Dobbeltstokastiske matriser
og
polytoper
og
majorisering

Geir Dahl

Norsk matematisk forening
29. mars 2001

Plan:

- dobbeltstokastiske matriser
- transportproblemet, polyedre
- Birkhoff polytopen
- majorisering: begrep og noen resultater
- Robin Hood !
- et par anvendelser
- egne bidrag

Mål:

- gi et lite innblikk i feltet ved å presentere noen sentrale begreper og resultater
- unngå tekniske ting
- vekt på tidlige resultater

Stokastiske og dobbeltstokastiske matriser

Stokastisk matrise:

- IKKE: matrise som er en stokastisk variabel
- En stokastisk matrise er en $m \times n$ matrise som er ikkenegativ og der alle radsummer er 1

$$\mathbf{M} = \begin{bmatrix} 0.3 & 0.2 & 0 & 0.5 \\ 0.6 & 0.3 & 0 & 0.1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

- Markov matrise
- Markov kjeder, statistiske eksperimenter

Dobbeltstokastisk matrise: en ikkenegative matrise der alle rad- og kolonnesummer er 1. Må være kvadratisk.

$$A = \begin{bmatrix} 0.3 & 0.2 & 0 & 0.5 \\ 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.1 & 0.4 & 0 & 0.5 \end{bmatrix}$$

- Romanovsky (1935): *stokastisk*
- Feller (1950): *dobbeltstokastisk*
- bistokastisk, Schur transformasjon

Noen klasser av dobbeltstokastiske matriser

- Matrisen av orden n med alle elementer lik $1/n$.
- Permutasjonsmatriser: kvadratisk $(0,1)$-matrise der hver linje (rad eller kolonne) inneholder nøyaktig én 1'er.

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

  Det er $n!$ permutasjonsmatriser.
- Ortostokastiske matriser: $A = [a_{i,j}]$ der $a_{i,j} = q_{i,j}^2$ og $Q = [q_{i,j}]$ er ortogonal. (Tilsvarende for Q unitær.)
- Sirkulantmatriser (visse):

$$\begin{bmatrix} c_0 & c_1 & c_2 \\ c_2 & c_0 & c_1 \\ c_1 & c_2 & c_0 \end{bmatrix}$$

  der c er stokastisk (dvs. c ikkenegativ og med sum 1).
- Latinske kvadrater (visse)

Et sentralt resultat

Teorem. [*Birkhoff (1946), von Neumann (1953)*]:

En matrise A er dobbeltstokastisk hvis og bare hvis A kan skrives som en konveks kombinasjon av permutasjonsmatriser:

Altså:

$$A = \sum_{j=1}^{N} \lambda_j P_j$$

der $P_1, \ldots, P_N$ er permutasjonsmatriser og

$$\lambda_j \geq 0 \quad (j \leq N), \qquad \sum_{j=1}^{N} \lambda_j = 1.$$

Birkhoff's teorem er et eksempel på et dekomposisjonsteorem for matriser.

Eksempel.

$$A = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.7 & 0.3 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix} =$$

$$0.3 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + 0.5 \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + 0.2 \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Skal snart:

- mengden av dobbeltstokastiske matriser: geometri
- se på et bevis for Birkhoff's teorem
- hvor mange permutasjonsmatriser kan vi klare oss med?

Transportproblemet

En enkel optimeringsmodell for transport.

Gitt:

- m lagre $i = 1, 2, \ldots, m$
- n kunder $j = 1, 2, \ldots, n$
- lager i har a_i enheter av en vare
- kunde j ønsker b_j enheter av varen
- det koster $c_{i,j}$ kr pr. varenhet som fraktes fra lager i til kunde j

Finn en transportplan, dvs. hvor mye som skal fraktes fra hver kunde til hvert lager, slik at

- alle behov dekkes og lagerkapasiteter overholdes og
- slik at den totale transportkostnad er minst mulig.

Må ha $\sum_i a_i \geq \sum_j b_j$ og kan faktisk anta likhet her:

$$\sum_i a_i = \sum_j b_j$$

Da benytttes hver lagerkapasitet fullt ut.

