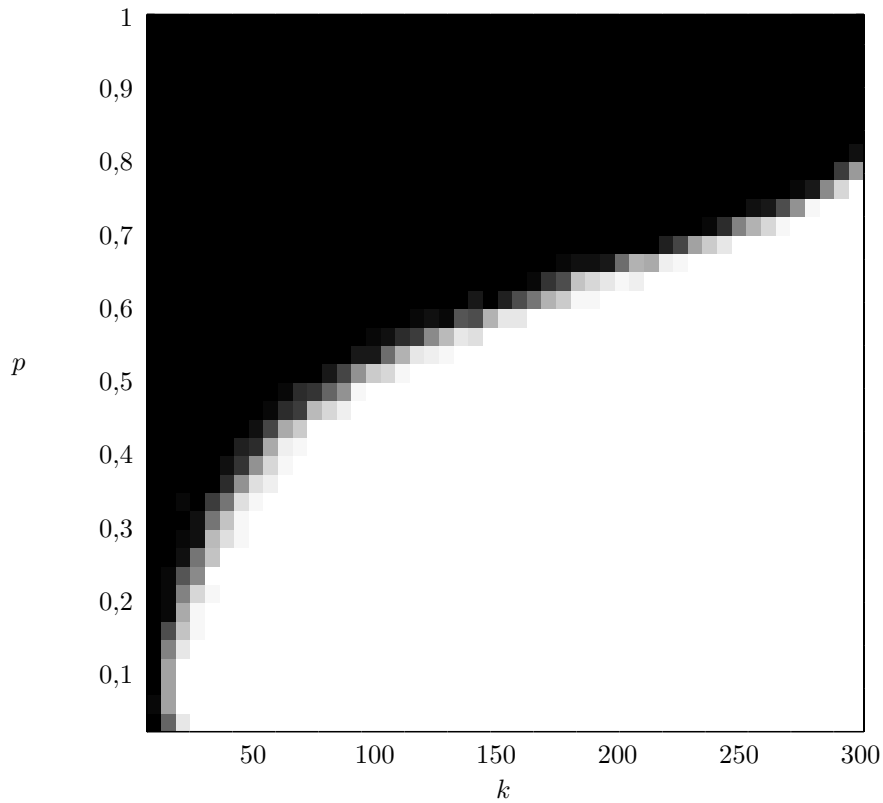
Jotta voisimme verrata onnistumistodennäköisyyttä suhteessa p :n ja k :n vaihteluihin, valitsemme tasaisin välein 40 eri p :n arvoa sekä 50 eri k :n arvoa ja laskemme jokaista paria (p, k) kohti onnistuneesti palautettujen vektorien määrän suhteessa kaikkien vektorien määrään, kun käytämme 30 vektorin otoskoko.

Kuvassa 9.4 on graafinen esitys testin tuloksesta: mitä vaaleampi väri, sitä useampi vektori palautui kyseisillä (k, p) onnistuneesti. Siten puhdas valkoinen väri tarkoittaa kaikkien vektorien palautuneen onnistuneesti ja musta kaikkien palautusten epäonnistuneen. Kuvasta näemme harmaan vyöhykkeen olevan kapea. Sitä on siis myös täydellisen onnistumisen ja epäonnistumisen välinen raja. Juuri tämä täydellinen onnistuminen oikeuttaa artikkelissa [8] todennäköisyydestä $1 - O_\alpha(n^{-\rho/\alpha})$ käytetyn luonnehdinnan ”murskaava”: emme nimittäin näe valkoisella alueella ainoatakaan harmahtavaa neliötä, mikä kielisi todennäköisyyden pettämisestä.

Valitsemamme yhden prosentin suhteellisen virheen raja onnistumisen määritelmässä näyttäisi kuvan perusteella aiheuttaneen tarpeeksi suurilla p kaikkien vektorien palautuksen epäonnistumisen riippumatta tehtyjen mitausten määrästä. Tämä

tukee viime testin lopulla tehtyä huomiota erityisesti $p:n$ vaikuttamisesta saatavan virheen suuruuteen.

Testin suorittavan tiedoston mkTest3.m sisältö on esitelty luvussa 11.



