

# Noen anvendelser av lineæralgebra

John Aslak Wee Kleven

13. februar 2023

I denne artikkelen skal vi se på noen anvendelser av lineæralgebra man kan gjøre på egen datamaskin: Hvordan projeksjonsformelen for indreproduktrom kan bli brukt til å tilnærme funksjoner, og SVD for minste kvadraters metode og for å komprimere bilder. Leseren oppfordres til å eksperimentere selv med det som står her.

Det som gjøres i her lærte jeg i MA1202 Lineær algebra og anvendelser (projeksjonsformelen og minste kvadraters med normallikningen), og TMA4145 Lineære Metoder (SVD og minste kvadraters-anvendelse). Læreboka i førstnevnte emne var [4]. I sistnevnte emne brukte vi forelesers notater [3] for det vi skal gjøre i denne artikkelen.

## 1 Projeksjon og tilnærming

I et indreproduktrom  $V$  over  $\mathbb{F}$  (hvor  $\mathbb{F}$  er  $\mathbb{R}$  eller  $\mathbb{C}$ ), har vi følgende projeksjonsformel:

**Teorem 1.1.** La  $V$  være et indreproduktrom og  $\{e_i\}_{i=1}^n$  en endelig ortonormal mengde. La  $M = \text{span}\{e_i\}_{i=1}^n$ . Da vil projeksjonen av  $x$  ned på  $M$ , for oss gitt ved<sup>1</sup>

$$P_M(x) = \sum_{i=1}^n \langle x, e_i \rangle e_i$$

være den unike vektoren i  $M$  som er nærmest  $x^2$ .

*Bevis.* Se [1, s. 196–198]. Beviset er elegant, men er ikke hva denne artikkelen handler om.  $\square$

For mengden  $C([a, b], \mathbb{C})$  av kontinuerlige funksjoner på formen  $f : [a, b] \rightarrow \mathbb{C}$ , et vektorrom, er følgende et indreprodukt:

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx$$

---

<sup>1</sup>Egentlig er det definert som at hvis  $x = a + b$  hvor  $a \in M$  og  $b$  er ortogonal på alle vektorene i  $M$ , så vil  $P_M(X) = a$ . Denne definisjonen viser seg å være veldefinert, og man kan vise at med denne definisjonen av  $P_M$  så vil vi få formelen som brukes i teoremet. Dette vises blant annet i [1, s. 196].

<sup>2</sup>Her defineres avstanden mellom to funksjoner  $f$  og  $g$  som  $\|f - g\|$ , hvor  $\|f\| = \sqrt{\langle f, f \rangle}$ , og  $\langle \cdot, \cdot \rangle$  er indreproduktet.

som vi kan bruke for projisering. Normen blir

$$\|f\| = \int_a^b |f(x)|^2 dx,$$

og avstanden mellom to vektorer  $f$  og  $g$  blir

$$\|f - g\| = \int_a^b |f(x) - g(x)|^2 dx.$$

Så hvis  $\{e_i\}_{i=1}^n$  er  $\{\sin(x), \cos(x)\}$ , kan vi ortonormalisere mengden slik at vektorene blir enhetsvektorer som står ortogonale vektorer. Da vil formelen over gi oss  $\alpha$  og  $\beta$  i  $\mathbb{C}$  slik at

$$\int_a^b |f(x) - \alpha \sin(x) - \beta \cos(x)|^2 dx$$

er minst, for hvilken som helst vektor  $f$  i  $C([0,1], \mathbb{C})$ . Geometrisk sett er dette det samme som å finne  $\alpha$  og  $\beta$  slik at  $\alpha \sin(x) + \beta \cos(x)$  er den beste mulige tilnærmingen til  $f(x)$  på  $[0, 2]$ .

Her er en skisse for hvordan man kan beregne dette. Prøv å implementere dette selv, for eksempel i Python. Hvis du vil ha eksakte, og ikke bare numeriske verdier må du ha et bibliotek eller denslags som kan regne integraler eksakt. For eksempel kan du bruke SageMath eller scipy.

Gitt  $\{e_i\}_{i=1}^n$  som en mengde av tilfeldige vektorer (med ingen krav på seg!)