Formulerer transportproblemet som et lineær optimeringsproblem:

minimer $\sum_{i,j} c_{i,j} x_{i,j}$

forutsatt at

(i) $\sum_{j=1}^{n} x_{i,j} = a_i \quad (i \leq m)$

(ii) $\sum_{i=1}^{m} x_{i,j} = b_j \quad (j \leq n)$

(iii) $x_{i,j} \geq 0 \qquad (i \leq m, j \leq n).$

En løsning x her kan vi oppfatte som en $m \times n$ matrise. En tillatt løsning (matrise) oppfyller kravene (i), (ii) og (iii).

Altså: de tillatte løsningene er ikkenegative $m \times n$ matriser med radsummer $a_1, a_2, \ldots, a_m$ og kolonnesummer $b_1, b_2, \ldots, b_n$. Disse utgjør en mengde $\mathcal{T}$ i $\mathbb{R}^{m \times n}$.

Hvis $m = n$ og hver a_i og hver b_i er lik 1, vil $\mathcal{T}$ være mengden av alle dobbeltstokastiske matriser.

Hva kan vi si om generelt om strukturen til $\mathcal{T}$?

Algoritme for transportproblemet?

Men hvor blir det av

ROBIN HOOD

????

Litt om Polytoper og polyedre

Et polyeder P (i $\mathbb{R}^n$) er løsningsmengden til et system av endelig mange lineære ulikheter:

$$P = \{x \in \mathbb{R}^n : Ax \leq b\}$$

der A er en $m \times n$ matrise og $b \in \mathbb{R}^m$ (og $\leq$ betyr komponentvis $\leq$). Altså: snitt mellom endelig mange halvrom

$$\{x \in \mathbb{R}^n : d_1 x_1 + \cdots d_n x_n \leq c\}.$$

Polyedre er spesielle konvekse mengder.

En polytop Q (i $\mathbb{R}^n$) er den konvekse innhylningen til en endelig mengde:

$$Q = \text{conv}(\{y_1, y_2, \ldots, y_r\})$$

Så: Q består av alle mulige konveks kombinasjoner av vektorene (punktene) $y_1, y_2, \ldots, y_r$.

Ekvivalent: Q er snittet av alle konvekse mengder som inneholder $y_1, y_2, \ldots, y_r$.

Teorem. [*Weyl (1935) Minkowski (1896)*]: En mengde P er en polytop hvis og bare hvis P er et begrenset polyeder.

Generalisering: [*Motzkin (1936)*]:

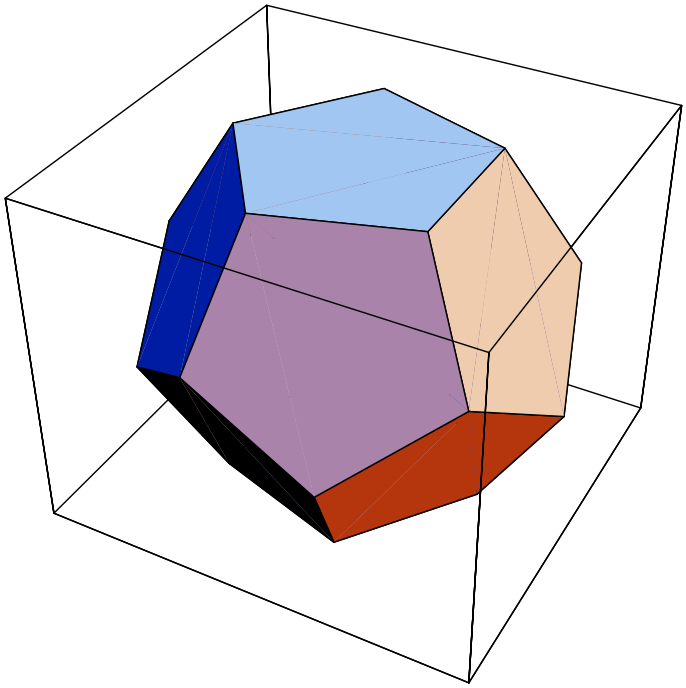
polyeder = polytop + endelig generert konveks kjegle.

