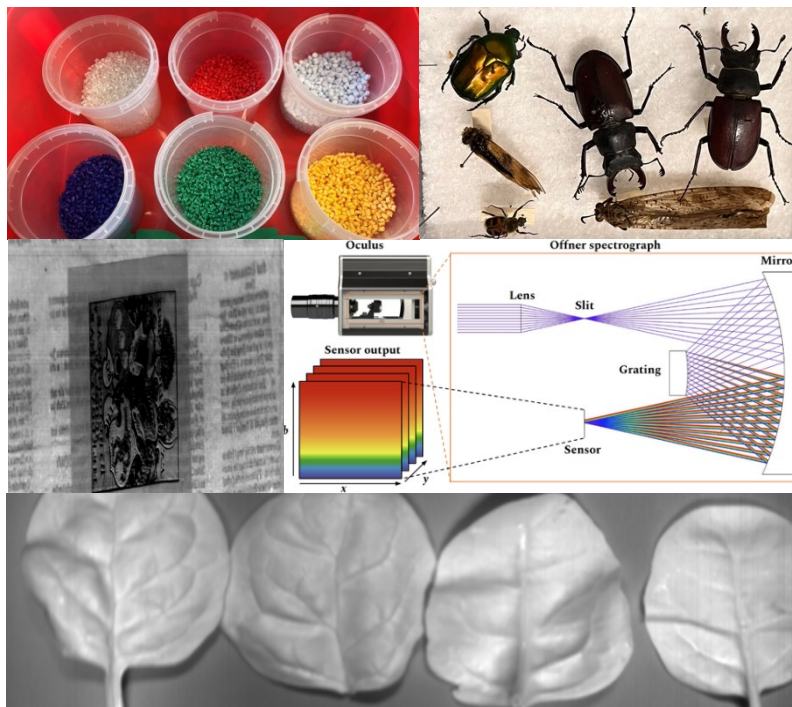




Praktisk Spektroskopi med Hyperspektrale Billeder



Indholdsfortegnelse

Introduktion	s. 3
Øjet, lys og farver	s. 4
Det hyperspektrale kamera	s. 6
Programmet "Hyperspectral Analyzer"	s. 9
Spinatbalde	s. 12
Kartofler	s. 13
Plastiktyper	s. 14
Biller	s. 16
Herlufsholm-samlingen og PCA	s. 17
Principal Component Analysis (PCA)	s. 18

Materialet er udarbejdet af Brian Skov Sørensen, SDU.
Morten Sielnik Andersen, SDU står bag programmet ”Hyperspectral Analyzer” og billederne. Mads Svanborg Peters fra Newtec har bidraget med billeder mens Mads Toudal Frandsen, Mikkel Theis Kristensen, Jakob Povl Holck, René Lyng Eriksen, Lilian Skytte, Peter Lund Boiesen, SDU og Bjarke Jørgensen, Newtec, alle har bidraget.

Link til program og data fås ved henvendelse til Brian på bsou@sdudk, eller Lilian på llsky@sdudk.

Udarbejdet med støtte fra Villum Fonden, februar 2026.

Link til program og data fås ved henvendelse til Brian på bso@sdu.dk, eller Lilian på lilsky@sdu.dk.

Udarbejdet med støtte fra Villum Fonden, februar 2026.

Introduktion

Ved at studere det lys, der reflekteres eller udsendes fra fysiske objekter, kan vi lære meget om deres egenskaber – for eksempel om en plantes sundhed, hvilke stoffer en plastikgenstand er lavet af, eller hvad der gemmer sig under overfladen på gamle bogomslag. Hyperspektrale kameraer gør det muligt at måle lysets intensitet ved hundredvis af forskellige bølgelængder – langt flere end det menneskelige øje eller almindelige kameraer kan opfatte.

Hyperspektral billedanalyse bruges i dag i mange forskellige sammenhænge. Teknologien kan fx:

- identificere sygdoms- eller tørketegn i planter (landbrug og fødevareproduktion),
- kortlægge mineraler, jordtyper og forurening (geologi og miljøovervågning),
- afsløre ændringer i vegetation eller overfladeforhold set fra satellitter (jordobservation og klimaforskning),
- skelne mellem materialer som plastik og metal (affaldssortering og industri).

Dette materiale giver dig mulighed for at arbejde med billeder taget med Newtecs hyperspektrale kamera [Buteo](#), som står på SDU i Odense.

Materialet er udviklet til fysikklubber i Region Syddanmark og til gymnasiets fysikundervisning. Det er opbygget, så elever kan arbejde selvstændigt med hele eller dele af forløbet.

Materialet består af tre dele:

1. Øjet, lys og farver

Du lærer, hvordan vi opfatter farver, og hvordan billedteknologien i et hyperspektralt kamera adskiller sig fra et almindeligt smartphone-kamera. Der er links til korte videoer, som uddyber de vigtigste begreber.

2. Dataanalyse med programmet *Hyperspectral Analyzer (Win)*

Du arbejder med ægte hyperspektrale billeder og analyserer spektre fra en række forskellige motiver. Du ser også, hvordan Principal Component Analysis (PCA) kan bruges til at fremhæve mønstre i billederne.

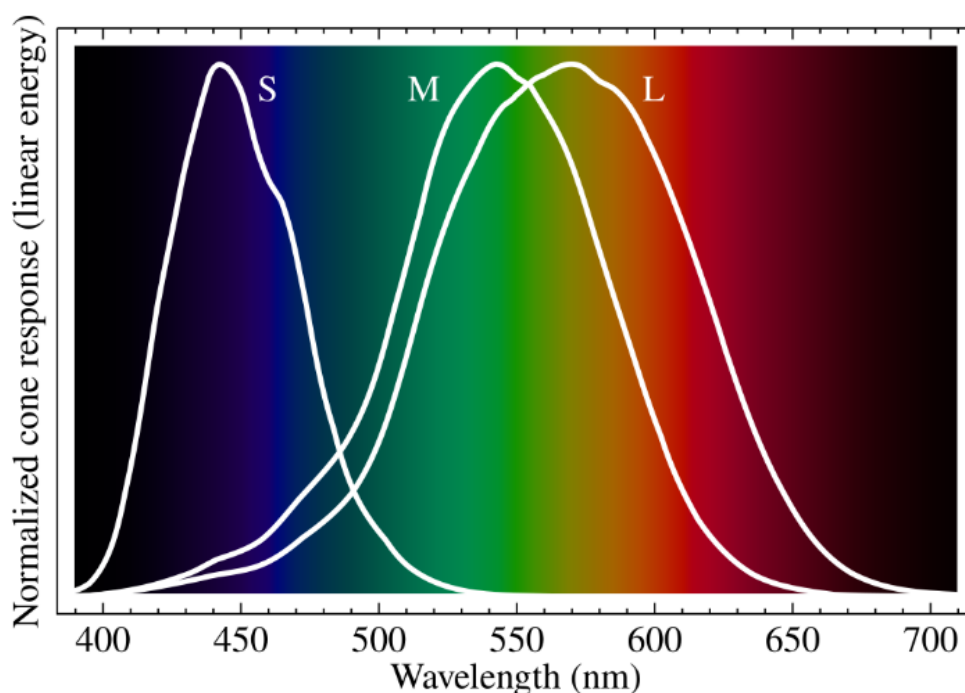
- Spinatblade
- Kartoffler
- Plastiktyper
- Biller
- Et gammelt bogomslag fra Herlufsholm-samlingen

3. Matematikken bag PCA (*Principal Component Analysis*)

Den sidste del indeholder et konkret regneeksempel på en pose avocadoer. Først forklares hvad PCA principielt gør ved data, hvorefter der er en mere teknisk regnegennemgang der involverer matricer. Sidste del er tænk som materiale til den matematikfaglige del af en SOP/SRP.