1. Ortonormaliser mengden med Gram-Schmidt-algoritmen<sup>3</sup>, og
2. Bruk formelen over.

**Eksempel 1.2.** La  $f(x) = e^x$  på intervallet. Finn  $\alpha$  og  $\beta$  slik at

$$\alpha + \beta x + \gamma x^2$$

er en best mulig tilnærming til  $f$  på intervallet  $[0, 2]$ .

Siden vi er i  $\mathbb{R}$ -land blir indreproduktet

$$\langle f, g \rangle = \int_0^2 f(x)g(x) dx.$$

Vi begynner med mengden

$$\{1, x, x^2\}$$

(hvor disse er kortform for *funksjoner*, ikke tall!). Gram-Schmidt gir følgende ortonormale mengde ut fra den over:

$$\left[ \frac{1}{2} \sqrt{2}, \quad \frac{3}{2} \sqrt{\frac{2}{3}}(x-1), \quad \frac{5}{4} \sqrt{\frac{2}{5}}(3x^2 - 6x + 2) \right] \quad (1)$$

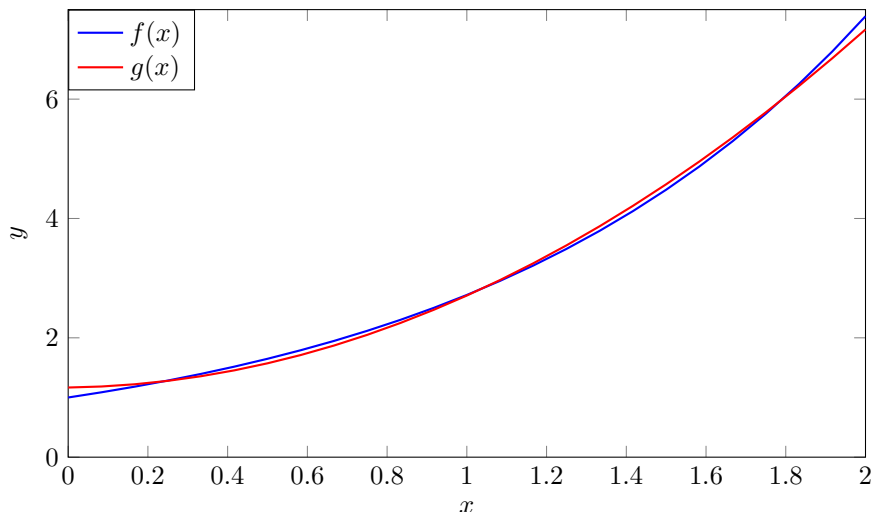
Altså er disse vektorene enhetsvektorer i indreproduktet  $\langle \cdot, \cdot \rangle$ , og de står alle ortogonalt på hverandre. Da blir tilnærmingen

$$\begin{aligned} & \left( \frac{15}{4}e^2 - \frac{105}{4} \right) x^2 + \left( -\frac{15}{2}e^2 + \frac{111}{2} \right) x + (3e^2 - 21) \\ & \approx 1.4590 x^2 + 0.0821 x + 1.1672 \end{aligned}$$

<sup>3</sup>For mer om dette, se en lærebok i lineær algebra, eller sjekk Wikipedia.

Kall denne tilnærmingen for  $g(x)$ .

For praktiske formål er det ingen grunn til å beregne (1) eksakt. Visuelt blir det slik som i figur 1.



Figur 1: Tilnærming av  $f(x) = e^x$  med  $g(x)$  gitt av teorem 1.1, altså  $g(x) \approx 1.4590x^2 + 0.0821x + 1.1672$ .

Ganske bra! Dette er bedre enn det et Taylorpolynom av samme grad kan få, siden Taylorpolynomet “fokuserer” på å få så bra nøyaktighet som mulig i ett punkt, mens her prøver vi å få en tilnærming som er så bra som mulig uniformt på et intervall. For å sitere [1, s. 200] har vi klart å bruke lineær algebra til å gjøre noe bedre enn det vi lærte i analysen!

## 2 Singulærverdidekomposisjon (SVD)

Når man først lærer om matriser og ulike måter å dekomponere dem, så ser man kun på dekomponeringer på kvadratiske matriser. For eksempel diagonalisering, ortogonal diagonalisering og Jordan Normal form. Matrisedekomponeringen vi skal se på nå, er mer generell i det at den kan bli gjort på alle matriser av alle størrelser.