Kuva 9.4: Kuvapisteen (k, p) sävy kertoo onnistettuneesti palautettujen vektorien määrän suhteessa kaikkien vektorin määrään: mitä vaaleampi sävy on sitä useamman vektorin palautus on onnistunut. Kuvan tulokset on muodostettu tiedostolla mkTesti3.m.

Luku 10

Tulosten esittely – Jälkisanat

Satunnaisprojektiosta kertovassa luvussa 3 käytiin läpi aiheen keskeistä teoriaa mukaanlukien projektion havaintojen väliset etäisyydet säilyttävää ominaisuutta. Tästä ominaisuudesta esiteltyä lausetta myös havainnollistettiin kahden eri testidatan tapauksessa ja tuloksia vertailtiin yleisesti käytetyn ulottuvuuksienvähentämismenetelmä PCA:n kanssa. Teorian lisäksi luvussa mainittiin menetelmän sovelluksia sekä kirjallisuutta.

Tutkielman pääaiheena oleva sovellus satunnaisprojektiolle eli Emmanuel Candèsin ja Terence Taon artikkelissaan *Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?* [8] käsittelemä tiivyyden havainnoinnin ongelma esiteltiin luvussa 4. Ongelmana oli niin kutsuttuun heikkoon kuulaan sisältyvän vektorin $x \in \mathbb{R}^n$ palauttaminen siitä tehdyistä k :sta, $k \ll n$, satunnaisesta mittauksesta Rx . Tässä $R \in \mathbb{R}^{k \times n}$ on mittaukset suorittava matriisi, ja vektorin heikkoon kuulaan $wl_p(L)$ sisältyminen taas tarkoittaa, että esitettäessä vektorin x koordinaatit itseisarvoltaan laskevassa järjestyksessä vektorina x_{arr} , suppenevat vektorin arvot kuulan säteen $L > 0$ ja suppenemisparametrin $0 < p < 1$ mukaisesti seuraavaa ylärajaa noudattaen:

$$|x_{\text{arr}}(l)| \leq L \cdot l^{-\frac{1}{p}}, \quad \text{jossa } l \in \{1, \dots, n\}.$$

Ratkaisuksi esiteltiin tulos, jonka mukaan vektorin palauttaminen onnistuu approksimaalisesti niin kutsutulla ℓ_1 -minimoinnilla eli valittaessa ratkaisuksi \hat{x} yhtälön $Rz = Rx$, toteuttavista vektoreista $z \in \mathbb{R}^n$ se, jonka ℓ_1 -normi on pienin. Tuloksessa ei luvata eksaktia palautusta, mutta virheen näytetään noudattavan suurella todennäköisyydellä ylärajaa

$$\|x - \hat{x}\|_{\ell_2} \leq C_{p,\alpha} \cdot L \cdot \left(\frac{k}{\beta}\right)^{\frac{1}{2} - \frac{1}{p}}.$$

Ylärajassa $C_{p,\alpha}$ on alaindeksin termeistä riippuva vakio ja β mittausmatriisikohtainen parametri. Tulosta ei siis todistettu yksittäiselle mittausmatriisityypille, vaan annettiin ominaisuudet, jotka mittausmatriisina käytettävältä matriisilta vaaditaan. Luvussa 5 ominaisuuksien merkitystä analysoitiin tarkemmin ja edelleen luvussa 8 todistettiin normaalilla satunnaismatriisilla olevan nämä ominaisuudet parametrin β arvolla $\log n$. Tällä mittausmatriisivalinnalla teorian toimivuutta esiteltiin numeerisia testejä sisältävässä luvussa 9. Kuten teorian antamasta virheen ylärajasta

voi päätellä, vaikuttaa heikon kuulan suppenemisparametri p huomattavasti virheen suuruuteen ja numeerisissa testeissä saatiinkin hyvin pieniä virheitä sopivan pienillä $p:n$ arvoilla.