Øjet, lys og farver

Det menneskelige øje er en optisk sensor, der kan registrere [elektromagnetiske bølger](#) med bølgelængder mellem ca. 400 og 700 nm. Vores opfattelse af [farver](#) er en kompleks proces, men de såkaldte [tapceller](#) (eller tapstave) i øjets nethinde spiller en central rolle. Hos mennesket findes tre typer tapceller – S, M og L – som hver især er følsomme over for forskellige bølgelængder, som vist på fig. 1.



[Fig. 1. Normaliserede responsspektre af menneskets tapstave, S-, M- og L-typer.](#)

L-celler reagerer mest på lys omkring 570 nm, M-celler omkring 540 nm, og S-celler omkring 440 nm. Når disse tre typer stimuleres i forskellig grad, kan hjernen kombinere signalerne og danne en farveoplevelse. Det gør det muligt for os at skelne op mod 10 millioner forskellige farver.

Den måde øjet opfatter farver på, udnyttes i både skærmt teknologi og kameraer, hvor man arbejder med kun tre farvekanaler: rød, grøn og blå. Dette kaldes [RGB-modellen](#).

I RGB-modellen kan næsten alle synlige farver dannes ved at kombinere rødt, grønt og blåt lys med forskellig intensitet. På fig. 2 ses til venstre, hvordan de tre farver blandes, og til højre hvordan komponenterne R, G og B kombineres til et samlet billede.

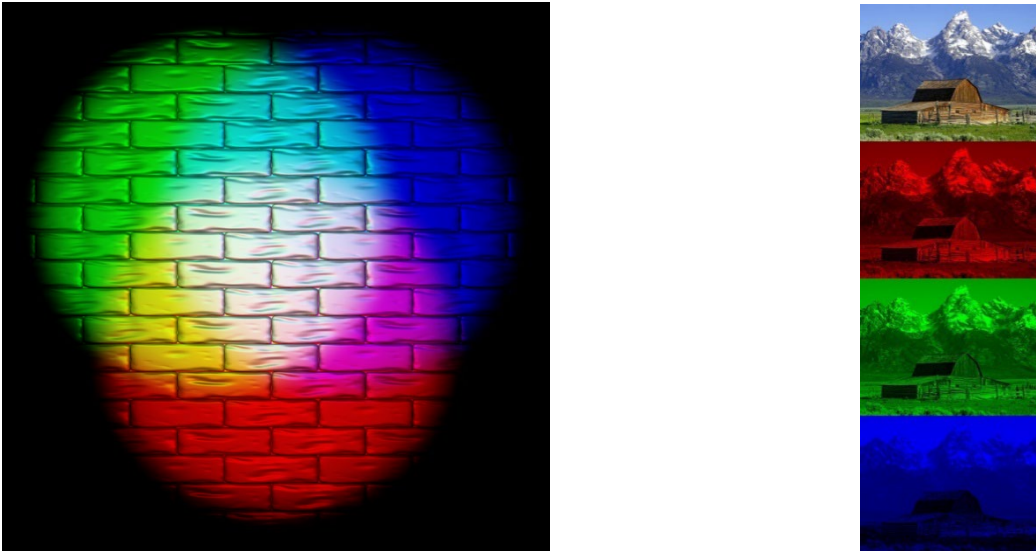


Fig. 2 Venstre: Additiv farveblanding af rød, grøn og blå. Bemærk at blandingen af alle tre bliver hvid. Højre: R, G, B komponenterne med det samlede billede øverst.

På digitale skærme, som fx din smartphone, består hver pixel af tre små lysdioder – én for hver farve: rød, grøn og blå. Sammen skaber de den farve, du ser.

I et digitalkamera sidder sensorer bag et mønster af farvefiltre (rød, grøn og blå), der sørger for, at hver pixel kun måler lys i én af de tre farver. På fig. 3 ses, hvordan hver farve registreres i sit eget pixelområde.

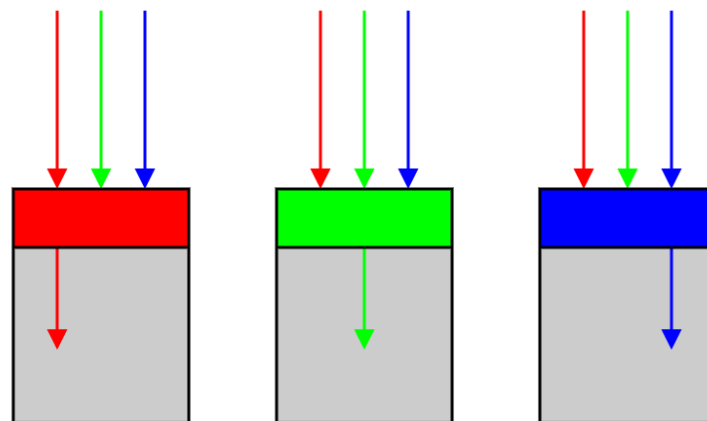


Fig. 3 Farvefiltre af forskellig farve, sikrer at sensoren til venstre kun opfanger intensiteten af rød, den næste af grøn osv.

Kort sagt: Når du tager et billede med et smartphone-kamera, bliver lysets intensitet i de tre farver registreret af kameraets sensorer og gemt som tre tal – én for rød, én for grøn og én for blå. Disse talværdier (R, G, B) omsættes til lys fra tre tilsvarende dioder i hver pixel på skærmen, og billedet vises.

I næste kapitel ser vi nærmere på dette sæt af tal (R,B,G).

Det hyperspektrale kamera

Du kan se en [video](#) om det hyperspektrale kamera, der står på SDU, og som er brugt til at tage de datacuber, du skal arbejde med i næste kapitel.

Når du tager et billede med et smartphone-kamera, registreres kun lysets intensitet ved tre forskellige bølgelængder – én for rød, én for grøn og én for blå. Det er ofte nok til at gengive virkeligheden i flotte farver, men ikke til at afsløre vigtige detaljer om materialer – detaljer, som også ofte ligger uden for det synlige område.

Hvis man vil undersøge materialer i dybden, kræver det målinger ved langt flere bølgelængder – og det er netop, hvad et hyperspektralt kamera kan.

Den centrale teknologi, der adskiller et hyperspektralt kamera fra et almindeligt kamera, bygger på **diffraktion og interferens**.

I fysiklokalet har du måske arbejdet med et forsøg, hvor man bestemmer bølgelængden af laserlys ved at sende det gennem et diffraktionsgitter og måle afstanden mellem diffraktionsordenerne (fig. 4). Hvis man sender hvidt lys gennem et diffraktionsgitter, bliver det splittet op i en regnbue af farver – én for hver orden. Du kan genopfriske, hvad der sker, når bølger interfererer, i en [video](#) om Youngs dobbeltspalte-eksperiment.

Nederst i fig. 4 ses, hvordan det hyperspektrale kamera er bygget op. Lyset fra objektet kommer ind fra venstre, samles af en linse og reflekteres ind på et [spejlgitter](#). I modsætning til det transmissionsgitter, man ofte bruger i undervisningen, reflekterer spejlgitteret lyset – men fysikken bag (diffraktion og interferens) er den samme.