La P være et polyeder. Da kalles $x \in P$ et ekstrempunkt i P hvis x ikke er midtpunkt mellom to andre punkter i P.

Dermed: en polytop er den den konvekse innhylningen av sine ekstrempunkter.

(Krein-Milman: tilsvarende for kompakte konveks mengder)

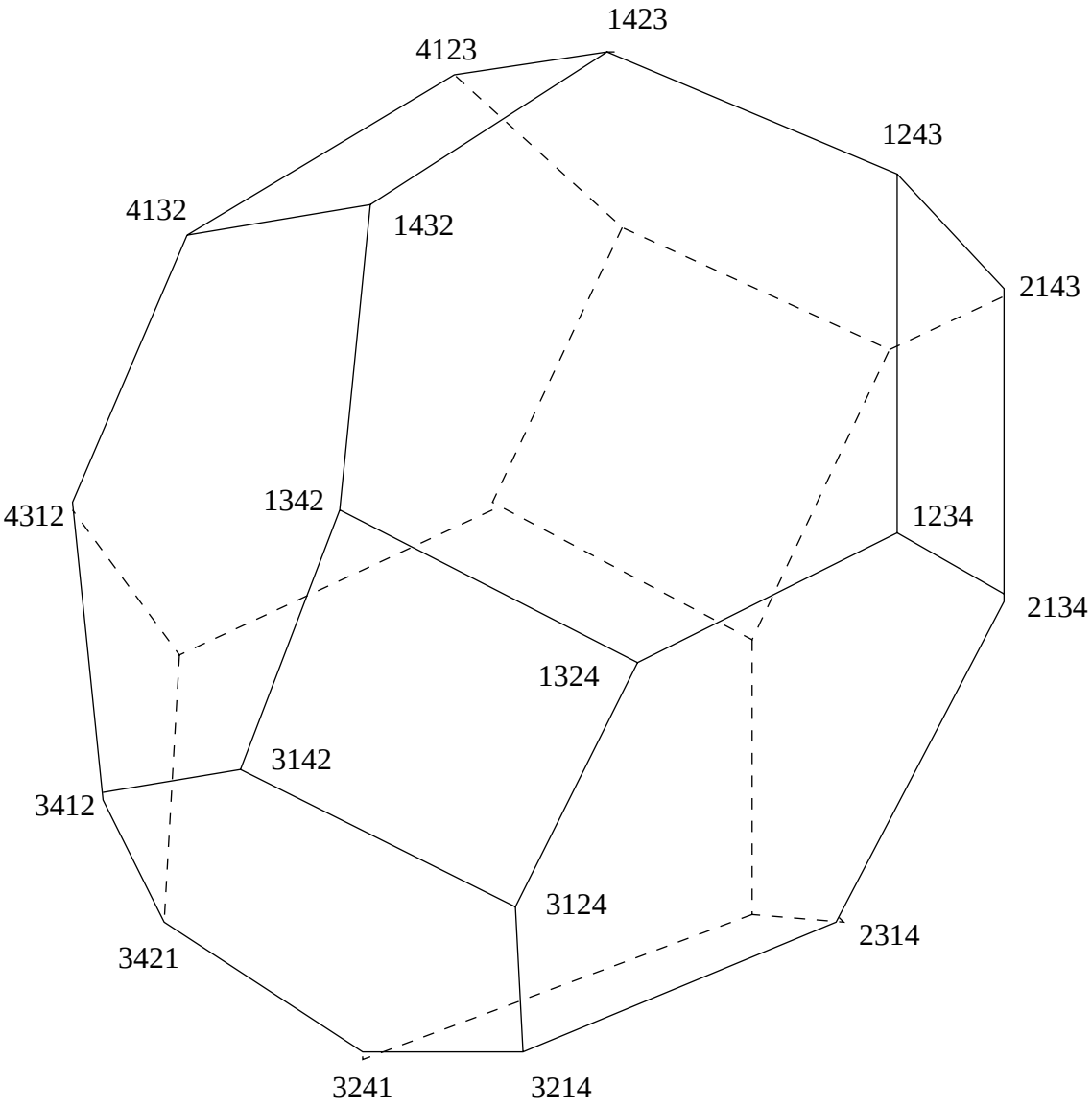
Dodekaeder



Euler's polyederformel: $v - e + f = 2$

- $v = 20$ ektrempunkter (hjørner)
- $e = 30$ kanter
- $f = 12$ fasetter

Permutaeder



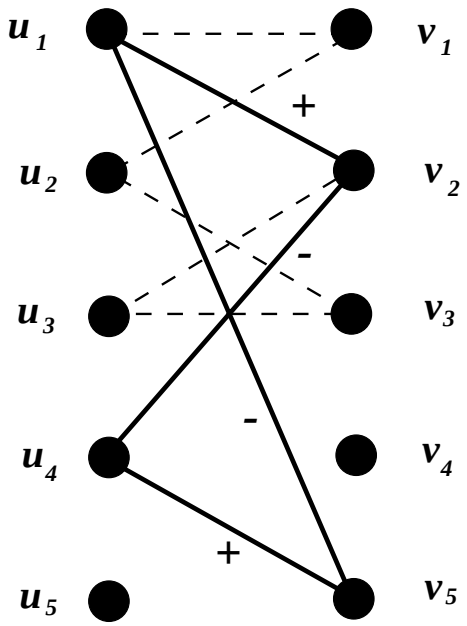
Birkhoff polytopen

$$\Omega_n = \{A \in \mathbb{R}^{n\times n} : A \text{ er dobbeltstokastisk}\}$$