Yksinkertaistaakseni sovelluksen teoriaa jätin muutamia artikkelin [8] tuloksia esittämättä. Teoria olisi nimittäin toiminut myös kompleksiarvoisille vektoreille ja heikon kuulan suppenemisparametrin arvolle $p := 1$. Lisäksi artikkelissa todistettiin normaalin satunnaismatriisin ohella kahdelle muullekin satunnaismatriisityypille mittausmatriisilta vaadittavat ominaisuudet: binääriselle otokselle eli matriisille $R \in \mathbb{R}^{k \times n}$, jonka alkiot ovat muotoa $R(i, j) = 1/\sqrt{n} \cdot B_{ij}$, jossa satunnaismuuttujat B_{ij} ovat riippumattomia ja noudattavat jakaumaa

$$B_{ij} \sim \begin{cases} +1, & \text{todennäköisyydellä } 1/2 \\ -1, & \text{todennäköisyydellä } 1/2 \end{cases} ,$$

ja mittausmatriisille, joka saadaan arvottaessa satunnainen määrä satunnaisia rivejä Fourier-matriisista $F \in \mathbb{R}^{n \times n}$,

$$F(s, t) = \frac{1}{\sqrt{n}} e^{-\frac{i2\pi st}{n}} .$$

Artikkelissa sivuttiin myös yleisemminkin tiiviiden havainnoinnin ongelmille keskeistä aihetta: vektoreita, jotka ovat harvoja jossain toisessa kannassa esitettyinä. Tiiviiden havainnoin teoria nimittäin toimii sellaisillekin vektoreille $x \in \mathbb{R}^n$, joille tiedetään ortonormaali kanta $W \in \mathbb{R}^{n \times n}$, jolla vektori Wx on harva (tai tutkielman ongelman yhteydessä $Wx \in wl_p(L)$). Tällaisen harvan kannan löytäminen ei kuitenkaan ole ollenkaan triviaali ongelma. Aihetta on yleisemmin käsitelty artikkelissa [9], kun olemme kiinnostuneet löytämään vektorille kannan, jossa sen ℓ_1 -normi on mahdollisimman pieni.

Luku 11

Testitiedostot

Tämä luku sisältää numeerisen laskennan kielellä GNU Octave [12] kirjoitetut testitiedostot, joita on käytetty luvuissa 3 ja 9 esitetyjen tulosten laskemiseen. Tiedostojen lukemisen helpottamiseksi käyn läpi tiedostoissa käytettyjen symbolien tarkoitukset. Pääosin merkinnät vastaavat suoraan teorialuvuissa käytettyjä merkintöjä. Sulkujen sisään olen merkinnyt satunnaisprojektiomenetelmän esittelyluvun merkityksen, jos se eroaa teorialukujen merkityksestä.

n: testivektorin ulottuvuuksien määrä,

P: sisältää käytettyjen suppenemisparametrien arvot,

p: vastaa yksittäisen suppenemisparametrin arvoa ja siten käytetään myös joukon **P** indeksinä,

L: heikon kuulan $wl_p(L)$ säde,

K: sisältää ne mittauksen määrät, joista palautukset (projisoinnit) tehdään,

M: jokaista suppenemisparametrin arvoa kohti luotavien testivektorien määrä,

m: käytetään joukon **M** indeksinä (testivektorien määrä datassa X),

randU: välin $(0, 1)$ tasajakauman satunnaislukugeneraattorin aloitustila, joka mahdollistaa testin toistamisen samalla lopputuloksella testivektoreiden satunnaisuudesta huolimatta,

randN: standardinormaalijakauman satunnaislukugeneraattorin aloitustila, joka mahdollistaa testin toistamisen samalla lopputuloksella mittausmatriisien satunnaisuudesta huolimatta,

x: yksittäisen saraketestivektorin säilyttämiseksi käytetty muuttuja,

X: kolmiulotteinen $(n \times M \times |P|)$ matriisi sisältäen kaikki käytetyt testivektorit (datamatriisi kokoa $(n \times m)$),

Xp: yksittäisen testivektorin palautukset kaikilla mittausmäärillä tietyllä suppenemisparametrin arvolla.