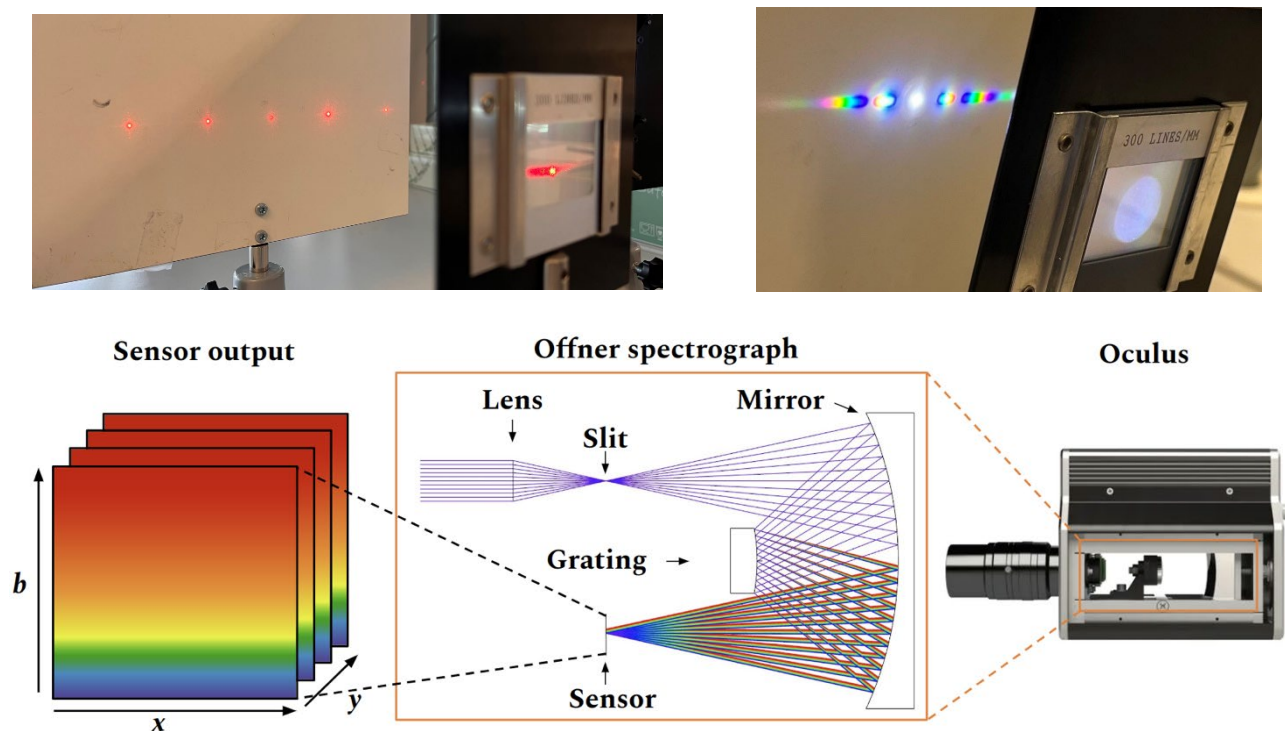


Fig. 4. Diffraction af laser- og hvidt lys i et optisk gitter, samt den optiske del af det hyperspektrale kamera nederst.

Lyset spaltes i sine bølgelængder og sendes videre ind på en række sensorer, som måler intensiteten af lyset ved hver bølgelængde. Resultatet gemmes i en såkaldt **datacube**, der på figuren er vist som et koordinatsystem hvor:

- **x- og y-aksen** svarer til det todimensionale billede (som et almindeligt foto)
- **b-aksen** angiver bølgelængden (eller spektralkanalen).

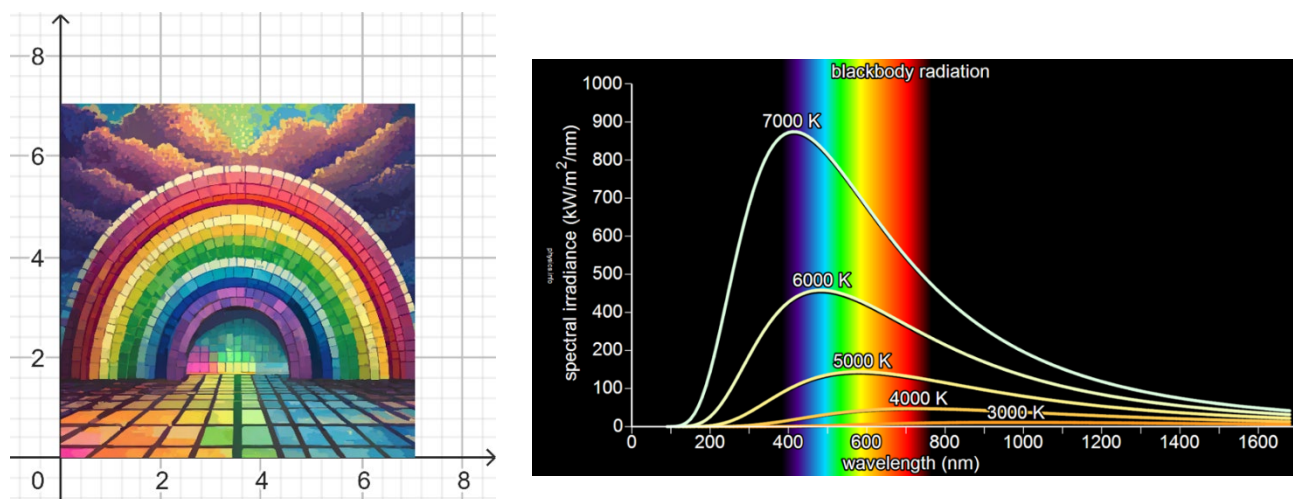


Fig. 5 Billede af en regnbue i et koordinatsystem og [spektre fra sortlegeme stråling ved forskellige temperaturer](#).

Det kan være svært at forestille sig, hvordan et billede med så mange oplysninger ser ud, så lad os sammenligne med et almindeligt RGB-billede fra en smartphone.

I et RGB-billede har hver pixel et koordinatsæt (x, y) og tre tal, der angiver intensiteten af rød (R), grøn (G) og blå (B). Et billede kan derfor beskrives som en række af 5 tal: (x, y, R, G, B).

F.eks. er pixlen i (x, y) = (4, 4) i fig. 5 beskrevet ved talsættet (4, 4, 22, 150, 63). Du kan selv eksperimentere med det i en [RGB color picker](#).

Tilbage til datacuben vist som et koordinatsystem på fig. 4., skal du forestille dig, at du ser billedet oppe fra. For et RGB-billede har **b**-aksen kun tre inddelinger (én for hver farvekanal). Men med det hyperspektrale kamera på SDU kan man måle intensiteten ved 900 forskellige bølgelængder i området fra 400 nm til 1700 nm.

Det svarer til, at **b**-aksen har 900 inddelinger. På grund af den høje opløsning i bølgelængderne, vises den som et kontinuert spektrum.

I fig. 5 ses fem forskellige [spektre fra sortlegemer](#) ved forskellige temperaturer. Det fremgår, at spektrets top forskydes mod kortere bølgelængder, når temperaturen stiger. Solen kan fx betragtes som et sortlegeme med en temperatur på ca. 5800 K. Det synlige område (400–700 nm) er vist som en regnbue.

Det hyperspektrale kamera kan altså ikke blot se meget finere spektrale detaljer end et almindeligt kamera - det kan også "se" ud over det synlige område, ind i det såkaldte nær-infrarøde område (NIRS). I dette område kan visse bølgelængder [absorberes af molekyler og få dem til at vibrere](#).