Begrenset polyeder og derfor polytop. *Skal finne ekstrempunktene.*

Observasjon: de *heltallige* dobbeltstokastiske matrisene er permutasjonsmatrisene. Så: nok å vise at ethvert ekstempunkt i Ω_n er heltallig.

Anta A ikke heltallig. Definer en bipartitt graf med noder $u_1, \ldots, u_n$ og $v_1, \ldots, v_n$ og en kant $[u_i, v_j]$ hvis $0 < a_{i,j} < 1$.

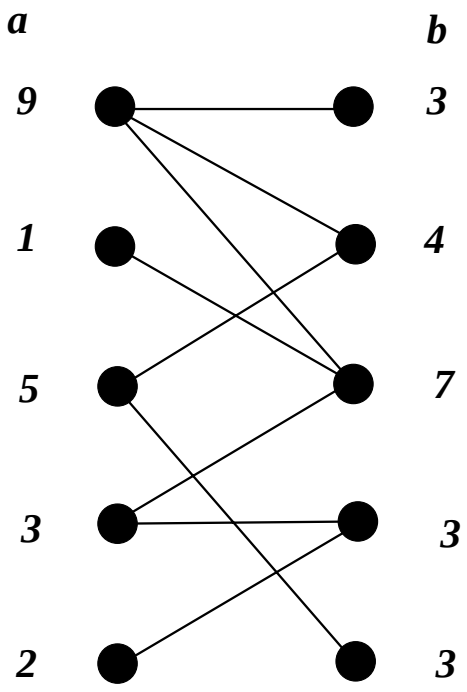


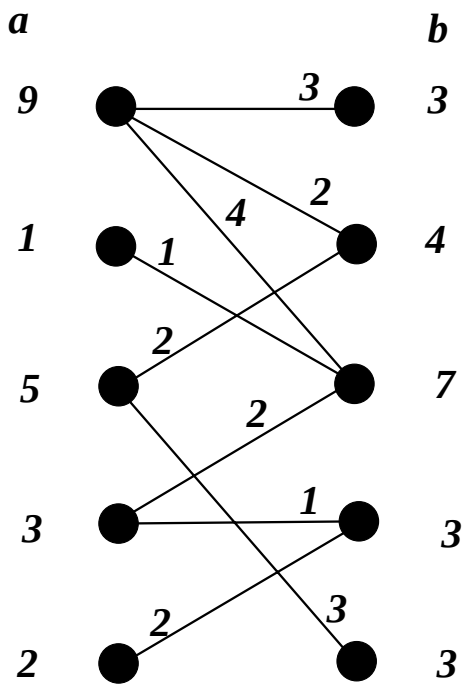
$$A = (1/2)(A + \epsilon D) + (1/2)(A - \epsilon D)$$

Dette viser Birkhoff's teorem.