`err`: muuttujaan lasketaan testissä mitattavat virheet tai suhteelliset virheet,

Funktiota `weaklp.m` käytetään heikon kuulan vektoreiden luomiseen varsinaisissa testitiedostoissa.

```
function [x, xarr] = weaklp(n, L, p, randU)
rand("state", randU);
xarr = L*(1:n).^(-1/p);
sgn = rand(1,n)-0.5; %Satunnaiset merkit x:lle
ind = randperm(n); %Satunnaiset paikat x:lle

x = xarr;
x(sgn<0) = -xarr(sgn<0);
x = x(ind);
x = x';
endfunction
```

Funktio `recover.m` palauttaa n -ulotteisesta syötevektorista tehdyt palautukset satunnaisista normaalista mittauksista. Käytettävät mittausten määrät voidaan erikseen valita. Funktio antaa itse palautuksen lähteen [6] yhteydessä saatavan algoritmin `lleq_pd.m` hoidettavaksi.

```
function X = recover(x,randN,K)
randn("state",randN);
n = length(x);
X = zeros(n,length(K));

for k = 1:length(K)
    R = 1/sqrt(n)*randn(K(k),n); %Satunnaismatriisi
    X(:,k) = lleq_pd(R'*R*x, R, [], R*x, 1e-3);
endfor
endfunction
```

Seuraavat tiedostot sisältävät luvun 9 yksittäisten testien suorittamiseen liittyvän koodin. Nämä tiedostot eivät luo testien yhteydessä näytettyjä kuvia, vaan suoritettuaan testit ne tallentavat tulokset. Näistä talletetuista tuloksista olen erikseen muodostanut testien kuvat.

`mkTesti1.m` suorittaa kuviin 9.1 ja 9.2 liittyvät testit ja tallentaa tulokset.

```
function mkTesti1()
tic;
n=300;
p=0.5;
L=1;
```

```

K=1:(n-1);
randU=10;
randN=1;

%Testivektorin luonti, palautukset ja virheen laskeminen
x      = weaklp(n,L,p,randU);
Xp     = recover(x, randN, K);
xp100  = Xp(:,100);
err    = norm((x*ones(1,length(K))-Xp),2,"cols");
err100 = err(100);

%Testitulosten tallentaminen
kesto = toc
save Testi1.mat n L K p randN randU x Xp xp100 err err100 kesto
endfunction

```

mkTesti2.m suorittaa kuvaan 9.3 liittyvän testin ja tallentaa tulokset.

```

function mkTesti2()
tic;
n = 300;
P = [0.1, 0.3, 0.6, 0.9];
L = 1;
K=1:(n-1);
M = 40;
randU = 20;
randN = 2;
rand("state" , randU);
randn("state", randN);

%Testivektorien luonti matriisiin X:
X=zeros(n, M, length(P));
for p = 1:length(P)
    for m = 1:M
        X(:,m,p)=weaklp(n,L,P(p),rand("state"));
    endfor
endfor

%Vektorien projisointi, palautus, sitten virhe matriisiin err:
err = zeros(length(P),length(K));
for m=1:M
    for p=1:length(P)
        x=X(:,m,p);
        Xp=recover(x, randn("state"),K);
        err(p,:) = err(p,)+norm(x*ones(1,length(K))-Xp,2,"cols")/norm(x);
    endfor
endfor

```

```

endfor
err = err/M;

%Tulosten tallentaminen
kesto = toc
save Testi2.mat n P L K M randU randN X err kesto
endfunction

```

mkTesti3.m suorittaa kuvaan 9.4 liittyvän testin ja tallentaa tulokset.

```

function mkTesti3()
tic;
n = 300;
P = 0.01:0.025:1; %length(P)=40
L = 1;
M = 30;
K = 1:6:300; % length(K)=50
randU = 30;
randN = 3;
rand("state" , randU);
randn("state", randN);

%Testivektorien generointi matriisiin X:
X=zeros(n, M, length(P));
for p = 1:length(P)
    for m = 1:M
        X(:,m,p)=weaklp(n,L,P(p),rand("state"));
    endfor
endfor

%Vektorien projisointi, palautus ja tn. laskeminen matriisiin err:
err=zeros(length(P),length(K));
for p =1:length(P)
    for m = 1:M
        x = X(:,m,p);
        Xp= recover(x, randn("state"),K);
        err(p,:) = err(p,.)+((norm((x*ones(1,length(K))-Xp),2,"cols")...
            /norm(x))<0.01);
    endfor
endfor
err = err/M;

%Testitulosten tallentaminen:
kesto = toc
save Testi3.mat n P L M K randU randN X err kesto
endfunction