Dette gør det muligt at identificere specifikke molekyler og materialer – en teknik du selv skal arbejde med i de kommende øvelser.

I næste kapitel får du en vejledning i brugen af programmet *Hyperspectral Analyzer*. Med dette program kan du selv ”skære” skiver ud af datacube’erne og analysere spektre fra billeder af forskellige materialer.

Programmet "Hyperspectral Analyzer"

I det forrige kapitel lærte du, hvordan et hyperspektralt kamera gemmer information i en såkaldt *datacube* – en tredimensionel datastruktur, hvor hver pixel ikke kun indeholder én farve, men et helt spektrum. Datacuben kan altså beskrives som en stak billeder taget ved forskellige bølgelængder – fra 400 nm til 1700 nm – hvor hvert lag svarer til en bestemt bølgelængde.

Nu skal du arbejde med datacuber taget med Newtecs hyperspektrale kamerasystem [Buteo](#), som dækker både det synlige og nær-infrarøde område. Til det formål skal du bruge programmet **Hyperspectral Analyzer**, som gør det muligt at visualisere, undersøge og analysere indholdet af disse datacuber.

I dette kapitel lærer du, hvordan du åbner en datacube, ser billederne igennem, udtrækker spektre, og anvender kalibreringsdata. Du vil også blive introduceret til en metode kaldet *Principal Component Analysis* (PCA), som kan bruges til at fremhæve mønstre og forskelle i billederne, der ellers er svære at se.

Installér først programmet "Hyperspectral Analyzer" med filen: *setup_Hyperspectral.exe*

Øvelse 1 Brug af programmet og kalibrering.

I denne øvelse skal du arbejde med et kalibreringsbillede og lære at bruge programmet *Hyperspectral Analyzer*. Du lærer at åbne en datacube, se billedet i forskellige bølgelængder, udtrække spektre og forstå, hvorfor kalibrering er vigtig.

Loade fil:

- *File* → *Load Cube* → kalibrering.pam
- **Yes** – du vil gerne loade en kalibreringscube
- Vælg igen kalibrering.pam

Se datacuben:

- *View Cube*

Pop-up vindue af datacuben kommer til syne. På billedet ser du *xy*-planet af datacuben. Det er et billede af en kalibreringsplade til kalibrering af det hyperspektrale kamera. Det består af sorte og hvide områder sammen med flere rækker af dioder (LEDs).

Med skyderen nederst i billedet kan du rulle dig gennem de 900 kanaler (*b*-aksen), som svarer til bølgelængderne i intervallet 400 nm til 1700nm.

1. Hvorfor blinker dioderne, når du bevæger skyderen gennem bølgelængderne – og hvad fortæller det om dem?

Udtræk spektre:

- *Extract Spectrum.*

Et nyt vindue med **xy**-planet åbnes.

Brug musen til at trække en firkant hen over:

1. Et hvidt område
2. Et sort område
3. En af dioderne

Luk pop-up vinduet

Du vil nu se tre spektre. y-aksen angiver intensiteten og x-aksen angiver kanalnummeret (**b**-aksen); dette kan du ændre til bølgelængder.

2. Svarer spektrene til det du vil forvente for hvidt, sort og en diode?

Det er muligt at slette spektrene igen (Clear Plot) eller eksportere dem i csv-format (Save Spectra), som kan importeres i Logger Pro til nærmere undersøgelse.

3. Undersøg spektrene af alle 20 lysdioder på én gang, og importer dem til Logger Pro.

Over tid kan både lysforhold og sensorernes følsomhed ændre sig, blandt andet på grund af temperaturændringer. Derfor er kalibrering vigtig – den sikrer, at de spektre, man måler på, er pålidelige, sammenlignelige og troværdige.

Lysdioder ([LED](#)) udsender typisk lys i en enkelt farve (monokromatisk), bestemt af halvledermaterialet de er lavet af og dets energi gab.

4. Betragt spektrene af de 20 lysdioder. Hvilke bølgelængdeområder mener du kan kalibreres med størst nøjagtighed og hvorfor?

Bemærk, at det spektrum du ser, er et gennemsnit af intensiteterne for alle de pixels, der ligger inden for det område, du har markeret.

Principal Component Analysis (PCA)

- **Faneblad:** *Principal Component Analysis* → *Compute PCA*

Et pop-up-vindue uden indhold kommer til syne.

5. Arrangér så du kan se de to vinduer samtidig. Undersøg hvad der sker når der ændres på de 4 vægte og kontrasten.

(Det er muligt at *slice data* - indtast et interval separeret med ":" i området 1296×1000×900)

Principal Component Analysis (PCA) komprimerer og fremhæver detaljer i et billede.

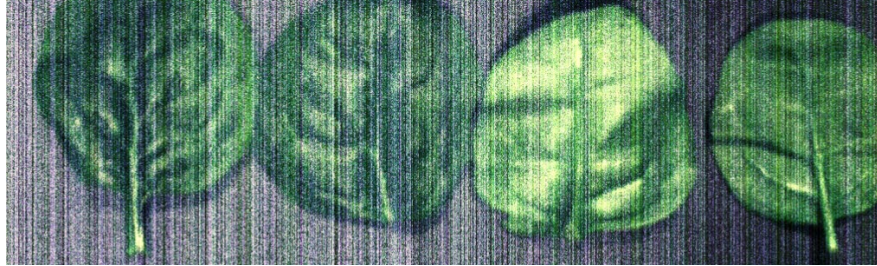
Et billede taget med det hyperspektrale kamera består af en datacube med meget store mængder data. I et almindeligt RGB-billede består hver pixel af fem informationer: (x, y, R, G, B) – altså et punkt i et 5-dimensionelt koordinatsystem. I en datacube fra det hyperspektrale kamera har hver pixel i stedet intensitetsværdier for 900 forskellige bølgelængder, plus positionen (x, y). Det svarer til et punkt i et 902-dimensionelt rum. PCA er en matematisk metode til at komprimere og forenkle denne store mængde data.

PCA finder de retninger i data, hvor der er størst variation mellem pixels – i vores tilfælde er det variation i intensitetsværdier ved forskellige bølgelængder. Når man bruger disse retninger til at danne nye billeder, fremhæves forskelle i materialer og strukturer, som ellers kan være svære at se i de originale billeder fra hver bølgelængde. Det kan fx gøre kontraster tydeligere og gøre det lettere at skelne objekter. I programmet *Hyperspectral Analyzer* er det disse retninger, du ser som PC1, PC2, PC3 og PC4. Når du trykker på *Compute PCA*, bliver billedet rekonstrueret ud fra de fire komponenter, i stedet for de oprindelige 900 kanaler. Det betyder, at hver pixel nu er beskrevet med fire tal i stedet for 900 – men på en måde, hvor man bevarer den vigtigste information i data.

Når du ændrer på de fire vægte i programmet, justerer du, hvor meget betydning hver af de fire PCA-komponenter skal have i det samlede billede. Du oplevede selv hvordan nogle strukturer blev tydeligere, mens andre forsvandt. PCA fremhæver ikke mønstre i sig selv, men organiserer data på en måde, så det bliver nemmere at se forskelle og finde strukturer i billedet.

PCA er en matematisk metode, som tager udgangspunkt i datasættets varians og bruger lineær algebra til at finde de retninger, hvor data varierer mest. Efter den sidste øvelse om Herlufsholm-samlingen, er der en matematisk gennemgang af metoden – den kræver at du også lærer lidt om vektorer og matriser.