Tilsvarende får vi at ekstrempunktene i transportpolytopen $\mathcal{T}$ svarer til (visse) delgrafer av G, nemlig spenntrær (sammenhengende og utspenner alle nodene, og ingen syklus).

Dette gir den kombinatoriske strukturen til ekstrempunktene, selv om man ikke finner dem eksplisitt.





$$\begin{bmatrix} 3 & 2 & 4 & & \\ & & 1 & & \\ & 2 & & & 3 \\ & & 2 & 1 & \\ & & & 2 & \end{bmatrix}$$

Dette gir strukturen til $\mathcal{T}$ og kan brukes til å gi en effektiv algoritme for transportproblemet. (Basert på dualitetsteori. Faktisk kan duale variable, én for hver node i grafen, beregnes fra spenntreet på en enkel måte.)

Tilbake til Birkhoff polytopen Ω_n:

La $A \in \Omega_n$. Så $A = \sum_{j=1}^{N} \lambda_j P_j$.

Hvor liten kan N velges?

$$N \leq (n-1)^2 + 1$$

Birkhoff algoritmen: finner en dekomposisjon.

Carathéodory's teorem.

Har noen sett noe til

ROBIN HOOD

????

Majorisering

Majorisering er en (pre-)ordning av vektorer som er nært knyttet til dobbeltstokastiske matriser.

La x og y være to vektorer i $\mathbb{R}^n$ og la $x_{[j]}$ betegne j'te største komponent i x. Så $x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}$.

Vi sier at x er majorisert av y dersom

$$\sum_{j=1}^{k} x_{[j]} \leq \sum_{j=1}^{k} y_{[j]} \qquad (k \leq n-1)$$

$$\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j$$

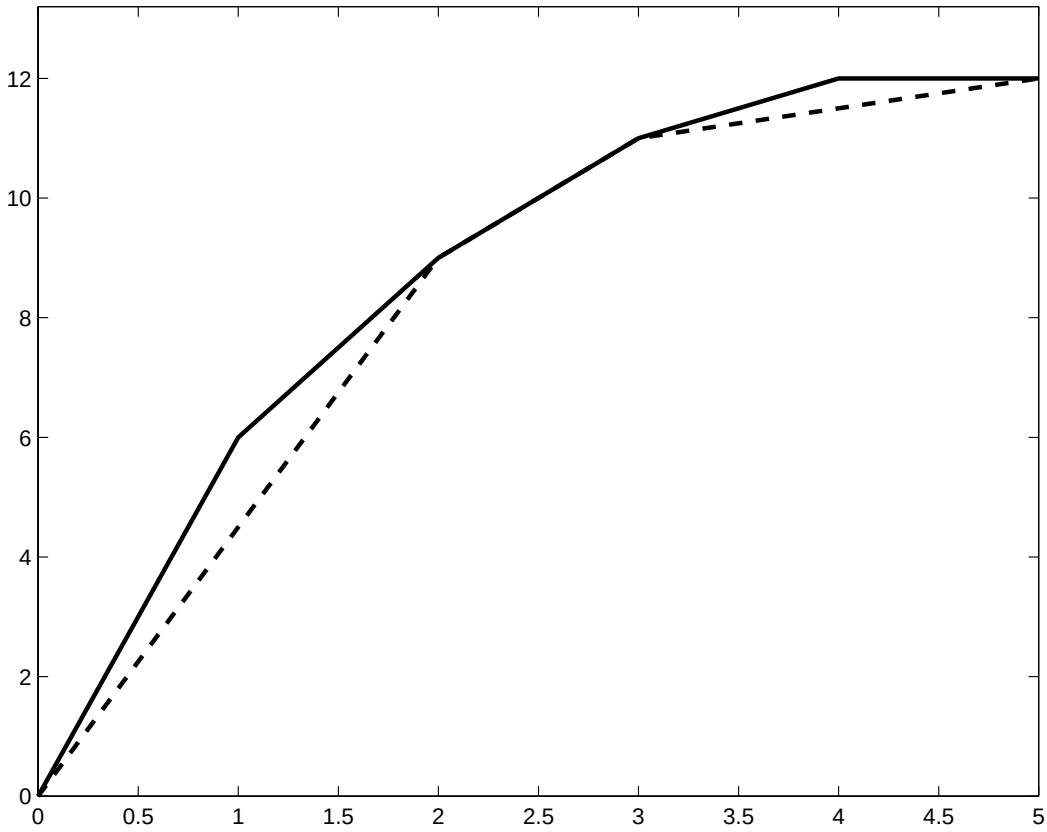
og vi skriver $x \prec y$.