```


Seuraavaa tiedostoa mkRPvsPCA.m on käytetty luvun 3 testitulosten laskemiseen. Tiedosto ei siis luo testien yhteydessä näytettyjä kuvia vaan tallentaa testitulokset muistiin. Parametrina tiedosto ottaa vastaan tiedon lasketaanko ilmastodataa ($temp := 1$) vai heikon kuulun dataa ($temp \neq 1$). Ilmastodatan tilanteessa data haetaan tiedostosta temp2Data.mat. Tämä ilmastodata on edelleen luotu tiedostolla mkTemp2Data.m alkuperäisestä lähteen [15] tiedostosta mil0010_EH5_mm_TEMP2_1-14472.nc.

```
function mkRPvsPCA(temp)
tic;
n=4608;
m=1000;
K=1:5:1500;

if(temp==1)
    load temp2Data
else
    X = zeros(n,m);
    for l=1:m;
        X(:,l)=weaklp(n,100,0.8,1);
    endfor
endif
[U, L, V] = svd(X);

randn("state", 4608);
rand("state", 1000);

errRPavg = zeros(1,n);
errPCavg = zeros(1,n);

for k=K;
    R=(1/sqrt(n))*randn(k,n);
    Uk=U(:,1:k);
    RP=R*X;
    PC=Uk'*X;

    ind = randperm(m);
    Xdist = X(:,ind);
    Xnorm = norm(Xdist-[Xdist(:,m), Xdist(:,1:(m-1))],2,"cols");

    RPdist = RP(:,ind);
    RPnorm = norm(RPdist-[RPdist(:,m), RPdist(:,1:(m-1))], 2,"cols");
    RPerr = abs( (Xnorm - sqrt(n/k)*RPnorm)./Xnorm );

    PCdist = PC(:,ind);
    PCnorm = norm(PCdist-[PCdist(:,m), PCdist(:,1:(m-1))], 2,"cols");
```

```

PCerr = abs( (Xnorm - PCnorm)./Xnorm );

errRPavg(k) = mean(RPerr);
errPCavg(k) = mean(PCerr);
endfor
kesto = toc;
if(temp==1)
    save RPsvsPCAtemp.mat X n m K L errRPavg errPCavg kesto;
else
    save RPsvsPCAweak.mat X n m K L errRPavg errPCavg kesto;
endif
endfunction

function mkTemp2Data()
tic;
pkg load octcdf
%Datan temp('time','lat','lon') hakeminen, jossa
%      (14472)* (48)* (96)= 66686976 alkiota
nc = netcdf('mil0010_EH5_mm_TEMP2_1-14472.nc','r');
longitudes = nc{'lon'}(:);
temp2 = nc{'temp2'}(:);

%Valitaan viimeiset 1000 kuukautta ja muunnetaan
%3-ulotteinen (time x latitude x longitude) matriisi
%2-ulotteiseksi (location x time) matriisiksi
temp2 = temp2(13473:14472,:,:);
tempAPU=temp2;
temp2(:,:, 1:48)=tempAPU(:,:,49:96);
temp2(:,:,49:96)=tempAPU(:,:, 1:48);
Time = size(temp2,1); %=1000
X=[];
for l=1:length(longitudes)
    X=[X,temp2(:,:,l)];
endfor
X = X';

%Datan tallentaminen
kesto = toc;
save temp2Data.mat X kesto
endfunction