Spinatblade



Vand og klorofyl absorberer lys i det nær-infrarøde område. Det ses som minima i spektret fra billeder taget med det hyperspektrale kamera, da kameraet måler intensiteten af det lys, der reflekteres fra spinatbladene i hele bølgelængdeområdet fra 400 nm til 1700 nm.

1. Træk ét eller flere spektre ud af datacuben for spinatblade.
2. Det nær-infrarøde område starter ved ca. 700 nm og dækker alle bølgelængder længere end det synlige lys. I dette område absorberer vandmolekyler lys, fordi deres bindinger begynder at vibrere. Studér spektret for absorption af bølgelængder i det nær-infrarøde område og undersøg hvilke [typer vibrationer af vandmolekylet](#) man kan se.
3. Ved hvilken bølgelængde ser du absorption af fotoner med en energi på 1,86 eV?

Dette svarer til det røde absorptionsmaksimum for klorofyl-a, som spiller en central rolle i fotosyntesen.

Jordobservationer fra satellitter udstyret med hyperspektrale kameraer gør det muligt at studere vegetation på jordens overflade og giver information om planterens helbred og produktivitet.

Du kan læse mere om jordobservationer på [NASAs hjemmeside](#), selv gå på opdagelse i data fra satellitter i "[The Copernicus Data Space Ecosystem Browser](#)" eller finde mere om vegetationsindeks i "[Fysik i Rummet](#)" fra serien *Fysik i der 21. århundrede*, som de fleste gymnasier har liggende, spørg din lærer.

Kartofler

En kartoffel udvikles i mørke, hvor cellerne nær skrællen primært fungerer som lagre for stivelse, de såkaldte *amyloplaster*. Når kartofflen udsættes for lys, fortæller kartofflens lyssensorer som *phytochromer* og *kryptochromer*, at *amyloplaster* skal omdannes til *kloroplaster*. Derved begynder der at opbygges *klorofyl* i de yderste cellelag af kartofflen. Efter noget tid ses det som en grønlig tone i skrællen.

Samtidig med dannelsen af *klorofyl*, skruer kartofflen op for sine naturlige forsvarsstoffer, de såkaldte *glykoalkaloide* – det er især *solanin* og *chaconin*. De laves primært i og lige under skrællen og koncentrationen stiger ved lys, sårning og andre former for stress. I høje mængder smager de bittert og kan give ubehag fra mave og tarm – derfor frarådes det at spise grønne kartofler.

Når man ser tegn på *klorofyl* i overfladen af en kartoffel, er det et tidligt fingerpeg om, at den har været udsat for lys længe nok til, at forsvarssystemerne også kan være i gang og have dannet *glykoalkaloide*. De to stoffer dannes altså uafhængigt af hinanden, men det er kun *klorofyl* man kan "se".

De første ændringer er målbare længe før farven bliver tydelig for det menneskelige øje; *klorofyl* absorberer især rødt lys omkring 640–670 nm, og reflektansen stiger brat lige over 700 nm – den såkaldte "røde kant". Det giver en meget karakteristisk spektral signatur, som et hyperspektralt kamera kan opfange tidligt.

Billederne er taget med Newtecs hyperspektrale kamera [Buteo](#) – de udvikler og sælger også hyperspektrale kameraer, der kan bruges til optisk sortering og detektering af *klorofyl*indhold.



1. Load *Kartofler.pam*-cuben sammen med *Kalibrering_kartofler.pam*.
2. Træk spektre ud for kartofflen i nederste højre hjørne, ved at markere flere små områder fra højre mod venstre.
3. Kan du genfinde de spektrale signaturer for *klorofyl* i dine spectra, som beskrevet i teksten?
4. Studér spektrene for absorption af bølgelængder i det nær-infrarøde område og undersøg hvilke [vibrationstyper i vandmolekylet](#) man kan se.

Den markante W-form i reflektansen omkring 900 – 1200 nm skyldes vandets absorptionsbånd (~970 og ~1200 nm) og tyder på højt vandindhold; det bekræftes ofte af et dybt fald ved ~1450 nm.

Plastiktyper



”Plastik.pam” er en datacube hvorfra man kan trække spektre ud for de 6 plastiktyper vist på RGB-billederne ovenfor.

Forkortelse	Typebetegnelse	Farve*	Typiske produkter
1. PET	Polyethylenterephthalat	trans.**	Sodavandsflasker
2. HDPE	Høj densitet polyethylen	rød	Spande og flasker
3. PVC	Polyvinylchlorid	hvid	Vinduer, tagrender, medicinske produkter
4. LDPE	Lav densitet polyethylen	blå	Poser f.eks. bæreposer
5. PP	Polypropylen	grøn	Hospitalsudstyr, emballage
6. PS	Polystyren	gul	Computerudstyr

Fra skoleboksen på plast.dk

Meget kort fortalt, består de seks plastiktyper af en kulstofkæde med C-H bindinger og en tilhørende funktionel gruppe. Det er typen af den funktionelle gruppe, der giver de forskellige plastiktyper deres karakteristika, som vi kender fra hverdagen. I bølglængdeområdet fra 700 nm til 1700 nm er de dominerende absorptioner knyttet til vibrationer i C-H bindingerne. De funktionelle grupper har mindre betydning, men kan påvirke graden af absorption samt placering i spektret.

Hyperspektral billedanalyse anvendes til at sortere plastiktyper ved [affaldssortering](#). Her skal du selv afprøve om det er muligt.

Forberedelse:

1. Load Plastik.pam-cuben sammen med Kalibrering.pam. Nummereringen, som er spejlvendt, svarer til tallene i boksen ovenfor. Rul gennem de 900 kanaler med skyderen nederst i billedet, og bemærk hvad der sker.
2. Træk alle spektre ud af datacube på én gang. Gem dem som CSV-fil, og importer filen i Logger Pro - så vises alle i samme grafvindue (dobbeltklik på y-aksen).
3. Navngiv hver type plastik og ændr farven, så den passer til plastikkens farve (dobbeltklik på grafen)

Undersøgelse af spektre:

4. Den synlige del af spektret ligger i bølgelængdeområdet fra 400 - 700 nm. Er reflektionsspektret i overensstemmelse med de plastikfarver, du kan se med det blotte øje?
5. I bølgelængdeområdet 1400 - 1700 nm er der to markante minima i alle spektrene. Disse skyldes C-H vibrationerne. Der er to plasttyper, der skiller sig ud fra de andre i dette område, - hvilke mener du, det er, og hvorfor?
6. For at kunne skelne disse to plasttyper fra hinanden spektralt, skal du finde et andet område i intervallet 400 – 700nm hvor de kan adskilles. Hvad er dit bud?
7. Nu kan to plasttyper spektralt kendes fra hinanden. I praksis trænes en algoritme (Machine learning – spørg din chatbot) til at genkende de forskellige typer af plastik. Vurdér, om det er muligt at adskille resten af plastiktyperne ved at se efter tydelige forskelle i spektrene.

Biller

På billedet nedenfor ses en dronebille sammen med et par hjortebiller og et par andre smådyr. Dronebillerne changerer i forskellige farver, mens hjortebillerne er ret matte. Refleksionen af de skiftende farver på dronebillerne skyldes [iridiserende](#). Iridisering er et fænomen, hvor fysiske mikro- og nanostrukturer i billens eksoskelet skaber lysinterferens. Effekten kan beskrives med den samme fysik som et **diffraktionsgitter** i fysiklokalet eller i et hyperspektralkamera. Billens farve skyldes derfor primært **strukturel farvedannelse** og ikke (kun) farvepigmenter: Nogle bølgelængder får konstruktiv interferens ved refleksion fra overfladen, mens andre ikke gør.