Så $\prec$ er binær relasjon på $\mathbb{R}^n$ og

- preordning: (i) $x \prec x$, (ii) $x \prec y$, $y \prec z$ medfører $x \prec z$.
- permutasjonsinvariant: hvis $x \prec y$ så er også enhver permutasjon av x majorisert av enhver permutasjon av y.
- ikke partiell ordning på $\mathbb{R}^n$
- men partiell ordning på restriksjon $\{x \in \mathbb{R}^n : x_1 \geq x_2 \geq \cdots x_n\}$. Faktisk: lattice.

Diagram: Lorentz kurve

Lineær interpolant til punktene $(k, \sum_{j=1}^{k} x_{[j]})$ for $k = 1, 2, \ldots, n$.



$x \prec y$ hviss Lorentz kurven til x ligger under Lorentz kurven til y.

I økonomi: majorisering brukt til å måle ulikhet i inntektsfordelinger i populasjoner. Lorentz (1905).

Lattice-operasjonene kan tolkes i diagrammet.

Karakteriseringer av $x \prec y$?

Teorem. [*Hardy, Littlewood, Polya (1929)*]: x er majorisert av y hvis og bare hvis det fins en dobbeltstokastisk matrise A slik at

$$x = Ay$$

Hvorfor?

Observasjon: For gitt vektor y er mengden

$$\{z : z \prec y\}$$

konveks. Faktisk et polyeder (hvorfor?).

Hvis A er dobbeltstokastisk, så er pga. Birkhoff's teorem A en konveks kombinasjon av permutasjonsmatriser

$$A = \sum_j \lambda_j P_j$$

Så hvis $x = Ay$ blir

$$x = (\sum_j \lambda_j P_j)y = \sum_j \lambda_j (P_j y)$$

Men $P_j y \prec y$ og $\{z : z \prec y\}$ er konveks, så da må $x \prec y$.

Hva med omvendingen?

Endelig: Robin Hood !!!!!!!!

Repetisjon: *Ta fra de rike og gi til de fattige!*

Robin Hood transformasjon: La Q være permutasjonsmatrisen til en transposisjon (bytter to koordinater) og la $0 \leq \lambda \leq 1$. Kaller

$$T = \lambda I + (1-\lambda)Q$$

en Robin Hood transformasjon (eller T-transform).

Merk: T er dobbeltstokastisk. Og T ligger på en kant som er nabo med hjørnet I i Birkhoff polytopen.

$$\begin{aligned} x &= (\ x_1,\ x_2\ ,\ x_3,\ x_4\ ) \\ Tx &= (\ x_1,\ (4/5)x_2 + (1/5)x_4\ ,\ x_3, (4/5)\ x_4 + (1/5)x_2\ ) \end{aligned}$$

De nye komponentene vil være nærmere hverandre.

Som vist: $Tx \prec x$.

Anta $x \prec y$ og at x, y begge er ikkevoksende. Velg j maximal med $x_j < y_j$ og velg $k > j$ minimal med $x_k > y_k$. Så

$$y_j > x_j \geq x_k > y_k$$

Robin Hood transformasjon: overføre mest mulig fra y_j til y_k, men slik at resultatet y^* oppfyller $x \prec y^*$.