Først, notasjon og noen begreper. Gitt en kompleks matrise  $A \in \mathbb{C}^{m \times n}$ , la

$$A^* := \overline{A}^T$$

være den konjugattransponerte til  $A$ . Hvis  $A$  er reell er dette bare den transponerte til  $A$ . Videre, en kvadratisk matrise  $U \in \mathbb{C}^{n \times n}$  er *unitær*<sup>4</sup> hvis  $UU^* = U^*U = I$ , altså hvis  $U$  er inverterbar og  $U^{-1} = U^*$ .

Singulærverdidekomposisjon lyder som følger:

**Teorem 2.1.** La  $A \in \mathbb{C}^{m \times n}$  være en matrise. Da fins unitære  $U \in \mathbb{C}^{n \times n}$  og  $V \in \mathbb{C}^{m \times m}$ , og en diagonalmatrise  $\Sigma \in \mathbb{R}^{m \times n}$  slik at  $A = U\Sigma V^*$ .

<sup>4</sup>I det reelle tilfellet kalles disse matrisene *ortogonale*, da radene er *ortonormale*, og samme med søylene.

Den geometriske tolkningen til at  $U$  og  $V$  er unitære, er at de bevarer lengder, og kan dermed kun rotere eller flippe akser.

Ved å se på algoritmen som beregner  $U, V$  og  $\Sigma$  ser man at hvis  $A$  er reell, så er  $U$  og  $V$  også det. Det gir den reelle versjonen av teoremet:

**Teorem 2.2.** La  $A \in \mathbb{R}^{m \times n}$  være en matrise. Da fins ortogonale  $U \in \mathbb{C}^{n \times n}$  og  $V \in \mathbb{R}^{m \times m}$ , og en diagonalmatrise  $\Sigma \in \mathbb{R}^{m \times n}$  slik at  $A = U\Sigma V^T$ .

Merk at  $\Sigma$  ikke er en kvadratisk matrise, så med “diagonal” menes at hele matrisen er null, utenom det på diagonalen som begynner øverst til venstre i matrisen. Elementene på diagonalen til  $\Sigma$  alle reelle, positive ( $\geq 0$ ). De ikkenull elementene på diagonalen kalles de *singulære verdiene* til  $A$ . De er unike uavhengig av hva slags singulærverdidekomposisjon man kan ha på  $A$  (dekomposisjonen er ikke unik!<sup>5</sup>), og er kvadratrøttene til egenverdiene til matrisene  $AA^*$ , som også er de samme som kvadratrøttene til egenverdiene til  $A^*A$ . Begge disse matrisene har bare reelle og positive egenverdier siden de er *normale*<sup>6</sup>.

Altså er  $\Sigma$  på formen

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_r & & & & \\ & & & 0 & & & \\ & & & & & \ddots & \\ & & & & & & \ddots \end{pmatrix}$$

hvor

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$$

er de singulære verdiene til  $A$ .

### 3 Minste kvadraters metode

I denne seksjonen vil  $\|\cdot\|$  være den euklidiske normen, altså den vanlige normen på  $\mathbb{C}^n$  og  $\mathbb{R}^n$ .

Man lærer vanligvis å løse minste kvadraters-problemet ved å løse den såkalte normallikningen

$$A^*Ax = A^*y$$

når man vil finne den  $x \in \mathbb{C}^n$  slik at  $\|Ax - y\|$  er minst.

Vi kan bruke SVD til å få en teoretisk overlegen løsning til problemet. Prisen vi betaler for dette er at det er mer kostbart å beregne SVD-er enn å løse likningen over, som kan blant annet bli løst ved QR-faktorisering<sup>7</sup>.

<sup>5</sup>Hvis man har  $A = U\Sigma V^*$  ser man at  $AA^* = U\Sigma\Sigma^*U^*$ , og tilsvarende for  $A^*A$ . Så SVD vil gi diagonaliseringer av  $AA^*$  og  $A^*A$ , men i en diagonalisering vil elementene på diagonalen nettopp være egenverdiene til matrisen, som er unike.