Fænomenet ses hos mange dyr, som f.eks. kolibrier, påfugle, sommerfugle og guldsmede. Blæksprutter og makrel er også iridiserende, ligesom visse blanke flader på kød kan være det.

1. Undersøg spektrene for de to forskellige biller og beskriv forskelle og ligheder. Bemærk, at billerne har deres egen kalibreringsfil: *Kalibrering_biller.pam*. Lav de markerede områder så små som muligt, når du trækker spektrene ud. Husk, LoggerPro giver dig flere muligheder for at gå på opdagelse i spektrene.
2. Hvilke bølgelængder i det synlige område (400–700 nm) bidrager til de farver, du ser på dronebillerne?
3. Diskutér billernes evne til at kamuflere sig – i hvilke omgivelser vil de kunne gemme sig bedst? (Tænk i synligt lys, 400–700 nm.)
4. Varmeafskærmning mod solens påvirkning er vigtig i bølgelængdeområdet 700–1700 nm, fordi en stor del af solens strålingsenergi ligger her, og vand absorberer stærkt i dette område. Er der tydelige forskelle i billernes spektre i 700–1700 nm? Siger billernes synlige farver (400–700 nm) noget om dette?

Du kan læse en videnskabelig artikel om emnet [her](#), eller se den fremragende populærvidenskabelige video: [Iridescence: Nature's Most Beautiful Physics Trick \(17 min.\)](#), hvis du vil dykke ned i hvor, hvordan og hvorfor iridisens optræder i naturen.

Herlufsholm-samlingen og PCA

Bogbindere i det 16. og 17. århundrede havde for vane at tage tidligere udgivelser, klippe dem op, og genbruge stykkerne som omslag til nye bøger. Det har medført at dele af disse tidligere udgivelser er velbevarede, men skjulte, på grund af den lufttætte indkapsling de har som bogbind. Der findes forskellige metoder til at genskabe sådanne tekster uden at ødelægge den bog de nu er bogbind på - hyperspektrale billeder er én af dem. Man kan læse nærmere om fund i Herlufsholm-samlingen i Jakob Povl Holcks artikel: "[Nye teknologier og Herlufsholm-samlingen](#)."

Orientér dig i kapitlet om brugen af programmet: "Hyperspectral Analyzer", hvis du ikke allerede har gjort det.

1. Load datacuben af et omslag fra Herlufsholm-samlingen sammen med en kalibreringsfil, og rul gennem de 900 kanaler, ved at bruge skyderen nederst i *View Cube*.
2. Hvad observerer du, når ruller gennem kanalerne? Kan du konkludere, hvordan bogomslagets refleksion afhænger af bølgelængden?
3. Udfør PCA på datacuben og brug vægtene (PC1,..) til at se, hvor tydeligt et billede af det bagvedlæggende dokument du kan skabe.

Betydningen af PCA ses i forskellen mellem de billeder der fremkom, da du rullede gennem kanalerne i opgave 1, og de billeder der fremkom efter PCA i opgave 3. Denne forskel er kort beskrevet i første del af næste afsnit, hvor matematikken bag PCA anvendes på en pose avocadoer.

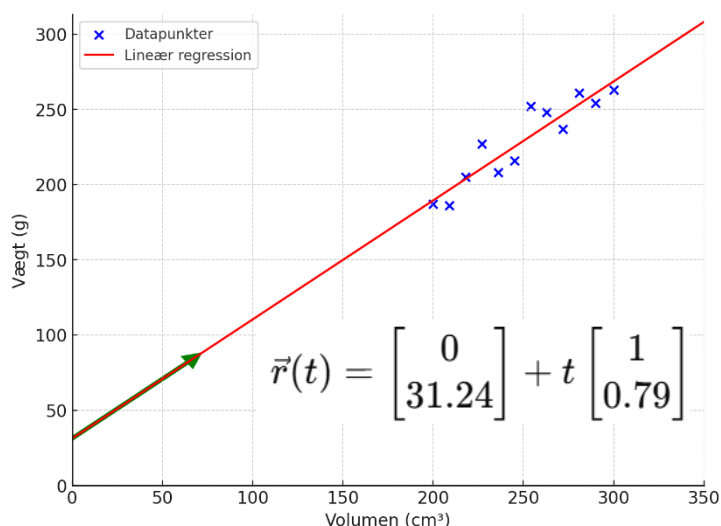
Principal Component Analysis (PCA)

Principal Component Analysis (PCA) har flere fordele i billedanalyse, især når det kommer til datakomprimering og at fremhæve detaljer i billedet. Det kan også anvendes til sortering. Datacubeerne i dette materiale kræver et 902-dimensionelt koordinatsystem (eller vektorrum) for at blive beskrevet, og fylder mere end 1 GB.

PCA er en metode, hvor antallet af dimensioner - og dermed datamængden, reduceres kraftigt, samtidig med man bevarer så meget information som muligt, fra det hyperspektrale billede.

Metoden kan illustreres meget kort med velkendt matematik, inden vi går videre... Nedenfor er vist et datasæt bestående af avocadoers volumen og vægt.

Volumen (cm ³)	Vægt (g)
200	204
209	194
218	188
227	198
236	191
245	232
254	219
263	228
272	229
281	263
290	249
300	261



De to variable for avocadoer er indtegnet i et koordinatsystem sammen med en ret linje, der fremkommer ved lineær regression. Retningsvektoren (forstørret for synets skyld) og parameterfremstillingen for den rette linje, er også vist.

Udgangspunktet, hvor hver avocado er beskrevet af et koordinatsæt (volumen, vægt), kan nu erstattes af en enkelt værdi for parameteren t , som tilnærmelsesvis beskriver hvor på den røde linje avocadoen er. På den måde er data nu reduceret fra 2 dimensioner (volumen, vægt) til 1 dimension (t).

Matematisk er retningen af den røde linje bestemt ved "mindste kvadraters metode", og man kan se at den ligger i en "fornuftig retning" når man vil beskrive avocadoer. Den ligger så at sige i den retning, hvor der "sker noget" i data. Men bemærk, at der umiddelbart går noget tabt ved dette skift fra (volumen, vægt) til t , når vi skal beskrive en avocado. I forsimplingen går man fra at referere til fysisk målbare størrelser, til tallet t . Sidst i dette afsnit vender vi tilbage til t , og ser på hvordan det kan anvendes til sortering.

I alle spektrene du har arbejdet med viste hver pixel, intensiteten af den pågældende bølgelængde, defineret med skyderen. Anvender man PCA-analyse bliver intensiteten af hver pixel et vægtet gennemsnit af intensiteterne for alle bølgelængderne i det 902-dimensionale koordinatsystem. De bølgelængder der dominerer i det vægtede gennemsnit, er bestemt af de retninger, hvor der er størst *varians*.

Det er ikke ”mindste kvadraters metode” man anvender, i PCA, men det minder om. Lineær regression minimerer fejl i én variabel (y-retningen), mens PCA finder retningen med størst variation i *alle* variable.

I det følgende er der vist et konkret eksempel med beregninger, der viser PCA udført på det 2-dimensionale datasæt af avocadoers volumen og vægt. Ved beregningerne bliver der anvendt begreber og matematik som sandsynligvis er nyt for dig. Mellemregningerne er nogle steder også sprunget over. Det er alt sammen med vilje – det giver dig mulighed for at bidrage med din egen viden, i forbindelse med en SOP eller SRP. Der bliver anvendt matricer og få matrixoperationer, som du kan læse mere om [her](#).