```

Lähdeluettelo

- [1] Achlioptas, D. 2003. *Database-friendly random projections: Johnson-Lindenstrauss with binary coins*. Journal of Computer and System Sciences 66, 671–687.
- [2] Aoki, M. 1998. *New approaches to macroeconomic modeling: evolutionary stochastic dynamics, multiple equilibria, and externalities as field effects*. Cambridge. Cambridge University Press.
- [3] Bingham, E. & Mannila, H. 2001. *Random Projection in Dimensionality Reduction: Applications to Image and Text Data*. Teoksessa Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD '01. New York (NY). ACM.
- [4] Bryan, K. & Leise, T. 2013. *Making Do with Less: An Introduction to Compressed Sensing*. SIAM Review 3, 547-566.
- [5] Cadima, J. & Jolliffe, I. 2009. *On Relationships Between Uncentred And ColumnCentred Principal Component Analysis*. Pakistan Journal of Statistics 4, 473–503.
- [6] Candes, E. & Romberg, J. 2005. *ℓ_1 -MAGIC: Recovery of Sparse Signals via Convex Programming*. Viitattu 23.5.2014. <http://users.ece.gatech.edu/justin/l1magic/>.
- [7] Candes, E., Romberg, J. & Tao T. 2006. *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*. IEEE Transactions on Information Theory 2, 489-509.
- [8] Candes, E. & Tao, T. 2006. *Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?*. IEEE Transactions on Information Theory 12, 5406–5425.
- [9] Chen, S., Donoho, D. & Saunders, M. 1998. *Atomic decomposition by basis pursuit*. SIAM Journal on Scientific Computing 1, 33–61.
- [10] Dasgupta, S. & Gupta, A. 1999. *An elementary proof of the Johnson-Lindenstrauss lemma*. Technical Reprint 99-006, UC Berkeley.

- [11] Demšar, U., Harris, P., Brunson, C., Fotheringham, S. & McLoone, S. 2013. *Principal Component Analysis on Spatial Data: An Overview*. Annals of the Association of American Geographers 1, 106–128.
- [12] Eaton, J., Bateman, D. & Hauberg, S. 2009. *GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations*. CreateSpace Independent Publishing Platform.
- [13] Goel, N., Bebis, G. & Nefian, A. 2005. *Face recognition experiments with random projection*. Teoksessa Biometric Technology for Human Identification II. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series.
- [14] Johnson, W. & Lindenstrauss, J. 1984. *Extensions of Lipschitz mappings into a Hilbert space*. Contemporary mathematics, 189–206.
- [15] Jungclaus, J. 2008. MPI-M Earth System Modelling Framework, millennium full forcing experiment (ensemble member 1). World Data Center for Climate. CERA-DB "mil0010". <http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=mil0010>.
- [16] Kurimo, M. 1999. *Indexing Audio Documents by using Latent Semantic Analysis and SOM*. Teoksessa Kohonen Maps. Elsevier.
- [17] Li, W., Bebis, G. & Bourbakis, N. 2004. *Integrating Algebraic Functions of Views with Indexing and Learning for 3D Object Recognition*. Conference on Computer Vision and Pattern Recognition Workshop. IEEE.
- [18] MATLAB. 2010. Natick (MA). The MathWorks Inc.
- [19] Qaisar, S., Bilal, R.M., Iqbal, W., Naureen, M. & Sungyoung L. 2013. *Compressive sensing: From theory to applications, a survey*. Journal of Communications and Networks 5, 443–456.
- [20] Seitola T., Mikkola V., Silen J. & Järvinen H. *Random projections in reducing the dimensionality of climate simulation data*. Hyväksytty julkaistavaksi lehdessä Tellus A.
- [21] Sipser, M. 2006. *Introduction to the Theory of Computation*. Boston (MA). Thomson Course Technology.
- [22] Strang, G. 2009. *Introduction to linear algebra*. Wellesley (MA). Wellesley-Cambridge Press. 4. painos.
- [23] Szarek, S. 1991. *Condition numbers of random matrices*. Journal of Complexity 2, 131–149.
- [24] Varmuza, K., Filzmoser, P. & Liebmann, B. 2010. *Random projection experiments with chemometric data*. Journal of Chemometrics 3–4, 209–217.

- [25] Tao, T. Sähköposti, joka sisältää vastauksen Visa Mikkolan esittämään kysymykseen artikkelista [8]. Vastaanotettu 6.8.2014.
- [26] Vempala, S. 2004. *The Random Projection Method*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science Volume 65.
- [27] Venkatesh, S. 2013. *The theory of probability*. Cambridge. Cambridge University Press.
- [28] Väisälä, J. 2007. *Topologia I*. Helsinki. Limes ry. 4. korjattu painos.