<sup>6</sup>En matrise  $X \in \mathbb{C}^{n \times n}$  er normal hvis  $XX^* = X^*X$ .

<sup>7</sup>Se [2, s. 78] for mer.

### 3.1 Pseudoinverser

Til vanlig er inverse matriser bare definert for kvadratiske matriser, fordi ellers kan man ikke gange fra begge sider. For matriser av alle størrelser kan vi definere en slags “psevdo-invers” som har mye av den samme oppførelsen som inverse matriser. Den egentlige definisjonen til denne psevdoinversen er ikke interessant for oss, men her er én måte å beregne den på:

**Teorem 3.1.** La  $A \in \mathbb{C}^{m \times n}$  ha SVD-faktorisering  $A = U\Sigma V^*$ , og la

$$\Sigma^+ = \begin{pmatrix} \frac{1}{\sigma_1} & & & & & \\ & \frac{1}{\sigma_2} & & & & \\ & & \ddots & & & \\ & & & \frac{1}{\sigma_r} & & \\ & & & & 0 & \\ & & & & & \ddots \end{pmatrix}$$

Da er *Moore-Pensore*-psevdoinversen til  $A$  gitt ved

$$A^+ = V\Sigma^+U^*.$$

Merk at siden SVD gir reelle dekomponering til reelle matriser, så vil dette gi at psevdoinversen til en reell matrise, er reell.

### 3.2 Pseudoinverser og minste kvadraters

For oss er psevdoinverser nyttige på grunn av følgende teorem. Det vil, i alle tilfeller, gi oss den “beste mulige” løsningen til  $Ax = b$  hvor  $A$  er en ikke-kvadratisk matrise.

**Teorem 3.2.** La  $A \in \mathbb{C}^{m \times n}$ , og betrakt likningssettet  $Ax = b$ .

1. Hvis likningssettet har nøyaktig én løsning, så vil  $z = A^+b$  være den.
2. Hvis likningssettet har uendelig mange løsninger, så vil  $z = A^+b$  være den løsningen med minst norm<sup>8</sup>.
3. Hvis likningssettet har ingen løsninger, så vil  $z = A^+b$  være den beste mulige løsningen i den forstand at

$$\|Az - b\| \leq \|Ay - b\| \quad \text{for hver } y \in \mathbb{C}^n.$$

Det er punkt 3 som lar oss bruke psevdoinverser til å løse “minste kvadraters”-problemstillinger, og punkt 2 er en forbedring ovenfor den naive løsningen av normallikningen som fås av å invertere  $A^*A$ , siden denne matrisen er ikke alltid inverterbar<sup>9</sup>. For å få til det samme med normallikningen vil vi trenge en metode for å finne løsningen av  $Ax = b$  med minst norm, når det er uendelig mange løsninger:

For mer om Moore-Pensore-psevdoinverser og teoremet over, se [3].

<sup>8</sup>Geometrisk sett blir dette den nærmest origo

<sup>9</sup>For anvendelsene til minste kvadraters er den imidlertid inverterbare ganske ofte. For å være presis er den det hvis og bare hvis  $\text{rank}(A)$  er maksimal, se for eksempel [4, s. 362].

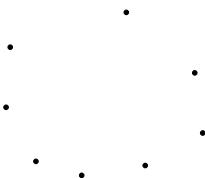
### 3.3 Bruk av minste kvadraters

Når man lærer om minste kvadraters metode, så er det nesten alltid i kontekst av lineær regressjon, og regresjon av andregraspolynomer. Vi kan gjøre *mye mer* enn dette. Husk minste kvadraters-problemet: *Gitt*  $A \in \mathbb{R}^{m \times n}$  *og*  $y \in \mathbb{R}^m$ , *finn*  $x \in \mathbb{R}^n$  *slik at størrelsen*  $\|Ax - y\|$  *er minst*. Regresjon blir opprinnelig involvert ved å finne på en lur måte å konstruere  $A$  på, og den samme teknikken kan man bruke på “annen regresjon” enn som så.

La oss si at vi er gitt følgende punkter i  $\mathbb{R}^2$ :

$$\begin{aligned} &(-1.68, 1.02), (1.39, -1.94), (-3.20, -0.34), (-3.41, -1.25), \\ &(1.89, -2.11), (-0.21, -2.37), (-1.00, -2.00), (-1.79, -0.57) \end{aligned}$$

Hvis man plotter disse punktene, vil de minne om en ellipse:



Figur 2: Minner om en ellipse?