Data opstilles i en matrix, som her kaldes A:

Volumen (cm³)	Vægt (g)
200	204
209	194
218	188
227	198
236	191
245	232
254	219
263	228
272	229
281	263
290	249
300	261

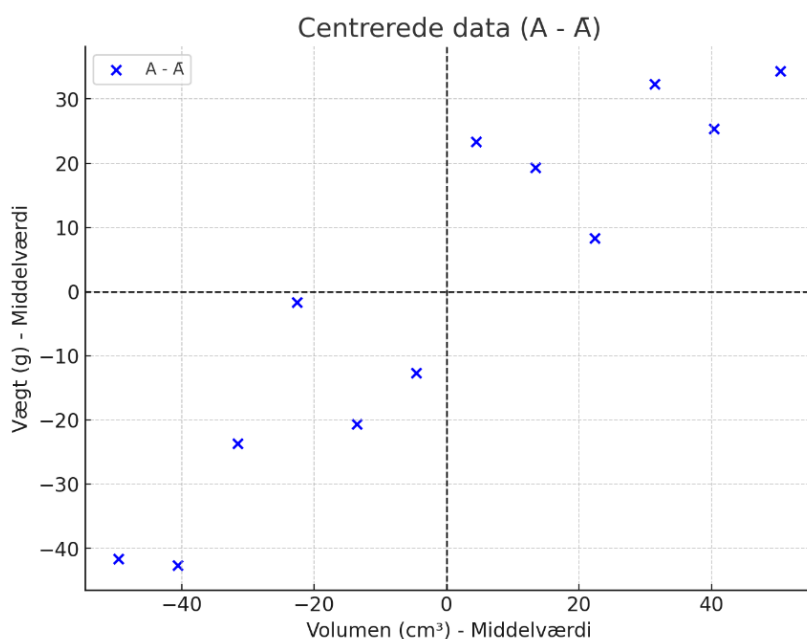
$$A = \begin{bmatrix} 200 & 187 \\ 209 & 186 \\ 218 & 205 \\ 227 & 227 \\ 236 & 208 \\ 245 & 216 \\ 254 & 252 \\ 263 & 248 \\ 272 & 237 \\ 281 & 261 \\ 290 & 254 \\ 300 & 263 \end{bmatrix}$$

Middelværdien af hver søjle kan beregnes og giver: $\bar{A} = [249.58 \quad 228.67]$

Den beregnede middelværdi trækkes fra A:

$$A - \bar{A} = \begin{bmatrix} -49.58 & -41.67 \\ -40.58 & -42.67 \\ -31.58 & -23.67 \\ -22.58 & -1.67 \\ -13.58 & -20.67 \\ -4.58 & -12.67 \\ 4.42 & 23.33 \\ 13.42 & 19.33 \\ 22.42 & 8.33 \\ 31.42 & 32.33 \\ 40.42 & 25.33 \\ 50.42 & 34.33 \end{bmatrix}$$

Data er nu centreret og man kan få en idé om effekten på grafen nedenfor. Akserne viser afvigelsen fra middelværdien for hver variabel. Overvej hvordan punkterne ville ligge, hvis alle avocadoerne vejede det samme.



Nu hvor retningerne for størst *varians* skal bestemmes, bliver der introduceret nye begreber. Nogle bliver forklaret her, andre skal du selv til at finde.

Retningerne for størst *varians* findes ved at bestemme egenvektorerne og de tilhørende egenværdier for kovariansmatricen af den centrerede matrix ($A - \bar{A}$).

Hvordan, vises på næste side.

Formlen for kovariansmatricen Σ er givet ved:

$$\Sigma = \frac{1}{n-1}(A - \bar{A})^T(A - \bar{A})$$

hvor $n = 12$ er antallet af observationer. T angiver transponering.

Med

$$(A - \bar{A})^T = \begin{bmatrix} -49.58 & -40.58 & -31.58 & -22.58 & -13.58 & -4.58 & 4.42 & 13.42 & 22.42 & 31.42 & 40.42 & 50.42 \\ -41.67 & -42.67 & -23.67 & -1.67 & -20.67 & -12.67 & 23.33 & 19.33 & 8.33 & 32.33 & 25.33 & 34.33 \end{bmatrix}$$

og brug af *matricemultiplikation* bliver kovariansmatricen: $\Sigma = \begin{bmatrix} 1062.08 & 840.12 \\ 840.12 & 778.24 \end{bmatrix}$

Kovariansmatricen viser, at variansen af volumen er 1062,08, variansen af vægt er 778,24 og kovariansen mellem volumen og vægt er 840,12. Den positive kovarians bekræfter, at når volumen stiger, stiger vægten typisk også.

Nu skal egenværdiligningen for kovariansmatricen opstilles og løses. En vektor v kaldes en egenvektor med egenværdien λ til matricen Σ , hvis den opfylder følgende ligning:

$$(*) \quad \Sigma v = \lambda v \Leftrightarrow \Sigma \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \Leftrightarrow \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \lambda \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

Her er σ_{11} variansen for volumen og σ_{22} variansen for vægt, mens $\sigma_{12} = \sigma_{21}$ er kovariansen mellem dem.

Egenværdierne λ_1, λ_2 findes ved at sætte det *karakteristiske polynomium* lig nul og løse ligningen på følgende måde:

$$\begin{aligned} \det(\Sigma - \lambda I) = 0 & \Leftrightarrow \det \begin{bmatrix} 1062.08 - \lambda & 840.12 \\ 840.12 & 778.24 - \lambda \end{bmatrix} = 0 \\ & \Leftrightarrow \lambda^2 - 1840.33\lambda + 120754.66 = 0 \end{aligned}$$

Løsningerne til denne andengradslikning er $\lambda_1 = 68,14$ og $\lambda_2 = 1772,19$. To forskellige egenværdier fortæller at der findes to egenvektorer. De to egenvektorer findes ved at indsætte én egenværdi i egenværdiligningen ad gangen. Den løses på følgende måde for $\lambda_2 = 1772,19$ ved at omskrive (*):

$$(\Sigma - \lambda_2 I)v = 0 \quad \Leftrightarrow \quad \begin{bmatrix} 1062.08 - 1772.19 & 840.12 \\ 840.12 & 778.24 - 1772.19 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