Så

$$x \prec y^* \prec y.$$

og enten er $y_j^* = x_j$ eller så er $y_k^* = x_k$. Kan gjenta prosessen og etter høyst $n-1$ iterasjoner har vi kommet til x.

Har vist: Hvis $x \prec y$, så kan x nåes fra y ved et endelig antall (faktisk $\leq n-1$) Robin Hood transformasjoner.

Og teoremet til Hardy, Littlewood, Polya (1929) følger da. Fordi produktet av dobbeltstokastiske matriser er dobbeltstokastisk.

Teorem. [*Schur (1929), Hardy, Littlewood, Polya (1929)*] x er majorisert av y hvis og bare hvis for hver konveks funksjon $g : \mathbb{R} \to \mathbb{R}$ så holder

$$\sum_{j=1}^{n} g(x_j) \leq \sum_{j=1}^{n} g(y_j)$$

Eksempel: $g(x) = -\log(x)$ gir $(1/n)\sum_i x_i \geq \prod_i x_i^{1/n}$: aritmetisk-geometrisk ulikhet.

En funksjon $\phi : \mathbb{R}^n \to \mathbb{R}$ kalles Schur-convex hvis

$$x \prec y \Rightarrow \phi(x) \leq \phi(y)$$

Ved å vise en viss majorizering og velge en passende Schur-convex funksjon kan man få interessante ulikheter. Se Marshall, Olkin: "Inequalities: Theory of Majorization and its applications" (1979).

Anvendelse i spektralteori: sammenheng mellom diagonalelementer og egenverdier.

Teorem. [*Schur (1923)*] La H være en reell symmetrisk (evt. Hermittisk) matrise med diagonal elementer $h_1, h_2, \ldots, h_n$ og egenverdier $\lambda_1, \lambda_2, \ldots, \lambda_n$. Da er

$$(h_1, h_2, \ldots, h_n) \prec (\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Så: summen av de k største diagonalelementene er høyst lik summen av de k største egenverdiene. Og $trH = \sum_j \lambda_j$.

Bevis: Ved spektralteoremet er $H = Udiag(\lambda)U^T$ som gir $h = A\lambda$ der $A = [u_{i,j}^2]$ er dobbeltstokastisk (ortostokastisk).

"Beste resultat": Omvending ved Horn (1954).

Anvendelser innen

- matriseteori: egenverdier, singulære verdier
- kombinatorikk: f.eks eksistens at grafer med gitte grader (evt. heltallige matriser med gitte linjesummer), farvelegging av grafer, turneringer.
- numerisk analyse: matriseapproksimasjon
- statistikk: multivariabel analyse, pålitelighetsteori, køteori, fordelingsfunksjoner.
- ...

Egen innsats her: studere spørsmål knyttet til majorisering og polyedre:

- generalisering av majoriseringsbegrepet
- struktur på polyedre knyttet til en majorisering
- kombinatoriske egenskaper
- optimering over polyedre definert via majorisering
- lattice egenskaper
- dobbeltstokastiske matriser der linjene er majorisert av (evt. majoriserer) en gitt vektor

- R.A. Brualdi, G.Dahl: "Majorization-constrained doubly stochastic matrices" (2001?)
- G. Dahl, "Principal majorization ideals and optimization" (2001)
- G. Dahl, "A note on diagonally dominant matrices" (2000)
- G. Dahl, "Majorization polytopes" (1999)
- G. Dahl, "Matrix majorization" (1999)
- G. Dahl, "Majorization, polyhedra and statistical testing problems" (1998).
- G. Dahl, F. Margot, "Weak k-majorization and polyhedra", (1998)
- G. Dahl, "Polyhedra and optimization in connection with a weak majorization ordering" (1994)

De fleste i tidskriftet *Linear Algebra and Its Applications*.

Generaliseringer, beslektet:

- Erik Torgersen: “Comparison of statistical experiments”. Cambridge (1991). Eget kapittel om majorisering for familier av mål.
- (?) utvidelser til W^* og C^* algebraer: Alberti, Uhlmann: “Stochasticity and partial order. Doubly stochastic maps and unitary mixing”. Erling Størmer: “Positive linear maps of C*-algebras”. Springer (1974).