Vi ønsker å finne ellipsen som best passer disse punktene. Likningen til ellipser som man ser først er noe i duren av

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

Denne likningen kan bli omformulert til å være mer av hva vi vil ha, men den kan ikke produsere roterte ellipser. Vi har lyst på de, siden det vil kunne gi flere frihetsgrader og (dermed) ellipser som passer dataen bedre. Så vi ser på likningen for en tovariabel kvadratisk form:

$$ax^2 + by^2 + cxy + dx + ey = 1$$

Spesialtilfeller av denne er ellipser, parabler og hyperbler, samt rotasjoner av disse. Siden vi skal bruke minste kvadraters på denne blir vi nødt til å håpe at likningen vi får ender opp med å bli en ellipse, og ikke noe annet<sup>10</sup>.

Hvis vi indekserer punktene våre  $(x_n, y_n)$ , ønsker vi altså å finne  $a, b, c, d$  og  $e$  slik at

$$ax_n^2 + by_n^2 + c_nxy + d_nx + e_ny \simeq 1$$

og at dette er en så god tilnærming som mulig for så mange  $n$  som mulig. (Den geometriske tolkningen til dette er at ellipsen går gjennom punktene  $(x_n, y_n)$ ). En måte å oppnå dette på er om

<sup>10</sup>I en mer virkelig anvendelse, for eksempel i et tegneprogram som gjenkjenner brukerens forsøk på sirkler/ellipser og erstatter dem med perfekte sirkler/ellipser, ville man nok sjekket at koeffisientene  $a, b, c, d$  og  $e$  virkelig lager en ellipse før de blir brukt til noe.

$$\begin{pmatrix} ax_1^2 + by_1^2 + c_1x_1y_1 + d_1x + e_1y \\ ax_2^2 + by_2^2 + c_2x_2y_2 + d_2x + e_2y \\ \vdots \\ ax_n^2 + by_n^2 + c_nx_ny_n + d_nx + e_ny \end{pmatrix} \simeq \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

som vektorer. Vi gjenkjenner høyresida som en matrisemultiplikasjon, og ser da at vi ønsker å finne  $a, b, c, d$  og  $e$  slik at

$$\begin{pmatrix} x_1^2 & y_1^2 & x_1y_1 & x_1 & y_1 \\ x_2^2 & y_2^2 & x_2y_2 & x_2 & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & y_n^2 & x_ny_n & x_n & y_n \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} \simeq \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

Så hvis vi lar  $A$  være matrisen,  $b$  være vektoren med bare énere i seg (venstresida over), og  $x$  være vektoren med koeffisientene  $a, b, c, d$  og  $e$ , så bør

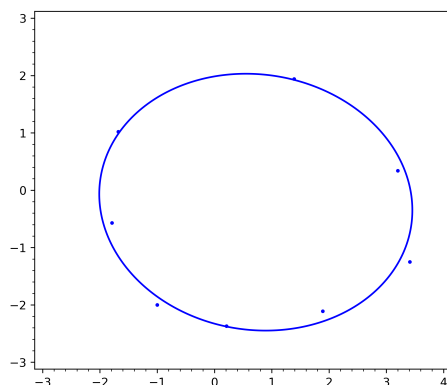
$$\|Ax - b\|$$

være liten hvis  $x$  skal gi en god ellipse-tilnærming. Men dette er et problem vi kan løse ved minste kvadraters!

For dataen gitt over, hvis vi setter opp matrisen  $A$  slik som gjort over, så vil minste kvadraters-tilnærmingen være, med forbehold om avrunding,

$$0.14532x^2 + 0.21640y^2 + 0.02232xy - 0.20461x + 0.07348y = 1.$$

Hvis vi plotter dette inn sammen med punktene over, får vi en ganske fin ellipse:



## 4 Bildekomprimering

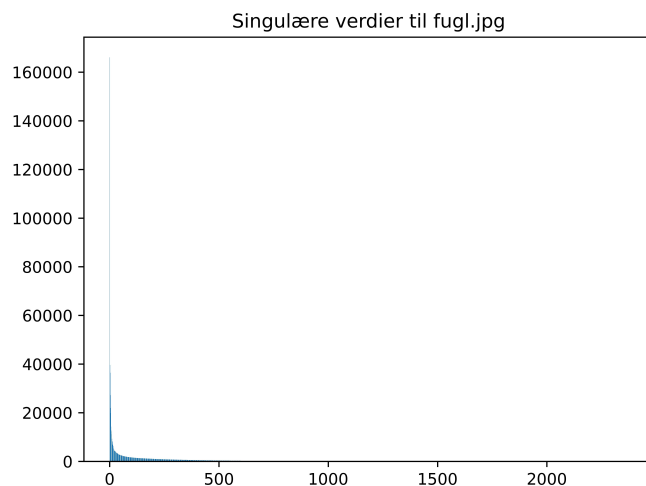
Vi avslutter dette notatet ved å involvere bilder. En bilde er bare en matrise (i hver fargekanal), så intuitivt bør vi kunne matrisedekomponere bilder

Så anta at vi har et bilde  $A \in \mathbb{R}^{m \times n}$ , og vi singularverdidekomponerer det:

$$A = U\Sigma V^T.$$

Vi antar at vi har konstruert  $A$  slik at alle elementer i  $A$  er mellom 0 og et eller annet tall, for eksempel 1 eller 255.

Her er et plott av sinulgærverdiene til et bilde jeg fant på nett:



Vi ser at det er noen få singulære verdier som er veldeig store. I rekkefølge er de de singulære verdiene omtrent lik

166082, 39599, 36446, 27205,  
21922, 19876, 14255, 12775,

og de minste er

0.194, 0.166, 0.142, 0.090,  
0.079, 0.060, 0.024, 0.005

De største singulære verdiene bør være de som bidrar mest til lysstyrken i  $A$  (siden de er størst), så hvis vi fjerner flere av de minste singulære verdiene i  $\Sigma$ , så bør den nye matrisen ha så mange nuller at vi kanskje slipper å lagre hele  $U$  og  $V$ , samtidig som det ikke burde endre bildet for mye.

Hvis man tenker litt på hvordan matrisemultiplikasjonen utføres på  $U\Sigma V^T$ , så ser man at hvis  $\Sigma$  inneholder  $r$  singulære verdier (ikke nødvendigvis alle!), så trenger man bare  $r$  rader/kolonner fra  $U$  og  $V^T$ . Man kan finne en fin illustrasjon av det her: <https://timbaumann.info/svd-image-compression-demo/>.

Så hvis  $A$  er en  $m \times n$ -matrise med  $r$  singulære verdier, og vi kapper av de unødvendige delene av  $U$  og  $V^T$ , så vil hele SVD-en kreve at vi lagrer

$$mr + r + nr = r \cdot (1 + m + n)$$

flyttall.

Dette er bra! Vi kan skalere andelen singulære verdier vi ønsker å beholde, og kaste bort resten for å få en god tilnærming til matrisen.



## 5 Avsluttende ord

Én annen referanse for SVD-bilde-greiene er <https://zerobone.net/blog/cs/svd-image-compression/>.

Kode for temaene diskuter her, kan man finne her: [https://folk.ntnu.no/jakleven/tma4106\\_linalg\\_anvend\\_kode.zip](https://folk.ntnu.no/jakleven/tma4106_linalg_anvend_kode.zip).

## 6 Referanser

- [1] Sheldon Axler. *Linear Algebra Done Right*. 3. utg. Springer, 2015. ISBN: 978-3-319-11079-0. URL: <https://linear.axler.net>.
- [2] Endre Süli and David Mayers. *An Introduction to Numerical Analysis*. Cambridge University Press, 2003. ISBN: 9780521007948.
- [3] Franz Luef. *TMA4145 – Linear Methods*. Nov. 2019. URL: [https://wiki.math.ntnu.no/\\_media/tma4145/2019h/course-notes\\_2019-10pt-bm-final.pdf](https://wiki.math.ntnu.no/_media/tma4145/2019h/course-notes_2019-10pt-bm-final.pdf).
- [4] S. Friedberg, A. Insel og L. Spence. *Linear Algebra*. 4. utg. Pearson, 2014.