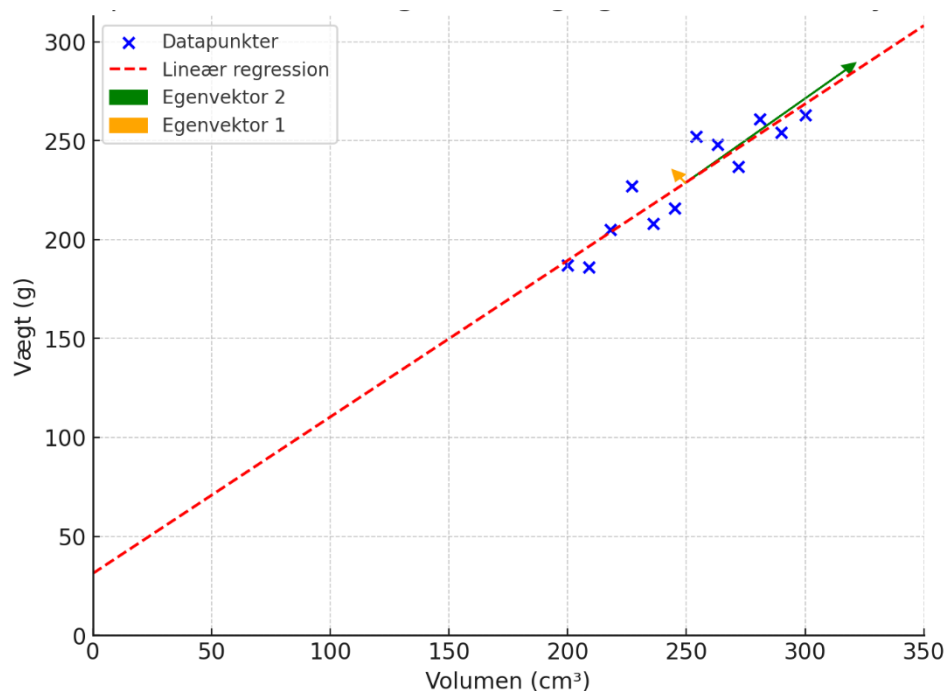
Det resulterer i to ligninger med to ubekendte, som kan løses.

$$\begin{aligned} -710.11v_1 + 840.12v_2 &= 0 \\ 840.12v_1 - 993.95v_2 &= 0 \end{aligned} \quad \Leftrightarrow \quad \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0.764 \\ 0.646 \end{bmatrix}$$

Dette er den normaliserede egenvektor for den største egenværdi. Den peger i den retning, hvor data varierer mest. Dette kaldes den primære principalkomponent og er betegnet "PC1 Weight" i "Hyperspectral Analyzer" programmet.

Beregning af den anden egenvektor med egenværdien λ_1 , overlades til læseren.

Herunder er de to egenvektorer indtegnet sammen med datapunkterne og den bedste rette linje. Egenvektorerne er skaleret i forhold til deres egenværdier. Bemærk at egenvektor 2 og den rette linje ikke er parallelle – den røde linje kommer fra lineær regression.



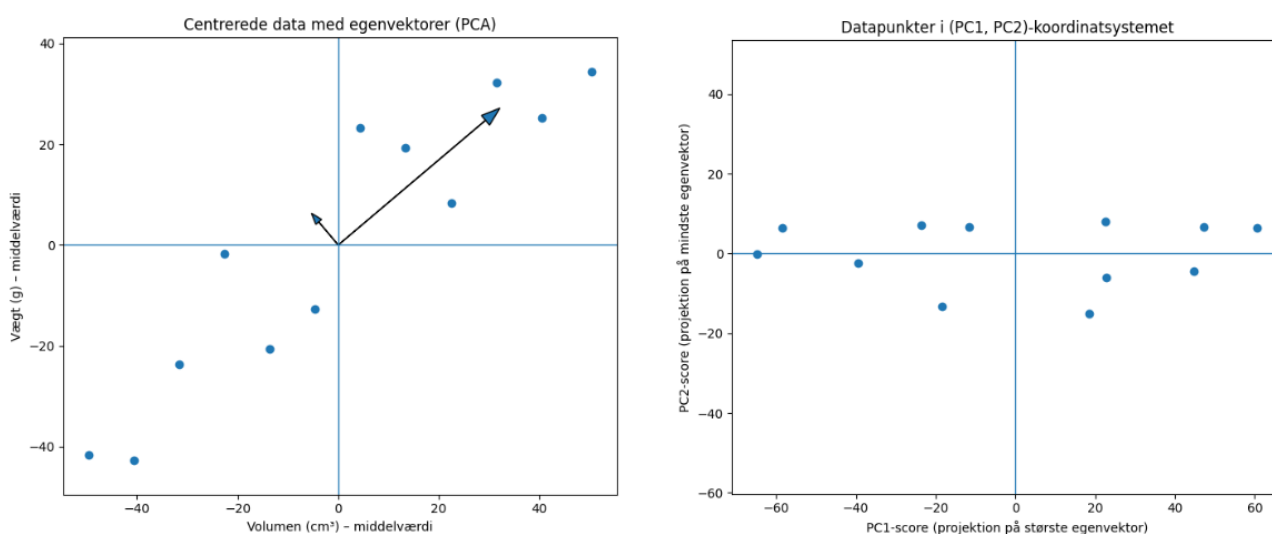
For at opsummere, har vi med PCA fundet to egenvektorer med tilhørende engenværdier.

De to egenvektorer udgør et nyt koordinatsystem, hvis akser peger i de retninger, hvor data varierer mest. Da den ene engenværdi er meget større end den anden, kan data med god tilnærmelse beskrives i én dimension. Vi er dermed tilbage ved udgangspunktet: en parameterfremstilling. Data kan nu med rimelig nøjagtighed repræsenteres ved én variabel t , hvor den dominerende egenvektor (PC1) fungerer som retningsvektor.

Afslutningsvis skal vi se på, hvordan dette kan anvendes til sortering.

I en parameterfremstilling kan man selv vælge sit nulpunkt. I denne sammenhæng giver det mening at vælge nulpunktet ved en "normal" størrelse avocado, som man selv definerer. Tallet t angiver da, hvor langt en given avocado ligger fra denne "normal"-avocado målt langs PC1-retningen forstået som stor og tung ($t > 0$) eller lille og let ($t < 0$).

Tallet t kan beregnes ved at bestemme avocadoernes *principal scores* (PC-scorer). På figuren nedenfor til venstre, er de centrerede data vist igen, sammen med de to principale retninger givet ved egenvektorerne, nu kaldet PC1 og PC2. De to PC-scorer for én avocado fås ved at projicere vektoren fra origo til datapunktet ned på henholdsvis PC1 og PC2. Matematisk beregnes scorene ved matrixproduktet mellem matricen af centrerede data og matricen bestående af egenvektorerne; herved fås for hvert datapunkt en PC1-score og en PC2-score. Resultatet ses herunder til højre.



Det fremgår af grafen til højre, at variationen i PC1-retningen er langt større end variationen i PC2-retningen. PC1-scoren udtrykker derfor i høj grad avocadoens samlede "størrelse" (stor og tung versus lille og let) og kan anvendes direkte til sortering efter størrelse. PC2-scoren varierer mindre og beskriver afvigelsen fra den overordnede størrelse-sammenhæng: om en avocado er relativt tung eller let i forhold til sin størrelse. Dette kan tolkes som et mål for "fylde" eller densitet, eller måske på forholdet avocadokød/avocadosten, og kan bruges til kvalitetskontrol eller til at udvælge avocadoer til et bestemt formål.

Med det hyperspektrale kamera arbejder man ikke med volumen og vægt, men med spektra. Her svarer *variablerne* til bølgelængderne (dvs. hver søjle/pixel i datasættet repræsenterer en bestemt bølgelængde), og *datapunkterne* svarer til de enkelte objekter eller prøver. Hvor vi her havde 12

avocadoer, kunne det i stedet være blade, kartofler eller forskellige typer plastik. PCA kan da bruges på samme måde: til at finde de vigtigste variationsretninger i spektrene, reducere dimensionaliteten og udtrække scorer, som kan anvendes til sortering og klassifikation af materialer eller til at opdage afvigere.

Hvor stor en andel af den samlede variation i data, der bliver ”fanget” af en given principal komponent, er givet ved:

$$\text{Forklaret varians (\%)} = \frac{\text{egenværdi for PC}}{\text{summen af alle egenværdier}} \times 100\%$$

Mere om PCA kan findes [her](#) og [her](#).